NURail Project
NURail2012-MIT-RO1

The final report for NURail project: **NURail2012-MIT-R01** consists of three distinct documents. The first 233 pages are titled “The Impact of Amtrack Performance in the Northeast Corridor” By Tolulope A. Ogunbekun. The second 174 pages are a report titled “Capacity Challenge on the California High-Speed Rail Shared Corridors: How Local Decisions Have Statewide Impacts” By Samuel J. Levy and the last 206 pages consists of “Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems” by Maite (Maria Teresa) Pena-Alcaraz. These were completed under grant number: DTRT12-G-UTC18.

NURail Center

National University Rail Center - NURail

US DOT OST-R Tier 1 University Transportation Center

NURail Project ID:

NURail2012-MIT-R01 High-Speed Rail as a Complex Sociotechnical System

The Impact of Amtrack Performance in the Northeast Corridor

By

Tolulope A. Ogunbekun

Supervised by Professor Joseph M. Sussman

10/1/2015

Grant Number: DTRT12-G-UTC18

DISCLAIMER
The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

National University Rail Center - NURail
US DOT OST-R Tier 1 University Transportation Center

TECHNICAL SUMMARY

Title

The Impact of Amtrak Performance in the Northeast Corridor

Author: Tolulope A. Ogunbekun

Introduction

The problem addressed in this report: develop an understanding of how Amtrak performs in the Northeast Corridor (NEC) of the US. The primary goal of this research is to study the impact of Amtrak's performance in the Northeast Corridor.

The performance of Amtrak's Acela and Regional services in the NEC is a topic that, while frequently discussed as substandard by some travelers, has received minimal attention in the compendium of open source research literature. Amidst discussions in U.S. Congress to reduce Amtrak's funding, the finances and policies required for track renovation, infrastructure maintenance and quality train operations are also compromised. This provides a backdrop and motivation for the work done in this thesis.

This report focuses on Amtrak's Acela and Regional passengers, as well as the travel time performance of these services in the last ten years (2005 to 2014). The thesis evaluates different factors that lead to variability in ridership and service performance, as well as the impact of service performance on ridership. Another objective of the thesis is to hypothesize about how service performance affects future demand on the Acela and Regional services. This research lays the foundation for future work on the impact of Amtrak's performance, and measures needed to strengthen and improve intercity passenger rail in the Northeast Corridor.

Approach and Methodology

This report is essentially empirical, using Amtrak data to ascertain current performance on the NEC, with particular emphasis on on-time performance, delays and variability in travel time along the corridor. Methods include standard statistical techniques with time series analysis playing an important role.

This report evaluates available Acela and Regional ridership data, as well as measures of performance using train operations data to assess Amtrak passenger rail service performance in the Northeast Corridor. The objectives of this research are to understand the factors that influence the service performance of Amtrak Acela and Regional in the Northeast Corridor, and to evaluate the impact of service performance on current and future ridership.

Findings

The analyses and discussions in this report showed that the Acela and Regional experienced large amount of delays each year between FY 2005 and FY 2014, and more so in FY 2014. In FY 2014, 58% of operated Acela and Regional trains arrived at their final destination later than the scheduled arrival time, causing delays to 66% of total Acela and Regional riders (about 7.7 million passengers). These delays resulted from variability in train travel times and service performance. The author investigated a number of factors that influence service performance and evaluated their impact in FY 2012 and FY 2014. FY 2012 and FY 2014 represented the years with best and worst annual service performance between FY 2005 and FY 2014. In FY 2012, although the percentage of trains impacted by delays were fewer (41% of trains arrived late), some of the factors causing routine delays appeared to be similar to those in FY 2014. The factors characterizing ridership and service performance variations discussed in this thesis were:

Seasonality and month of year

Day of week and time of day

Administration, management and control elements

Capacity levels on trains

Accidents and incidents

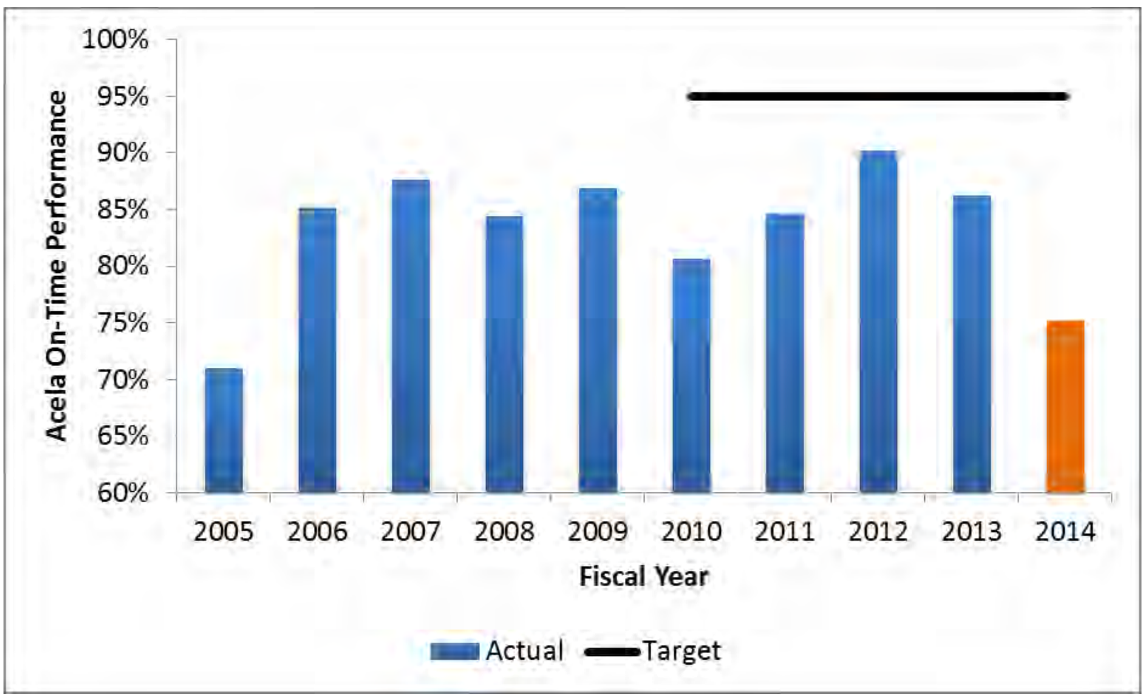
Interference from other trains

Demand response to service performance

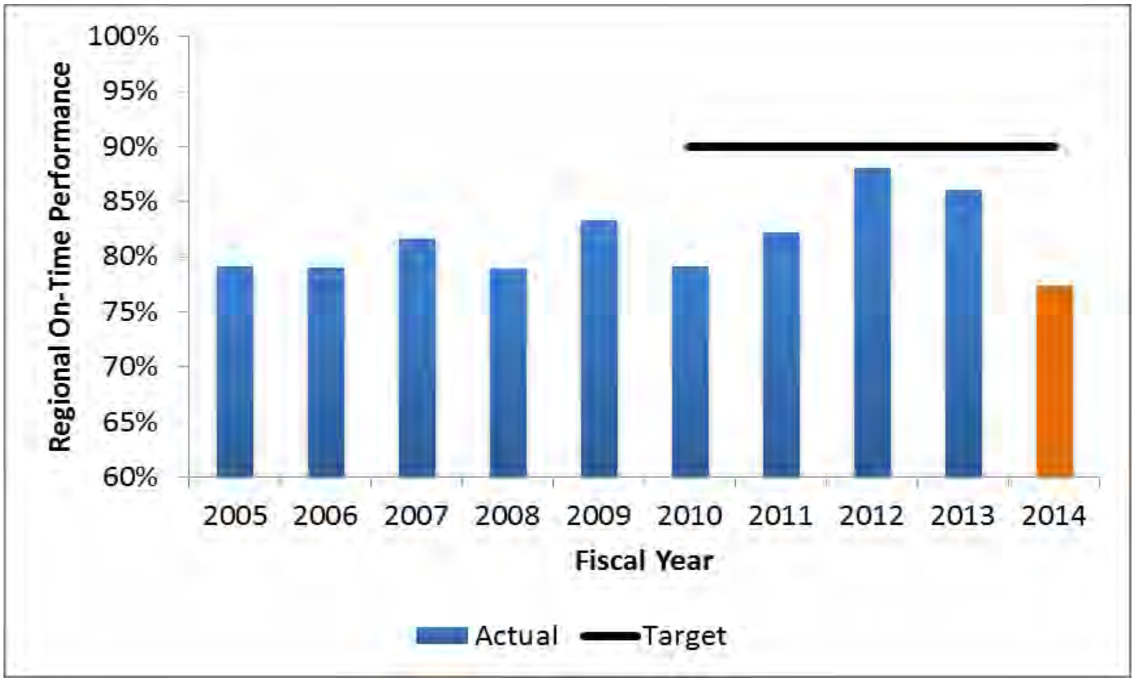
Conclusions

In this section, we present two figures that characterize the large picture of performance on the Acela and Regional services respectively, in the Northeast Corridor (NEC). The black line represents the PRIIA on-time performance goals, which Amtrak set together with the Federal Railroad Administration (FRA) in FY 2010. In both figures, the blue bars show that the actual annual on-time performance on the Acela and Regional services trended towards the PRIIA goals until FY 2012, and then performance deterioration was sustained and amplified between FY 2013 and FY 2014. Consequently, the regrettable conclusion is that both Acela and Regional are currently underperforming.

endence.



Acela FY 2005 – FY 2014 On-Time Performance



Regional FY 2005 - FY 2014 On-Time Performance

Recommendations

This section provides recommendations and active steps for how Amtrak can monitor and improve actual service performance, as well as service performance records for the Northeast Corridor.

Refine data records: Cleaning the dataset cleaner will greatly add to the understanding of performance on the NEC

Refine timetables: Although the stopping pattern of the trains have changed slightly, Amtrak has not implemented any significant train schedule adjustments in more than 25 years, even when train technology changed. In this report, daily routine delays were attributed to timetable artifacts, which affected both Acela and Regional trains. Amtrak should plan to fine-tune the train timetables on an annual basis, and especially when new trains sets and train slots are introduced.

Educate about and reinforce on-time culture: Under PRIIA requirements, Amtrak currently has at least 18 delay codes categorized under Amtrak- and third-party-responsible delays, used to report causes of delays and responsible party. Although this database was not provided to the author, other public reports[i] indicate that Amtrak-responsible delays include delays causes by passengers boarding and alighting, delays caused by crew lateness, etc, while third-party delays include delays caused by weather-related issues, police-activity issues, etc. While the third party delays are for the most part unavoidable, some Amtrak-responsible delays should continue to be managed and reduced by educating and reinforcing on-time culture for operating and managing Amtrak crewmembers. Furthermore, Amtrak should monitor and utilize the information in the cause of delay and responsible party database in order to tackle, reduce and eliminate some of the minor causes of delays, such as late departure from originating station. Furthermore, delays attributed to train, crew and control center personnel's can be made managed through a monthly or quarterly review process for train managing and operating personnel, or on-board visual aids to help train drivers monitor deviations from scheduled timetables.

Management, Policies and Programs: Strict policies and programs like PRIIA Section 207 are required to help Amtrak meet established goals. The analyses in this thesis showed that even though Amtrak owns most of the track in the Northeast Corridor, both Acela and Regional services experienced an unprecedented high in service performance in FY 2011 and FY 2012 while PRIIA Section 207 was active, and both services have been encountering performance deterioration since PRIIA Section 207 was overturned in FY 2013. In addition to statutory laws like PRIIA, other proven techniques like Six Sigma and Lean (used predominantly in manufacturing systems and more recently in health care systems) could be utilized to improve quality output of train operations by identifying and removing the causes of errors and minimizing variability in service operations.

Upgrade Infrastructure: In the long-term, Amtrak requires adequate funding to address essential track alignment/ curvature renovations, catenary maintenance, bridge and tunnel restorations, and rolling stock and signal improvements along the Northeast Corridor. In FY 2012, there was only one major delay caused by operational issues, however, by FY 2014, about 50% of the severe daily delays (greater than 1,000 minutes) were attributed to equipment and infrastructural issues, which evidences the deteriorating infrastructure in the Northeast Corridor in recent years. Consequently, Amtrak's Northeast Corridor Capital Investment Program[ii], which was designed to achieve a state of good repair and facilitate performance enhancement in the corridor requires a stable, multi-year funding program as opposed to the current unpredictable annual appropriation of funds.

Publications

The Impact of Amtrak Performance in the Northeast Corridor (attached)

Tolulope Ogunbekun and Joseph Sussman

Primary Contact

Professor Joseph M. Sussman
sussman@mit.edu
(617) 253-4430

Principal Investigators

Names: Tolulope A. Ogunbekun, Joseph M. Sussman

Title: **The Impact of Amtrak Performance in the Northeast Corridor**

Department: Civil and Environmental Engineering

University: Massachusetts Institute of Technology

Tolulope A. Ogunbekun
tolu.ogunbekun@gmail.com
(617) 253-4430

Joseph M. Sussman
sussman@mit.edu
(617) 253-4430

Other Faculty and Students Involved

Name: Joseph M. Sussman

Title: JR East Professor of Civil and Environmental Engineering and Engineering Systems

Department: Civil and Environmental Engineering

University: Massachusetts Institute of Technology

sussman@mit.edu
(617) 253-4430

NURail Center
217-244-4999
nurail@illinois.edu
http://www.nurailcenter.org/

The Impact of Amtrak Performance in the Northeast Corridor

by

Tolulope A. Ogunbekun

B.A. Physics

Mount Holyoke College, 2009

B.S. Mechanical Engineering

University of Massachusetts, Amherst, 2010

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Engineering in Civil and Environmental Engineering
at the

MASSACHUSETTS INSITUTE OF TECHNOLOGY

June 2015

Signature of Author:

Department of Civil and Environmental Engineering

May 20, 2015

Certified by:

Joseph M. Sussman
JR East Professor of Civil and Environmental Engineering and Engineering Systems
Thesis Supervisor

Accepted by:

Heidi M. Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering
Chair, Departmental Committee for Graduate Students

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

The Impact of Amtrak Performance in the Northeast Corridor

By

Tolulope A. Ogunbekun

Submitted to the Department of Civil and Environmental Engineering
On May 20, 2015 in partial fulfillment of the requirements for the degree of
Master of Engineering in Civil and Environmental Engineering

ABSTRACT

The performance of Amtrak's Acela and Regional services in the Northeast Corridor (NEC) is a topic that, while frequently discussed as substandard by some travelers, has received minimal attention in the compendium of open source research literature. Amidst leading discussions in U.S. Congress to reduce Amtrak's funding, the finances and policies required for track renovation, infrastructure maintenance and quality train operations are also compromised. This provides a backdrop and motivation for the work done in this thesis.

Amtrak is a vital transportation provider on the Northeast Corridor serving travelers between Boston, MA and Washington, DC, including major cities such as Providence, RI; New Haven, CT; New York, NY; Philadelphia, PA; and Baltimore, MD. In Fiscal Year 2014, Amtrak had a record high of 11.6 million passengers on the Acela and Regional services combined. However, in FY 2014 only 3.9 million passengers arrived at their destination at the scheduled arrival time, that is, 7.4 million passengers experienced delays for a myriad of reasons. Furthermore, in 1981, Amtrak advertised Acela's predecessor (Express Metroliner) as trains that made the trip between Washington, D.C. and New York in 2 hours, 59 "civilized" minutes with a 92% on-time performance. Thirty-three years later, travel times in the NEC have barely improved; the Washington, DC – New York trip currently takes 2 hours 44 minutes on Acela and 3 hours 24 minutes on the Regional. Additionally, in FY 2014 overall on-time performance on the Acela and Regional services were 74% and 77%, respectively, despite a 10-minute delay threshold.

This thesis focuses on Amtrak's Acela and Regional passengers, as well as the travel time performance of these services in the last ten years (2005 to 2014). The thesis evaluates different factors that lead to variability in ridership and service performance, as well as the impact of service performance on ridership. Another objective of the thesis is to hypothesize about how service performance affects future demand on the Acela and Regional services. This research lays the foundation for future work on the impact of Amtrak's performance, and measures needed to strengthen and improve intercity passenger rail in the Northeast Corridor.

Thesis Supervisor: Joseph M. Sussman

Title: JR East Professor of Civil and Environmental Engineering and Engineering Systems

<This page intentionally left blank>

ACKNOWLEDGEMENT

I am blessed to have been surrounded by an amazing family, and a great group of friends and colleagues who encouraged and supported me throughout my time at MIT. Specifically, I would like to thank mom and dad, Oyinade, Mayokun, Segun, Foluso, Amy, Abigail, Ibukun, Tomisin and Joanna – I could not have done this without you.

In particular, I would like to thank:

My thesis and academic advisor, Joe Sussman for his time and support, and for guiding me through this thesis research.

Bill Sheridan, Vice President, Market Research and Analysis at Amtrak for providing the data necessary for this research.

Steer Davies Gleave in Boston for their support.

In addition, I would like to acknowledge the people and families who were affected by the tragic Amtrak train derailment, which occurred near Philadelphia on May 14, 2015. The incident further emphasizes the need to work towards a more reliable and safer passenger rail system in the Northeast Corridor and the U.S. at large.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

This thesis aims to assess the performance of the Amtrak passenger services in the Northeast Corridor. It focuses only on performance measures related to travel time reliability of the scheduled services. Furthermore, the thesis aims to evaluate the effect (if any) of performance on the demand for Amtrak services based on the fact that travelers place a high value on being able to get to their destinations "on time".

The introduction begins by presenting an overview of the study area – the Northeast Corridor in S**ection 1.1**. **Section 1.2** introduces the specific topic and provides a background for the research. The motivations for the thesis topic are discussed in **Section 1.3**. And the specific research objectives are outlined in **Section 1.4.** Lastly, **Section 1.5** presents the organization of each chapter in the thesis.

1.1 Area of Analysis

The Northeast Corridor (NEC) is a 457-mile stretch of fully electrified railway line between Boston and Washington D.C (mainline). The NEC network includes the additional feeder corridors to Springfield, MA and Harrisburg, PA. It crosses eight states – Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Delaware, Maryland, as well as the District of Columbia and many other cities. Amtrak owns 79% (363 miles) of the track between Boston and Washington D.C., while New York Metropolitan Transportation Authority (NYMTA) and Connecticut Department of Transportation (ConnDOT) own the New Haven Line, a 56-mile section between New Rochelle, NY. and New Haven, CT., and the Massachusetts Bay Transportation Authority (MBTA) owns the 38-mile section between the Massachusetts-Rhode Island border and Boston-South Station, known locally as the Attleboro Line.

Figure 1.1 NEC Infrastructure Ownership (Source: NEC Commission)[iii]

The section owned by NYMTA and ConnDOT is operated and controlled by Metro-North Railroad (MNR); however, Amtrak operates (dispatches and maintains the right-of-way) in the section owned by the MBTA under an agreement with the MBTA. The Northeast Corridor is a shared-use rail corridor; it operates a variety of services, including freight, commuter, and intercity "higher" speed services on the same track. Amtrak operates 153 trains a day on portions of the Northeast Corridor; in addition, there are more than 2,000 daily commuter trains, and 70 daily freight trains that utilize the corridor. The shared use character leads to a number of physical and operational challenges. The two main Amtrak services along the mainline (BOS – WAS) of the Northeast Corridor are the Acela Express (Acela) and Northeast Regional (Regional/NER). However, there are other Amtrak routes that operate on shorter segments of the NEC, and some others that do not run on the Northeast Corridor but terminate at a station in the NEC. The four groups of Amtrak services that are active in the NEC are introduced below:

i. **Acela Express (Acela)** is Amtrak's fastest rail service. It operates along the Northeast Corridor mainline between Boston and Washington. It began operations in December 2000 and serves major cities along the corridor, including Providence, New Haven, New York City, Philadelphia and Baltimore. The Acela Express service is currently the only existing "high-speed" operation in the United States and can attain a speed of 150mph (240km/h) but only for a relatively short distance due to infrastructure constraint. The term "high-speed" is used because there are different definitions for the speed criteria both within the U.S. and internationally. In brief, the most general consensus requires a sustained speed of more than 125mph (201km/h). Although Acela can theoretically attain a speed of 150mph, in reality the maximum operational speed is about 110mph and the average speed is in the 60mph to 70mph range. The Acela service is scheduled to cover the 457 miles between Boston and Washington in 413 minutes (~7 hours, average speed ~65 mph), the 226 miles between New York and Washington, D.C. in 171 minutes (~3 hours, average speed ~75 mph), and the 231 miles between New York and Boston in 223 minutes (~ 3.7 hours, average speed ~60 mph). Amtrak currently schedules 32 daily Acela trains on weekdays; 16 northbound and 16 southbound trains. The Acela currently stops at 13 main stations in the NEC, namely (Amtrak station codes provided in parenthesis): Boston MA (BOS), Back Bay MA (BBY), Route 128 MA (RTE), Providence RI (PVD), New Haven CT (NHV), Stamford CT (STM), New York NY (NYP), Newark NJ (NWK), Philadelphia PA (PHL), Wilmington DE (WIL), Baltimore MD (BAL), BWI Airport (BWI) and Washington DC (WAS). In addition, a few Acela trains also stop at New London CT (NLC) and Metropark NJ (MET).

ii. **Northeast Regional (Regional/NER)** is Amtrak's moderate speed rail service with more frequent local stops along the Northeast Corridor. It serves the same major markets between Boston and Washington D.C. as the Acela Express, and in addition serves local markets in

Massachusetts, Rhode Island, Connecticut, New Jersey and Maryland. Some NER trains also extend beyond the Northeast Corridor to Richmond, Newport News and Lynchburg in Virginia. There are actually different Amtrak services that are considered as Northeast Regional – the regular NER (described above), includes some Keystone trains that serve local markets west of Philadelphia but also operate between Philadelphia and New York, some Vermonter trains that serve local markets in Vermont but also operate between Springfield, MA and Washington, the Carolinian that serves local markets in North Carolina but also operate between New York and Washington, and the Pennsylvanian that serves local markets in Pennsylvania but also operate between New York and Philadelphia. Altogether, Amtrak currently schedules 64 daily Regional trains on weekdays, 32 northbound and 32 southbound. The regular Regional services are scheduled to cover the 457 miles between Boston and Washington in 470 minutes (~8 hours, average speed ~59 mph), the 226 miles between New York and Washington, D.C. in 202 minutes (~3.4 hours, average speed ~66 mph), and the 231 miles between New York and Boston in 250 minutes (~ 4.2hours, average speed ~55 mph). Finally, the Regional service currently stops at 38 stations in the NEC.

iii. **Other Amtrak routes that use shorter portions of the Northeast Corridor** are the Cardinal, which runs from New York to Chicago, the Crescent service, which operates between New York and New Orleans, and the Palmetto, Silver Meteor and Silver Star, which operate between New York and Florida. The above-mentioned Amtrak routes operate one train each day in both directions, and they all use the NEC portion between New York and Washington. While these trains stop to receive and discharge passengers at some of the stations in the Northeast Corridor, Amtrak passengers wishing to travel solely within the Northeast Corridor are not able to purchase tickets on the trains.

iv. **Other Amtrak routes that do not use the Northeast Corridor but terminate at a station in the NEC** are the Downeaster, which originates in Portland, ME and terminates in Boston, MA. The Adirondack (Montreal, QC to New York, NY), the Empire Services (Albany-Rensselaer, NY and Buffalo-Exchange St., NY to New York, NY), the Ethan Allen Express (Rutland, VT to New York, NY), the Maple Leaf (Toronto, ON to New York, NY), and the Lake Shore Limited (Chicago, IL to Albany, NY and then separate trains go to Boston, MA and New York, NY), which all terminate in New York, NY, and Capitol Limited (Chicago, IL to Washington, DC), which terminates in Washington, DC.

There are eight **Commuter Railroads** that operate on some sections of the Northeast Corridor. Collectively, they operate about 2,000 trains daily.

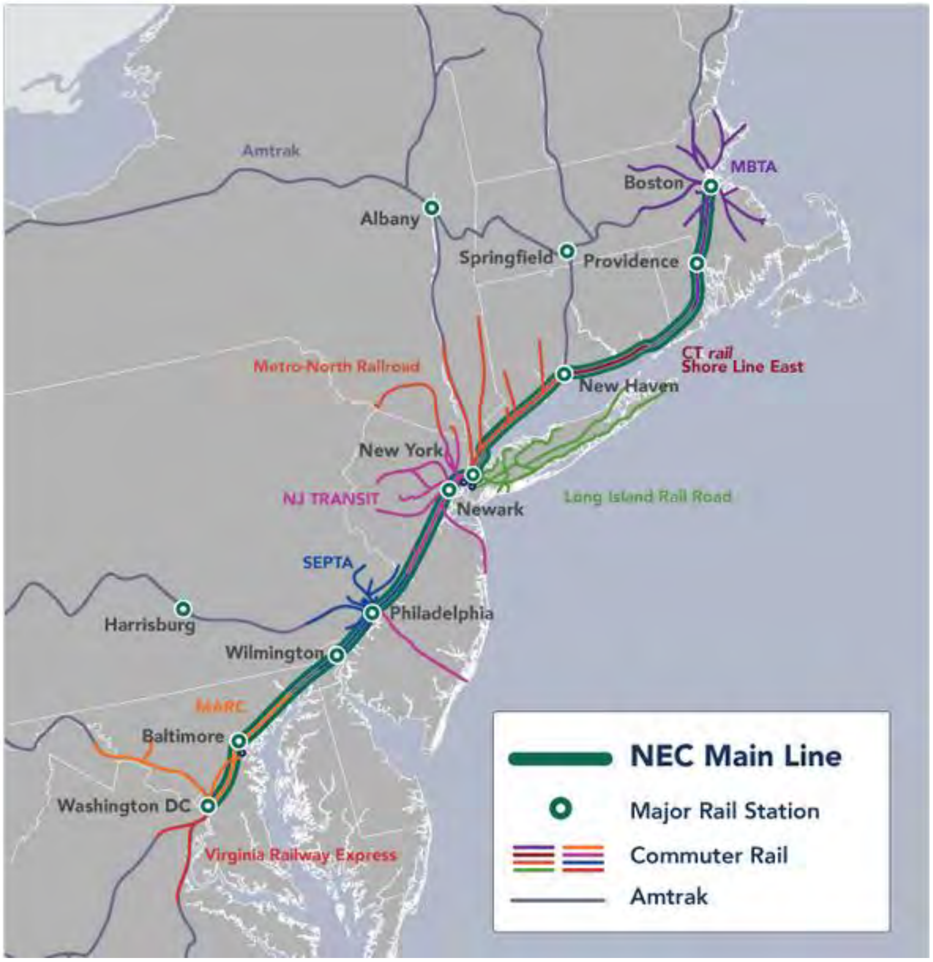


Figure 1.2 Commuter Railroads along the NEC (Source: NEC Commission)

The following commuter services operate in the NEC – including the cities served.

i. Massachusetts Bay Transportation Authority (MBTA) - Boston, MA
ii. Shore Line East (SLE) - New York City, NY; New Haven, CT; New London, CT
iii. Metro-North Railroad (MNR) - New York City, NY; New Haven, CT
iv. Long Island Rail Road (LIRR) - New York City, NY; Long Island, NY
v. New Jersey Transit (NJT) - New York City, NY; Newark, NJ; Trenton, NJ
vi. Southeastern Pennsylvania Transportation Authority (SEPTA) - Philadelphia, PA
vii. Maryland Area Regional Commuter (MARC) - Baltimore, MD; Washington, DC
viii. Virginia Railway Express (VRE) - Washington, DC

Although it is not a major freight corridor, there are four **freight railroads** that operate on parts of the Northeast Corridor Main Line. There are about 70 daily freight trains on the corridor, with the heaviest tonnage flows in Maryland and Delaware[iv]

i. **Norfolk Southern Railway –** operates south of Philadelphia
ii. **CSX Transportation** – operates between New York and New Haven, and in sections in Maryland
iii. **Conrail –** operates between Philadelphia and New York
iv. **Providence and Worcester Railroad** – operates between New Haven and Rhode Island.

1.2 Amtrak Performance Background

The primary goal of this research is to study the impact of Amtrak's performance in the Northeast Corridor. The two keys words in the subject statement are "impact" and "performance". Firstly, "impact" refers to any consequence or effect experienced by the stakeholders of Amtrak's service. The primary stakeholders include the service providers, Amtrak, the commuter operators, and the freight operators, and the service consumers, Amtrak passengers. This thesis focuses on Amtrak Acela and Regional train operations on the supply side and Amtrak passengers on the demand side. Secondly, "performance" refers to degree to which Amtrak achieves the task of service provision measured against some predetermined standards. In principle, the arrival and departure times of each train at each station should be in accord with the published timetables. In other words, performance is a measure of the repeatability and predictability of Amtrak's service compared to a given schedule and/or expectation.

The performance of Amtrak's system largely influences users' perception of the quality of the services, as well as the Amtrak brand. Travel surveys, including Amtrak in-house surveys show that the expected arrival time is a main driver of people's travel choices. Users across all transportation modes place a high value on being able to get to their destination "on time". Similar to travelers on other modes, Amtrak passengers incur some negative effect or disutility from delays and unreliable travel time experienced during their Amtrak trip. Amtrak's Northeast Corridor (NEC) served approximately 32,000 daily Amtrak riders (Acela and Regional), and in total, there were about 11.6 million total Acela and Regional passengers in Amtrak's Fiscal Year 2014 (October 2013 to September 2014). Furthermore, in FY 2014, the daily Acela ridership was as high as 18,600, although the median daily weekday ridership was about 11,300. And for Regional service in FY 2014, the highest daily ridership was 34,800, and the median was about 22,100 riders.

Amtrak recognizes the value of punctual performance of its services. For one, travel time reliability influences a traveler's decision to ride an Amtrak train, which impacts Amtrak's ridership and revenue totals. Amtrak currently measures the reliability of its service, and routinely publishes the following performance measures:

i. Delay minutes,
ii. End-point on-time, and
iii. All station on-time.

The **delay minutes** calculates the difference between the actual arrival time and scheduled arrival time for each train. Figure 1.3 shows the distribution of the delay minutes at the endpoint terminals for Amtrak Acela and Regional trains in FY 2014. If all trains arrived at the scheduled arrival time, we would see a spike up to 100% at the 0 delay minutes mark, and say the train had 100% schedule adherence. However, Amtrak is routinely not able to achieve arrivals at the scheduled time, and Figure 1.3 shows the distribution of actual FY 2014 deviations of Acela and Regional trains from the scheduled arrival times. The FY 2014 delay distribution shows a short spike at the 0 delay minutes mark and a very long tail. It illustrates the variability in arrival times across multiple trains and days, showing that a few trains are able to achieve on-time arrivals but many trains suffer from some delay, and further that a few trains suffer significantly large amounts of delay. In FY 2014, only 41% of all Amtrak trains arrived at the terminal station before or at the scheduled arrival time. An additional, 19% of trains arrived late but within 5 minutes of the scheduled arrival time, that is in total 60% arrived either on-time or within a 5-minute threshold. In total, 72% of Acela and Regional trains arrived within a 10-minute threshold, while the remaining 28% experienced more than 10 minutes of delay. Moreover, about 5% of trains in FY 2014 experienced delays greater than one hour.

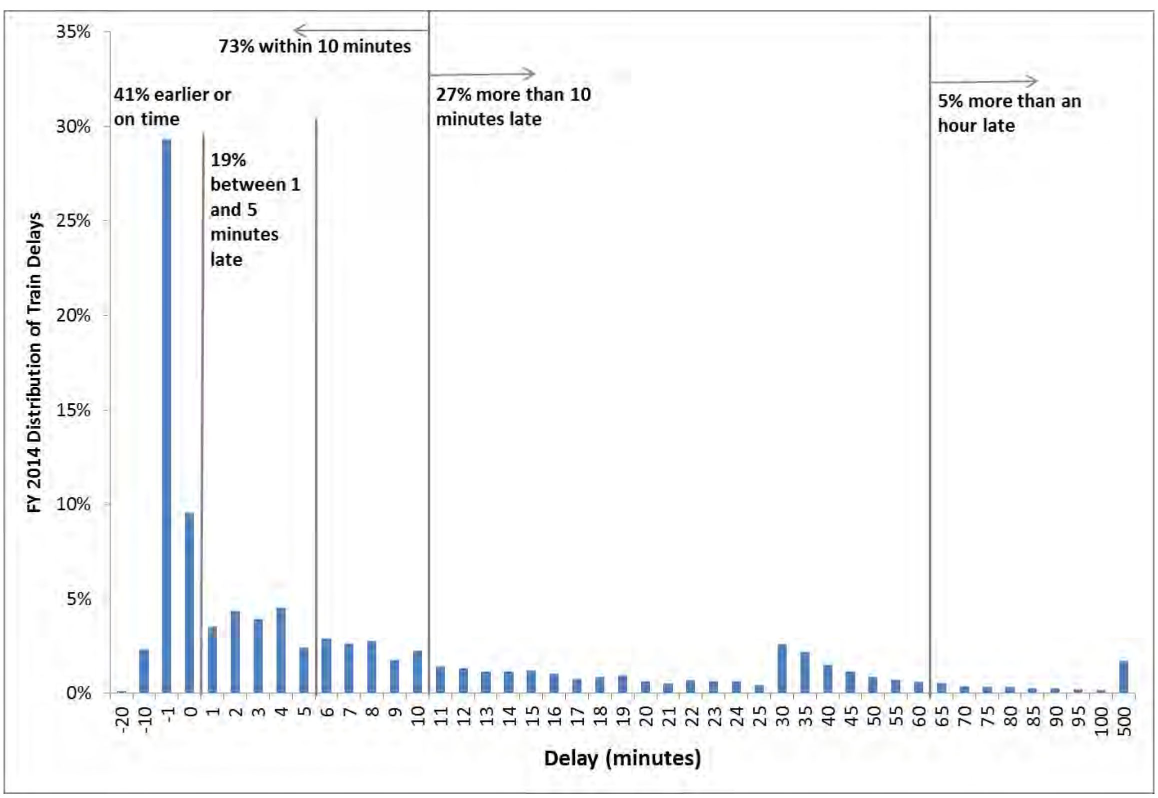


Figure 1.3: FY 2014 Distribution of Endpoint Delay

The **end-point on-time performance** measures the percentage of trains that arrive at their final destination on time, while the **all station on-time performance** measures the percentage of trains that arrive at each en-route station on time. In the Northeast Corridor, an Acela train is classified as “on time” if it arrives within 10 minutes of its scheduled arrival time, while a Northeast Regional train is classified “on time” if it arrives within 10 minutes for trips less than 250 miles, 15 minutes for trips between 251 and 350 miles, and 20 minutes for trips between 351 and 450 miles. In FY 2010, Amtrak set the endpoint on-time performance (OTP) target at 95% for Acela and 90% for Regional.

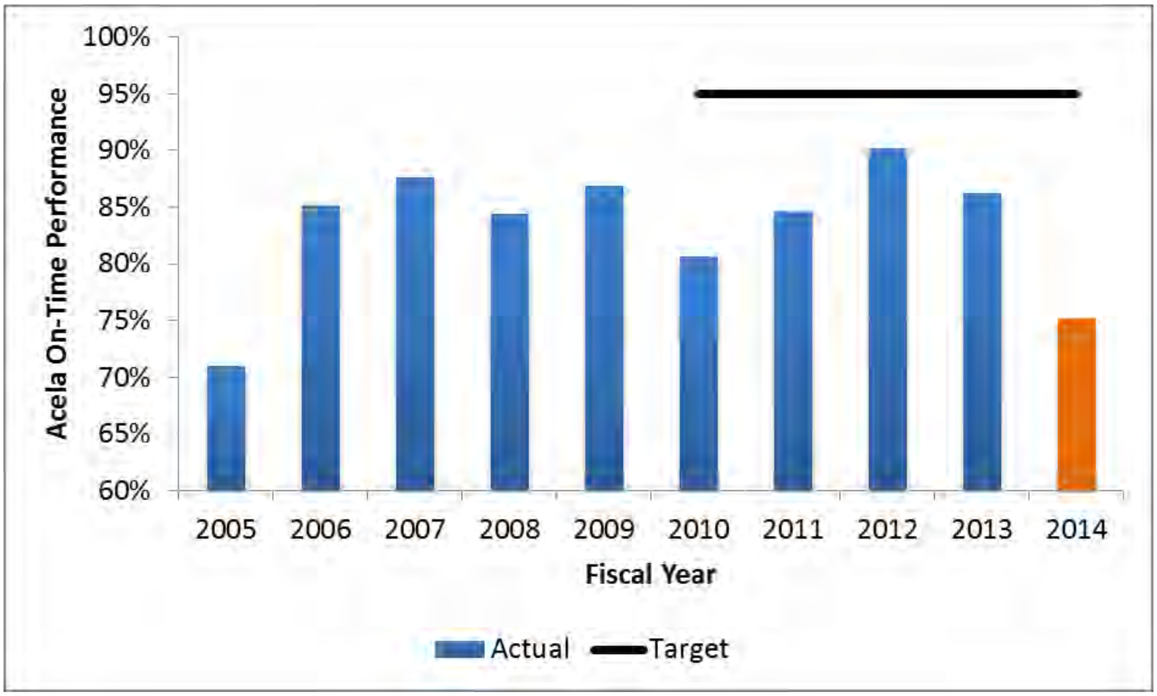


Figure 1.4: Acela Annual On-Time Performance

Figure 1.4 shows the annual average end-point OTP for the Acela service compared to the performance target. The black line represents the OTP target for Acela; prior to FY 2010, Amtrak did not have an established on-time target. The most recent fiscal year (FY 2014) is highlighted in orange. Between FY 20010 and FY 2014, despite the 10-minute arrival buffer, the performance on the Acela service has been about 5 to 20 percentage points below the target. Furthermore, compared to the 95% OTP target, Fiscal Year 2012 experienced the best service performance at 90% while FY 2014 experienced the worst at 75%. In other words, in FY 2014, 1 in 4 Acela trains arrived at their final destination more than the 10 minutes after the scheduled arrival time.

Under Amtrak policies, the Regional service is considered “on-time” if it arrives at the endpoint terminal within 10 to 20 minutes (depending on trip distance) of its scheduled arrival time. In addition, the OTP target for the Regional service was set to 90%. However, despite the lenient buffer and lower OTP goal, Figure 1.5 shows that the Regional service also underperformed by about 5to 20 percentage points between FY 2010 and FY 2014. Although still below target, similar to the Acela, the Regional service experienced the best performance in FY 2012 with an annual average OTP of 88%, and worst performance in FY 2014 with an annual average OTP of 77%.

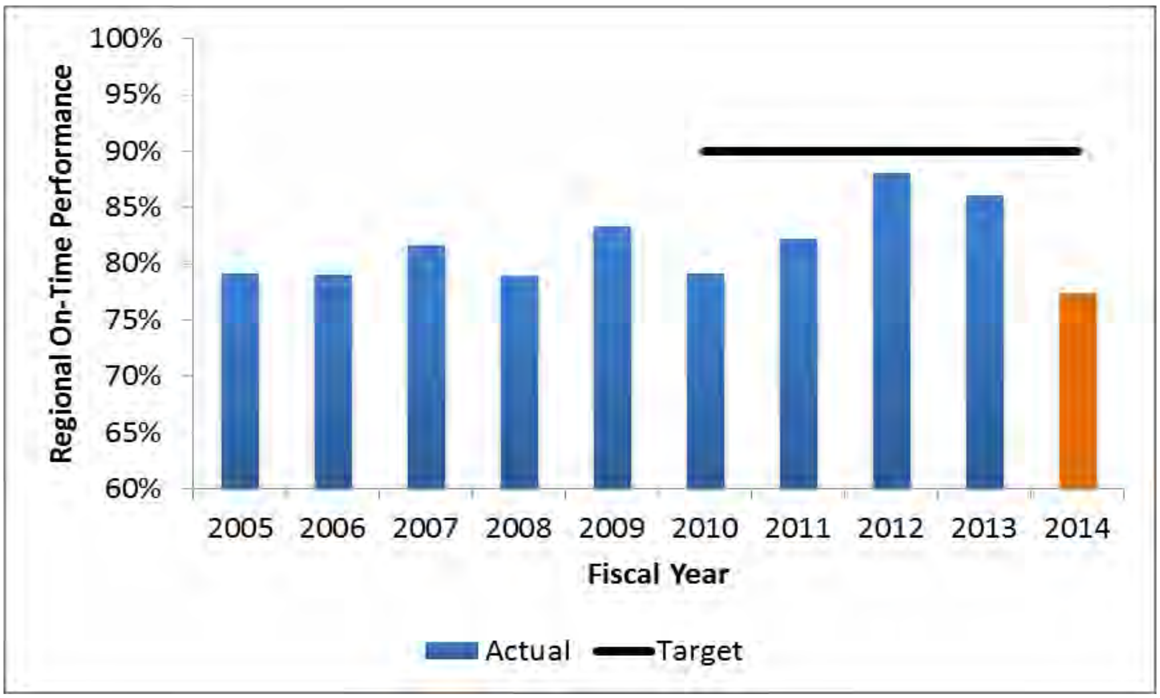


Figure 1.5: Regional Annual On-Time Performance

In comparison to the commuter services that operate on segments of the Northeast Corridor, Amtrak’s performance is lower. The commuter operators reported better on-time performance than both Amtrak Acela and Regional services. Generally speaking, commuter trains are able to perform better than Amtrak services because they typically travel much shorter distances. That said, Table 1.1 reports the on-time performance rate for the commuter services compared with Acela and Regional on-time performance. The on-time performance for the commuter services ranged from 90% to 97%, in comparison with Acela’s 75% and Regional’s 77%. The “on-time” threshold for commuter services is within 5 and 6 minutes of the scheduled arrival time.

Agency		FY 2014 On - Time Performance	OTP Definition
Amtrak	Acela	75%	"On time" if train arrives terminal point within 10 minutes of scheduled arrival.
	Regional	77%	"On time" if train arrives terminal point within 10 - 20 minutes of scheduled arrival.
Commuter	MBTA	90%	"On time" if train arrives terminal point within 4 minutes and 59 seconds of scheduled arrival.
	SLE	92%	"On time" if train arrives terminal point within 5 minutes and 59 seconds of scheduled arrival.
	MNR	97%	
	LIRR	92%	
	NJT	94%	
	SEPTA	91%	
	MARC	92%	
	VRE	92%	

Table 1.1: FY 2014 Amtrak and Commuter On-Time Performance

Furthermore, compared to Amtrak Acela and Regional services, roughly similar rail services in Japan and China routinely achieve on-time performance greater than 99% with a zero-minute delay tolerance. For example, the Tokaido Shinkansen service in Japan that travels 320 miles between Tokyo and Shin-Ōsaka experienced only 0.9 minutes of delay per operational train on average in FY 2014. In Europe, Deutsche Bahn's punctuality for passenger rail transport in Germany was 95.6% in FY 2014. A Deutsche Bahn train is "on-time" if it arrives at each station within 6 minutes of its scheduled arrival time. In terms of performance, Amtrak appears to be behind its international counterparts.

Although Amtrak monitors the performance of its services, it has not been able to implement long-term operational changes due to political, infrastructure, financial and technical constraints. In general, this thesis aims to discuss some of the major constraints that undermine Amtrak's ability to provide reliable service to its passengers in the Northeast Corridor. A specific goal of the thesis is to analyze Amtrak's historical ridership and operations data to characterize the level of performance of Acela and Regional services, and to identify some causes of delays on Amtrak trains, including any catalytic effect of delays across the consecutive trains within a day. An additional goal of the thesis is to estimate the impact of service quality on demand.

1.3 Motivation

The Northeast Corridor (NEC) is the busiest railroad in the U.S, both in terms of service frequency and number of passengers. 2,200 trains operate on the corridor daily; 2,000 commuter trains, 153 Amtrak intercity passenger trains, and 70 freight trains. Between Amtrak and the commuter trains, there were approximately 750,000 daily riders, and in total about 260 million passenger trips made on the NEC in FY 2014. The dynamics of scheduling trains given demand growth, limited capacity, and different train speeds, while striving to achieve the diverse and sometimes conflicting objectives of different service operator is a complex challenge. NEC's multi-functional characteristic poses issues regarding train scheduling, timetable coordination, and train priority, among others. Furthermore, some segments along the NEC are presumed to be operating at or near capacity, and train slots are tightly scheduled with little-to-no slack or recovery leeway. As a result, theoretically minor disruptions on one train could propagate quickly, causing significant delays and deterioration of service quality in the system. There are some opinions that Amtrak's Acela and Regional services are of utmost importance, and thereby should always receive train priority. The counter argument to this school of thought is that the commuter rail and Amtrak services have a source-sink relationship. In other words, they supply and distribute traffic on each other's routes. Consequently, substandard service on commuter rail would likely negatively impact Amtrak demand, and vice versa. Altogether, the NEC is a complex sociotechnical system that deserves a great deal of attention. Furthermore, the NEC is the flagship of rail in the U.S. and at the core of one of 11 megaregions designated for high-speed rail, such that current performances on the corridor influences the view of passenger rail operations in the country, and the hopes for high-speed rail in the Northeast as well as states like Texas, California, and Florida.

From an economic point of view, the Amtrak services in the Northeast Corridor have played a role in the burgeoning economic development of the northeast region, which is one of the densest regions in the United States. The region includes major economic hubs that together had an estimated population of 56 million and generated about 20% of U.S. GDP in 2013. In comparison with the total population in the region, Amtrak served 11.6 million passengers in the NEC in Fiscal Year 2014. Part of the appeal of the Northeast region is the relative proximity between the major cities. The good transportation connectivity in the region widens the radius within which people and businesses can interact. As a transportation link connecting people and businesses, the NEC influences the attractiveness of the Northeast region as a place to live and do business. The labor force in the region have access to businesses, and vice versa, in addition businesses can reach customers and suppliers, and they both have access to other businesses, including retail and public services such as health care, education and entertainment. Amtrak's role, as the only intercity rail connection between the major cities in the NEC is important. Consequently, its performance in the NEC is crucial because it directly affects the economic development in the region.

Also from an economic point of view, Acela Express service generated $585.8 million and the Regional service generated $603.5 million in Amtrak Fiscal Year 2014. Both services currently achieve an operating (not including infrastructure costs) cost recovery ratio considerably greater than 1. It can be expected that an improvement in Amtrak services might reduce costs from inefficient operations, as well as increase demand (given sufficient capacity), thus reducing costs and increasing revenue over time.

The NEC is one of the oldest rail corridors in the U.S., which affords a wealth of valuable historical data for research. It is important to ensure that future developments in passenger rail keep up with innovation and advancing technology, and more importantly with the changes in the transportation environment and travel pattern. This research aims to study the performance of Amtrak in the Northeast Corridor, through the lens of the research community keen on solving complex problems. Given the current capacity and shared-use constraints in the Northeast Corridor, tackling service performance is a complex challenge. The research aims to develop a case study of Amtrak's current performance in the Northeast Corridor, investigate pros and cons of programs that have been tested to enhance service performance, and come up with new solutions that could provide incremental or perhaps transformational improvements. In addition, this research has the potential to highlight lessons that other intercity passenger rail networks might be able to learn from.

Due to the high density and productivity in the city hubs along the NEC, many of the trips are concentrated at major nodes. A significant number of the 56 millions people who live and travel in the corridor are making trips around the same time, for similar reasons, and with roughly similar origins and destinations. This is evident from peak period congestion across all modes of transportation, and around the central business districts in the corridor. Currently, the main travel alternatives in the Northeast region are rail, air, and highway, which include auto and bus. Rail has a significant advantage here because it has a much higher capacity than the other modes. The current Acela trainsets have a seating capacity of 304 (44 first class; 260 business class), while the Regional trainsets have a capacity ranging from about 500 to 750 seats, which both provide much higher capacities than the other modes. Consequently, intercity passenger trains will continue to be indispensable as the population in the northeast region continues to grow, assuming similar trends in future trip. Additionally, improvements in the service quality of Amtrak rail will further strengthen its competition in the NEC.

The transport sector is responsible for 23% of global CO_2 emissions and is also the fastest growing source of CO_2 emissions. At the same time, population in the Northeast region is expected to increase by an additional 15 million resident, and intercity travel is expected to increase by 45% to 75% between 2014 and 2040/2050. Many of the discussion to mitigate the current trend in CO_2 emissions, and lower carbon emission per trip have included an improvement in the quality of "low emission" modes such as walking, cycling and some common-carrier modes. Amtrak fits under the common-carrier modes umbrella. In addition, the NEC as it stands today is fully electrified, which makes it a sustainable mode of the future. For example, trains consume 17% less energy per passenger-mile than airline, and 34% less than automobile. Improving Amtrak's service performance has the potential to reduce Amtrak's energy consumption and CO_2 emissions by reducing unnecessary braking, and also produce mode shift from air, auto and bus. The need to reduce the transportation sector's CO_2 footprint further emphasizes the need for an improvement in Amtrak's performance in the Northeast Corridor.

All of the above motivations provide a glimpse as to why the Northeast Corridor and Amtrak services are vital and important to study. Finally, the performance of Amtrak's Acela and Regional services in the Northeast Corridor (NEC) is a topic that, while frequently discussed as substandard by some travelers, has received minimal attention in the compendium of open source research literature. Amidst leading discussions in U.S. Congress to reduce Amtrak's funding, the finances and policies required for track renovation, infrastructure maintenance and quality train operations are also compromised, further motivating this research.

1.4 Research Objectives

The author will evaluate available Acela and Regional ridership data, as well as measures of performance using train operations data to assess Amtrak passenger rail service performance in the Northeast Corridor. The objectives of this research are to understand the factors that influence the service performance of Amtrak Acela and Regional in the Northeast Corridor, and to evaluate the impact of service performance on current and future ridership. The research focuses on the variability in ridership and service performance between Fiscal Year 2005 and Fiscal Year 2014, as well as discusses specific factors that lead to travel time variability and delays. In the process, studies regarding service performance within the transportation context in both the U.S. and internationally will be reviewed, and different metrics to quantify service performance will be examined. Additionally, the thesis will provide an overview of Passenger Rail Investment and Improvement Act (PRIIA) Section 207, the most recent measure taken to address the service performance of Amtrak trains in U.S. Overall, the research aims to understand what causes delays on Acela and Regional trains, and subsequently to provide suggestions and active steps on how Amtrak can monitor and improve service performance in the Northeast Corridor.

1.5 Thesis Organization

This thesis is organized into eight chapters including the introduction. The **second chapter** reviews the literature relevant to the subject of performance of transportation services. The **third chapter** examines the performance metrics and standards that were introduced under Passenger Rail Investment and Improvement Act (PRIIA) Section 207, as well as the effectiveness and current status of the program. The **fourth chapter** provides an overview of the data on Acela and Regional ridership and service performance received from Amtrak. The **fifth chapter** and **sixth chapter** examine the factors leading to ridership and service performance fluctuations on the Acela and Regional services, respectively. The **seventh chapter** discusses the time series and regression analysis. Finally, the **eighth chapter** concludes the thesis with a summary and recommendations.

The next chapter reviews the literature relevant to the topic of service quality and performance in the transportation context.

<This page has been intentionally left blank>

2 Literature Review

In order to integrate service performance metrics into transportation models, it is important to understand how variability in service performance impact demand. This section is divided into three parts. **Section 2.1** briefly discusses various definitions and measures of reliability. **Section 2.2** reviews previous research on methods to evaluate reliability and its effect on demand. And **Section 2.3** presents a summary of the results obtained in the literature.

2.1 Definitions and measurements of reliability in different contexts

Accuracy measures how close a measured value is to the standard (true) value, while precision measures how close sets of measured values are to each other. Consequently, precision is synonymous with repeatability. Following an analogous concept, a train service that arrives at a destination at the exact scheduled time is considered accurate, while a train service that always arrives, for example, five minutes late, though inaccurate is precise, due to the repeatability of the service. These two schools of thought represent the general concept by which people often describe the service performance of passenger rail systems. On one hand, reliability could refer to consistent on-time rail service, and on the other hand to predictable service, which could include some expected level of delay. Consequently, travel time reliability studies measure the variability in travel times across multiple scheduled trips.

As a result of unreliable service quality, individuals spend a longer time traveling than expected or desired, which leads to stress associated with the uncertainty and sometimes consequences of late arrivals at their destination (e.g. reduced pay or work termination). To account for expected travel time variability, travelers can decide to change their departure time, but sometimes also change their routes and/or mode as a result of unreliable service quality.

In transportation systems, travel time variability introduces uncertainty to travel plans, and therefore is often formulated as additional costs and disutility. In the U.K., the Passenger Demand Forecasting Handbook (PDFH) measures passenger rail performance by using the terms ‘reliability’ and ‘punctuality’, where ‘reliability’ refers to the rate of cancellation, and ‘punctuality’ refers to the percentage of operated trains (trains that were not cancelled) that arrive under a given 'lateness' threshold. 'Lateness' measures the difference between the actual departure/arrival times and the published timetables, while 'delay' measures the difference between the actual time and en-route adjusted schedules. The Public Performance Measure (PPM) is a synthesis of both 'reliability' and 'punctuality', and it measures the percentage of all scheduled trains that are operated and arrive under a given 'lateness' threshold. A train that arrives under the given 'lateness' threshold is considered 'on-time'. However, globally, the 'on-time' notion varies across train service, total travel time or distance, operator and country. It ranges from exactly on time to within five and/or ten minutes of published timetables. Furthermore, some long distance services in the U.S. even consider trains that arrive under 15-20 minutes of the published schedule as 'on-time'. In the U.S., Amtrak measures on-time performance (OTP) as the percentage of trains that achieve the 'on-time' target compared to the total number of trains in service. Furthermore, Amtrak defines the acceptable 'on-time' target differently for end-point and station-to-station OTP. All of this is to say that measures of travel time reliability using PPM and OTP, while useful in characterizing a specific system/service, are not consistent with each other, and across different passenger rail systems, and therefore not directly comparable. In addition, the most obvious metric, delay-minutes, the difference between the actual time and the published timetables can also be misleading because a 20 minute delay on a 30-minute trip is very different from a 20 minute delay on a 6-hour trip. To account for this, in the U.S. Amtrak normalizes the delay-minutes metric to 10,000 train miles to provide a consistent reporting basis, and to allow for comparison across services.

Additionally, across the literature on travel time variability, reliability is measured as the standard deviation (or variance) of the travel time distribution or by the number of minutes travelers are willing to arrive earlier or later than a preferred arrival time (PAT). Another ambiguity is found in this definition; in some studies, travelers experience a disutility from arriving either early or late, while in others disutility is only as a result of a late arrival.

A final reliability metric that is common in the literature on travel time variability is the value of reliability (VOR). VOR is similar to the concept of value of time (VOT), and measures a traveler's willingness to pay for reduction in travel time variability (increase in travel time reliability). In addition, the reliability ratio (RR) evaluates the tradeoff between mean travel time and travel time variability (VOR/VOT). The reliability ratio is sometimes alleged to offer a more consistent and comparable metric for assessing the impact of travel time variability.

In summary, the different performance metrics across the transportation literature are:

Performance Metric	Definition
Reliability	Rate of cancellation
Lateness	Difference between the actual arrival/departure times and the published timetables
Punctuality	Percentage of operated trains that arrive under a given 'lateness' threshold
Public Performance Measure (PPM)	Percentage of all scheduled trains that are *operated* and arrive under a given "lateness" threshold
On-Time Performance (OTP)	Percentage of trains that achieve the "on-time" target compared to the total number of trains in service (*All-station vs End-point)*
'On-time'	Varies across train service, total travel time or distance, operator and country from no tolerance to a 10-15 minute threshold
Total Delay Minutes	Difference between the actual arrival/departure times and published timetables
Delay-minutes Per 10,000 train miles	Difference between the actual time and published timetables, normalized for train miles to ensure a consistent reporting basis
Effective Speed	Ratio between train mileage and the total travel time planned into the timetable, to ensure that service speeds are not simply reduced on the timetables in order to achieve a given performance standard
Schedule Delay Early/Late (SDE/SDL)	Disutility from arriving either early or late
Value of Reliability (VOR)	Willingness to pay for reduction in travel time variability (increase in travel time reliability) similar to the concept of value of time (VOT)

Reliability Ratio (RR)	Tradeoff between mean travel time and travel time variability (VOR/VOT)

Table 2.1: Performance Metrics Definitions

2.2 Theories/Models/Methods to measure the effect of unreliability

The theories in the literature used to measure the effect of service performance on demand can be categorized under three main methods:

i. Econometric model
ii. Mean-variance model
iii. Scheduling model

Although the underlying theory of most studies is rooted in one of these three approaches, each study tends to adopt variations in the formulations within each method.

2.2.1 Econometric model

In the U.K., the relationship between demand and average lateness at destination station is based on the formulation established in the British Passenger Demand Forecasting Handbook (PDFH). In this context, lateness refers to the difference between the actual and public timetable arrivals at destination stations:

$$I = \left[1 + \frac{w(\bar{L}^{+}_{scenario} - \bar{L}^{+}_{base}}{GJT_{base}}\right]^{\varphi}$$

where,

I is change in rail demand

$\bar{L}^{+}_{base}$ is average lateness in the 'base' case

$\bar{L}^{+}_{scenario}$ is average lateness in the 'scenario' case being forecasted

GJT is Generalized Journey Time in the base case

w is lateness multiplier to convert average lateness to equivalent GJT

φ is elasticity of rail demand to GJT

The plus notation $\bar{L}^{+}_{*}$ imposes a non-negativity constraint on the average lateness variable. That is, on-time or early arrivals at the destination station are given a value of 0. Under the forecasting framework, Generalized Journey Time (GJT) is the sum of the service characteristics (in-vehicle time, frequency, and transfer time) converted into time units. The main criticism of this approach is that it might not adequately estimate the demand elasticity to lateness because the elasticity of GJT is fixed and the demand response is largely given by adjustments to the lateness multiplier more than the magnitude of $(\bar{L}^{+}_{scenario} - \bar{L}^{+}_{base})$.

Batley et al. (2011) developed an econometric model to estimate the elasticity of demand for rail in terms of ticket sales at O–D levels with respect to changes in performance metrics. Three performance metrics were modeled:

i. **Average Lateness Minutes (ALM)** - an average of the difference between actual running time and <u>en-route adjusted schedules</u> weighted by passenger loadings at each station
ii. **Average Performance Minutes (APM)** - an average of the difference between actual running time and all <u>published schedules</u> (that is, including canceled services)
iii. **Public Performance Measure (PPM)** - the percentage of trains that arrive at their destination within a specific margin of the publicly available timetables (five minutes for short distance and ten minutes for long distance)

Batley et al. (2011) used panel data, a combination of cross-section data and time-series data accumulated from several O-D pairs over a number of years, and subdivided into 13 4-weekly periods. They formulated a static and dynamic model. The static model was a constant elasticity demand model

$$\ln V_{ijt} = \mu_{ij} + \sum_{k=2}^{13} \kappa_{\kappa} D_{\kappa} + \phi \ln F_{ijt} + \varphi \ln GJT_{ijt} + \gamma \ln G_{it} + \eta \ln R_{ijt} + \varepsilon_{it}$$

where,

V_{ijt} is the volume of rail demand between stations i and j at period t

μ_{ij} are O–D-specific effects which account for time-invariant differences between flows not specified in the other variables of the model, i.e. ‘fixed-effects’;

D_{κ} are 12 dummy variables to account for seasonality in the 4-weekly data;

F_{ijt} is rail fare for O–D ij at period t divided by the retail price index;

GJT_{ijt} is generalized journey time for O–D ij at period t

G_{it} is Gross Value Added at Government Office Region of origin i at time t

R_{ijt} are the performance metrics for O–D ij at time t

The dynamic model was an autoregressive distributed lag model (ADL), which included lags of both the dependent and independent variables

$$\ln V_{ijt} = \mu_{ij} + \sum_{k=2}^{13} \kappa_\kappa D_\kappa + \sum_{s=0}^{S} \phi_s \ln F_{ijt-s} + \sum_{l=0}^{L} \varphi_l \ln GJT_{ijt-l} + \sum_{m=0}^{M} \gamma_m \ln G_{it-m} + \sum_{u=0}^{U} \eta_u \ln R_{ijt-u} + \sum_{w=0}^{W} \lambda_w \ln V_{ijt-w} + \varepsilon_{ijt}$$

where,

s, l, m, u *and* w are the lags for the associated variable at period 0 to time maximum lag length denoted by the upper case letter.

The results from both models revealed a marginal but statistically significant effect of lateness and reliability on rail passenger demand, ranging from -0.01 to -0.06 for ALM and APM, and from 0.02 to 0.27 for PPM. In the dynamic model, the lag for each performance metric was 1 * 4-weekly period.

Halse et al. (2014) developed a panel data fixed-effect model (treating O-D pairs as separate observations) to estimate the demand effects of rail reliability in Norway. The fixed-effect regression method was used to control for unobserved factors that account for cross-sectional differences in reliability across segments. The results showed low elasticity (~-0.01) of demand to delay. The specifications of the static and dynamic models used are as shown below:

Static model:

$$Q_t = \alpha + \beta_s Delay_t + \beta_2 Canc_t + \theta_1 DOW_t + \theta_2 Week_t + \delta \cdot t + \gamma \mathbf{X_t} + \varepsilon_t \quad (1)$$

With lags:

$$Q_t = \alpha + \beta_1 Delay_t + \beta_2 Canc_t + \sum_{k=1}^{K} \lambda_{1k} Delay_{t-k} + \sum_{k=1}^{K} \lambda_{2k} Canc_{t-k} + \theta_1 DOW_t + \theta_2 Week_t + \delta \cdot t + \gamma \mathbf{X_t} + \varepsilon_t \quad (2)$$

Paul Schimek (TRB 2015) applied a similar econometric model to estimate the elasticity of demand to fare changes. He analyzed panel data (cross-sectional time-series data) using both fixed- and random-effect models, and found the fixed-effect models to be better. The models were estimated in double log form, that is, both independent and dependent explanatory variables were log-transformed in order to better capture large variability, and directly interpret coefficients as elasticities. He also included lagged dependent variables, and estimated a dynamic pooled model using the software Gretl.

2.2.2 Scheduling Approach

This approach was motivated by discrete choice models and utility maximization theory under the assumption travelers attach a utility to arriving at their destination at a particular time, commonly referred to as the preferred arrival time (PAT). Under this framework, travelers associate a cost due to an early or late arrival at their destination that leads to a reduction in utility.

Empirical research by Small (1982) mostly based on Gaver (1968) and Vickery (1969) suggest a travel disutility influenced by the departure time and given by:

$$U(t_d) = \alpha T + \beta(SDE) + \gamma(SDL) + \theta D_L$$

where,

t_d is the travelers departure time choice

T is travel time

SDE is schedule delay early and given by $Max(0, PAT - [T + t_d])$

SDL is schedule delay late and given by $Max(0, [T + t_d] - PAT)$

D_L is a binary variable equal to 1 when SDL >0 and 0 otherwise

α, β, γ and θ are cost of travel time, cost per minute of arriving early or late, and lateness penalty, respectively, and expected to be negative.

Noland and Small (1995) further developed this framework to include the effect of uncertainty based on maximum expected utility theorem. As a result of travel time uncertainty and a traveler's inability to plan exactly, a traveler taking into account travel time uncertainties will choose the option that has the highest value of expected utility. The expected utility is thus given by:

$$E[U(t_d)] = \alpha E[T] + \beta E[SDE] + \gamma E[SDL] + \theta P_L$$

Where, $P_L = E[D_L]$ is the probability of experiencing a late arrival, that is, the proportion of time in which a late arrival occurs (it is independent of the magnitude of the late arrival). Moreover, a more variable travel time results in a larger expected value of SDE and SDL. The scheduling approach treats SDE and SDL separately and tends to capture greater disutility from late arrival, and is thus cited by some studies as a better evaluation of variability

The publications by Noland et al. (1998) and Small et al. (1999), also included in the overview by Noland and Polak (2002), are among the best examples of the schedule delay function approach.

2.2.3 Mean-variance (Centrality-dispersion) approach

Under this framework, a traveler maximizes utility by minimizing both travel time and variability in travel time (unreliability), based on the formulation:

$$U = \beta_T \mu_T + \beta_\sigma \sigma_T$$

where,

μ_T is average travel time

σ_T is travel time variability (standard deviation or variance)

α, β are model coefficients.

The average travel time is included as the mean (centrality) variable, while the travel time variability represents the variance (dispersion) term. Travel time variability is typically measured by standard deviation or variance of travel time, as well as by travel time distribution percentiles. An alternative formulation includes the expected utility as well as travel fares:

$$E[U] = \beta_T E[T] + \beta_\sigma \sigma_T + \beta_C C$$

where, $E[T]$ is expected time, σ_T is standard deviation of travel time, C is travel cost, and $\beta_T, \beta_\sigma, \beta_C$ are estimated coefficients. The value of time (VOT) is measured as β_T/β_C, to estimate the willingness to pay for reduction in travel time. Similarly, the value of reliability (VOR) is the ratio of β_σ/β_C, and represents a travellers willingness to pay for reduction in travel time variability. In addition, the reliability ratio (RR), β_σ/β_T is defined as the marginal rate of substitution between average travel time and travel time variability (VOR/VOT).

Publications by Varian (1978), Jason and Jucker (1982), Senna (1994), Copley et al. (2002) and Hollander (2006) are among the best examples of the mean-variance model

2.3 Overview of Different Approaches and Results in Different Studies

As highlighted in Section 2.2 the various studies reviewed were based on the different theoretical frameworks. The estimated reliability ratios (RR) from the studies ranged from 0.1 to 2.51, and the estimated value of reliability (VOR) ranged from \$0.79 to \$56 per hour in 2009\$. These results suggest that on one hand, travelers place barely any value on reliable travel, and on the other hand, travelers are willing to pay up to \$56 to reduce unexpected travel delays by an hour. Similarly, a reliability ratio of 2.5 suggests that travelers are willing to pay 2.5 times more for an hour reduction in variability than for an hour reduction in total travel time. The significant variation among the results is a consequence of the different theoretical approaches, as well as the various data sources used, year of study, transportation mode, time of day (peak versus off-peak), trip purpose, and study country. A comprehensive review of value of reliability studies, including theoretical approaches and results is presented in Carrion and Levinson (2012), Li et al. (2010), and Noland and Polak (2002). Overall, De Jong et al. (2009) proposed some recommended reliability ratios based on available international evidence, especially from the UK, The Netherlands and Sweden; a reasonable reliability ratio for car travel was 0.8, and for interurban train and public transport (bus, tram, metro) was 1.4. Bates et al. (2001) concluded from their research that reliability is highly valued by travelers, and that a plausible reliability ratio for car travel was around 1.3, and no more than 2.0 for public transport.

To the best of the author of this thesis' knowledge, there has not been any study to identify the value of reliability, reliability ratio, or elasticity of demand to reliability/service performance for rail in the U.S. More specifically, to the best of the author's knowledge, there is no prior research studying the value Amtrak travelers in the Northeast Corridor place on service performance, or to quantify the effect of service performance on demand in the Northeast Corridor. This void in open-source literature further motivated the work done in this thesis.

The next chapter discusses recent measures of service performance of Amtrak trains in the Northeast Corridor introduced under the Passenger Rail Investment and Improvement Act (PRIIA) of 2008.

<This page has been intentionally left blank>

3 Passenger Rail Investment and Improvement Act (PRIIA) Section 207

3.1 Overview

The Passenger Rail Investment and Improvement Act (PRIIA) of 2008 is the most recent measure taken to address the service performance of passenger trains in the U.S.. PRIIA was established as a platform through which the National Railroad Passenger Corporation (Amtrak), the U.S. Department of Transportation (US DOT), Federal Railroad Administration (FRA), states, and other stakeholders could collaborate to strengthen and improve intercity passenger rail. The 'other' stakeholders included the Surface Transportation Board, freight railroads, rail labor unions, and passenger rail organizations. PRIIA Section 207 focused on metrics and standards to measure the performance and service quality of intercity passenger rail operations. The two main performance indicators utilized were on-time performance (OTP) and train delay.

3.2 Background

The FRA and Amtrak were in charge of establishing performance indicators to monitor the reliability of intercity passenger rail operations. Although Amtrak is the sole intercity passenger rail operator in the U.S., and it owns 79% of track in the NEC, outside the NEC, Amtrak primarily runs on tracks owned and operated by freight railroad companies. Consequently, the main groups that would be affected by the metrics, in terms of administering changes were Amtrak as the rail operator, and the host railroads, which include both Amtrak and the freight railroads.

In 1973, Congress granted Amtrak the right of preference, which obligated host railroads to grant dispatching preference to Amtrak passenger service over freight operations. The host railroads were also required to pay Amtrak in the event of any violations of the right of preference that led to Amtrak train delays. However, Amtrak and the host railroads measured delays and violations of the right of preference differently, which led to disagreements and difficulty in enforcing the rule. In addition, Amtrak and the host railroads had other service quality agreements that differed by state and host railroad. Furthermore, each of the states served by the State-supported routes had the right to negotiate state operating and performance contracts with Amtrak and the other railroads. As a result, the existing performance indicators lacked uniformity. A main motivation for PRIIA Section 207 was to create a standardized method to measure the performance and service quality of intercity passenger train operations using on-time performance and minutes of delay. The statute was designed to improve any existing metrics and develop new standards and minimum standards that were consistent and comparable over time, and across Amtrak routes, states, and host railroads. The stricter laws with monetary penalties for not meeting the standards were expected to transform train priority and lead to a dramatic improvement in passenger rail performance.

The FRA and Amtrak jointly developed performance measures for Amtrak and the host railroads based on historical operational and performance data provided by Amtrak. On March 13, 2009, FRA and Amtrak released the first draft of proposed metrics and minimum standards, and solicited feedback from the other stakeholders and invested parties. The stakeholders and invested parties included freight railroad companies that host intercity passenger trains, state department of transportation, commuter passenger rail agencies, labor unions that represent Amtrak employees, and groups to represent Amtrak and commuter passengers. After receiving feedback between March 13 and March 27, 2009, the proposed metrics and standards for intercity passenger rail service were clarified or revised, and finalized.

3.3 Final Metrics and Standards for Intercity Passenger Rail Service

The metrics and standards addressed on-time performance and minutes of delay separately. The on-time performance metric dealt with the repeatability of reliable service, while the minutes of delay metric dealt with the precision of service or amount of deviation from the published schedules. Although PRIIA Section 207 was established for all Amtrak services in the U.S., only the metrics and standards relevant to the Northeast Corridor routes will be highlighted in the rest of this section. Furthermore, in terms of the enforcement of PRIIA standards and the penalties for not meeting the standards, because Amtrak owns 79% of track in the NEC, it serves as both rail operator and host railroad in the NEC. The other host railroad is Metro North Railroad (MNR), which operates and controls a 56-mile section on the NEC between New Rochelle, NY. and New Haven, CT. Consequently, Amtrak would essentially be meting out penalties to either MNR or itself if service quality failed to meet the standards established under PRIIA.

3.3.1 On-Time Performance

The on-time performance (OTP) metric monitors the repeatability of service. In other words, it measures the variability of travel time over repeated trips, and the percentage of Amtrak trains that achieve the 'on-time' target compared to the total number of trains in service. Under PRIIA Section 207, three metrics were established and used to monitor OTP:

Percent on-time at the endpoint

Under PRIIA, an Acela train was considered ‘on-time’ if it arrived its endpoint terminal within 10 minutes of the scheduled arrival time. Comparably, a Northeast Regional train was considered ‘on-time’ if it arrived within 10 minutes for trips less than 250 miles, 15 minutes for trips between 251 and 350 miles, and 20 minutes for trips between 351 and 450 miles. Table 3.1 summarizes the on-time definitions for Acela and Regional. Starting in FY 2010, endpoint OTP was required to be at least 90% for Acela and 85% for Regional. By FY 2014, this threshold was required to increase to at least 95% for Acela and 90% for Regional. These thresholds were measured against either the published timetables or adjusted schedules (Amtrak occasionally releases an adjusted timetable for a specified period of days due to major renovation, track work, or other major time-impeding projects on the corridor).

Table 3.1: “On-Time” Definition for Acela and Regional

“On-time”	
Acela	**Regional**
10 min	10 min; < 250 mi 15 min; 251 and 350 miles 20 min; 351 and 450 miles

Percent on-time at all-stations served

The tolerances for the all-station OTP ensured that the scheduled times at all stations (departure time from origin station and arrival times at all subsequent stations) were within 10 minutes for Acela trains, and 15 minutes for Northeast Regional trains. Starting in FY 2012, all-station OTP was also required to be at least 90% for Acela and 85% for Regional. By FY 2014, this threshold was required to increase to at least 95% for Acela and 90% for Regional. Similar to the endpoint OTP, these thresholds were measured against either the published timetables or adjusted schedules.

Change in effective speed

Effective speed was defined as the ratio between train mileage and the total travel time. The total travel time was calculated as the sum of the scheduled end-to-end travel time and average endpoint delay. It was calculated on a rolling four-quarter basis, and compared to a fixed FY 2008 baseline. The effective speed metric was included to ensure that train schedules were not simply lengthened in order to achieve the OTP standards. This metric emphasized the importance of shortening (or at least preventing deterioration in) end-to-end travel times, and guarded against schedule creep.

Table 3.2 shows a summary of the three metrics that were established to monitor the on-time performance of Amtrak trains in the NEC, and the standards that were fixed under PRIIA:

Metric	Endpoint OTP		All-station OTP		Change in Effective Speed vs. FY 08
	Acela	Regional	Acela	Regional	
First Year (FY 2012)	90%	85%	90%	85%	>=0
By FY 2014	95%	90%	95%	90%	>=0

Table 3.2: On-Time Performance Metrics and Standards

3.3.2 Train Delays – On NEC

The minutes of train delay metric dealt with amount of deviation from the published schedules. This metric was required under PRIIA section 207, in addition to the OTP metrics in order to keep track of all delays encountered even when trains arrived 'on-time'. The train delay metric was measured in minutes per 10,000 train miles to ensure a normalized reporting basis because some trains in the NEC operate on only half of the corridor (i.e. Boston, MA to New York, NY or New York, NY to Washington, DC) while other trains operate on the full length of the corridor (i.e. Boston, MA to Washington, DC). In addition, all Amtrak routes operated well over 10,000 train-miles every month. Specifically, both Acela and Regional services operated over 10,000 train-miles daily. Table 3.3 shows that total daily train-miles on the Acela services is almost 12,000.

Direction	Acela Markets	Distance (miles)	# of daily Trains	Train-Miles
SB	Boston - Washington	457	10	4,570
	New York - Washington	226	6	1,356
NB	Boston - New York	231	1	231
	New York - Washington	226	7	1,582
	Boston - Washington	457	9	4,113
	Total train-miles :			**11,852**

Table 3.3: Acela Daily Total Train-Miles

The acceptable minutes of delay per train-miles metric was calculated separately for Acela and Northeast Regional services. The FY 2008 data was used for the analysis, to evaluate the amount of delay minutes were incurred on days when the OTP was above a specified minimum OTP standard. The OTP standards used in the calculations were 90% for Acela, and 85% for Northeast Regional.

For Acela, an existing mathematical regression model was used to estimate a delay threshold of 285 minutes of delay/10,000 train-miles. This value was the mid-point of the high-low delay minutes range that corresponded with a 90% endpoint on-time arrival rate. The threshold for Acela was later adjusted to 265 minutes/10,000 train-miles, to include only Amtrak-responsible and host-responsible delays (but not third-party responsible delays). For Northeast Regional, the FY 2008 OTP versus delay data were plotted, and used to calculate a delay threshold of 470 minutes/10,000 train-miles. This value was the mid-point of the high-low delay minutes range that corresponded with a 85% endpoint on-time arrival rate. For similar reasons to those above, the threshold for Regional was also later adjusted to 475 minutes/10,000 train-miles.

The 265-minute (Acela) and 475-minute (Regional) standards were intended to absorb routine/seasonal maintenance, track work, and other routine construction projects[v]. However, in the event of a major construction or maintenance project, an additional delay buffer (above the allowable threshold) was permitted. Amtrak owns or operates most of the track in the NEC, except the section owned and operated by Metro North. As a result, Amtrak serves as the host-railroad along the NEC and thus was required to follow the strict standards outlined above. The standard for Metro North railroad was more lenient, with a delay threshold of 900 minutes/10,000 train-miles. Table 3.4 summarizes the minutes of delay standards established under PRIIA.

Delay Metric (Minutes per 10,000 Train-Miles)	Acela	Regional	Metro-North
Total train delays	265	475	900
Cause of delay	Yes	Yes	Yes

Table 3.4: Minutes of Delay Metrics and Standards

In addition to the minutes of delay metric, under PRIIA regulations, the cause of delay and the responsible party, categorized by Amtrak-, host- and third-party-responsible delay had to be reported.

i. **“Amtrak-responsible”** refers to delays coded on *Amtrak Conductor Delay Reports* as Passenger-Related (ADA, HLD), Car Failure (CAR), Cab Car Failure (CCR), Connections (CON), Engine Failure (ENG), Injuries (INJ), Late Inbound Train (ITI), Service (SVS), Crew and System (SYS), or Other Amtrak-Responsible (OTH).

ii. **“Host-responsible”** refers to delays coded on *Amtrak Conductor Delay Reports* as Freight Train Interference (FTI), Passenger Train Interference (PTI), Commuter Train Interference (CTI), Slow Orders (DSR), Signals (DCS), Routing (RTE), Maintenance of Way (DMW), Debris Strikes (DBS), Catenary or Wayside Power System Failure (DET, used in electrified territory only), or Detours (DTR).

iii. **“Third-party”** refers to delays coded on *Amtrak Conductor Delay Reports* as Unused Recovery Time (NOD), Customs (CUI), Police-Related (POL), Trespassers (TRS), Drawbridge Openings (MBO), Debris (DBS), or Weather-Related (WTR).

3.4 Other issues regarding metrics and standards

The issues raised regarding the administration of the metrics and the standards were as follows:

A number of the comments received were with regards to the administration of the metrics and standards. Some of the host railroads noted that the proposed performance measures would present an administrative burden and would require significant operational changes to make current Amtrak schedules realistic. Furthermore, they stated that Amtrak schedules would need to be revised using computer modeling techniques that account for current traffic and seasonal pattern in order to meet the performance standards. From an implementation and data collection point of view, they argued that automated and technically advanced data collection mechanisms would be needed in place to reliably track the performance on actual operations. In addition, they cited the poor infrastructure, and the need for ongoing and future rail infrastructure improvements. FRA and Amtrak responded to the issues raised citing them as valid reasons confirming the importance of the metrics in assisting to detect areas and causes of poor performance, in order to ultimately strengthen and improve intercity passenger rail.

3.5 Performance following PRIIA Section 207

Both the OTP and delay minutes metrics were established to provide a comprehensive indicator of intercity train performance and service quality. The metrics and standards were in effect as of May 12, 2010. Following PRIIA Section 207 in 2010, all Amtrak routes experienced a record high in on-time performance. Although Amtrak owns most of the track in the Northeast Corridor, Acela and Regional also experienced performance improvements due to the stricter rules introduced under PRIIA Section 207. That said, there were other interrelated factors that contributed to the improved performance. The beginning of PRIIA coincided with the economic recession, which led to less freight rail traffic interfering with passenger rail services. Figure 3.1 shows the on-time performance for Acela between FY 2004 and FY 2014. The blue markers indicate the annual average OTP while the vertical lines represent the highest and lowest monthly OTP in each fiscal year. Following the establishment of PRIIA in 2010, Acela experienced an improvement in OTP in FY 2011, further improvement in the following years, and a record high performance in FY 2012.

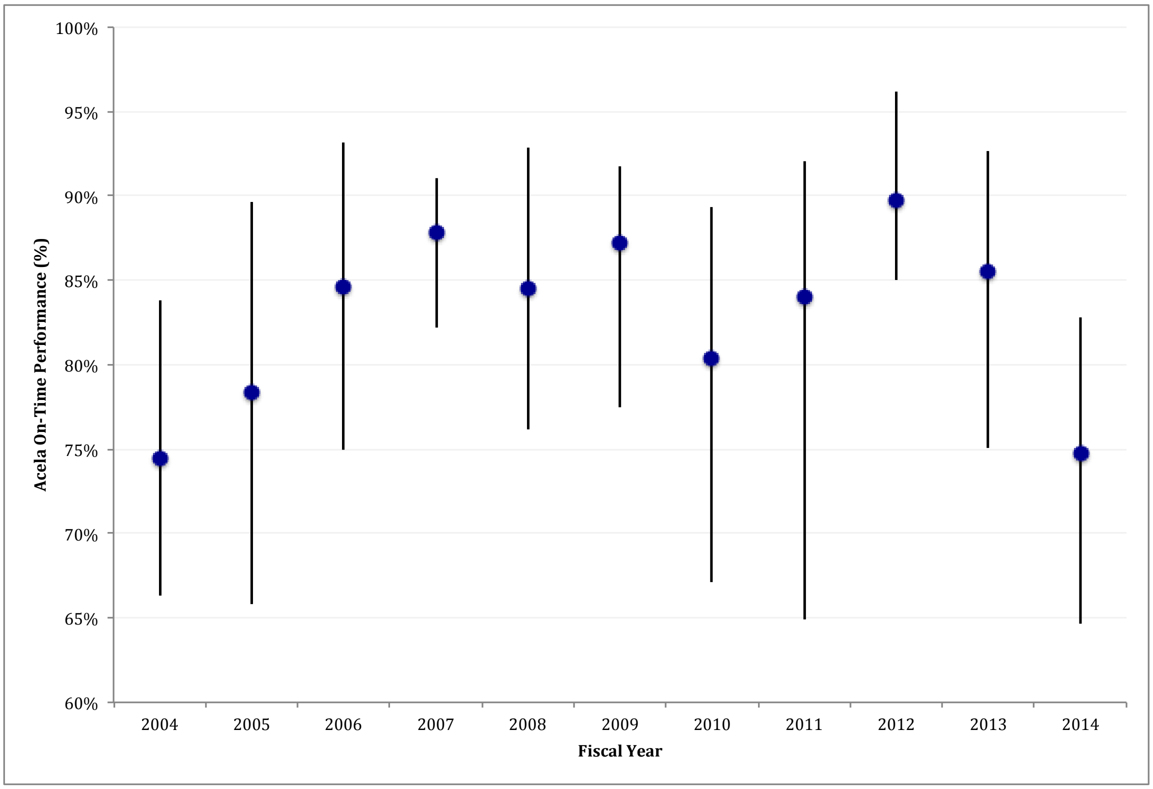


Figure 3.1: Acela FY 2004 to FY 2014 On-Time Performance

Figure 3.2 is similar to Figure 3.1 and shows the on-time performance for the Regional service between FY 2004 and FY 2014. The Regional service also experienced OTP improvements between FY 2010 and FY 2011, and a record high OTP of 88% in FY 2012, only 2 percentage points lower than the PRIIA standard. Following FY 2012, the on-time performance on the Regional service has been deteriorating.

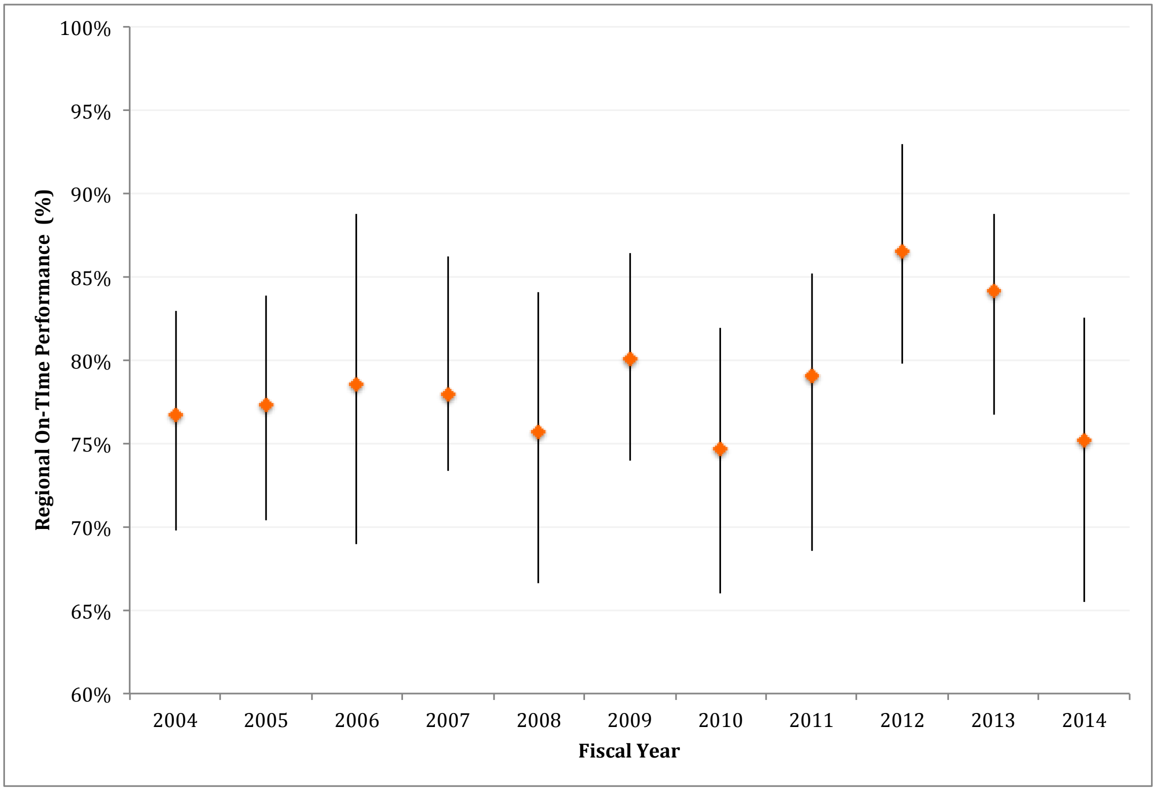


Figure 3.2: Regional FY 2004 to FY 2014 On-Time Performance

3.6 Current Status of PRIIA Section 207

On May 31st, 2012, the Association of American Railroads (AAR) filed a suit in the U.S. District Court for the District of Columbia – *Association of American Railroads (AAR) versus Department of Transportation, et. al., Civil Action 11-1499*. The AAR represents the railroad companies that host both Amtrak and freight trains around the U.S. They stated that it was "unconstitutional delegation of lawmaking and rulemaking permitting the FRA and Amtrak to jointly set the metrics and minimum standards for measuring Amtrak passenger train performance and service quality." They argued that Amtrak was created by Congress to operate and be managed as a for-profit corporation that would benefit financially when host railroads were unable to meet the strict rules under PRIIA Section 207. On May 2012, the Federal District Court dismissed the charges ruling that Amtrak is a government entity whose top goal is to strengthen and improve intercity passenger rail in the U.S. prior to any profit-making motives. The AAR appealed to the U.S. Court of Appeals for the D.C. Circuit (*Association of American Railroads v. U.S. Department of Transportation, et al., No. 12-5204*), and PRIIA Section 207 was eventually overturned on July 2, 2013. Amtrak is currently preparing a counter-suit alleging that the railroads have inadequate dispatching practices, and showing performance improvements while PRIIA section 207 was in place, and worsened performance since it was overturned.

3.7 General discussion of PRIIA and Amtrak performance pre/post PRIIA

Even though Amtrak owns most of the track in the Northeast Corridor, and Metro North Railroad (MNR) is the sole host railroad in the corridor, both Acela and Regional services experienced an unprecedented high in performance in FY 2011 and FY 2012 while PRIIA Section 207 was active. And surprisingly as shown in Figure 3.1 and Figure 3.2, the performance on both Acela and Regional services deteriorated since PRIIA Section 207 was overturned in FY 2013. This suggests that the existence of the metrics and standards were beneficial in improving Amtrak's performance and service quality even within the NEC. As part of this thesis, the performance of Acela and Regional will be discussed in detail, and a particular focus would be given to FY 2012 and FY 2014, which experienced the record best and worst in Amtrak performance in the last ten years.

The next chapter presents an overview and description of the data on historical ridership, operations, and performance provided by Amtrak, which were used for the analysis in this thesis.

<This page has been intentionally left blank>

4 Description of Amtrak Data

The purpose of this chapter is to describe the reporting procedure of the data provided by Amtrak, which were used in this thesis. Amtrak provided a detailed **demand database**, which included ridership, revenue and passenger-miles for a 10-year period dating from October 2004 to September 2014. They also provided a detailed **train operations database**, which included scheduled and actual running times for each scheduled Amtrak train between October 2004 and September 2014.

This Chapter describes the structure and level of detail in the demand and train operations databases received from Amtrak. **Section 4.1** describes the data that shows Amtrak ridership on the Acela and Regional and **Section 4.2** describes the data that shows the performance of each scheduled Acela and Regional train. Finally **Section 4.3** discusses how both datasets were combined for the purpose of analyzing and portraying various aspects of Acela and Regional service performances.

4.1 Overview of Ridership, Revenue and Passenger-miles Database

Amtrak's demand database includes the following columns:

i. Route Code
ii. Passenger Train Number
iii. Date
iv. Bi-directional Station Code
v. Class
vi. Trip Type
vii. Number of Passengers
viii. Total Revenue
ix. Passenger-miles

Descriptions of the data contained in these nine columns are discussed in the following subsections**.**

4.1.1 Route Code

The system of Amtrak trains is organized by routes; each route has a name and route code number. The routes indicate various travel options offered by Amtrak, and are differentiated by the U.S. region and markets they serve, as well as the train stopping patterns within the region/market. The two routes relevant to this thesis work are Route code 1, which is named Acela Express, and Route code 5 named Northeast Regional.

4.1.2 Passenger Train Number

All Amtrak trains are identified by a passenger train number that indicates its direction, the day(s) of the week it operates, and an associated train schedule in the timetable. Acela trains are given numbers between 2100 and 2299. The first two-digits give information about the day of week the Acela train operates on. Trains that begin with 21 are weekday trains operating Monday through Friday, while those beginning in 22 are weekend trains operating on Saturday and/or Sunday, as well as on public holidays that fall on weekdays. Northeast Regional (Regional) trains are given numbers between 100 and 199. However, Amtrak trains that serve the Northeast Corridor (NEC) but extend to markets outside the NEC (e.g. Vermonter, Pennsylvanian and Carolinian) have numbers under 100, and the Keystone trains that extend to Harrisburg, PA have numbers in the 600s. Note that for the Amtrak trains in the 100s and 600s, the ridership in the segments on the NEC mainline between Boston and Washington, DC are considered a part of the Regional service. Unlike the Acela train numbers, Regional train numbers are not formulated to give additional information indicating the day of week service pattern. Furthermore, the last digit of **all** Amtrak train numbers reveals the train direction. Northbound and eastbound trains end in even numbers, while southbound and westbound trains end in odd numbers. For example, one can tell that train 151 is a south-west bound train because it ends with a 1, while train 2228 is a north-east bound train because it ends with an 8. Although trains are actually defined as north-east bound or south-west bound, the convention in this thesis and generally in the Northeast corridor is to identify north-east bound trains as *northbound* trains, and the south-west bound trains as *southbound* trains.

Altogether, an Amtrak train number provides substantial information on the operating pattern of the train that is unique to the train and does not change over time. For example, train 151 is a Northeast Regional southbound train operating on Mondays through Fridays and corresponds to the train on the timetable scheduled to depart from New York Penn Station at 4:40am and terminate in Washington Union Station at 8:15am. Likewise, train 2228 is an Acela northbound train that operates only on Sundays, and corresponds with the train on the timetable scheduled to depart from Washington Union Station at 8pm and terminate in New York Penn Station at 10:55pm.

4.1.3 Class

The class represents the quality of seating and service, and gives an indication of ticket fares. The two class options available on Acela trains are First class (F) and Business class (B). First class is the highest class of service available on the Acela and offers an exclusive seating area and premium amenities, including spacious seating configurations, at-seat attendant service, at-seat meal and beverage service, etc. Business class is the minimum and general seating available on the Acela, featuring wide comfortable seats with moderate spacing configurations. Acela trains have one exclusive first class car with 44 seats, and four general business class cars with 65 seats each, summing to a total of 260 business class seats, and a train capacity of 304 seats. As expected, passengers pay a higher fare for first class seats than business class seats.

Alternatively, the two class options available on the Regional are Business (B) and Coach (C). Business class is the highest seating class available on the Regional, while Coach class is the minimum and more general seating. Regional trains usually have one dedicated business car, but sometimes Business class is just a dedicated seating area with limited seating capacity. Business class passengers pay a higher fare and receive complimentary soft drinks and a newspaper, in addition to the premium and more spacious seating configuration. There are various types of trainsets used to operate the Regional service, and they vary in the number of passenger cars (ranging from having 7 to 10 cars) and seating capacity (ranging from about 500 to 750 seats). Consequently, unlike the Acela that has a definitive capacity of 304 seats per train each day, the capacity on the Regional depends on the train equipment operated on different days.

4.1.4 Date

The date field indicates day, month and year. The dates can be aggregated into either Calendar year or Amtrak Fiscal year (FY). Unlike the calendar year that goes from January to December, Amtrak Fiscal years is the same as the Federal Fiscal year that starts in October of the prior year and end in September. For example, the 2014 Calendar year is from January 2014 to December 2014, while the Amtrak Fiscal Year 2014 (FY 2014) goes from October 2013 to September 2014.

4.1.5 Bi-directional Station Code

The bi-directional station code has the format XXX – YYY and it shows station pair ridership for each northbound and southbound Amtrak train. Each station code (XXX) is a unique three-letter code tied to a specific geographical location and used to identify each Amtrak station. The station pair code is 'bi-directional' because it does not distinguish the origin from the destination. For example, the BOS-NYP code represents the station pair for a passenger on a northbound train who boards in New York Penn (NYP) and alights in Boston (BOS), as well as for a passenger on a southbound train who boards in Boston (BOS) and alights in New York Penn (NYP). The bi-directional code is organized in alphabetical order, and therefore, continuing with the same example, would be presented as BOS-NYP (as opposed to NYP-BOS). Given the train number and direction, and the bi-directional code, one can decipher the actual origin and destination of each passenger.

Furthermore, in the Northeast Corridor, the bi-directional codes are typically grouped into one of three markets, depending on the origin and destination station – north-end, south-end and through markets. As the names imply, the north-end market includes station pairs with both the origin and destination north of New York Penn (NYP) e.g. BOS-NHV, on the other hand, if both origin and destination stations are south of NYP they are grouped in the south-end market e.g. BAL-WAS. The through market includes station pair with an origin North of NYP and a destination south of NYP, and vice versa e.g. BOS-WAS. It should be noted that the through trains serve both the north end (BOS-NYP) and south end (NYP-WAS) markets.

4.1.6 Trip/Ticket Type

There are two types of trips based on the Amtrak ticket structure in the Northeast Corridor – Single-Ride tickets (ST) and Multi-Ride tickets (MR). As the names imply, Single-Ride tickets can only be used once for a particular trip, while Multi-Ride tickets can be used to take multiple trips within a set amount of time using the same ticket. Multi-Ride tickets are not available on any Acela trains, and are available only for certain destinations and time of day on Regional trains. Thus, the majority of trips in the Northeast Corridor are on Single-Ride tickets.

4.1.7 Summary of Demand Database

Amtrak provided the detailed **demand database** from October 2004 to September 2014. The demand data shows station pair ridership, revenue and passenger-miles by route, as well as train number, date, class and trip type on all Acela and Regional train operated. The ridership column indicates the number of passengers who purchased a ticket for a given origin-destination pair, the revenue column indicates the total fare paid by all passengers, and the passenger-mile is a product of the number of passengers and the distance (in miles) between the origin and destination station.

4.2 Overview of Train Operations and Performance Database

This section discusses the **train operations database** showing scheduled and actual running times for each scheduled Acela and Regional train. For the Acela, Amtrak currently schedules 16 daily southbound trains; 10 through trains (BOS-WAS) and 6 south end only trains (NYP-WAS). All Acela trains serve New York Penn Station. As such, from an Amtrak customer's point of view, there are 10 Acela trains between BOS and NYP, and 16 Acela trains between NYP and WAS, since the through trains serve the south-end markets also. In the northbound direction, Amtrak currently schedules 17 Acela trains; 1 north end only train (NYP-BOS), 7 south end only trains (WAS-NYP), and 9 through trains (WAS-BOS). From an Amtrak customer's perspective, there are 10 northbound Acela trains serving the north-end market, 16 serving the south-end market, and 9 serving the through markets.

Service	Route	Market	Distance (miles)	Weekday Round Trips	Scheduled Travel Time (hr:min)
Acela	Boston - New York	North End	231	10	3:25 to 3.35
	New York - Washington	South End	226	16	2:44 to 2:50
	Boston - Washington	Through	457	9 to 10	6:30 to 6:40
Regional	Boston - New York	North End	231	9	4:00 to 4:20
	New York - Philadelphia	South End	91	33	1:20 to 1:25
	New York - Washington	South End	226	23	3:12 to 3:39
	Boston - Washington	Through	457	9	7:40 to 8:05

Table 4.1: Acela and Regional Scheduled Services

For the Northeast Regional, Amtrak currently schedules 63 to 64 trains between Mondays and Thursdays, 67 trains on Fridays, 51 trains on Saturdays, and 53 trains on Sundays (total trains in both directions). On a typical weekday there are 33 scheduled southbound trains; 9 through trains (BOS–WAS), 11 south-end only trains (NYP-WAS), 1 Carolinian train between NYP and WAS, 10 Keystone and Pennsylvanian trains between NYP and PHL, and 2 trains between SPG and WAS in both directions. Hence there are 9 Regional trains that serve the north-end market, 33 that serve the south-end market between NYP and PHL, 23 that serve the south-end market between NYP and WAS, and 9 that serve the through market.

Table 4.1 shows a summary of the Acela and Regional scheduled services. The scheduled travel times are provided as ranges to indicate that trains are sometimes scheduled with different end-to-end travel times based on the actual stopping patterns of the train.

Amtrak's **train operations database** indicates location, day and time information for each Amtrak train organized under the following columns – Train Number, Calendar Date, Location, Scheduled Departure Time, Actual Departure Time, Departure Performance in minutes, Scheduled Arrival Time, Actual Arrival Time, Arrival Performance in minutes, Scheduled Dwell in minutes, and Actual Dwell in minutes.

The Train Number corresponds to the train numbers described in Section 4.1. Amtrak train numbers are used to identify passenger trains and are typically associated with a unique timetable schedule. Amtrak's published timetable shows the Scheduled Departure Time from and Scheduled Arrival Time at each consecutive station along a train route. The scheduled times reflect either Amtrak's published or temporary adjusted timetables. On occasion, due to major renovation, track work, or other major time-impeding projects, Amtrak will release an adjusted timetable for a specified period of days. Amtrak does this as a means of factoring in additional time incurred due to major track-related projects, and also to inform Amtrak passengers to plan for the modified schedules and additional travel time. The adjusted timetables are usually pre-released, publicized, and given with advanced notice, thus real time or en-route travel time changes are not considered here. The provided data also includes a Scheduled Dwell Time in minutes, which is scheduled terminal time at certain stations to ensure enough time for passengers to board and alight at large stations, as well as for occasional train checks and safety measures. Next, as the names imply, the Actual Departure Time indicates the time each train actually departed from each station, and the Actual Arrival Time provides the same level of information in terms of arrivals at stations for each operated train. All departure and arrival time information in the database are given with to-the-minute-precision (e.g. 2:24PM as opposed to 2:24:30PM). Finally, two performance indicators are provided in the database - the Departure Performance in minutes and the Arrival Performance in minutes, which both calculate the difference between actual and scheduled departure and arrival times, respectively.

In addition to the performance indicators provided in the train operations database, On-Time Performance and Total Delay minutes were calculated. The On-Time Performance (OTP) evaluates the percentage of trains that achieve an 'on-time' target at the final destination compared to the total number of trains in service that day. Under PRIIA regulations (see Chapter 3), an Acela train is considered 'on-time' if it arrives at its endpoint terminal within 10 minutes of the scheduled arrival time, while a Regional train is considered 'on-time' if it arrives within 10 minutes for north-end and south-end only trains, and 20 minutes for through trains. In this thesis Amtrak's current on-time performance standard is referred to as Amtrak OTP, and is distinguished from the Pure OTP, which refers to a performance standard that categorizes a late train as one that arrives later than the scheduled arrival time. In other words, the Pure OTP is a no-tolerance delay metric while the Amtrak OTP is a 10 to 20-minute delay tolerance metric. Thus, the Pure OTP provides an objective perspective on the on-time metric. The Total Delay minutes measures the difference between the scheduled and actual arrival at the endpoint terminal (final station). Altogether, Amtrak's scheduled and operated train database is valuable for characterizing Amtrak performance in the Northeast Corridor.

4.3 Combining the ridership and performance databases

For the analysis in this thesis, the train performance data was joined to the *station pair* ridership data for each train operated in the last 10 years. The combined data was then aggregated to show total ridership, total number of operated trains, total minutes of delay, total number of delayed trains under the Pure 'on-time' metric (0 minutes threshold), total number of delayed trains under Amtrak's 'on-time' metric (10 - 25 minutes threshold), OTP (Pure and Amtrak), and the total number of passengers on delayed trains (Pure and Amtrak) for each day over the last 10 years. The daily data was further aggregated into a monthly database. The station pair combined data was also aggregated into a similar train level summary for each month between FY 2005 and FY 2014. In addition, station level arrival and departure train performance data were collated over the course of a year for certain trains. The station level granularity of the data provided location-specific trends in Amtrak delays on the NEC, based on the stations where the delay started and how the delay propagated through subsequent stations on the corridor.

The discussion in Chapter 5 (focus on Acela) and Chapter 6 (focus on Northeast Regional) were based on analysis of the newly generated combined data sets. The data analysis was used to identify the days and trains on which delays were incurred, which are useful in separating systematic trends from the random components in the delays. Some of the interesting questions regarding service performance of Amtrak trains in the Northeast Corridor discussed in Chapter 5 and Chapter 6 include: What percentage of scheduled Amtrak trains and passengers experience delays? Are some trains more susceptible to service disruptions? Are there characteristic times of day, days of week, and months of year when most delays are encountered? Are there certain stations that experience more delays? etc. Another interesting analysis highlighted in Chapter 5 and 6 is on quantifying and evaluating the impact of a wide spectrum of service disruptions, ranging from catenary wire issues to severe weather conditions. These questions and considerations are important because they provide an assessment of how Acela and Regional services have performed in the last 10 years, and might shed light on major causes of delays and inform the discussion on prescriptive measures to mitigate any systematic components of service disruptions.

NOTE: The analyses in the following sections are only as good as the data used to analyze them. The author cleaned the dataset prior to analyzing; however, there were a few minor outstanding issues in the datasets. The main issue with the data relates to data coding. In the train operations database for example, there are some stations with missing information suggesting that the train either did not make a stop at the station or that the train personnel failed to record the actual arrival and departure times of the train on the given day. This affects the analysis because the missing records are sometimes represented as a train with zero delays, which is incorrect. Another example from the ridership database relates to the coding of riders on a train that was cancelled. There are a few times when ridership details are included in the demand data for a train coded as cancelled in the train operations database. While the entire datasets are not filled with these types of errors, they are present and do slightly impact the analysis shown. That said, because most of the analysis are presented as aggregates or averages over multiple days, they are less biased by single errors.

The next chapter presents and discusses on Acela ridership and service performance and their impacts on the distribution of delay.

5 Acela Ridership and Service Performance

The chapter examines factors that lead to ridership and service performance fluctuations on the Amtrak's Acela service.

Travel time distribution is a measure of day-to-day fluctuations in demand and service performance. The key factors causing demand and service performance variations discussed in this chapter are shown in Figure 5.1. The demand fluctuations are captured under four main categories: i) seasonality and month of year, ii) day of week, and time of day, iii) demand fluctuations due to capacity levels on the trains, and iv) demand variations in response to travel information and service quality. The service performance variations are captured under six main categories: i) seasonality and month of year variations, ii) day of week, and time of day differences, iii) service performance fluctuations due to capacity levels on trains, iv) service disruptions due to accidents and incidents (e.g. signal failures, weather related, track work, etc.), v) disturbances due to interference from other trains, and vi) performance variations due to administration, management and control elements. Each of these factors will be further discussed in the rest of this chapter. The double directional arrow connecting demand and service performance fluctuations reflects how ridership fluctuations could theoretically cause service performance fluctuations, and vice versa. For example, high demand during peak hour could lead to additional delays on the train as a result of the large number of people in the system.

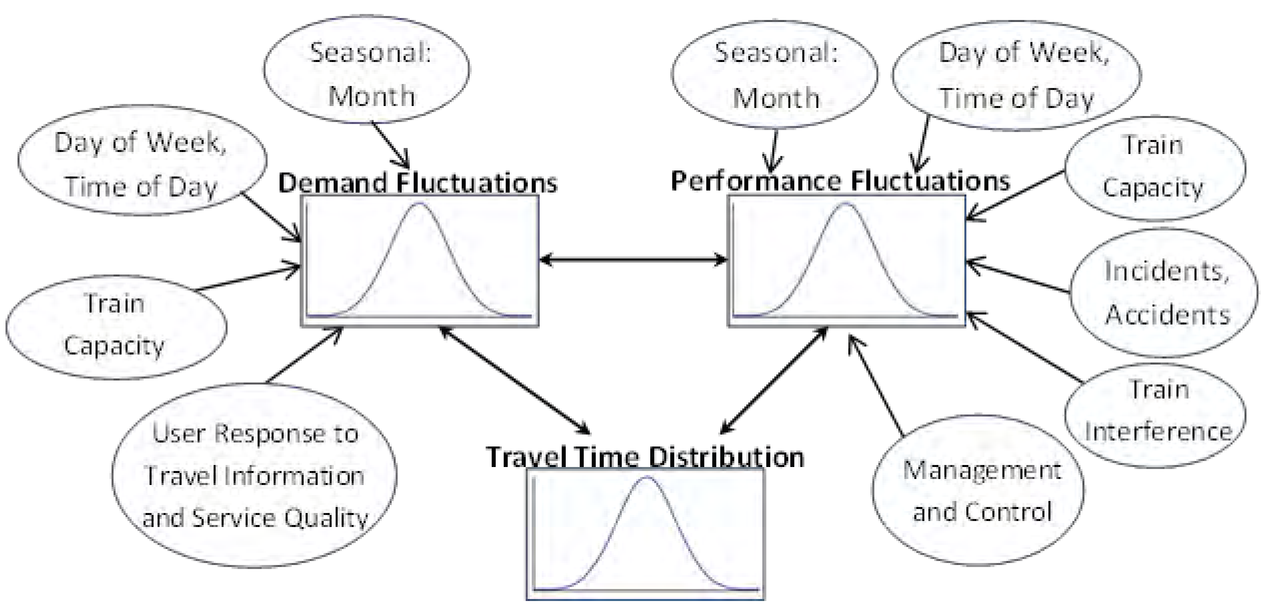


Figure 5.1: Factors Affecting Travel Time Distribution on the Acela service

The rest of this chapter is organized as follows. **Section 5.1** presents some performance metrics that characterize the travel time variability on the Acela service. Specifically, it shows the distribution of actual delays on Acela trains, and examines the relationship between delays and on-time performance. This section will directly answer the question about what metrics are useful in characterizing service performance. **Section 5.2** shows the annual and monthly fluctuations in Acela ridership and service performance between FY 2005 to FY 2014. **Section 5.3** drills further down into day of week and time of day variations in Acela ridership and service performance, and highlights any relationships between them. This section will directly answer the question about whether poor performance leads to even poorer performance. **Section 5.4** is on the First Train Analysis, which quantifies the effect of management and controls on service quality. It examines the delays on the first train of the day as a proxy of Amtrak train operator's culture and principle, as well as any inherent effects of Acela train timetables. **Section 5.5** investigates service disruptions caused by trains interfering with one another and how poor performance can cascade. **Section 5.6** focuses on service performance fluctuations due to accidents and incidents (e.g. signal failures, weather related, track work, etc.), and includes their effects on delays and train cancellations. **Section 5.7** analyzes the capacity on Acela trains between FY 2005 and FY 2014, which has the potential to affect both demand and supply variations. **Section 5.8** presents a preview of the impact of Acela rider's responses to service quality. Finally, **Section 5.9** concludes the chapter with a summary of the prior sections, and a discussion about what causes poor performance.

5.1 Performance Metrics Characterizing Travel Time Variability

Amtrak uses two performance metrics to quantify service performance – delay minutes and on-time performance (OTP) – however both metrics capture different aspects of performance. While the delay minutes reveal the **magnitude** of delay, the OTP indicates the **frequency** of good performance. Both metrics are discussed and compared in this section. The total end-to-end delay measures the difference between actual and scheduled travel times on Acela trains. The distribution of delays characterizes travel time variability, and indicates the degree to which actual end-to-end travel times deviate from the scheduled travel times. The OTP measures the percentage of trains that meet a certain 'on-time' standard, and indicates how often Acela trains attain the set standards. Both metrics highlight different perspectives of service performance but do each of them capture a substantive indication of service performance? This section shows the distribution of actual delays on Acela trains, and also examines the relationship between the delay minutes and on-time performance metrics.

5.1.1 Distribution of delay

Figure 5.2 shows the distribution of end-to-end delays encountered on the 9,604 Acela trains operated in FY 2014. Negative delay values (<0 minutes) represent arrivals that occur before the scheduled arrival times, while delay values equal to zero represent on-time train arrivals. Positive delays (>0 minutes) indicate trains that arrived after the scheduled arrival time. In FY 2014, delays encountered on Acela trains ranged from 1 minute to >100 minutes. The Unknown delay category represents train arrival records that are missing from the dataset either because the train was cancelled or the data was not recorded by the Amtrak crew (cancelled trains would be discussed in section 5.6.2). The delay distribution peaks at the zero-delay value and has a long right tail, representing a large proportion of Acela trains that arrive late. In FY 2014, 42% of Acela trains arrived earlier than or at the scheduled arrival time, and 58% of all scheduled Acela trains arrived at their final destination later than the scheduled arrival time. However, Amtrak's 'on-time' threshold for Acela trains currently includes trains that arrive within 10 minutes of the scheduled time. Under this classification, an additional 29% of FY 2014 trains would be categorized as 'on-time' as they arrived equal to or less than 10 minute late. Altogether, 71% of Acela trains arrived within 10 minutes of the scheduled arrival time, and 29% arrived more than 10 minutes late. The annual average OTP for FY 2014 was about 75%, and as one would expect it gives a rough indication of the delay distribution given the 'on-time' definition.

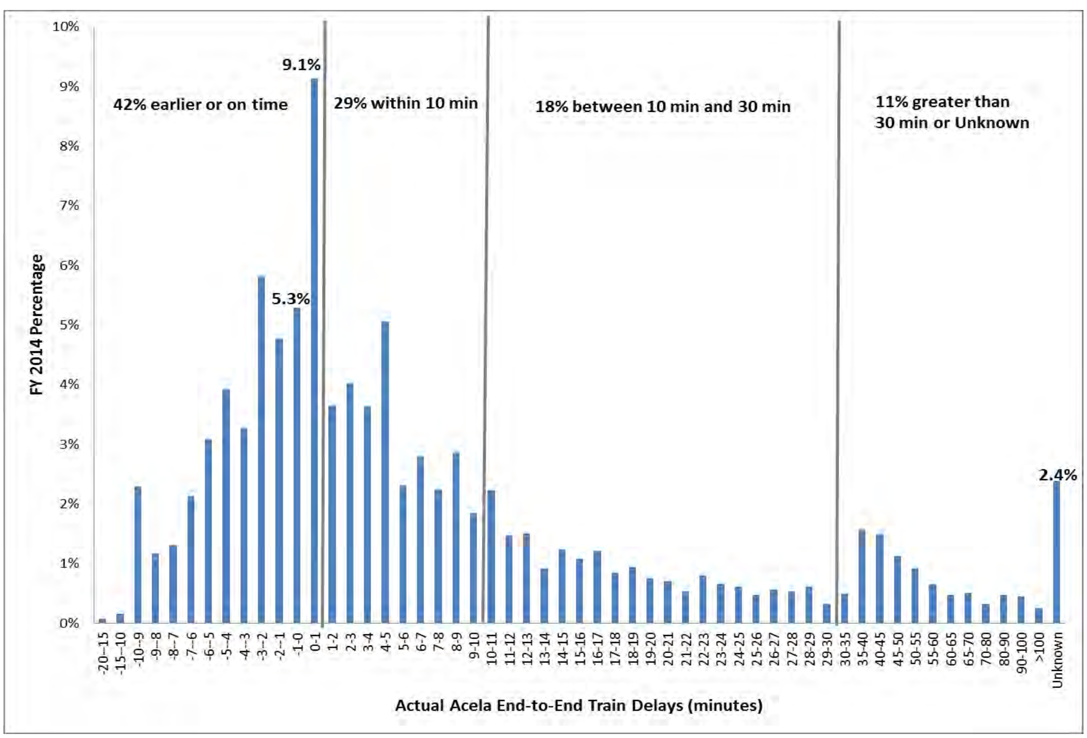


Figure 5.2: Distribution of Actual End-to-End Acela Train Delays

5.1.2 On-Time Performance and Delay Metrics

This section examines the relationship between the delay and on-time performance metrics, and measures the correlation between them. Figure 5.3 shows the relationship between the monthly on-time performance and average delay per train metric for Acela between FY 2005 and FY 2014. The blue and red data points distinguish between Pure OTP and Amtrak OTP. Pure OTP refers to a performance standard that categorizes a late train as one that arrives later than the scheduled arrival time, while Amtrak OTP refers to Amtrak's current performance standard that defines a late Acela train as one that arrives more than 10 minutes after the scheduled arrival time. In other words, the Pure OTP is a no-tolerance delay metric while the Amtrak OTP is a 10-minute delay tolerance metric. Although the difference between the Pure OTP and Amtrak OTP ranged from 16% to 34% between FY 2005 and FY 2014, the average, median and mode of the difference was about 29%. This suggests that while the Amtrak OTP metric is biased to portray a better service performance, the Pure OTP metric (zero-delay tolerance) for Acela can be quickly approximated as 29% lower.

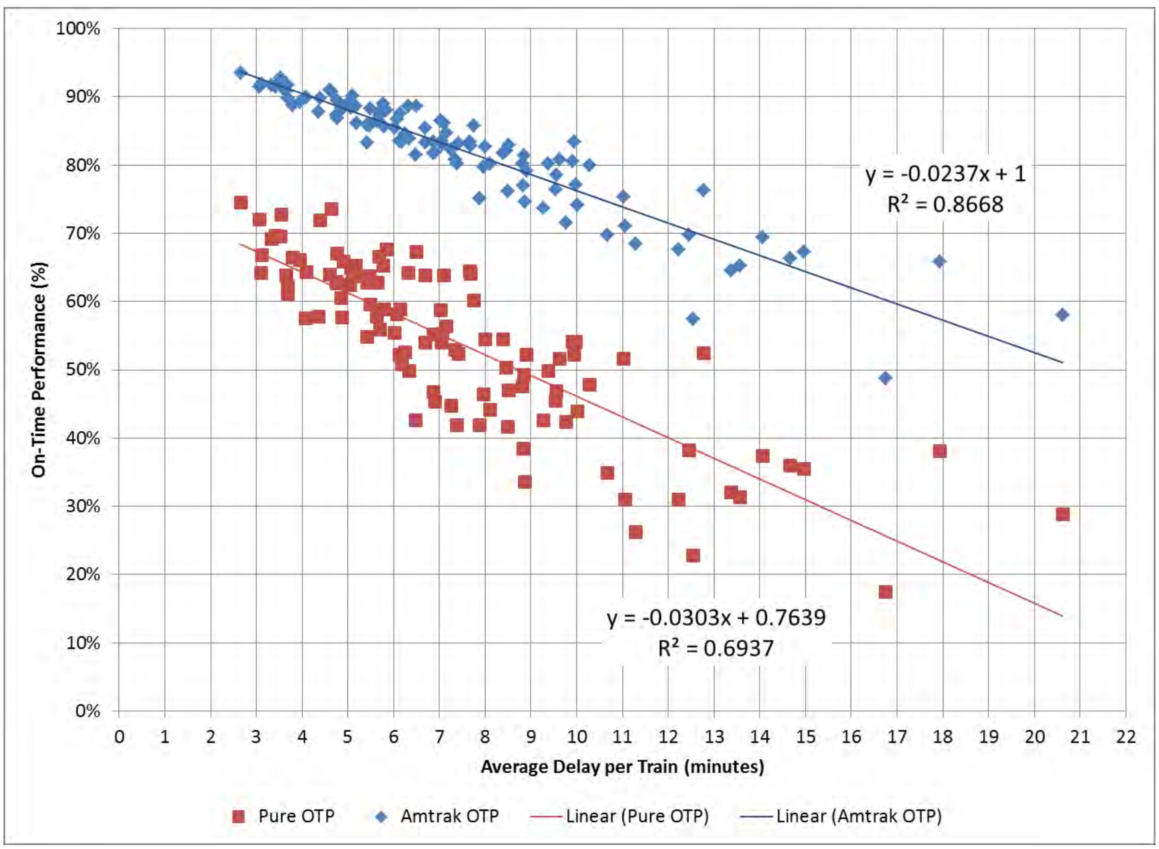


Figure 5.3: On-Time Performance versus Average Delay/Train

Furthermore, the simple regression shows a linear relationship between OTP and average delay per train, and includes the linear trend line and correlation factor. The correlation factor (R^2) between on-time performance (both Pure and Amtrak OTP) and average delay minutes per train is greater than or roughly equal to 70%. Although this correlation is not perfect (equal to 1), it is sufficiently high to propose that the two service quality indicators (on-time performance and delay) are related and can be substituted for one another if needed.

In summary, both on-time performance and average delay are useful in characterizing service performance, and in addition because they are correlated, thus can serve as substitutes for one another in characterizing service performance.

5.2 Acela Annual and Monthly Ridership, On-time Performance and Delays

There is a vast amount of valuable information that can be gleaned from analyzing Acela ridership and operations data over an extended period of time. However, prior to assessing any impacts on Acela ridership, it is important to evaluate annual trends and to account for certain systematic variations such as those introduced by seasonal and holiday factors, as well as one-time shocks or trends produced by economic factors. This section shows the annual and monthly fluctuations in Acela ridership and service performance between FY 2005 to FY 2014.

5.2.1 Annual Ridership

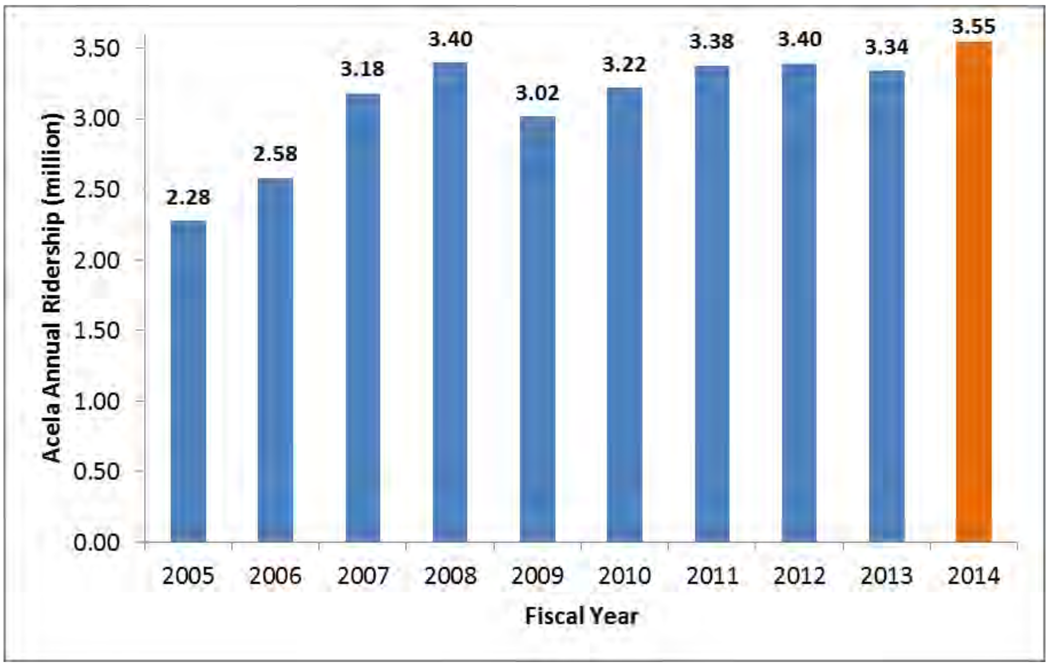


Figure 5.4: Acela Annual Ridership

Figure 5.4 shows the total Acela ridership between FY 2005 and FY 2014, highlighting the year-to-year variations. It shows a steady increase in total Acela ridership from 2.28 million to 3.40 million between FY 2005 and FY 2008. The year-over-year ridership growth rate was 13% between FY 2005 and FY 2006, 23% between FY 2006 and FY 2007, and 7% between FY 2007 and FY 2008. Adverse effects of the economic recession that lasted through 2008 and 2009 were likely responsible for the 11% drop in total Acela ridership to 3 million in FY 2009. The ridership increased again at a lower year-over-year growth rate between FY 2009 and FY 2014, and Acela experienced a record high of 3.5 million riders in FY 2014. However compared to FY 2008, the FY 2014 growths are modest. The variation in monthly ridership within each year is explored below.

5.2.2 Ridership by Month

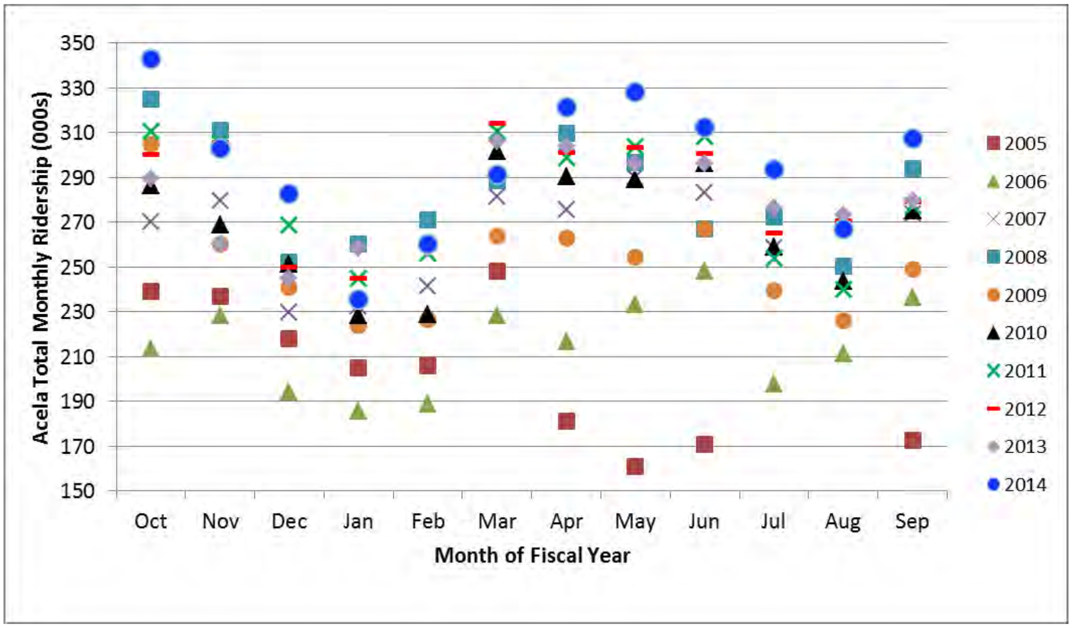


Figure 5.5: Acela Monthly Ridership FY 2005 to FY 2014

The seasons and holidays influence travel patterns and in turn, the monthly demand for Amtrak's Acela service. Figure 5.5 shows the total Acela ridership aggregated by month from FY 2005 to FY 2014. It compares total Acela ridership for a particular month from 2005 to 2014. The months with the highest ridership are in the fall months at the beginning of the fiscal year, overlapping with the start of the school year and New England's foliage season, and also during the spring months. Conversely, January and August typically have the lowest ridership likely due to vacation during the winter and summer holidays. It is important to note that the monthly ridership characteristics are not only due to seasonal trends but also include the effect of other external factors such as service performance, gas prices, economic indicators, unemployment rate, changes air fare, changes in Amtrak fares, etc.

5.2.3 On-Time Performance by Month

Amtrak monthly on-time performance (OTP) from one year to another reveals some inherent seasonal patterns as well. Figure 5.6 shows Acela on-time performance by month for FY 2005 to FY 2014 under the 10-minute delay tolerance standard. Unsurprisingly, the service performances in the winter months were largely dependent on the severity of the weather. One of the first noticeable features of the historical on-time performance is that the winter months - December, January, February and March had some extreme figures. For example, the Acela OTP in December and January FY 2005 (red squares) were as low as 57%, likely due to the impact of a severe winter. Quite surprisingly, the data shows that, the best service performances also occurred in the winter months (December, January, February and March), likely during mild winter conditions. The worst Acela service performances occurred in the summer months (June, July and August).

In terms of the year-over-year OTP variations for a given month, May and October (excluding the extreme figures in FY 2005) exhibited the least variance in year-over-year OTP, ranging from about 78% to 90% each year. Furthermore, comparing across years, the OTP in FY 2012 (red dashes) and FY 2007 (purple X's) were the highest, while the OTP in FY 2005 (red squares) and FY 2014 (blue dots) were the lowest. In FY 2014 the OTP for Acela was relatively one of the worst each month over the last 10 years – Acela experienced OTP as low as 65% in February and the highest OTP in FY 2014 was 81% in April. Conversely, the Acela performance in FY 2012 was one of the best each month between FY 2005 and FY 2014. It is also interesting to note that in the last 10 years, the OTP in May was never below 80% likely because of the favorable spring weather and minimal scheduled track work while the OTP in July was never above 86%, likely due to the larger amount of scheduled routine track work in the summer.

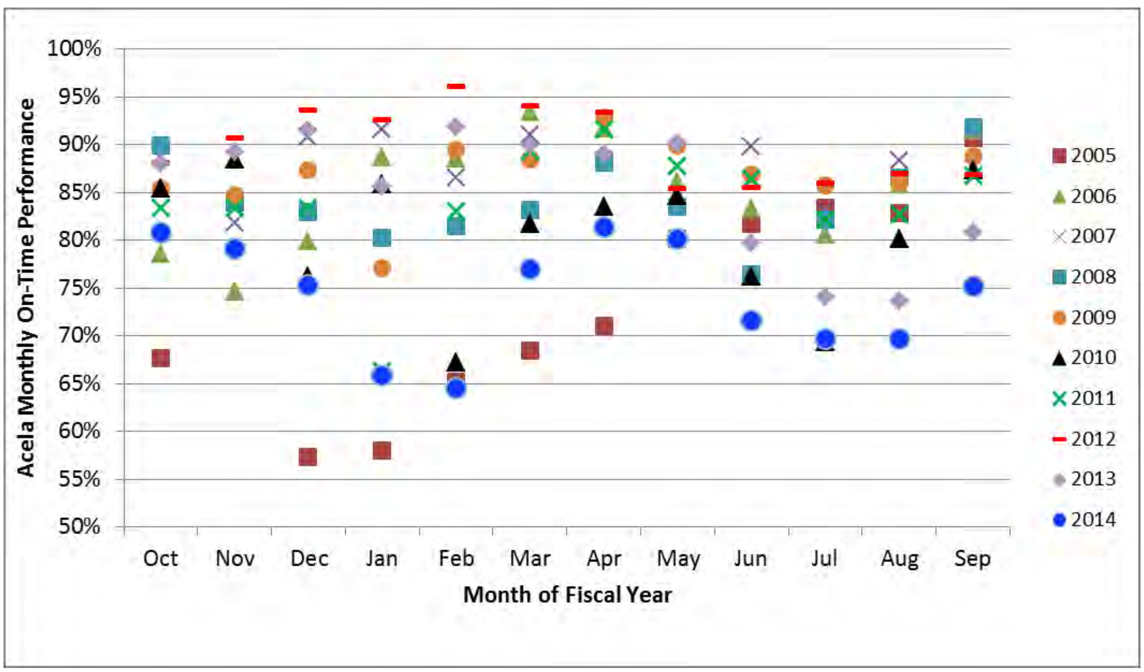


Figure 5.6: Monthly Acela On-Time Performance

5.2.4 Total Minutes of Delay by Month

Figure 5.7 shows total monthly delay for FY 2005 to FY 2014. The Acela delay shows characteristically similar trends but of course in reverse compared to the OTP. That is, months with high OTP show low total delay, and vice versa. Similar to the OTP trends, the winter months of January, February, and March historically experienced a wide variation in the total amount of delay (assuming the FY 2005 extreme figures in October and November are ignored). For example, Acela incurred about 15,000 minutes of total delay in January 2005 but only a total of 2,000 delay minutes in January 2012. Similar to the OTP data, the year-to-year variation of total monthly delay in October, November and May was relatively low, ranging from about 3,000 and 7,000 total delay minutes (again assuming the FY 2005 extreme figures in October and November are ignored).

Overall, the total monthly delay in FY 2012 (red dashes) and FY 2007 (purple X's) were the lowest (that is best performance-wise), while the total monthly delay in FY 2005 (red squares) and FY 2014 (blue dots), were the highest (that is worst performance-wise). The total monthly delay for Acela in FY 2014 was one of the highest with delays ranging from 6,100 in September to 12,000 minutes in January. In the last 10 years, the worst delay experienced on Acela was in January FY 2005; the second worst was in January 2014. On the other hand, the best Acela performance was in February 2012 and the second best were in March 2006 and March 2012.

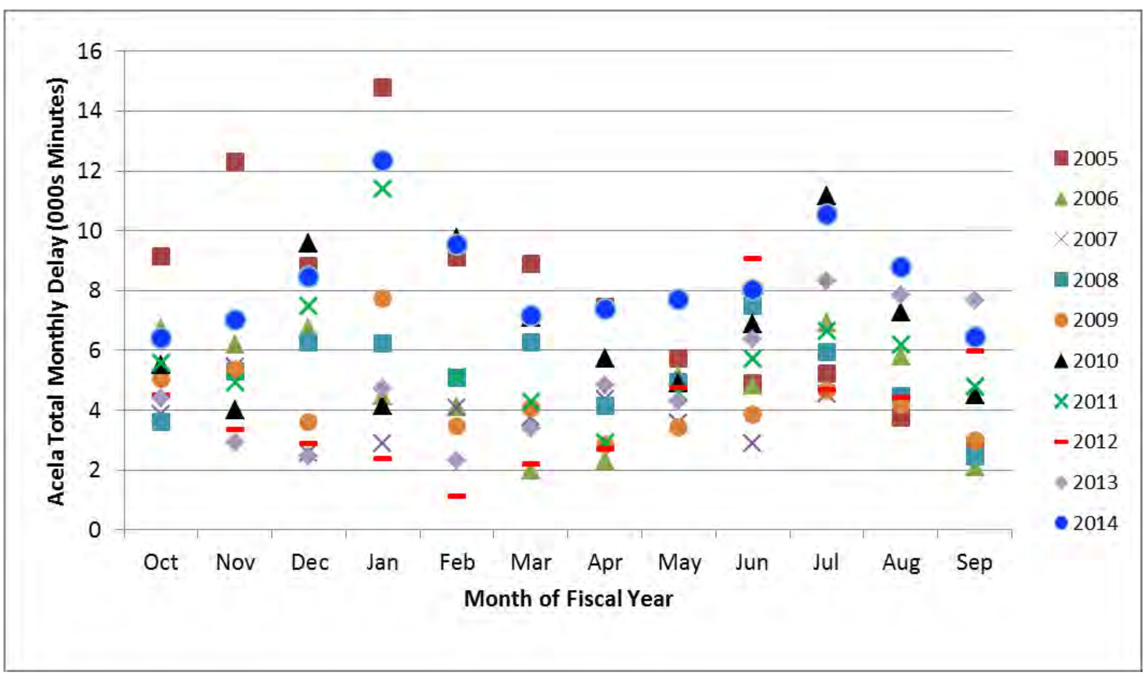


Figure 5.7: Total Monthly Delay

5.2.5 Monthly Ridership versus Monthly Delay

Figure 5.8 shows total monthly ridership (blue) and total monthly delay (orange) for Acela trains between FY 2005 and FY 2014. The vertical gridlines indicate the beginning of a fiscal year. The correlation coefficient (R^2) measures the strength and direction of the linear relationship between ridership and delay each month within each fiscal year. Correlation coefficients close to zero indicate a weak linear relationship between the ridership and delay in the same month, while correlations coefficients equal to 1 represent a perfect linear relationship between both variables. Additionally, negative R^2 values denote an inverse linear relationship between the demand and delay, and suggest that months with high demand are associated with low delays, and vice versa. The fiscal years with negative R^2 values are highlighted in red. For example, in FY 2014, the correlations coefficients was equal to -0.79, revealing a substantial correlation between low ridership volumes and high delays in months like January, and high ridership volumes and low levels of delay in months like May. It is important to note that these are not cause-effect relationships but are simply observations from the data. The fiscal years with positive R^2 values are highlighted in black –excluding FY 2005 (for same reasons as noted above), the R^2 values were all less than 0.5, indicating weak linear trends between high ridership and high levels of delay in the same month. Considering all, having more people in the system was sometimes associated with poor service performance.

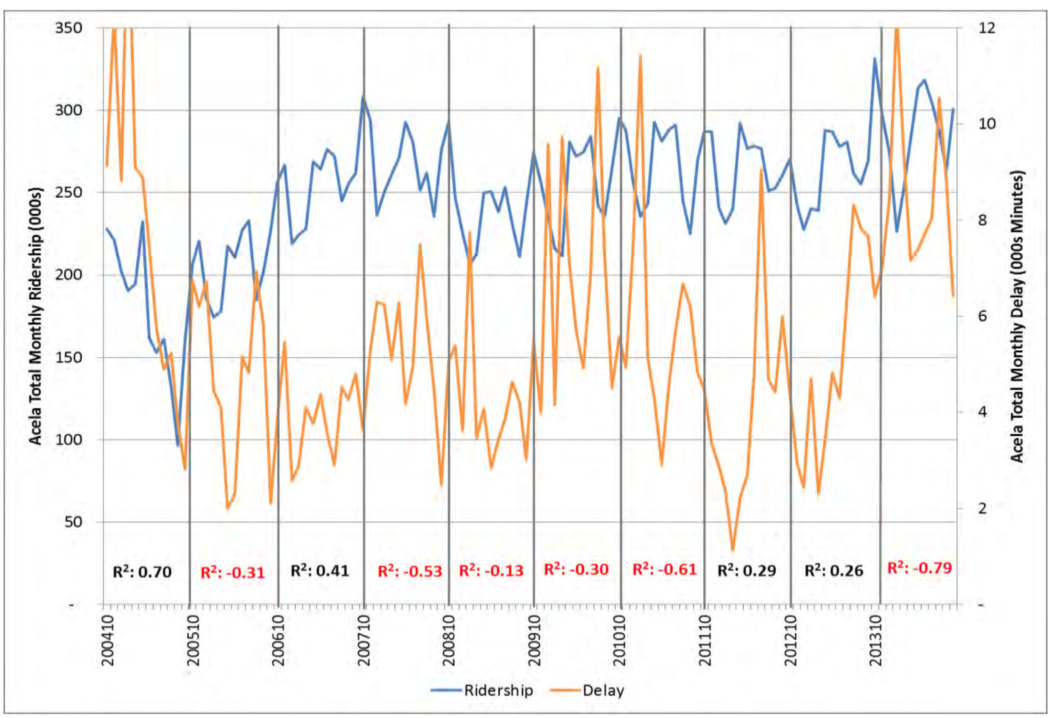


Figure 5.8: Monthly Ridership versus Total Delay

In summary,

i. In FY 2014 Acela service experienced a record high of 3.55 million riders

ii. The months with the highest Acela ridership are in the fall months at the beginning of the fiscal year, overlapping with the start of the school year and New England's foliage season, and also during the spring months. Conversely, January and August typically have the lowest ridership likely due to vacation during the winter and summer holidays.

iii. Similar to the ridership, the performance on the Acela service varied within the same month in different years. Compared to summer months, the winter months exhibited a large variance due to the effects of mild and severe winter seasons on performance. Nonetheless, the best on-time performances were usually in the winter months.

iv. Additionally, the performance on the Acela service varied across different months in the same fiscal year. Although the summer months exhibited less year-to-year variance, on a month-to-month comparison (excluding the severe winter months), the amount of delays were usually higher in the summer, and especially in July due to routinely scheduled track works, heat restrictions and infrastructures issues (catenary wire drooping). October and May exhibited the least year-to-year variance likely because of the favorable weather (fall and spring) and minimal scheduled track work.

v. Between FY 2005 and FY 2014, the best Acela performance was in FY 2012 while the worst performance was in FY 2014.

vi. The ridership to performance correlations suggests that having more people in the system was sometimes associated with poor service performance.

5.3 Daily Variations in Performance and Ridership

The travel time distribution on Acela trains is impacted by the day-to-day fluctuations on both the demand side and the supply side. On the demand side, the number of daily riders can be expected to vary by day of week (especially weekday versus weekend) and time of day (especially peak hours versus off-peak hours). On the supply side, there might also be systematic variations in the levels of delay and on-time performance by day of week and time of day. On any given day, on one hand, high demand might be associated with days with low levels of delays because riders choose to make trips on days with low delays based on past experiences, but on the other hand, high demand might lead to higher levels of delay because there are too many riders in the system. In reality, the Amtrak system probably experiences some level of both of these trends.

The analysis in this section focuses on the fiscal years with the best and worst performance. This is because unlike the monthly data, which can effectively be condensed into meaningful charts and discussions, the sheer amount data points contained in all days between FY 2005 and FY 2014 does not afford the same luxury. Furthermore, the systematic portion of daily trends can be understood from analyzing the averaged over all the days in any one given year. Regarding the use of the years on both extremes of service performance, the daily trends that are influenced by service performance can be captured by analyzing both extremes and assuming that the intermediate years lie somewhere in-between. From the earlier discussion, the overall Acela performance appeared to be the best in FY 2012 and the worst in FY 2014. Using the daily on-time performance, total delay and ridership values in FY 2012 and FY 2014, this section further explores and compares the distribution of end-to-end delays encountered on Acela trains.

The section begins by presenting the number of daily delayed trains and daily delayed riders on the Acela in FY 2012 and FY 2014. It then drills further down into day of week and time of day variation in ridership and service performance in both years, and highlights any relationships between them.

5.3.1 FY 2012 and FY 2014 Daily Delays: Trains

Figure 5.9 and Figure 5.10 show daily Acela scheduled and delayed trains in FY 2012 and FY 2014, respectively. The gray area shows the total number of scheduled trains. In FY 2012, there were 32 scheduled Acela weekday (Monday to Friday) trains while in FY 2014, there were 33 scheduled weekday trains. In both years, there were 9 Acela trains scheduled on Saturdays and 19 Acela trains on Sundays. The lower weekend schedules are portrayed in both figures as spaces between the high weekday peaks. Furthermore, there are usually fewer trains scheduled during the winter holiday between the weeks including December 25th and January 1st. Finally, the spike in both figures occurs on the days before Thanksgiving on which Amtrak typically schedules additional trains.

The blue area represents the total number of trains that arrived at their final destination with a delay greater than 0 minutes. In FY 2012, about 35% of all scheduled trains arrived at their final destination after the scheduled arrival time, while in FY 2014, about 56% of all scheduled trains arrived at their final destination after the scheduled arrival time. In other words, of the 32 scheduled daily weekday trains in FY 2012, about 11 on average were routinely late, and of the 33 scheduled trains in FY 2014, on average 18 trains were routinely late.

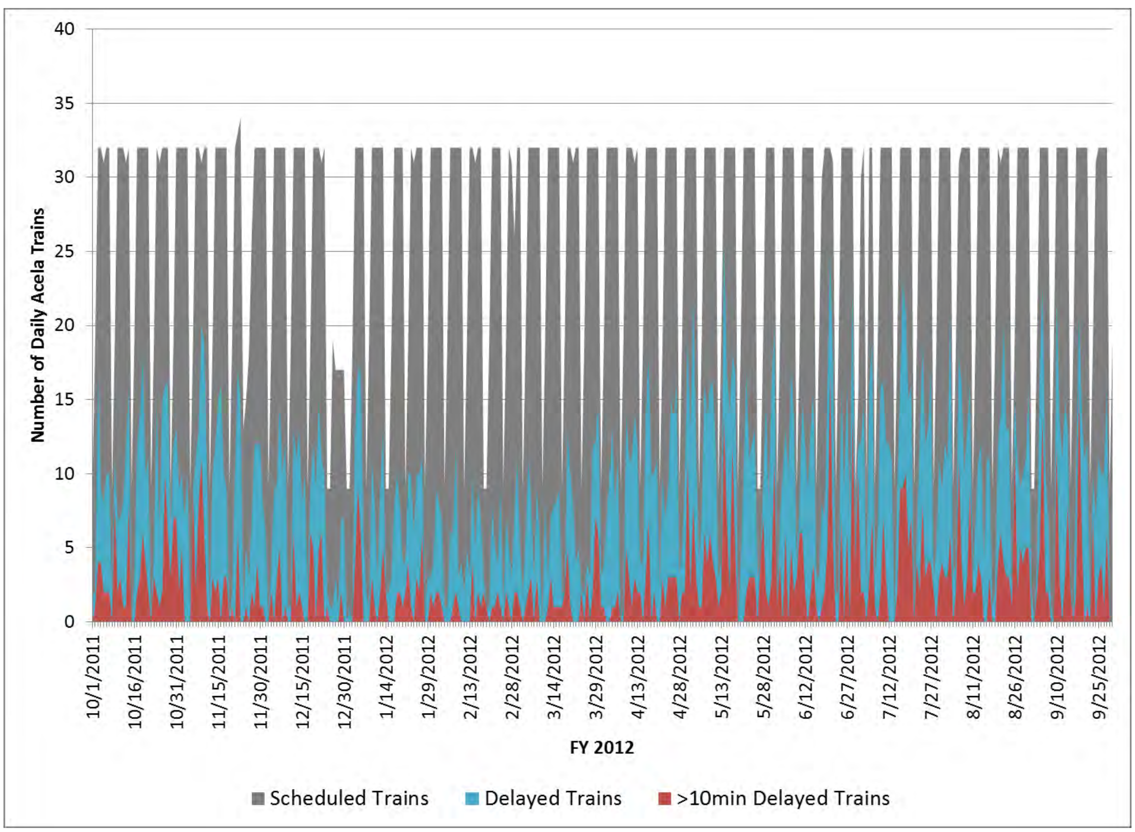


Figure 5.9 FY 2012 Scheduled and Delayed Acela Trains

Given Amtrak's 10-minute 'on-time' threshold, the red area represents the number of "late" trains by Amtrak's standards. In FY 2012, roughly 10% of trains (on average 3 of 32) and in FY 2014, roughly 26% of trains (on average 8 of 33) experienced delays greater than 10 minutes.

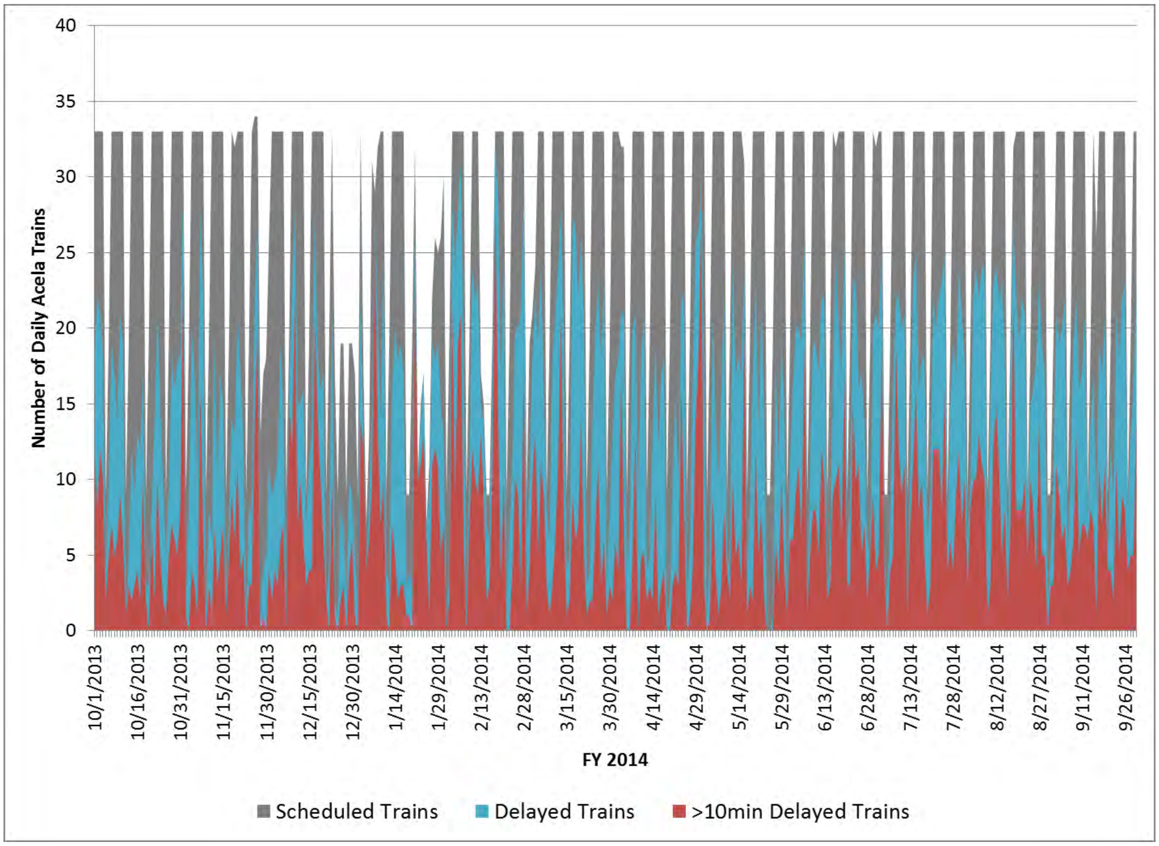


Figure 5.10: FY 2014 Scheduled and Delayed Acela Trains

5.3.2 FY 2012 and FY 2014 Daily Delays: Riders

Figure 5.11 and Figure 5.12 show the total number of daily and delayed Acela riders in FY 2012 and FY 2014, respectively. The gray area corresponds with the total number of daily riders. Firstly, the high weekday ridership trends versus the low weekend ridership are depicted as high peaks and the spaces between the high peaks. In both years, the lowest Acela ridership occurred on the days between Christmas and the New Year, which corresponds to the days with fewer scheduled Acela trains (as shown in Figure 5.9). Comparing Figure 5.11 and Figure 5.12, the gray regions also highlight the growth in ridership between FY 2012 and FY 2014. While in FY 2012, the daily ridership exceeded 12,000 on a few days, in FY 2014, it regularly exceeded 12,000, and even exceeded 14,000 occasionally.

In Figure 5.11 (FY 2012) and Figure 5.12 (FY 2014), the blue area shows the number of daily riders that arrived at their destination after the scheduled arrival time, and the red area shows the number of daily riders that arrived at their destination more than 10 minutes late. In FY 2012, 44% of the 3.4 million Acela passengers experienced delays, and 10% experienced delay greater than 10 minutes. In actual values, that corresponds to about 1.4 million delayed passengers and about 326,000 passengers experiencing delays greater than 10 minutes in FY 2012. In comparison, in FY 2014, the blue area almost overlaps entirely with the grey area suggesting that almost all Acela riders experienced delays. 2.3 million (66%) of the 3.55 million Acela riders arrived at their destination late, and 916,000 passenger (27%) arrived at their destination more than 10 minutes late.

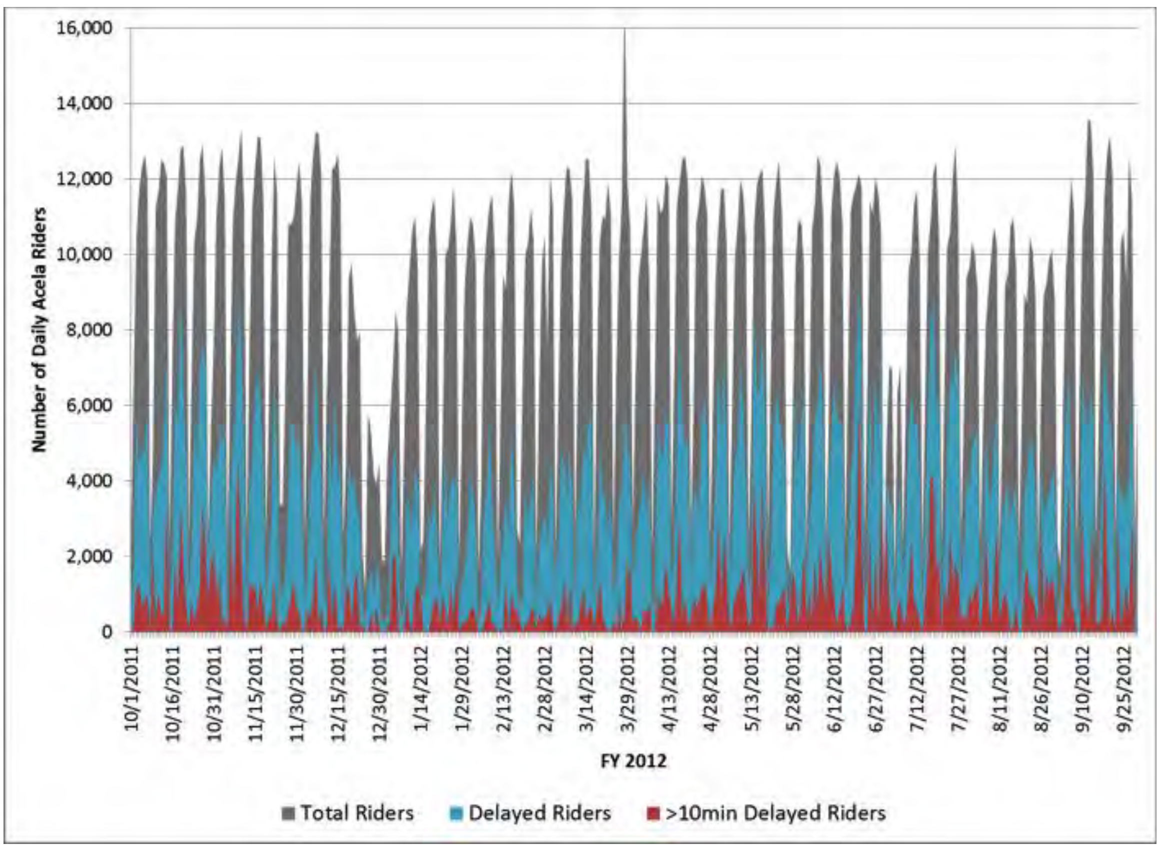


Figure 5.11: FY 2012 Scheduled and Delayed Acela Riders

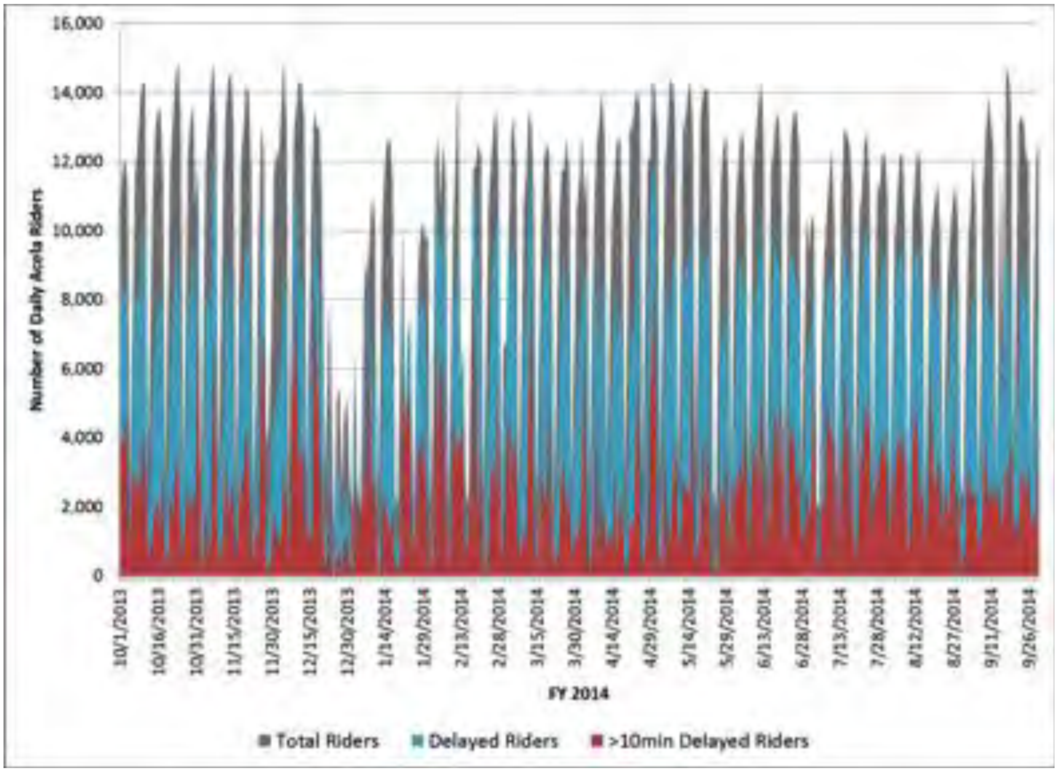


Figure 5.12: FY 2014 Scheduled and Delayed Acela Riders

In summary, even in the best performing year, FY 2012 as many as 35% of scheduled trains (11 of 32) and 44% of traveling passengers (1.4 million passengers) arrived at their final destination after the scheduled arrival time, and as many as 10% (3 of 32) of trains and 10% of passengers arrived late with delays greater than 10 minutes. By FY 2014, the numbers of late Acela trains and late Acela passengers had grown to about 56% of trains (18 of 32) and about (66%) of riders (2.3 million riders).

5.3.3 Day of Week Performance

Figure 5.13 shows the average daily ridership (dotted lines) and average daily delay per train (solid lines) on Acela trains by day of week for FY 2012 (in blue), FY 2014 (in red), and averages over FY 2005 to FY 2014 (in green). The averages over FY 2005 to FY 2014 were included in this analysis to ensure that random disruptions or calendar effects that might have affected Acela operations and service performance on a specific weekday did not bias the day of week patterns. Furthermore, the FY 2005 to FY 2014 values were likely more representative of true day of week performances since they reflect averages over many more days (there were 523 Mondays between FY 2005 and FY 2014 as opposed to only 53 Mondays in FY 2012 or FY 2014).

On the demand side, the average weekend ridership on Acela was significantly lower than the average weekday ridership in both FY 2012 and FY 2014, as well as over the 10-year period (FY 2005 – FY 2014). The average ridership was typically lowest on Saturdays. In FY 2012, the average daily ridership on Saturdays was 2,300 compared to 5,600 on Sundays, and 10,600 on weekdays. In FY 2014, the average daily ridership on Saturdays was 2,600 compared to 6,200 on Sundays, and 11,500 on weekdays. Of the weekdays, the ridership on Wednesdays and Thursdays were typically the highest, followed by Tuesdays, Friday, and lastly, Mondays.

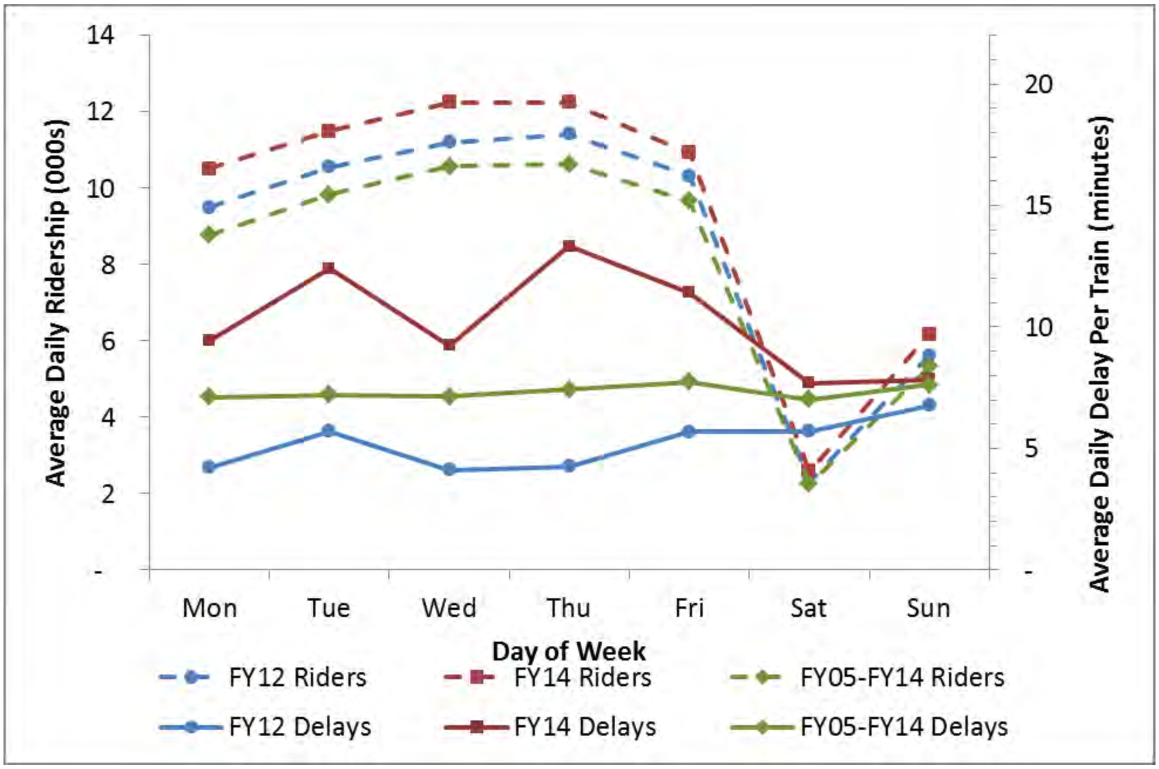


Figure 5.13: Average Ridership and Delay by Day of Week

On the delay side (solid lines), even though the average delay per train appeared to vary by day of week in FY 2012 and FY 2014, the averages over the 10-year period suggest that all weekdays and Sundays experience the same level of delays, while the delays on Saturdays were usually slightly lower. This implies that higher demand on certain days of the week did not lead to additional delays.

Figure 5.14 shows a similar chart (to Figure 5.13) but with average on-time performance instead of average delay per train by day of week. The dotted lines represent average daily ridership and the solid lines represent average on-time performance (OTP). Although the average OTP in FY 2014 (red) shows large variation ranging from 69% on Tuesdays to 81% on Saturday, it trended towards a flat line in both FY 2012 (blue) and over the 10-year period (FY 2005 – FY 2014 in green). Although the average OTP in FY 2012 and between FY 2005 and FY 2014 was slightly higher on Wednesdays and slightly lower on Fridays, the ranges were under 5% such that we can assume that weekday variations are trivial. However, similarly to the average delay, Acela OTP was significantly better on Saturdays, thus we can safely assume Acela service performance is best on Saturdays. This assumption is plausible since both Amtrak and commuter services usually have fewer trains scheduled on Saturdays.

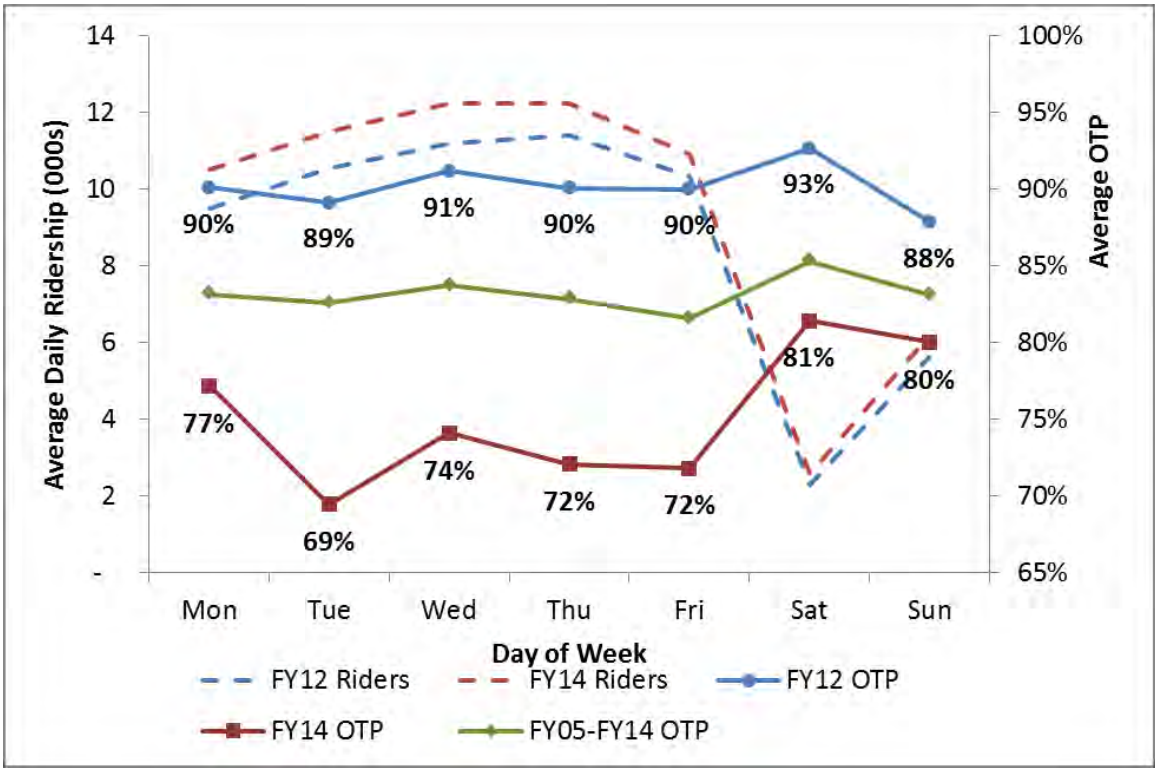


Figure 5.14: Average Ridership and OTP by Day of Week

In summary, Acela ridership volumes were lowest on Saturdays and highest on Wednesdays and Thursdays, likely an artifact of weekday business travel patterns. Additionally, of the weekdays, Mondays had the lowest ridership. In terms of service quality, performance appeared to be roughly the same on all weekdays and was usually slightly better on Saturdays. The Saturday performance improvement was likely because fewer Amtrak and Commuter services operate on weekends. Since days with higher ridership did not appear to have worse performance, it implies that higher demand on certain days of the week did not lead to additional delays.

5.3.4 Time of Day Performance

This section explores the performance of Acela trains by time of day. The Northeast Corridor spans 457 miles, and as such trains typically cross multiple time periods between departure at origin station and arrival at the terminating station. As a result, the time of day analysis in this section is based on train departure times by direction and originating station.

Table 5.1 shows the average Amtrak OTP (10-minute delay threshold) and average delay per train (no threshold) for Acela trains on weekdays in FY 2012 and FY 2014. The averages are over the 261 weekdays in both years. The first half of the table shows the Acela trains that travel southbound on the Northeast Corridor while the second half shows the northbound trains. The trains by direction are further categorized by originating-terminating station and departure times. In the southbound direction, the trains are grouped into BOS-WAS, through trains that operate between Boston South Station (BOS) and Washington Union Station (WAS), and NYP-WAS, south-end only trains that originate in New York Penn station (NYP) and terminate in WAS. Note that, the BOS-WAS trains also stop in NYP and as such serve the south-end (NYP-WAS) markets as well (see detailed explanation in Section 4.2). The BOS-WAS trains are scheduled to travel the 457-mile stretch in about 6hr 30min, including a roughly 15 minute dwell time for the train in NYP, while the NYP-WAS trains are scheduled to travel the 226-mile stretch in under 3 hours (about 2hr 50min).

Intuitively, the BOS-WAS through trains are likely to encounter more delays than the NYP-WAS south-end only trains since they longer travel times, travel double the distance, and stop at many more stations. The empirical data confirms this expectation showing that the SB BOS-WAS Acela trains had approximately 6% lower OTP on average in FY 2012, and 20% lower OTP on average in FY 2014. The BOS-WAS trains also encountered double the amount of delay than on the NYP-WAS trains in both years. This gives an impression that delays are accumulated along the length of the NEC, as opposed to being concentrated in certain segments. This will be further discussed in Section 5.4.

Regarding time of day performance, the performance of Acela SB trains deteriorated through the day; morning trains usually encountered fewer delays than afternoon trains. In addition, not surprisingly, trains scheduled to depart during the afternoon peak periods experienced the highest amount of delays.

Weekdays Only (261/365 days)					FY 2012		FY 2014	
Train No	Route	Direction	O-D	Departure Time	Amtrak OTP*	Avg. Delay per train (min)	Amtrak OTP*	Avg. Delay per train (min)
2151	Acela	SB	BOS-WAS	5:05:00 AM	92%	4.11	72%	11.09
2153	Acela	SB	BOS-WAS	6:05:00 AM	95%	2.42	77%	10.68
2155	Acela	SB	BOS-WAS	7:15:00 AM	93%	3.67	75%	10.12
2159	Acela	SB	BOS-WAS	9:15:00 AM	87%	5.16	67%	11.65
2163	Acela	SB	BOS-WAS	11:10:00 AM	86%	5.29	66%	14.88
2165	Acela	SB	BOS-WAS	12:10:00 PM	86%	5.47	65%	14.31
2167	Acela	SB	BOS-WAS	1:10:00 PM	84%	6.45	59%	18.58
2171	Acela	SB	BOS-WAS	3:10:00 PM	81%	12.33	61%	15.41
2173	Acela	SB	BOS-WAS	4:15:00 PM	90%	9.08	49%	17.59
2175	Acela	SB	BOS-WAS	5:20:00 PM			70%	17.22
2103	Acela	SB	NYP-WAS	6:00:00 AM	96%	2.23	89%	4.84
2107	Acela	SB	NYP-WAS	7:00:00 AM	96%	2.44	90%	5.68
2109	Acela	SB	NYP-WAS	8:00:00 AM	95%	3.62	78%	9.45
2117	Acela	SB	NYP-WAS	12:00:00 PM	94%	3.75	83%	6.02
2121	Acela	SB	NYP-WAS	2:00:00 PM	91%	3.76	89%	4.61
2119	Acela	SB	NYP-WAS	6:00:00 PM	93%	4.08	80%	9.41
2190	Acela	NB	NYP-BOS	6:20:00 AM	91%	3.96	87%	6.38
2100	Acela	NB	WAS-NYP	6:00:00 AM	87%	6.59	70%	10.96
2104	Acela	NB	WAS-NYP	8:00:00 AM	94%	2.63	79%	6.50
2110	Acela	NB	WAS-NYP	11:00:00 AM	98%	1.07	90%	3.96
2122	Acela	NB	WAS-NYP	5:00:00 PM	90%	5.08	87%	6.44
2124	Acela	NB	WAS-NYP	6:00:00 PM	89%	5.73	67%	12.58
2126	Acela	NB	WAS-NYP	7:00:00 PM	91%	5.29	83%	10.15
2128	Acela	NB	WAS-NYP	8:00:00 PM			79%	9.59
2150	Acela	NB	WAS-BOS	5:00:00 AM	81%	6.61	63%	16.00
2154	Acela	NB	WAS-BOS	7:00:00 AM	89%	4.22	63%	13.19
2158	Acela	NB	WAS-BOS	9:00:00 AM	91%	3.37	66%	12.70
2160	Acela	NB	WAS-BOS	10:00:00 AM	90%	3.71	66%	14.94
2164	Acela	NB	WAS-BOS	12:00:00 PM	93%	5.30	76%	8.97
2166	Acela	NB	WAS-BOS	1:00:00 PM	87%	6.16	81%	7.51
2168	Acela	NB	WAS-BOS	2:00:00 PM	90%	4.50	78%	9.31
2170	Acela	NB	WAS-BOS	3:00:00 PM	83%	7.21	55%	17.00
2172	Acela	NB	WAS-BOS	4:00:00 PM	81%	8.05	51%	17.96

*Amtrak OTP for the longest segment with passengers, which is not necessarily the end point station

Table 5.1: FY 12 and FY 14 Weekday OTP and Average Delay per Train

Continuing with the southbound discussion, in FY 2012, the 3:10PM BOS-NYP train (Train 2171) exhibited the worst performance with an average OTP of 81% and an average delay per train of 12 minutes. Likewise, in FY 2014, the trains departing BOS at 1:10PM (Train 2167), 3:10PM (Train 2171), 4:15PM (Train 2173), and 5:20PM (Train 2175) all exhibited the worst performance with average delays greater than 15 minutes and OTP less than 60% (except, Train 2175 which had an average OTP of 70%).

For the southbound trains departing NYP, in FY 2012, the 2:00PM train (Train 2121) exhibited the worst performance with an average OTP of 91% and average delay of 4 minutes. In FY 2014, the morning peak train departing NYP at 8:00AM (Train 2109) and the PM peak train departing NYP at 6:00PM (Train 2119) both experienced the worst performance with OTP less than 80% and average delays greater than 9 minutes.

Now focusing on the second half of Table 5.1, the northbound trains are also organized by O-D station and departure time of day. The trains are grouped into NYP-BOS trains (north-end), WAS-NYP trains (south-end), and WAS-BOS trains (through). Again, the WAS-BOS through trains serves both the north-end and south-end markets.

Similarly to the southbound trains, the longer through NB trains (WAS-BOS) performed worse than the south-end only trains (WAS-NYP) or the north-end only trains (NYP-BOS). In addition, trains that departed during the morning and afternoon peak periods had the worst performance, and experienced the highest amount of delays. In fact, the first northbound train departure at 5:00AM from WAS (Train 2150) appeared to have one of the worst performances in both years. And within the WAS-BOS trains, the PM peak trains departing from WAS at 3:00PM (Train 2170) and 4:00PM (Train 2172) also experienced higher delays on average in both years. In the WAS-NYP section, the AM peak train, Train 2100 departing WAS at 6:00AM and the PM peak train, Train 2124 departing WAS at 6:00PM exhibited the worst performance, in both years.

In summary, Acela trains that departed during AM peak periods (5:00AM to 8:00AM) or during the PM peak periods (3:00PM to 6:00PM) tended to have worse performance. Additionally, Acela trains exhibited distance-related deteriorations, suggesting that delays accumulate as trains served more stations along the length of the corridor. Furthermore, the first Acela train of the day was not able to achieve consistent on-time arrivals and encountered considerable amount of delays even though there were no other trains or at least fewer trains in the system to slow them down. Consequently, a detailed station-level analysis of the first Acela trains of the day is discussed in Section 5.4

5.4 First Train Analysis

This section examines the first train of the day in each direction within each market (north-end, south-end and through). In a system where trains interfere with one another, one would expect to see better performing trains in the morning when there were either no other trains or fewer trains in the system to slow each other down, and possibly performance deterioration through the day as more trains entered the system. However, Table 5.1 shows this is not the case in Amtrak's Acela system, and the first train of the day routinely encounters a significant amount of delays. Consequently, this section attempts to track the detailed station-to-station arrivals and departures of the first Acela trains over multiple days to identify causes of delay. The apriori assumption is that the findings in this section might serve as a proxy to quantify the effect of management and controls on service performance.

The discussion is divided into four subsections: i) northbound departures from New York Penn Station (NYP), ii) northbound departures from Washington Union Station (WAS), iii) southbound departures from New York Penn Station (NYP), and iv) southbound departures from Boston South Station (BOS). In each section, the average arrival and departure delays for the train at successive stations on the corridor are presented. The delays are calculated as differences between the actual and scheduled arrival and departure times of the train, and are averaged over the 261 weekdays of the year. In the delay calculations, Amtrak is not credited for 'negative' delays, and arrivals that occurred before the scheduled arrival time are treated as on-time arrivals with zero delays.

5.4.1 First <u>Northbound</u> Acela Train From NYP - Train 2190 (6:20AM)

Train 2190 is the first northbound Acela train scheduled to depart NYP at 6:20AM and arrive at BOS at 10:05AM on weekdays. In FY 2014, it had an on-time performance of 87% and average delay of 6 minutes. Figure 5.15 shows the average FY 2014 arrival and departure delays for the train at each successive station on the corridor. The light blue bars represent delays in arrivals while the dark blue bars represent departure delays.

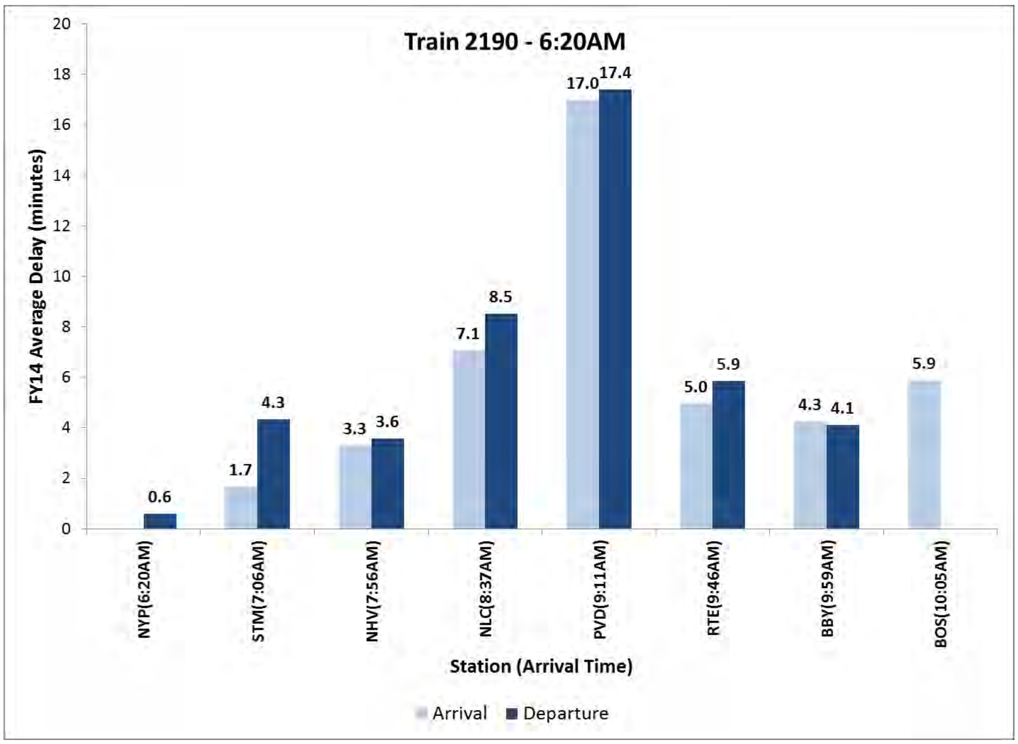


Figure 5.15: FY 2014 Average Station Delays for Train 2190

Overall, the dark blue bars are higher than the light blue bars, that is, trains typically accumulated additional delays between arrival at a station and departure from the same station. This is likely due to the fact that Amtrak does not schedule terminal time at all stations to account for the time it takes for passenger to get on and off the trains. There are no scheduled dwell times for Train 2190.

On average, Train 2190 departed NYP almost a minute later than scheduled, arrived at Stamford, CT (STM) about 2 minutes late, and had accumulated about 4 minutes of delay by the time it departed its second station (STM). The highest amount of delay seemed to accrue in the 62-mile segment between New London, CT (NLC) and Providence, RI (PVD) but the train seemed to recover from a substantial amount of the accrued delay in the 32-mile segment between PVD and Route 128, MA (RTE). On average, Train 2190 arrived at BOS 6 minutes later than the scheduled time.

Table 5.2 shows the first train analysis from a segment level and timetable point of view. A segment refers to the portion of the corridor between successive stations. It includes the segment distance, scheduled departure time from and arrival time at the stations in the segment, as well as total accumulated segment delays. The next set of columns in the table show the segment travel time (scheduled, average actual, and difference), and finally the last set of columns show segment travel speeds (scheduled, average actual, and difference). The data in each column are further clarified in the highlighted example below.

Focusing on the 62-mile NLC-PVD segment, Acela Train 2190 was scheduled to depart from NLC at 8:37AM and arrive at PVD at 9:11AM. The delay shown corresponds with the station arrival delays shown in light blue in Figure 5.15, which indicates that by the time Train 2190 arrived PVD, it had accumulated 17 minutes of delays. The next set of columns show that although the train was scheduled to travel the NLC-PVD segment in 34 minutes, on average, it actually took 42 minutes, 8 minutes slower than scheduled. Correspondingly, the scheduled travel speed in the NLC-PVD segment was 109 mph but on average, the train travelled at 88 mph, 22 mph slower than scheduled. It is interesting to notice that the scheduled travel speed in the NLC-PVD segment (109mph) is the fastest, 31 mph faster than the next highest (NHV-NLC: 98mph).

Train 2190		Scheduled		*Delay (min)	Segment Travel Time (min)			Segment Travel Speed (mph)		
Segment	Distance (miles)	Departure Time	Arrival Time		Scheduled	Avg. Actual	Actual-Scheduled	Scheduled	Avg. Actual	Actual-Scheduled
NYP-STM	36	6:20 AM	7:06 AM	1.3	46	47	1	47	46	-1
STM-NHV	39	7:06 AM	7:56 AM	2.6	50	49	-1	47	48	1
NHV-NLC	51	7:58 AM	8:37 AM	6.8	39	43	4	78	72	-6
NLC-PVD	62	8:37 AM	9:11 AM	17.0	34	42	8	109	88	-22
PVD-RTE	32	9:11 AM	9:46 AM	2.2	35	23	-12	55	85	30
RTE-BBY	10	9:46 AM	9:59 AM	1.4	13	11	-2	46	53	6
BBY-BOS	1	9:59 AM	10:05 AM	1.6	6	8	2	10	8	-2

Table 5.2: Train 2190 Segment Level Performance

The reverse situation seems to occur in the PVD-RTE section, where the trains traveled 30 mph faster on average than scheduled, thus being able to traverse the PVD-RTE segment 12 minutes faster than scheduled, and recover from the delays accumulated in the NLC-PVD segment. This suggests that the

Acela train timetables are not consistent with actual operations and further that the Acela timetables are padded to provide an opportunity for trains to "catch up".

The next section focuses on Train 2150, which is the first train also traveling in the northbound direction between Washington and Boston.

5.4.2 First Northbound Acela Train From WAS - Train 2150 (5:00AM)

Train 2150 is the first weekday northbound Acela train scheduled to depart WAS at 5:00AM and arrive at BOS at 11:40AM with a 15-minute en-route dwell time at NYP. It is important to note that unlike Train 2190, Train 2150 is actually not the first Amtrak train on the corridor; it is the first Acela train on the corridor but there are four Northeast Regional trains preceding it. Also unlike Train 2190, Train 2150 does not have a scheduled stop at New London, CT (NLC) but travels directly from New Haven, CT (NHV) to Providence, RI (PVD). In FY 2014, Train 2150 had an on-time performance of 63% and average delay of 16 minutes. Figure 5.16 shows the average FY 2014 arrival and departure delays for the train at each successive station on the corridor.

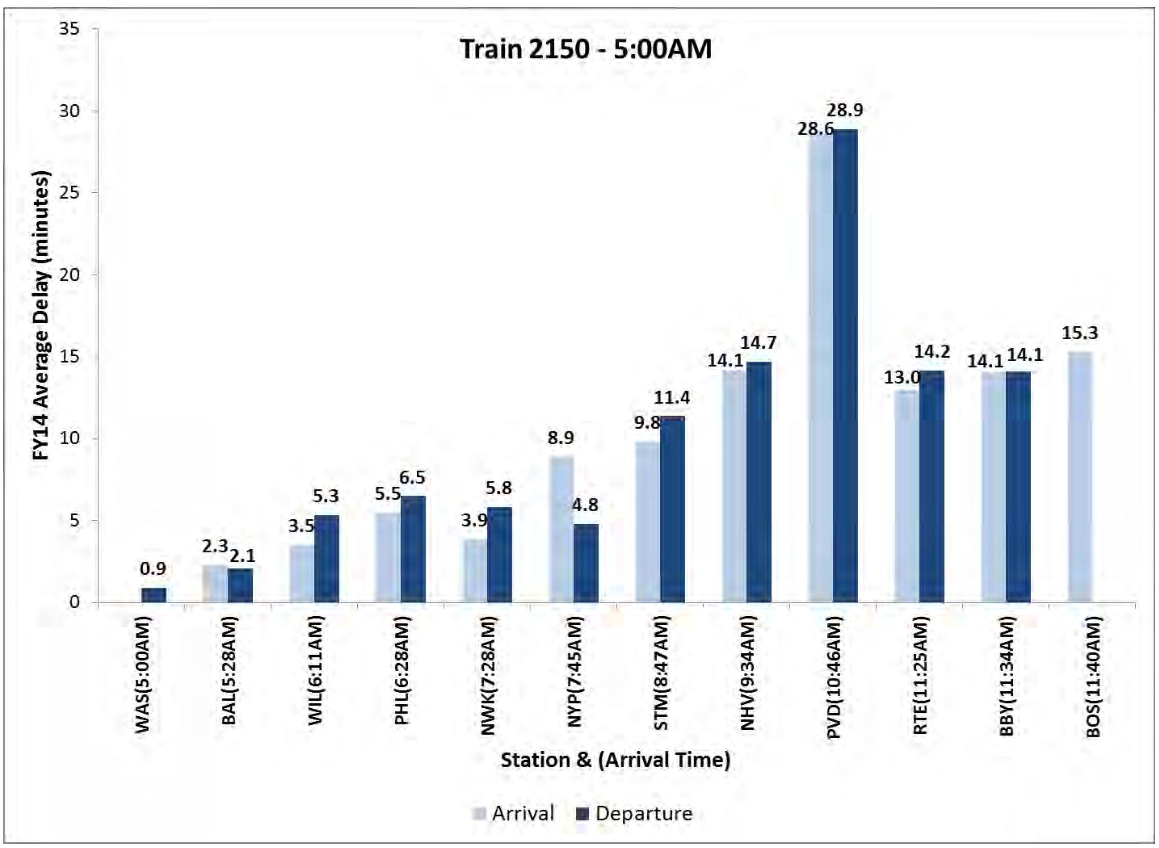


Figure 5.16: FY 2014 Average Station Delays for Train 2150

In FY 2014, the delay on Train 2150 grew incrementally between each segment along the corridor, starting with a one-minute on average late departure from WAS. The train typically accumulated additional delays between arrival at a station and departure from the same station, except at NYP, where the 15-minute scheduled dwell time also appeared to provide some buffer to recover from the upstream delays. The longest segment on Train 2150 is the 113-mile section between NHV and PVD, which also seemed to be the section where the highest amount of delay was accumulated. In FY 2014, Train 2150 arrived and departed PVD on average, 29 minutes later than the scheduled time. However, the delays at the station immediately north of PVD – Route 128, MA (RTE) – were much lower (about 13 minutes), suggesting timetable paddings in the segment to provide an opportunity for trains to "catch up". Altogether, even though the delays were significantly lower south of PVD, some delays were still present and the train typically terminated at BOS with an average delay of 15 minutes.

Train 2150		Scheduled		*Delay (min)	Segment Travel Time (min)			Segment Travel Speed (mph)		
Segment	Distance (miles)	Departure Time	Arrival Time		Scheduled	Avg. Actual	Actual-Scheduled	Scheduled	Avg. Actual	Actual-Scheduled
WAS-BAL	41	5:00 AM	5:28 AM	2.04	28	29	1	88	84	-4
BAL-WIL	69	5:30 AM	6:11 AM	3.05	41	42	1	101	98	-3
WIL-PHL	25	6:11 AM	6:28 AM	5.38	17	17	0	88	88	-1
PHL-NWK	81	6:30 AM	7:28 AM	2.85	58	55	-3	84	88	4
NWK-NYP	10	7:28 AM	7:45 AM	8.49	17	20	3	35	30	-5
NYP-STM	36	8:03 AM	8:47 AM	9.80	44	49	5	49	44	-5
STM-NHV	39	8:47 AM	9:34 AM	14.13	47	50	3	50	47	-3
NHV-PVD	113	9:36 AM	10:46 AM	28.56	70	84	14	97	81	-16
PVD-RTE	32	10:46 AM	11:25 AM	12.00	39	23	-16	49	83	34
RTE-BBY	10	11:25 AM	11:34 AM	13.56	9	9	0	67	67	1
BBY-BOS	1	11:34 AM	11:40 AM	14.43	6	7	1	10	8	-2

Table 5.3: Train 2150 Segment Level Performance

Table 5.3 shows the performance of Train 2150 from the segment level and timetable point of view. In the 113-mile NHV-PVD segment, although the train was scheduled to take 70 minutes, on average, it actually took 84 minutes, 14 minutes slower than scheduled. Correspondingly, the scheduled travel speed in the NHV-PVD segment was 97 mph but on average, the train travelled at 81 mph, 16 mph slower than scheduled. It is interesting to note that the scheduled travel speed in the BAL-WIL and NHV-PVD segments were the fastest, at 101mph and 97 mph, respectively. Although theoretically Acela is able to reach such high speed, in reality it usually did not. Again, the reverse situation seems to occur in the PVD-RTE section with the trains traveling 34 mph faster than scheduled, thus being able to traverse the segment 16 minutes faster than scheduled, and recover from the delays accumulated in the NHV-PVD segment. Once again, this suggests timetable padding and flawed timetables.

The first southbound trains are presented in the next sections.

5.4.3 First Southbound Acela Train From NYP - Train 2103 (6:00AM)

Train 2103 is the first southbound Acela train out of NYP. It is scheduled to depart NYP at 6:00AM and arrive WAS at 8:55AM. Although it is the first Acela train on the corridor, there are three Northeast Regional trains preceding it. In FY 2014, Train 2103 had an on-time performance of 89% and average delay of 5 minutes. There are no scheduled dwell times at en-route stations for Train 2103.

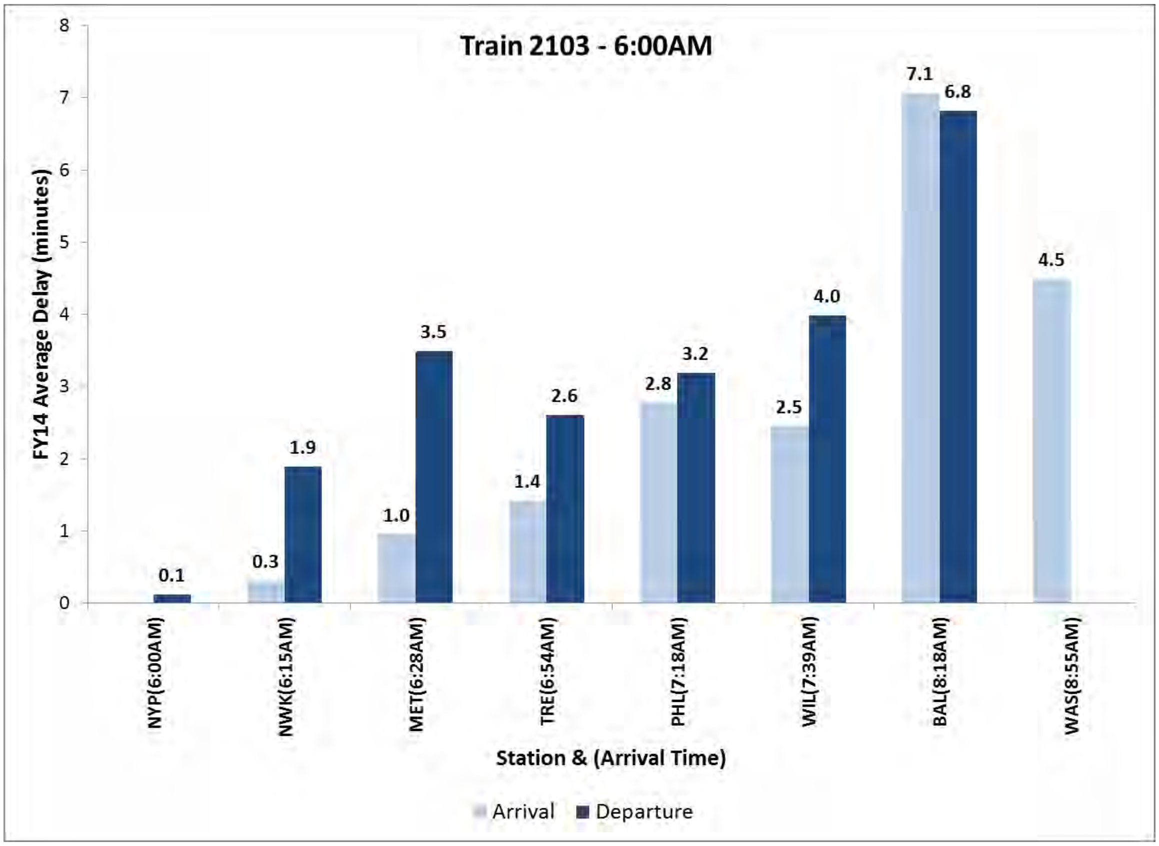


Figure 5.17: FY 2014 Average Station Delays for Train 2103

As shown in Figure 5.17, on average, the train departed NYP relatively on schedule but accumulated 2 minutes of delays as it departed Newark, NJ (NWK), and about 3.5 minutes as it departed Metropark, NJ (MET). Regarding terminal time, the train accumulated about 1.5 minutes of unscheduled time discharging and receiving passengers at NWK, and about 2.5 minutes at MET. The delays on the train typically continued to add up downstream, with the segment between Wilmington, DE Station (WIL) and Baltimore Penn Station (BAL) consistently experiencing the highest amount of delays (about 7 minutes).

Train 2103	Scheduled			*Delay (min)	Segment Travel Time (min)			Segment Travel Speed (mph)		
Segment	Distance (miles)	Departure Time	Arrival Time		Scheduled	Avg. Actual	Actual-Scheduled	Scheduled	Avg. Actual	Actual-Scheduled
NYP-NWK	10	6:00 AM	6:15 AM	0.32	15	15	0	40	39	-1
NWK-MET	14	6:15 AM	6:28 AM	0.96	13	12	-1	65	70	5
MET-TRE	34	6:28 AM	6:54 AM	1.43	26	24	-2	78	85	7
TRE-PHL	33	6:54 AM	7:18 AM	2.78	24	24	0	83	82	-1
PHL-WIL	25	7:20 AM	7:39 AM	2.46	19	18	-1	79	82	3
WIL-BAL	69	7:39 AM	8:18 AM	7.07	39	42	3	106	98	-8
BAL-WAS	41	8:20 AM	8:55 AM	4.49	35	33	-2	70	75	5

Table 5.4: Train 2103 Segment Level Performance

Table 5.4 shows the performance of Train 2103 from the segment level and timetable point of view. In the 69-mile WIL-BAL segment, although the train was scheduled to take 39 minutes, it actually took 42 minutes on average, 3 minutes slower than scheduled. Correspondingly, the actual travel speed in the segment was 8 mph slower than scheduled. The scheduled speed in the segment was 106 mph but on average Train 2103 traveled at a speed of 98 mph in the WIL-BAL segment.

Table 5.4 also shows that on average, Train 2013 traveled in the NWK-MET segment about 5 mph (1 minute) faster than scheduled, 7mph (2 minutes) faster in the MET-TRE segment, and 5mph (2 minutes) fasters in the BAL-WAS segment.

5.4.4 First <u>Southbound</u> Train From BOS – Train 2151 (5:05AM)

Train 2151 is the first southbound Acela train scheduled to depart from BOS at 5:05AM and arrive at WAS at 11:53AM with a 15 minute en-route dwell time at NYP. Unlike Train 2103, Train 2151 does not have a scheduled stop at Metropark (MET) but travels directly from Newark, NJ(NWK) to Philadelphia, PA (PHL). Additionally, it is the first Acela train on the corridor, and there are no other Amtrak trains preceding it. In FY 2014, Train 2151 had an on-time performance of 72% and average delay of 11 minutes. Figure 5.18 shows the average FY 2014 arrival and departure delays for the train at each successive station on the corridor.

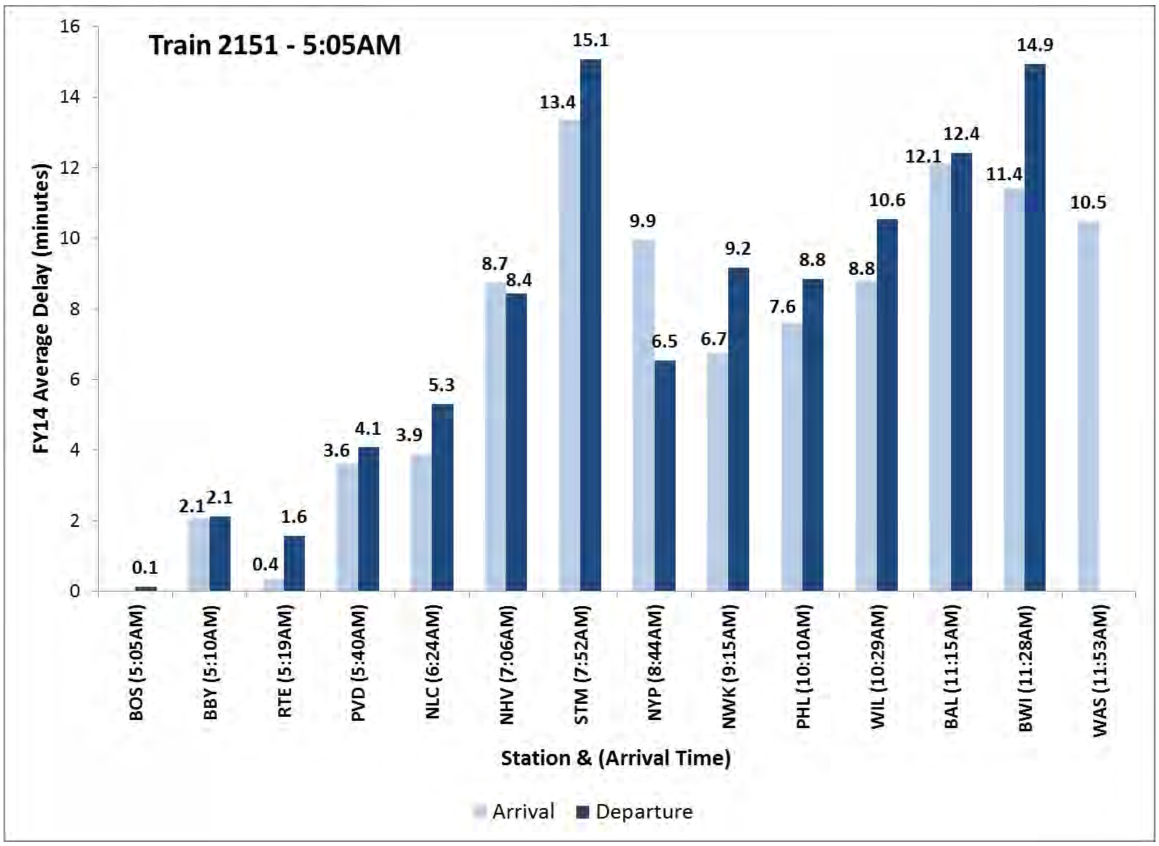


Figure 5.18: FY 2014 Average Station Delays for Train 2151

On average, in FY 2014, Train 2151 departed the originating station (BOS) relatively on time. Although scheduled to arrive at Boston Back Bay Station (BBY), 1 mile away at 5:10AM, Train 2151 didn't usually arrive until around 5:12AM. The train performance improved at RTE but after RTE, the delays on the train typically continued to grow, peaking in the New Haven - Stamford, CT segment (NHV-STM), where it usually arrived 13 minutes later and departed 15 minutes later than scheduled. Train 2151 often recovered from some of the built-up delays at NYP, likely an advantage of the scheduled 15-minute train dwell time. The delays typically continued to grow again downstream, with the train terminating at WAS on average, 11 minutes later than scheduled.

Table 5.5 shows the performance of Train 2151 from the segment level and timetable point of view. The train typically traveled about 3 minutes slower than scheduled in the NLC-NHV segment, and about 5 minutes slower than scheduled in the NHV-STM segment. The segments highlighted in green show segments where the train travelled faster than scheduled, which happened in the BBY-RTE segment, the STM-NYP segment, and the BWI-WAS segment.

Train 2151	Scheduled			*Delay (min)	Segment Travel Time (min)			Segment Travel Speed (mph)		
Segment	Distance (miles)	Departure Time	Arrival Time		Scheduled	Avg. Actual	Actual-Scheduled	Scheduled	Avg. Actual	Actual-Scheduled
BOS-BBY	1	5:05	5:10	2	5	7	2	12	9	-3
BBY-RTE	10	5:10	5:19	0	9	7	-2	67	83	16
RTE-PVD	32	5:19	5:40	4	21	23	2	91	83	-8
PVD-NLC	62	5:40	6:24	4	44	44	0	85	85	0
NLC-NHV	51	6:24	7:04	9	40	43	3	77	70	-6
NHV-STM	39	7:06	7:52	13	46	51	5	51	46	-5
STM-NYP	36	7:52	8:44	10	52	47	-5	42	46	5
NYP-NWK	10	9:00	9:15	7	15	15	0	40	39	-1
NWK-PHL	81	9:15	10:08	8	53	51	-2	92	94	3
PHL-WIL	25	10:10	10:29	9	19	19	0	79	79	0
WIL-BAL	69	10:29	11:13	12	44	46	2	94	91	-3
BAL-BWI	11	11:15	11:28	11	13	12	-1	51	55	4
BWI-WAS	30	11:28	11:53	10	25	21	-4	72	88	16

Table 5.5: Train 2151 Segment Level Performance

To clarify that the station level performances observed so far are not unique to FY 2014, the next section shows the station level performance for Train 2151 in FY 2012.

5.4.5 First Southbound Acela Train From BOS - Train 2151 (FY 2012)

Fiscal Year 2012 was one of the best years in terms of service performance between FY 2005 and FY 2014. The Amtrak OTP for Train 2151 in FY 12 was 92% and the average delay per train was 4 minutes (compared to OTP of 72% and average delay of 11 minutes in FY 2014).

Figure 5.19 shows the FY 2012 station level average delay per train for Train 2151. Although the average delays at each station were lower in FY 2012 compared to FY 2014, the average delay trend looks very comparable. Similar to FY 2014, the delays on the train arriving at STM appeared to be the largest. The average delay peaked around STM, recovered in NYP but grew incrementally at each successive station on the corridor, and arrived WAS with 4 minutes of delay on average. The similarities between the FY 12 and FY 14 average delay for Train 2151 indicate that the delay trend on the first train is more of a regular occurrence than a unique or one-time event.

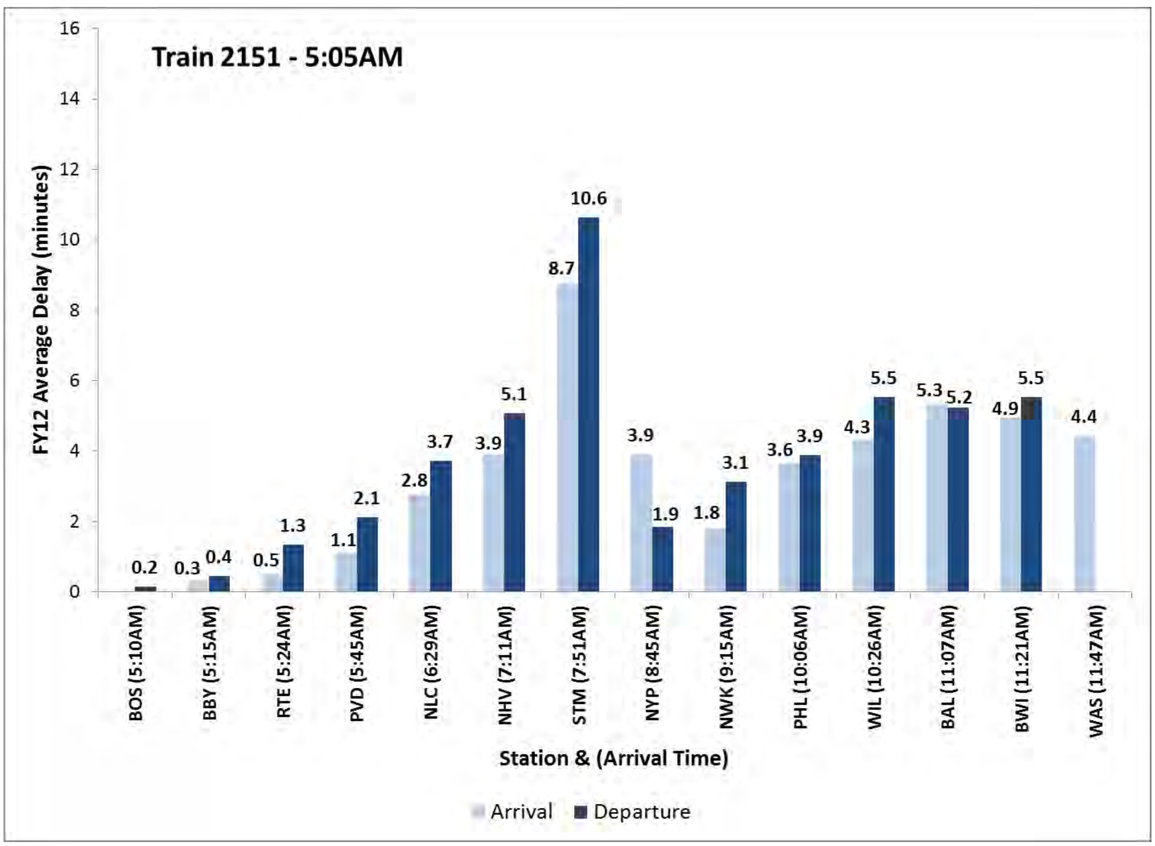


Figure 5.19: FY 2012 Average Station Delays for Train 2151

In summary, upstream delays on the Acela seemed to accumulate along the corridor as a result of unscheduled terminal time at many stations. In addition, flawed segment train speeds and travel times led to additional en-route delays on some segments, and catch-up on other segments.

5.5 Train Interference Analysis

Theoretically, delays on one Acela train could be large enough to spill over to the next scheduled train. This phenomenon is typically referred to as train interference, where one train interferes with smooth operations and performance of another train. This section investigates interference within the Amtrak system but does not include the effect of commuter or freight train interference.

Figure 5.20 shows the delay per mile for successive northbound Amtrak trains (including Acela and Regional). Assuming Amtrak trains interfered with one another, one would see an increasing trend in the delays on consecutive trains; however the delays between consecutive trains in Figure 5.20 appear to be random. This manifestation is not surprising because Amtrak usually spaces trains in intervals greater than 30 minutes, and Acela trains typically experience delays less than 20 minutes on an average day. Again, from the travel time distribution, we estimated that about 92% of Acela trains arrive within 30 minutes of the scheduled arrival time. This suggests that on an average day, delays do not propagate from one train to the next; however on a bad day, they likely do. Also, due to limited data on locomotives, the cascading effects of delay on round trips were not studied.

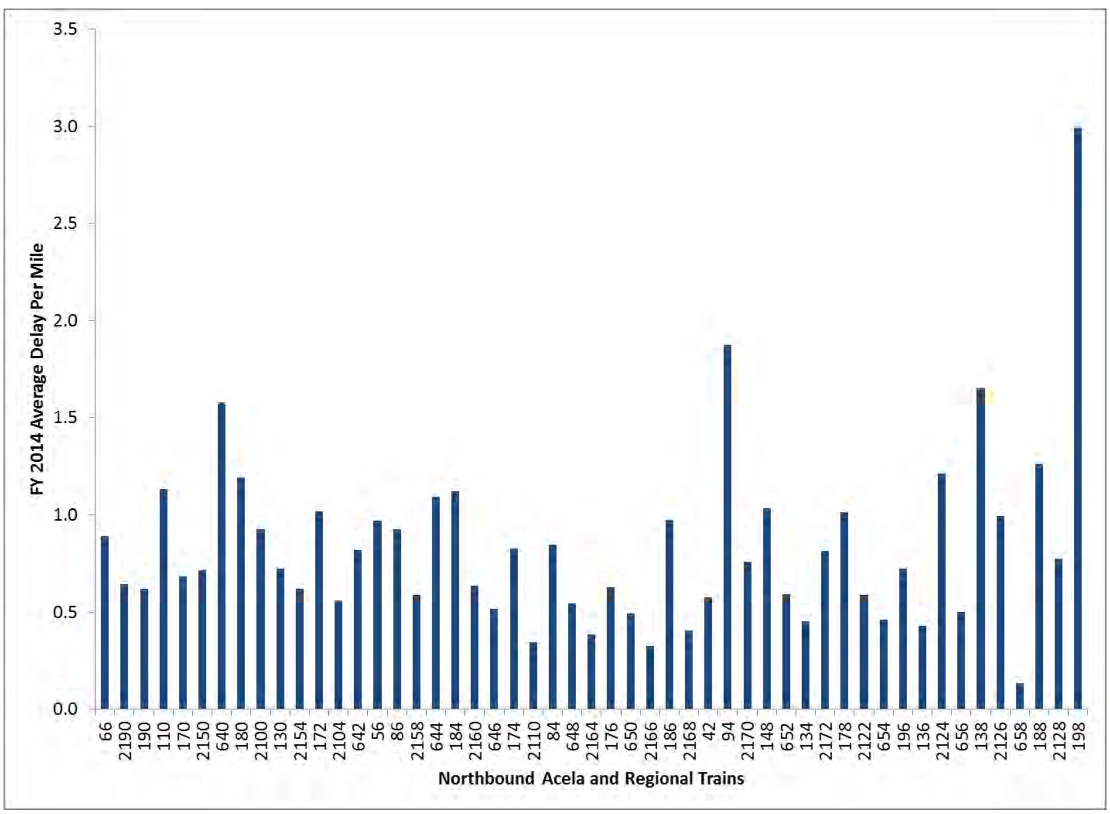


Figure 5.20: FY 2014 Average Delay Per Mile

5.6 Supply Fluctuations Due to Weather, Accidents and Other Incidents

In the Northeast Corridor, the performance of Acela service is occasionally affected by unanticipated changes in travel conditions such as bad/extreme weather, crashes and incidents, equipment failure or other unexpected infrastructure malfunctions. This section investigates supply fluctuations due to accidents and incidents (e.g. signal failures, weather related, track work, etc.), and includes their effects on travel delays and train cancellations.

Under PRIIA Section 207, Amtrak was required to report the total delay minutes apportioned into Amtrak-responsible, Host-responsible (Metro-North Railroad-responsible), and Third Party-responsible category (see Section 3.3.2 for additional details of PRIIA requirements regarding delays). However the author of this thesis did not have access to this data. Consequently, the inferences made in this section are based on observing the total daily delay minutes on the Acela services. The total daily delay minutes were calculated by aggregating end-point delays on all the Amtrak train that were operated each day. Following that, the author went through a manual process of searching through Amtrak's Twitter[vi] and Breaking News[vii] feeds for reports of major accidents and incidents on days with significant delays.

<u>FY 2012 Daily Delays</u>

Figure 5.21 shows the total daily end-point delays in minutes experienced on Acela trains in FY 2012. There were four days of the year on which Acela experienced severe delays greater than 1,000 minutes. 1,000 minutes was chosen as an arbitrary boundary line separating 'regular' delays from severe delays. The Twitter and Breaking News feeds indicated Severe Storms as the main cause of the 1,939 minutes of delay on 6/29/2012 and 1,510 minutes of delay on 6/30/2012. In addition, the 2,566 minutes of delay on 9/18/2012 were attributed to service disruptions from Hurricane Sandy. The author was not able to identify the cause of the 1,346 minutes of delay on 6/1/2012. Not counting the severe delays (>1,000 minutes), there were 10 other days when total delays exceeded 500 minutes. Overall, the total daily delays on Acela trains in FY 2012 were fairly consistently around 100 minutes. To comprehend this value, divide 100 minutes by 32 trains on a given Monday, for example, which equates to Acela trains being on average 3 minutes late in arrivals at its final station.

5.6.1 Total Minutes of Delay

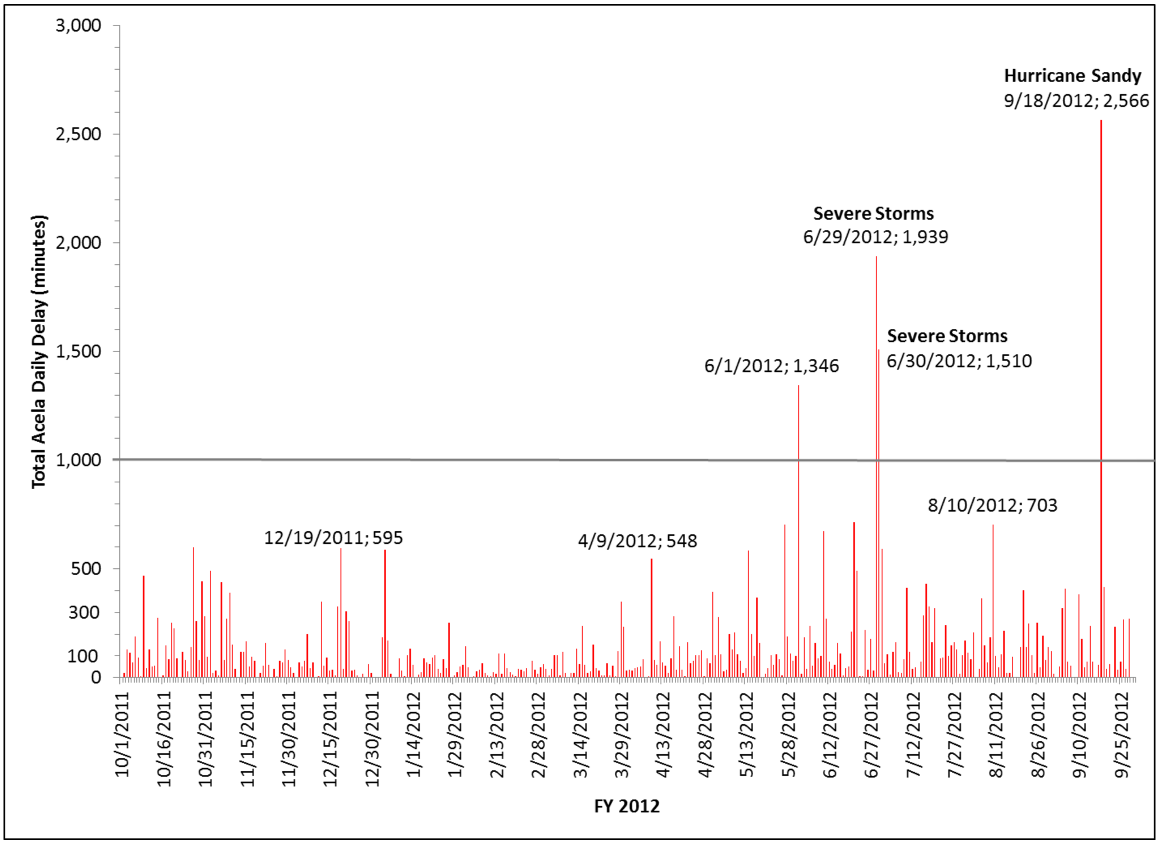


Figure 5.21: FY 2012 Acela Daily Delays

From this proxy analysis, the major delays in FY 2012 appear to be as a result of unanticipated weather disruptions, and accounted for 15% of FY 2012 total Acela delays.

FY 2014 Daily Delays

The same analysis process was undertaken for total daily delays in FY 2014. Figure 5.22 shows the total daily end-point delays in minutes experienced on Acela trains in FY 2014. There were 14 days of the year on which Acela experienced severe delays greater than 1,000 minutes.

The largest delay in FY 2014 was experienced on July 3rd, 2014 attributed to Hurricane Arthur. For example, Train 2124, which was scheduled to depart Washington Union station (WAS) at 6:00pm and arrive in New York Penn (NYP) at 8:45pm, did not pull into the terminal in New York until 2:47am with 150 passengers on board. Due to the severe weather condition, the train experienced a 6-hour delay, in addition to the scheduled 2hr 45min travel time. Amtrak's on-time performance metric for Acela trains on that day was 42%. In other words, 19 of the 33 scheduled Acela trains on July 3rd, 2014, arrived at their destination more than 10 minutes later than the scheduled arrival time.

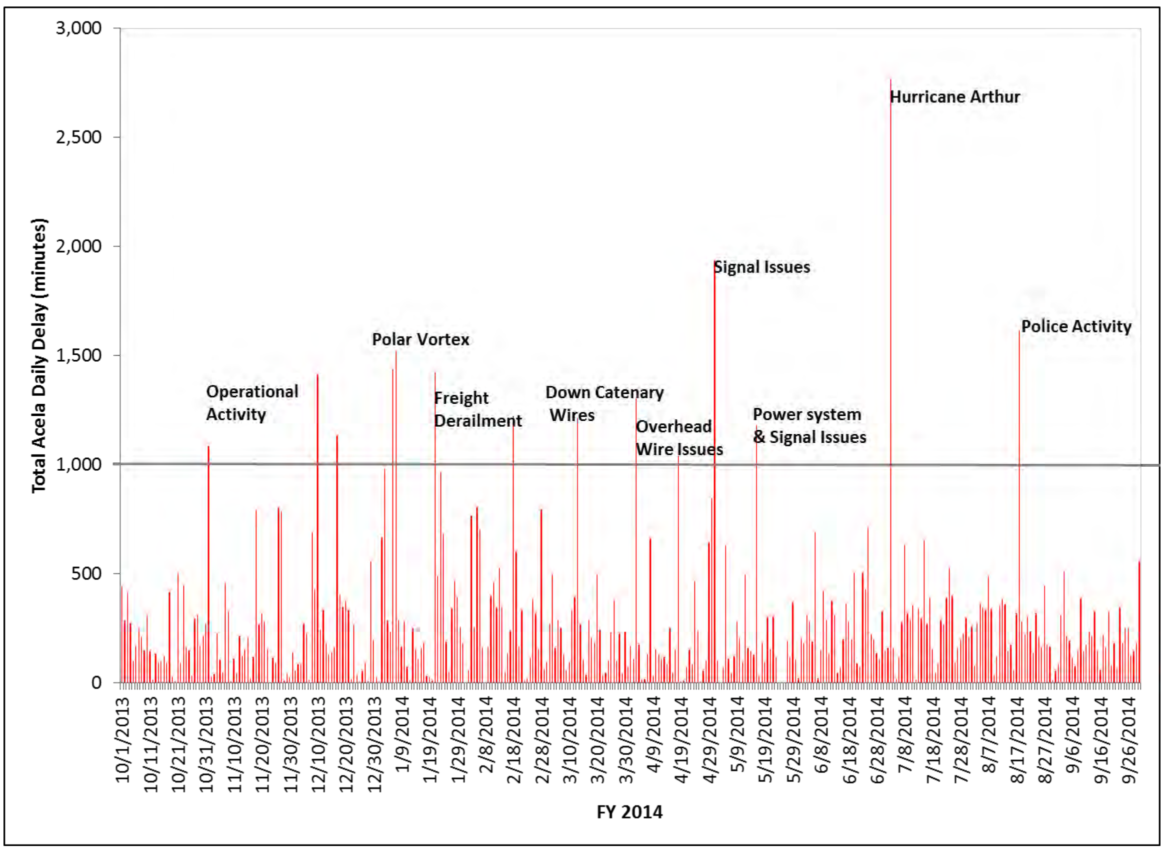


Figure 5.22: FY 2014 Acela Daily Delays

Signal issues on the section of track between Washington, DC and Baltimore, MA were responsible for the second most delayed day on Acela in FY 2014. The severe winter weather aka Polar Vortex in

January 2014 also resulted in higher than average delays, especially on the 5th, 6th and 21st of the month. 7 of the 14 days in FY 2014 with significantly high delays were due to power system, signal issues, overhead wire issues, and operational issues. Compared to FY 2012, which did not experience any days with severe delays due to train or infrastructural malfunctions, this suggests that the track and equipment in the NEC corridor have deteriorated

Table 5.6 shows the 14 days in FY 2014, on which experienced total delay was greater than 1,000 minutes, including the date, day of week, reason for the delay, and the OTP achieved on that day. Although the total delay was abnormally high, the OTP on some days was still high (e.g. 4/18/2014), which suggests that the delay was concentrated at a specific section of the corridor and time of the day, and had minimal impact on most trains scheduled to operate on that day. Days with high delays and high OTP (e.g. 1/7/14) suggest that the delay propagated throughout the corridor and day, and affected majority of the scheduled trains.

Delay Reason	Date	Weekday	Total Delay Minutes	OTP
Hurricane Arthur	7/3/2014	Thu	2,767	42%
Signal issues WAS-BAL	5/1/2014	Thu	1,937	18%
Police activity PHL-TRE	8/18/2014	Mon	1,613	31%
Polar vortex	1/7/2014	Tue	1,523	10%
Polar vortex	1/6/2014	Mon	1,439	68%
Polar vortex	1/21/2014	Tue	1,424	28%
Operational activity	12/10/2013	Tue	1,412	33%
Overhead wire issues WAS-BAL	4/3/2014	Thu	1,306	47%
Freight derailment	2/18/2014	Tue	1,223	15%
Overhead wire Issues WIL-BAL	3/13/2014	Thu	1,214	36%
Power system issues NYP-STM; Police activity north of WAS	5/16/2014	Fri	1,182	35%
Disabled Amtrak train	12/17/2013	Tue	1,134	36%
Unknown	11/1/2013	Fri	1,086	27%
Overhead wire issues WAS-WIL	4/18/2014	Fri	1,040	88%

Table 5.6: Acela Daily Total Delay Greater Than 1,000 minutes

The severe delays were responsible for 24% of total FY 2014 Acela delays. Of the 24%, half were attributed to weather-related, police activity, or unknown sources while the other half were associated with equipment and infrastructural issues.

5.6.2 Cancelled Trains

It is important to investigate supply side fluctuations that result in train cancellations firstly because they pose huge difficulties to travelers, and secondly because they are currently not accounted for in the existing delay and OTP metrics.

Although Acela service performance was best in FY 2012 (of performance between FY 2005 and FY 2014), there were 14 days in the year on which Acela trains were cancelled, and 16 total cancelled trains. Figure 5.23 shows ten days on which one Acela train was cancelled, and two days with two Acela trains cancellations. Altogether, less than 1% (0.2%) of total FY 2012 scheduled trains was cancelled.

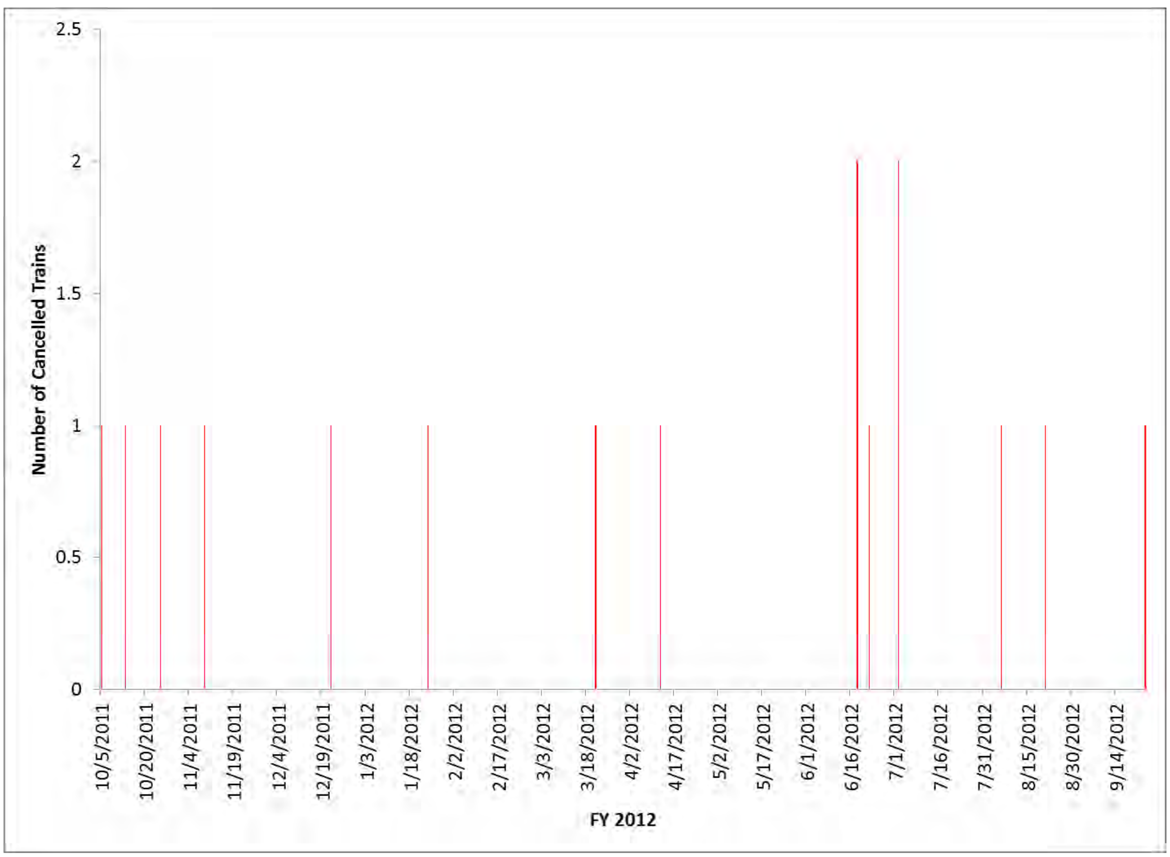


Figure 5.23: FY 2012 Number of Cancelled Acela Trains

In FY 2014, Figure 5.24 shows that there 28 days in the year on which Acela trains were cancelled, and a total of 187 cancelled trains. The data labels in Figure 5.24 are formatted as: date; # of cancelled trains (e.g. 1/22/2014; 23). The major train cancellations occurred in January and February. Altogether, 2% of total scheduled Acela Train in FY 2014 was cancelled.

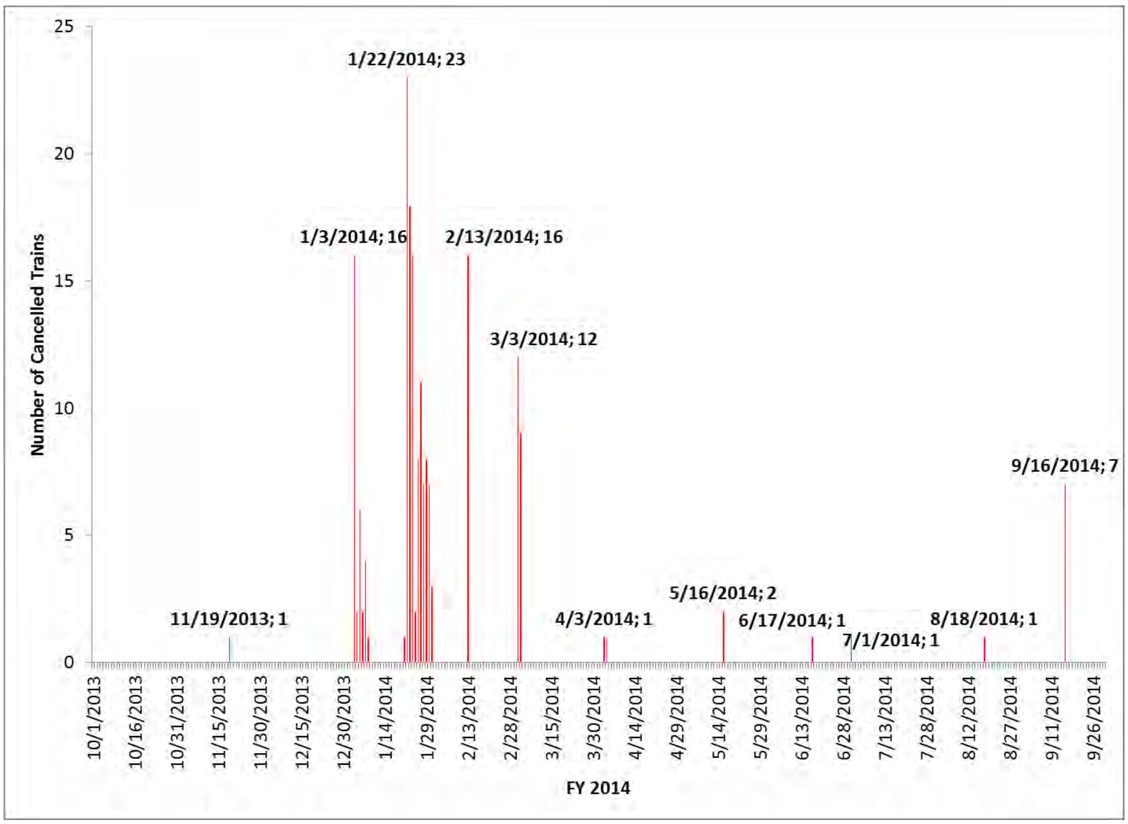


Figure 5.24: FY 2014 Number of Cancelled Acela Trains

Table 5.7 shows a list of the 28 days along with data on the number of scheduled trains, the number of cancelled trains, the total delay minutes on operated Acela trains, the Amtrak OTP, and the cause of the train delay/cancellations (from the authors manual process explained in Section 5.5.1). 84% of FY 2014 Acela train cancellations were attributed to weather-related issues, 7% to down catenary/overhead wires issues, 7% to Police Activity, and the last 2% to disabled Amtrak trains or unknown sources.

The days with high cancellations are also days with severe weather or other third-party-related issues. Furthermore, excluding the days with severe delays from external factors such as weather-related, infrastructure or police activity, Amtrak rarely cancels trains.

Date	Day of week	# Trains scheduled	# Trains Cancelled	Total Delay minutes	Amtrak OTP	Delay/Cancellation Reason
11/19/2013	Tue	33	1	269	81%	Disabled Amtrak train
1/3/2014	Fri	33	16	982	29%	Weather-Related
1/4/2014	Sat	9	2	286	43%	Weather-Related
1/5/2014	Sun	19	6	234	54%	Weather-Related
1/6/2014	Mon	33	2	1,439	68%	Weather-Related
1/7/2014	Tue	33	4	1,523	10%	Down Cantenary Wires
1/8/2014	Wed	33	1	290	59%	Down Cantenary Wires
1/21/2014	Tue	33	1	1,424	28%	Weather-Related
1/22/2014	Wed	33	23	492	0%	Weather-Related
1/23/2014	Thu	33	18	964	27%	Weather-Related
1/24/2014	Fri	33	16	682	24%	Weather-Related
1/25/2014	Sat	9	2	189	14%	Weather-Related
1/26/2014	Sun	19	8	51	91%	Weather-Related
1/27/2014	Mon	33	11	343	50%	Weather-Related
1/28/2014	Tue	33	7	470	54%	Weather-Related
1/29/2014	Wed	33	8	394	56%	Weather-Related
1/30/2014	Thu	33	7	255	81%	Weather-Related
1/31/2014	Fri	33	3	182	77%	Weather-Related
2/13/2014	Thu	33	16	530	24%	Weather-Related
3/3/2014	Mon	33	12	270	52%	Weather-Related
3/4/2014	Tue	33	9	498	46%	Police Activity
4/3/2014	Thu	33	1	1,306	47%	Down Cantenary Wires
4/4/2014	Fri	33	1	175	81%	Down Cantenary Wires
5/16/2014	Fri	33	2	1,182	35%	Police Activity
6/17/2014	Tue	33	1	366	69%	Unknown
7/1/2014	Tue	33	1	148	88%	Police Activity
8/18/2014	Mon	33	1	1,613	31%	Police Activity
9/16/2014	Tue	33	7	60	96%	Down Cantenary Wires

Table 5.7: FY 2014 Acela Train Cancellation Summary

In summary, although the train cancellations on the Acela service were not disturbingly high on an average day, cancellations cause a substantial amount of inconvenience and disutility to travelers and thus must be accounted for in service performance metrics. The need for a more precise cumulative service performance index is discussed in Section 8.4.

5.7 Capacity Analysis

This section analyzes the capacity on Acela trains between FY 2005 and FY 2014, which has the potential to influence variations both in demand and supply that impact the assessment of service performance.

The goal of the capacity analysis for this thesis is to provide a caveat to the analysis on the impact of Amtrak service performance on demand. Demand for Acela service is currently alleged to be above the capacity of Acela trains, which might dampen the actual effect of demand responses to service performance. For example, if a passenger on a train experienced major delays during an Acela trip and therefore decided to use a different mode of travel on their next trip, the demand data might not show the effect of the lost passenger because a traveler who was initially not able to purchase a ticket due to the capacity constraints, might replace the lost passenger. The section begins by showing the daily ridership profile and average weekday ridership on Acela trains between FY 2005 and FY 2014. Then it delves into the actual segment level volumes on each train for some chosen sample dates.

Amtrak ridership in the Northeast Corridor has continued to grow strongly, despite the effect of the economic recession that lessened ridership in FY 2009. Total Acela ridership grew from 2.3 million in FY 2005 to a record high of 3.5 million riders in FY 2014. The annual growth in passenger demand was reflected in daily ridership as well. The daily ridership volumes were used to pick the dates for which the Acela capacity analysis was done.

Figure 5.25 shows the number of Acela riders each day between FY 2005 and FY 2014. The top legend in the figure shows the number of days on which ridership exceeded the 14,000, 15,000 and 16,000 mark, respectively. Figure 5.25 also shows the day and month (format: day/month) each year with the highest ridership. In FY 2005, there was only one day where ridership exceeded 14,000 (and 15,000) on March 31, 2005 (3/31). However, by FY 2008, there were 13 days where ridership exceeded 14,000, 7 days where ridership exceeded 15,000 and 4 days with Acela ridership over 16,000. In FY 2008, April 30th (4/30) had the most riders (20,800). The daily ridership was significantly impacted in FY 2009 by the recession; in FY 2009, there were only 5 days with ridership above 14,000, 4 days with ridership above 15,000 and 1 day with ridership above 16,000, and May 29th 2009 (5/29) had the most riders (17,000). By FY 2014, there were 35 days on which the number of daily riders on Acela exceeded 14,000, compared to 14 times in FY 2013. In FY 2013, the day with highest demand was February 28 while in FY 2014, October 31ST and February 28th experienced the highest daily demand.

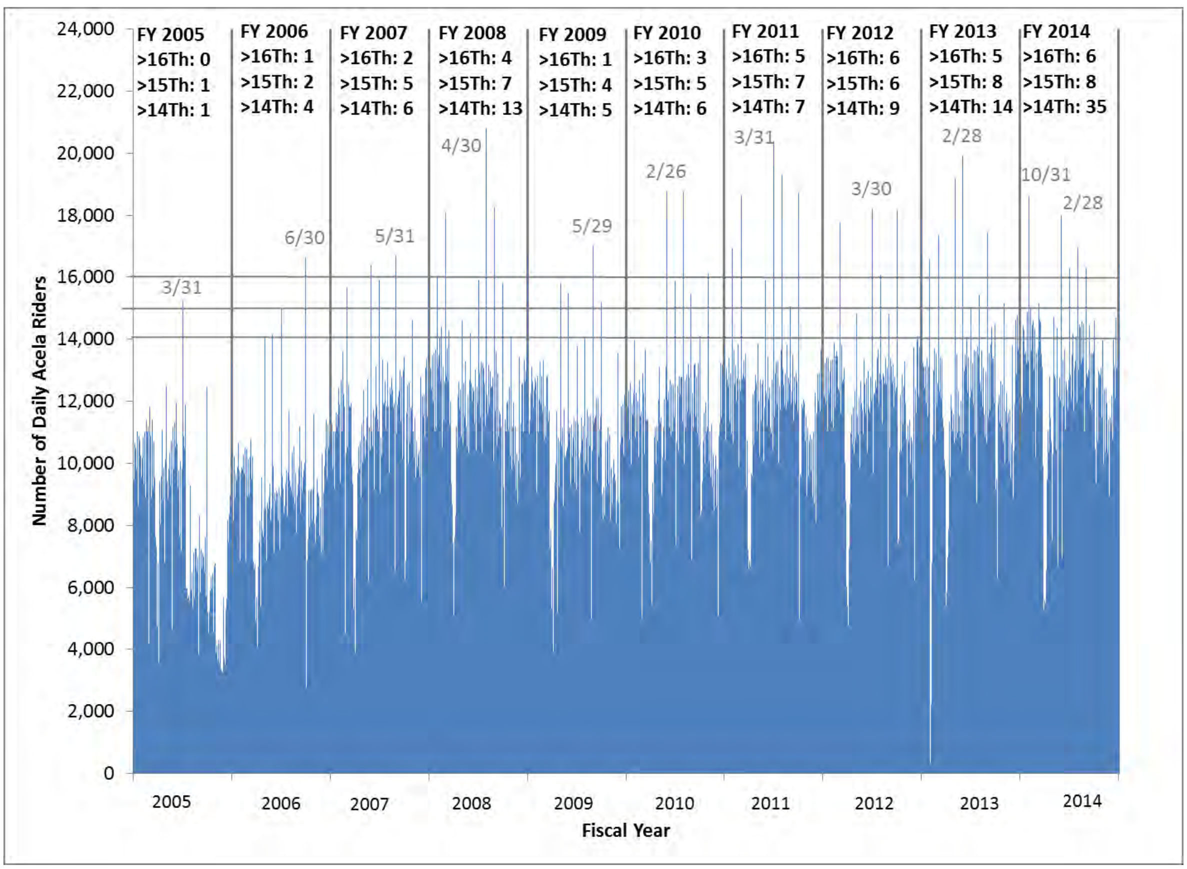


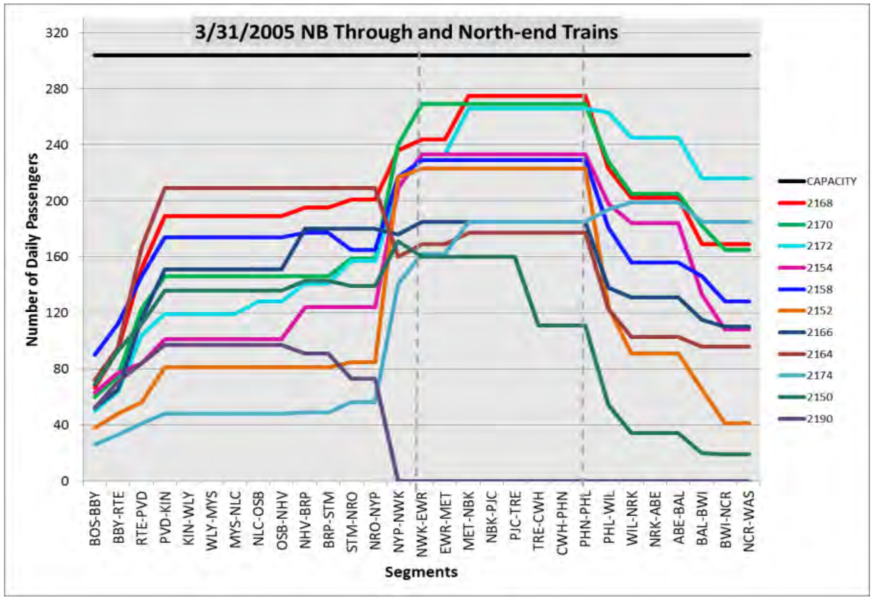
Figure 5.25: FY 2005 - FY 2014 Daily Acela Riders

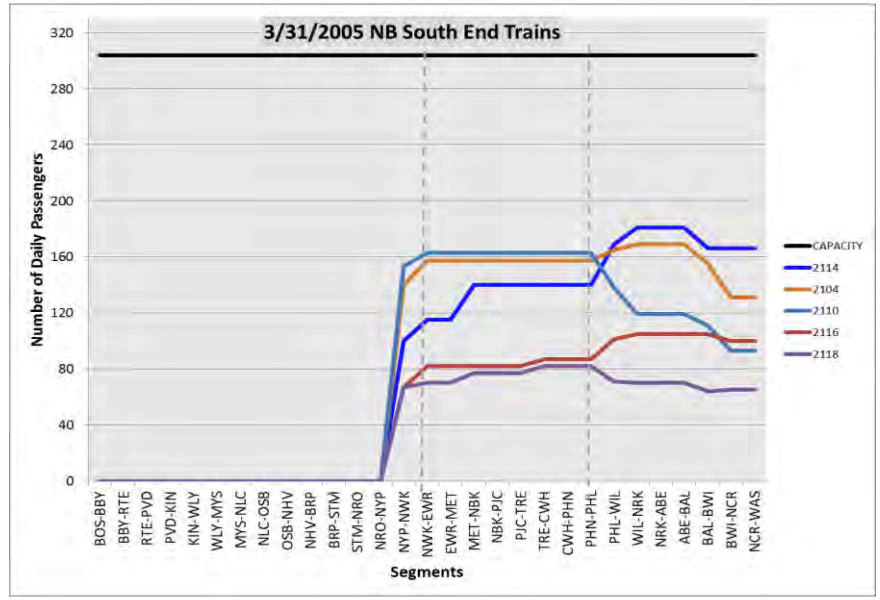
5.7.1 FY 2005

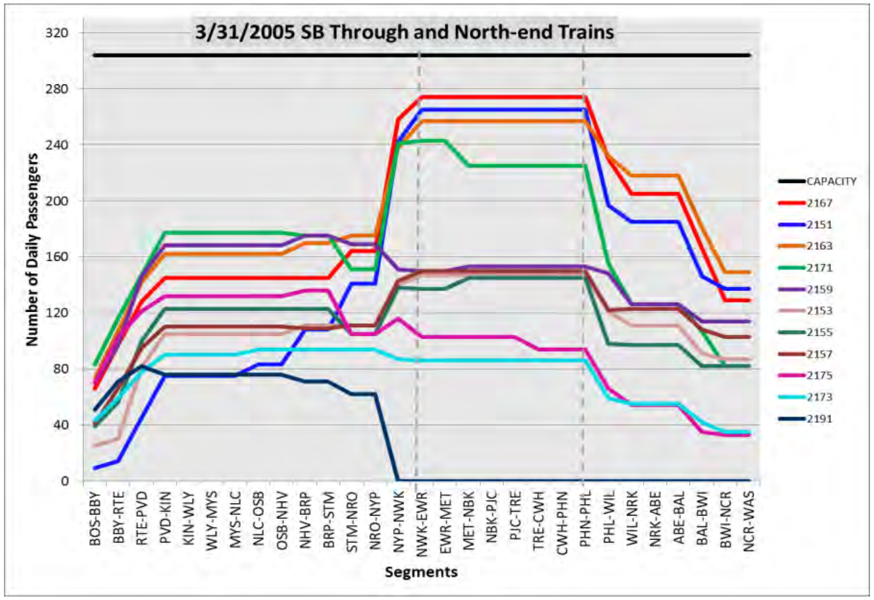
Since the growth in passenger demand between FY 2005 and FY 2014 has not been matched with a proportionate increase in infrastructure, some Acela trains have approached and other might have reached capacity in some segments of the Northeast Corridor. In FY 2005, the day with the highest ridership was March 31st (3/31) with a demand of 15,000 riders. The four figures in Figure 5.26 show segment level demand for each Acela train that was operated on March 31st 2005. The top plots show segment volumes on the northbound (NB) trains while the bottom plots show the volume on the southbound (SB) trains. The capacity of all Acela trains is 304, which corresponds to a load factor of 100%, that is, a segment level volume of 230 corresponds with a 75% load factor.

In the top left corner, the NB through trains (WAS-BOS) show peak loading in the segments between Newark, NK (NWK) and North Philadelphia, PA (PHN). Train 2168 (in red), which departed from WAS at 2:00PM had a load factor of 90% as it crossed the PHN - NWK section (between 3:30PM and 4:30PM). Train 2168, 2170 and 2172 are all PM peak trains, departing from WAS at 2:00PM, 3:00PM and 4:00PM, respectively. Train 2154 and 2158 are AM peak trains, departing from WAS at 7:00AM and 9:00AM, respectively. Train 2152 is also an AM peak train that departs from WAS at 6:00AM.

Similarly, in the bottom left figure, the through southbound trains show peak loading in the same section between NWK and PHN. Train 2167 (in red) departed from Boston (BOS) at 1:20PM, and had a peak loading of 90% as it crossed the section between NWK and PHN in the PM peak (between 5:00PM and 6:00PM). Train 2151 (in blue) which was the AM peak train departing from BOS at 5:00AM had the second highest segment load as it crossed the NWK - PHN section (between 9:00AM and 10:00AM). Train 2167 and 2171 are PM peak trains departing BOS at 1:00PM and 3:10PM, while Train 2151, 2163 are AM peak trains departing BOS at 5:00AM and 11:00AM, respectively.


3/31/2005 NB Through and North-end Trains
Number of Daily Passengers
Segments
CAPACITY
2168
2170
2172
2154
2158
2152
2166
2164
2174
2150
2190
BOS-BBY
BBY-RTE
RTE-PVD
PVD-KIN
KIN-WLY
WLY-MYS
MYS-NLC
NLC-OSB
OSB-NHV
NHV-BRP
BRP-STM
STM-NRO
NRO-NYP
NYP-NWK
NWK-EWR
EWR-MET
MET-NBK
NBK-PJC
PJC-TRE
TRE-CWH
CWH-PHN
PHN-PHL
PHL-WIL
WIL-NRK
NRK-ABE
ABE-BAL
BAL-BWI
BWI-NCR
NCR-WAS


3/31/2005 NB South End Trains
Number of Daily Passengers
Segments
CAPACITY
2114
2104
2110
2116
2118



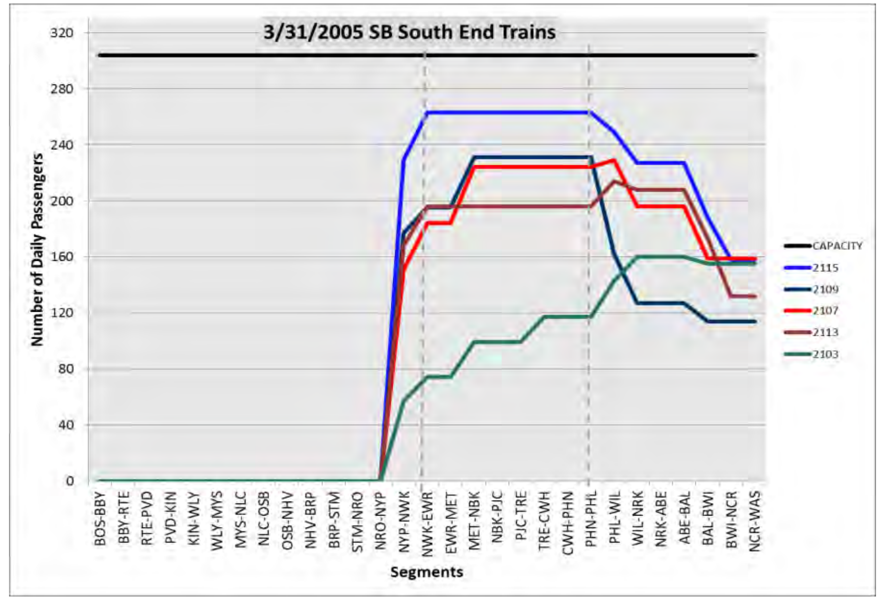


Figure 5.26: FY 2005 Segment Level Train Capacity

5.7.2 FY 2012

Figure 5.27 shows similar figures of segment loading for 3/30/2012, the day with peak ridership volume in FY 2012. Compared to the FY 2005 plots, all the trains show an upward shift suggesting more trains approaching capacity. Furthermore, by FY 2012, the segment loading in the north-end section of the Northeast Corridor between BOS and NYP saw a significant increase in ridership. In FY 2005, the peak volume in the North-end segments were about NB: 210 riders (70% load factor) and SB: 180 riders (60% load factor), however, by FY 2012, the peak volume in the north section of the NEC had increased to 220 riders (85% load factor) in both NB and SB.

In the top left figure, the through NB WAS-BOS trains show peak loading in the segments between Newark, NK (NWK) and North Philadelphia, PA (PHN). Train 2168 (in red), which departed from WAS at 2:00PM had a load factor of 96% as it crossed the PHN - NWK section (between 3:30PM and 4:30PM). Similarly, in the bottom left figure, the through southbound trains show peak loading in the same section between NWK and PHN. Train 2163 (in red) departed from Boston (BOS) at 11:00AM, and had a peak loading of 99% as it crossed the section between NWK and PHN in the PM peak (between 3:00PM and 4:30PM).

In the northbound direction for the through trains, Train 2168, 2170 and 2172 are all PM peak trains, departing from WAS at 2:00PM, 3:00PM and 4:00PM, respectively. Train 2160 and 2164 are late morning trains, departing from WAS at 10:00AM and 12:00PM, respectively, while for NB south-end only trains, Train 2124 and 2126 were scheduled to depart from WAS at 6:00PM and 7:00PM, respectively.

In the southbound direction for the through trains, Train 2163 is an AM peak train departing BOS at 11:00AM, Train 2171 is a PM peak train departing BOS at 3:00PM, and Train 2151 and 2153 are the early morning trains departing from BOS at 5:00AM and 6:00AM, respectively. For the SB south-end only trains, Train 2121 and 2119 were scheduled to depart NYP at 2:00PM and 6:00PM respectively.

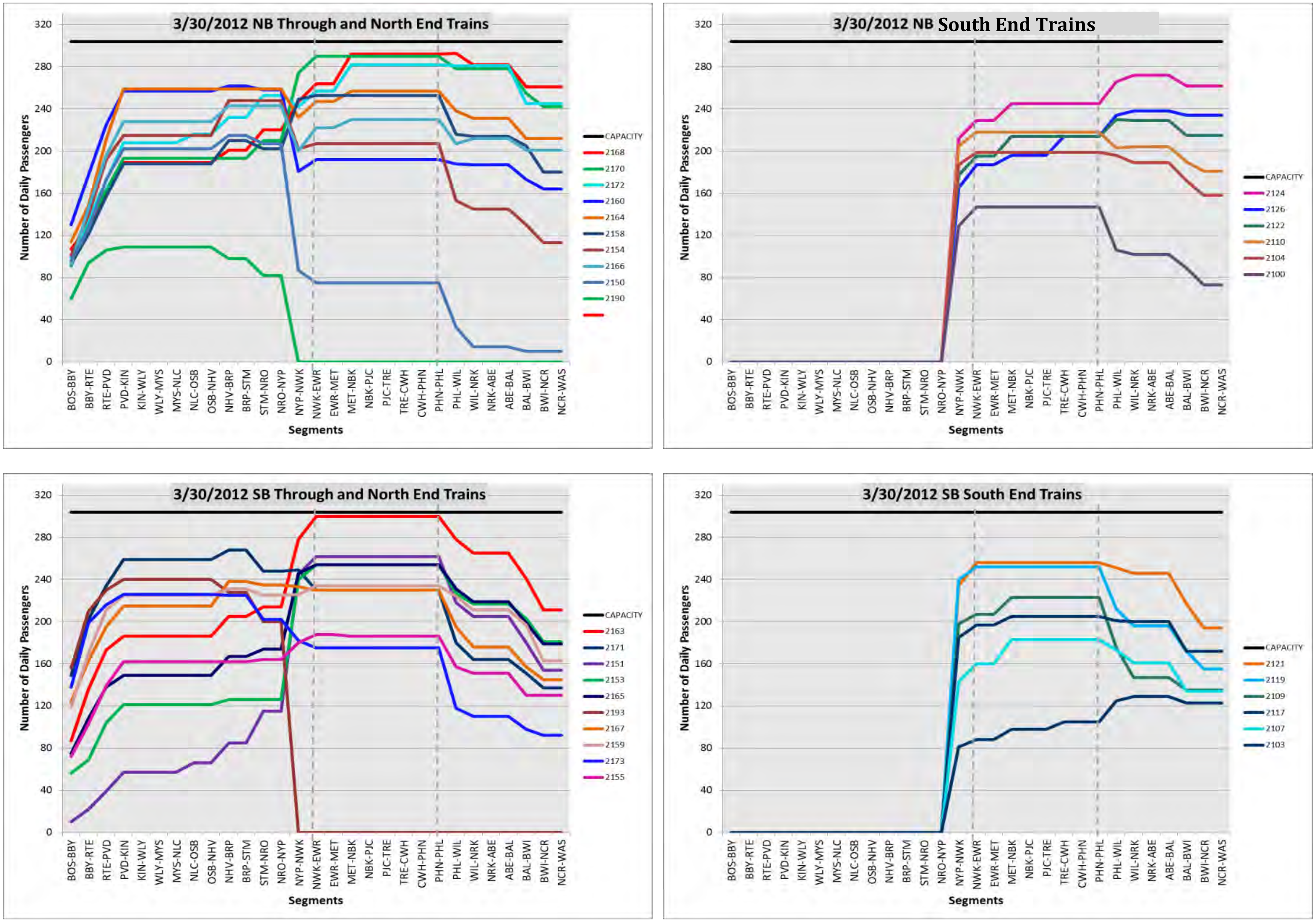


Figure 5.27: FY 2012 Segment Level Train Capacity

5.7.3 FY 2014

Figure 5.28 shows the segment volumes on February 28th 2014, the day with the highest daily ridership in FY 2014. In the top left figure, northbound Train 2168, 2170 and 2172 ran with a peak segment load factor of 100%, 98% and 95%, respectively. Train 2168, 2170 and 2172 are scheduled to depart WAS at 2:00PM, 3:00PM and 4:00PM, respectively. Train 2166, 2164 and 2160 are scheduled to depart from WAS but seem to have a higher loading in the north-end section after departing NYP at 1:00PM, 3:00PM and 4:00PM, respectively.

In the bottom left figure, the southbound through trains have a very interesting characteristic with the trains either loaded in the south-end or north-end sections. For example, Train 2165 scheduled to depart from BOS at 12:10PM was about 65% loaded in the north section; however the loading spiked up to 95% as the train departed from NYP at 4:00PM. Train 2167 (scheduled to depart from BOS at 1:10PM) and Train 2163 (scheduled to depart from BOS at 11:10AM) exhibited a similar pattern. On the other hand, Train 2171 exhibited the opposite pattern with a 94% loading in the north section and a 50% loading in the south section. Train 2171 departed from BOS at 3:10PM and from NYP at 7:00PM. The next train departures from BOS at 4:15PM (Train 2173) also exhibited a similar pattern.

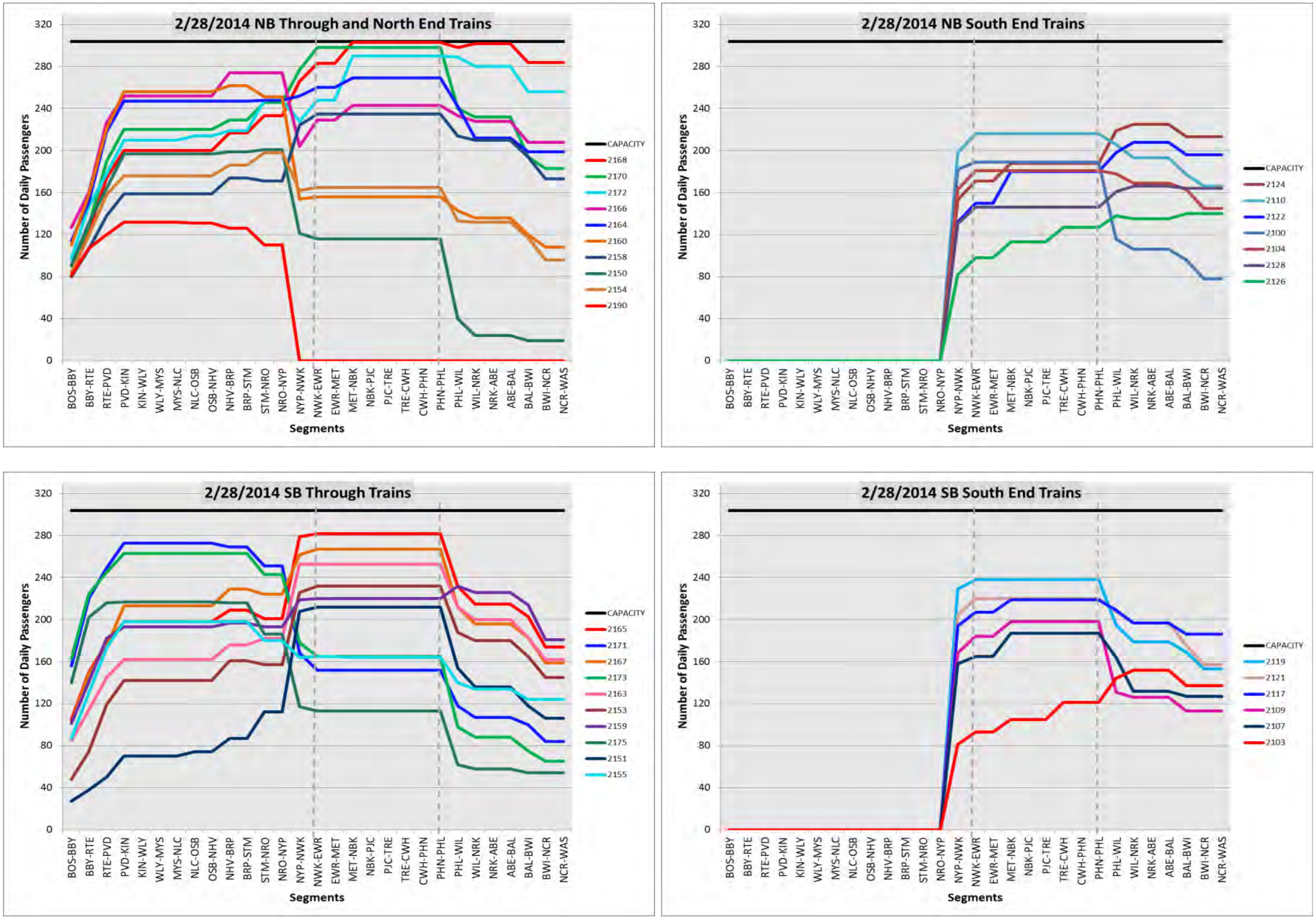


Figure 5.28: FY 2014 Segment Level Train Capacity

In summary, the capacity analysis shows that Acela peak hour trains were indeed at or near capacity, with load factors greater than 94%. This indicates that the full impact of service performance on demand might be dampened when using the historical data due to the capacity constraints. In other word, because demand on Acela trains appears to be more than the train capacity, if hypothetically Acela lost some demand due to poor performance, the full impact might not be evident as the demand backlog might replace the lost demand.

5.8 Demand Response to Acela Service Performance

The main question examined in this section is: Do Amtrak Acela passengers modify their future travel choices in response to past Acela performance that either they experienced or were informed about? As a practical constraint, there needs to be a period of time between when travelers experience or learn about performance information and when they make future travel decisions. This time is usually referred to as a lag period. Following this concept, a simple analysis was conducted to correlate total annual ridership each year with Acela train performance from the prior year, that is, assuming a lag of one year. In other words, the goal of the analysis was to test the assumption that a relationship exists between ridership in a given year and OTP or total delay from the prior year.

The correlation coefficient was used to measure the strength and direction of the relationship between annual ridership and performance. Correlation coefficients close to zero indicate a weak relationship between the ridership and performance from the prior year, while correlations coefficients equal to 1 represent a perfect relationship between both variables. Additionally, positive correlation values denote a direct relationship between the demand and performance, suggesting that years with high demand are associated with high prior year performances, and vice versa.

5.8.1 Annual Ridership to Annual Average OTP

Figure 5.29 shows the total annual Acela ridership (blue) between FY 2005 and FY 2014 and the annual average OTP (red) from the prior year. For example, the first data points show total FY 2006 ridership (2.6 million) and FY 2005 average OTP (71%). The relationship between OTP in two consecutive years is typically associated with a similar relationship between the ridership in the following year. For example, an upward trend in average OTP between FY 2005 and FY 2006 is followed by a similar upward trend between the annual ridership in FY 2006 and FY 2007. This relationship is observed between FY 2005 and FY 2010. However, in FY 2010, even though the performance deteriorated in the prior year, Acela ridership continued to grow. The relationship appeared the continue again between FY 2011 and FY 2013 but broke down again in FY 2014 with ridership increasing despite performance deteriorations in FY 2013. The correlation coefficient between the annual ridership and lagged annual on-time performance between FY 2005 and FY 2014 is 0.75. Although this correlation is not perfect (equal to 1), it is sufficiently high to propose that the ridership in a given year is associated with service performance from the prior year. Furthermore, as expected the correlation between FY 2005 and FY 2010 was much higher and equal to 0.97.

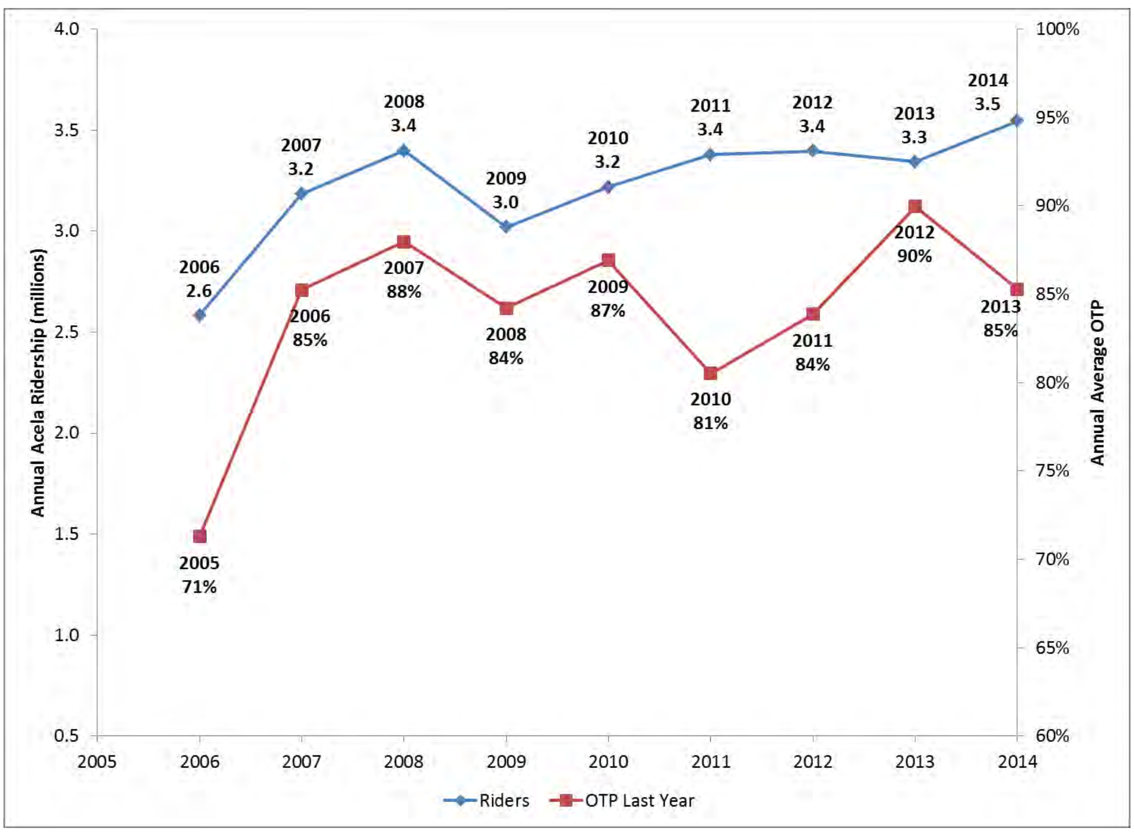


Figure 5.29: Lag Annual Ridership to OTP

5.8.2 Annual Ridership to Annual Delay

Figure 5.30 shows the total annual Acela ridership (blue) and the total annual delay (red) from the prior year. The annual ridership and annual delay exhibited a similar lagged relationship as the annual ridership and annual average OTP. Improvements in performance (that is reduction in total annual delay) appeared to be associated with ridership increase the following year. The correlation between the annual ridership and lagged annual total delay between FY 2005 and FY 2014 was 0.58. Although this correlation is not as high as the correlation coefficient for annual on-time performance, it suggests that Acela ridership in a given year is also associated with delays from the prior year. Furthermore, as expected the correlation between FY 2005 and FY 2010 was much higher and equal to 0.99.

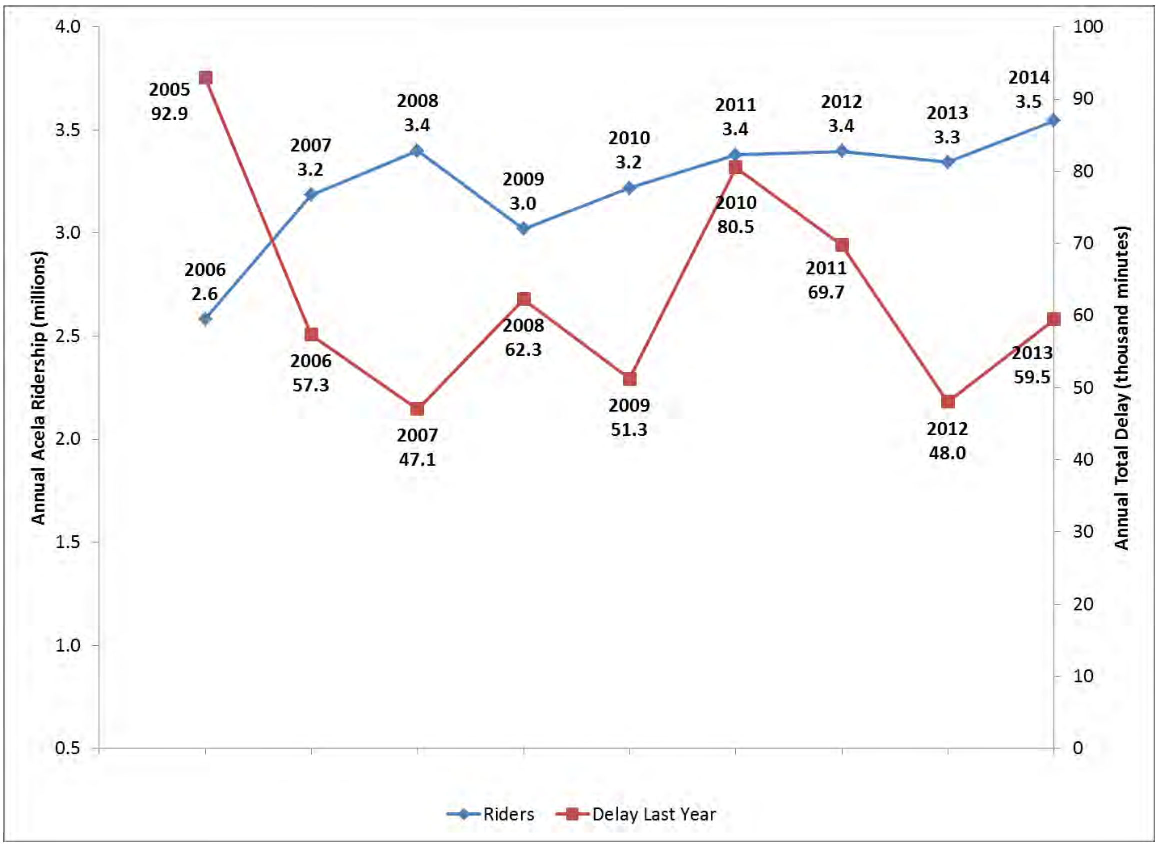


Figure 5.30: Lag Annual Ridership to Delay

It is unclear whether the one-year lagged associations exhibited between OTP or total delay and ridership are cause-effect relationships between service performance and demand. The correlation between Acela ridership and lagged service performance would be investigated further in the time series analysis in Chapter 7.

5.9 Acela Summary

This section summarizes the analysis in Chapter 5.

In FY 2014, 42% of Acela trains arrived on-time, 47% arrived late but within 30 minutes, and the remaining 11% experienced delays greater than 30 minutes. Altogether, 71% of trains arrived within 10 minutes and 89% within 30 minutes.

On-time performance and delay minutes are 70% correlated, and both are useful metrics in quantifying performance. However, neither of them includes the effect of cancelled trains, motivating the need for a cumulative service performance index.

Acela ridership and service performance exhibit seasonal variations. Acela ridership is usually highest in the fall months at the beginning of the fiscal year (October, November), and also during the spring months (April, May, June), and lowest in January and August. Acela performance is usually worst in January (especially during extreme winters) and in the summer (July, August), and best in October, April and May.

Both Acela ridership and service performance vary by day of week. Ridership was typically higher on the weekdays and service performance was typically better during the weekend. However, during the week (Monday – Friday), while Acela ridership exhibited weekday variations, Acela service performance appears to be independent of the day.

In terms of time of day performance variations, AM and PM peak trains were usually more prone to delays than trains during other off-peak times of the day.

The first train analysis revealed that a major portion of Acela delays appeared to be attributable to late departures from originating station, which accumulated and propagated at each consecutive station downstream. In addition, unscheduled dwell time at many stations and poorly estimated segment train speeds and travel times lead to additional en-route delays. On an average day, delays attributed to interference from other Amtrak trains were limited because of the scheduled spacing between trains throughout the day.

The performance of Acela service was occasionally affected by unanticipated changes in travel conditions such as bad/extreme weather, crashes and incidents, equipment failure or other unexpected infrastructure malfunctions. Unanticipated weather-related and third-party events were responsible for about 12% of Amtrak train delays (based on FY 2012 and FY 2014 analysis). Furthermore, in FY 2012, there were no major delays caused by equipment failure, however, in FY 2014, 12% of delays were associated with equipment and infrastructural issues, suggesting rolling-stock and infrastructure deterioration. Lastly, not including days with severe weather issues, Amtrak rarely cancels Acela trains.

The capacity analysis revealed that Amtrak peak hour trains were indeed at or near capacity, with load factors greater than 94% in either the north-end or south-end section of the Northeast Corridor. In addition, a few peak period trains were capacity constrained throughout the entire corridor.

Finally, Acela ridership appeared to exhibit a one-year lagged correlation with service performance. However, it is unclear whether the one-year lags exhibited between OTP or total delay and ridership were due to cause-effect relationships between service performance and demand or just correlations.

6 Northeast Regional Ridership and Service Performance

Figure 5.1 used to describe the variability in travel times on the Acela service is repeated here because of its equal applicability in describing travel time variability on the Northeast Regional service. Figure 6.1 shows different factors that cause fluctuations of Regional demand and service performance, which subsequently lead to variability in travel times on the Regional service. The factors discussed here are those that influence variability in the number of Regional passengers, on the demand side, as well as those that influence variability in the service performance on Regional trains.

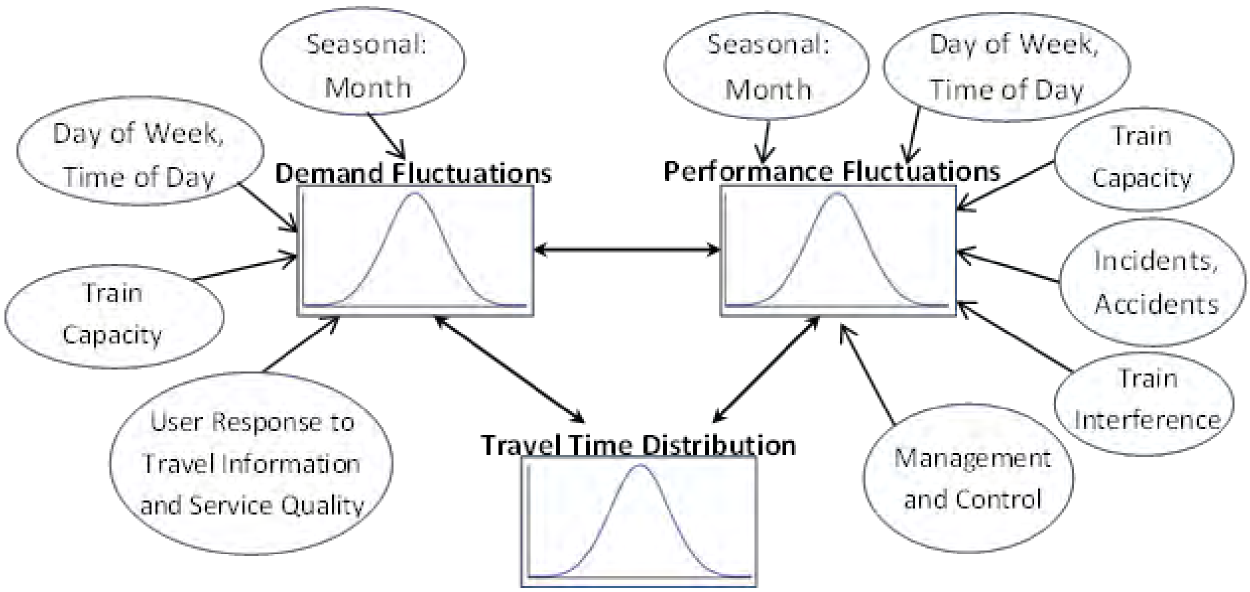


Figure 6.1: Factors affecting Travel Time Distribution on the Regional Service

In this chapter, the factors that cause fluctuations of both Regional ridership and service performance area captured under three main categories:

i) Seasonality and month of year,
ii) Day of week, and time of day, and
iii) Capacity levels on the trains.

Other factors considered in this chapter causing either ridership or performance fluctuations are:

i) Performance disruptions due to accidents and incidents (e.g. signal failures, weather-related, track work, etc.),
ii) Performance interruptions as a result of train interferences,
iii) Performance variations due to administration, management and control elements, and
iv) Demand responses to travel information and service quality.

Each of these factors will be individually investigated in the different sections of this chapter. Firstly, **Section 6.1** shows the distribution of actual delays on Regional trains, and examines the relationship between delays and on-time performance. **Section 6.2** presents the annual, seasonal and month of year fluctuations in the ridership (demand) and performance of Amtrak's Regional service (supply) between FY 2005 to FY 2014. **Section 6.3** drills further down into day-of-week and time-of-day variations in both ridership and service performance, and highlights any relationships between them. This section will directly answer the question about whether poor performance leads to even poorer performance. **Section 6.4** is on the First Train Analysis, which examines the en-route delays on the first trains of the day, comparing the scheduled performance to the average actual performance at each station on Regional trains. **Section 6.5** focuses on service performance disruptions caused by accidents and incidents (e.g. signal failures, weather related, track work, etc.), including their effects on delays and train cancellations. **Section 6.6** analyzes the capacity on Regional trains between FY 2005 and FY 2014, which on the demand side, affects the number of people able to ride Regional trains, and on the supply side, affects the overall performance of operated Regional trains. **Section 6.7** presents a preview of the relationship between the service quality and demand of the Regional service. Finally, **Section 6.8** concludes the chapter with a summary of the prior sections.

6.1 Regional Travel Time Variability

As discussed in the context of Acela trains, travel time variability is a measure of service performance, and is often measured by on-time performance (OTP) or minutes of delay. However both metrics capture different aspects of performance; while the minutes of delay reveals the **magnitude** of delay, the OTP shows the **frequency** of good performance. Both metrics are discussed and compared in this section.

6.1.1 Distribution of Delay

Figure 6.2 shows the distribution of end-to-end delays encountered on all Regional trains operated in FY 2014. The performances on the ~22,000 Regional trains operated in FY 2014 are reflected in the delay distribution. On time arrivals at the terminating station are represented as trains with negative (<0) and zero delays; late trains have positive delay values (>0). The distribution of the delays on Regional trains in FY 2014 is very similar to that of Acela trains. In both, 42% of trains arrived earlier than or at the scheduled arrival time, and 58% of all scheduled trains arrived at the terminating station later than scheduled. In addition, on both the Acela and Regional services, 29% of trains arrived within 10 minutes of the scheduled arrival time. However, compared to Acela, more Regional trains arrived with a delay greater than 30 minutes. Altogether, 71% of Regional trains arrived within 10 minutes of the scheduled arrival time, and 29% arrived more than 10 minutes late.

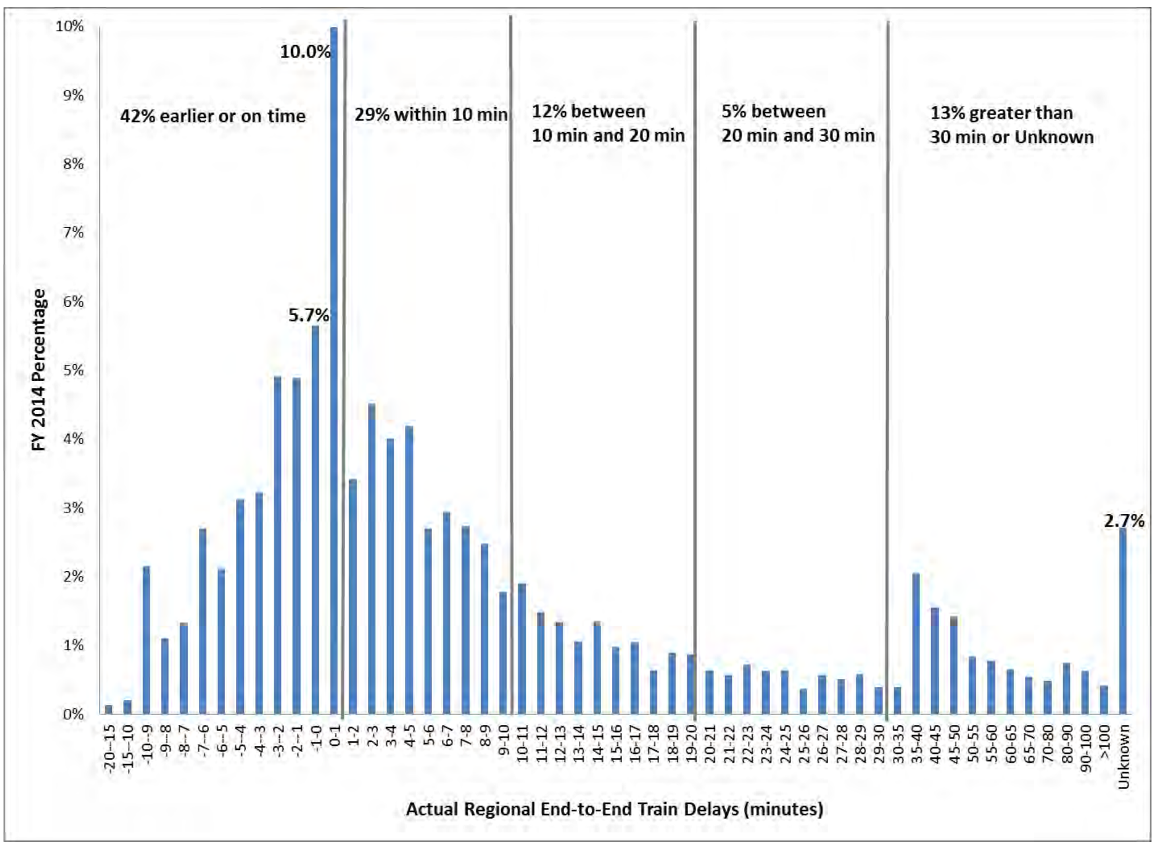


Figure 6.2: Distribution of Actual FY 2014 End-to-End Regional Train Delays

6.1.2 On-Time Performance versus Average Delay

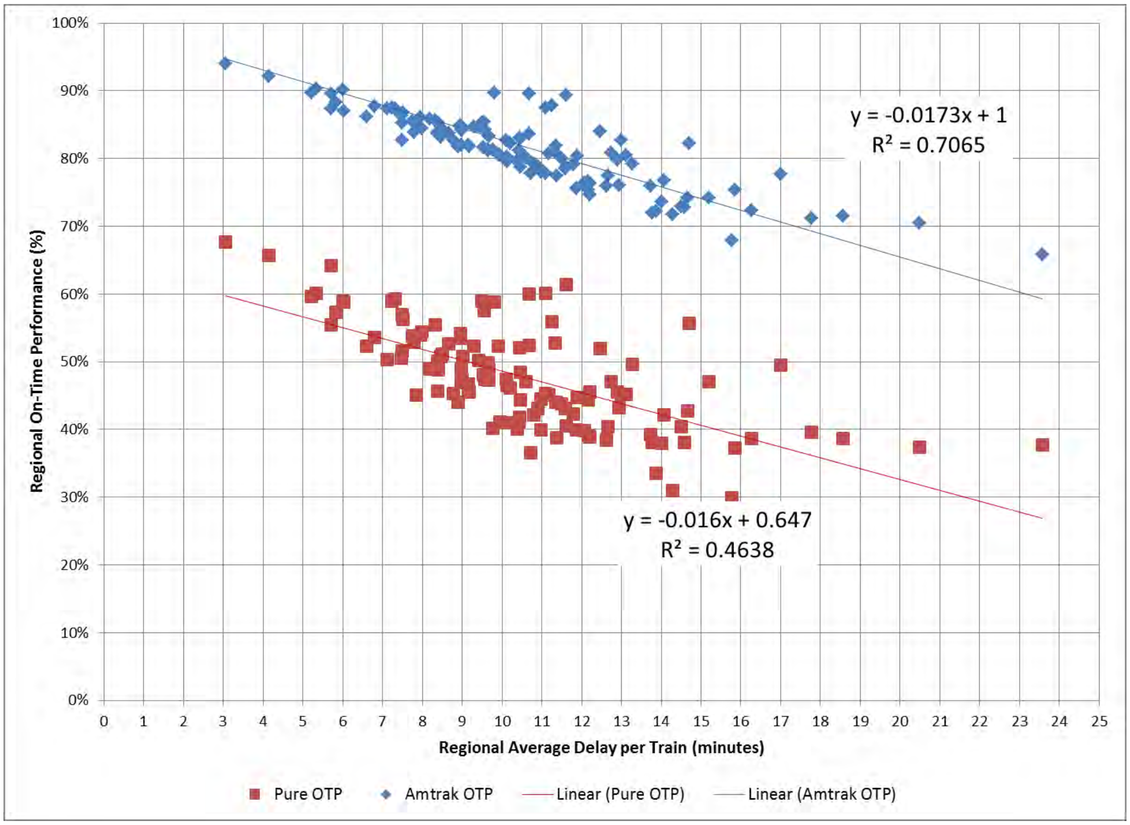


Figure 6.3: Regional On-Time Performance versus Average Delay/Train

Figure 6.3 shows the relationship between the monthly on-time performance and average delay per train on the Regional trains operated between FY 2005 and FY 2014. Currently, Amtrak's on-time threshold for Regional trains is 10 minutes for trains under 250 miles, and 20 minutes for trains over 250 miles. The blue data points represent Amtrak OTP (assuming a 10-20 minute delay tolerance) while the red data points represent Pure OTP (assuming a zero-delay tolerance). The simple regression shows a linear relationship between OTP and average delay per train, including the linear trend line equation and correlation factor (R^2). Unlike the Acela, some of the data points for Regional do not fall close to the linear function, which suggests that the monthly average delay per train and monthly OTP are not highly correlated. This is likely due to the fact unlike the Acela service, the on-time threshold on the Regional service varies by distance, and the number of scheduled Regional services also varies by day of week. That said, the R^2 between the monthly average delay per train and monthly Amtrak OTP is about 70% (compared to 87% on the Acela), which suggests that although the correlation is not perfect (equal to 1), the metrics are still sufficiently correlated and can be substituted for one another if needed.

6.2 Regional Annual and Monthly Riders, On-Time Performance and Delays

This section investigates the annual, seasonal and month of year fluctuations in the ridership and performance of Amtrak's Regional service between FY 2005 to FY 2014.

6.2.1 Annual Regional Riders

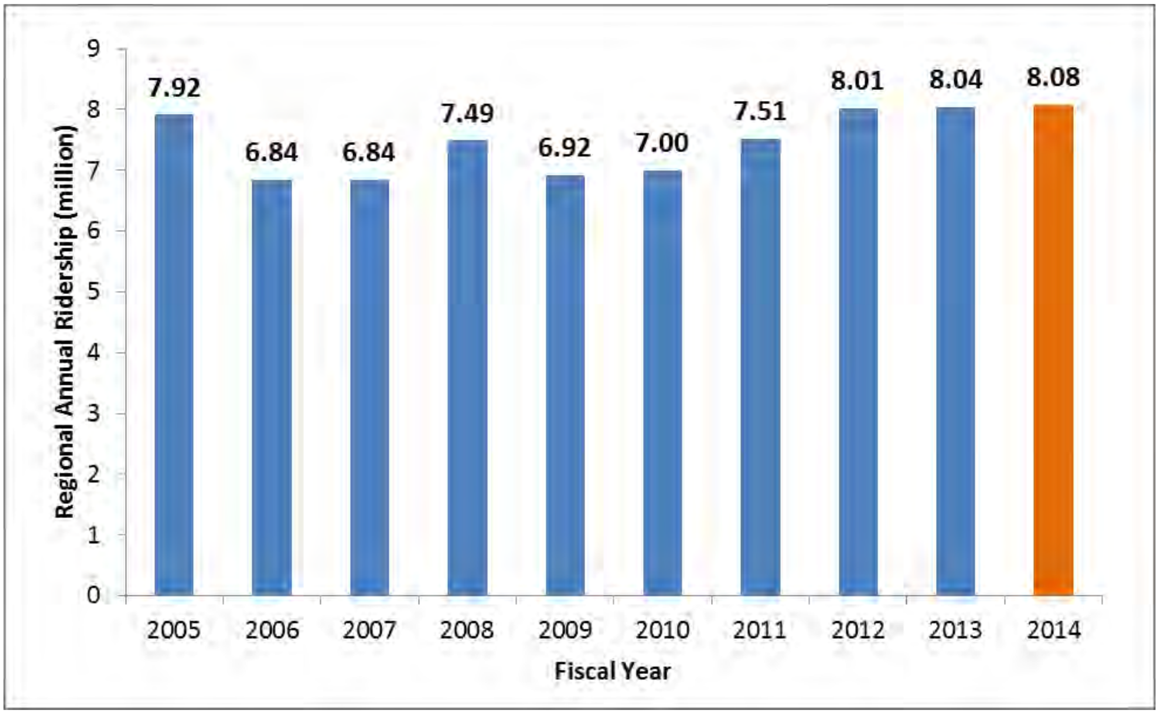


Figure 6.4: Regional Annual Ridership

Figure 6.4 shows the total Regional ridership between FY 2005 and FY 2014, highlighting the year-to-year variations. Unlike Acela that saw a 13% increase in ridership between FY 2005 and FY 2006, the Regional service experienced a 14% ridership decrease between FY 2005 and FY 2006. This was likely due to cannibalization of Regional demand by Acela service as it continued to grow in popularity around FY 2006. Regional ridership remained steady at 6.8 million in FY 2006 and FY 2007, and increased by 10% in FY 2008. Similar to the Acela service, Regional demand was greatly impacted by the recession that lasted through late 2008 and 2009, which led to an 8% drop in FY 2009 ridership. Following the recession, Regional ridership grew by 7% in both FY 2011 and FY 2012, but seems to have plateaued in the last three years. However, in FY 2014 Regional service experienced a record high of 8.1 million riders, which compared to pre-recession in FY 2008 was an 8% growth. The seasonal and month of year variations are explored in the next section.

6.2.2 Riders by Month

Regional demand exhibits fairly consistent patterns within the months of each year. The seasons and holidays largely influence travel patterns and in turn, the demand and ridership on Regional trains. Figure 6.5 shows the total ridership by month on the Regional service from FY 2005 to FY 2014. Typically, the months with the lowest travel on the Regional are in January and February, coinciding with the end of the winter holiday and vacation, while the ridership at the beginning of each fiscal year (October, November and December), and through the spring and summer months are usually much higher. Excluding January and February, the monthly demand for Regional does not appear to vary very much. For example, the total demand in FY 2014 was roughly 700,000 each month (except in January and February). This suggests that Regional passengers might be regular riders, who have similar travel patterns throughout the year.

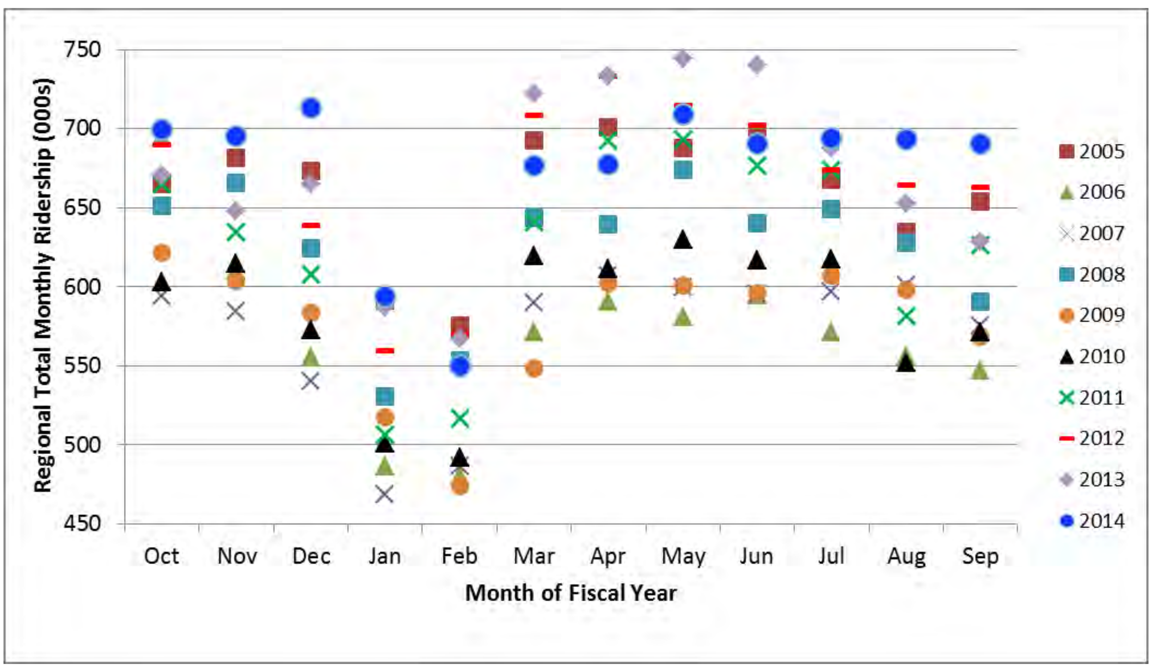


Figure 6.5: Regional Monthly Ridership

6.2.3 On-Time Performance by Month

Figure 6.6 shows Regional on-time performance by month for FY 2005 to FY 2014. Unlike Regional ridership, the on-time performance appears to vary between months. The winter months (November, December, January and February) experienced a wide variation in OTP. For example, the Regional OTP for the month of January was as low as 66% in 2014 but was as high as 89% in 2012. February also experienced a large variation, with an OTP range from 72% to 94% between FY 2005 and FY 2014. Unsurprisingly, the years with severe winters (e.g. 2005, 2010 and 2014) experienced worse performance. Essentially, the service performance in the winter months is largely dependent on the severity of the weather. Although the winter months suffered the worst performance, compared to other months, they also experienced the best service performance in years with less severe winters, while the summer months usually experienced worse performance. In other words, if the severe winter outliers are excluded in each fiscal year, the months with the highest OTP are usually in December, January, February or March, while the months with lowest OTP are usually June, July or August. This is likely because heat restrictions are usually imposed in the summer when the temperature of the track exceeds 120 degrees. The heat restrictions require trains to run at a slower speed, which leads to delays. Furthermore, the absence of constant tension catenary in the Northeast Corridor means that on hot days the overhead power lines droop, which is a safety hazard and also causes additional delays in summer months.

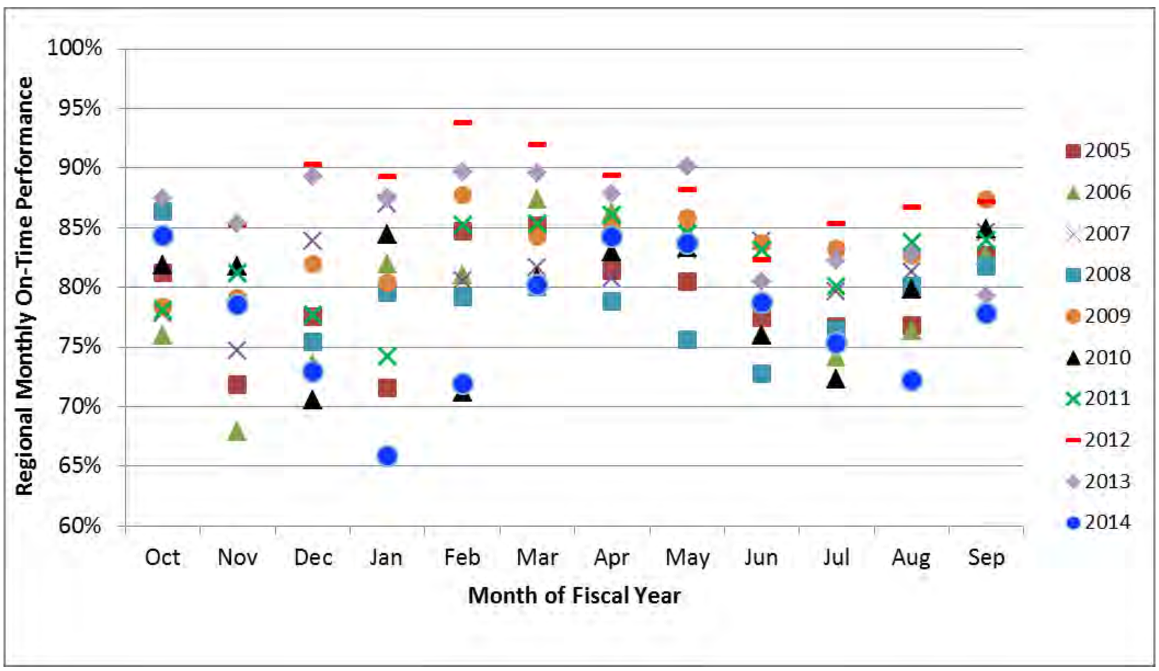


Figure 6.6: Monthly Regional On-Time Performance

Collectively, the Regional OTP in FY 2012 (red dashes) and FY 2013 (purple diamonds) were the highest, while the OTP in FY 2008 (blue squares) and FY 2014 (blue dots), were the lowest. Furthermore, comparing over all years, the Regional OTP in all months in FY 2014 was one of the worst, even after excluding the effect of the harsh winter weather – the Regional OTP was below 85% in all months of FY 2014.

6.2.4 Total Minutes of Delay by Month

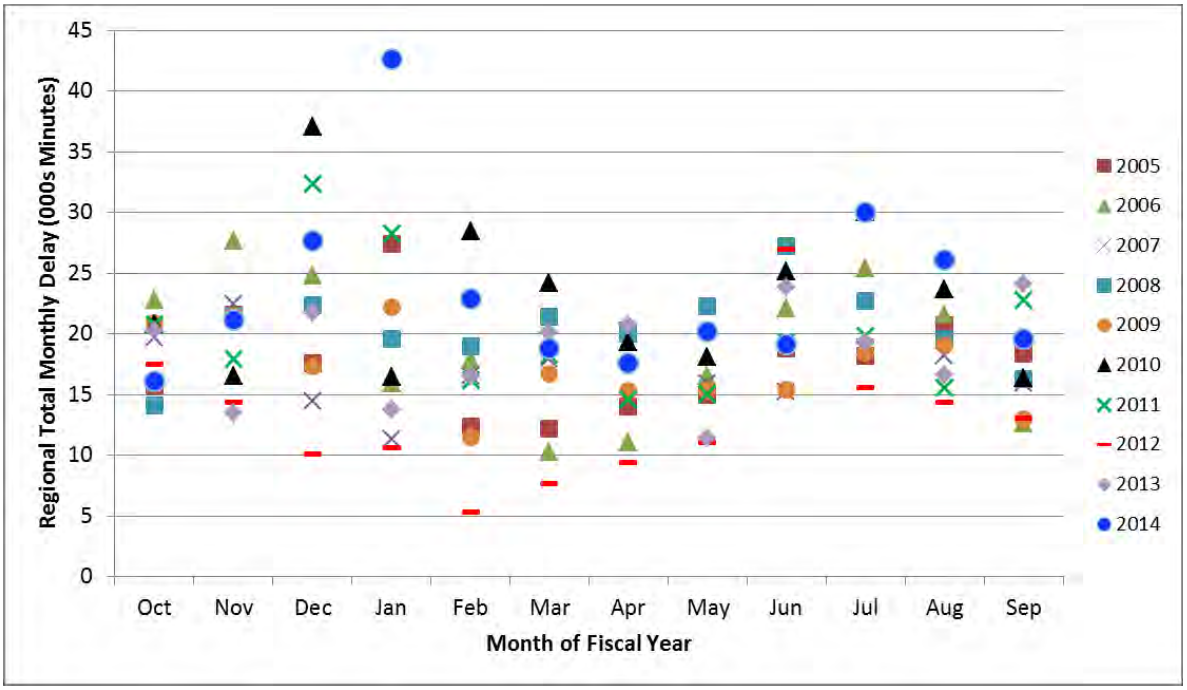


Figure 6.7: Regional Total Monthly Delay

The second supply side performance indicator that was used to investigate seasonal and day of month trends was the delay minutes incurred on operated trains. Figure 6.7 shows total Regional service delay by month for FY 2005 to FY 2014. Similar to the OTP trends, the winter months - December, January and February exhibited a wide variance in the total delay while the summer months exhibited less variance. For example, comparing across January, Regional trains incurred about 42,700 minutes of total delay in January 2014 but only a total of 10,600 delay minutes in January 2012, which is roughly equivalent to 23 minutes per train in January FY 2014 and only 6 minutes per train in January FY 2012. Similar to the observations from the OTP, although the winter months suffered the worst delays, compared to other months, they also experienced the least delays in years with less severe winters, while the summer months typically experienced higher delays. Furthermore, even though the performance in July was generally not the worst, the amount of delay incurred was usually relatively high (>15,000) compared to other months, likely due to summer track work.

Collectively, the total monthly Regional delay in FY 2012 (red dashes) and FY 2006 (green triangles) were the lowest (best), while the total monthly delay in FY 2010 (black triangles) and FY 2014 (blue dots) were the highest (worst).

In summary,

vii. In FY 2014 Regional service experienced a record high of 8.1 million riders

viii. Excluding January and February, Regional ridership exhibited minimal variation between the months of the year, which suggests that Regional riders have similar travel patterns throughout the year.

ix. Unlike the ridership, the performance on the Regional service varied within the same month in different years. Compared to summer months, the winter months exhibited a large variance due to the effects of mild and severe winter seasons on performance. Nonetheless, the best on-time performances were usually in the winter months.

x. Additionally, the performance on the Regional service varied across different months in the same fiscal year. Although the summer months exhibited less year-to-year variance, on a month-to-month comparison (excluding the severe winter months), the amount of delays were usually higher in the summer due to routinely scheduled track works, heat restrictions and infrastructures issues (catenary wire drooping).

xi. Between FY 2005 and FY 2014, the best Regional performance was in FY 2012 while the worst performance was in FY 2014.

xii. Compared to the Acela service, Regional service had almost 2.5 more riders in FY 2014. Additionally, Regional trains experienced more delays, probably because the Regional is a more local service with more frequent stops along the corridor.

6.3 Daily Variations in Performance and Ridership

This section begins with an overview of the impacts of service performance on daily train operations and daily ridership, and then focuses on the day-of-week and time-of-day variations on both Regional ridership and service performance. Except for the day of week analysis, the daily variations are presented only for the years where Regional service experienced the best performance (FY 2012) and the worst performance (FY 2014). The reasoning behind this is that the systematic portion of the day-of-week and time-of-day fluctuations in the other years not shown are likely similar to either FY 2012 or FY 2014, thus presenting the redundant information might not add additional value to the discussion.

6.3.1 FY 2012 and FY 2014 Daily Delays: Trains

In both FY 2012 and FY 2014, Amtrak scheduled 64 Regional trains on Mondays, Tuesdays and Wednesdays, 65 trains on Thursdays, 67 trains on Fridays, 51 trains on Saturdays, and 53 trains on Sundays, except on holidays. Figure 6.8 shows the daily total and delayed Regional trains in FY 2012, with the gray area representing the total number of scheduled trains. The spike in the chart occurs on the days before Thanksgiving on which Amtrak scheduled 9 additional Regional trains.

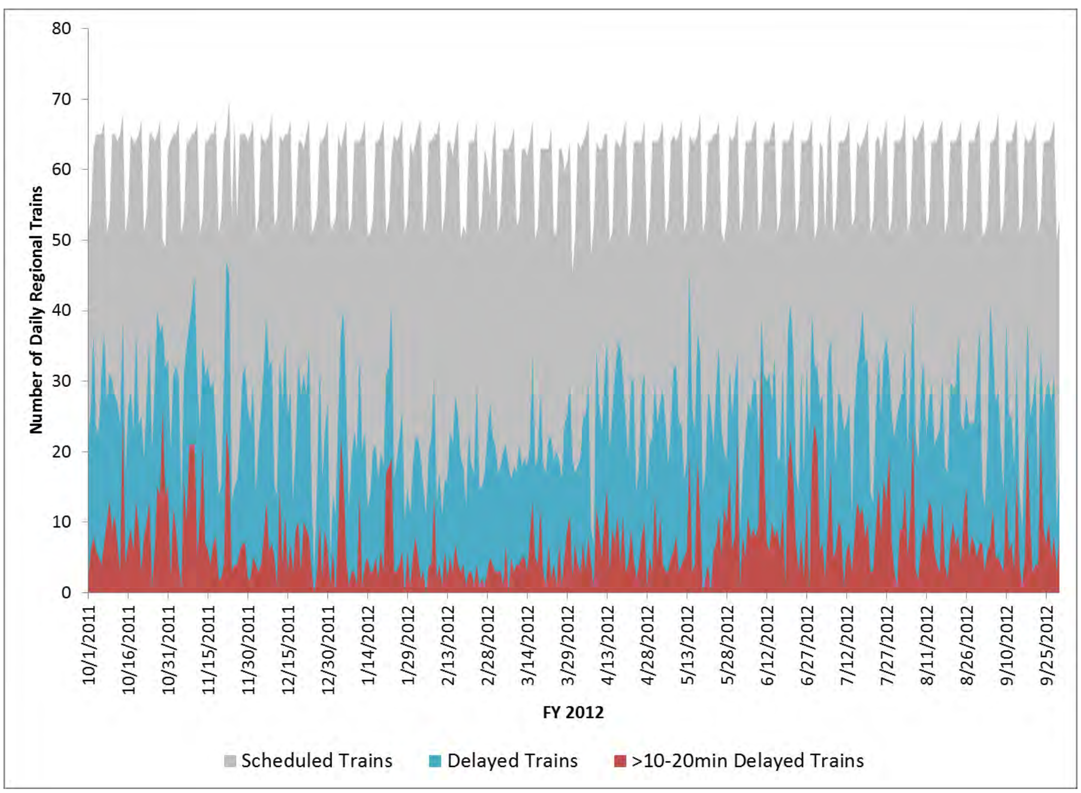


Figure 6.8: FY 2012 Total and Delayed Regional Trains

The blue bars represent the total number of trains in FY 14 that arrived at their final destination after the scheduled arrival time, while the red bars represent those that arrived late under Amtrak's PRIIA standards with a delay greater than 10 minutes for Regional trips under 250 miles and greater than 20 minutes for Regional trips over 250 miles.

In FY 2012, about 41% of all scheduled Regional trains (e.g. 26 of 64 trains on a given Monday) arrived at their final destination after the scheduled arrival time, and about 12% (e.g. 8 of 64 trains on a given Monday) arrived with a delay greater than the PRIIA standards (10 to 20 minutes depending on trip distance). Compared to Acela trains in FY 2012, about 6% more Regional trains were late.

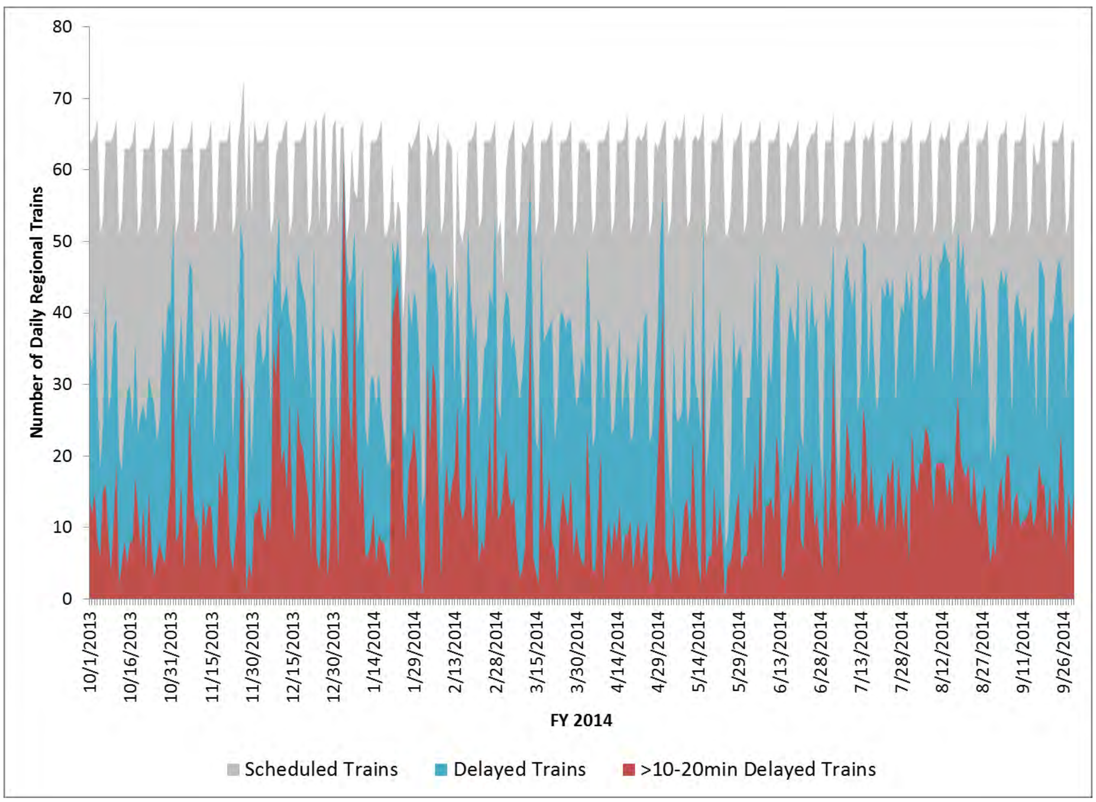


Figure 6.9: FY 2014 Total and Delayed Regional Trains

Figure 6.9 shows the blue and red bars in FY 2014 are much longer than in FY 2012, indicating a higher number of delayed trains compared to schedule. In FY 2014, 58% of all scheduled Regional trains (e.g. 37 of 64 trains on a given Monday) arrived at their final destination after the scheduled arrival time (blue bars), and roughly 23% of Regional trains (15 of 64) arrived more than 10 or 20 minutes later than the scheduled arrival time.

6.3.2 FY 2012 and FY 2014 Daily Delays: Riders

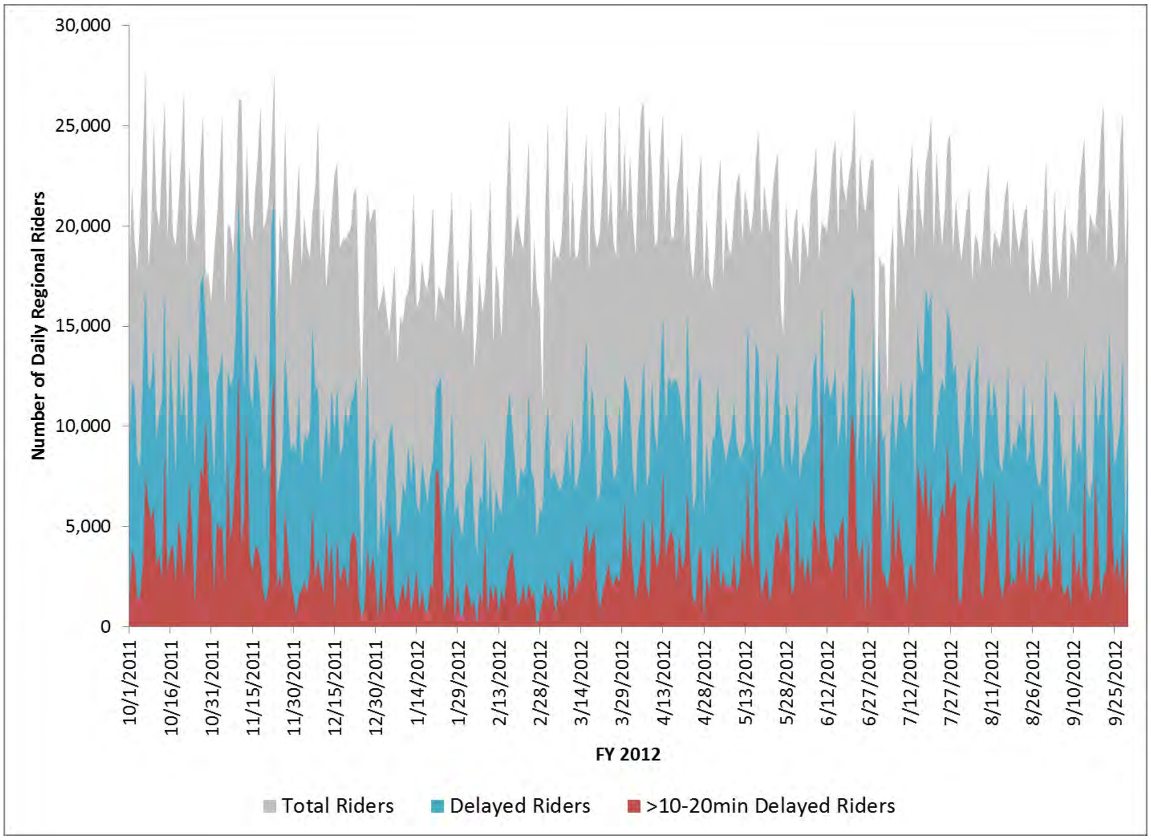


Figure 6.10: FY 2012 Total and Delayed Regional Riders

In Figure 6.10 the gray bars corresponds with the total number of daily Regional riders In FY 2012, the blue bars corresponds with the number of daily riders that arrived at their destination after the scheduled arrival time, and the red bars show the number of daily Regional riders that experienced delays greater than the PRIIA threshold, that is, 10 minutes for trips under 250 miles and greater than 20 minutes for trips over 250 miles.

The gray area shows low daily ridership between 12/15/2013 and 2/13/2014, and peak daily ridership around the 10/13/2014 and 11/30/2012 over the Columbus and Thanksgiving holidays. In FY 2012, half of all Amtrak Regional riders, that is about 4 million riders arrived at their destination after the scheduled arrival time, and 18% of Amtrak Regional passengers, that is about 1.5 million riders arrived at their destination on a train that was more than 10 or 20 minutes late. Compared to Acela riders in FY 2012, about 6% more Regional riders experienced delays.

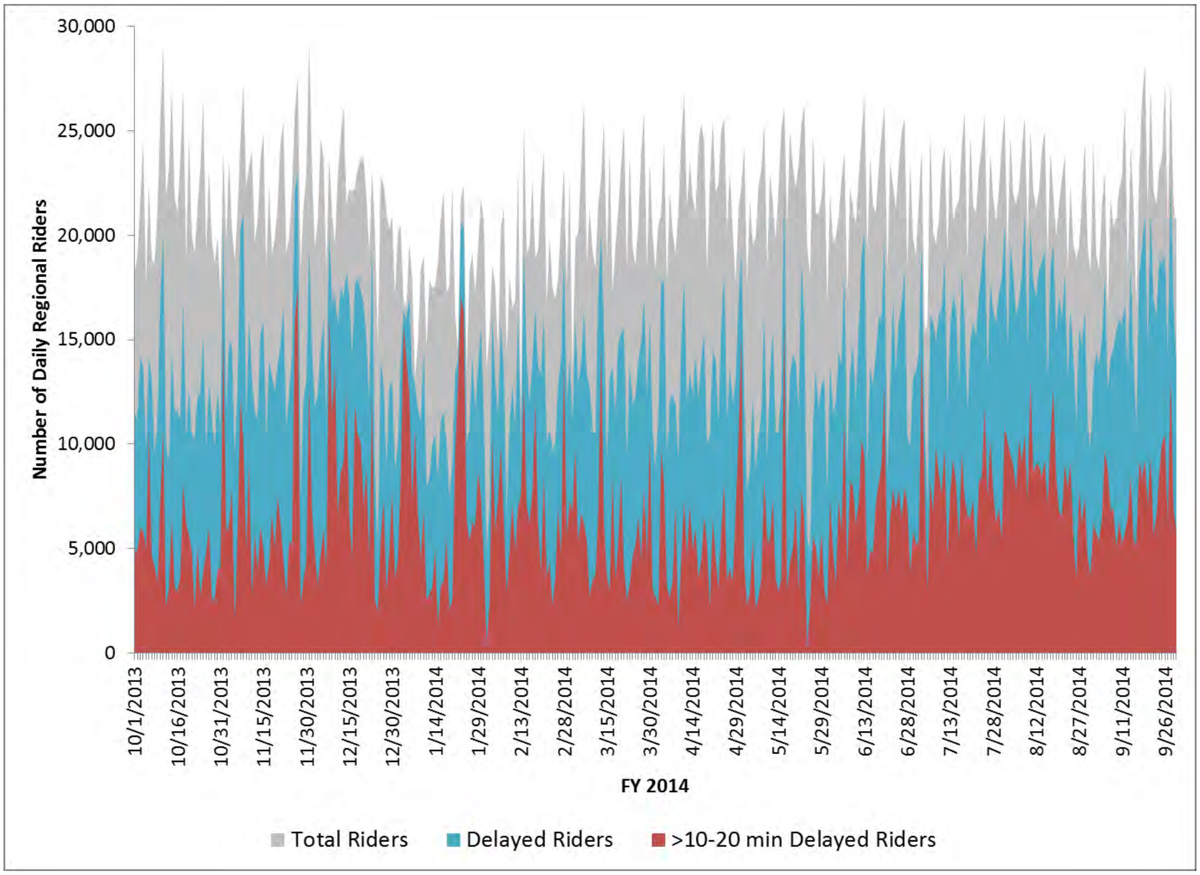


Figure 6.11: FY 2014 Total and Delayed Regional Riders

Figure 6.11 shows the FY 2014 total and delayed Regional riders. Similar to the earlier observation, the gray area shows low ridership over the winter holiday, and peak ridership around the Columbus and Thanksgiving holidays. In FY 2014, 66% of all Amtrak Regional riders, that is about 5.3 million riders arrived at their destination after the scheduled arrival time, and 31% of Amtrak Regional passengers, that is about 2.5 million passengers arrived at their destination on a train that was more than 10 to 20 minutes late. Compared to the Acela riders in FY 2014, roughly the same proportion of Regional riders experienced delays. Furthermore on the Regional, some days in January (red spikes between 12/30/2014 and 1/29/2014) impacted by the Polar Vortex had more than 90% of Regional passengers experiencing some delay, and more than 70% experienced delays above the 10 to 20 minute PRIIA threshold.

In summary, even in the best performing year, FY 2012 as many as 41% of scheduled trains (26 of 64) and 50% of traveling passengers arrived at their final destination after the scheduled arrival time, and as many as 12% (8 of 64) of trains and 18% of passengers arrived late with delays greater than 10-20 minutes. By FY2014, the numbers of late trains and late passengers had almost doubled.

6.3.3 Day of Week Performance

Figure 6.12 shows the average daily ridership (dotted lines) and average daily delay per mile (solid lines) on Regional trains by day of week for FY 2014 (in red), FY 2012 (in blue), and averages over FY 2005 to FY 2014 (in green). The averages over FY 2005 to FY 2014 were included in this analysis to ensure that random disruptions or calendar effects that might have affected Regional operations and service performance on a specific weekday did not bias the day of week patterns. In other words, the FY 2005-FY 2014 values reflect averages over many more days.

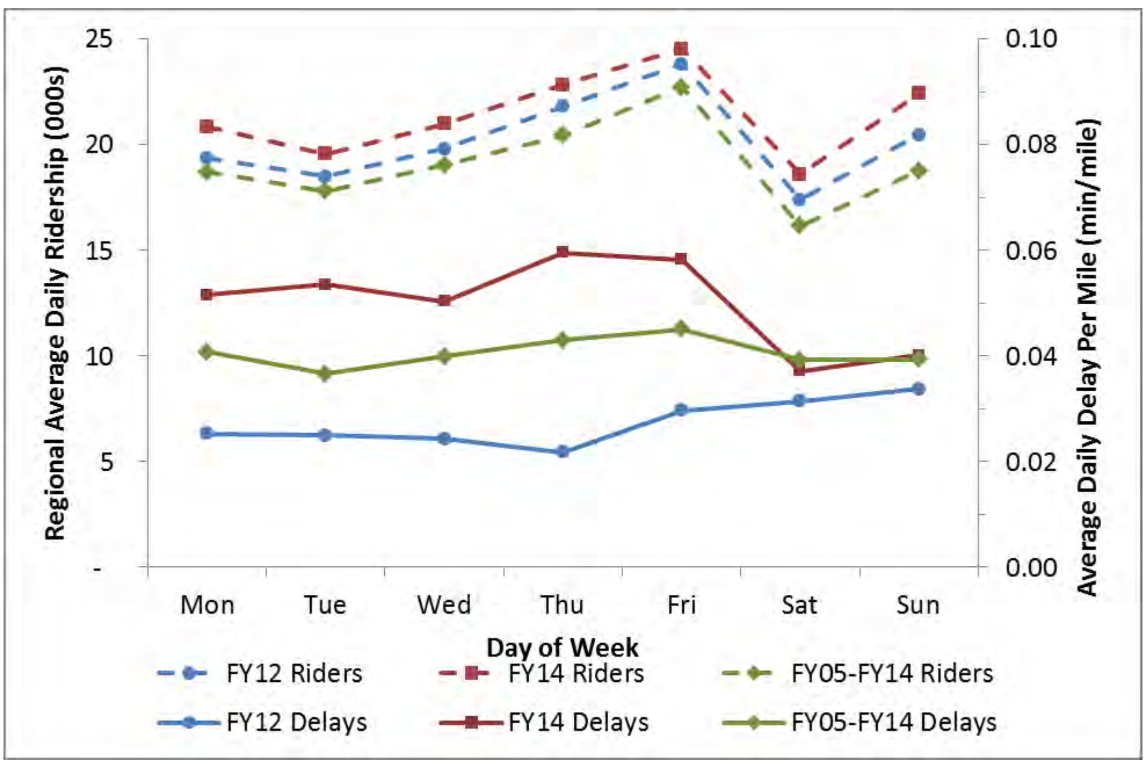


Figure 6.12: Regional Average Ridership and Delay by Day of Week

Firstly, focusing on the ridership (dotted lines), the Regional trend by day of week exhibited the same pattern in FY 2012, FY 2014, as well as over the 10-year period (FY 2005 – FY 2014). The average ridership on Regional was typically lowest on Saturdays compared to the average weekday and Sunday ridership. Unlike the Acela, which had low ridership volumes on both Saturdays and Sundays, the Regional ridership level on Sundays was comparable to weekday ridership. This is likely due to the fact that Acela riders are mainly business passengers traveling during the week while Regional riders are more leisure passengers with majority of trips at the end of the week on Fridays and as the week begins on Sundays. Of the weekdays, the ridership on Tuesdays was typically the lowest, followed by similar ridership levels on Mondays and Wednesdays, higher ridership on Thursdays, and peak ridership on Fridays. In FY 2012, the average daily ridership on Saturdays was 17,400 compared to 20,500 on Sundays, and on average 20,700 on weekdays. In FY 2014, the average daily ridership on Saturdays was 18,600 compared to 22,400 on Sundays, and on average 21,800 on weekdays.

The average delay per mile (solid lines in Figure 6.12) exhibited different day of week patterns in FY 2012 and FY 2014. This emphasized the value of presenting the averages over the 10-year period (FY 2005 to FY 2014), which might reflect patterns closer to the true day of week proportions since they represent averages over many more days. For example, in FY 2012 and FY 2014, there were only 53 Mondays; however, there were 523 Mondays between FY 2005 to FY 2014. From the FY 2005 to FY 2014 averages (solid green line), although the delays on the Regional service were similar for all days of the week, Regional trains appeared to encounter slightly more delays on Mondays and Fridays and slightly fewer delays on Tuesdays and Saturdays. The OTP by day of week on the Regional service were explored to see if they exhibited a day of week pattern similar to average delay.

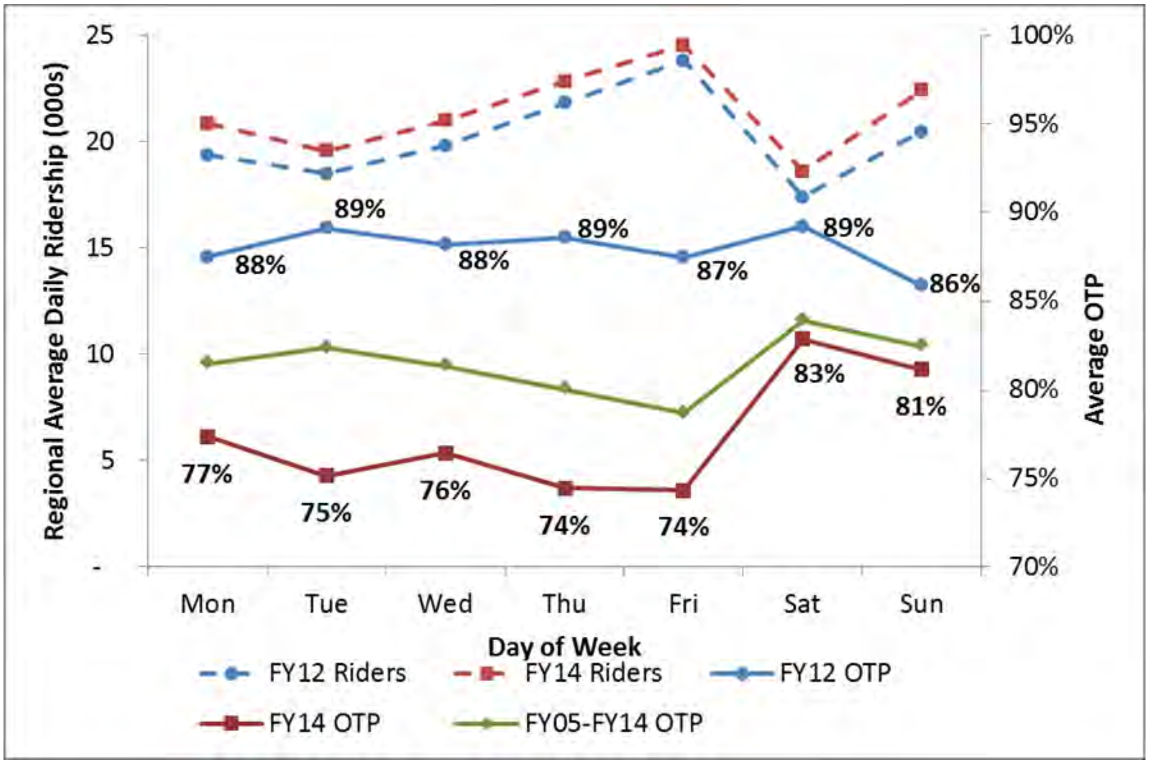


Figure 6.13: Average Ridership and OTP by Day of Week

Figure 6.13 shows a similar chart (to Figure 6.12) but with average on-time performance instead of average delay by day of week. The dotted lines represent ridership and the solid lines represent on-time performance (OTP). In FY 2012 the average OTP (solid blue line) was roughly the similar on all days of the week, while in FY 2014 the average OTP (solid red line) varied considerably by day of week. In comparison, over the 10-year averages (FY 2005 – FY 2014 in green), which theoretically provide a more accurate representation of actual day of week performances, Regional performance appeared to be the best on Saturday and Sundays, and worst on Fridays. This seems reasonable since the number of scheduled Amtrak trains is lower and commuter services usually also have fewer trains in operation on weekends.

Furthermore, the performance on the Regional service looked like the reverse of the Regional ridership. This suggests that higher Regional demand on certain days of the week might be associated with poorer performance.

In summary, Regional ridership volumes were lowest on Saturdays and highest on Fridays. Additionally, unlike Acela, Regional ridership on Sundays was relatively high. The high ridership volumes are likely because Regional riders are more leisure passengers with majority of trips at the end of the week on

Fridays and as the week begins on Sundays. Finally, of the weekdays, Tuesdays had the lowest ridership. In terms of performance, overall, performance appeared to be roughly the same on all days of the week; however, it was usually slightly better on Tuesdays, Saturdays and Sundays, and slightly worse on Fridays. The weekend improvements were likely because fewer Amtrak and Commuter services operate on weekends.

6.3.4 Time of Day Performance

This section explores the performance of Regional trains by time of day. However the Northeast Corridor is very long and each train crosses multiple time periods between departure at originating station and arrival at the terminating station. Consequently, the time of day analysis in this section is based on the departure times of train by direction and originating station. A more comprehensive time of day analysis might include train departure and arrival times at each station on the NEC.

In the time of day analysis, each Regional train is categorized by market and further classified into origin-destination routes as: NYP-WAS, NYP-PHL (Keystone trains that serve NYP and PHL but terminate in the Keystone segment west of PHL), BOS-NYP, and BOS-WAS. Table 6.1 summarizes the level of service characteristics of the Regional train routes. The scheduled travel times are provided as ranges to indicate that trains are sometimes scheduled with different end-to-end travel times based on the actual stopping patterns of the train.

Service	Route	Market	Distance (miles)	Scheduled Travel Time (hr:min)
Regional	New York – Washington (NYP–WAS)	South End	226	3:12 to 3:39
	New York – Philadelphia (NYP–PHL)	South End	91	1:20 to 1:25
	Boston - New York (BOS-NYP)	North End	231	4:00 to 4:20
	Boston – Washington (BOS-WAS)	Through	457	7:40 to 8:05

Table 6.1: Level of Service for Regional Train Route Groups

Table 6.2 shows the average Amtrak OTP (10-minute delay threshold) and average delay per train (no threshold) for southbound (SB) Regional trains on weekdays in FY 2012 and FY 2014. The averages are over the 261 weekdays each year. The southbound trains are further classified into origin-destination routes (as summarized in Table 6.1). Intuitively, the NYP-PHL trains are likely to have the best performance due to the short trip length and fewer stops, and the BOS-WAS trains are likely to encounter the most delays since they travel the longest distances and have more stops. This speculation is confirmed in Table 6.2, which shows that the southbound NYP-PHL trains had the best performance in both FY 2012 and FY 2014, and the BOS-WAS trains had approximately 10% lower OTP than the NYP-WAS trains, and also roughly double the amount of average delay per train on the NYP-WAS trains. This suggests that delays accumulate along the length of the corridor.

Weekdays Only (261 of 365 days)					FY 2012		FY 2014	
Train No	Route	Direction	O-D	Departure Time	Amtrak OTP*	Avg. Delay per train (min)	Amtrak OTP*	Avg. Delay per train (min)
151	Regional	SB	NYP-WAS	4:40:00 AM	96%	1.73	86%	6.14
111	Regional	SB	NYP-WAS	5:30:00 AM	94%	2.68	85%	7.21
181	Regional	SB	NYP-WAS	6:10:00 AM	81%	5.95	78%	7.98
79	Regional	SB	NYP-WAS	7:05:00 AM	77%	8.41	66%	13.69
183	Regional	SB	NYP-WAS	7:17:00 AM	91%	4.67	72%	10.72
185	Regional	SB	NYP-WAS	8:10:00 AM	88%	4.64	69%	14.67
125	Regional	SB	NYP-WAS	11:35:00 AM	85%	5.76	63%	15.86
133	Regional	SB	NYP-WAS	1:09:00 PM	63%	12.01	53%	17.02
85	Regional	SB	NYP-WAS	3:05:00 PM	78%	8.42	67%	15.32
127	Regional	SB	NYP-WAS	4:05:00 PM	88%	4.77	76%	11.86
129	Regional	SB	NYP-WAS	4:42:00 PM	80%	7.77	69%	18.58
193	Regional	SB	NYP-WAS	5:39:00 PM	85%	7.52	68%	16.89
187	Regional	SB	NYP-WAS	9:20:00 PM	85%	9.29	75%	13.21
641	Regional	SB	NYP-PHL	7:25:00 AM	93%	3.84	87%	6.65
643	Regional	SB	NYP-PHL	9:30:00 AM	94%	2.49	88%	5.88
43	Regional	SB	NYP-PHL	10:52:00 AM	92%	3.40	72%	10.65
645	Regional	SB	NYP-PHL	12:05:00 PM	94%	2.61	85%	6.09
647	Regional	SB	NYP-PHL	2:11:00 PM	92%	3.65	89%	5.05
649	Regional	SB	NYP-PHL	2:44:00 PM	90%	3.50	89%	3.94
651	Regional	SB	NYP-PHL	4:03:00 PM	93%	1.80	95%	2.39
653	Regional	SB	NYP-PHL	5:10:00 PM	92%	5.24	83%	8.31
655	Regional	SB	NYP-PHL	6:35:00 PM	84%	6.48	69%	11.96
639	Regional	SB	NYP-PHL	11:15:00 PM	92%	3.73	87%	6.54
95	Regional	SB	BOS-WAS	6:10:00 AM	87%	10.40	65%	27.79
195	Regional	SB	BOS-WAS	6:40:00 AM	90%	10.04	64%	25.30
171	Regional	SB	BOS-WAS	8:15:00 AM	90%	7.98	71%	24.28
93	Regional	SB	BOS-WAS	9:30:00 AM	86%	9.00	57%	29.63
83	Regional	SB	BOS-WAS	9:30:00 AM	56%	24.89	48%	29.86
173	Regional	SB	BOS-WAS	11:15:00 AM	89%	9.28	64%	25.73
137	Regional	SB	BOS-WAS	1:40:00 PM	86%	12.31	76%	17.33
175	Regional	SB	BOS-WAS	3:20:00 PM	89%	8.67	69%	22.90
177	Regional	SB	BOS-WAS	5:35:00 PM	85%	12.95	77%	23.84
179	Regional	SB	BOS-NYP	6:45:00 PM	76%	10.15	67%	9.72
67	Regional	SB	BOS-WAS	9:30:00 PM	96%	2.80	92%	8.36
55	Regional	SB	SAB-WAS	8:58:00 AM	88%	14.07	59%	30.23
141	Regional	SB	SPG-WAS	5:55:00 AM	86%	8.54	69%	18.74

*Amtrak OTP for the longest segment with passengers, which is not necessarily the end point station

Table 6.2: Regional SB FY 12 and FY 14 Weekday OTP and Average Delay per Train

In the NYP-WAS group, Train 151 departing NYP at 4:40AM exhibited the best performance, while Train 133 departing at 1:09PM exhibited the worst performance in both FY 2012 and FY 2014.

In the NYP-PHL group, Train 651 departing NYP at 4:03PM exhibited the best performance while Train 655 departing at 6:35PM exhibited the worst performance in both years.

In the BOS-WAS group, the 9:30PM departure from BOS, Train 67 exhibited the best performance while the 9:30AM departure from BOS, Train 83 exhibited the worst performance in both years.

Firstly, it is interesting that even though the overall performance of Regional trains in FY 2012 was much better than in FY 2014, the best and worst performing trains were the same. This suggests that there are systematic disturbances that disrupt the trains on a regular basis. Furthermore, the best and worst performing trains were not time-of-day consistent across the different groups. For example, the first train of the day at 4:40AM was the best in the NYP-WAS group but one of the last trains of the day at 9:30PM performed the best in the BOS-WAS group. The fact that some of the morning trains exhibited bad performance while some of the afternoon and end of day trains exhibited better performance suggests that the cascading effects from trains interfering with each other might not be present in the data.

Table 6.7 shows the Regional trains that travel northbound (NB) on the Northeast Corridor. Although similar to the SB trains, the NB trains exhibited worse performance as trip length increased; collectively, the NB trains performed better than the SB trains. In addition, comparisons between the best and worst performing trains in FY 2012 and FY 2014 show similar relative performances, again suggesting systematic disturbances on the NEC. Lastly, likewise the best and worst performing trains in each segment were not time-of-day consistent, which suggest that trains are not necessarily interfering with one another

In summary, similar to the Acela service, the Regional trains exhibited distance-related deteriorations, suggesting that delays accumulate as trains stopped to serve more stations along the corridor. Furthermore, just like the Acela, the first Regional train of the day was not able to achieve consistent on-time arrivals even though there was no other train or at least fewer trains in the system to slow them down. Consequently, a detailed station-level analysis of the first Regional trains of the day was done and discussed in Section 6.4

Weekdays Only (261 of 365 days)					FY 2012		FY 2014	
Train No	Route	Direction	O-D	Departure Time	Amtrak OTP*	Avg. Delay per train (min)	Amtrak OTP*	Avg. Delay per train (min)
190	Regional	NB	WAS-BOS	3:15:00 AM	91%	8.36	83%	14.14
170	Regional	NB	WAS-BOS	4:52:00 AM	87%	9.23	76%	15.71
172	Regional	NB	WAS-BOS	7:25:00 AM	87%	9.62	68%	21.97
56	Regional	NB	WAS-SPG	8:10:00 AM	84%	8.61	69%	16.88
86	Regional-RVR	NB	WAS-BOS	8:40:00 AM	90%	6.52	73%	18.34
174	Regional-NFK	NB	WAS-BOS	10:20:00 AM	91%	9.59	75%	18.36
176	Regional-LYH	NB	WAS-BOS	12:02:00 PM	87%	12.71	83%	13.70
94	Regional-NPN	NB	WAS-BOS	2:02:00 PM	86%	10.42	46%	38.93
148	Regional	NB	WAS-SPG	3:02:00 PM	82%	11.01	65%	17.98
178	Regional	NB	WAS-BOS	4:02:00 PM	91%	5.87	81%	20.95
136	Regional	NB	WAS-SPG	5:05:00 PM	88%	11.96	79%	27.38
66	Regional-NPN	NB	WAS-BOS	10:10:00 PM	91%	7.24	87%	11.42
110	Regional	NB	WAS-NYP	4:00:00 AM	84%	6.76	67%	12.58
180	Regional	NB	WAS-NYP	5:30:00 AM	85%	7.58	66%	13.47
130	Regional	NB	WAS-NYP	6:30:00 AM	86%	5.28	77%	8.31
184	Regional	NB	WAS-NYP	9:20:00 AM	81%	8.25	68%	11.37
84	Regional-RVR	NB	WAS-NYP	11:02:00 AM	90%	3.78	75%	9.45
186	Regional	NB	WAS-NYP	1:02:00 PM	90%	4.26	68%	10.64
134	Regional	NB	WAS-NYP	3:30:00 PM	80%	5.89	77%	9.12
80	Carolinian	NB	WAS-NYP	4:06:00 PM	79%	17.32	33%	49.30
196	Regional	NB	WAS-NYP	5:05:00 PM	89%	5.40	79%	9.61
138	Regional	NB	WAS-NYP	6:05:00 PM	79%	9.50	73%	15.99
188	Regional	NB	WAS-NYP	7:10:00 PM	92%	4.57	76%	13.45
198	Regional	NB	WAS-NYP	9:05:00 PM	85%	23.03	70%	22.13
640	Keystone	NB	PHL-NYP	7:00:00 AM	82%	7.89	59%	15.17
642	Keystone	NB	PHL-NYP	9:45:00 AM	90%	4.30	79%	7.65
644	Keystone	NB	PHL-NYP	10:55:00 AM	93%	3.16	74%	10.68
646	Keystone	NB	PHL-NYP	11:45:00 AM	93%	2.36	89%	5.05
648	Keystone	NB	PHL-NYP	1:00:00 PM	97%	1.01	86%	5.41
650	Keystone	NB	PHL-NYP	2:05:00 PM	92%	2.80	89%	4.47
42	Pennsylvanian	NB	PHL-NYP	3:25:00 PM	91%	4.19	81%	8.36
652	Keystone	NB	PHL-NYP	5:18:00 PM	84%	7.92	81%	5.69
654	Keystone	NB	PHL-NYP	6:50:00 PM	96%	2.23	91%	4.56
656	Keystone	NB	PHL-NYP	7:40:00 PM	97%	1.25	88%	4.70
658	Keystone	NB	PHL-NYP	8:35:00 PM	91%	2.78	89%	5.26

Table 6.3: Regional NB FY 12 and FY 14 Weekday OTP and Average Delay Per Train

6.4 First Train Analysis

This section examines the first train of the day in each direction within each market (north-end, south-end and through). In a system where trains interfere with one another, one would expect to see better performing trains in the morning when there are either no other trains or fewer trains in the system to slow each other down. However, the discussion in Section 6.3 showed that this is not the case in Amtrak's Regional system, and the first train of the day routinely encounters considerable amounts of delay. Consequently, this section attempts to track the detailed station-to-station arrivals and departures of the first Regional trains over multiple days to identify causes of delay.

The discussion is divided into three subsections: i) northbound departures from Washington, DC (WAS), (since the first NB Regional train of the day is a through train which serves both the north and south segment, it was not necessary to analyze other NB trains), ii) southbound departures from New York Penn Station (NYP), and iii) southbound departures from Boston South Station (BOS). In each subsection, the average arrival and departure delays for the train at successive stations on the corridor are presented. The delays are calculated as differences between the actual and scheduled arrival and departure times, and are averaged over the 261 weekdays of the year. In the delay calculations, Amtrak was not credited for 'negative delays', that is, arrivals that occurred before the scheduled arrival time are treated as on-time arrivals with zero delays.

6.4.1 First Northbound Regional Train From WAS - Train 190 (3:15AM)

Train 190 is the first northbound Regional train scheduled to depart WAS at 3:15AM and arrive BOS at 11:05AM on weekdays. In FY 2014, it had an on-time performance of 83% and average delay of 14 minutes. Amtrak's scheduled dwell times for Train 190 are – 2 minutes in BAL, 5 minutes in PHL, and 15 minutes in NYP. Figure 6.14 shows the average FY 2014 arrival (light blue) and departure (dark blue) delays for the train at each successive station in the corridor. Firstly, overall the dark blue departure bars are higher than the light blue arrival bars, likely because Amtrak does not schedule terminal time at all stations. Secondly, as assumed in Section 6.3, the delays appear to accumulate along the length of the corridor. Regarding delay accumulation, two unusual delay reductions occur en-route and are highlighted in the chart. The first is at NYP, which shows a 2-minute drop between NWK and NYP, and the second is at RTE, which shows an 8-minute drop between PVD and RTE, suggesting that the timetables are padded to provide an opportunity for the train to "catch up". It was also unusual that despite the dwell time at NYP, the amount of delay jumped by about 4 minutes immediately north of NYP in the segment entering Stamford, CT (STM).

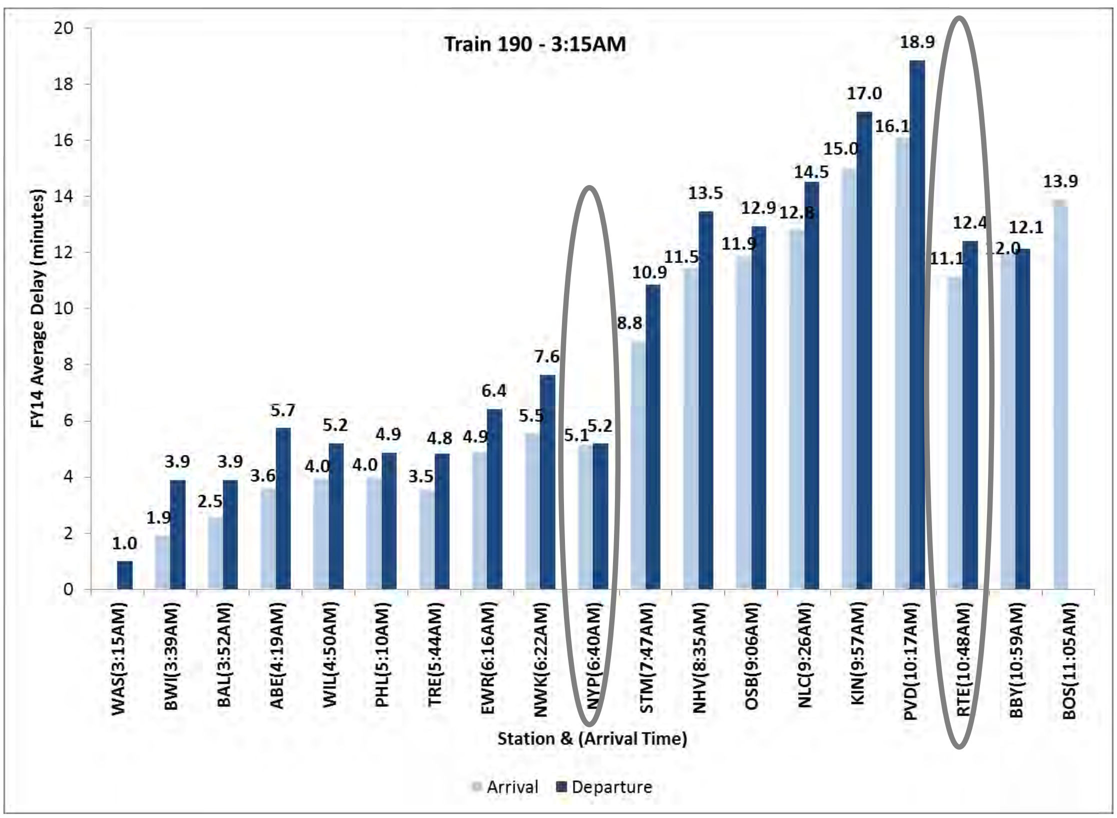


Figure 6.14: FY 2014 Average Station Delays for Train 190

Table 6.4 shows the performance of Train 190 from the segment level and timetable point of view. It shows that in the 32 mile PVD-RTE segment, although the train was scheduled to travel at 62mph, on average it traveled 21mph faster and thus was able to traverse the segment about 8 minutes faster. A similar effect was observed between the scheduled and actual performance on Acela trains in the PVD-RTE segment. This explains the 8-minute delay reduction at RTE that was observed in Figure 6.14. Additionally, although the discrepancy is not as much, Table 6.4 also shows that the train on average travelled about 5mph (1 to 2 minutes) faster than scheduled in some of the south-end segments, including the NWK-NYP segment where the other 2-minute travel delay reduction was observed (at NYP). This further supported the timetable-padding idea that provided an opportunity for trains to "catch up".

Train 190		Scheduled		*Delay (min)	Segment Travel Time (min)			Segment Travel Speed (mph)		
Segment	Distance (miles)	Departure Time	Arrival Time		Scheduled	Avg. Actual	Actual-Scheduled	Scheduled	Avg. Actual	Actual-Scheduled
WAS-BWI	30	3:15 AM	3:39 AM	1.9	24	25	1	75	72	-3
BWI-BAL	11	3:39 AM	3:52 AM	2.5	13	12	-1	51	57	6
BAL-ABE	30	3:54 AM	4:19 AM	3.6	25	25	0	72	73	1
ABE-WIL	39	4:19 AM	4:50 AM	4.0	31	29	-2	75	80	5
WIL-PHL	25	4:50 AM	5:10 AM	4.0	20	19	-1	75	80	5
PHL-TRE	33	5:15 AM	5:44 AM	3.5	29	28	-1	68	72	3
TRE-EWR	45	5:44 AM	6:16 AM	4.9	32	32	0	84	84	0
EWR-NWK	3	6:16 AM	6:22 AM	5.5	6	5	-1	30	35	5
NWK-NYP	10	6:22 AM	6:40 AM	5.1	18	16	-2	33	39	5
NYP-STM	36	6:55 AM	7:47 AM	8.8	52	56	4	42	39	-3
STM-NHV	39	7:47 AM	8:35 AM	11.5	48	49	1	49	48	-1
NHV-OSB	33	8:37 AM	9:06 AM	11.9	29	27	-2	68	72	4
OSB-NLC	18	9:06 AM	9:26 AM	12.8	20	20	0	54	54	0
NLC-KIN	35	9:26 AM	9:57 AM	15.0	31	31	0	68	67	-1
KIN-PVD	27	9:57 AM	10:17 AM	16.1	20	19	-1	81	85	4
PVD-RTE	32	10:17 AM	10:48 AM	11.1	31	23	-8	62	82	21
RTE-BBY	10	10:48 AM	10:59 AM	12.0	11	11	0	55	57	2
BBY-BOS	1	10:59 AM	11:05 AM	13.9	6	8	2	10	8	-2

Table 6.4: Train 190 Segment Level Performance

6.4.2 First Southbound Regional Train From NYP - Train 151 (4:40AM)

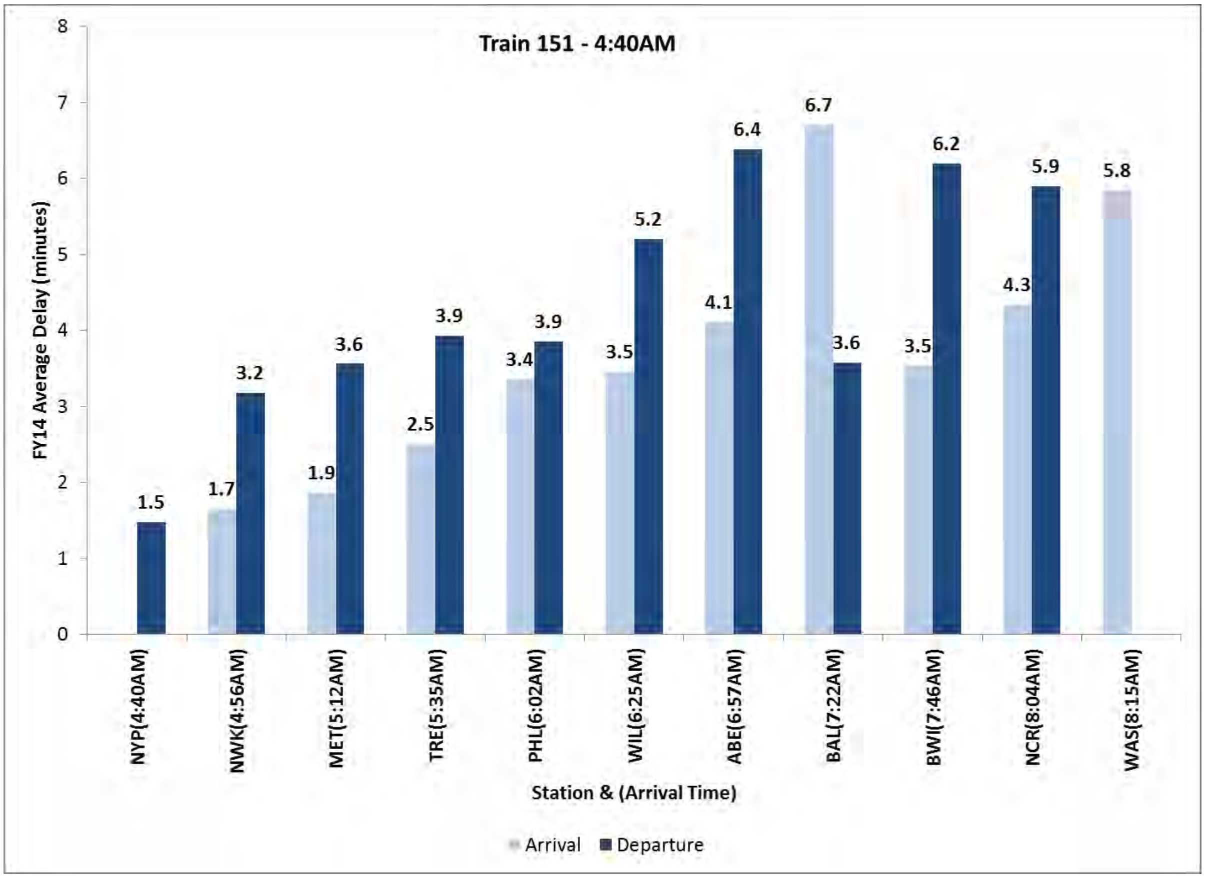


Figure 6.15: 2014 FY 2014 Average Station Delays for Train 151

Train 151 is the first southbound Regional train scheduled to depart from NYP at 4:40AM and arrive at WAS at 8:15AM. There are no Amtrak trains scheduled to operate ahead of this train (The overnight Regional train, Train 67 is scheduled to arrive NYP at 2:15AM and depart NYP at 3:00AM, an hour and 40 minutes before Train 151). Amtrak's scheduled dwell time for Train 151 is 10 minutes at BAL. Figure 6.15 shows the average FY 2014 arrival and departure delays for Train 151. In FY 2014, two of every five trains arrived the end-point terminal later than the scheduled time. On average the actual departure time for the train was almost two minutes later than scheduled. In FY 2014, the delays on Train 151 grew incrementally between each segment along the corridor, peaking around BAL, where it also recovered from some of the upstream delays, likely an advantage of the 10-minute scheduled train dwell time. Nonetheless, the train typically accumulated additional delays south of BAL, and terminated at Washington Union Station (WAS) with an average delay of about 6 minutes.

The Regional trains travelled about 8 mph (2 minutes) faster than scheduled in the BWI-NCR segment as shown in Table 6.5. Like the Acela trains, on average it also travelled faster than scheduled in the NWK-MET segment as well as the MET-TRE segment.

Train 151	Scheduled			*Delay (min)	Segment Travel Time (min)			Segment Travel Speed (mph)		
Segment	Distance (miles)	Departure Time	Arrival Time		Scheduled	Avg. Actual	Actual-Scheduled	Scheduled	Avg. Actual	Actual-Scheduled
NYP-NWK	10	4:40 AM	4:56	1.7	16	16	0	37	37	0
NWK-MET	14	4:56 AM	5:12	1.9	16	15	-1	53	57	5
MET-TRE	34	5:12 AM	5:35	2.5	23	22	-1	89	93	4
TRE-PHL	33	5:35 AM	6:02	3.4	27	26	-1	73	75	2
PHL-WIL	25	6:05 AM	6:25	3.5	20	20	0	75	77	2
WIL-ABE	39	6:25 AM	6:57	4.1	32	31	-1	73	76	3
ABE-BAL	30	6:57 AM	7:22	6.7	25	25	0	72	71	-1
BAL-BWI	11	7:32 AM	7:46	3.5	14	14	0	47	47	0
BWI-NCR	21	7:46 AM	8:04	4.3	18	16	-2	70	78	8
NCR-WAS	9	8:04 AM	8:15	5.8	11	11	0	49	49	0

Table 6.5: Train 151 Segment Level Performance

6.4.3 First Southbound Regional Train From BOS - Train 95 (6:10AM)

Train 95 is the first southbound Regional through-train departure from BOS scheduled to depart from BOS at 6:10AM and arrive at WAS at 2:00PM. Amtrak's scheduled dwell times for the train are – 2 minutes in NHV, 15 minutes in NYP, 3 minutes in PHL, and 8 minutes in BAL. In FY 2014, Train 95 had an on-time performance of 65% and an average delay of 28 minutes.

Figure 6.16 shows that the train usually departed from BOS on average almost a minute later than scheduled, and usually arrived at WAS on average about 28 minutes later than the scheduled time. Overall, the dwell times did not seem to provide enough buffer for the train to recover from upstream delays. The upstream delays appeared to propagate and heighten along the length of the corridor. The delays are typically almost as high as 15 minutes in the first half of the trip, north of NYP, and almost doubles in the south end segments, south of NYP.

Table 6.6 shows the segment level performance for Train 95, which shows that on average, the train travelled about 7minutes faster than scheduled in the OSB-NHV segment, and about 4 minutes faster in the BWI-NCR segment. The Acela trains also appeared to travel faster than schedule in the BWI-WAS segment.

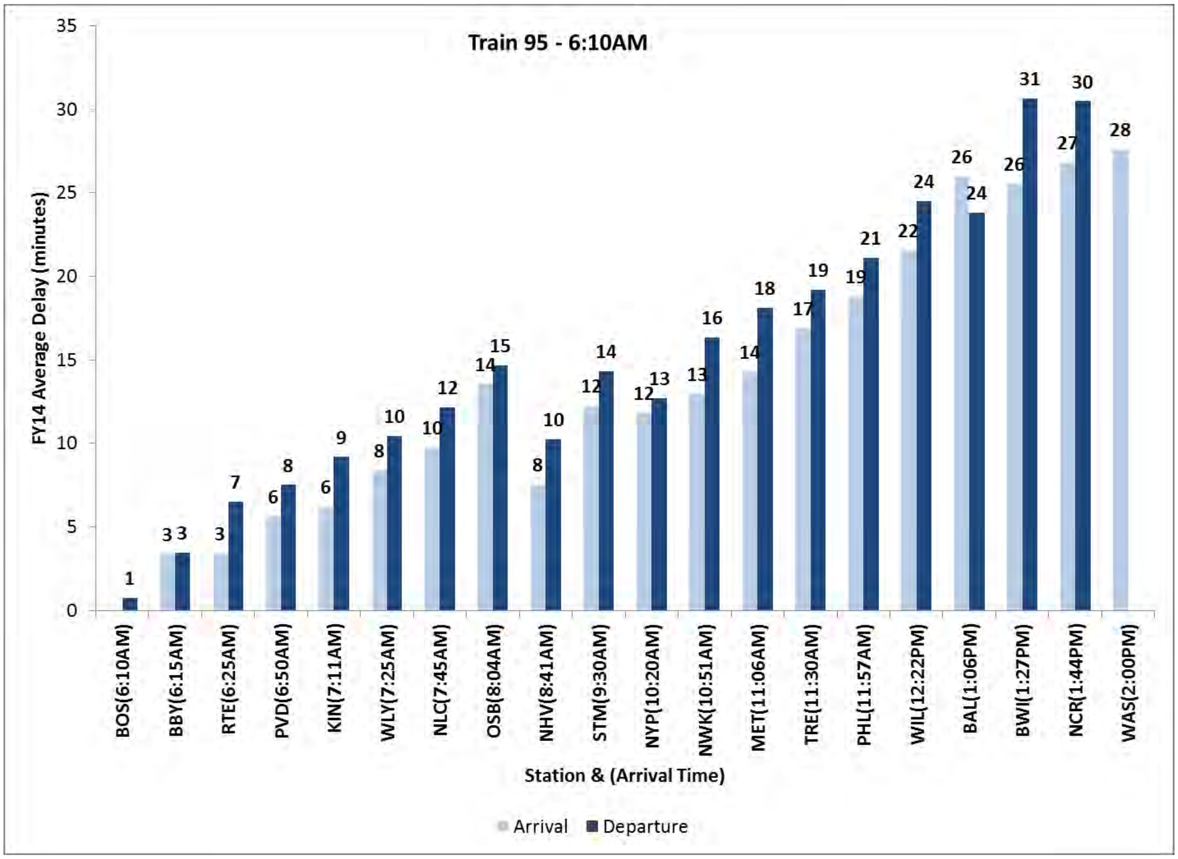


Figure 6.16: FY 2014 Average Station Delays for Train 95

Train 95		Scheduled		*Delay (min)	Segment Travel Time (min)			Segment Travel Speed (mph)		
Segment	Distance (miles)	Departure Time	Arrival Time		Scheduled	Avg. Actual	Actual-Scheduled	Scheduled	Avg. Actual	Actual-Scheduled
BOS-BBY	1	6:10 AM	6:15 AM	0.0	5	8	3	12	8	-4
BBY-RTE	10	6:15 AM	6:25 AM	3.4	10	10	0	60	60	0
RTE-PVD	32	6:25 AM	6:50 AM	3.4	25	24	-1	77	80	3
PVD-KIN	27	6:50 AM	7:11 AM	5.6	21	20	-1	77	82	5
KIN-WLY	17	7:11 AM	7:25 AM	6.2	14	13	-1	73	77	4
WLY-NLC	18	7:25 AM	7:45 AM	8.4	20	19	-1	54	56	2
NLC-OSB	18	7:45 AM	8:04 AM	9.7	19	20	1	57	53	-4
OSB-NHV	33	8:04 AM	8:41 AM	13.6	37	30	-7	54	66	13
NHV-STM	39	8:43 AM	9:30 AM	7.6	47	49	2	50	48	-2
STM-NYP	36	9:30 AM	10:20 AM	12.3	50	48	-2	43	45	2
NYP-NWK	10	10:35 AM	10:51 AM	11.8	16	16	0	38	37	-1
NWK-MET	14	10:51 AM	11:06 AM	13.0	15	13	-2	56	65	9
MET-TRE	34	11:06 AM	11:30 AM	14.3	24	23	-1	85	90	5
TRE-PHL	33	11:30 AM	11:57 AM	16.9	27	27	0	73	74	1
PHL-WIL	25	12:00 PM	12:22 PM	18.8	22	22	0	68	67	-1
WIL-BAL	69	12:22 PM	1:06 PM	21.6	44	45	1	94	91	-3
BAL-BWI	11	1:14 PM	1:27 PM	26.0	13	15	2	51	45	-6
BWI-NCR	21	1:27 PM	1:44 PM	25.6	17	13	-4	74	96	22
NCR-WAS	9	1:44 PM	2:00 PM	26.8	16	13	-3	34	41	7

Table 6.6: Train 95 Segment Level Performance

In summary, the first train of the day routinely encountered considerable amounts of delays because:

i. The trains departed from the originating station late, and upstream delays propagated downstream and in addition, accumulated along segments in the corridor, as well as at each station on the corridor

ii. Amtrak does not schedule dwell time at many of the stations, thus delays add up between arrival at and departure from each station in the corridor.

iii. Overall, the trains usually ran behind schedule though on occasion, some trains were able to recover from some of the built-up upstream delays by traveling faster than scheduled in certain segments.

   a. Northbound: on occasion, the trains traveled on average 1 to 2 minutes faster in some of the south-end segments, including the NWK-NYP segment, and on average 8 minutes faster than scheduled in the PVD-RTE segment

   b. Southbound: on occasion, the trains traveled on average 2 minutes faster than scheduled in the NWK-MET segment and MET-TRE segments, as well as on average 5 to 7 minutes in the OSB-NHV segment and the BWI-WAS segment.

iv. Altogether, late departures, terminal time not accounted for at stations, and poorly estimated segment train speeds and travel times led to significant en-route delays.

6.5 Fluctuations Due to Weather, Accidents and Other Incidents

In the Northeast Corridor, the performance of Regional service is occasionally affected by random variations including delays caused by incidents/accidents, engine failure, injuries, signal failures, broken rails, construction, crew-related, or weather-related issues. This section investigates supply fluctuations due to these random, generally one-time disruptions.

Under PRIIA Section 207, Amtrak was required to report the total delay minutes apportioned into Amtrak-responsible, Host-responsible (Metro-North Railroad-responsible), and Third Party-responsible category (see Section 3.3.2 for additional details of PRIIA requirements regarding delays). However the author of this thesis did not have access to this data. Consequently, the inferences made in this section are based on observing the total daily delay minutes on the Regional services. The total daily delay minutes were calculated by aggregating end-point delays on all the Amtrak train that were operated each day. Following that, the author went through a manual process of searching through Amtrak's Twitter[viii] and Breaking News[ix] feeds for reports of major accidents and incidents on days with significant delays.

6.5.1 Total Minutes of Delay

FY 2012 Daily Delays

Figure 6.17 shows the total daily end-point delays in minutes experienced on Regional trains in FY 2012. There were seven days of the year on which Regional experienced severe delays greater than 2,000 minutes. 2,000 minutes was chosen as an arbitrary boundary line separating 'regular' delays from severe delays. The Twitter and Breaking News feeds indicated Severe Storms as the main cause of the delays in June. In addition, the 2661 minutes of delay on 9/18/2012 were attributed to service disruptions from Hurricane Sandy. The author was not able to identify the cause of the 1,346 minutes of delay on 6/1/2012. Overall, the total daily delays on Regional trains in FY 2012 were consistently around 400 minutes. To comprehend this value, divide 400 minutes by 64 trains on a given Monday, for example, which is equivalent to Regional trains being on average 6 minute delayed in arriving at their final stations. From this proxy analysis, overall the major delays in FY 2012 appeared to be as a result of unanticipated weather disruptions, and on one occasion as a result of operational issues.

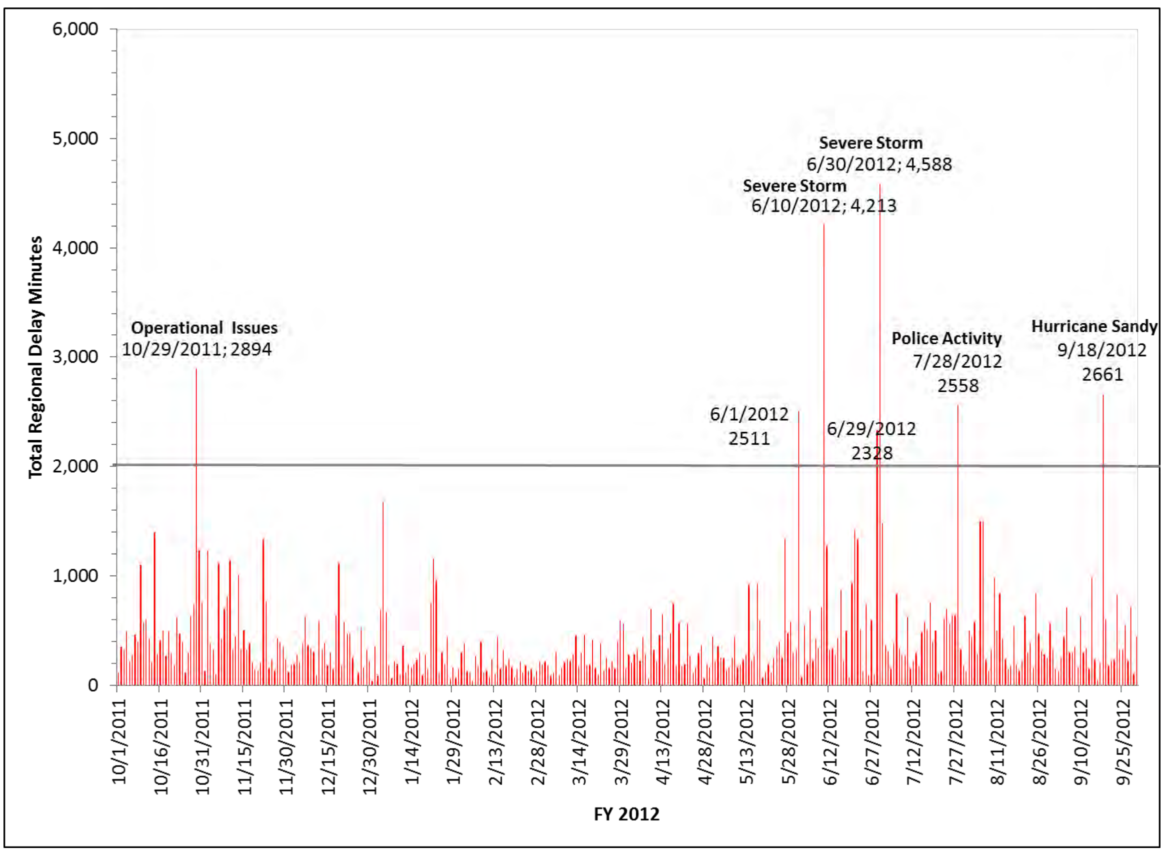


Figure 6.17: FY 2012 Regional Daily Delays

FY 2014 Daily Delays

Figure 6.18 shows the total amount of daily delays in minutes experienced by all Regional trains in FY 2014. The most significant Regional delays were due to the Polar Vortex and sub-zero weather conditions in the Northeast Corridor on multiple days in January 2014. Hurricane Arthur on July 3rd, 2014 also caused major disruption and delays on the Regional service. Due to these severe weather conditions, some trains experienced delays as high as 3hours, and Amtrak's on-time performance metric for Regional trains on some of those days were as low as 8%. For example, 64 of the 65 scheduled Acela trains on July 3rd, 2014, arrived at their destination more than 10 to 20 minutes later than the scheduled arrival time. The other major causes of Regional delay were operational issues, down catenary wires, signal issues and power issues. These operational and infrastructure issues led to 9 of the 16 major delays.

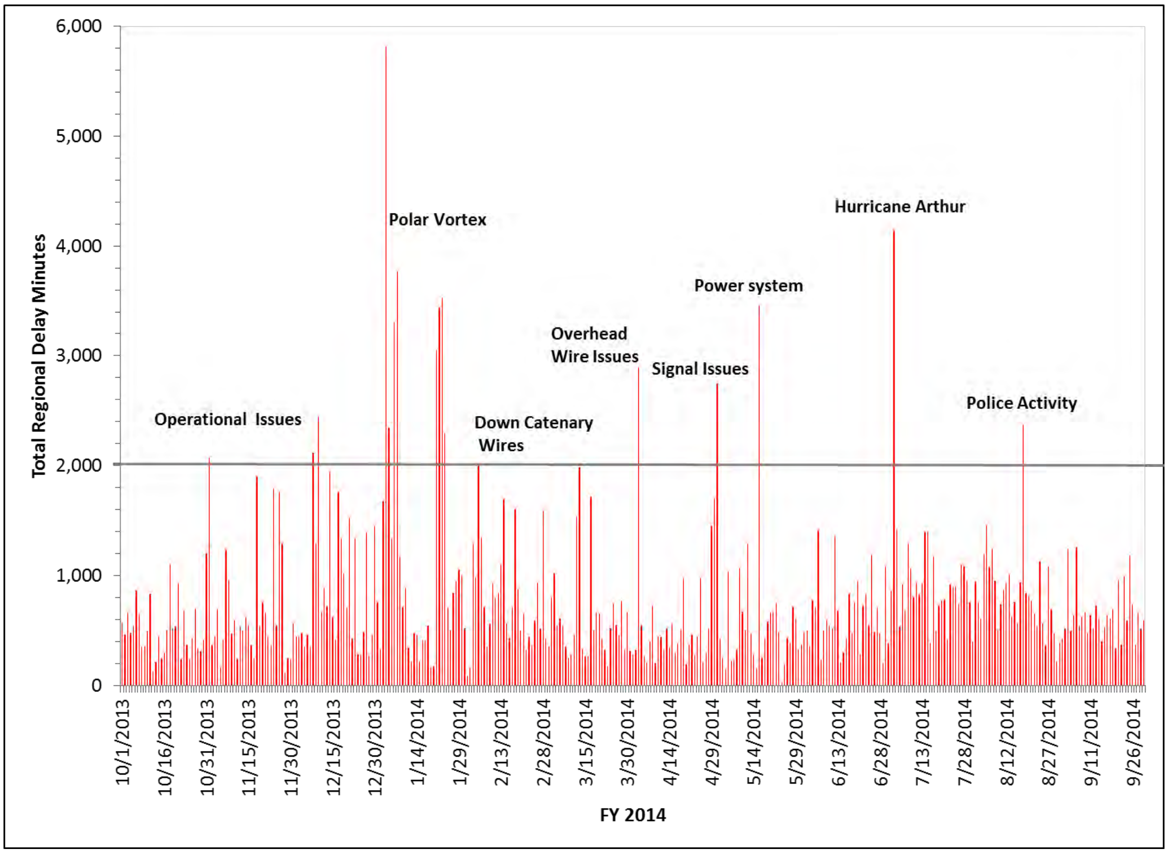


Figure 6.18: FY 2014 Regional Daily Delays

Table 6.7 shows all the days, which experienced total delays greater than 2,000 minutes, including the date, day of week, reason for the delay, and the OTP achieved on that day. Although the total delay was abnormally high on 1/4/2014 and 1/24/2014, the OTP on these days were still greater than 60%, which suggests that the delay was concentrated in specific sections of the corridor and times of the day, and affected a select few of the trains scheduled to operate on that day. Days with high delays and low OTP (e.g. 1/7/14 Polar Vortex) suggest that the delay expanded throughout the corridor and day, and affected majority of the scheduled trains.

Delay Reason	Date	Weekday	Total Delay Minutes	OTP
Operational Issues	11/1/2013	Fri	2,073	43%
Operational Issues	12/8/2013	Sun	2,114	34%
Operational Issues	12/10/2013	Tue	2,452	39%
Polar Vortex	1/3/2014	Fri	5,820	8%
Polar Vortex	1/4/2014	Sat	2,346	18%
Polar Vortex	1/6/2014	Mon	3,307	67%
Polar Vortex	1/7/2014	Tue	3,766	25%
Polar Vortex	1/21/2014	Tue	3,053	36%
Polar Vortex	1/22/2014	Wed	3,437	18%
Polar Vortex	1/23/2014	Thu	3,525	21%
Polar Vortex	1/24/2014	Fri	2,298	31%
Catenary/Power Issues	4/3/2014	Thu	2,894	62%
Signal Issues WAS-BAL	5/1/2014	Thu	2,746	36%
Power system issues NYP-STM; Police activity north of WAS	5/16/2014	Fri	3,456	49%
Hurricane Arthur	7/3/2014	Thu	4,142	50%
Police Activity	8/18/2014	Mon	2,368	54%

Table 6.7: Regional Daily Total Delay Greater Than 2,000 minutes

In summary, the major disruptions on Regional service in both FY 2012 and FY 2014 were caused by weather related factors (severe winter weathers and hurricanes). However, in FY 2014, in addition to the weather related issues, there were significantly more delays attributed to infrastructural issues compared to FY 2012. This suggests the state of infrastructures and equipment in the Northeast Corridor has deteriorated in recent years.

6.5.2 Cancelled Trains

The existing performance metrics do not account for train cancellations, even though they pose huge inconveniences to travelers. In this section, Regional train cancellation in FY 2012 and FY 2014 are presented.

Although Regional service performance in FY 2012 was relatively high (compared to other fiscal years), there were 54 days in the year on which Regional trains were cancelled, and a total of 73 cancelled trains. Figure 6.19 shows that only one train was cancelled on most days, and at most 3 trains cancelations happened once. Altogether, less than 1% of total FY 2012 scheduled Regional trains were cancelled.

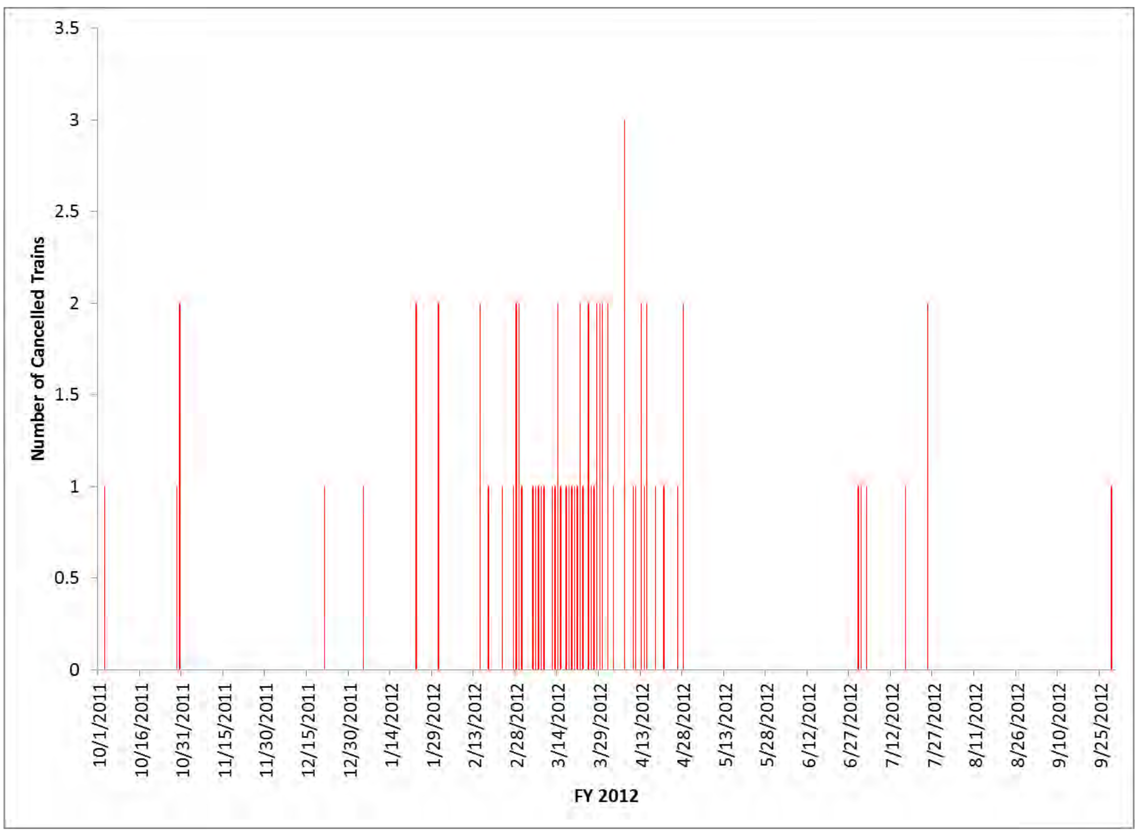


Figure 6.19: FY 2012 Number of Cancelled Regional Trains

Figure 6.20 shows that in FY 2014, there were 48 days in the year on which Regional trains were cancelled, and there were a total of 167 cancelled trains. The major train cancellations occurred in January and February due to the severe winter weather attributed to the Polar Vortex[x]. 51% of FY 2014 Regional train cancellations were attributed to weather-related issues, 6% to down catenary/overhead wires issues, 2% to Police Activity, and 40% to unknown sources. Altogether, 1% of total scheduled Regional trains in FY 2014 were cancelled.

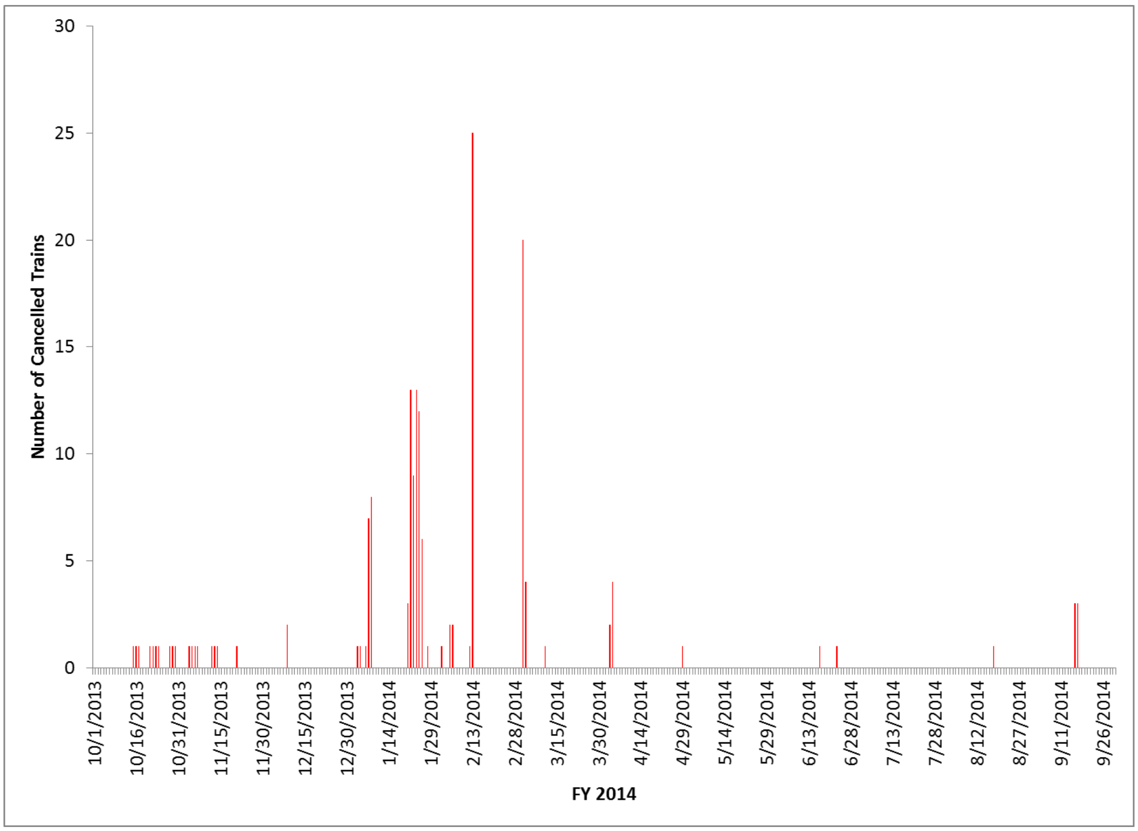


Figure 6.20: FY 2014 Number of Cancelled Regional Trains

In summary, although the train cancellations on Regional are not disturbingly high on an average day, they cause a substantial amount of inconvenience and disutility to travelers and thus must be accounted for in service performance metrics. The need for a more precise cumulative service performance index is discussed in Section 8.4.

6.6 Capacity Analysis

This section analyzes the capacity on Regional trains between FY 2005 and FY 2014. The goal of the capacity analysis for this thesis was to provide a caveat to the analysis on the impact of Amtrak service performance on demand. Demand in the Northeast Corridor is currently alleged to be above the capacity of Amtrak trains, especially Acela trains. However, the capacity analysis was expanded to Regional services to provide a comprehensive and comparative analysis of the Regional service as the Acela in Chapter 5. The section presents actual segment level volumes on each Regional train operated in both directions for some chosen dates with peak ridership in FY 2005, FY 2012, FY 2013 and FY 2014.

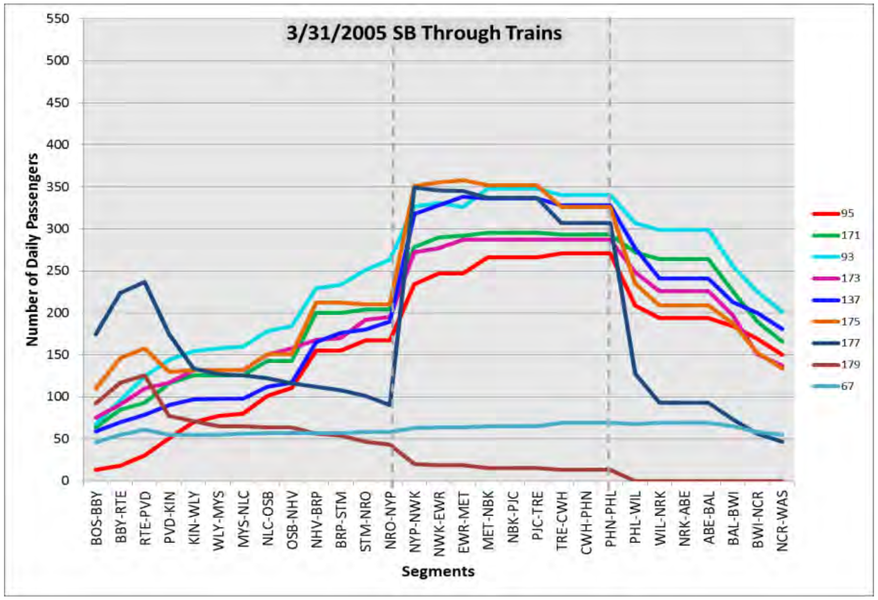
Capacity analysis of the Regional service is more complex than on the Acela service. The Regional system is made up of an assortment of trainsets having 7 – 10 passenger cars and with seating capacity ranging from about 450 to 550. In addition, the number of Regional trains in operation varied by day of week from about 51 trains on Saturdays to 64 trains on Mondays and 67 trains on Fridays. Consequently, unlike the Acela that has a definitive capacity of 304 seats per train each day, the capacity on Regional depends on the day of week and what train equipment were operated on the given day.

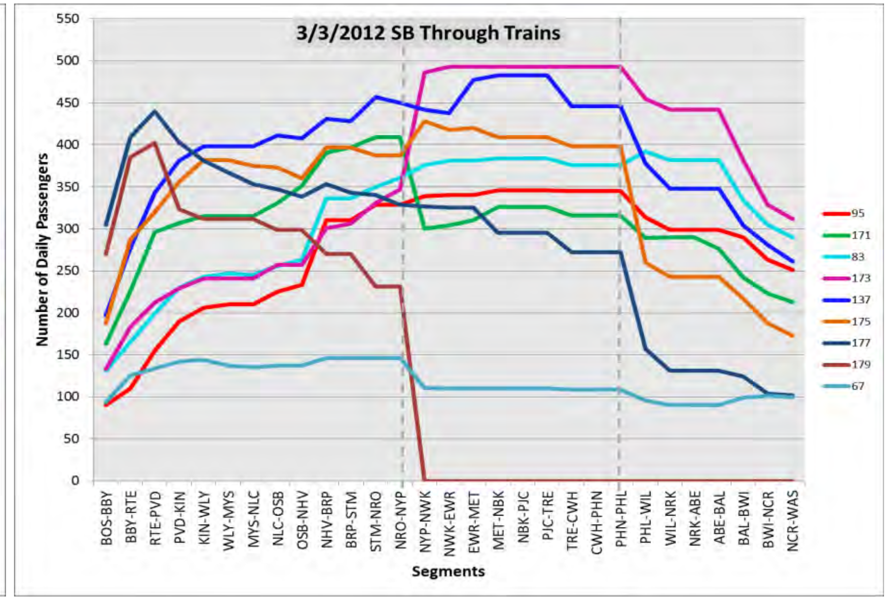
6.6.1 FY 2005 to FY 2014 Southbound Through Regional Trains

The four plots in Figure 6.21 show segment level ridership for each southbound through (BOS-WAS) train that was operated on the specific dates (indicated in the figure) in FY 2005, FY 2012, FY 2013 and FY 2014 with the highest ridership. The legend shows the train numbers in order by departure time from BOS, such that the colors that correspond with each train is the same across the different years. For example, the red line represents Train 95 on all charts.

Train 95 departs at 6:10AM, Train 171 departs at 8:15AM, Train 83 departs at 9:30AM, Train 173 departs at 11:15AM, Train 137 departs at 1:40PM, Train 175 departs at 3:20PM, Train 177 departs at 5:35PM, Train 179 departs at 6:45PM, and Train 67 departs at 9:30PM. Because these are through trains, although the departure time from the originating station was in an off-peak time of day, the arrival times at other stations en-route could be during the peak periods. For example, Train 173 and Train 137 departs from NYP at 3:35PM and 6:25PM, respectively and arrives in the south end of the corridor during the PM peak time of day.

In FY 2005, the maximum passenger loading was around 350 riders in the south section of the corridor between NYP and WAS. However the passenger loading in the south section increased to almost 500 in FY 2012, but dropped to about 400 in FY 2013, and further to under 400 by FY 2014. The ridership decline in the south section (NYP-WAS) on Regional trains between FY 2012 and FY 2014 is opposite of the growth observed on the Acela trains. However, similar to the Acela trains, the load factor in the north section of the corridor between BOS and WAS grew steadily between FY 2005 and FY 2014. The FY 2005 chart (top left) shows low ridership between BOS and NYP, and a spike in ridership at NYP, suggesting low ridership in the north section and high ridership in the south section of the corridor. In comparison, by FY 2014 (bottom right) the ridership in both the north section (BOS-NYP) and south section (NYP-WAS) of the corridor were high. Overall, the through trains are now more frequently near-fully loaded along the entire length of the corridor. Assuming a through train capacity of 500 seats, the load factor of the most loaded train in FY 2014 was about 80%.


3/31/2005 SB Through Trains
Number of Daily Passengers
Segments
95
171
93
173
137
175
177
179
67


3/3/2012 SB Through Trains
Number of Daily Passengers
Segments
95
171
83
173
137
175
177
179
67

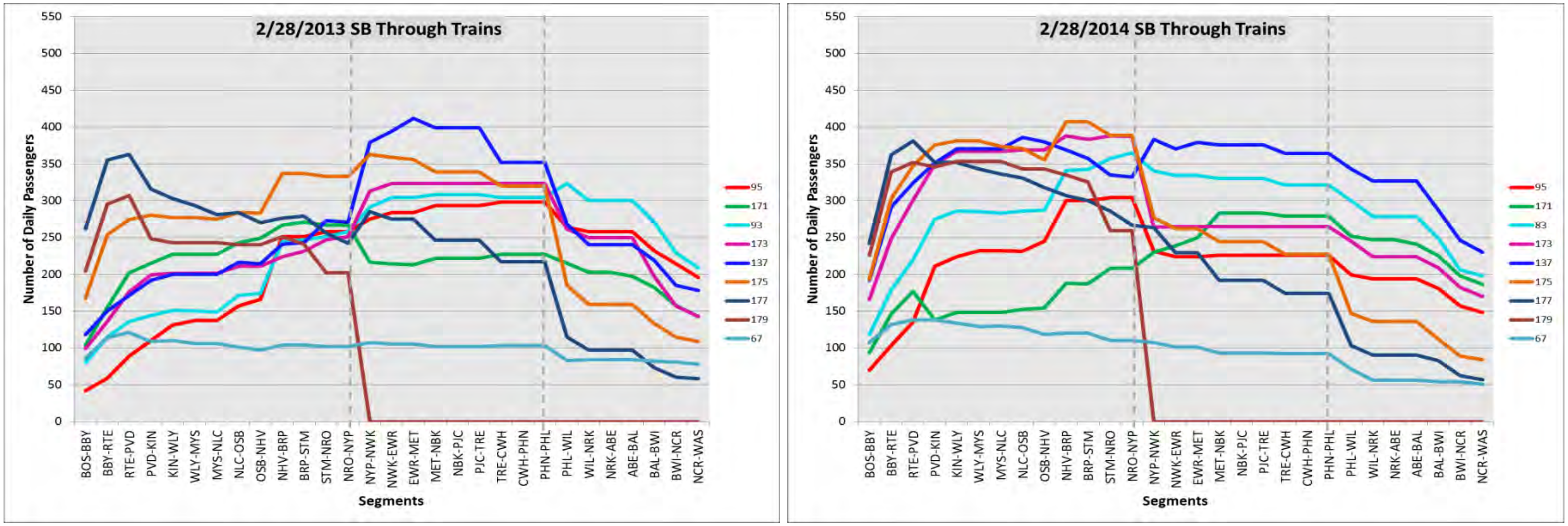


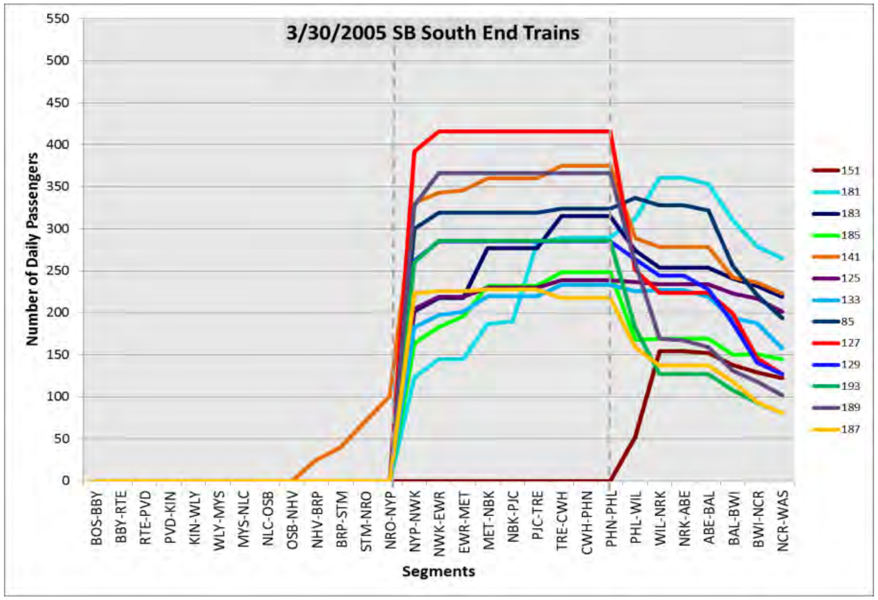
Figure 6.21: FY 05 - FY 14 SB Segment Level Through Train Capacity

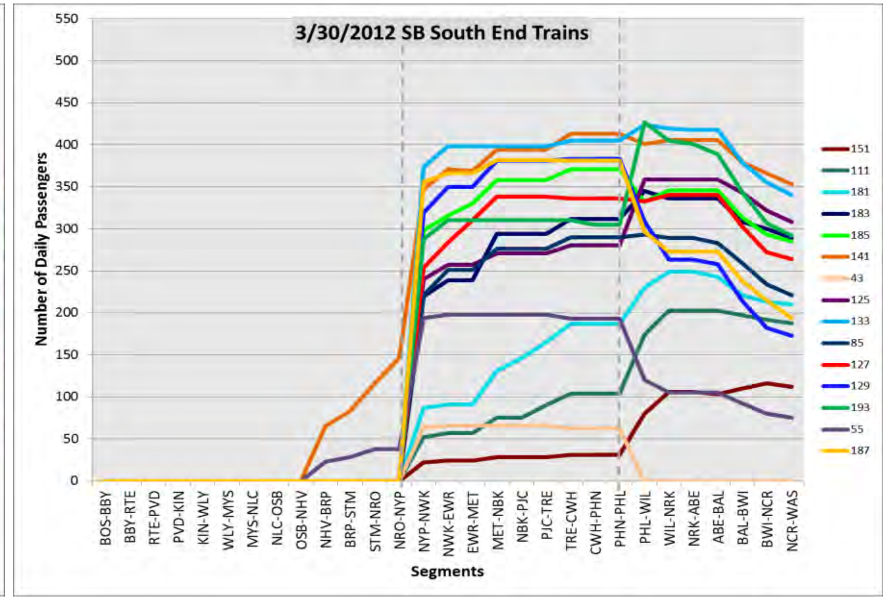
6.6.2 FY 2005 to FY 2014 Southbound South-End only Regional Trains

The four figured in Figure 6.22 show segment level demand for each southbound south-end only Regional service that was operated on the specific dates in FY 2005, FY 2012, FY 2013 and FY 2014 with the highest ridership. The southbound south end only Regional services originate in NYP and terminate in WAS. The legend shows the train numbers in order by departure time from NYP such that the colors that correspond with each train is the same across the different years. The additional Regional-Keystone trains that operate in the south end between NYP and PHL are not included in this analysis because they operate well below capacity.

Train 151, Train 181, Train 183, Train 185, Train 141, Train 125 and Train 133 are AM trains while Train 85, Train 127, Train 129, Train 193, Train 189 and Train 187 are PM trains. From the figure, in FY 2013 and FY 2014, the trains with the highest ridership volumes were Train 127 (in red), which departed from NYP at 4:05PM and Train 129 (in blue), which departed from NYP at 4:42PM.

Similar to the through trains, the number of Regional riders on the south end trains declined between FY 2005 and FY 2014. The maximum passenger loading was around 430 in FY 2005, and remained at that level in FY 2012 but dropped to about 400 in FY 2013 and further down to under 350 passengers by FY 2014. Again, the ridership decline on Regional trains between FY 2005 and FY 2014 in the load factor in the south section (NYP-WAS) is opposite of the growth observed on the Acela trains.


3/30/2005 SB South End Trains
Number of Daily Passengers
Segments
151
181
183
185
141
125
133
85
127
129
193
189
187
BOS-BBY
BBY-RTE
RTE-PVD
PVD-KIN
KIN-WLY
WLY-MYS
MYS-NLC
NLC-OSB
OSB-NHV
NHV-BRP
BRP-STM
STM-NRO
NRO-NYP
NYP-NWK
NWK-EWR
EWR-MET
MET-NBK
NBK-PJC
PJC-TRE
TRE-CWH
CWH-PHN
PHN-PHL
PHL-WIL
WIL-NRK
NRK-ABE
ABE-BAL
BAL-BWI
BWI-NCR
NCR-WAS


3/30/2012 SB South End Trains
Number of Daily Passengers
Segments
151
111
181
183
185
141
43
125
133
85
127
129
193
55
187
BOS-BBY
BBY-RTE
RTE-PVD
PVD-KIN
KIN-WLY
WLY-MYS
MYS-NLC
NLC-OSB
OSB-NHV
NHV-BRP
BRP-STM
STM-NRO
NRO-NYP
NYP-NWK
NWK-EWR
EWR-MET
MET-NBK
NBK-PJC
PJC-TRE
TRE-CWH
CWH-PHN
PHN-PHL
PHL-WIL
WIL-NRK
NRK-ABE
ABE-BAL
BAL-BWI
BWI-NCR
NCR-WAS

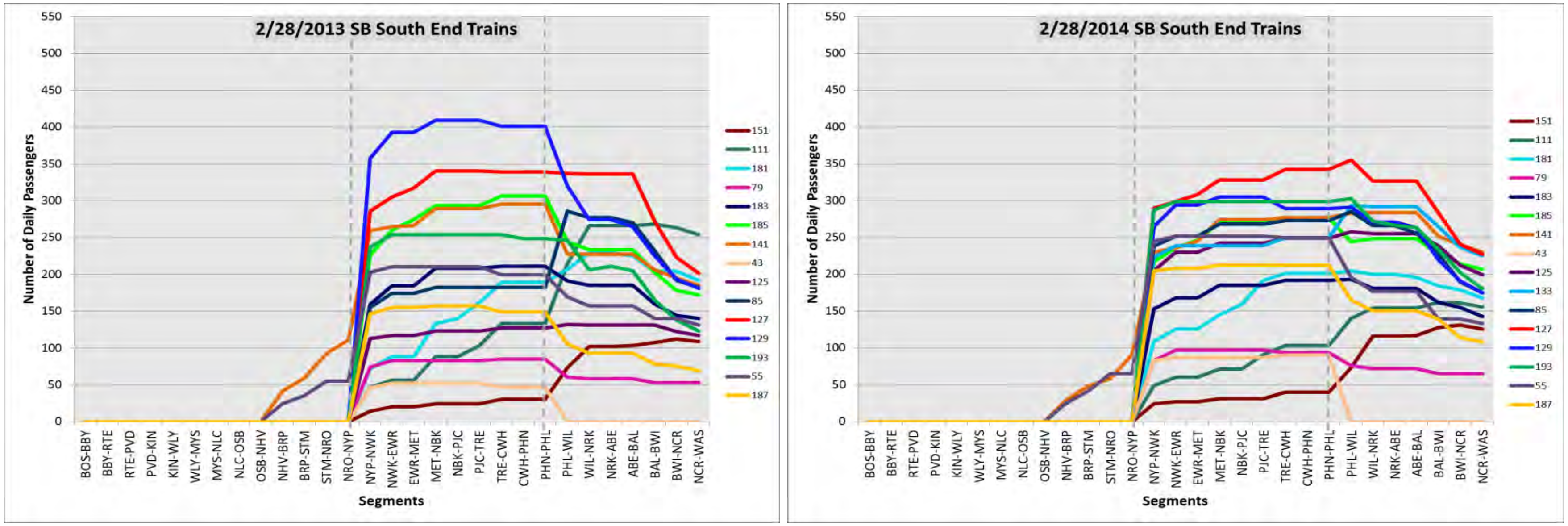


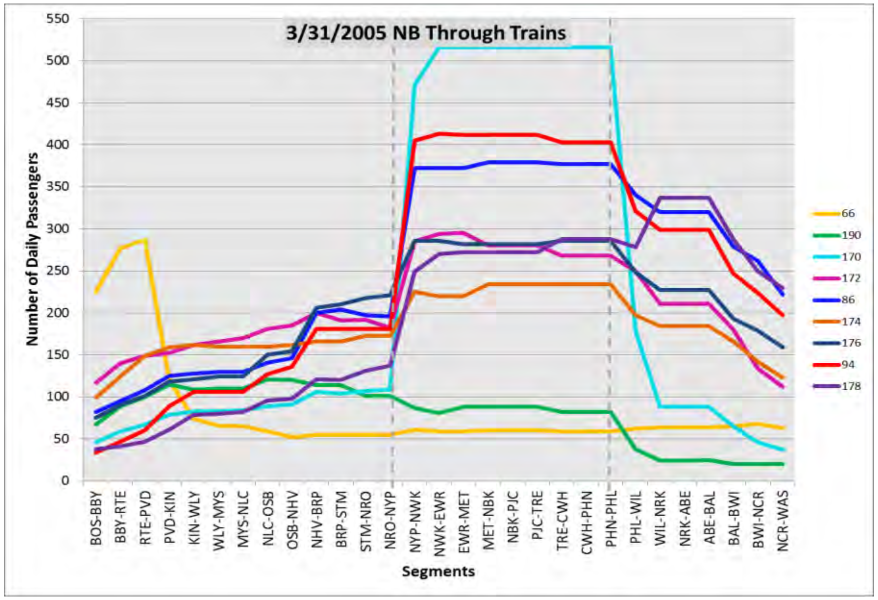
Figure 6.22: FY 05 - FY 14 SB Segment Level South End Train Capacity

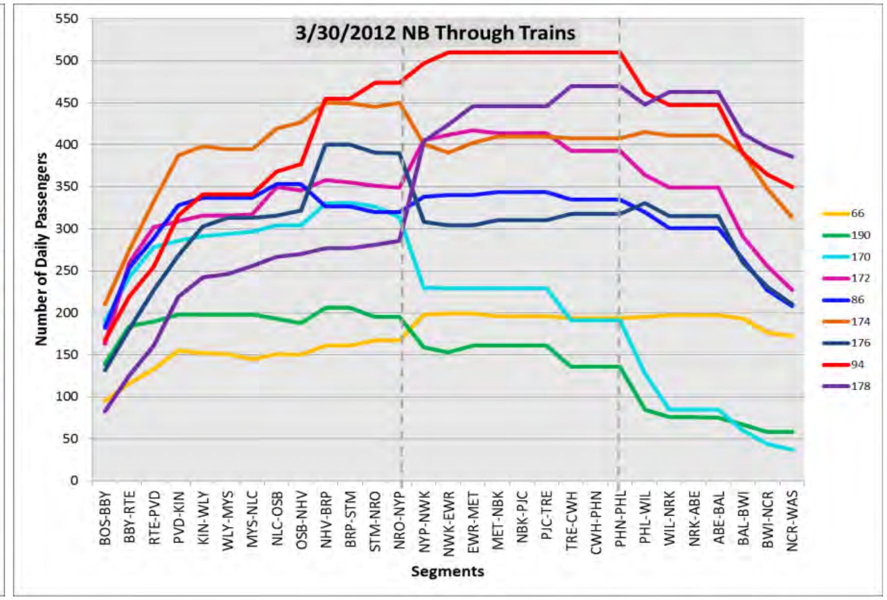
6.6.3 FY 2005 to FY 2014 Northbound Through Regional Trains

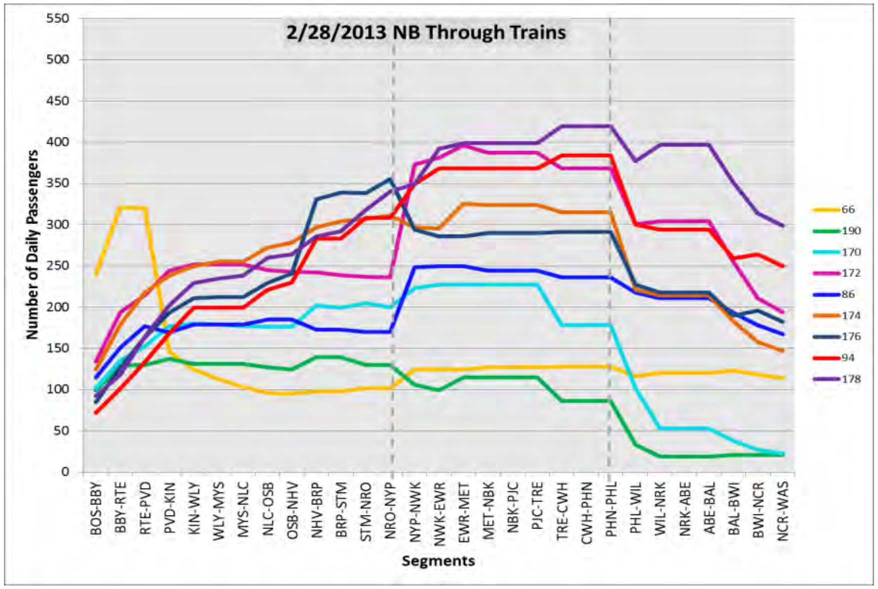
The four plots in Figure 6.23 show segment level demand for each northbound through Regional service operated on the specific dates (indicated in the charts) in FY 2005, FY 2012, FY 2013 and FY 2014 with the highest ridership. The northbound through Regional services originate in WAS and terminate in BOS. The legend shows the train numbers in order by departure time from WAS so that the colors corresponding with each train is the same across the different years. For example, Train 66 is shown in yellow in all the charts.

The Regional northbound through trains exhibit similar patterns as the southbound trains. The late night/early morning trains departing from WAS at 10:10PM (Train 66), 3:15AM (Train 190) and 4:52Pm (Train 170) show the lowest segment loading of about 230 riders each year, except in 2005 where Train 170 had a maximum passenger load of about 520 riders in the south section. Train 174 departing WAS at 10:20AM, Train 176 departing WAS at 12:25PM, Train 94 departing WAS 2:02PM and Train 178 departing WAS at 4:02PM have the highest segment loading ranging from 400 to 450 passengers in the south section in 2014.

Again comparing FY 2005 to FY 2014, Regional trains exhibit decline in segment ridership on each peak train.


3/31/2005 NB Through Trains
Number of Daily Passengers
Segments
66
190
170
172
86
174
176
94
178
BOS-BBY
BBY-RTE
RTE-PVD
PVD-KIN
KIN-WLY
WLY-MYS
MYS-NLC
NLC-OSB
OSB-NHV
NHV-BRP
BRP-STM
STM-NRO
NRO-NYP
NYP-NWK
NWK-EWR
EWR-MET
MET-NBK
NBK-PJC
PJC-TRE
TRE-CWH
CWH-PHN
PHN-PHL
PHL-WIL
WIL-NRK
NRK-ABE
ABE-BAL
BAL-BWI
BWI-NCR
NCR-WAS


3/30/2012 NB Through Trains
Number of Daily Passengers
Segments
66
190
170
172
86
174
176
94
178



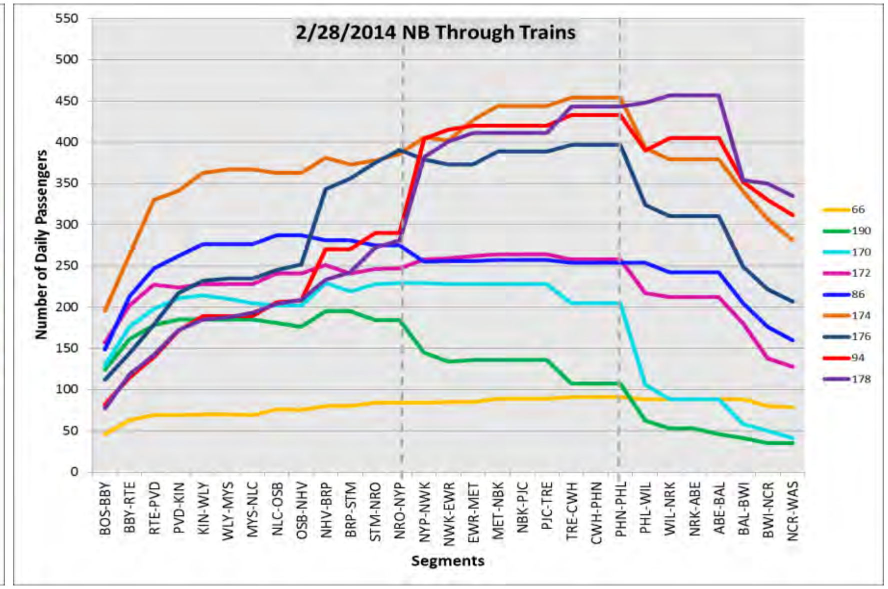


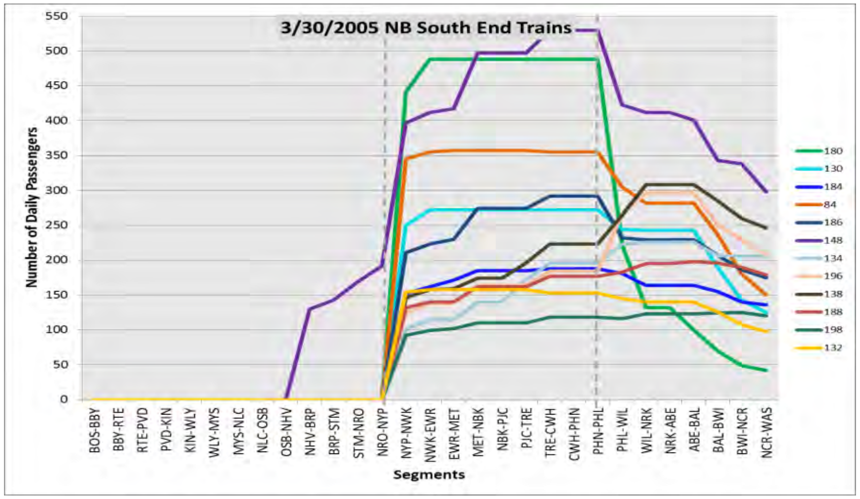
Figure 6.23: FY 05 - FY NB Segment Level Through Train Capacity

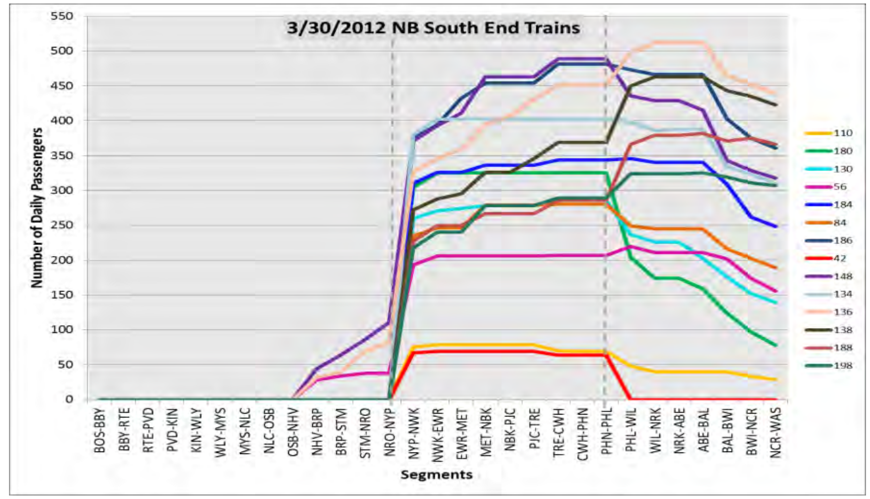
6.6.4 FY 2005 to FY 2014 Northbound South-end Only Regional Trains

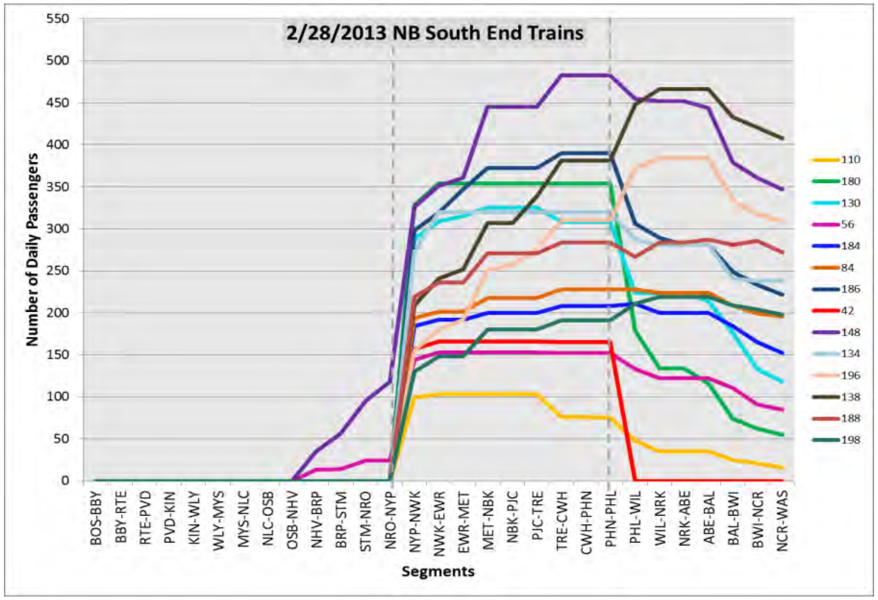
The four plots in Figure 6.24 show segment level demand for each northbound south-end only Regional service operated on the specific dates (indicated in the charts) in FY 2005, FY 2012, FY 2013 and FY 2014 with the highest ridership. The northbound south-end only Regional services originate in WAS and terminates in NYP. The legend shows the train numbers in order by departure time from WAS so that the colors corresponding with each train is the same across the different years.

The northbound south-end only trains appear to have the highest segment load factor. The peak loading ranged between 450 and 550 passengers between BWI and PJC (Princeton Junction). Unlike all the other Regional trains, the northbound south-end only trains show segment-loading growth on the peak trains between FY 2005 and FY 2014

Overall, in summary, except the northbound south-end only Regional trains, the Regional service do not appear to be currently capacity constrained.


3/30/2005 NB South End Trains
Number of Daily Passengers
Segments
180
130
184
84
186
148
134
196
138
188
198
132


3/30/2012 NB South End Trains
Number of Daily Passengers
Segments
110
180
130
56
184
84
186
42
148
134
136
138
188
198



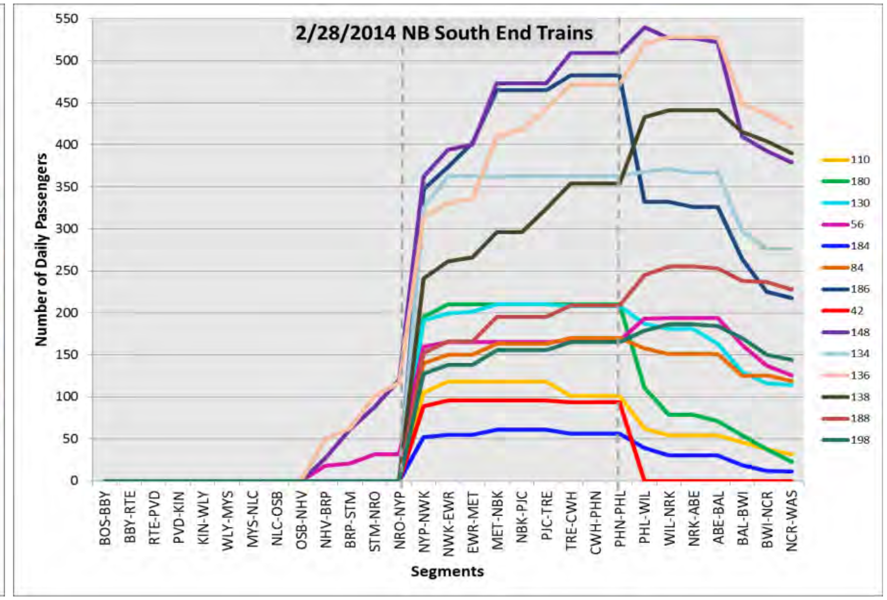


Figure 6.24: FY 05 - FY 14 SB Segment Level South End Train Capacity

6.7 Demand Response to Regional Service Performance

The main question examined in this section is: Do Amtrak Regional passengers modify their future travel choices in response to past Regional performance that either they experienced or were informed about? As a practical constraint, there needs to be a period of time between when travelers experience or learn about performance information and when they make future travel decisions. This time is usually referred to as a lag period. Following this concept, a simple analysis was conducted comparing total annual ridership each year with Regional train performance from the prior year. This assumes a lag of one year. In other words, the goal of the analysis was to test the assumption that a relationship exists between ridership in a given year and OTP or total delay from the prior year

6.7.1 Annual Ridership to Annual OTP

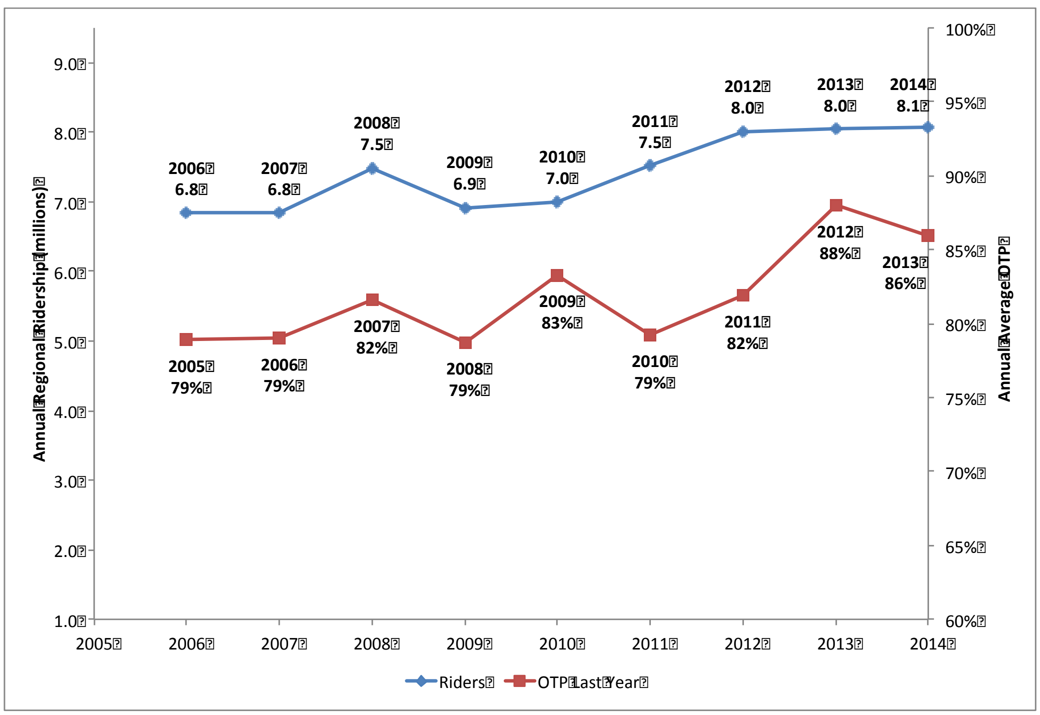


Figure 6.25: Lag Annual Ridership to OTP

Figure 6.25 shows the total annual Regional ridership (blue) between FY 2005 and FY 2014 and the annual average OTP (red) from the prior year. For example, the first data points show total FY 2006 ridership (6.8 million) and FY 2005 average OTP (79%). The relationship between OTP in two consecutive years is typically associated with a similar relationship between the ridership in the following years. For example, an upward trend in average OTP between FY 2005 and FY 2006 is followed by a similar upward trend between the annual ridership in FY 2006 and FY 2007. This correlated relationship is observed between FY 2005 and FY 2010. However, in FY 2010, even though the performance deteriorated between FY 2009 and FY 2010, Regional ridership continued to grow between FY 2010 and FY 2011. The relationship continued again between FY 2011 and FY 2013 but broke down in FY 2014 with ridership increasing despite performance deteriorations in FY 2013. As discussed in Section 6.5, two of the major causes of poor performance in FY 2014 were as a result of the Polar Vortex and Hurricane Arthur, both of which had a worse impact on airline services, perhaps explaining why the ridership grew in FY 2014 despite poor performance in the prior year. The correlation coefficient between the annual ridership and lagged annual on-time performance on the Regional service between FY 2005 and FY 2014 is 0.74. Although this correlation is not perfect (equal to 1), it is sufficiently high to propose that the ridership in a given year is associated with service performance from the prior year.

6.7.2 Annual Ridership to Annual Delay

Figure 6.26 shows the total annual Regional ridership (blue) and the total annual delay (red) from the prior year. The annual ridership and annual delay exhibit a similar lagged relationship as the annual ridership and annual average OTP. Improvements in performance (that is reduction in total annual delay) appear to be associated with ridership increase the following year. However, unlike the high correlation coefficient between annual ridership and lagged annual OTP, the correlation between annual ridership and lagged annual total delay between FY 2005 and FY 2014 is 0.2. Although, as expected the correlation between FY 2005 and FY 2010 is much higher 0.46, it is still comparatively low.

Figure 6.26: Lag Annual Ridership to Delay

It is unclear whether the one-year lagged correlation exhibited between performance and ridership is a cause-effect relationship or simply a correlation caused by other external factors. The Acela service also exhibited a similar lag of one year, which suggests that the correlations are worth looking into.

6.8 Regional Summary

This section summarizes the analyses in Chapter 6.

In FY 2014, 42% of Regional trains arrived on-time, 29% arrived late but within 10 to 20 minutes, and the remaining 29% experienced delays greater than 10 to 20 minutes. Altogether, similarly to the Acela service, 71% of trains arrived within 10 minutes of the scheduled time. However, compared to Acela (11%), more Regional trains (13%) arrived with a delay greater than 30 minutes.

On-time performance and delay minutes are 70% correlated, and both are useful metrics in quantifying performance. However, neither of them includes the effect of cancelled trains, which are important to quantify from a passengers point of view.

In terms of seasonality, excluding January and February, the monthly ridership for Regional appeared to be roughly the same, suggesting that Regional passengers might be regular riders, who have similar travel patterns throughout the year. Additionally, the ridership pattern on the Regional service was unlike that of the Acela service, which exhibited clear seasonal variations. Similarly to Acela performance, Regional performance exhibited seasonal variations with summer months usually being the worst due to heat restrictions and infrastructural issues (catenary wire drooping).

Regarding day of week performance, Regional ridership volumes were lowest on Tuesdays and Saturdays and highest on Fridays and Sundays. In terms of performance, overall, performance appeared to be roughly the same on all days of the week, however, it was usually slightly better on Tuesdays, Saturdays and Sundays, and slightly worse on Fridays. The weekend improvements were likely because fewer Amtrak and Commuter services operate on weekends.

In terms of time of day performance variations, AM and PM peak trains are usually more prone to delays than trains during other off-peak times of the day.

The first train analysis revealed that a major portion of Regional delays appeared to be attributable to late departures from originating station, which accumulate and propagate at each consecutive station downstream. In addition, unscheduled terminal time at stations and poorly estimated segment train speeds and travel times lead to additional en-route delays. On an average day, delays attributed to interference from other Amtrak trains are limited because of the scheduled spacing between trains throughout the day.

The performance of Regional service is occasionally affected by unanticipated changes in travel conditions such as bad/extreme weather, crashes and incidents, equipment failure or other unexpected infrastructure malfunctions. Unanticipated weather-related and third-party events are responsible for about 12% of Amtrak train delays (based on FY 2012 and FY 2014 analysis). Furthermore, in FY 2012, there was only one major delay caused by operational issues, however, by FY 2014, 12% of delays were associated with equipment and infrastructural issues. Lastly, not including days with severe weather issues, Amtrak rarely cancels Regional trains.

The capacity analysis revealed that Regional trains are not at capacity but some peak period trains operate close to load factors of about 80%.

Finally, Regional ridership appears to lag service performance by a year. However, it is unclear whether the one-year lagged associations exhibited between OTP or total delay and ridership are cause-effect relationships between service performance and demand or just correlations.

<This page has been intentionally left blank>

7 Time Series Analysis

In Chapter 5 and 6, the relationship between demand and service performance was investigated. Amtrak ridership appeared to be correlated with service performance from the prior year. However, it was unclear whether the one-year lagged associations exhibited in the FY 2005 to FY 2014 dataset between OTP or total delay and ridership were cause-effect relationships between service performance and demand. The correlation between Amtrak ridership and lagged service performance of Amtrak Acela services will be investigated further in this chapter.

The collection of Amtrak ridership and service performance data between FY 2005 to FY 2014 can be considered as a time series dataset because it contains an ordered sequence of observations at equally spaced time intervals. Moreover, time series modeling provides techniques to abstract the underlying structures of a dataset, as well as meaningful relationships, characteristics and statistics of data attributes. In the most disaggregate form, the demand and train operation data over the ten-year period can be modeled as time progressions at the minute or hourly level. On one hand, the disaggregate level could highlight fine details in Amtrak ridership and performance, but on the other hand, the detail might be unnecessary and conclusions about the system at such a disaggregate level would likely be noisy and highly volatile. As such, there are tradeoffs between the level of disaggregation chosen for a time series model and the stability of the model. Keeping the tradeoffs in mind, for the purpose of this research, the times series observations of Amtrak FY 2005 to FY 2014 ridership and service performance would be evaluated at the daily, weekly, monthly or annual levels, until an informative and stable model is obtained.

The objective of the time series analysis is to explore the characteristics of the datasets, in order to account for the underlying structure (autocorrelation, trends and seasonality) in the data points over time. The time series analysis would explore:

i. Serial Correlation - statistical dependence between observations of an attribute in a dataset at different times as a function of time lags. Other terms used to describe this dependence are Autocorrelation or Seasonality. In this context, seasonality refers to regular systematic or periodic variations in a dataset that are typically calendar-related. For example, monthly seasonality is observed when data twelve months apart are related. Seasonality exists if there are regular spaced and reasonably consistent variations in terms of the timing, direction and magnitude of a data attribute. For example, if delays have been consistently low in January of prior years, they can be expected to be low in current observation of January delays. Nevertheless, seasonal data could still exhibit random variations within a particular season when compared over multiple years. Seasonal variations are usually an effect of natural conditions (e.g. weather) or social events (e.g. holidays and vacation).

ii. Trends - long-term directions in the time series that are not calendar related. For example, in a given time period, a positive trend could be an increase in total year-over-year ridership while a negative trend could be a decrease in total year-over-year delay. Trends are usually an effect of changes in population, employment and other socioeconomic characteristics

iii. Cross Correlation - the condition that a current observation of an attribute in the dataset is a function of lagged observations of other attributes. For example, the ridership in a given time period could depend on performance in prior time periods. Cross Correlation can be a result of cause-effect relationships between attributes.

7.1 Time Series Analytical Methods

In the literature the following methods are frequently used to analyze time series data:

i. Smoothing

ii. The family of ARMA - Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), ARIMAX (dynamic regression)

iii. Transfer Function Models

The next section provides a brief overview of each method as well as a selection of an appropriate technique to be used in this thesis.

7.1.1 Smoothing

As discussed earlier, even though the demand and train operation data exhibit seasonal patterns, the clarity of the pattern can sometimes be muffled by the random variations. A technique called "smoothing" is often times used to highlight trends, seasonality, or cyclic components of times series data. Smoothing involves some form of statistical combination of proximal data points such that the nonsystematic components of individual observations cancel each other out[xi]. There are two different statistical measures typically used to estimate the central tendency of the dataset- averages and medians. Although averages are used more frequently, the median is sometimes preferred because it is not affected by unusually high or low values of the observation (outliers). An advantage of using means over median is that it allows individual observations to be weighted differently based on proximity of the data points. In moving average smoothing, each observation in time is replaced with the simple or weighted means of N surrounding observations, where N refers to the width of the smoothing window[xii]. In single moving average smoothing, simple averages are applied and all observations in the smoothing window are weighted equally. In comparison, in exponential moving average smoothing exponentially decreasing weights are assigned to observations within the smoothing window that are older. The equation for moving average smoothing is:

$$M_t = \frac{w_t Y_t + w_{t-1} Y_{t-1} + \cdots + w_{t-N+1} Y_{t-N+1}}{N}$$

Where,

Y_t are the original time series observations

w_t are the weight applied to time dependent observations

N is the number of observations in the smoothing window

M_t is the moving average

7.1.2 Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), ARIMAX

Box and Jenkins (1976) developed the family of ARMA models to explore the characteristics of time series datasets made up of systematic variations as well as large errors or random variations in order to reveal hidden or unclear patterns in the data. An ARMA model consists of two parts, an autoregressive process (AR) and a moving average process (MA). Both processes are described below. However, the ARMA model is much simpler if the time series are stationary

Stationarity requirement

The basic idea of the stationarity requirement is that the time series observations do not accumulate past effects over time. In essence, time series that exhibit growth or decay over time are not stationary. Specifically, a process is said to be stationary if the joint distribution of random variables is the same irrespective of time. In other words, the probability laws that govern the behavior of the time series process do not change over time. The trend effects in a non-stationary time series can be eliminated through mean adjustments. The mean-adjusted series, $y_t = Y_t - \bar{Y}$, where Y_t is the original time series, $\bar{Y}$ is the sample mean.

Autoregressive Process: $\boldsymbol{AR(p)}$

The autoregressive process controls the serial correlation between current values and past values in a time series, in that at any given time, the process "remembers" some of the past values in the series. The autoregressive model expresses a time series as a linear function of one or more time-lagged elements (past values). In effect, the model is simply a linear regression of current values of the series against prior values of the series, including an error component. The *order* of the AR model, p is the number of lagged elements in the model, such that,

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t$$

Where Y_t is the original time series, α_p is the autoregressive coefficient associated with each time-lagged element, and ε_t is white noise (also called error, random shock or residual). An autoregressive coefficient (α_p) value equal to zero indicates that there is no temporal dependence between the time series elements, while large α_p values indicate that current values in the series are highly influenced by the past values. A useful autoregressive model captures the dependence structure in the data accurately enough so that the series of error component is random (not autocorrelated) and normally distributed.

Moving Average Process: $\boldsymbol{MA(q)}$

In addition to the autoregressive process, each observation in the time series can be affected by past errors (or random shocks or residuals) not accounted for by the autoregressive component. The process is a moving average of a series of shocks such that current values of the series can be found from current shocks as well as past shocks/errors ε. In other words, each observation in the model is made up of a random error component and a linear combination of prior random shocks/errors. The order of the MA model, q is the number of lagged shocks included in the model, such that,

$$Y_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Where Y_t is the original time series, ε_t is the shock/error, and θ_i is the moving average coefficient.

Autoregressive Moving Average Process: $ARMA(p, q)$

An $ARMA(p, q)$ process is a combination of the $AR(p)$ and $MA(q)$ processes, which includes lagged terms of the time series itself and lagged terms of the series of error components.

Autoregressive Integrated Moving Average Process: $ARIMA(p, d, q)$

The "I" in ARIMA stands for "Integrated", which refers to the process of differencing non-stationary series through one or more mean-adjustments to achieve stationarity. In essence, instead of using the original time series, the differenced or mean-adjusted series are used in the ARMA processes. The first differenced series, $y_t = Y_t - \bar{Y}$, where Y_t is the original time series, $\bar{Y}$ is the sample mean. Consequently, the ARIMA process is made up of three separate processes i) the autoregressive process with p lags of the series, ii) differencing process with d mean-adjustments, and iii) the moving average process with q lags of the error series. For example, an $ARIMA(1,1,3)$ refers to a model with 1 autoregressive parameter, p and 3 moving average parameters, q which were computed for the series after it was differenced once. Observing the autocorrelation plots provide a good rule of thumb in deciding which process to use. The expected pattern of the autocorrelation and partial autocorrelation plots for the AR(p), MA(q) and ARMA (p,q) are shown in Table 7.1 and described in detail in Section 7.2.

Test	AR(p)	MA(q)	ARMA(p,q)
Autocorrelation	Tails off	Cuts off after lag **q**	Tails off
Partial Autocorrelation	Cuts off after lag **p**	Tails off	Tails off

Table 7.1: Autocorrelation and Partial Autocorrelation Tests for Identifying AR, MA and ARMA models

Seasonal Autoregressive Integrated Moving Average Process: $SARIMA(p, d, q)$x$(P, D, Q)_m$

Seasonal ARIMA models rely on seasonal lags and differences to fit the seasonal pattern. The seasonal part of the ARIMA model has three additional components, where P is the number of seasonal autoregressive terms, D is the number of seasonal differences, Q is the number of seasonal moving average terms, and m is the number of periods per season.

7.1.3 Transfer Function Models

Transfer function models allow for lagged and decaying effects of covariates. The ARIMAX and regression with ARMA errors are special cases of transfer functions.

ARIMAX (p, q, d) includes the linear effect that one or more exogenous series have on the stationary response series of Y.

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t + \sum_{k=1}^{r} \beta_k x_{t-k}$$

where, the first sum is the AR component, the second sum is the MA component, and the third sum is the dependence on other variables component.

7.2 Time Series Analysis of Amtrak Data

The time series analyses of the Amtrak data are presented in this section. The section is organized by the different statistical methods used in identifying the time series data. The time series analysis was conducted in R statistics software. Some section of the .R script are presented in this section in text boxes formatted as shown below:

```
>#Welcome to R

>AmtrakMonthlyData <- read.csv("~../Analysis/AcelaRegionalMonthlySeries.csv")
```

The methods used in identifying the time series processes are organized under the following subsections: Section 7.1.1 discusses linear additive decomposition; Section 7.1.2 discusses the Dickey-Fuller Test and Augmented Dickey Fuller Test for Stationarity. Section 7.1.3 is on autocorrelation functions (ACF), partial autocorrelation functions (PACF), and cross correlation function (CCF). And finally Section 7.1.4 is on the family of ARMA models.

7.2.1 Linear Additive Decomposition

A time series can be additively decomposed into a set of independent functional forms referred to as trend, seasonal and random components. The trend is the deterministic, non-seasonal component. The seasonality or periodicity captures systematically repeating processes between groups of successive observations. And the random component contains the stochastic or irregular elements of the series, but could exhibit autocorrelations with the series. The linear additive decomposition is written as:

$$Y_t = T_t + S_t + \varepsilon_t$$

where Y_t is the original time series, T_t is the trend component, S_t is the seasonal component and ε_t random or error component.

The Amtrak Acela ridership and performance series were additively decomposed in R.

```
>AcelaMonthlyRiderscomp <- decompose(AmtrakMonthlyTS[,4])

>Plot(AcelaMonthlyRiderscomp)
```

Figure 7.1 shows the decomposition of Acela monthly ridership into four blocks; the first block shows the original time series, the second block shows the trend, the third block shows the seasonal component, and the last block shows the random variations. The trend line exhibits a steep upward slope till 2008, a negative trend between 2008 and 2010, a positive and then relatively flat slope between 2010 and 2013, and finally a slight upward slope between 2013 and 2014. The seasonal portion shows Acela ridership series with peaks at the end of winter/beginning of spring and fall, and troughs in the summer and end/beginning of the year. The random component represents a combination of unsystematic variations in the Amtrak ridership in response to various external events.

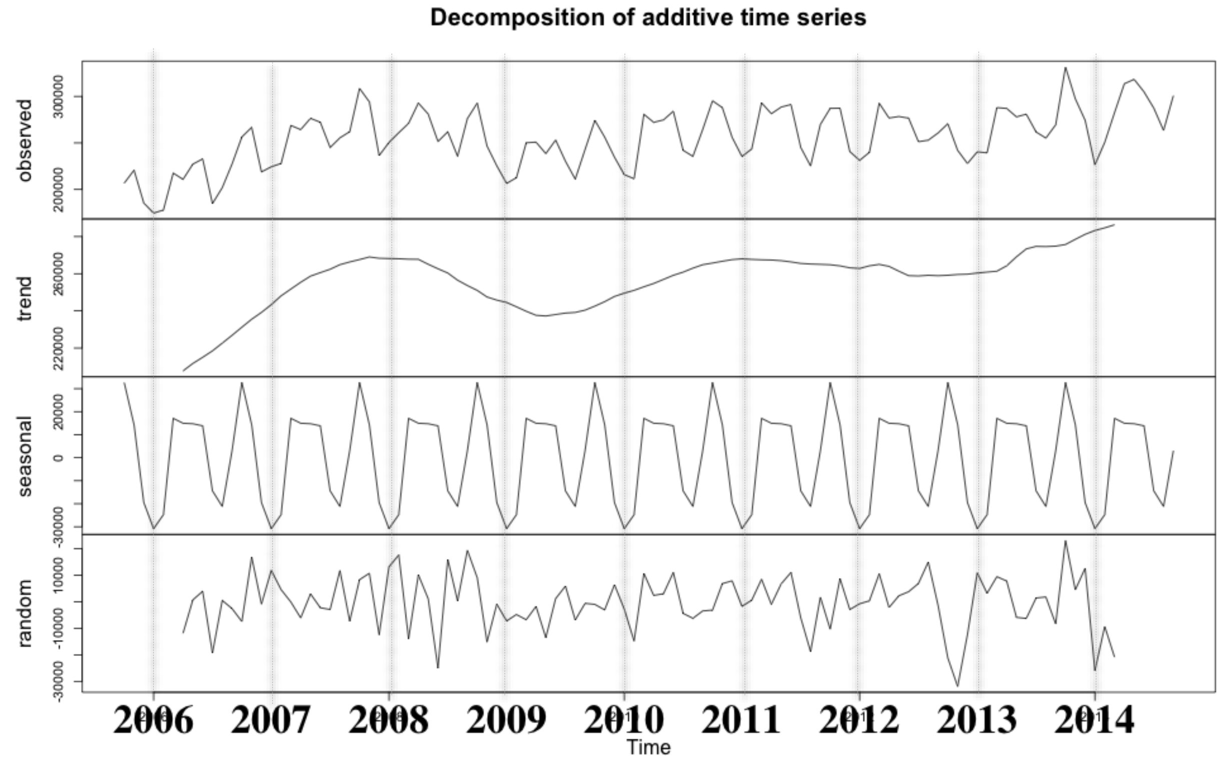


Figure 7.1: Acela Monthly Ridership Additive Decomposition

Figure 7.2 and Figure 7.3 show the decomposition of Acela monthly OTP and average delay per mile time series into the trend, seasonal and random components. The trends of both series show performance deterioration between 2006 and 2014 with OTP decreasing and delays increasing through the series. The seasonal components show high performance in February and September corresponding with peaks in the OTP chart and troughs in the average delay chart, and vice versa in the summer and at the end/beginning of the year. The random shock components show the residuals not explained by the trend or seasonal components.

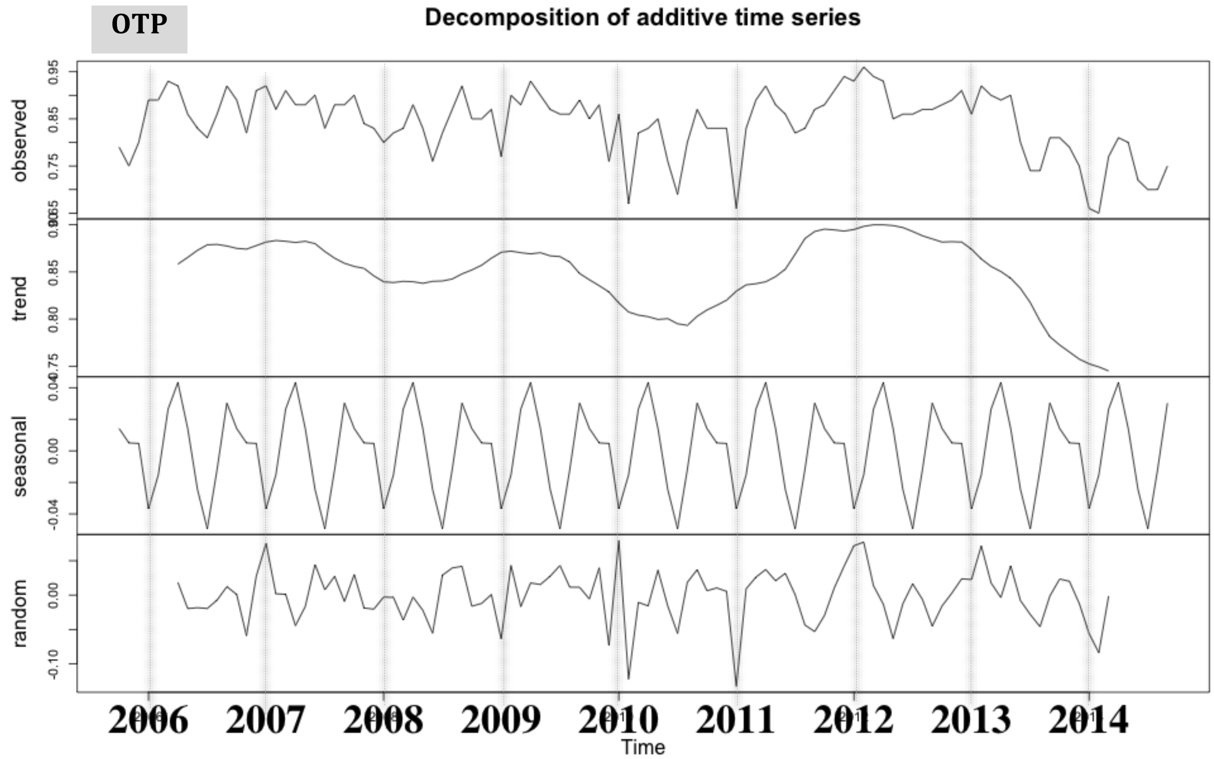


Figure 7.2: Acela Monthly OTP Additive Decomposition

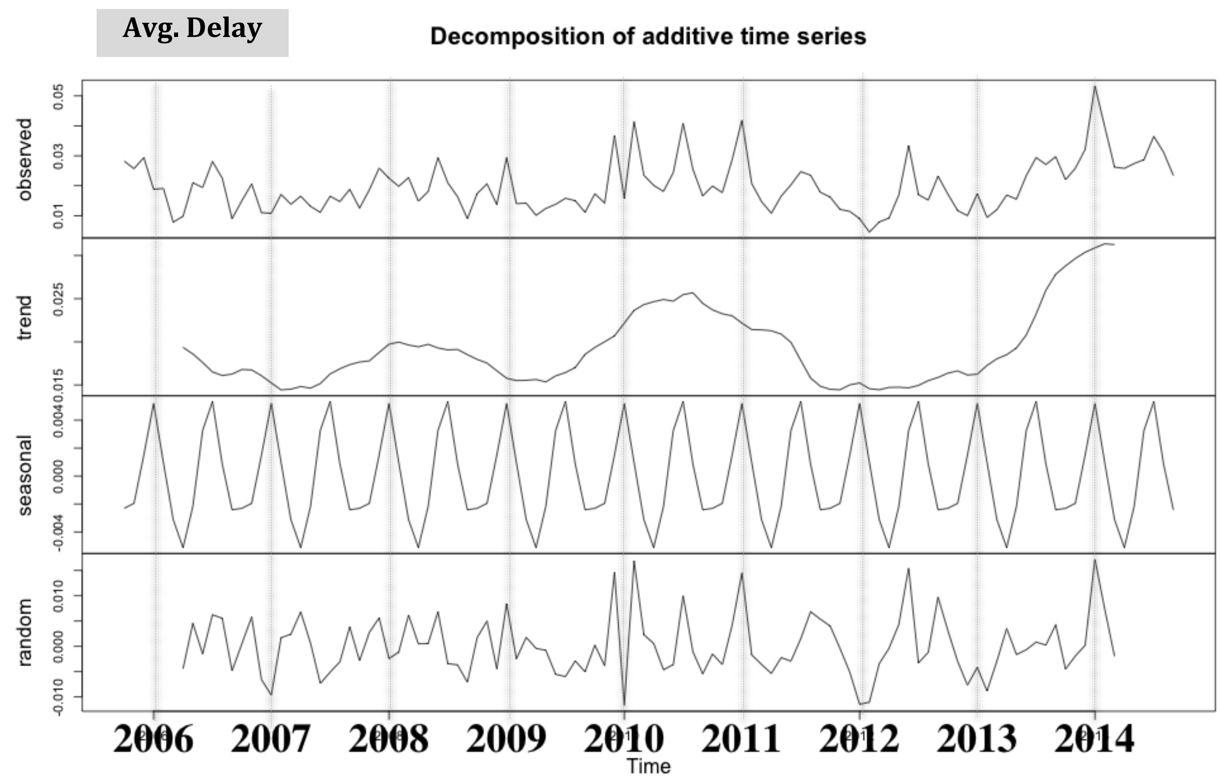


Figure 7.3: Acela Monthly Average Delay Per Mile Additive Decomposition

7.2.2 Dickey-Fuller Test (DF) and Augmented Dickey-Fuller (ADF) Test for Stationarity

The Dickey-Fuller Test was used to check for stationarity, that is, whether increasing or decreasing effects in the time series aggregate or die out over time. The Dickey-Fuller (DF) Test is expressed as

$$Y_t = \alpha Y_{t-1} + \varepsilon_t$$

$$\nabla Y_t = (\alpha - 1)Y_{t-1} + \varepsilon_t = \gamma Y_{t-1} + \varepsilon_t$$

where ∇ is the first difference operator and $\gamma = \alpha - 1$. When $\gamma = 0$, $\alpha = 1$ (unit root), and the series is not stationary. A non-stationary process where $Y_t = Y_{t-1} + \varepsilon_t$ is also referred to as a random walk. However, if serial correlation also exists in the original times series then the Augmented Dickey-Fuller (ADF) test is used to test for stationarity. Said and Dickey (1984) augmented the basic autoregressive unit root test to accommodate general ARMA models with unknown orders. The ADF Test is expressed as:

$$Y_t = \alpha Y_{t-1} + \sum_{i=1}^{p-1} \varphi_i Y_{t-i} + \varepsilon_t$$

$$\nabla Y_t = (\alpha - 1)Y_{t-1} + \sum_{i=1}^{p-1} \varphi_i Y_{t-i} + \varepsilon_t = \gamma Y_{t-1} + \sum_{i=1}^{p-1} \varphi_i Y_{t-i} + \varepsilon_t$$

where p is the order of the autoregressive component. The null hypothesis for non-stationarity is H_0: $\gamma = 0$ ($\alpha = 1$, unit root), and the hypothesis that there is a unit root is rejected if the test statistic is significant. The expression for the ADF test in R is shown in black below and the R output is shown in blue. If the p-value is small when the alternative is stationary, reject the null

```
>adf.test(AcelaMonthlyRiders, alternative="stationary", k=0)

Augmented Dickey-Fuller Test

Dickey-Fuller = -5.6596, Lag order = 0, p-value = 0.01

alternative hypothesis: stationary
```

hypothesis.

The output (in blue) shows that the p-value is small so the Acela ridership series from FY 2005 to FY 2014 is overall stationary.

The series for the OTP and average delay were also stationary.

7.2.3 Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF)

After the time series was stationarized, the next step was to determine whether AR or MA terms were needed to correct any autocorrelation that existed in the series. The autocorrelation function (ACF) between Y_t and Y_{t-k} is given by:

$$\frac{Covariance(Y_t, Y_{t-k})}{Variance(Y_t)}$$

The Partial Autocorrelation Function (PACF) captures the correlation between Y_t and Y_{t-k} that is not explained by the correlations at all lower-order lags (lag 1 through k-1). Specifically, the ACF and PACF are useful in identifying the order of the autoregressive (AR) and moving average (MA) models.

ACF and PACF plots are typically used as visual aids to identifying the orders of an ARMA model. The plots show the correlation coefficients for a specified number of lags in the series, as well as 95% confidence interval bands for statistical significance. For example, Figure 7.4 shows the ACF plot for the Acela monthly ridership up to a lag of 120 months (10 years).

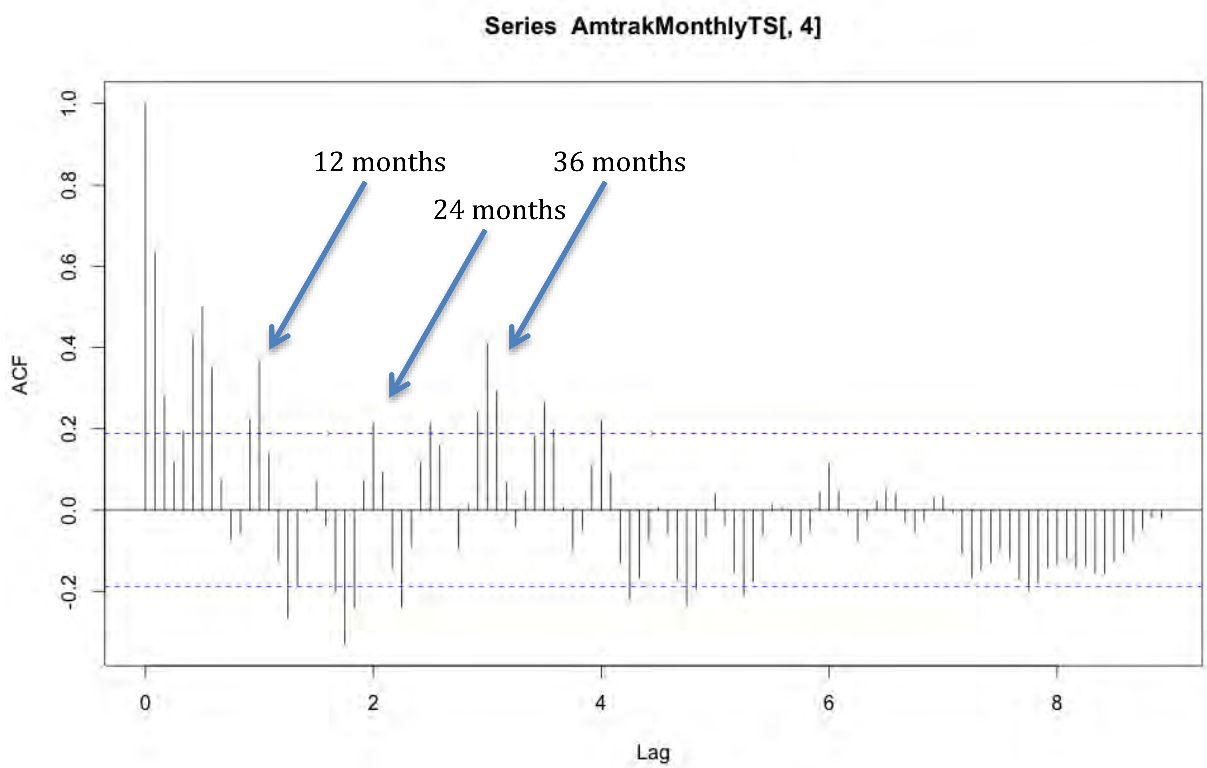


Figure 7.4: ACF for Monthly Acela Ridership at Max Lag = 120 months

The first bar in the autocorrelation plot (ACF) shows the correlation with itself at 0 lag and thus always has a coefficient equal to 1. The second bar shows a 0.6 correlation coefficient between the ridership series and the 1-month lagged series. The plot also shows an alternating pattern of positive and negative spikes, and significant correlations every 12 lags, which indicates the presence of seasonal effects. To help identify the non-seasonal components, the seasonal component was removed and the ACF plot was generated on the seasonally differenced series.

```
>Ridersseasonallyadjusted <- AmtrakMonthlyRiders - AcelaMonthlyRiderscomp$seasonal

>acf(Ridersseasonallyadjusted, lag.max=120)
```

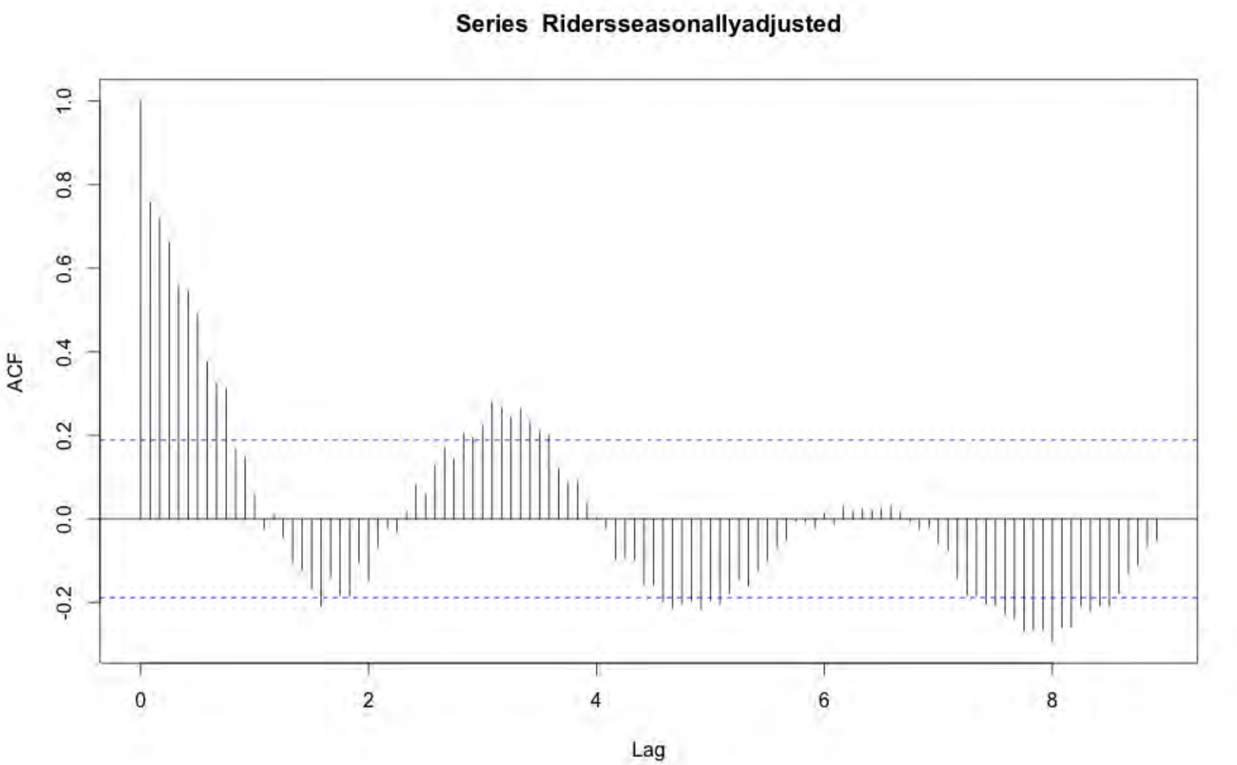


Figure 7.5: ACF of Seasonally Differenced Monthly Acela Ridership Series

Figure 7.5 shows a mixture of exponential decay and a sinusoidal pattern, which indicates that an seasonal ARMA model with order greater than one may be appropriate. The partial autocorrelation function (PACF) plot was also generated on the seasonally differenced ridership series. Figure 7.6 shows the PACF plot. Unlike the ACF plot, in the PACF plot, the first bar at lag 0 shows the correlation after the first lag is removed. Consequently, the coefficient of lag 0 in the PACF is proportional to the coefficient of lag 1 in the ACF.

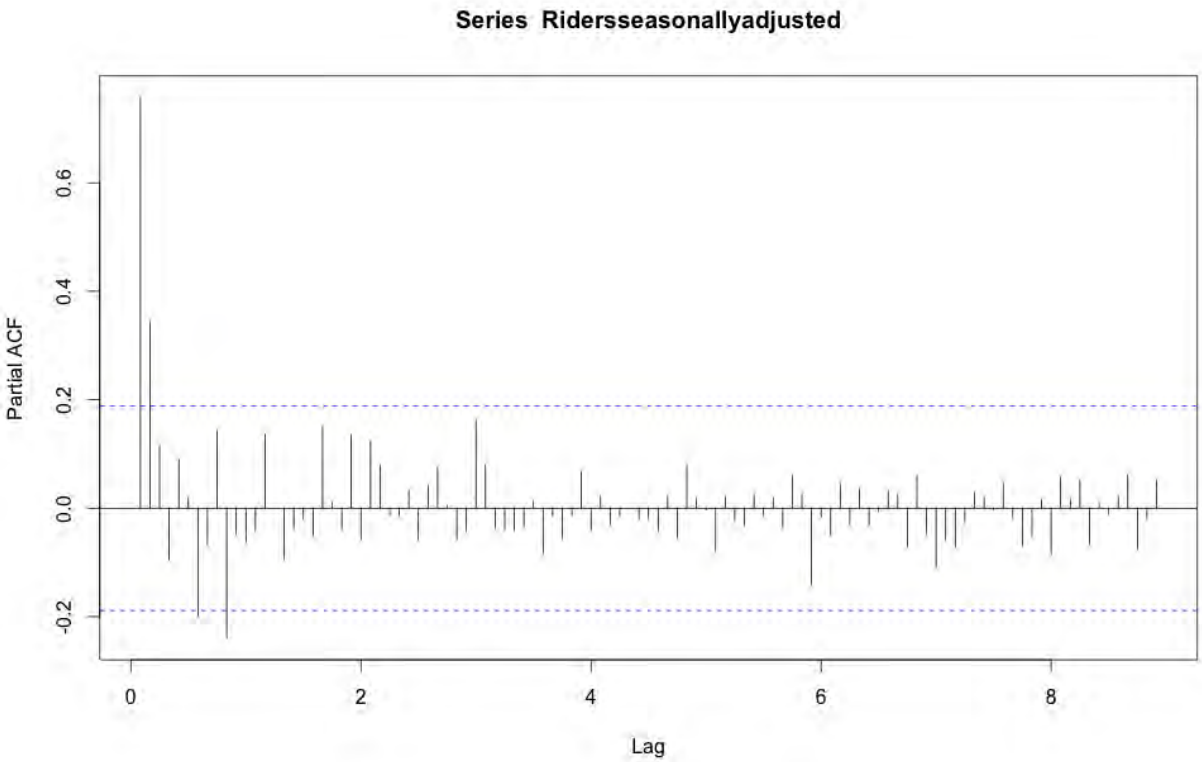


Figure 7.6: PACF of Seasonally Differenced Monthly Acela Ridership Series

The partial autocorrelation (PACF) plot shows significance on the second lag and pretty much non significance in the rest of the series, which suggests that an MA(2) model might be appropriate. The 7th and 10th lag were also significant, indicating some remaining seasonality.

Figure 7.7, Figure 7.8, Figure 7.9 and Figure 7.10 show the ACF and PACF of the Monthly Acela OTP and Average Delay time series. Both ACF plots exhibit a mixture of exponential decay and a damped sinusoidal pattern, while the PACF plots show significance at the 1st and 2nd lags, and again at the 12th lag, which suggest an MA(2) process and some remaining seasonality.

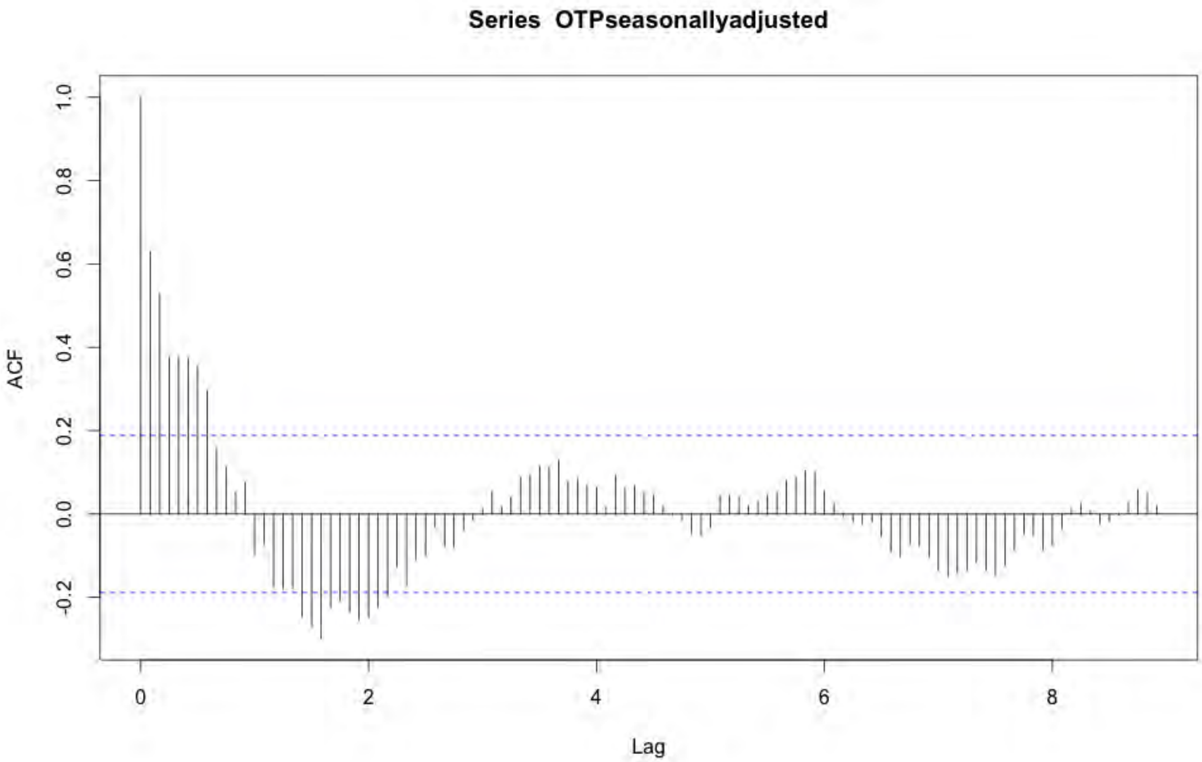


Figure 7.7: ACF of Seasonally Differenced Monthly Acela OTP Series

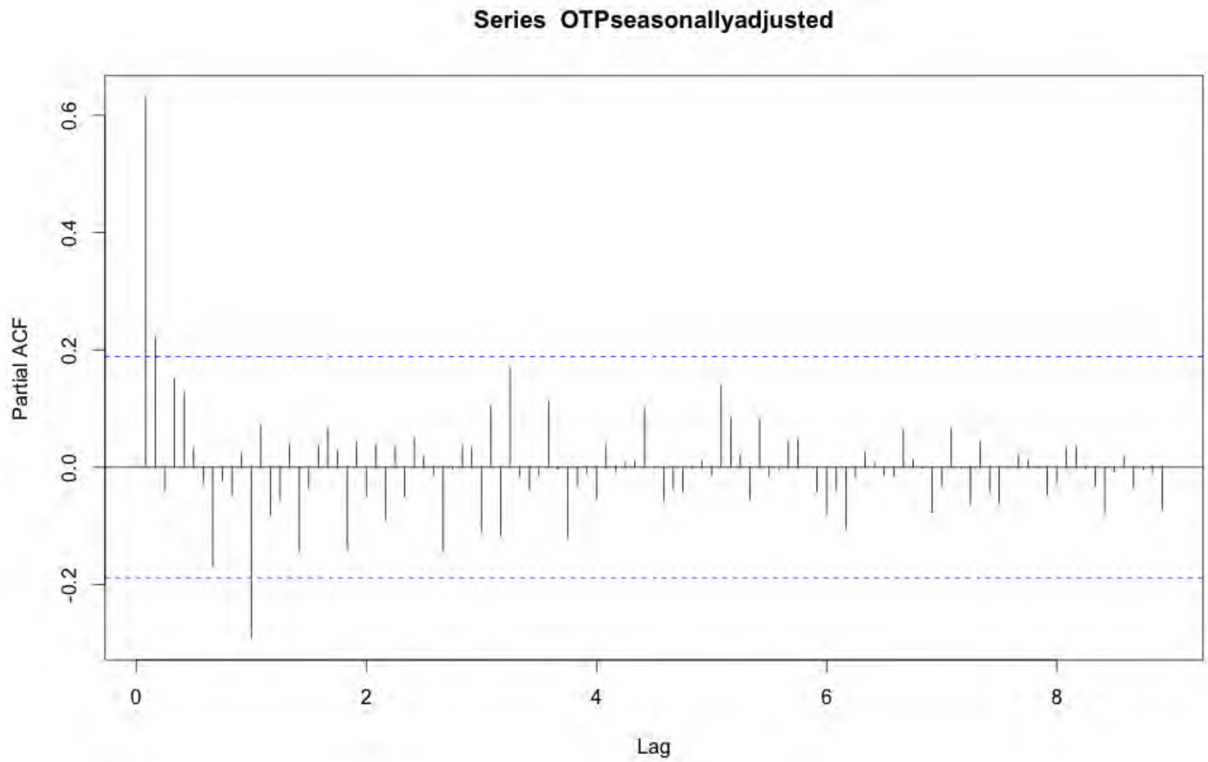


Figure 7.8: PACF of Seasonally Differenced Monthly Acela OTP

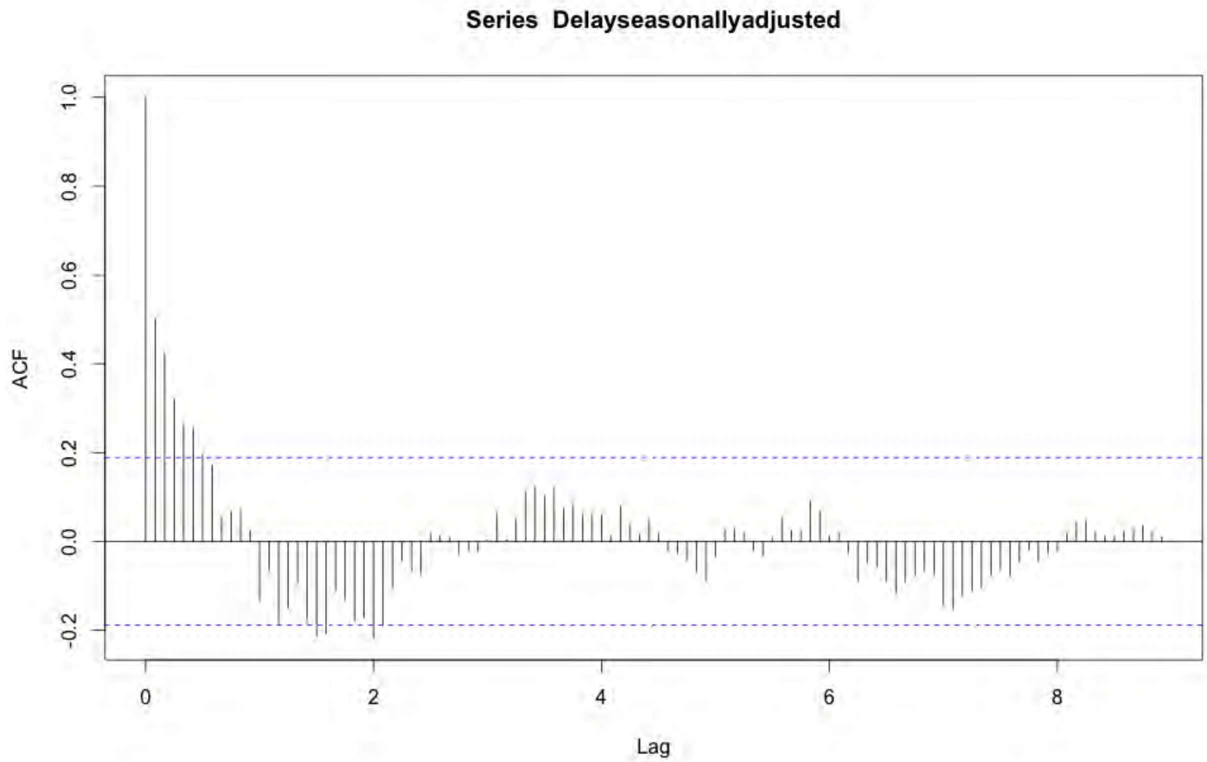


Figure 7.9: ACF of Seasonally Differenced Monthly Acela Average Delay

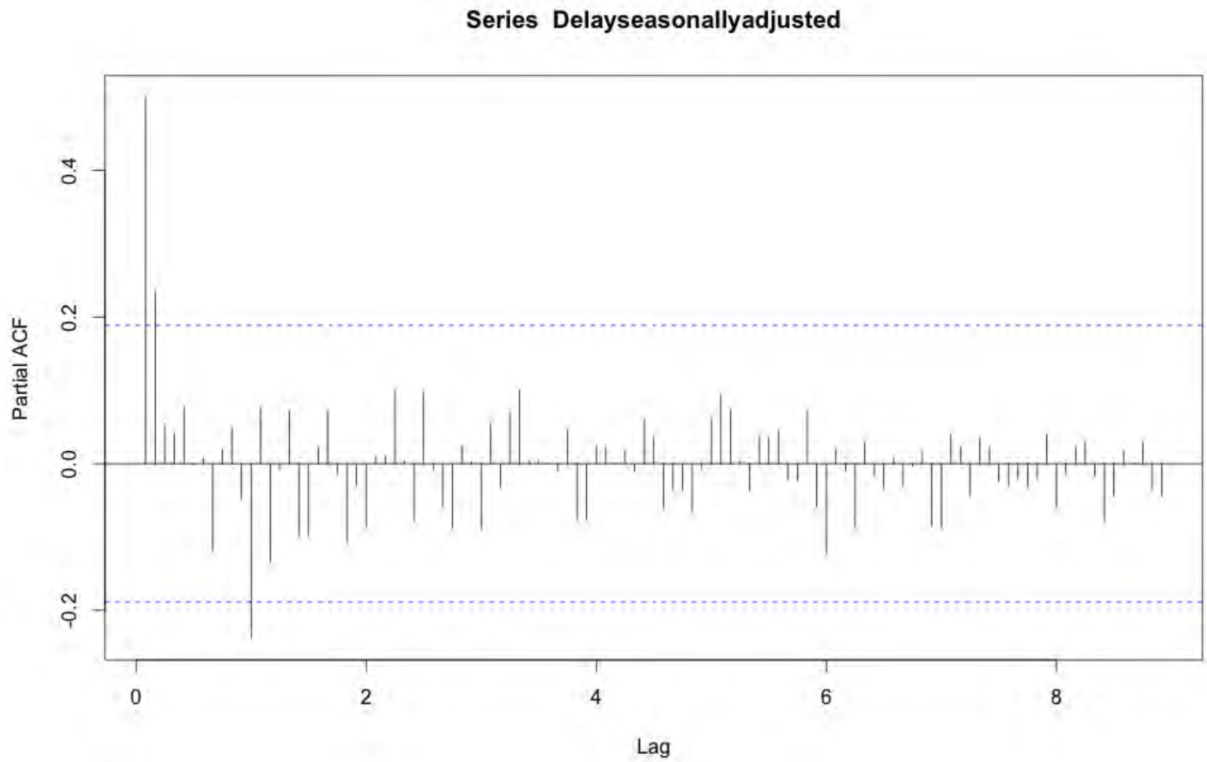


Figure 7.10: PACF of Seasonally Differenced Monthly Acela Average Delay

Starting with the performance metrics, based on the observations of the ACF and PACF plots, a seasonal ARIMA(1,1,1) model was used to fit the original Monthly Acela OTP time series data

$$X_t - X_{t-1} = \theta_1(X_{t-1} - X_{t-2}) + \alpha_1\varepsilon_{t-1} + \psi_1(X_{t-12} + X_{t-13}) + \varphi_1\varepsilon_{t-12} + \varepsilon_t$$

Where θ_1 is the AR(1) parameter, α_1 is the MA(1) parameters, and ψ_1 and φ_1 represent the

```
>ma = arima(AcelaMonthlyOTP order = c(1, 1, 1), seasonal=list(order=c(1,0,1), period=12))

>ma

Coefficients:

      ar1     ma1       sar1      sma1

    0.2941  -0.7544   0.9943  -0.9553

s.e.  0.1541   0.1062   0.0375   0.1483
```

seasonal parameter. The R expressions (in black) and outputs (in blue) are shown below:

Then the Box-Ljung Test was used to test that the residuals from the model up to 30 lags are random with adjusted degrees of freedom to account for the 3 estimated parameters. Under the Box-Ljung Test, the null hypothesis is that the residuals are random.

```
>BLT = Box.test(ma$residuals, lag=30, type = "Ljung-Box", fitdf=3)

>BLT

X-squared = 35.2654, df = 27, p-value = 0.1323

# To determine critical region:

>qchisq(0.95,27)
```

Since the X-squared statistic is less than the qchisq critical value, the null hypothesis of the Box-Ljung test was not rejected and the fitted model was adequate.

A seasonal ARIMA(1,1,1) model was also used to fit the original Monthly Acela Average Delay time series data and is shown below:

```
>ma_delay = arima(AcelaMonthlyDelay, order = c(1, 1, 1), seasonal=list(order=c(1,0,1), period=12))

>ma_delay

Coefficients:

        ar1       ma1      sar1      sma1

     0.2916   -0.8287   0.9915   -0.9646

s.e. 0.1600    0.1073   0.0564    0.1245

sigma^2 estimated as 5.384e-05:  log likelihood = 372.17,  aic = -734.34


>BLT_delay = Box.test(ma_delay$residuals, lag=30, type = "Ljung-Box", fitdf=3)

>BLT_delay

X-squared = 27.5248, df = 27, p-value = 0.4358
```

Similarly, the Box Ljung test was not rejected since the X-squared statistic was less than the qchisq critical value.

There were a lot of complex trends in the ACF and PACF of the ridership and seasonally different ridership series, thus the auto.arima function in R was used to identify the seasonal model useful for fitting the original Monthly Acela Ridership time series. The R expression and outputs are shown below. An ARIMA(1,0,1)(2,1,1)[12] with drift was identified.

```
> auto.arima(AcelaMonthlyRiders)


ARIMA(1,0,1)(2,1,1)[12] with drift

Coefficients:

       ar1     ma1     sar1     sar2     sma1     drift

    -0.4550  0.6425  -0.4626  -0.6308  -0.3357  369.3629

s.e.   0.1874  0.1749   0.0967   0.0602   0.2088   47.8278

sigma^2 estimated as 170176354:  log likelihood=-915.19

AIC=1870.53   AICc=1871.8   BIC=1888.48


>BLT_delay = Box.test(ma_delay$residuals, lag=30, type = "Ljung-Box", fitdf=3)

>BLT_delay

X-squared = 22.7545, df = 27, p-value = 0.6981
```

7.3 Transfer Functions and Cross Correlation Function (CCF)

This section focuses on modeling the relationship between two time series, where one could be related to past lags of the other series. The cross correlation function (CCF) is often used to identify lags of one series that might be useful predictors of the other series. The CCF is defined as the set of correlations between the predictor series X_{t+h} and the "independent" series Y_t for h lags. Negative values of h suggest that the X_t series leads the Y_t series, while positive h values suggest the opposite. For example, $h = -1$ indicates that current values of Y_t are influenced by X_{t-1}, the values of the X_t series from the prior period.

The cross correlation plots of the seasonally differenced monthly OTP and ridership series is shown in Figure 7.11 and that for the seasonally differenced monthly average delay and ridership series is shown in Figure 7.12.

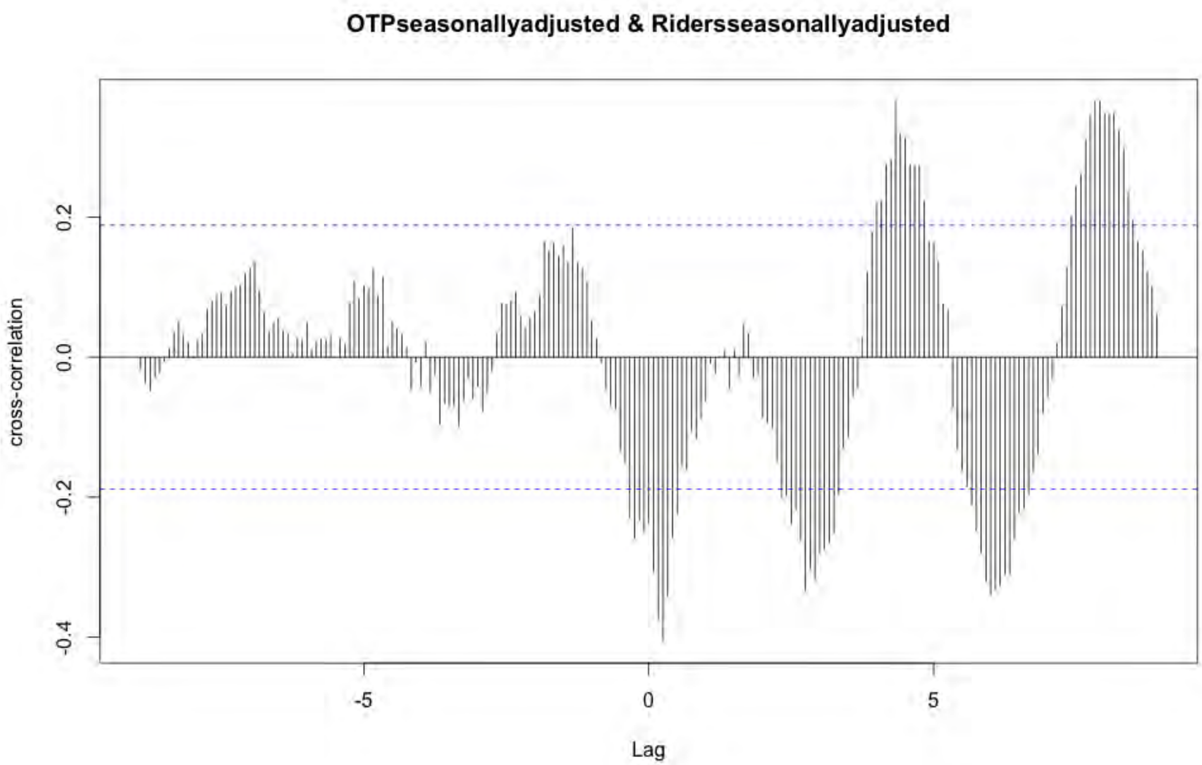


Figure 7.11: CCF between Monthly Acela OTP and Ridership Series

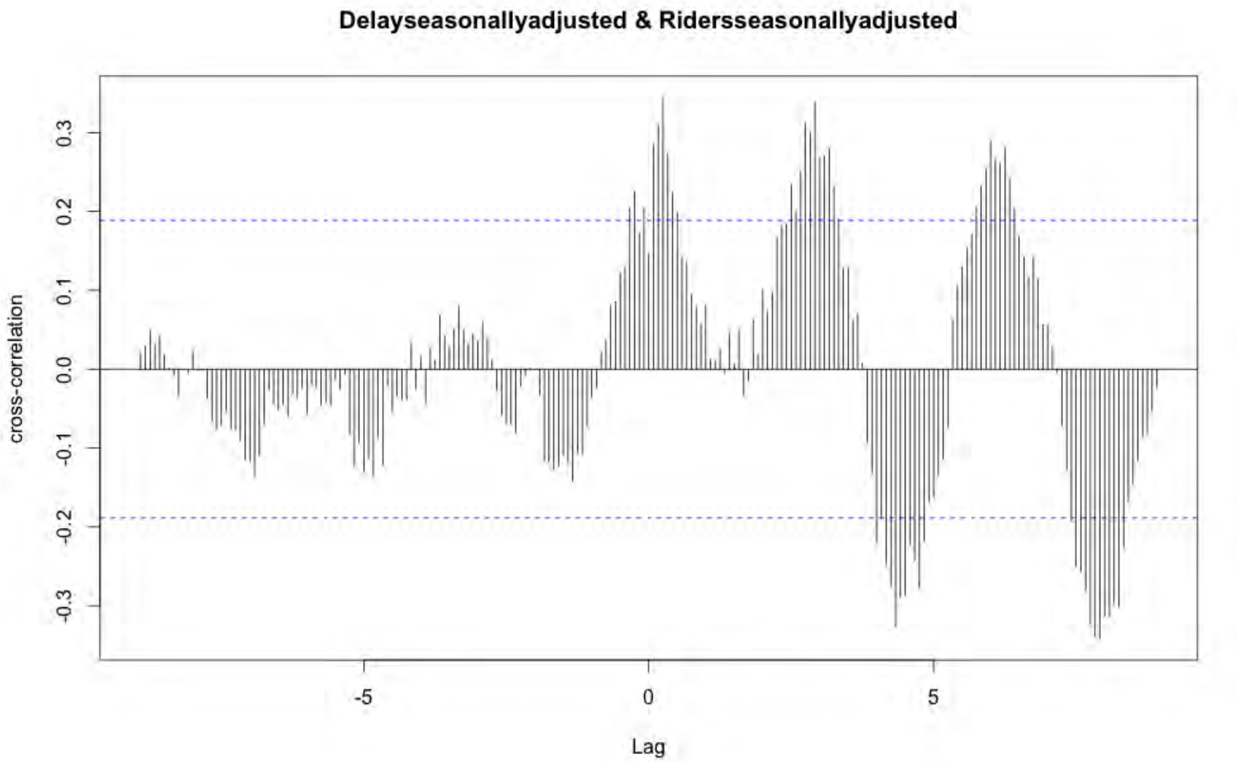


Figure 7.12: CCF between Monthly Acela Average Delay and Ridership Series

The significant and dominant cross correlations in both figures occur for positive values of h, which suggests that the performance series lags the ridership series. The CCF plots of the seasonally differenced series also exhibit a periodic pattern.

Transfer functions are used to model times series e.g. Y_t as a function of past lags of Y_t as well as current and past lags of another series, X_t. Again the auto.arima function in R was used to fit relationship between Acela monthly ridership series and the monthly OTP series. However, the results were complex to interpret and thus are not shown in this thesis.

7.4 Discussion of Time Series Analysis

The time-series analysis of Acela ridership had complex autocorrelations, seasonality, and cross correlations with the performance series, which made it difficult to quantify the effect of performance on ridership. Furthermore, the results obtained are inconclusive and given the time constraint for this thesis, the author was not able to dig deeper into the analysis and thus has made suggestions regarding time series modeling for future work.

<This page has been intentionally left blank>

8 Summary, Conclusion and Recommendations

This chapter provides a summary of findings in **Section 8.1,** thesis conclusions in **Section 8.2** and final recommendations regarding performance of the Acela and Regional services in **Section 8.3**. In addition, **Section 8.4** includes a list of potential topics that could supplement the work done in this thesis for future consideration.

8.1 Summary

The analyses and discussions in Chapter 1 to 7 showed that the Acela and Regional experienced large amount of delays each year between FY 2005 and FY 2014, and more so in FY 2014. In FY 2014, 58% of operated Acela and Regional trains arrived at their final destination later than the scheduled arrival time, causing delays to 66% of total Acela and Regional riders (about 7.7 million passengers). These delays resulted from variability in train travel times and service performance. The author investigated a number of factors that influence service performance and evaluated their impact in FY 2012 and FY 2014. FY 2012 and FY 2014 represented the years with best and worst annual service performance between FY 2005 and FY 2014. In FY 2012, although the percentage of trains impacted by delays were fewer (41% of trains arrived late), some of the factors causing routine delays appeared to be similar to those in FY 2014. The factors characterizing ridership and service performance variations discussed in this thesis were:

8.1.1 Seasonality and month of year

On both Acela and Regional services, there were clear signs of monthly performance variations; winter months typically experienced the best performance, except during severe winter weathers and summer months typically suffered poorer performance due to heat restrictions, infrastructural issues (catenary wire drooping) and track work (usually scheduled in good weather months). Similarly, the ridership profile exhibited seasonal patterns. For Regional, January and February regularly had the lowest ridership, which coincided with the end of the winter holiday and vacation, while all other months had relatively similar ridership levels. For Acela, the lowest ridership were observed in January and August likely due to vacation during the winter and summer holidays and highest ridership were usually in the fall and spring months.

8.1.2 Day of week and time of day

Other than Saturday, which usually experienced lower levels of delays, service performance on the Acela and Regional did not appear to vary considerably by day of week. Conversely, average ridership on both the Acela and Regional appeared to be much lower on Saturdays. Furthermore, of the weekdays, the ridership was typically the highest on Wednesdays and Thursdays on the Acela and on Thursdays and Fridays on the Regional.

Additionally, time of day was shown to impact Acela and Regional performances as certain morning and evening peak hour trains experienced both high ridership and significant delays. Also, the best and worst performing Acela and Regional trains in FY 2012 and FY 2014 were usually the same suggesting systematic issues. Overall, all trains experienced a significant amount of delays, including the first train of the day, which theoretically should be able to achieve on-time arrivals regularly.

8.1.3 Administration, management and control elements (e.g. operating crew, timetable construction, etc.)

iv. The first train analysis attributed routine delays on Acela and Regional trains to 'avoidable' delays and timetable construction artifacts. Firstly, the analyses in both FY 2012 and FY 2014 revealed that the first train of the day often times departed from the originating stations with about 1 to 3 minutes delays on average, which accumulated and propagated at each consecutive station downstream. Late departures of the first train of the day from the originating station should be avoidable. In terms of timetable construction artifacts, the author could not detect scheduled train dwell time at each station and the scheduled segment speeds and travel times appeared to be different from actual speeds and travel times on average, which led to odd accumulation and dissipation of en-route delays. These effects were noticed on both Regional and Acela, but the segment level deviations were more pronounced on the Acela service. On the Acela, the difference between scheduled and actual average segment travel times and speeds led to delay accumulation in some segments (where average actuals were slower than scheduled), and delay reduction in others segments (where average actuals were faster than scheduled). So for example, end-to-end travel time between Boston and Washington would be close to the timetable but could be late for intermediate stations, suggesting timetable padding to provide an opportunity for a train to “catch up”.

8.1.4 Capacity levels on trains

Compared to FY 2012, in FY 2014, more Acela trains appeared to be near or at capacity in both the north-end and south-end segments of the corridor, especially the trains operated during morning or evening peak periods. Both Acela and Regional trains operated during the peak time of day exhibited poorer performance.

8.1.5 Accidents and incidents (e.g. signal failures, weather-related, track work, etc.)

Unanticipated weather-related and third-party events were responsible for about 12% of Acela and Regional train delays (based on FY 2012 and FY 2014 analysis). Furthermore, in FY 2012, there was only one major delay caused by operational issues; however, in FY 2014, about 50% of severe daily delays (delays >10,000 minutes) were associated with equipment and infrastructure issues. This evidences the impact of inadequate track renovation and infrastructure maintenance on the Northeast Corridor in recent years[xiii]. Furthermore, not including days with severe weather issues, Amtrak rarely cancels Acela and Regional trains.

8.1.6 Interference from other trains

Interference from other trains occurred on days with severe delays caused by accidents or weather disruptions. However, on an average day most Acela and Regional trains experienced delays less than 30 minutes and since Amtrak usually spaces trains in intervals greater than or equal to 30 minutes, routine delays on one train usually did not cascade to the next scheduled train. The train interference analysis was done by observing consecutive train departures from each station; however a more detailed analysis for future work would be to observe complete daily round-trip of train sets/locomotives and train crews.

8.1.7 Demand response to service performance

The annual demand response to annual delay and annual OTP showed one-year lagged correlations as high as 0.74 indicating that current Acela and Regional on-time performance levels were associated with ridership the following year. It is unclear whether the one-year lagged associations exhibited between service performance and ridership on both the Acela and Regional services are cause-effect relationships between service performance and demand or simply correlations. Furthermore, the times series analysis indicated autocorrelations between Acela ridership in adjacent months, and a similar autocorrelated relationship for on-time performance and delays. The author believes that the time series analyses are inconclusive and could be enhanced to fully understand the impact of service performance on demand.

8.2 Conclusions

In this section, the author refers back to the figures introduced in Chapter 1. Figure 8.1 and Figure 8.2 characterize the large picture of performance on the Acela and Regional services respectively, in the Northeast Corridor (NEC). As a reminder, the black line represents the PRIIA on-time performance goals, which Amtrak set together with the Federal Railroad Administration (FRA) in FY 2010. In both figures, the blue bars show that the actual annual on-time performance on the Acela and Regional services trended towards the PRIIA goals until FY 2012, and then performance deterioration was sustained and amplified between FY 2013 and FY 2014. Consequently, the regrettable conclusion is that both Acela and Regional are currently underperforming. If the positive trend between FY 2010 and FY 2012 had continued in FY 2013 and FY 2014, the conclusions will likely have been different. The major causes of delay and travel time variability were attributed to (i) timetable construction artifacts such as poorly estimated segment speeds and no dwell time allowances at many stations, (ii) minor avoidable issues such as late departure from originating station, (iii) absence of policies and programs to keep Amtrak accountable to established goals, and (iv) infrastructure and rolling stock deterioration. Recommendations are made in Section 8.3 regarding each of these major issues. Furthermore, the relationship between service performance and ridership was assessed through a preliminary correlation analysis, which indicated a one-year lagged correlation but further research is required to ascertain actual cause-effect dependence.

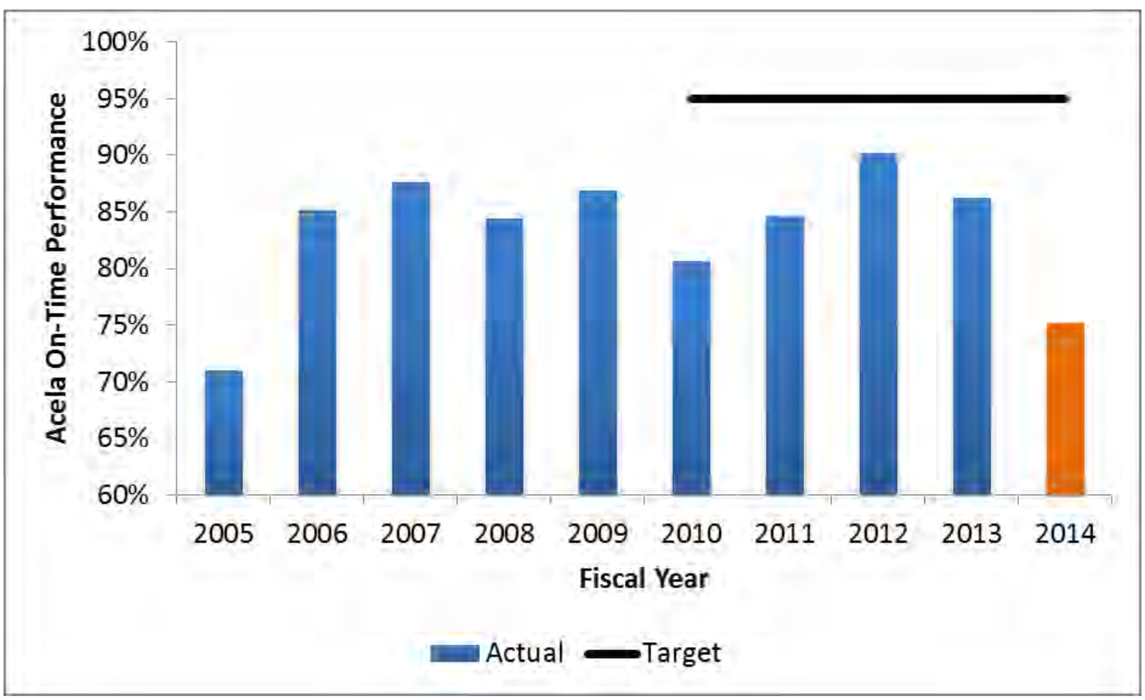


Figure 8.1: Acela FY 2005 – FY 2014 On-Time Performance

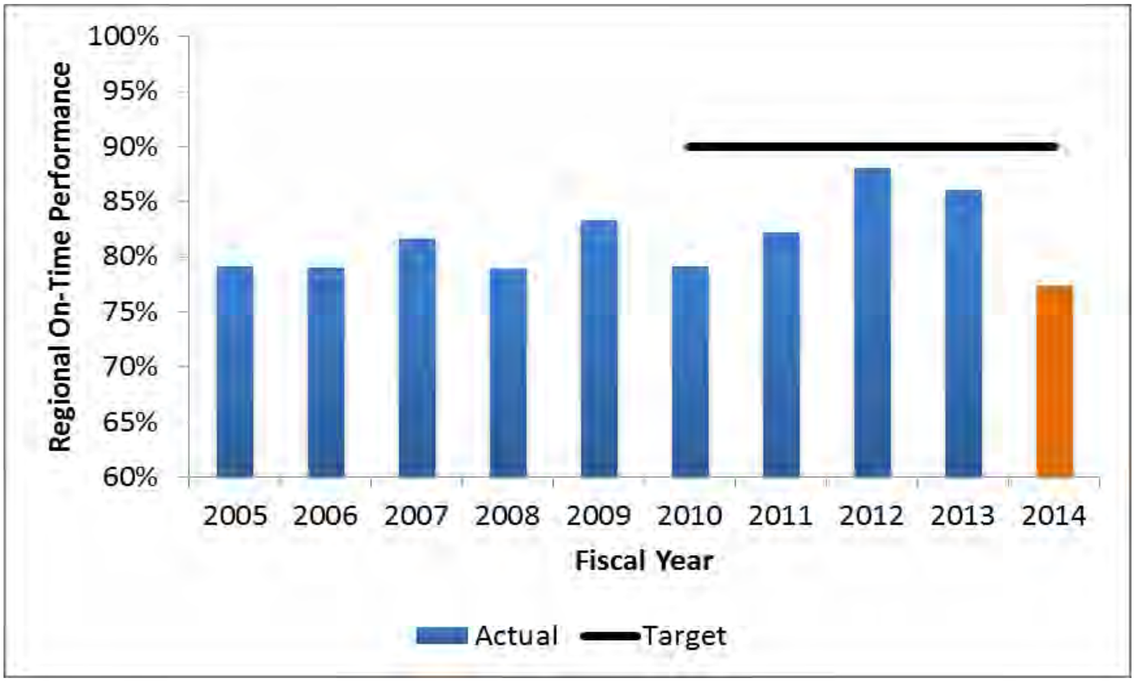


Figure 8.2: Regional FY 2005 - FY 2014 On-Time Performance

The findings presented in this thesis are drawn from analyzing Acela and Regional ridership and service operations databases. Although the findings are reliable, they are preliminary and further research at a more detailed level is required to make precise conclusions about each cause or effect factor discussed. Additionally, the data sources required to make precise judgments will need to be broadened to include not just Acela and Regional ridership and train operations data but also data on NEC commuter and freight service operations, data on the competitive balance between rail and the air, auto and bus modes in the NEC, as well as in-person interviews of Amtrak management and train and operations crews, etc. Some suggestions for future work are made in Section 8.4.

Lastly, the ridership and train operations data used for the thesis analyses were the best of Amtrak's in-house Acela and Regional records. However, the train operations data had shortcomings such as missing or incorrectly entered train arrival or departure times, which were attributed to human error. The defects of the data will likely not change the major conclusions of this thesis, since the conclusions reflect aggregates (totals) and averages over multiple trains and days. However, for more accurate and sophisticated bookkeeping, the actual train operations data entry should be upgraded from a manual method to a more automatic process. The first recommendation in Section 8.3 relates to refining the data records.

8.3 Recommendations

This section provides suggestions and active steps on how Amtrak can monitor and improve actual service performance, as well as service performance records for the Northeast Corridor.

8.3.1 Refine data records

In September 2013, Amtrak launched a new interactive tool to track daily train operations in real-time. The train tracker provides accurate information about the location of each train en-route, including train speeds, departure time from originating station, and scheduled and actual arrival times at each successive station. The automatic vehicle locators (AVL) data should replace Amtrak's current train operations database to provide a more accurate record of train performances. This would eliminate the shortcoming of the current dataset, which had missing train information at some stations because the train personnel failed to record the actual arrival and/or departure times of the train. This would make the dataset cleaner and further make train cancellations or stations that were not served clearer in the data records.

8.3.2 Refine timetables

The author found Amtrak historical timetables[xiv] that were effective April 1, 1990 through October 27 1990. In the historical timetables, Train 107 on the Metroliner Service (Acela predecessor) was scheduled to depart from New York at 9:00AM and arrive in Washington, DC at 11:49AM. In comparison, Amtrak current timetables effective January 12, 2015[xv] shows that Acela Express Train 2151 is also scheduled to depart from New York at 9:00AM and arrive in Washington, DC at 11:49AM. Although the stopping pattern of the trains have changed slightly, this shows that Amtrak has not implemented any significant train schedule adjustments in more than 25 years, even when train technology changed. In this thesis, daily routine delays were attributed to timetable artifacts, which affected both Acela and Regional trains. Routine delays appeared to stem from deviations of operated train from scheduled segment-level speed and travel times and undetectable terminal time at stations, which accumulated at subsequent segments and stations downstream. Consequently, the refined data obtained from the automatic vehicle locator should be used to revamp or in the very least, adjust train timetables. Furthermore, Amtrak should plan to fine-tune the train timetables on an annual basis, and especially when new trains sets and train slots are introduced.

8.3.3 Educate and reinforce on-time culture

Under PRIIA requirements, Amtrak currently has at least 18 delay codes categorized under Amtrak- and third-party-responsible delays, used to report causes of delays and responsible party. Although this database was not provided to the author, other public reports[xvi] indicate that Amtrak-responsible delays include delays causes by passengers boarding and alighting, delays caused by crew lateness, etc, while third-party delays include delays caused by weather-related issues, police-activity issues, etc. While the third party delays are for the most part unavoidable, some Amtrak-responsible delays should obviously continue to be managed and reduced by educating and reinforcing on-time culture for operating and managing Amtrak crewmembers. Furthermore, Amtrak should monitor and utilize the information in the cause of delay and responsible party database in order to tackle, reduce and eliminate some of the minor causes of delays, such as late departure from originating station. Furthermore, delays attributed to train, crew and control center personnel's can be made managed through a monthly or quarterly review process for train managing and operating personnel, or on-board visual aids to help train drivers monitor deviations from scheduled timetables.

8.3.4 Management, Policies and Programs

Strict policies and programs like PRIIA Section 207 are required to help Amtrak meet established goals. The analyses in this thesis showed that even though Amtrak owns most of the track in the Northeast Corridor, both Acela and Regional services experienced an unprecedented high in service performance in FY 2011 and FY 2012 while PRIIA Section 207 was active, and both services have been encountering performance deterioration since PRIIA Section 207 was overturned in FY 2013. In addition to statutory laws like PRIIA, other proven techniques like Six Sigma and Lean (used predominantly in manufacturing systems and more recently in health care systems) could be utilized to improve quality output of train operations by identifying and removing the causes of errors and minimizing variability in service operations.

8.3.5 Upgrade Infrastructure

In the long-term, Amtrak requires adequate funding to address essential track alignment/ curvature renovations, catenary maintenance, bridge and tunnel restorations, and rolling stock and signal improvements along the Northeast Corridor. In FY 2012, there was only one major delay caused by operational issues, however, by FY 2014, about 50% of the severe daily delays (greater than 1,000 minutes) were attributed to equipment and infrastructural issues, which evidences the deteriorating infrastructure in the Northeast Corridor in recent years. Consequently, Amtrak's Northeast Corridor Capital Investment Program[xvii], which was designed to achieve a state of good repair and facilitate performance enhancement in the corridor requires a stable, multi-year funding program as opposed to the current unpredictable annual appropriation of funds.

8.4 Future Work

Although this thesis presented a wide range of analyses, the author believes the research only scratched the surface in terms of truly understanding and improving the service performances of Amtrak's Acela and Regional services in the Northeast Corridor. Some suggestions for future research include:

i. Train interference and cascading delays analysis by observing complete daily round-trips of Amtrak train sets/locomotives.
ii. Compare service performance of Amtrak trains with air, auto and bus modes in the Northeast Corridor, especially on the days with really poor performance attributed to weather condition
iii. Cumulative service performance metric that captures not only the magnitude (delay minutes) and frequency of delay occurrence on operated trains (on-time performance), but also the effect of cancelled trains.
iv. Time series and regression analysis to study demand response to service performance using the cumulative service performance metric.

v. A comprehensive impact of performance in the Northeast Corridor, including not just Amtrak Acela and Regional services, but also other Amtrak services, commuter rail services and freight rails services that share the track in the Northeast Corridor.

vi. Investigating Amtrak's on-time culture given that circumstantial evidence suggesting that some delays could be attributed to a poor on-time culture.

vii. Correlation between lateness leaving first station and late arrivals at the final destination.

viii. Analysis on the impact of timetable padding that results in end-to-end travel time being close to the timetable but higher delays at intermediate stations.

Looking into the future: High-speed passenger rail in the Northeast Corridor

The Northeast Corridor is a vital segment of U.S. rail, enhancing connectivity, mobility and economic productivity in the Northeast Region, the densest of the 11 U.S. megaregions. The author believes the ambitious plans to build "true" Next Generation high-speed rail in the Northeast Corridor is viable from a ridership and revenue point of view. In FY 2014, both the Acela and Regional set a record high of 3.55 and 8.08 million annual passengers, respectively and combined annual revenue record of $8.1 billion. Conversely, the rolling stock and service performance of Acela and Regional, as well as the corridor infrastructure exhibited signs of deterioration in the last two years, and especially in FY 2014. Consequently, significant infrastructure investments and service improvements will need to be in place before true international high-speed rail standards can be attained in the corridor.

Coda

Prior to him becoming my academic advisor, I met with Professor Sussman in his office at the beginning of the school year. The first thing I remember noticing was a placard quote attributed to Albert Einstein, prominently displayed on his shelf, that read: "If we knew what we were doing, it wouldn't be called research." In that moment, I would not have guessed that the quote would become such a key source of inspiration and encouragement for me as I worked on my thesis throughout the year. A few weeks after my first meeting with Professor Sussman, I asked him, and he agreed to be my thesis advisor. Every time I went into his office for yet another discussion of my thesis, I took a second to re-read that quote, and to remind myself that having all the right answers during the process would have defeated the very point of doing the research.

Finally, I would like to thank [you] the reader, for paying attention to the analyses and discussions presented in this thesis. I hope the thesis has been informative and useful, and provides a good background to build on in the future.

9 References

Amtrak. (2009a). *An Interim Assessment of Achieving Improved Trip Times on the Northeast Corridor* (Vol. 212).

Amtrak. (2009b). *Northeast Corridor State of Good Repair Spend Plan* (Vol. 2008).

Amtrak. (2012). The Amtrak Vision for the Northeast Corridor 2012 Update Report, (July).

Amtrak. (2013). Amtrak teams with google to create interactive train locator map. *Amtrak News Release*.

Asensio, J., & Matas, A. (2006). AN EMPIRICAL ESTIMATION OF THE VALUE OF TRAVEL TIME. *Association for European Transport and Contributors*.

Asensio, J., & Matas, A. (2008). Commuters' valuation of travel time variability. *Transportation Research Part E: Logistics and Transportation Review*, *44*(6), 1074–1085. http://doi.org/10.1016/j.tre.2007.12.002

Bates, J., Polak, J., Jones, P., & Cook, A. (2001a). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2-3), 191–229. http://doi.org/10.1016/S1366-5545(00)00011-9

Bates, J., Polak, J., Jones, P., & Cook, A. (2001b). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2-3), 191–229. http://doi.org/10.1016/S1366-5545(00)00011-9

Batley, R., Dargay, J., & Wardman, M. (2011a). The impact of lateness and reliability on passenger rail demand. *Transportation Research Part E: Logistics and Transportation Review*, *47*(1), 61–72. http://doi.org/10.1016/j.tre.2010.07.004

Batley, R., Dargay, J., & Wardman, M. (2011b). The impact of lateness and reliability on passenger rail demand. *Transportation Research Part E: Logistics and Transportation Review*, *47*(1), 61–72. http://doi.org/10.1016/j.tre.2010.07.004

Batley, R., Ibáñez, J. N., Wardman, M., Shires, J., & Gerard, W. (2007). A Discrete Choice Study to Addresss the Impact of Reliability on Passenger Rail Demand. *Association for European Transport and Contributors*, 1–21.

Bayissa, T. F. (2013). *How do Delays Along the Way Influence Delays to Final Destination ?*.

Börjesson, M., Eliasson, J., & Franklin, J. P. (2012). Valuations of travel time variability in scheduling versus mean–variance models. *Transportation Research Part B: Methodological*, *46*(7), 855–873. http://doi.org/10.1016/j.trb.2012.02.004

Bougas, C. (2013). *Forecasting Air Passenger Traffic Flows in Canada : An Evaluation of Time Series Models and Combination Methods*. Unversite Laval.

Carrion, C., & Levinson, D. (2012a). Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice*, *46*(4), 720–741. http://doi.org/10.1016/j.tra.2012.01.003

Carrion, C., & Levinson, D. (2012b). Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice*, *46*(4), 720–741. http://doi.org/10.1016/j.tra.2012.01.003

Chui, C., & Chen, G. (n.d.). *Kalman Filtering with Real-Time Applications* (Fourth Edi). Springer.

Chung, S., Kang, K., & Engineering, T. (2013). A Comparison Study of Forecasting Accuracy of Econometric Models on Domestic Freight Demand in South Korea. *Proceedings of the Eastern Asia Society for Transportation Studies*, *9*(2003).

Clements, M. P., & Henry, D. F. (2004). *A Companion to Economic Forecasting*. Blackwell.

Cryer, J. D., & Chan, K.-S. (n.d.). *Time Series Analysis With Applications in R* (Second Edi). Springer.

Dargay, J., Batley, R., Ibáñez, J. N., Wardman, M., & Shires, J. (2008). AN ECONOMETRIC ANALYSIS OF THE IMPACT OF RELIABILITY ON. *Association for European Transport and Contributors*, 1–17.

De Jong, G., Kroes, E., & Sanders, P. (n.d.). *The Value of Reliability*.

Duchateau, H., Debrincat, L., & Goldberg, J. (2005). On the value of punctuality on suburban trains to and from Paris. *Association for European Transport and Contributors*.

Farnsworth, G. V. (2009). *Econometrics in R*.

Federal Railroad Administration (FRA), & (DOT), D. of T. (2009). *Metrics and Standards for Intercity Passenger Rail Service*.

Fosgerau, M., & Engelson, L. (2011). The value of travel time variance. *Transportation Research Part B: Methodological*, *45*(1), 1–8. http://doi.org/10.1016/j.trb.2010.06.001

Fosgerau, M., & Karlström, A. (2010). The value of reliability. *Transportation Research Part B: Methodological*, *44*(1), 38–49. http://doi.org/10.1016/j.trb.2009.05.002

Fourer, R., Gertler, J., & Simkowitz, H. (1977). Models of Railrad Passenger-Car Requirments in the Northeast Corridor: An Application of Sesame. *Annals of Economics and Social Measurement*, *6*(4), 367–398.

Fridström, L., & Thune-Larsen, H. (1989). An econometric air travel demand model for the entire conventional domestic network: The case of Norway. *Transportation Research Part B: Methodological*, *23*(3), 213–223. http://doi.org/10.1016/0191-2615(89)90004-0

Guiraud, L., Gayda, S., Cabrita, I., & Kroes, E. (2014). Value of travel time reliability on French high-speed and regional services. In *Transport Research Arena* (pp. 1–10).

Hallowell, S. F., & Harker, P. T. (1998). Predicting on-time performance in scheduled railroad operations: methodology and application to train scheduling. *Transportation Research Part A: Policy and Practice*, *32*(4), 279–295. http://doi.org/10.1016/S0965-8564(97)00009-8

Halse, A. H., Killi, M., & Østli, V. (2014). *Estimating demand effects of rail reliability using disaggregated data*.

Hansen, I. (2001). Improving railway punctuality by automatic piloting. *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*, 792–797. http://doi.org/10.1109/ITSC.2001.948761

Harvey, A. C. (1990). Estimation Procedures for Structural Time Series Models. *Journal of Forecasting*, *9*(June 1988), 89–108.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(Series D), 35–45.

Kamga, C., & Yazıcı, M. A. (2014a). Temporal and weather related variation patterns of urban travel time: Considerations and caveats for value of travel time, value of variability, and mode choice studies. *Transportation Research Part C: Emerging Technologies*, *45*, 4–16. http://doi.org/10.1016/j.trc.2014.02.020

Kamga, C., & Yazıcı, M. A. (2014b). Temporal and weather related variation patterns of urban travel time: Considerations and caveats for value of travel time, value of variability, and mode choice studies. *Transportation Research Part C: Emerging Technologies*, *45*, 4–16. http://doi.org/10.1016/j.trc.2014.02.020

Kato, H., Kaneko, Y., & Soyama, Y. (2010). Influence of Chronic Delays in Tokyo Urban Rail Service on Passenger Behavior and Perceptions. *Association for European Transport and Contributors*, 1–14.

Kato, H., Kaneko, Y., & Soyama, Y. (2014). Economic benefits of urban rail projects that improve travel-time reliability: Evidence from Tokyo, Japan. *Transport Policy*, *35*, 202–210. http://doi.org/10.1016/j.tranpol.2014.06.004

Kim, J., & Mahmassani, H. S. (2014). A finite mixture model of vehicle-to-vehicle and day-to-day variability of traffic network travel times. *Transportation Research Part C: Emerging Technologies*, *46*, 83–97. http://doi.org/10.1016/j.trc.2014.05.011

Kouwenhoven, M., de Jong, G. C., Koster, P., van den Berg, V. a. C., Verhoef, E. T., Bates, J., & Warffemius, P. M. J. (2014). New values of time and reliability in passenger transport in The Netherlands. *Research in Transportation Economics*, *47*, 37–49. http://doi.org/10.1016/j.retrec.2014.09.017

Kouwenhoven, M., & Kroes, E. (2009). Preliminary Monetary Values for the Reliability of Travel Times in Freight Transport. *EJTIR*, *9*(2), 83–99.

Li, Z., Hensher, D. a., & Rose, J. M. (2010a). Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Transportation Research Part E: Logistics and Transportation Review*, *46*(3), 384–403. http://doi.org/10.1016/j.tre.2009.12.005

Li, Z., Hensher, D. a., & Rose, J. M. (2010b). Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Transportation Research Part E: Logistics and Transportation Review*, *46*(3), 384–403. http://doi.org/10.1016/j.tre.2009.12.005

Lin, J. (2006). *A Practical Guide to State Space Modeling: An important lemma for deriving the Kalman Filter*.

Litman, T. (2008). Valuing Transit Service Quality Improvements. *Journal of Public Transportation*, *11*(2), 43–63.

Noland, R. B., & Polak, J. W. (2002). Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*, *22*(1), 39–54. http://doi.org/10.1080/01441640010022456

Noland, R. B., Small, K. a., Koskenoja, P. M., & Chu, X. (1998). Simulating travel reliability. *Regional Science and Urban Economics*, *28*(5), 535–564. http://doi.org/10.1016/S0166-0462(98)00009-X

Office, B. C. (1979). *The Current and Future Savings of Energy Attributable to Amtrak*.

Okutani, I., & Stephanedes, Y. J. (1984). Dynamic prediction of traffic volume through kalman filtering theory. *Transportation Research Part B: Methodological*, *18B*(1), 1–11.

Pichler, P. (2007). *State Space Models and the Kalman Filter Seminar paper prepared for* (pp. 1–7).

Prato, C. G., & Rasmussen, T. K. (2014). *Estimating Value of Congestion and Value of Reliability from the Observation of Car Drivers ' Route Choice Behavior*. *Transportation Research Board 2014 Annual Meeting* (Vol. 5996).

Renne, J. (n.d.). State-Space Models and Applications.

Schimek, P. (2015a). Dynamic Estimates of Fare Elasticity Using Panel Data for U . S . In *Transportation Research Board 2015 Annual Meeting*.

Schimek, P. (2015b). *Dynamic Estimates of Fare Elasticity Using Panel Data for U.S. Transportation Research Board 2015 Annual Meeting*.

Shumway, R., & Stoffer, D. (n.d.). *Time Series Analysis and Its Applications With R Examples* (Third Edit). Springer.

Small, K. a. (2012). Valuation of travel time. *Economics of Transportation*, *1*(1-2), 2–14. http://doi.org/10.1016/j.ecotra.2012.09.002

Smith, B. L., & Demetsky, M. (1997). TRAFFIC FLOW FORECASTING : COMPARISON OF MODELING APPROACHES. *Journal of Transportation Engineering*, *i*(August), 261–266.

Smith, B. L., Williams, B. M., & Keith Oswald, R. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, *10*(4), 303–321. http://doi.org/10.1016/S0968-090X(02)00009-8

Song, H., & Li, G. (n.d.). *Tourism Demand Modelling and Forecasting A Review of Recent Research* (pp. 1–28).

Stadtler, D. J. (2014). *TESTIMONY OF D . J . STADTLER , VICE PRESIDENT OF OPERTIONS AMTRAK BEFORE THE SURFACE TRANSPORTATION BOARD*.

Stathopoulos, A., & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, *11*(2), 121–135. http://doi.org/10.1016/S0968-090X(03)00004-4

The Aerospace Corporation. (1976). *Northeast Corridor Passenger Transportation Data Study*.

The NEC Master Plan Working Group. (2010). *The Northeast Corridor Infrastructure Master Plan*.

Van Loon, R., Rietveld, P., & Brons, M. (2011). Travel-time reliability impacts on railway passenger demand: a revealed preference analysis. *Journal of Transport Geography*, *19*(4), 917–925. http://doi.org/10.1016/j.jtrangeo.2010.11.009

Wardman, M., & Nicolás Ibáñez, J. (2012). The congestion multiplier: Variations in motorists' valuations of travel time with traffic conditions. *Transportation Research Part A: Policy and Practice*, *46*(1), 213–225. http://doi.org/10.1016/j.tra.2011.06.011

Welch, G., & Bishop, G. (2006). *An Introduction to the Kalman Filter* (pp. 1–16).

Williams, B. M., Asce, M., Hoel, L. A., & Asce, F. (2003). Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process : Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*, (December 2003), 664–672.

Additional Publications

The Impact of Amtrak Performance in the Northeast Corridor (attached)

Tolulope Ogunbekun and Joseph Sussman

The Impact of Amtrak Performance in the Northeast Corridor

Tolulope Ogunbekun & Joseph Sussman
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA

The performance of Amtrak's Acela and Regional services in the Northeast Corridor (NEC) is a topic that, while frequently discussed as substandard by some travelers, has received minimal attention in the compendium of open source research literature. This brief focuses on Amtrak's Acela and Regional travel time performance in the last ten years (2005 to 2014).

Amtrak is a vital transportation provider on the Northeast Corridor serving travelers between Boston, MA and Washington, DC, including major cities such as Providence, RI; New Haven, CT; New York, NY; Philadelphia, PA; and Baltimore, MD. In Fiscal Year 2014 (FY 2014), Amtrak had a record high of 11.6 million passengers on the Acela and Regional services combined. However, only 3.9 million passengers arrived at their destination at the scheduled arrival time; that is, 7.4 million passengers experienced delays for a myriad of reasons. This brief evaluates different factors that led to variability in service performance, as well as the impact of service performance on ridership. The authors hopes that this research will inform the ongoing discussion on measures needed to strengthen intercity passenger rail in the Northeast Corridor.

NEC Background

The Northeast Corridor (NEC) is a 457-mile stretch of fully electrified railway line between Boston and Washington D.C. Figure 1 shows the NEC infrastructure ownership. Amtrak owns 79% (363 miles) of the NEC track, and operates (dispatches and maintains the right-of-way) throughout the corridor, except the 56-mile section between New Rochelle, NY. and New Haven, CT. controlled by Metro-North Railroad. The NEC is the busiest railroad in the U.S. both in terms of service frequency and demand. Over 2,200 trains operate on the corridor daily; 2,000 commuter trains, 153 Amtrak intercity passenger trains, and 70 freight trains. Between Amtrak and the commuter trains, there were approximately

750,000 daily riders, and in total about 260 million passenger trips made on the NEC in FY 2014.

Figure 1. NEC Infrastructure Ownership (Source: NEC Commission)[xviii]

Acela and Regional Performance Background

The Acela Express (Acela) and Northeast Regional (Regional) are the two main Amtrak services operating along the NEC mainline, and the focus of this research.

Amtrak routinely reports the **end-point on-time performance (OTP)**, which measures the percentage of Acela and Regional trains that arrive at their final destination at the scheduled arrival time. An Acela train is classified as "on time" if it arrives within 10 minutes of its scheduled arrival time, while a Regional train is classified "on time" if it arrives within 10 minutes for trips less than 250 miles, 15 minutes for trips between 251 and 350 miles, and 20 minutes for trips between 351 and 450 miles. In FY 2010, Amtrak together with the Federal Railroad Administration (FRA) set the OTP target at 95% for Acela and 90% for Regional. However, each year, both Acela and Regional have performed below the target.

Figure 2 shows the annual average end-point OTP for Acela compared to the performance target. The black line represents the OTP target for Acela (prior to FY 2010, Amtrak did not have an established on-time target). Between FY 2010 and FY 2014, despite the 10-minute arrival buffer, the annual on-time performances on the Acela service were about 5 to 20 percentage points below the target. The best Acela service performance was experienced in FY 2012 with an annual average OTP of 90% while FY 2014 experienced the worst with an annual average OTP of 75%. In other words, in FY 2014, 1 in 4 Acela trains arrived at their final destination more than the 10 minutes after the scheduled arrival time.

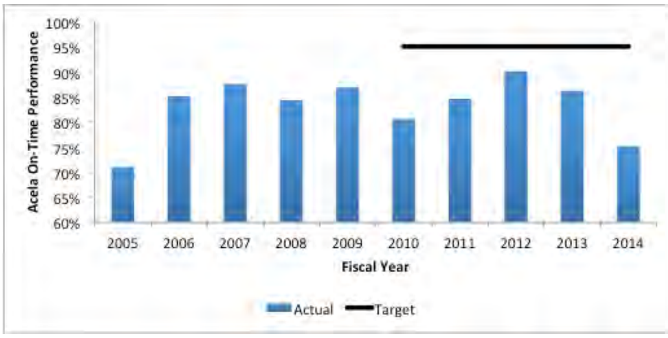


Figure 2. Acela Annual OTP

Despite the lenient buffer and lower OTP goal, Figure 3 shows that the Regional service also underperformed by about 5 to 20 percentage points between FY 2010 and FY 2014. Similar to the Acela, the Regional service experienced the best performance in FY 2012 with an annual average OTP of 88%, and the worst performance in FY 2014 with an annual average OTP of 77%. In this brief, some of the factors leading to travel time variability and underperformance on Amtrak's Acela and Regional services are discussed.

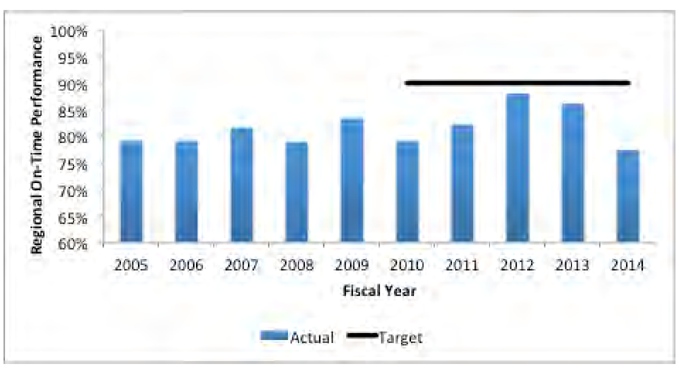


Figure 3. Regional Annual OTP

Factors Influencing Performance Fluctuations

The variation in Amtrak's service were studied under six main categories: i) seasonality and month of year variations, ii) day of week, and time of day differences, iii) performance variations due to administration, management and control elements, iv) service disruptions due to accidents and incidents (e.g. signal failures, weather-related, track work), v) disturbances due to interference from other trains, and vi) service fluctuations due to capacity levels on trains. The impacts of each of these factors were investigated by analyzing the FY 2005 to FY 2014 historical ridership and train operations data. The data analysis was used to identify the days and trains on which delays were incurred, which are useful for separating systematic trends from the random components in the delays.

Discussion of Results

Seasonality and Month of Year

Figure 4 shows the monthly on-time performance for Acela between FY 2005 and FY 2014. On both Acela and Regional services, there were clear signs of monthly performance variations; winter months typically experienced the best performance, except during severe winter weathers, and summer months typically suffered poorer performance due to heat restriction (which requires trains to run at a slower speed when track temperature exceeds 120 degrees), infrastructural issues (catenary wire drooping) and track work (usually scheduled in good weather months).

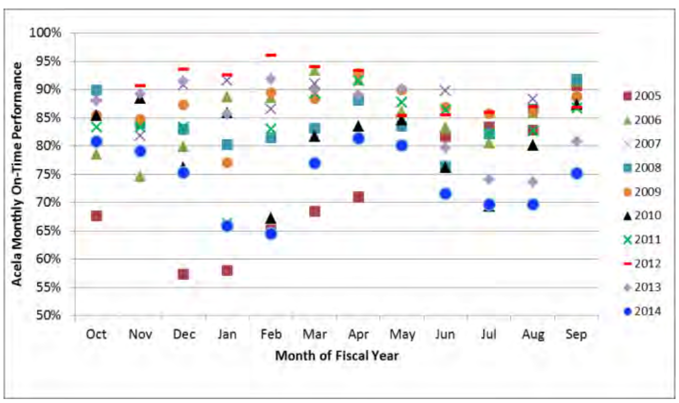


Figure 4. Acela Monthly On-Time Performance

Day of week and Time of Day

Other than Saturday, which usually experienced lower levels of delays, service performance on the Acela and Regional did not appear to vary considerably by day of week.

Time of day was shown to impact Acela and Regional performances as certain morning and evening peak hour trains experienced both high ridership and significant delays. Also, the best and worst performing Acela and Regional trains were usually the same across the years suggesting some systematic issues. Overall, all trains experienced a significant amount of delays, including the first train of the day, which theoretically should be able to achieve on-time arrivals regularly.

Administration, Management and Control Elements (e.g. operating crew, timetable construction, etc.)

The first train analysis served as the proxy to evaluate the effect of administration, management and control elements. The analysis attempted to track detailed station-to-station arrivals and departures of the first Acela and Regional trains over multiple days to identify causes of delay. The analysis revealed that the first train of the day often departed from the originating stations with about 1 to 3 minutes of delay on average, which accumulated and propagated at each consecutive station downstream.

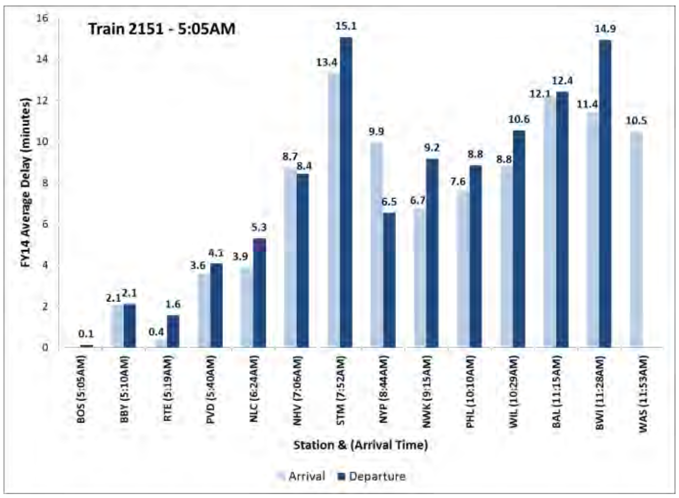


Figure 5. Train 2151 Average FY 14 Delay

For example, Figure 5 shows the average FY 2014 arrival and departure delays for Train 2151 at each successive station on the corridor. Train 2151 is the first southbound Acela train scheduled to depart from Boston South Station (BOS) at 5:05AM and arrive at Washington Union Station (WAS) at 11:53AM with a 15 minute en-route dwell time at New York Penn Station (NYP). In FY 2014, Train 2151 had an OTP of 72% and average delay of 11 minutes. The light blue bars are arrival delays while the dark blue bars represent departure delays.

Overall, the dark blue bars are higher than the light blue bars suggesting that trains accumulated additional delays between arrival at a station and departure from the same station. This is likely due to the fact that terminal time is not scheduled at all en-route stations to account for the time it takes for passenger to get on and off the trains. Additionally, upstream delays on the Acela and Regional seemed to accumulate on some segments and lessen on other segments as a result of flawed timetable segment train speeds and travel times.

Accidents and Incidents (e.g. Signal Failures, Weather-related, Track Work, etc.)

Unanticipated weather-related and third party events were responsible for about 12% of Acela and Regional train delays (based on FY 2012 and FY 2014 analysis). Furthermore, in FY 2012, there was only one major delay caused by an operational issue; however, by FY 2014, about 50% of severe daily delays (delays >1,000 minutes) were associated with equipment and infrastructure issues. This evidences the impact of inadequate track renovation and infrastructure maintenance on the Northeast Corridor in recent years[xix]. Furthermore, excluding days with severe weather issues, Amtrak rarely cancels Acela and Regional trains.

Interference From Other Amtrak Trains

On days with severe service interruptions caused by accidents or weather disruptions, delays on one Amtrak train usually cascaded to other trains. However, on a given average day, most Acela and Regional trains experienced delays less than 30 minutes and since Amtrak schedules trains in intervals greater than or equal to 30 minutes, routine delays on one train usually did not cascade to the next scheduled train.

Capacity Levels on Trains

Compared to prior years, in FY 2014, more Acela trains appeared to be near or at capacity in both the north-end and south-end segments of the corridor, especially the trains operated during morning or evening peak periods. The capacity-constrained Acela and Regional trains operated during the peak periods also exhibited poorer performance.

Demand Response to Service Performance

Figure 5 shows the total annual Acela ridership (blue) between FY 2005 and FY 2014 and the annual average OTP (red) from the prior year. The annual demand response to annual performance showed a one-year lagged correlation greater than 0.75. On both the Acela and Regional, the correlation coefficients were sufficiently high to propose that the ridership in a given year are typically associated with service performance from the prior year. However, it is unclear whether the one-year lagged associations are cause-effect relations or simply correlations.

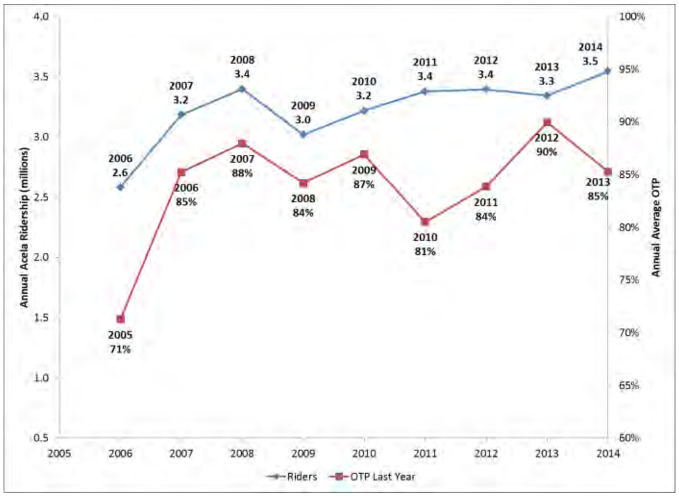


Figure 5. Lagged Annual Ridership to OTP

Conclusions and Recommendations

Figure 2 and Figure 3 highlight the conclusion of the research that both Acela and Regional services are underperforming. The major causes of delays and poor performance were attributed to: (i) timetable construction artifacts such as poorly estimated segment speeds and no dwell time allowances at many stations, (ii) minor avoidable issues such as late departure from originating station, (iii) infrastructure and rolling stock deterioration, and (iv) absence of policies and programs to keep Amtrak accountable to established goals. The findings from the research were used to recommend active steps for Amtrak regarding monitoring and improving service performance in the Northeast Corridor.

Refine Data Records

In September 2013, Amtrak launched a new interactive tool to track daily train operations in real-time. The train tracker provides accurate information about the location of each train en-route, including train speeds, departure time from originating station, and scheduled and actual arrival times at each successive station. The automatic vehicle locators (AVL) data should replace Amtrak's current train operations database to provide a more accurate record of train performances. This would eliminate the shortcoming of the current dataset, which had missing train information at some stations because the train personnel failed to record the actual arrival or departure times of the train. The AVL data would refine Amtrak's current data records.

Refine Timetables

The author compared Amtrak historical timetables[xx] with current timetables[xxi], which showed that Amtrak has not implemented any significant train schedule adjustments in more than 25 years, even when train technology changed. Daily routine delays on both Acela and Regional trains were attributed to timetable artifacts as discussed in the results section. Consequently, Amtrak should use the refined data obtained from the automatic vehicle locator to revamp or in the very least, adjust train timetables. Furthermore, Amtrak should plan to fine-tune the train timetables on an annual basis, and especially when new trains sets and train slots are introduced.

Educate and Reinforce On-Time Culture

While third party delays (e.g. delays caused by weather-related issues and police-activity issues) are for the most part unavoidable, Amtrak-responsible delays (e.g. delays caused by passengers boarding and alighting, delays caused by crew lateness, etc.) should continue to be managed and reduced by educating and reinforcing an on-time culture for operating and managing Amtrak crewmembers. Additionally, Amtrak should continue to monitor day-to-day causes of delays in order to reduce and eliminate minor causes of delays, such as late departure from originating station.

Management, Policies and Programs

Strict policies and programs like PRIIA Section 207 are required to help Amtrak meet established goals. The research showed that even though Amtrak owns most of the track in the Northeast Corridor, both Acela and Regional services experienced an unprecedented high in service performance in FY 2011 and FY 2012 while PRIIA Section 207 was active, and both services have been encountering performance deterioration since PRIIA Section 207 was overturned in FY 2013. In addition to statutory laws like PRIIA, other proven techniques like Six Sigma and Lean could be utilized to improve quality output of train operations by identifying and removing the causes of errors and minimizing variability in service operations.

The authors would like to thank NURail and USDOT for sponsoring this research.

References

[1] Federal Railroad Administration (FRA) & Department of Transportation (DOT), (2009). *Metrics and Standards for Intercity Passenger Rail Service*.

[2] The NEC Master Plan Working Group. (2010). *The Northeast Corridor Infrastructure Master Plan*.

[3] Amtrak (2012). The Amtrak Vision for the Northeast Corridor. 2012 Update Report.

[4] Ogunbekun, T. A. (2015). The Impact of Amtrak Performance in the Northeast Corridor (Master's Thesis, Massachusetts Institute of Technology)

---

[i] http://www.amtrak.com/servlet/ContentServer?c=Page&pagename=am%2FLayout&cid=1241245669222

[ii] http://www.amtrak.com/ccurl/412/537/Amtrak-FY2015-Federal-Budget-Request-ATK-14-028,0.pdf

[iii] http://www.nec-commission.com/

[iv] State of the Northeast Corridor Region Transportation System, Cambridge Systematics, Inc. February 2014

[v] Metrics and Standards for Intercity Passenger Rail Service, Department of Transportation, Federal Railroad Authority, 2009

[vi] Amtrak's Twitter feed - https://twitter.com/AmtrakNEC

[vii] Amtrak's Breaking News feed - http://www.breakingnews.com/topic/amtrak/

[viii] Amtrak's Twitter feed - https://twitter.com/AmtrakNEC

[ix] Amtrak's Breaking News feed - http://www.breakingnews.com/topic/amtrak/

[x] Referring to the artic blast across many states in the U.S. that occurred in January 2014

[xi] http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm

[xii] Box & Jenkins, 1976; Velleman & Hoaglin, 1981

[xiii] http://www.amtrak.com/ccurl/412/537/Amtrak-FY2015-Federal-Budget-Request-ATK-14-028,0.pdf

[xiv] http://www.timetables.org/

[xv] http://www.amtrak.com/train-schedules-timetables

[xvi] http://www.amtrak.com/servlet/ContentServer?c=Page&pagename=am%2FLayout&cid=1241245669222

[xvii] http://www.amtrak.com/ccurl/412/537/Amtrak-FY2015-Federal-Budget-Request-ATK-14-028,0.pdf

[xviii] http://www.nec-commission.com/

[xix] http://www.amtrak.com/ccurl/412/537/Amtrak-FY2015-Federal-Budget-Request-ATK-14-028,0.pdf

[xx] http://www.timetables.org/

[xxi]http://www.amtrak.com/train-schedules-timetables

National University Rail Center - NURail
US DOT OST-R Tier 1 University Transportation Center

NURail Project ID:

NURail2012-MIT-R01 High-Speed Rail as a Sociotechnical System

Capacity Challenge on the California High-Speed Rail Shared Corridors: How Local Decisions Have Statewide Impacts

By

Samuel J. Levy

Supervised by Professor Joseph M. Sussman

9/30/2015

Grant Number: DTRT12-G-UTC18

TECHNICAL SUMMARY

Title

Capacity Challenges on the California High-Speed Rail Shared Corridors:

How Local Decisions Have Statewide Impacts

Author: Samuel J. Levy

Introduction

In 2012, as a cost-control measure and in response to local opposition in the San Francisco Bay Area, the California High-Speed Rail Authority (CHSRA) adopted a "blended system" at the north and south bookends of the planned first phase of its high-speed rail line. In this blended operation, the high-speed rail line will share track and other infrastructure with commuter rail, intercity rail, and freight on the 50-mile Peninsula Corridor in Northern California and on 50 miles of right-of-way between Burbank, Los Angeles, and Anaheim in Southern California. This thesis provides a critical review of the blended system and discusses the level of cooperation and coordination necessary between host railroads and the high-speed rail tenant operator.

Approach and Methodology

In Chapter Two of this report, we perform a literature review and discuss the practical challenges railroads face when sharing corridors. In Chapter Three we will "zoom in" on the San Francisco-San Jose Peninsula Corridor and discuss the capacity challenges the railroads are facing today and into the future. In this chapter, we will present the crucial need for coordination between high-speed rail, commuter rail, and freight on the corridor. We also review in detail the concept of a "blended system." In Chapter Four, we will discuss how this coordination is currently progressing and *apply a train-operator financial model to understand the relationship between the agencies on the Peninsula Corridor*. Next, in the Chapter Five, we introduce the Southern California rail network and upcoming challenges with the introduction of high-speed rail in the region. Finally, in Chapter Six we present a California rail network "wish-list;" that is, a list of statewide rail goals. We then evaluate four localized decisions and discuss how the impact of those decisions affect the fruition of the aforementioned "wishes."

Findings

The essential findings of this report are that

- HSR is a necessary component for a long-term plan for intercity transportation in California

-A blended system is the only feasible financial way forward for California HSR.

-However a blended system requires a degree of institutional and physical cooperation and coordination that will be difficult to achieve

-Nonetheless, feasible institutional and physical paths forward do exist and with leadership can be implemented, if all concerned keep their collective eyes on the overall goals for the rail transportation system of California

In the conclusions section, immediately below, we trace the implications of these findings.

Conclusions

Conflicting institutional priorities stand in the way of a unified California rail network

The California rail network is currently fragmented and there is much opportunity for interagency coordination. The Rail2Rail program in Southern California is a positive first step, but that program could be improved through expansion. The Southern California Regional Rail Authority (SCRRA) could work together with the North County Transit District (NCTD) to ensure timed transfers at Oceanside, or better yet, through-run trains to provide commuters options that are more affordable than buying an Amtrak California ticket. Agencies are often protective of their own assets and there is no overseeing agency that implores agencies to work together; as Clever notes in his paper on integration, "By dividing up project planning into separate professional disciplines studying engineering/capital costs, ridership/operating costs, and environmental impacts, sight of the system as a whole is lost. System wide ridership studies are completed without knowing the exact station locations or the level of integration with other modes" (Clever 12). This individualistic agency mindset will not work in a blended system in the future.

Northern California is further along in the planning process than Southern California

While the CHSRA announced the change to a blended system for both Northern and Southern California at the same time, the southern blended system lags the northern counterpart in definition. Granted, the Southern California network is much more complex than its northern counterpart; but since the Los Angeles Basin will see high-speed rail service six years before the Bay Area does, it is surprising that the north leads the south in this regard. Perhaps this is due to higher levels of political pressure in Northern California, but the fact remains that there is no blended operations analysis or significant memorandum of understanding signed between SCRRA, the CHSRA, Amtrak California, or the Class I freight railroads. In Northern California, this early planning has been beneficial because it has brought to the forefront questions such as platform compatibility and ultimate track layout. However, the current fuzziness of the Southern California system is also an opportunity, since all options are still available for consideration.

Shared corridor challenges most important to California

Congestion and Delays

The blended system will put high demands on the existing infrastructure. Commuter rail or freight delays can propagate to the HSR system. As the ratio of service volume to capacity tends towards 1, the system loses stability and on-time performance suffers. With degraded on-time performance and uncertainty regarding arrival times, schedule padding becomes necessary and makes rail as a mode less attractive to time-sensitive consumers.

The Caltrain corridor today experiences delays due to two main factors: aging equipment malfunctions and on-track suicides[1]. While some of the oldest equipment will be retired once the electric equipment arrives in 2020, Caltrain plans to continue operating some diesel-electric equipment concurrently with the electric vehicles. Aside from the two different fleet types making maintenance more time-consuming, this could lead to delays when aging equipment breaks down and a specific fleet type needs to be substituted. And since grade crossings will not be eliminated, on-track suicides will continue to cause significant delays, though ongoing prevention programs could help minimize this impact.

CPUC Requirements

There are two major California Public Utility commission regulations that stand in the way of an integrated rail system. The first is a regulatory framework for high-voltage operation on shared corridors. The CHSRA developed a framework for high-voltage operations on its dedicated corridors, but operations in blended corridors with freight and grade crossings have yet to be addressed. If this fails to be resolved, the CHSRA would be forced to either build dedicated right-of-way as originally planned or truncate its operations in Burbank and San Jose and forced travelers to transfer to commuter rail systems.

Overhead catenary wire presents clearance issues for freight railroads in Southern California. Freight railroads cannot operate in certain tunnels with double-stack containers if overhead catenary wires are not high enough to provide adequate clearance. It will also be challenging to construct the improvements necessary for the high-speed rail system on the BNSF-owned right of way in Southern California as freight railroads demand high levels of track availability.

The second CPUC requirement is the freight train lateral clearance requirement. This standard requires adequate lateral clearance for freight trains at platforms. The platforms of the SPRINTER light rail system in San Diego County are evidence of this requirement in action. The SPRINTER shares track with freight

---

[1] Though the suicide rate fluctuates from year-to-year, Caltrain claims it has the highest suicide rate in the nation (Brotherhood of Locomotive Engineers and Trainmen 2003)

railroads but also uses high platforms. To meet the ADA maximum 3” gap requirements and satisfy the CPUC minimum gap requirements, the SPRINTER employs mechanical platform “gangways” that descend during light-rail operation during the day, and retract at night to allow for freight trains to pass. The CHSRA has opted for separate stations to avoid this CPUC requirement. Using the same stations as the commuter rail lines would mean that the CHSRA must find a compatible platform height and urge for CPUC rule changes. Freight railroads will likely contest rule changes because it will limit their operational flexibility to haul wide loads. However, the cost savings of not having to build HSR-specific stations are substantial and the rule changes are something the author believes are worth working hard to achieve.

Track Ownership and Priority

The blended system dictates that the CHSRA will operate on a multi-owner network, much like Amtrak does today. Instead of having sole control of its infrastructure, the CHSRA will have to work with the TJPA, PCJPB, SCRRA, and BNSF Railway to ensure smooth operation. As a result, the CHSRA will face many of the issues Amtrak faces today regarding train priority. The ability for the CHSRA to operate in a reliable fashion will depend on the priority rules that the CHSRA can negotiate with its host railroads on the blended corridors.

The blended service on the Peninsula as it stands today is infeasible

While neither agency would describe their relationship as a competitive one, the HSR operator and Caltrain will be competing for track access and access to the downtown San Francisco terminal. The blended operations analysis, though PCJPB emphasizes that it is a feasibility study, not a service plan, shows that Caltrain will likely have limited access to its most important station and that the HSR operations create uneven headways on the corridor. Furthermore, without passing track construction on the mid-peninsula (through the very same communities that fought against the four-track corridor), CHSRA can only operate two trains per hour, per direction into the terminal. The separate stations for high-speed rail up and down the Peninsula will make it difficult for passengers to “interline” or use both systems as envisioned in the 2012 CHSRA Business Plan describing blended service. A renewed dedication to blended operations is beginning to form with the current discussion on Caltrain’s new electric equipment; the two agencies need to keep this momentum while moving forward towards a true shared system.

The blended system as planned does not match the former aspirations of Caltrain or HSR

The blended system was a significant scale-back of the high-speed rail system sold to voters in 2008, and it is evident that both the service plans of Caltrain and the CHSRA are nowhere nearly as customer-friendly as they were before 2012. Caltrain has shared aspirations of 10 trains per hour, per direction on the corridor, while the high-speed rail authority has shown service frequency as high as eight trains per hour per direction The LTK engineering study reported that six Caltrain trains per hour was feasible and four HSR trains was possible only with significant construction. Due to the importance of Caltrain to everyday commuters, significant construction of passing tracks is exceptionally challenging: the Ponderosa Project to build the passing tracks for Baby Bullet service required weekend shutdowns of the Caltrain system for two years This means that once a high-speed operator is running revenue service along with increased Caltrain service, performing track construction to expand capacity on the line or in the Transbay Transit Center will be more challenging than ever before.

The CHSRA asserts that they are still meeting the bond measure requirements that were approved by voters in 2008. However, the language of the measure suggests a very different system than what is on the table today. When the voter read the text "Achievable operating headway (time between successive trains) shall be five minutes or less" it would have been reasonable for the voter to assume that service levels would be somewhat higher than two high-speed trains per hour (30 minute headway) into San Francisco.

The impact of the phased approach has consequences that go beyond the phase

Truncating the initial operating segment in the San Fernando Valley for six years is a hugely important decision, especially if the CHSRA chooses to operate in separate station facilities than Metrolink. Metrolink service will need to match high-speed rail trains with much higher off-peak service levels than today to ensure a timed transfer. Even so, a smooth transfer will be very difficult because Metrolink will be at a separate platform, meaning that without costly facilities, baggage transfer will be difficult. Because transit connections are poor in the San Fernando Valley, the high-speed operator will demand high levels of parking to enable access for residents across the Southland. Building these parking facilities for a temporary terminal will impact land-use around the station site; the positive impacts of transit-oriented development, a key driver in ridership, will not be realized. This could affect long term ridership and the ability of the high-speed rail operator to generate revenues to cover operating costs.

Freight Impacts

It could be argued that a strong rail freight network is just as important to California's economy and congestion mitigation as a strong passenger one. The freight railroads provide a key role in transcontinental goods movement and one freight train removes hundreds of trucks from our nation's highways. It is for this reason that the freight railroads cannot be "pushed aside" for passenger service. On the Peninsula Corridor, freight will seek to continue operation on the corridor; however, increased corridor use from high-speed rail will narrow operating windows for freight service and finding time for track maintenance will be more difficult than it is today. In Southern California, BNSF will likely require high-speed rail to expand capacity on BSNF's San Bernardino subdivision between Los Angeles and Fullerton, CA. Negotiations for capacity with freight railroads can become contentious when freight railroads feel that they are losing flexibility to run trains efficiently through their network when necessary. However, if passenger rail can use freight infrastructure in a way that does not impede freight trains from operating efficiently, freight railroads would stand to benefit financially from increased access charges.

Recommendations

Agency-specific

California High-Speed Rail Authority

Procure low-floor vehicles if manufacturers can supply them

The CHSRA is currently committed to purchasing service-proven technology. There are no high-speed trains currently in service that are 1) capable of 220mph operation and 2) a low-floor vehicle. If the commuter rail agencies are to pursue level boarding and compatibility with high-speed rail, having a low-floor vehicle has the potential to reduce cost across the network. This is because Caltrain and Metrolink currently operate the same Bombardier equipment and could simply raise platforms to accommodate their existing equipment[2,3]. If funding for level boarding falls short, it could be implemented at individual stations that would stand to benefit the most from its implementation. Even if Caltrain and Metrolink never raise platforms, high-speed rail could operate at commuter rail stations (without level boarding). This will save Caltrain the expense of procuring specialized electric vehicles with two sets of doors that match both existing and future platform heights.

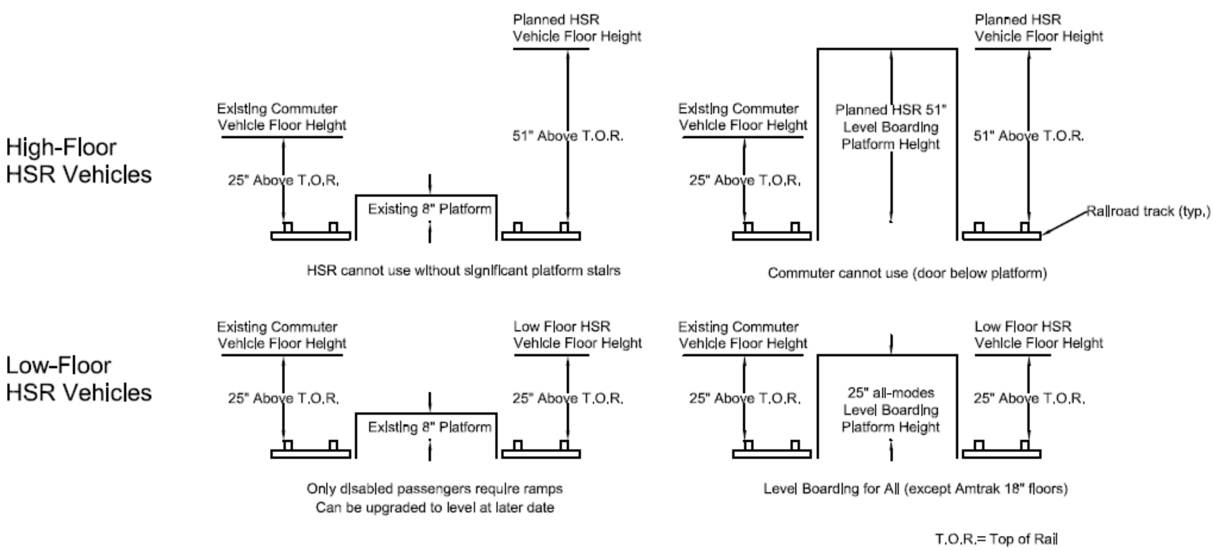


Figure 0-1: Low floor HSR vehicles would allow for shared infrastructure and allow for gradual transition of existing commuter rail platforms to level boarding. Amtrak California's 18" floors would require car reconstruction or a step system for higher platforms
SOURCE: Diagram by Author, Floor heights from PCJPB,

There are clear risks for the CHSRA to try and conform to existing California vehicle floor heights as opposed to existing HSR vehicle floor heights. Vehicle costs of a new technology could be much higher due to reduced competition of vehicle suppliers. Lack of service-proven technology means that California would be the

---

2 Metrolink's new Hyundai Rotem equipment that is replacing the Bombardier equipment has the same 25" floor height

3 The *FrontRunner* commuter rail service in the Salt Lake City, Utah area currently operates the same Bombardier coaches and has level boarding.

“guinea pig;” that is, they would be the first to experience any issues with reliability and safety of new equipment.

The Parsons Brinckerhoff technical memorandum advocating for high platforms for high-speed rail was released in 2009 when the plan for HSR service was to operate in a completely dedicated right-of-way. Since that time, the blended system demands a fresh approach and vision for the entire system and how it will integrate with existing rail services. By adopting a low-floor vehicle specification for its trainsets, the CHSRA will show its agency partners that it is committed to the success of blended operations and encourage them to follow suit and develop a similar attitude. Some risks like new low-floor high-speed rail vehicles would be worth taking.

Avoid terminating service temporarily in Burbank

The CHRSA’s plan of terminating service from 2022 to 2028 before extending service to Anaheim and Los Angeles poses a risk. We show in Chapter Five that the transferring to conventional diesel service from Burbank to Los Angeles and Anaheim will add a non-trivial amount of time in addition to a perceived “transfer penalty.” There are also associated costs with operating a temporary terminal station. Maintenance and baggage transfer facilities will be necessary; and because a Burbank/San Fernando Valley station has relatively minimal transit connectivity, the station will require a lot of additional parking for high-speed rail customers. The relative lack of access from the southern portion of the Los Angeles Basin might initially crimp ridership and revenue. Either a private investor or the CHSRA (i.e. the State) itself will have to fund the cost of additional time for ridership and revenue to materialize after the section to Anaheim and Los Angeles is finished (presumably in 2028).

The CHSRA will have project acceleration costs if it chooses to build the initial operating segment system to downtown Los Angeles. The additional cost will have to be borne by the Authority and not a private investor, since the Authority claims that investment is contingent upon successful revenue service. At the time of writing, the Authority has only $12.5 billion in committed funding (in addition to carbon tax revenues, which were $250 million last year). This means the CHSRA only currently has 40% of the Phase 1 cost, estimated at $31 billion. Pursuing more public funding on top of the $31 billion to finish the railroad to Los Angeles might be political infeasible. However, if the CHSRA determines it is possible, it should prioritize completing the connection into Los Angeles Union Station.

Formulate a coherent blended service strategy at both ends of the line

In order to reduce the risk for the private operator that will ultimately run the system, the CHSRA needs to understand exactly what blended will entail. Will there be a Metrolink train meeting every HSR train in Burbank and will that train continue on to Anaheim from Los Angeles? Will Caltrain customers be able to use HSR trains instead of Caltrain trains between San Jose and San Francisco? How many HSR trains per hour will the high-speed operator be allowed to run during the peak hour into downtown San Francisco? The CHSRA needs to resolve these questions before making infrastructure decisions such as passing tracks in Northern California or a temporary Burbank Terminal, and definitely before it tries to seek a private operator to invest in the capital construction costs.

Peninsula Corridor Joint Powers Authority

Consider the impacts of competing for access with a statewide HSR operator

The Peninsula Corridor Joint Powers Board owns the right-of-way and track between San Francisco's 4th and King Station and San Jose and the Transbay Joint Powers Authority will own 1.3-mile Downtown Extension into the Transbay Transit Center. This seemingly puts the PCJPB/Caltrain on "home turf" when it comes to access negotiations since they represent San Francisco's 150-year old commuter railroad. However, the CHSRA serves a much larger, statewide constituency, than the three-counties that enjoy Caltrain service. Furthermore, the PCJPB structure is a weak because it requires voluntary contributions from the three counties; it has no dedicated funding source. The PCJPB and CHSRA currently have a healthy partnership, but ultimately, Caltrain will be working in partnership with the private HSR operator, not the CHSRA. Since the operator will be facing revenue and ridership pressure, and since Caltrain operates at their peak service levels at the most valuable times for high-speed rail, this relationship could potentially become more adversarial. In order to avoid this, the PCJPB and CHSRA need to carefully negotiate access, or at a minimum, agree on a method for objectively allocating and pricing capacity.

Grow stronger working relationship with Southern California

Whether or not Metrolink or Amtrak California decide to electrify operations on the blended corridor in Southern California, they will encounter many of the same issues faced by the Caltrain in the north. For example, there will be negotiations regarding service levels and access to railroad assets such as the run-though tracks at Los Angeles Union Station. As owners of tracks and right-of-way that will see frequent use from high-speed rail, both the Caltrain and Metrolink stand to benefit by sharing effort in developing the ultimate form of the blended system. Both agencies have overlapping interests: they want to protect rail service for the everyday commuter, they want to maximize the benefit of HSR connectivity funding, and they want to keep positive relationships with the freight railroads and other passenger railroads with whom they share infrastructure. Man-hours could be saved by working through HSR compatibility challenges together.

Service plan as soon as possible (try examples of better integration) and develop metrics to evaluate output timetable

In their blended service planning document, Caltrain writes that their next step is to look at service plan potions prior to fleet need. However, Caltrain is already beginning the procurement process for their vehicles. Service planning needs to come first and Caltrain needs to develop these service plans with the CHSRA. Planning should include multiple levels of integration, from shared platforms all the way to shared trainsets so the two agencies can understand the value of integration investments. For example, some HSR trains could replicate Caltrain's Baby Bullet service and stop a five or six intermediate stations between San Francisco and San Jose instead of two. Even though HSR trains acting as "Baby Bullets" making limited stops would hurt the utility of the high-speed service, it is the superior alternative to truncating service in San Jose and forcing HSR travelers to transfer to commuter rail. As LTK concluded, the capacity is simply not there for greater than two HSR trains per hour and Caltrain Baby Bullet service.

To evaluate the timetable from a customer perspective, the CHSRA and Caltrain should develop metrics which factor customer travel time, headway uniformity, and passenger residential and job data among other measures. Caltrain's origin-destination data is not reflective of consumers' preferred stations; rather, it is reflective of a Baby Bullet service which pulled riders to high-service stations. Timetable metrics would allow the two agencies to objectively compare one service plan to another and quantify the benefits of one timetable that may require the construction of passing tracks versus another.

Address the freight issues—CPUC horizontal clearance and operating windows

The freight issues are not going to disappear any time soon—the PCJPB has affirmed through its statements and in the memorandum of understanding with the high-speed rail authority that freight is going to remain on the Peninsula Corridor. By not addressing the freight issues, the PCJPB is implicitly encouraging the CHSRA to use separate facilities since the waiver is a necessary condition for Caltrain to have level boarding, but not one for HSR if it believes that separate facilities are sufficient for its own operation.

Currently Union Pacific is guaranteed one 30-minute window between 10:00 A.M. and 3 P.M. each day to run freight trains on each of the northbound and southbound tracks of the Peninsula Corridor. Between midnight and 5:00 A.M., one main track is reserved for freight use. With increased service due to HSR, these window capacity allocations may need to be revisited. As owner of the right of way, it is the PCJPB's responsibility to manage these operating window constraints before these issues affect Caltrain and HSR's service commitments.

Transbay Joint Powers Authority

Consider impact of reduced Caltrain service to Transbay Transit Center

Limited Caltrain service to the Transbay Transit Center will be as painful to the developers and the Transbay Transit Center as it will to the commuters themselves. Being able to boast a frequent one-seat ride to the points south on the Peninsula will raise rents for the property owners and developers as well as revenues for retail spaces, and as a result, revenue for the City of San Francisco. It is in the best interest of the developers at Transbay Transit Center to ensure that Caltrain has adequate access to the terminal.

Ensure there is a path for improvements to the Downtown Extension

One of the primary concerns raised in the environmental impact report for the Transbay Transit Center was the small size of the six track terminal. For a terminal station, six tracks is minimal—Los Angeles Union Station has 14 tracks and Penn Station in New York City has 21. In the 2002 Environmental Impact Report/Environmental Impact Statement, the Transbay Joint Powers Authority wrote that "additional turnaround or "tail" tracks will greatly assist in relieving congestion at the platform tracks." In 2008, the Transbay JPA announced that the "tail tracks would be deferred until operationally required." This means that the TJPA anticipates a need for track capacity, but will seek funding sources later. This raises interesting questions about who will be responsible for increasing capacity and who will fund that additional capacity. Will one operator be forced to choose between a large capital expenditure and service reductions, and if so, which operator? Because the TJPA is a separate entity, neither the CHSRA nor Caltrain has a "right" to access the vital terminal. If the CHSRA is required to fund capacity improvements, should the cost fall on the private operator or the public? The Transbay Transit Center should answer these questions sooner rather than later to help both the Caltrain and the CHSRA plan for service and future expansions of service.

Rail Agencies in Southern California

Quantify to the greatest extent possible, the benefits and costs of level boarding, platform sharing, electrification, and increased integration in terms of increased ridership and connectivity

The Southern California Regional Rail Authority, Amtrak California (the LOSSAN JPA), and even NCTD should, along with the CHSRA, evaluate the benefits and costs of level boarding, platform sharing, electrification, and increased integration across the greater Los Angeles area. Level boarding and platform sharing clearly provide a benefit for transfers to-and-from high-speed rail and allow for maximum use of limited capacity at Union Station, but the benefits of level boarding outside the blended corridor are markedly less. Increased integration could result in benefits such as a one-seat ride between Los Angeles and San Diego or Las Vegas, and could help commuter rail serve as a better first-mile and last-mile connection for high-speed intercity travelers; that is, both modes will see ridership gains with an integrated system. If "Blended Service" means integration between HSR and commuter rail operators, can Southern California operators integrate non-HSR service as well? Perhaps cooperation with the high-speed rail authority can help foster increased cooperation between Metrolink, Amtrak California, and COASTER.

Protect the run-through tracks at Los Angeles Union Station

In the author's opinion, the run-through tracks at Union Station will be a hugely important asset. In terms of cost per minute of travel time saved for through trains, the Southern California Interregional Connector Project is one of the most cost-effective rail projects under construction in the state. Unfortunately, only four tracks of the 14 at Union Station are going to "run-though" and it is likely that high-speed rail will need to use at least two of them. These tracks are going to be more valuable for regional commuter services as well: overnight, Metrolink will gain much more flexibility with service planning and timed-transfers and Amtrak California can now offer customers travelling from the northern half of the LOSSAN corridor to the southern half nearly 15 minutes off of their trip. It is important that the return on this $350 million capacity investment is maximized; CHSRA should demonstrate their plan to achieve high levels of track utilization before LACMTA (the owner of Union Station) allows high-speed rail to monopolize half of the run-though track improvements. However, it would be ideal to maintain all four tracks for general use through a consistent low-floor platform height.

General Recommendations

Integrated operations have a potential to bring HSR a revenue source; like regional airlines feed into international hubs, so too can commuter rail services feed into interregional high-speed rail services. To that end, service planning should drive infrastructure decisions; in an era where infrastructure such as new HSR stations or electrification are expensive and the public is leery of megaprojects, California can set an example with a well-conceived (i.e. well service planned), integrated rail system.

This integration is important from a risk management perspective as well. The CHSRA is fortunate to have their lead cheerleader in the statehouse: Governor Jerry Brown is the by far the most important political ally that the Authority has at its disposal. Unfortunately for the CHSRA, Governor Brown concludes his term in 2018 and it is unlikely that his successor—whether Democrat or Republican—will be such an ardent supporter of the project. At the time of writing, there are court cases that threaten the CHSRA's ability to use Proposition 1A funding on the grounds that the current system is not what was promised to voters. If the funding were to evaporate, it would be much better to be left with a useable system than a standalone section of high-speed track in the Central Value with little independent utility. The CHSRA was right to start construction in the Valley as they are filling key a rail gap between the Central Valley and the Los Angeles Basin; however, if this piece does not fit into the rest of the California network; this will truly be a "train to nowhere" as project critics attest.

Publications

Capacity Challenges on the California High-Speed Rail Shared Corridors: How Local Decisions Have Statewide Impacts (attached)

Samuel Levy and Joseph M. Sussman

Challenges and Opportunities in Implementation of the Future California Rail Network (TRB, 2016)

A. Awadagin Faulkner, Samuel Levy, and Joseph M. Sussman

Analyzing the Financial Relationship between Railway Industry Players in Shared Railway Systems: The Train Operator's Perspective (TRB, 2015)

Samuel Levy, Maite Pena-Alcaraz, Aleksandr Prodan, and Joseph M. Sussman

Primary Contact

Professor Joseph M. Sussman
sussman@mit.edu
617.253.4430

Principal Investigators

Name: Samuel J. Levy, Joseph M. Sussman

Title: **Capacity Challenges on the California High-Speed Rail Shared Corridors: How Local Decisions Have Statewide Impacts**

Department: Civil and Environmental Engineering

University: Massachusetts Institute of Technology

Samuel J. Levy
samlevy90@gmail.com

Joseph M. Sussman
sussman@mit.edu

Other Faculty and Students Involved

Name: Professor Joseph M. Sussman

Title: JR East Professor of Civil and Environmental Engineering and Engineering Systems

Department: Civil and Environmental Engineering

University: Massachusetts Institute of Technology

sussman@mit.edu
617.253.4430

NURail Center
217-244-4999
nurail@illinois.edu
http://www.nurailcenter.org/

Capacity Challenges on the California High-Speed Rail Shared Corridors:
How Local Decisions Have Statewide Impacts

by

Samuel J. Levy

B.S. Civil Engineering, Minor in Economics
University of Southern California, 2013

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

Massachusetts Institute of Technology

June 2015

Signature of Author: ……………………………………………………………………………………

Department of Civil and Environmental Engineering

May 19, 2015

Certified by: ……………………………………………………………………………………

Dr. Joseph M. Sussman
JR East Professor of Civil and Environmental Engineering and Engineering Systems
Thesis Supervisor

Accepted by: ……………………………………………………………………………………

Heidi Nepf

Donald and Martha Harleman Professor of Civil and Environmental Engineering

Chair, Graduate Program Committee

Capacity Challenges on the California High-Speed Rail Shared Corridors:

How Local Decisions Have Statewide Impacts

by

Samuel J. Levy

Submitted to the Department of Civil and Environmental Engineering

On May 19, 2015 in Partial Fulfillment of the

Requirements for the Degree of Master of Science in Transportation

Abstract

In 2012, as a cost-control measure and in response to local opposition in the San Francisco Bay Area, the California High-Speed Rail Authority (CHSRA) adopted a "blended system" at the north and south bookends of the planned first phase of its high-speed rail line. In this blended operation, the high-speed rail line will share track and other infrastructure with commuter rail, intercity rail, and freight on the 50-mile Peninsula Corridor in Northern California and on 50 miles of right-of-way between Burbank, Los Angeles, and Anaheim in Southern California. This thesis provides a critical review of the blended system and discusses the level of cooperation and coordination necessary between host railroads and the high-speed rail tenant operator.

In Northern California, the Peninsula Corridor Joint Powers Board's *Caltrain* commuter rail service between San Francisco and San Jose is experiencing record levels of ridership. This thesis explores the impact of both the electrification of the line and its extension into San Francisco's central business district on future ridership demand. With the California High-Speed Rail Authority competing spatially and temporally with Caltrain for access to high-revenue and high-cost infrastructure, we review different strategies for coordination and integration between the two agencies.

In Southern California, the final form of the blended system is more nebulous than its northern counterpart. For the first few years of high-speed rail service, the *Metrolink* service operated by the Southern California Regional Rail Authority is expected to complement the high-speed rail system. However, since Metrolink operates on congested rail infrastructure, some of it owned by capacity-conscious freight railroads, there will exist the challenge of providing quality service and transfer opportunities for time-sensitive high-speed rail customers.

The change to a blended system was a dramatic change of direction for the CHSRA; as a result, a new paradigm is needed for implementation of the system over the next 15 years. This thesis reviews the upcoming local design choices to be made on the local rail corridors and evaluates them from the perspective of the future statewide rail network. We find that the decisions made on the local blended corridor level will affect both the financial viability of the overall project and the quality of service experienced by customers across the entire California rail system.

Thesis Supervisor: Joseph M. Sussman
Title: JR East Professor of Civil and Environmental Engineering and Engineering Systems

Acknowledgements

I want to thank everyone who helped make this thesis possible. I appreciated all of the encouragement I received along the way from my friends here at MIT and those at home in California.

To Ulrich Leister, thank you for talking through many of the ideas presented in this thesis and challenging me to perform an original analysis.

To Dave Schlesinger, thank you for introducing me to valuable contacts from across the rail industry in California and doing a fantastic industry proofread of my draft thesis. Fight On, Dave!

To Maite and Alex, thank you for being great collaborators on my first published paper and for offering sound advice and support throughout my two years at MIT.

Ryan, Joel, Iori, Soshi, Bruno, Rebecca, Joanna, and Patton: I will miss spending time with you and I will miss our conversations in 1-151. It has been a pleasure working with all of you and I wish you good fortune in your bright careers.

This thesis would have been impossible without all of the help I received from industry. Drew Cooper from the San Francisco County Transportation Authority, Mel Corbett from Metro-North Railroad, Roderick Diaz from the Southern California Regional Rail Authority, Paul Dyson from the Rail Passengers Association of California and Nevada, David Hunt from Oliver Wyman, Martha Martinez and Mark Simon from the Peninsula Corridor Joint Powers Board, Andy Nash from Green City Streets, Clem Tillier from the Caltrain-HSR Compatibility Blog, Frank Vacca from the California High-Speed Rail Authority, and Dominic Spaethling from HNTB all made considerable effort to help get me the answers I was looking for. Mary Seerveld provided the fantastic graphics in the sixth chapter and Courtney Albini provided some excellent copy editing as well.

Thank you to Dagin Faulkner, my undergraduate research assistant, who helped map out the complex ecosystem of California passenger rail among many other very important tasks that were necessary in completing this thesis. You are an all-star and I hope you will always keep your enthusiasm for all things transportation.

And to my advisor, Joe, thank you for and inviting me to join your research group. Your encouragement and enthusiasm helped me understand the value of my work. I will miss our conversations about baseball and your undying surprise at which weather I considered “shorts-appropriate.” You are what makes MIT a special place for so many of us.

Finally, I would like to thank Veronica and my family for being a great support system and sending me lots of candy and warm clothing during my time here in Cambridge. My love goes to all of you.

Sam Levy
Cambridge, MA
May 19, 2015

Contents

List of Abbreviations

ACE—Altamont Commuter Express

BART—Bay Area Rapid Transit

BNSF—Burlington Northern and Santa Fe Railway

CHSRA—California High-Speed Rail Authority

CS—Cambridge Systematics

HSR—High-speed Rail

IOS—Initial Operating Segment (for high-speed rail)

LACMTA/Metro—Los Angeles County Metropolitan Transportation Agency

LAUS—Los Angeles Union Station

LOSSAN—Los Angeles-San Luis Obispo-San Diego

LTK—Louis T. Klauder Engineering Services

NCTD—North County Transit District

PB—Parsons Brinckerhoff

PCJPB—Peninsula Corridor Joint Powers Board

RailPAC—Rail Passengers Association of California and Nevada

SCAG—Southern California Association of Governments

SCRIP—Southern California Regional Interconnector Project

SCRRA—Southern California Regional Rail Authority

SFMTA—San Francisco Municipal Transportation Agency

SJRRC—San Joaquin Regional Rail Authority

TJPA—Transbay Joint Powers Authority

UP—Union Pacific Railroad

VCRR—Ventura County Railroad

VTA—(Santa Clara) Valley Transportation Agency

1 Introduction and Motivation

1.1 Problem Statement

In 2012, as a cost-control measure and a response to local opposition in the San Francisco Bay Area, the California High-Speed Rail Authority (CHSRA) adopted a "blended system" at the bookends of the state's planned high-speed rail line (CHSRA 2012). In this blended operation, the high-speed rail line will share track and other infrastructure with commuter rail, intercity rail, and freight on the 50-mile Peninsula Corridor in Northern California and on 50 miles of track between Burbank, Los Angeles, and Anaheim in Southern California (ibid). This change to "blended" from "dedicated" reflects the truth that the costs and challenges associated with constructing new, dedicated rail infrastructure are enormous, especially in urban areas. Shared rail corridors represent the possibility of more efficient use, that is, higher utilization, of precious rail infrastructure. Multiple railroads can share the burden of track maintenance and traffic control, both of which require high fixed costs. Sharing track, when done properly, is an attractive option for both passenger rail agencies and freight railroads and increasingly common in the United States. However, sharing track comes with challenges for all participating railroad operators as well.

1.2 The Importance of Capacity Management on Shared Rail Corridors

Of course, sharing track requires coordination and more often than not, it is between non-homogenous rail traffic. Rail capacity is not a fixed quantity; it depends on how the infrastructure is used. In this thesis, we discuss freight traffic (typically few stops, slow speeds), commuter rail traffic (many stops, medium speeds), and high-speed traffic (few stops, high speeds). A rail line can accommodate much more homogenous traffic (same speed and stopping patterns) than a rail line with heterogeneous traffic. In the next section, we will discuss the growth in demand for *all* types of railway traffic in the United States.

1.2.1 Increasing Demand for Passenger and Freight Service

In the United States, managing rail capacity is increasingly important as demand for rail infrastructure grows. This growth arises from increases in both passenger and freight traffic over the last several decades. On the supply-side, capacity improvements have not kept pace. The combination of these two developments means that railroads—both freight and passenger—are increasingly competing for rail capacity on corridors across the United States. California, which is the focus of this thesis, will be at the forefront of addressing this challenge.

Freight Rail

In 2012, freight rail carried 40% (measured by ton-miles) and 16% (measured by tons) of America's

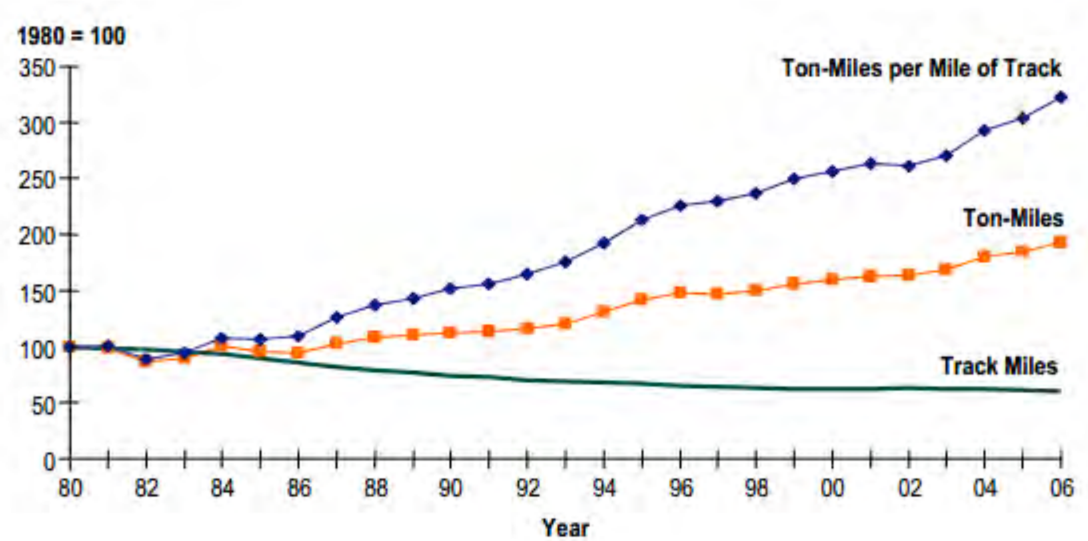


Figure 1-1: Changes in rail freight traffic since the 1980 Staggers Act
Source: Cambridge Systematics

intercity cargo (Grunwald 2012). The $65 billion US freight rail industry has posted strong growth in revenue and efficiency since the US Congress passed the Staggers Act in 1980 which eliminated regulation on shipping rates and contracts (Palley 2011). The industry has consolidated to seven Class I railroads (systems with an annual operating revenue of $433.2 million or more) from 14 in 1990; these railroads account for more than 90% of industry revenue—California freight rail is dominated by two Class I railroads, the Union Pacific Railroad (UP) and the Burlington Northern Santa Fe Railway (BNSF), both of which had operating revenues of approximately $22 billion in 2013 (Grenzenback et al. 2007, Knight 2013, BNSF 2013). While rates have climbed with rising fuel prices in the last half-decade, freight rates adjusted for inflation were the same in 1981 as they were in 2011 (Lawrence 2015). Most importantly for this thesis, Class I railroads have realized an 88% growth in train-mile traffic in that period while shrinking their networks on a track-mile basis by 42% (Grenzenback et al 2007). While some of this efficiency can be attributed to longer train lengths, railroads are seeing their networks carry higher and higher amounts of traffic (Connell 2010). According to FRA data (as seen in Figure 1-1), ton-miles of freight doubled between 1980 and 2006 while traffic density (measured in ton-miles per mile of track) has tripled (Grenzenback et al 2007). Forecast rail traffic growth between 2006 and 2035 is projected to grow another 88% (ibid).

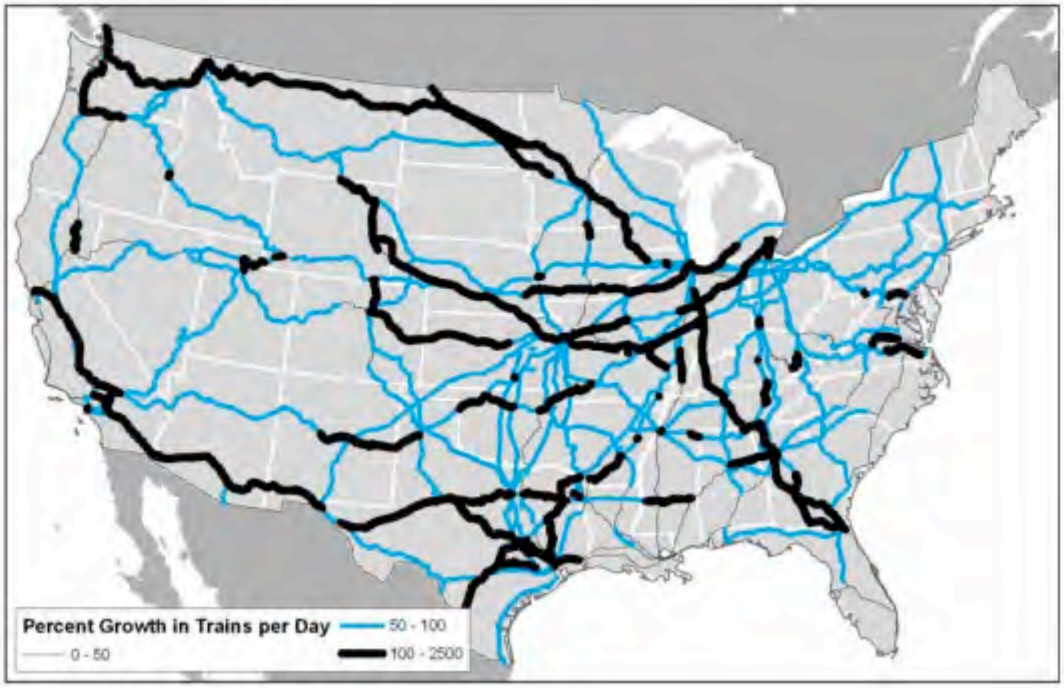


Figure 1-2: Percentage growth in trains per day from 2005 to 2035 by Primary Rail Corridor
Source: Cambridge Systematics

Freight rail traffic will continue to grow because of the relative affordability of rail transportation, especially for longer trips. It takes less time to ship from Asia to ports in California and ship cross country than to ship from Asia to the East Coast of the US; for this reason, much of this growth with be concentrated around California's Los Angeles-Long Beach port complex (and to a lesser extent, the Port of Oakland). Some of California's busiest freight subdivisions see traffic volumes as high as 64 trains per day[4].

Passenger Rail

We are seeing increased demand on the passenger side as well. Between 2000 and 2010, both unlinked commuter rail passenger trips and commuter rail passenger miles have increased 66% while total vehicle miles has nearly doubled (Neff and Dickens 2012). The reason behind this growth has been attributed, in part, to rising fuel prices which has pushed commuters from automobiles to more affordable public transportation alternatives[5] (Haire and Machemehl 2007). Aside from commuter rail, intercity passenger rail, operated by Amtrak and state agencies, has increased from roughly 21 million passenger miles in 2000 to 31.2 million in 2012 (Neff and Dickens 2013). In short, this means that track occupancy time and capacity usage by commuter rail has increased substantially.

---

[4] Recent expansions in the Panama Canal have been matched by expansions at California's main ports in Los Angeles and Long Beach and should counter potential Panama Canal-based market share loss.

[5] Higher traffic congestion, the opportunity to be productive in-vehicle, and increased quality-of-life as well as many other factors have contributed to transit ridership growth

A Brief Introduction to California Geography

Before we proceed further, we feel it is necessary to introduce the reader to the general geography of California. At nearly 40 million residents, California is the most populous state in the United States and on its own, the world's eighth largest economy. However, population is concentrated in several urban areas with lots of open terrain in between. The state is often divided both politically and culturally into two regions: Northern California and Southern California. San Francisco (837,000 people) and the surrounding San Francisco Bay Area serves as the cultural capital of the north and Los Angeles (3,840,000 people), the largest city in the state and second largest in the nation, as the hub of the south (Census Bureau 2015)[6]. San Diego, the state's second largest city (1,356,000 people with 3 million in its metropolitan area) sits adjacent to the Mexican border at the southwestern corner of the state (ibid). The capital city of Sacramento (475,000 people) serves as the anchor of the San Joaquin and Sacramento River Valleys, collectively referred to as the Central Valley and a vital farming region that accounts of 8% of the United States' agricultural output (ibid). While the state population currently sits at approximately 40 million persons, the state is expected to add 10 million more by 2050 with growth concentrated in the Central Valley (CA Department of Finance 2014).

Figure 1-3: Topographic Map of California. Source: worldatlas.com

---

[6] The San Francisco-Oakland-San Jose region combined statistical area has a population of approximately 7.5 million (about 20% of California's total population); the Los Angeles metropolitan area has a population of approximately 13 million (about 35%)

Following the same trend as the rest of the nation, both intercity and passenger rail have increased their ridership in California in the past 20 years without any significant increases to their network size. The intercity routes operated by Amtrak California, in particular, have increased service to the tune of nearly 50% more annual train miles on a route network where Amtrak is always a tenant railroad on another host railroad's (freight or government agency-owned) network[7]. The *Capitol Corridor*, an intercity rail service between northern Sacramento, Oakland, and San Jose has nearly doubled its services. Fortunately, the Capital Corridor's route is almost entirely on a single railroad's network (Union Pacific) thereby simplifying capacity negotiations. The *Pacific Surfliner (*or *Surfliner)*, one of Southern California's many responses to regional automobile traffic congestion, has increased its annual train mileage 26% between 2000 and 2012. Unlike the *Capitol Corridor* which operates predominantly on Union Pacific tracks, however, the *Pacific Surfliner* uses tracks owned by multiple railroads—both public and private—so train scheduling and use of capacity is a much more complex challenge.

Figure 1-4: California Passenger Rail Network. Note the "rail gap" between Bakersfield and Greater Los Angeles
Source: Rail Passenger Association of California (RailPAC)

[7] A "tenant" railroad is a guest on a "host" railroad's infrastructure. The host owns the track and the tenant accesses through "trackage rights," though the tenant may be responsible for track maintenance, signals, etc.

Route Name	Pacific Surfliner	San Joaquin	Capitol Corridor	All Lines
Network Owners	*SCRRA, UP, NTCD, VCRR*	*UP,BNSF*	*UP, PCPJB*	
Passenger (Pax) Miles				
2000	1,661,704	733,152	1,030,837	3,425,693
2012	2,664,935	1,133,654	1,770,616	5,569,205
Percent Change	*60%*	*55%*	*72%*	63%
Pax Miles/Train Mile				
2000	106.2	103.7	106.0	105.6
2012	135.4	124.0	93.7	116.7
Percent Change	*27%*	*20%*	*-12%*	11%
Train Miles				
2000	15,647	7,070	9,725	32,442
2012	19,682	9,142	18,897	47,721
Percent Change	*26%*	*29%*	*94%*	*47%*

Table 1-1: Amtrak California's Passenger Rail Growth, 2000 to 2012

Source: 2013 California State Rail Plan

Commuter rail systems have also grown in in the past decade. Ridership on the Caltrain (San Francisco-San Jose) and Metrolink (Los Angeles metropolitan area) networks have increased nearly 50% while California's other two commuter services, ACE (Bay Area-Central Valley) and the COASTER (San Diego County) have seen significant growth as well. Train-miles are increasing across these systems as well. Having seen double digit annual ridership growth for the last several years, Caltrain has reached a practical capacity on its network and is now planning to add to its train length instead of adding frequency to meet passenger demand (Keller 2015). In Los Angeles, service expansions on the COASTER and Metrolink systems creates new capacity management challenges with the Class I freight railroads.

Like it does with Amtrak California intercity rail, network ownership plays an important, though different, role in the commuter rail capacity management narrative. Unlike the intercity services, however, Californian commuter rail agencies own a significant portion of the track that they use.

Table 1-2: California Commuter Rail Passenger Rail Growth, 2001-2012

Commuter Rail System	Caltrain[8]	Altamont Commuter Express (ACE)	COASTER	Metrolink	Totals (All Commuter)
Operating Agency	PCJPB	SJJPA	NCTD	SCRRA	
Geographic Location	San Francisco Peninsula	East Bay Area Counties	South Coast (Oceanside-San Diego)	L.A. Metro Area	
2001 Ridership	9,942,082	738,969	1,281,124	8,510,558	20,472,733
2012 Ridership	14,134,117	786,947	1,624,211	11,977,540	28,522,815
% Chg. Ridership	42%	6%	27%	41%	39%
2001 Revenue Train-Miles	1,249,980*	86,000*	206,000*	1,792,000	3,333,980
2012 Revenue Train-Miles	1,337,548	160,677	277,179	2,677,136	4,452,540
% Chg. Train-Miles	7%	87%	35%	49%	34%
2012 Network Size	77.5 miles	86 miles	62 miles	512 miles	737.5 miles
Mileage Owned (% Owned)	51 (66%)	0 (0%)	62 miles (100%)	216 (42%)	329 (45%)

***estimated from timetable analysis, all other data from National Transit Database**

Caltrain runs the lion's share of its services on its own track; only six of its nearly one hundred weekday runs use another railroad's (Union Pacific's) territory. The COASTER in San Diego runs entirely on its own trackage and Metrolink owns nearly half of its network. Commuter rail agencies, then, have a degree of control over access to their network, though much of this control is dictated by long-term trackage-rights agreements.

[8] The PCJPB (Peninsula Corridor Joint Powers Board) operates Caltrain, the NCTD (North County Transit District) operates COASTER, and the SCRRA (Southern California Regional Rail Authority) operates Metrolink. To reduce confusion for the reader, we will try to use the service name (e.g. Caltrain) instead of the owner when possible.

What is driving this growth in passenger rail demand and services? The simplest answer would be population growth, and as we noted earlier, California is expected to grow nearly 25% by 2050 (CA Department of Finance 2014).

Figure 1-5: Growth in Interregional Personal Travel (all modes)
Source: California High-Speed Rail Ridership and Revenue Model, 2012

We have also seen trends that suggest that rail will see an increasing share of trips. Car ownership, once a rite-of-passage into American adulthood, is no longer a priority among many millennials (Ball 2014)[9]. In 2013, the portion of Americans age 16-24 with driver's licenses fell to its lowest level in almost 50 years (ibid). While fuel price increases have a positive cross-elasticity with transit use and helped pushed vehicle-miles travelled down to the same level as 2004, there is evidence to show that once users change to public transportation due to a fuel price spike, it is less likely they will return to their automobiles after prices fall; that is, ridership responds asymmetrically to a rise in fuel prices versus a fall (Chen et al 2011). Fuel price volatility, then, will continue to drive rail trips upward. It is factors like these that suggest that California's passenger rail growth seen over the last 10 years will continue into the middle of the 21st century.

---

[9] Millennials are people born between 1980 and the early 2000s, according to the Census Bureau

One solution to capacity constraints on a railroad is to simply run longer trains. As noted earlier, the freight rail industry has managed demand growth to some extent in this manner. However, for passenger rail service, it is not as easy. Consumers value frequent service; all else being equal, it is more likely that passenger rail agencies would prefer to increase service frequency than simply make trains longer. The U.S. air industry has realized the importance of frequency and serve some of their busiest markets with smaller planes in order to maximize traveler convenience and capture market share (Belobaba 2009). European railroads also aim to provide frequent service between major city pairs at hourly intervals or less (Nash 2003). The challenge of using frequency to provide capacity is that, due to signaling constraints, six, one-car trains require even more than six times the track occupancy time than does one, six-car train. As passenger rail agencies look to increase frequency to provide more attractive options for customers, the importance of capacity management increases as well.

1.2.2 Proliferation of Shared Track Corridors

Amtrak operates over about 22,000 miles of track, most miles of which are owned by Class I freight railroads. According to Liu et al., all of the 18 existing and 38 proposed commuter railroads operate on shared tracks with freight railroads or other transit agencies; that is, even though some passenger railroads may own all of their right-of-way, they still open their track to freight traffic (Liu, Yang, and Chen, 2011).

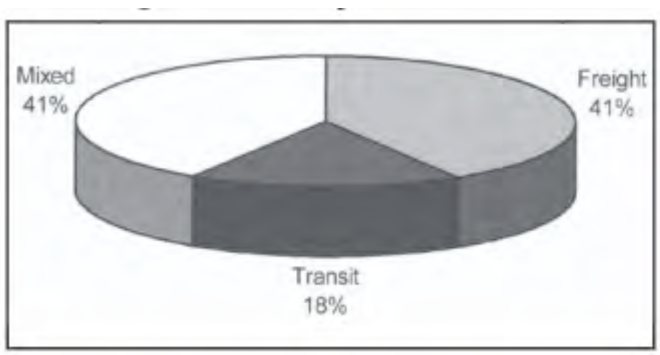


Figure 1-6: Ownership structure of commuter rail tracks in the U.S. For example, 41% of commuter railroads operate on lines owned by both freight railroads and transit agencies.

Since freight railroads own key sections of right-of-way, regions are finding it more cost-effective to build stations on existing freight rights-of-way rather than build new lines. Oftentimes, commuter rail agencies will acquire track from freight railroads; however, freight railroads often require continued access to the track through easements or trackage rights.

The California State Rail Plan lists no less fewer than 13 planned intercity and commuter rail projects, all of which would share track with freight rail lines or operate in territory owned by freight companies. While it is safe to assume that not all of these projects will push the track's capacity to its limit, there will be unique challenges in managing the operation and administration of each of these new shared-use corridors.

1.2.3 Delay Costs

Poor capacity management can make both passenger and freight rail operators experience delays. Delays are a cost to both passenger and freight operators, but their profit incentive makes freight railroads especially sensitive to delays. As a result, there has been much research on costs of railroad delay in the freight sector. Cost of rail delays to freight railroads range anywhere from $200 to $1400 dollars per hour depending on the type of train—intermodal trains tend to have higher delay costs than trains shipping bulk commodities (Lovett 2014). Components of these delay costs include wasted fuel, crew compensation, deprecation and rental expenses of locomotives and cars, as well as the costs of lost revenue or unsatisfied customers. Passenger rail operators experience much of these same costs, however, as subsidized entities and the fact that containers cannot themselves complain, the impact of delays on lost cargo (passenger) goodwill is of more importance than financial costs to passenger rail operators.

Shared rail capacity is subject to peaking much like highways, airports, or other transportation systems. While some sections of track are almost always have excess capacity, certain critical rail junctions or specific urban areas at almost always congested (McClellan 2006). Historical population centers that drive passenger rail traffic are, for many of the same reasons, historical centers of freight traffic. Chicago, Illinois, for example, is a hub for both national intercity, local commuter passengers and freight traffic and experiences extreme rail congestion—it can take two-thirds as long to move a freight train across Chicago as it does to move a freight train from outer Chicago to Los Angeles (Moser 2012).

How do commuter and intercity passenger railroads address these bottlenecks? Because on-time-performance is an oft-sought-after goal, agencies have simply added time to their schedules—a practice known as "schedule padding." As shown in Figure 1-7, between the second-to-last station (Burbank-Bob Hope Airport) and the terminal station in Los Angeles, the Coast Starlight's inbound trip is scheduled to take 10 minutes longer than the outbound trip on the same 14-mile segment; the Coast Starlight's most frequency cause of delays was freight train interference in March 2015. While this improves the on-time performance by accounting for potential delays, this costs the agency in that it makes their services appear less competitive with alternative modes. The industry standard for schedule padding is 6% of the running time (LTK 2012)

11 48A	1021		Salinas, CA Monterey, Carmel —see back			6 28P
1 38P	1119		Paso Robles, CA			4 37P
3 20P	1157		San Luis Obispo, CA (Morro Bay)			3 35P
6 02P	1274		Santa Barbara, CA			12 40P
D7 05P	1310		Oxnard, CA			11 44A
D7 48P	1341	Ar	Simi Valley, CA			R11 11A
D8 22P	1358	Ar	Van Nuys, CA–Amtrak Station			R10 40A
D8 31P	1363	Ar	Burbank-Bob Hope Airport, CA			R10 29A
9 00P	1377	Ar	Los Angeles, CA (PT)		Dp	10 10A

Figure 1-7: Example of schedule padding on Amtrak's Coast Starlight. The travel time from Los Angeles to Burbank is 10 minutes shorter than the travel time from Burbank to Los Angeles
Source: Amtrak

Improved management of capacity can lead to a more efficient rail line that allows for better, more accurate schedule planning which in turn would allow public agencies to offer more competitive options to customers. Furthermore, effective capacity management will provide opportunities for growth of not only shared-use commuter rail services but also expansion of energy- and logistically-efficient freight rail. Perhaps most importantly, it provides a buffer before the last-resort capacity increase solution: construction.

1.2.4 High Capital Costs and Led Times for Additional Capacity

While the capacity gains are likely more significant, an engineering solution is an expensive solution to increasing capacity and requires longer lead times to realize the capacity gains than capacity management solutions (Hunt 2014). While costs vary depending on existing conditions, a new controlled rail siding costs multiple millions of dollars per mile and can be even higher if significant grading or property acquisition is needed to create space for the track (McClellan 2006). In addition to the capital costs, any new length of track requires a larger maintenance budget so operating costs increase as well (ibid).

The long lead times to implement engineering capacity solutions also presents a challenge. Timescales are on the level of multiple years meaning that significant confidence about the state of future demand is required. Or if capacity expansion is reactive, the long lead times result in several years of rail congestion before the solution is operational.

We are in the middle of moving from an “infrastructure era” and into a “systems era,” shifting away from a physical “build what we want” perspective to a “manage what we have” paradigm. In a time of limited resources, in terms of capital and operational funding as well as physical rights-of-way, capacity management becomes all the more important—a “just build it” attitude is both costly and difficult to implement in the current fiscal situation.

That is not to say California agencies are no longer pushing for large capital infrastructure investments. California is ambitiously planning a bevy of rail projects. For example, the San Diego Association of Governments is planning on making its portion of the Los Angeles-San Diego rail corridor 97% double-tracked (up from 50% today) by 2034 at a cost of $800 million (Eschenbach 2014). This effort is composed of 19 separate rail projects, 15 of which are already funded (ibid). In Los Angeles, Metrolink is rehabilitating freight rail track for passenger service as part of its $250 million Perris Valley Line extension (Riverside County Transportation Commission 2015). Finally, in Northern California, the Altamont Corridor Express service is planning to increase to service from 4 daily round trips to 10 by 2022 by adding track in certain locations (San Joaquin Joint Powers Authority 2015). The biggest rail capital expenditure by far, however, is the $68 billion California High-Speed Train Project, which we will now introduce in further detail.

1.3 High-Speed Rail as a Transportation Capacity Solution

California, along with many other states, sees high-speed rail as a one of the most effective alternatives to meet rising transportation demand. The State argues that the cost of equivalent airport and highway capacity would be $158 billion with operation and maintenance costs of $132.8 billion over 50 years; this is opposed to $68 billion for a high-speed rail line that will theoretically operate without subsidy (CHSRA 2014b). Furthermore, in a state very concerned about climate change, high-speed rail is seen as a more environmentally-friendly solution than either automobile or air travel.

In his doctoral dissertation, “The Competitive Advantage of High-Speed Rail,” Reinhard Clever identified five advantages high-speed rail holds over air transportation. First, on distances up to 250 miles, high-speed rail is just as fast as aircraft service when comparing gate-to-gate and platform-to-platform time. Second, HSR trains can act as “subways” in metropolitan areas and take customers directly into downtown areas. Third, trains can split up so one train can serve multiple destinations. Fourth, is the “quick-stop” advantage—in rural areas, high-speed trains can serve smaller communities without the incurring the cost that an individual, separate flight would require. And finally, trains are much easier to automate than planes and hold to on-time performance standards. It is no surprise that a large portion of California’s projected high speed rail travelers are air travelers in the Southern California-Northern California market and road travelers that cannot afford the expensive airfares out of the relatively rural Central Valley.

1.4 High-Speed Rail and Shared Corridors in California

1.4.1 California High Speed Rail: A Brief History

The California High Speed Rail Authority (CHSRA) was established in 1996 under the governorship of Pete Wilson with the mission of directing development and implementation of an “intercity high speed rail service fully integrated with the state’s existing intercity rail and bus network,” for the purpose of meeting future transportation demand without putting excessive demand on California’s intercity air and road network (Kopp 2006). It was not until 2008, however, that Governor Arnold Schwarzenegger placed Proposition 1A on the ballot, a $9.95 billion bond issue that, if passed, promising construction of a line capable of transporting passengers between Los Angeles and San Francisco in no more than 160 minutes with trains capable of running at least 200 miles per hour (Davis and Parra 2008). $9 billion of the bond issue would be available for direct construction of the line itself, the remaining $950 million specified for improvements connectivity to the existing California rail network (ibid).

Figure 1-8: California High-Speed Train Project according to its 2014 business plan
Source: CHSRA

In November 2008, Proposition 1A passed with a 52.2% majority, with 10 of the 16 counties containing the proposed alignment voting for the bond issue, including 7 of the 8 counties through which the San Francisco-Los Angeles alignment was to traverse (L.A. Times 2008). In 2010, various California rail projects were awarded an additional $2.334 billion in federal funding from the American Recovery and Reinvestment Act, including $2.25 billion for California High-Speed Rail (White House 2009). Additional funds from the FRA’s High Speed Intercity Passenger Rail (HSIPR) program—including high-speed rail funding “returned” from the governors of Wisconsin and Florida—have brought the total federal commitment to $3.5 billion (Fleming 2012). The deadline to spend this federal stimulus money is September 2017; and though the CHSRA awarded two billion-dollar design-build contracts for the first 94 miles of the alignment in California’s Central Valley, amid eminent domain lawsuits and other issues to be described later in this thesis, at the time of writing, construction (aside from utility relocation and demolition) has yet to begin (Rosenberg 2015).

According to the latest business plan, the project will be phased in sections. The summary of the phasing for the project is shown below in Table 1-3:

Table 1-3: California High-Speed Rail Authority Phased Implementation Plan (2014)
Source: CHSRA

SECTION	LENGTH (APPROX)	ENDPOINTS	SERVICE DESCRIPTION	PLANNING SCHEDULE	CUMULATIVE COST (YOES, BILLIONS)
Initial Operating Section (IOS)	300 miles	Merced to San Fernando Valley	→ One-seat ride from Merced to San Fernando Valley. → Closes north-south intercity rail gap, connecting Bakersfield and Palmdale and then into Los Angeles Basin. → Begins with construction of up to 130 miles of high-speed rail track and structures in Central Valley. → Private sector operator. → Ridership and revenues sufficient to attract private capital for expansion. → Connects with enhanced regional/local rail for blended operations with common ticketing.	2022	$31
Bay to Basin	410 miles	San Jose and Merced to San Fernando Valley	→ One-seat ride between San Francisco and San Fernando Valley. → Shared use of electrified/upgraded Caltrain corridor between San Jose and San Francisco Transbay Transit Center. → First high-speed rail service to connect the San Francisco Bay Area with the Los Angeles Basin.	2026	$51
Phase 1	520 miles	San Francisco to Los Angeles/ Anaheim	→ One-seat ride between San Francisco and Los Angeles/ Anaheim. → Dedicated high-speed rail infrastructure between San Jose and Los Angeles Union Station. → Shared use of electrified/upgraded Caltrain corridor between San Jose and San Francisco Transbay Transit Center. → Upgraded Metrolink corridor from LA to Anaheim.	2028	$68

In this thesis we will discuss the implications of this phased implantation; that is, how the phasing affects the decisions of other entities that touch the network, be it freight, commuter, or intercity rail. We will also discuss the "blended system" in detail and the implications it brings for both Metrolink in Southern California and Caltrain in the north.

1.5 Motivation: How Might We Improve Capacity Management in California?

The overarching goal of this thesis is to offer recommendations on how California's rail capacity might be managed better and to develop a methodology to approach design decisions that affect capacity. As travel patterns and freight demands move across rail corridors with different owners, capacity planning becomes increasingly important. The following ideas will guide this thesis's recommendations in helping California realize not only its passenger rail service goals, but also provide the goods movement capabilities that allow it to maintain its leadership within the United States and the global economy.

1.5.1 Integration and the Impact of Local Decisions on "System Optimal"

The demand for freight and passenger rail services in the 2nd (Los Angeles-Long Beach) and 11th (San Francisco-Oakland) largest metropolitan areas in the United States has surpassed the supply of existing and undeveloped right-of-ways; freight and passenger sharing of corridors and track in California is a financial imperative. Demand patterns for both freight and passenger traffic have compelled railroads to operate in the "territory" of other host railroads. This context demands that agencies and freight railroads work together to ensure service quality for both cargo and passengers. Agencies need to adopt a systems perspective and understand how their decisions regarding capacity improvements and service changes affect the service possibilities for other agencies: a decision made in San Francisco regarding capacity improvements can affect the quality of service experienced by travelers in Los Angeles and vice versa. This will be especially true as high-speed rail arrives in California directly linking the two regions and putting capacity demands on both rail networks. This thesis strives to explore the upcoming local rail decisions and discuss their effect on the California rail network.

Regarding public and private coordination, freight railroads and passenger railroads need to plan capacity improvements concurrently. This is easier said than done. Since track ownership—especially in Southern California—is fragmented between freight and local agencies, corridor planning often involves multiple owners. Freight carriers have final approval on any new operations on their track, but capacity decisions made outside of freight rail territory could very well affect their ability to allow new passenger rail service (Fox et al 2014). In this thesis we hope to present the key benefits of integration and warn of the risks of agencies taking an individualistic perspective.

1.5.2 Should California follow the European Union's lead in capacity allocation codification?

Over the last 20 years, the European Commission, through a series of directives known as the railway packages, has pushed member states to have vertical separation of infrastructure ownership and operations and allow for open access to railway networks (Levy et al 2015). Actual separation between railroad operators and infrastructure managers varies from country-to-country, but nevertheless, every railroad operator is in theory competing for track access against all other operators. The railway packages have mandated that the infrastructure managers must establish a set of rules governing access to the network. This comes in the form of an annually-published "Network Statement," a document with standardized sections that describes how operators access the network. In particular, in its network statements, the infrastructure manager must specify how to allocate capacity when more train-paths are requested than the network can accommodate.

While it might be unrealistic to expect freight railroads or even public owners in the United States to voluntarily surrender ownership of track to a third-party, we might consider the possibility of multi-lateral agreements on rules governing track access. This is especially possible in regions like Southern California where freight and passenger rail trade off roles between host and tenant railroad.

On the San Francisco Peninsula, railway capacity will need to be allocated between high-speed service and local commuter service. The California High-Speed Rail Authority intends to sell the right to operate the high-speed line as a concession. Since this requires the operating portion of the railroad to be profitable, we could expect any high-speed operator will expect a certain amount of power in gaining track access (or at least have a guarantee of a profit-enabling amount of track access). This will involve a compromise with the commuter rail authority who has a public service obligation to provide service on the same corridor. A set of rules for allocating track capacity might not only reduce the risk for the consortium bidding on the concession contract, but also help the commuter rail agency better plan capacity improvements. We will explore this idea later in this thesis.

1.5.3 Present a path forward for the California Rail Network

Finally, another key goal of this thesis is presenting the CHSRA and other passenger rail agencies in the state a way forward for the development of not just high-speed rail, but for the California transportation network. To this end, we want to present the options and recommend certain actions for specific agencies.

1.6 An Outline of the Rest of This Thesis

In Chapter Two, we perform a literature review and discuss the practical challenges railroads face when sharing corridors. In Chapter Three we will "zoom in" on the San Francisco-San Jose Peninsula Corridor and discuss the capacity challenges the railroads are facing today and into the future. In this chapter, we will present the incredible need for coordination between high-speed rail, commuter rail, and freight on the corridor. We also review in detail the concept of a "blended system." In Chapter Four, we will discuss how this coordination is currently progressing and apply a train-operator financial model to understand the relationship between the agencies on the Peninsula Corridor. Next, in the Chapter Five, we introduce the Southern California rail network and upcoming challenges with the introduction of high-speed rail in the region. Finally, in Chapter Six we present a California rail network "wish-list;" that is, a list of statewide rail goals. We then evaluate four localized decisions and discuss how the impact of those decisions affect the fruition of the aforementioned "wishes."

We thank the reader for joining us and hope this thesis will provide not only a valuable overview of the challenges and opportunities faced by rail agencies and freight operators in California, but also a new paradigm to evaluate the choices being made today by the California High-Speed Rail Authority and other entities in the state.

2 California High-Speed Rail and Sharing Capacity—Literature and Practice

In this chapter, we will introduce capacity allocation mechanisms, discuss the specific challenges of shared corridors from a theoretical and practical standpoint, review the contemporary research on the California High-Speed Rail project, and explain the debate between vertical separation and vertical integration in the rail industry. Finally, we will look at capacity sharing in Europe where vertical separation is written into law.

2.1 Forms of Capacity Allocation and Rail-Specific Challenges

In this next section we will discuss different methods of allocating rail capacity and the specific challenges of distributing rail capacity versus other public utilities.

2.1.1 Stephen Gibson's Mechanisms for Allocating Capacity

Stephen Gibson defines three overarching mechanisms for allocating rail capacity: 1) administrative mechanisms, 2) cost-based mechanisms, and 3) market-based mechanisms (2003). Gibson's writing is focused on rail capacity allocation in the United Kingdom where, unlike in the United States, track ownership and train operators are completely separate entities. However, Gibson's definitions can be modified to generally characterize mechanisms available to railroad operations in the United States.

Administrative (or rule-based) mechanisms are the most prevalent in the United States and Europe and are typically the result of historic precedents and agreements between rail operators. Decisions are not made based on prices or values of the specific rail capacity, but rather a set of rules such as "commuter rail gets priority over freight" or "freight will have track access between 12 A.M. and 5 A.M." These rules might take the form of track-use agreements and could be changed with the consent of both the track owner and the railway operator.

Cost-based mechanisms attempt to "signal to capacity users and providers the costs of consuming or providing" additional services; that is, if a freight operator wanted to use the track during a peak passenger service period, it would need to pay the infrastructure owner a higher price. The difference between this and the administered mechanism is that here the railroad operator has unlimited access to track assuming it can afford the access charges offered by the infrastructure owner.

Finally, market-based mechanisms would attempt to reveal the true value of a particular train to an operator through some form of auction for capacity or track access-trading scheme. Due to the complexity of capacity, which will be described later, market-based mechanisms are particularly difficult to implement in the rail industry.

2.1.2 Rail-Specific Challenges

Capacity is shared between operators in other industries such as utilities, radio bands, and airports. However, several key aspects of rail operations make sharing rail capacity a unique challenge.

Patricia Perennes highlights some of the key challenges that are rail-specific (2013). First, unlike in the telecom industry where phone calls use the same line simultaneously, trains cannot pass one another—in fact, there is a finite and substantial required separation between trains on the same track. Airports gates are similar in this regard—there is a finite and substantial required separation between planes at the same gate (i.e. two planes cannot be at the same gate at the same time). However, rail capacity is a bit more rigid—a stopped train (whether from a mechanical breakdown or stopping at a station) cannot simply be taken apart and moved somewhere else while an airplane could be moved to a different gate or an outlying area on the tarmac. As Perennes puts it, "airport capacity is a one-dimensional process (time) where rail capacity allocation is two-dimensional (time and location) (2). While not the case across all gates at an airport, planes can be switched to different gates; because the number of tracks on a rail line (or even at stations[10]) is orders of magnitude smaller than the number of gates at an airport, rail capacity is much more dependent on how the entire network used.

Another key-element of rail capacity is that railways are not nearly as homogenous as other forms of infrastructure. A kilowatt-hour of power is the same kilowatt-hour whether or it came from a coal-fired power plant or a wind turbine and all phone calls take up more or less the same amount of cable. In aviation, we see bigger heterogeneity, but ultimately, these only amount to a few extra minutes of separation upon takeoff and landing (to avoid wake turbulence effects for safety reasons) or a few extra minutes at a gate to unload and load a larger number of passengers. The heterogeneity of rolling stock in railroad operations, however, with trains operating with different speeds, power-sources, dynamic envelopes, and weight all using the same infrastructure (railway track) create a many shared-use challenges. We group these shared-use challenges into three categories: physical, operation, and institutional. However, these buckets are not in silos—decisions made to address to the challenges in one category affect how operators face challenges in others. We will discuss these shared-use challenges in more detail later in this chapter. First, however, we will introduce some of the research that explores more efficient (in terms of higher infrastructure utilization) strategies to manage capacity.

2.1.3 Novel Approaches to Managing Capacity

Academic research has put forward other possible methods of allocating capacity. In his paper, "Towards a welfare enhancing process to manage railway infrastructure access," Jan-Eric Nilsson develops an allocating mechanism in which operators submit "bid functions" for their preferred train path as well as an indication of their lessened willingness to pay for an alternative path (2002). Nilsson draws some inspiration from the Federal Communications Commission's auctioning of radio frequency which he claims proves that "it is indeed feasible to give particular services a preferential treatment as well as to handle the risk that some operator comes to dominate a market" (434). However, Nilsson leaves the task of prioritizing bids to the government as he admits it is a political problem, not an optimization one.

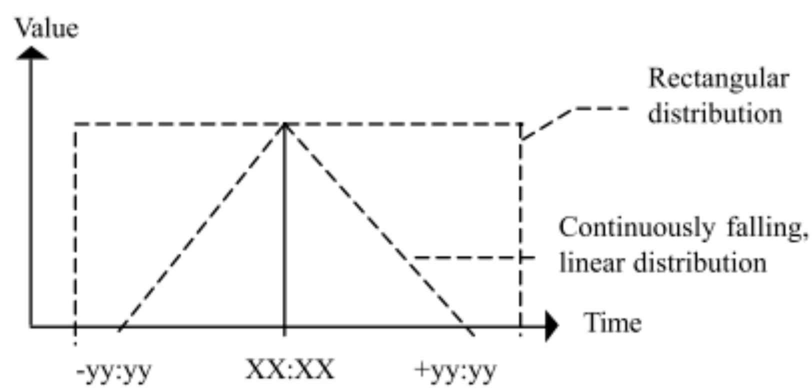


Figure 2-1: Nilsson's operator bid function distribution
Source: Jan Eric Nilsson

Nash and Johnson suggest an alternative where congestion charges are based on the opportunity cost of an incumbent operator not being able to run a train at a desired time (2008). The opportunity cost is based on 1) lost revenue

---

10 The US's busiest rail station, Penn Station, has 21 tracks, while the busiest airport, Hartsfield-Jackson Atlanta International Airport, has 207 gates.

from not running a certain train, 2) the consumer's surplus resulting from the "additional quality" and capacity provided by the new train, 3) the savings of external costs to road users, and 4) less the operating expenses of the incumbent running the train. This could conceivably work for a private operator operating on publically-owned rights-of-way as would be the case in California; and while the incumbent data would be available since these are public services, agreeing on a cost methodology will be challenging.

One of the biggest challenges in allocating capacity in a railroad is the network effects of passenger services. One train may be of great value if it can connect to another train at a railroad hub, but be of almost no value otherwise. Borndorfer attempts to rectify this by proposing a combinatorial auction in which railroad operators bid on bundles of slots with "positive synergies between the slots" (2006). Borndorfer suggests an iterative programming approach to solve the auction. While he proves that this is implementable in the Hannover-Fulda-Kassel area in Germany and "can induce a more efficient use of capacity", it might be more complicated a procedure than California agencies and freight railroads are willing to accept.

In her doctoral thesis, Maite Pena-Alcaraz evaluates price-based and auction-based capacity allocation mechanisms on the Central Corridor in Tanzania and the Northeast Corridor in the United States (2015). In the price-based mechanism, the infrastructure manager sets a price and operators choose which trains to operate based on priority rules. In the auction-based mechanism, the operators indicate the price they are willing to pay for each train as well as their flexibility if the schedule is adjusted (similar to Nilsson's flexibility bid function. The infrastructure manager then chooses the profit maximizing solution (solved via an adaptive relaxed linear program). Pena-Alcaraz finds that the auction performs better from an infrastructure manager cost-recovery standpoint, but operator profits and service levels are decreased.

Ahmendreza Talebian presents a sequential bargaining game for allocating capacity in vertically integrated systems, using the proposed Chicago-St. Louis high-speed line as a case study (2015). We will discuss the distinction between vertical integration and vertical separation in the railway industry later in this chapter. Talebian's approach is as follows: the tenant operator will present their ideal timetable along with a schedule-delay penalty (its compensation for an alternative train slot). The infrastructure owner/host railroad will offer the tenant railroad an access charge. The tenant railroad can accept and the "game" is over. Or the railroad can reject in which case the host railroad offers a new timetable and access charge scheme. The tenant railroad can now offer a new schedule-delay penalty and the host railroad can accept or reject. The equilibrium result is a utility-maximizing schedule for both railroads.

2.2 Integration and Institutions

Reinhard Clever, in his paper, promotes integration in transit agencies. He outlines his Six States of Integration and argues that increasing levels of integration are directly correlated to ridership in transit systems (2013). The lowest level of integration (no integration) means passengers are forced on poorly-timed, long walking-distance rail-to-rail transfers. The highest level of integration, or what Clever describes as "Convergence" is a non-transfer, one seat ride from suburb to urban center. This no doubt comes with institutional and regulatory challenges such as allowing inter-urban lines to run on urban transit infarstrcuture (consider the Acela running on New York Subway tracks, for example), but Clever argues the rewards could be enormous. Clever presents a potential solution for convergence with the Northern California end of the high-speed rail line, an idea we will discuss later in the thesis.

In a similar sense, Meyer et al. discuss collaboration in transportation, calling it "the key to success" (2005). The authors look at various settings to look for collaboration in the form of "efforts in transportation systems management and operations, responding to disruptions caused by unexpected or unusual events, managing transportation assets across modal boundaries, integrating traveler information systems, and integrating transportation and land use strategies." Collaboration versus competition is a very important theme in this thesis and Meyer et al explain the distinction between the two in Table 2-1.

How does collaboration and competition unfold when one entity is a for-profit entity (such as freight operators or high-speed rail) and one is a subsidized public entity? A for-profit entity will ostensibly have a higher-willingness to pay for what is necessary to be profitable. Baritaud and Stefanescu identify three issues with subsidizing public operators against private operators (2000). First, there is the "loose budget" effect where extra subsidies can drive up costs for everyone since the subsidized operator has an increased willingness to pay to access infrastructure. Second, there are distributive issues in that *all* tax payers have to subsidize users of the public transport system. Finally, there is a lack of information about users' willingness to pay for the fixed costs of a transportation system. Baritaud and Stefanescu concede that "any attempt to clarify subsidies to the rail industry…will remain biased" and that access pricing needs to "consider the existence of externalities" (20). Collaboration between a for-profit operator and a subsidized commuter operator will require careful regulation to ensure that the private operator is not a beneficiary of subsidies (or lack thereof).

2.3 Research on the California High-Speed Train Project

There is much literature already written on the California High-Speed Train project. Most of the research has focused on predicting ridership and feasibility of the system, particularly the dedicated system that was planned until the 2012 transition to "blended system."

2.3.1 Reviews of Revenue, Ridership, and Feasibility

There have been numerous studies done (both independently and for the California High-Speed Rail Authority) examining ridership and revenue potential of the system. In 2000, Charles River Associates (CRA) examined revenue and ridership for various alignments and technologies such as steel wheel/steel rail or magnetic levitation. CRA also forecasted diverted demand from air and road as well as stimulated demand from the new system. In preparation for the 2008 bond measure, Cambridge Systematics (CS) provided a ridership and revenue forecast for the Authority, including performing a comprehensive stated-preference (SP) survey at California airports. David Brownstone, Mark Hansen, and Samer Madanat of UC Berkeley's Institute for Transportation Studies completed a peer review of CS's forecast in 2010 and identified thirty areas of concern that included questioning the firm's methodology. Dr. Lance Neumann and Dr. Kimon Proussaloglou of CS, however, stood behind their procedure claiming that it followed generally accepted best practices and provided "objective and unbiased results" (2010). After hearing support for CS's forecast from several members of the public as well as public agencies, the CHSRA sided with CS claiming this is a "classic disagreement between the academician and the industry practitioner" (Van Ark 2010). Since 2011, perhaps in response to the UC Berkeley's criticism, the Authority itself has convened a peer review group of academics and industry practitioners to evaluate revenue and ridership projections furnished by CS and others.

California Assembly Bill 3034, the bill that became Proposition 1A, established an independent peer review group to evaluate "funding plans and prepare its independent judgment as to the feasibility and the reasonableness of the Authority's plans, appropriateness of assumptions, analyses and estimates, and any observations or evaluates the Group deems necessary" (Davis and Parra 2008). In their latest review of the 2014 CHSRA Business Plan, the Peer Review Group identified several issues, some of which we will review in this thesis. These include a funding shortfall of $20.934 billion, the requirement to spend $2.5 billion worth of American Recovery and Reinvestment by September 2017, and a critical look at the blended system (including platform heights, electrification compatibility with freight, and positive train control interoperability), and the costs of revenue risks that would need to be passed to the private operator of the system (Thompson 2014).

2.3.2 High-Speed Rail's Relationship with Air Transportation

In her PhD dissertation, Regina Clewlow compares the environmental consequences of high-speed rail versus air transportation (2012). While this author is not concerned with the environmental implications, Clewlow does discuss the degree to which high-speed rail can capture market share from air travel. She asserts that high-speed rail's market share is more related to journey time rather than distance and that short-haul air-travel demand in Europe "has clearly been affected by the introduction of improved rail travel times" (24) This loss of market share to rail may not necessarily be a loss to the airlines, writes Clewlow. Airlines can then focus on longer haul traffic, more profitable long-haul traffic and use HSR as a complement to their network.

Daniel Albalate, Germa Bel, and Xavier Fageda take a detailed look at competition and cooperation between HSR and airlines and find that while the two modes are often in direct competition, there is "some evidence that HSR can provide feeding services to long haul air services in hub airports, particularly in hub airports with HSR stations" (2013). This could be of particular importance in California where SFO serves as both a hub airport for United Airlines and an HSR station. In Los Angeles, LAX serves as a hub for United Airlines, Delta Airlines, American Airlines as well as a focus city for Southwest Airlines; while LAX is not directly connected to the HSR system, the airport and the HSR station will be a mere 25-minute shuttle ride transfer apart.

2.3.3 California Rail Network Service Planning

In the spirit of the blended system (in which high-speed rail coexists with conventional services), Ulrich Leister applies what he describes as a "Swiss Approach" to high-speed rail in California by emphasizing a lean approach that features a high-speed "trunk" in the middle of the state connecting to major transfer stations that serve the metropolitan areas of San Francisco and Los Angeles (2011). Leister proposes an approach which moderately increases travel time between California's major metropolitan areas, but also (according to Leister's own estimations), greatly reduces the project's overall cost. Ross Maxwell, in his 1999 TRB paper, also emphasizes the Swiss strategy of timed transfers and provides a conceptual schedule map for northern California operations with a San Jose hub. Writes Maxwell, "Like the Swiss, North Americans need build only what is needed, but they should build what is truly needed to develop a fully integrated system. To build only to meet the market on a corridor-by-corridor basis runs the risk of underinvesting and thus failing to maximize the public good." Maxwell and Leister both espouse the importance of cooperation and integration in their service planning for California HSR.

Shuichi Kasuya, in his 2005 master's thesis, also discusses the opportunity for California HSR to act as a commuter service at those major metropolitan areas. He suggests that the low cost of rolling stock relative to the total cost of the infrastructure means that sharing rolling stock for intercity and commuter service is "essential for efficient rail transportation." Sakura performs a case study of Metrolink's Antelope Commuter Rail Line in the Los Angeles basin and concludes that for little additional cost, the HSR operator can increase profits by acting as a commuter service and commuters can benefit up to a 50% travel time savings.

As far as research regarding the current plan for the blended system, we have not been able to find any further academic work. LTK Consulting produced a short feasibility study of the blended system in 2012 and concluded that the system is in fact feasible; however, the author contends that the report did not fully explore the implications of the blended system and will discuss this later in the thesis.

2.4 Physical Challenges

2.4.1 Rail Infrastructure and Equipment Design

While all railroad tracks connected to the United States' national railroad system are standard gauge (4' 8.5" between rails), rolling stock is not nearly as compatible. In this paper, we split the physical challenges in sharing capacity into two broad categories: rail geometry and above-rail geometry.

2.4.2 Rail Geometry

In spite of the American standard gauge, track designed for conventional passenger rail differs from freight rail which in turn differs from high-speed rail. Take, for example, superelevation, which is the practice of "banking" curves to allow for higher speeds. High-speed rail might demand superelevations of up to seven inches; that is, the upper rail in a curve would be seven inches higher than the lower rail (CHSRA 2009). If a heavy, lower-speed freight train were to run consistently through that sever-inch superelevated curve, the lower rail would deteriorate due to "crushing" and rail ties and track surface would be quickly degraded (Zarembski, Blaze, and Patel 2011). This would lead to additional maintenance expense and cause tracks to be out of service while rail and fasteners get replaced. Conversely, if we designed all track for the freight trains and removed superelevation, high-speed trains would be limited to slower speeds through curves causing network delays. Operating high-speed passenger trains too fast through non-superelevated curves can lead to derailments, centripetal lateral forces stresses on the outside rails, as well as decreased passenger comfort (ibid).

If we say that superelevation causes compatibility issues laterally; grade requirements cause compatibility issues longitudinally. Freight trains cannot climb as quickly as passenger trains requiring any shared tracks be brought down to the lowest-common denominator (i.e. freight) in terms of elevation changes. While lower grades do not pose a maintenance or safety risk for passenger train, they do increase capital costs by adding constraints in construction such as building more significant approaches to overpasses and underpasses.

2.4.3 Above-Rail Geometry

The operating envelope of rolling stock creates challenges in shared corridors above the rails. The method with which trains receive their power create compatibility issues in shared corridors. The three basic types of traction power in use today are 1) internally powered (i.e. diesel-electric), 2) third-rail, and 3) overhead catenary systems.

Diesel-electric trains do not require wayside power generation equipment however they do emit carbon monoxide exhaust. This presents a ventilation challenge in tunnels and subways.

Third-rails are physical rails that power trains via an electrical pickup shoes. Third rail power for commuter railroads is rare in the US outside of the New York City area. Because its proximity to the track, the third-rail makes it difficult for freight trains to share tracks. Freight trains in New York that use third-rail tracks need to be modified to avoid scrapping the third-rail and some freight railcars are not permitted whatsoever (Rail New York 2014). Regardless of whether or not freight rail can use the tracks, third-rail operation requires a wider right-of-way than diesel-electric operation.

Overhead catenary poses a similar challenge to third-rail except the clearance issue is vertical as opposed to lateral. Overhead catenary systems could require tunnels to be rebuilt to accommodate the additional height required from the overhead wire (McCallon 2014). Like third-rail power, in catenary power systems catenary poles and wayside equipment such as traction power substations take up increased right-of-way.

Train operating envelope requirements differ between passenger rail and freight rail. To account for brakemen riding on the edges of freight cars or for moving potentially hazardous materials, freight rail requires lateral clearances between the side of the train and platforms. This presents compatibility issues in US shared systems where passenger rail services have high-level platforms. High-level platforms provide for faster boarding and alighting because passengers do not need to step up or down into passenger carriages. However, Americans with Disability Act requirements mandate a maximum gap between rail 1car and platform, directly conflicting with freight lateral clearance requirements (Harkin 1988). Solutions to this require operationally cumbersome or expensive gauntlet or bypass tracks, retractable platforms or mini-high platforms with car-borne, conductor-operated, wheelchair ramps.

2.5 Operational Challenges

Assuming railroads can rectify the physical design challenges associated with running shared track operation, there are still many challenges associated with the actual operation of the shared-use line.

2.5.1 Variables Train Speeds and Lengths

One of the most unique challenges to sharing rail infrastructure versus other infrastructure is the non-homogeneity of capacity: some trains run faster than each other; and unlike cars and trucks on a freeway, passing opportunities are rare. Imagine a single-track rail line with no passing sidings. Passenger trains following slower freight trains would be forced to run slower as well. An alternative would be for passenger trains to allow the slow moving freight trains a "head-start" so the passenger train could run at normal speed as it closes the gap between it and the freight train ahead. This arrangement though means that a large amount of capacity between the freight train and passenger train goes unused.

On systems using block signaling, trains may not enter a section of track (block) until the train ahead of it has left the block. Train spacing is then dictated by the size of the block. Long freight trains with significant braking distances running under block signaling control require larger blocks which limit the capacity of a line. An all-passenger line might be able to accommodate more trains per hour since block lengths can be shorter with shorter trains. Freight railroads claim that one Amtrak long distance train uses between two and five freight train slots given the different speed and length characteristics of Amtrak trains versus freight trains (Wilner 2013).

In California, high-speed rail and commuter rail will be sharing a predominantly two track, bi-directional corridor. In Northern California, even though they will be operating with virtually identical performance characteristics on the corridor (top speed, braking ability, etc.), because the high-speed train will be making fewer stops, it is essentially operating at a higher speed. In her doctoral thesis, Maite Pena-Alcaraz elegantly depicts the loss of capacity induced by mixed traffic operations.

The diagrams on the left are time-space diagram. Each line represents a train travelling across an 80-unit corridor during a 100-unit time period. At unit 40, there are passing tracks that allow trains to pass one another. There is a 10 time-unit space restriction between each successive train. The top diagram represents a system with only fast trains; the capacity of the system based on these constraints is 10 trains per 100 units of time. The middle diagram represents a system with only slow trains; the capacity of the system is also 10 trains per 100 units of time. The bottom diagram represents a shared corridor with mixed traffic. Because of the spacing and passing constraints, the capacity of the mixed traffic system is only 7 trains per 100 units of time.

There is a technique called “timetable compression” in which the timetable is modified and optimizes a train schedule to maximize the throughput of trains per hour. However, timetable compression represents a tradeoff between maximizing utilization of the corridor and presenting an appealing timetable for customers. In summary, mixed operations limits the overall capacity of a rail line.

2.5.2 Safety

Safety is a major concern on shared rail corridors—since different operators use different rolling stock and carry different types of cargo or passengers add to the complexity of safety rules, regulations, and operating procedures. The FRA is in the process of reforming its crashworthiness standards from one based on "prescriptive regulations to a more performance-based regulatory environment. This should allow for shared track operations of light-weight, higher-efficiency railcars (Nunez 2013). Liability and insurance becomes increasing complex as well. The cost of liability for passenger injury or death can be exorbitant for new operators and keep potential host freight railroads from allowing passenger rail access rights (TRB Intercity Passenger Rail Committee 2011).

2.5.3 Positive Train Control

Positive train control, a GPS-based technology that is capable of preventing train-to-train collisions, over-speed derailments, and work-zone incursions was federally mandated after a September 2008 train-to-train collision in Chatsworth, California (Metrolink 2014). The largely unfunded federal mandate, which applies nationwide on all shared-track corridors, did not specify a specific supplier of the technology yet required interoperability so that one railroad could operate on any track of another (Baker 2012). In response, the four Class I railroads—BNSF, UP, CSX, and NS—formed the Interoperable Train Control Committee and agreed to use Westinghouse Air Brake Technology's (Wabtec)'s Interoperable Electronic Train Management System (I-ETMS) as the supplier of their PTC system (ibid). While the Wabtec system is being adopted by many other freight and passenger railroads such as Metra, Sound Transit, Metrolink, it is not universal and integration challenges lie ahead.

2.5.4 Grade Crossings

Every day in the US, trains travel across more than 212,000 highway-rail grade crossings; collisions at grade crossings cause about 270 deaths a year, or about 35% of all rail-related deaths in the United States (Federal Railroad Administration 2015a). Highway-rail grade crossing regulations differ for services at different speeds, so naturally high-speed rail and commuter rail or freight rail will face different regulatory conditions. The Federal Railroad Administration (FRA) sets a maximum highway-rail grade crossing speed of 110 mph, but grants waivers with increased protection at speeds up to 125mph (FRA 2015b). In order to share tracks with commuter rail, high-speed rail typically needs to reduce speeds to commuter rail speeds. This increases travel time for high-speed rail services and makes it less competitive with other intercity modes.

2.5.5 Train Derailments

Freight train derailments pose an additional concern in shared rail corridors. Even if freight trains and passenger trains do not share track, freight derailments can "spill" onto parallel passenger tracks and cause further damage. The FRA has safety standards the require passenger trains to be able to withstand collisions with heavy freight trains. In California for example, Caltrain will use partially-compliant FRA vehicles that will share track with conventional compliant commuter rail rolling stock as well as limited freight traffic (Saat and Barkan 2013). In order to acquire this waiver, Caltrain had to show that there would be temporal separation between its trains and freight equipment; however, this temporal separation leads to track downtime issues as we describe in the next section.

2.5.6 Maintenance Windows

One challenge that railroads operating on shared use track and corridors face is finding time windows to perform routine maintenance. Because closing sections of track requires significant set-up and clean-up times, closing track sections for very short windows is unproductive. This challenge is exacerbated by requirements on many railroads requiring temporal separation of freight and passenger rail; that is, freight is only allowed on the track after passenger rail has completed the day's services. This means that maintenance periods have a higher likelihood of affecting the operations of one of the operators.

2.6 Institutional Challenges

All of the aforementioned shared-use challenges are shaped by the institutions—railroad operators, infrastructure owners, local, regional, and federal government entities—however, the challenges in this section reflect a more fundamental question about how a corridor is to be administered and run. Train priority, for example, undoubtedly has to do with the operations of the rail line, but the decision about which trains (if any) get priority comes from an institutional level. Institutional questions range from who will pay for capacity expansions and investments to who will clean up a fallen tree on the railroad tracks There may be no perfect shared use solution or agreement, but the more open communication there is between operators the more likely the institutional relationships will be successful and long-lasting.

2.6.1 Track Ownership, Dispatch, and Priority

The details regarding the ownership of the tracks and right-of-way on a shared-use corridor, the dispatcher of trains on that corridor, and the priority of certain operating companies versus other operators have a profound effect on the ultimate operation of the corridor.

Owners, that is, the entity that has the property rights associated with the track and/or right-of-way dictate (within regulations and previously agreed-upon contracts) which operators can access the track. With ownership comes the duty to perform or contract out track maintenance.

Dispatching means being in control of train movements along the corridor. For example, dispatchers can choose whether Train A should wait for Train B at a siding or if Train A should proceed potentially delaying Train B. Dispatching does not necessarily come with ownership: for example, Shore Line East and Amtrak in Connecticut runs on Northeast Corridor track owned by the Connecticut Department of Transportation yet trains are dispatched by Metro-North Railroad from Grand Central Station in New York City.

While dispatchers can give priority to certain trains in certain situations, priority often is codified as well. This can be very important in timetable and service planning. Having a lower priority will lead to delays when there are capacity demands that exceed that which the rail line can supply. In many railway systems, priority determines which railroads are able to schedule services at which times. Later in this chapter, we will explore the differing rules European countries maintain for allocating capacity on congested, high-demand rail infrastructure.

2.6.2 Competition versus Collaboration

When there are multiple passenger railroads operating on the same corridor, there is the challenge of determining how customers perceive and use the two services—are these train services complementary or substitutes for one another? In the case of complements, there is a potential for collaboration which adds challenges such as interline ticketing and schedule coordination. However, if the rail products are similar enough and there is no collaboration between the operations, the rail operators could also find themselves competing for the same customers.

In California, Amtrak often finds itself in a unique situation with respect to the railroads whose track it uses to provide its intercity services. Frequently, there is much overlap between Amtrak routes and commuter rail services. Amtrak's product, aside from a slightly fewer station stops, on-board food service, and the opportunity for in advance (yet unreserved seat) reservations, is practically the same as the services offered on its commuter rail hosts. This relationship plays out differently in different parts of the state. In Southern California, Amtrak accepts passengers on certain trains to supplement the commuter rail services in a program called Rail2Rail. On the Peninsula in Northern California, the San Joaquin Regional Rail Commission (operator of Altamont Corridor Express) are not allowed to even sell tickets let alone provide intra-corridor origin-destination trips on its shared track with Caltrain (PCJPB 2013a).

2.6.3 Special Services

There will be times that rail operators will be interested in running additional services whether it is a special train bringing passengers sporting event or an extra freight run to satisfy a time-sensitive shipment demand. These special trains will be subject to previously-agreed upon priority rules. On corridors with scheduled services, special trains will not only need to fit within the operating schedule, but other operators will need confidence about the preservation of their own timetables. If the additional train causes delays, there needs to be a way of compensating other operators—a problem specific to shared-use rail infrastructure.

2.7 Vertical Separation and Vertical Integration

2.7.1 The EU Railway Packages

In 1991 the European Union's European Commission, citing that the "future development and efficient operation of the railway system may be made easier if a distinction is made between the provision of transport services and the operation of infrastructure," issued its First Railway Directive (Holvad 2009) This directive along with additional directives issued in 2001 that outlined requirements for infrastructure charging and capacity allocation systems collectively known as the First Railway Package. The Second Railway Package in 2004 created the European Railway Agency to oversee technical interoperability across EU states without technical, regulatory, or operating constraints. The Third Railway Package in 2007 and Fourth Railway Package proposals in 2013 continued to move towards a Single European Transport Area with full separation between infrastructure managers and railroad operators (Temple 2014). Outside of Europe, vertical separation is relatively rare with the exception of some countries with very low levels of rail traffic.

2.7.2 Comparing Institutional Structures

As evidenced by the Railway Packages, the European Commission strongly believes in vertical separation, also known as "unbundling," of infrastructure and operations. Since rail infrastructure is a natural monopoly, vertical separation supposedly encourages competition at the operator level. This competition leads, in theory, to operating efficiencies. The operating efficiencies, however, could be outweighed by the transactional costs of negotiation between now separate operation and infrastructure entities.

Most of the papers in the literature analyze access charges in railway systems from the perspective of the IM. Maria Herrero provides an economic model for railway capacity costs as a method for (1) railway congestion pricing of scarce railway capacity, (2) timetable optimization between train operators, and (3) quantifying the benefit of railway capacity investments (2014). Her paper draws a parallel between road congestion and rail congestion noting that rail (at least in the European Union) has special characteristics in that railway services are scheduled and that infrastructure is separate from operations. Herrero suggests that it may be possible to find an "optimal toll" for rail capacity, but her paper does not provide a formulation that accounts for the unique characteristics of rail. As Chris Nash and his colleagues explain, determining "optimal charges" that reflect the marginal social cost of railway capacity is not easy (2011).

Drew and Nash, however, warn against the EU moving ahead quickly with unbundling. They point to limited evidence that vertical separation improves rail performance. They note that typically "countries with vertically separated railways have issued fewer operating licenses than those with vertical integration" which runs contrary to the claim that unbundling reduces barriers to entry and allows for increased competition (Drew and Nash 2011). In a report, Steer Davies Gleave points towards high "transaction costs" (e.g. access charges and administrative costs) that result when operators and infrastructure managers are separate. Gomez-Ibanez, once a staunch critic of unbundling points to success in Germany "the costs of the subsidized regional services and commuter services have gone down around 20% since Deutsche Bahn [Germany's national train company] was split up" (Kille 2014). However, Gomez-Ibanez agrees that results have not been as positive in other vertically-separated systems.

In his policy paper for the World Bank, Louis S. Thompson points out some of the benefits of separating tracks from trains (1997).

First, Thompson points out, is the reduction in unit costs; as traffic increases on the network, the cost on a per train basis decreases. This, Thompson claims, is why railroads are keener on serving as "tenant railroads" on tracks owned by other railways than by construction and maintaining their own track to create a redundant network. Outside of the Northeast Corridor and parts of Michigan, Amtrak functions in this manner; rather than building new track, Amtrak is a "tenant" to freight railroads and local commuter rail agencies.

Second, with separation intra-rail competition is now possible, provided there is some level of neutrality between the competing railroads. For example, if one railroad owned the track and the other one was a tenant, then the host railroad could make it nearly impossible to gain access (e.g. the host could charge an exorbitant access charge). While this will be discussed later in the thesis, this neutrality is relevant to the San Francisco Peninsula corridor where the CHSRA operator will compete for track access with freight railroads and the owner of the corridor and operator of commuter rail services.

Next, Thompson says that separation allows for specialization—operators can focus on providing top-quality service and owners can focus on safety, state-of-good-repair maintenance, and corridor upgrades. Again, Amtrak was created under this idea; as Thompson writes, "The only hope for sustaining national passenger services [in the United States] seemed to be to create a separate company (Amtrak) focused entirely on passenger services"

Finally, Thompson claims that "infrastructure separation can also help improve the balance between the public and private sector…critical infrastructure can continue to be public planned and provided, but rail services can be divvied up between public and private agencies." In California, there is a blurring of lines between public and private which will continue with high-speed rail. Today, almost the entirety of the state's key passenger rail infrastructure was built by private railroads and the CHSRA hopes that a private operator will finance the public infrastructure based on returns from operating the high-speed rail line.

Academic studies analyzing the results of vertical separation and integration in railways have come to different conclusions, but one paper — published earlier this year by Japanese professors Fumitoshi Mizutani and Shuji Uranishi in the *Journal of Regulatory Economics* — adds nuance to the question by considering the influence of traffic density (2013). The authors conclude that when traffic density is low, separation reduces costs while the opposite is true with high traffic density. The reason this happens because it high traffic system, vertical separation requires high coordination costs (i.e. maintenance scheduling, safety during busy trains operation that outweigh the lowered production costs resulting from specialization of activities (i.e. operating trains and managing infrastructure) (16).

Vertical separation even among the passenger services without the presence of freight can lead to some of the aforementioned shared corridor challenges. For example, SNCF, France's main rail operator bought trains that were too wide for the existing platforms because of lack of communication with RFF, the infrastructure manager (Lewis 2015). In the United Kingdom, Virgin Trains announced plans for a new fleet of electric tilting trains, designed to operate at up to 140mph, but trains could not realize their full potential because necessary signal upgrades never materialized (Railway Technology 2010). While vertical separation does offer the chance for competitive efficiencies, it does not obviate the challenges of sharing infrastructure; in fact, it may magnify them.

2.8 Shared Corridors in Practice

Shared corridors exist across the world. However, in many countries rail services, the challenges are diminished for various reasons. Sometimes rail services are often government-controlled eliminating many institutional challenges. Or passenger rail services are operated by the same entity as freight services minimizing operational challenges. Sometimes one service type comprises such a small share of traffic that capacity allocation schemes are not necessary. For these reasons, we will take a look at Europe and the United States where shared corridors are not only prevalent with high levels of mixed traffic, but also where freight operators are typically distinct from passenger service providers.

2.8.1 Shared Corridors in the United States

In the United States, most of capacity is shared though trackage rights agreements in which the "host railroad" allows a "tenant railroad" access to the track for either a specified time period or for a specified number of trains. Tenant railroads pay host railroads on a negotiated per-train-mile or per-ton-mile basis that is agreed upon at the time of contract. Agreements are amended as needed, and, in the case of service expansions, require the host railroad to evaluate whether or not its services can sustain the additional trains without causing delays. Amtrak is required by law to receive dispatch priority in freight rail-owned territory and in utilizing freight-railroad owned track, Amtrak pays a user fee that is reduced in the event of excessive delays (Wilner 2013). Making significant changes to services requires detailed capacity studies which are often funded by the tenant railroad and the host railroad has no obligation to accept tenant railroad services increases even if the studies find that delay increases are negligible (though the host railroad in this case has an incentive to collect additional track-access charges).

2.8.2 Shared Corridors in Europe

The aforementioned railway packages managed that the infrastructure managers in EU member states must establish a set of rules governing access to the network. As mentioned in Chapter 1, this comes in the form of an annually-published "network statement," a document with standardized sections that describes how train operators access the network. While each infrastructure manager's regulatory framework for capacity allocation is different, we will examine a select group of states to showcase the variability of these allocation mechanisms and priority rules for congested rail infrastructure.

In Europe, a number of studies on capacity pricing have been conducted, following the Railway Reform in the early 2000s. The International Union of Railways (UIC) has performed a number of longitudinal studies in 2005, 2007 and 2012, looking at charges for passenger high-speed and intercity services (Teixeria and Prodan 2012). The studies provided an overview of the difference in pricing used in each European country, and also provided a longitudinal evaluation of charging systems, concluding that charging mechanisms are getting more heterogeneous. The 2012 UIC study also considered the importance of having charges that are consistent across borders in sending the right incentives to the train operators. It concluded that there was a lack of coordination for charges across borders between infrastructure managers. The International Transport Forum (formerly known as the European Conference of Ministers of Transport) conducted a number of studies in 2005 and 2007, with similar conclusions to the UIC studies, stating that cost recovery principles differ from country to country (EU 2005, Thompson 2008). Both of these sources concluded that there is no agreement between European countries on either how much to charge for capacity or how to structure the charges. Furthermore, a lack of coordination between different infrastructure managers on infrastructure charges had a potential negative effect on cross-border train services due to difficulties of capacity allocation and capacity pricing.

At the time of writing, there is little actual sharing of track in Europe as most of the formerly state-owned services still dominate the market. The degree of intra-rail competition within countries depends largely on their regulatory policies and the market power of incumbent railway operators (which are often former state-owned undertakings). In some states, long-established incumbent railway operators have absorbed smaller ones. In the United Kingdom, competition takes the form of competition "for the track" versus competition "on the track;" that is to say, rail operators bid against each other for the right to operate franchises that have no direct competition.

In Italy in 2012, NTV, a private high-speed operator (of which the French operator, SNCF, owns a 20% share), entered the Italian market. NTV's entry marked the first case of competition in the HSR market in Europe. In September 2014, NTV requested intervention from the Italian government in the high-speed rail market in response to Italian infrastructure manager, RFI's plan to increase track access charges (Day 2014). NTV alleges that RFI which is controlled by the same group that controls Trenitalia, the main incumbent provider of passenger rail service in Italy, is trying to force the competition out of the market (Faiola 2013). In 2011, NTV had accused RFI of making last-minute changes to its network statement that delayed NTV's inauguration of service (Railway Gazette 2011).

Belgium (Source: 2013 Network Statement)

If there is a conflict between rail operator requested train-paths that Infrabel (the Belgian infrastructure manager) and the operators cannot solve through negotiation, the following guidelines, not in ranked in a particular order, apply:

- Operational utilization of the track; that is, the rate at which track is used by operating trains under different operator train-paths
- If both operators had services in the previous timetable, the effective utilization rate of train path reservations by the operator in the previous timetable
- The number of hours reserved and the number of train paths

Operators can appeal Infrabel's decisions if they choose. In order to encourage amicable train-path resolution, Infrabel charges the operators the administrative costs of settling the dispute.

There are some sections of the Belgian railway network that Infrabel has designated as "congested." In that case, Infrabel awards train-paths based on the train type priorities based on the line type; that is, high speed trains will have priority on congested high speed lines. In the event of two high speed trains (or any two similar trains) competing for access in a congested zone, Infrabel awards the train-path to the applicant that will pay Infrabel the higher amount of infrastructure access charges (based on the rail operator's planned services).

Germany (Source 2013 Network Statement)

DB Netz, a subdivision of DB, is the infrastructure manager in Germany. There is a multi-step capacity allocation process. Operators send their requests for trains ("proposals") to DB Netz which is responsible for identifying conflicts. DB Netz then negotiates with operators. Then the final scheduled is published and the operators have 5 days to accept it.

In the coordination period, DB Netz suggests resolutions to conflicts. If a new conflict arises it will be solved separately, but it is not possible to solve the new conflict without disrupting a previously solved conflict. The negotiation process is anonymous if the conflict is small or it implies a high cost for one party. Otherwise it is an open negotiation and alternate routes and services within 1 hour for passengers and 2 hours for freight are discussed. There is a maximum of two conversations to discuss each alternative. If no agreement is reached, then these priority rules will be applied:

1. Express services
2. Hourly or better services
3. Regularly-scheduled freight services
4. Special event services

If there is still a conflict after these priority rules have been applied, then each operator will submit a bid to the regulatory office. The highest bid will get the train scheduled.

Netherlands (Source: 2013 Network Statement)

ProRail is the Dutch infrastructure manager in charge of capacity allocation. For a large number of corridors, ProRail sets minimum frequencies for high speed service. In the event of the conflict, then, between a high speed operator and a conventional operator, the high speed service will win the slot if minimum frequencies on that line are not met. After these high speed frequency conditions are met, ProRail refers to a priority list of train services starting with metropolitan public transportation service as the highest priority going down to private passenger transportation service as the lowest priority. In the case of unresolved conflicts, ProRail then favors train paths with highest utilization (for revenue service operations) of its network.

Portugal (Source: 2012 Network Statement)

REFER, the Portuguese infrastructure manager, works with operators to solve conflicting train-path requests. If a resolution is not found during negotiations with operators, REFER will make a decision based on these qualitative considerations (ranked in order of importance):

1. Overall impact on timetable structure (higher impacts less likely to win train-path)
2. Optimization of capacity use, particularly in terms of quality
3. Priority rules in congested areas (see next paragraph)
4. Number of used identical paths (operators that run frequent service are favored)

5. Chronological order in which requests were received (earlier requests are favored)

In declared congested zones, REFER first allocates capacity to public service operators over private freight service. Next REFER uses a priority table based on service types and time bands. Service types range from high frequency suburban passenger services to international freight to dead-head trains. Typically international and intercity passenger services have highest priority, expect during peak hours, when high frequency (six trains per hour) commuter services have priority. Finally, if there still are conflicts, REFER will apply the following criteria (in descending order of priority):

1. Requests which cause less relative network impact (measurement of this is not defined)
2. Requests which use the highest number of identical paths
3. Requests which use the most train-kilometers on the network
4. The chronological order in which requests were received

Spain (Source: 2013 Network Statement)

ADIF (the infrastructure manager) has the ability to make schedule adjustments. The applicants will be consulted and modifications shall be agreed upon by each applicant, although the infrastructure manager may exercise its power to coordinate, study possible technical solutions, and eventually mediate between applicants. Applicants' requests are considered fulfilled if 1) they are rewarded the paths they requested; 2) the timetable of commercial stops for passenger trains is not altered; and 3) for freight trains, the path does not vary by more than 15 minutes at any point en route.

If operators request the same time slot, the following priority rules apply:

1. Priorities, if any, established by the Ministry of Public Works for the different types of service of each line.
2. Those services declared of public interest by the Ministry of Public Works
3. Train paths allocated and used effectively during the term of the preceding service timetable
4. Requests which are subject to the existence of an previous agreement
5. The highest frequency operators to request a path within the timetable

Switzerland (2013 Network Statement)

The infrastructure manager, Trasse Schwiez, puts together the timetable on an annual basis. Reflecting to Switzerland's highly coordinated, cadenced timetabling, regularly scheduled trains; that is, trains that have "clock-face" service on 30-, 60-, 120-minute intervals take priority. After this, the rail operator that announces the higher projected contribution from that train wins the path. However, in order to win the path, the rail operator pledges 8% of projected passenger revenues for regional trains and 13% of project passenger revenues for franchised long distance trains. This 8%/13% pledge helps prevent operators from overestimating revenues in order to win paths.

United Kingdom (2013 Network Statement, 2012 Network Code)

Network Rail, the infrastructure manager, awards franchises in a competitive proposal process for specific lines, called "track access agreements," on the following basis (in a specifically non-hierarchal order):

- developing innovative timetables which build on the core train service requirement published by the Department for Transport
- investment in innovative ways to transform the customer experience on trains and at stations
- identifying further opportunities for investment along the route, particularly at stations
- making the route and train operations more considerate of the environment
- involving communities along the route in local decision making
- demonstrating how their proposals will support economic growth along the route

While the rail operator has sole authority to operate trains on the line for a set number of years, many of these lines overlap with other franchises. In these cases, Network Rail does not have clear-cut capacity priority rules in the event of conflicting train paths, but has what it describes as a "flexing right" to adjust train slots within track access agreements with rail franchises. Network Rail uses this flexing right in accordance with its "Objective" to "share capacity on the Network for the safe carriage of passengers and goods in the most efficient and economical manner in the overall interest of current and prospective users and providers of railway services." In order to accomplish this Objective, Network Rail has a list of eleven (non-ordered) "Considerations" such as keeping journey times as short as possible or enabling operators of trains to utilize their assets efficiently.

In the case that Network Rail cannot include all train slots even after using its flexing right, it will prioritize the slots as follows:

1. Timetable participants that exercised rights to run trains for the entire timetable period.
2. Timetable participants that have exercised timetable rights in force at the Priority Date and expect to have similar timetable rights continue after their current rights expire.
3. Timetable participants that applied for rights before the Priority Date (40 days before the timetable period begins) that Network Rail expects to be in force during the timetable period.
4. First-come, first-serve for Timetable participants that applied for rights after the Priority Date but before 26 days before the timetable period.

Nash and Johnson use the British rail network as a case study for congestion charging. The authors lament that "once an operator gains a slot, they have no incentive to seek to make more efficient use of capacity, for instance by retiming, rerouting or accelerating their train" (Johnson and Nash 2008).

2.8.3 Is the European Experience Applicable to the United States?

Europe offers a multitude of examples of codified capacity allocation strategies that could theoretically be implemented on corridors in the United States. There are two significant challenges standing in the way of this implementation. First and foremost is the integration of infrastructure ownership and operations on U.S. freight railroads and many U.S. commuter railroads. There is no incentive for a freight railroad or commuter railroad to relinquish control of capacity on its tracks—the loss of control presents an increased risk of operators losing the ability to run trains at profit- or customer-critical windows. The second challenge is one of freight railroads running timetable-free operations in the United States. Planning specific slots (as opposed to windows of operation) means that freight railroads have to have a strong sense of their future demand over the timetable horizon.

Both of these challenges could be overcome, however, and both passenger railroads and freight railroads might agree to some version of capacity allocation where it is in their best interests. This is most likely to occur in corridors where there is mixed freight and passenger ownership. As we will discuss later in this thesis, even if a certain corridor has mixed ownership, freight rail and long-distance passenger rail planning on a corridor affects the nature of service in other parts of the network. Timetable planning and capacity allocation, therefore, needs to be network-wide; or at the very least, timetable the networks effects of a capacity allocation plan should be understood by all parties.

This network effects concept begs the question: what is the extent of the network? In California, do we consider the effect a capacity allocation scheme on the Peninsula has on train services in Southern California? How about freight or long-distance services that run to Chicago? And what about the connecting freight trains and passenger trains in

Chicago that move goods and people to the Eastern Seaboard? For the purposes of this thesis, we consider California the extent of the network, but we keep in mind that operators (especially freight railroads) will likely have out-of-network constraints.

2.9 Conclusions

In this chapter, we discussed current literature on capacity allocation as well as the constraints that make capacity allocation in the rail sector unique. We broke down rail-specific challenges into three broad categories: physical, operational, and institutional challenges. We discussed the philosophies of vertically-integrated and vertically-separated railroad structures, the latter undergoing a full-scale trial in the European Union for the last two decades. Finally, we take a look at capacity allocation mechanisms in the United States and Europe, both of which would fall under Gibson's "administered" category, though European capacity rules are typically much more complex given the vertically-separated environment. In comparing capacity allocation on both sides of the Atlantic, we introduce one of the key concepts of this thesis: it is unwise to confine capacity allocation to a single corridor without realizing the effects of any capacity allocation scheme on the rest of a network.

3 Northern California Blended Service Capacity Challenges

3.1 Introduction

In this chapter, will "zoom in" on the challenges that the northern end of the California HSR line faces in operating their service in conjunction with the commuter service operated on the Peninsula rail corridor between San Francisco and San Jose. We will discuss institutional relationships, some of the planned improvements on the corridor and the importance of the corridor in the future for local commuters and intra-California travelers alike.

3.2 The Peninsula Corridor Commute: History and Institutions

The 51.5-mile Peninsula Corridor between San Francisco and San Jose is poised to become the West Coast's premier example of a shared-use corridor. Five operators already use the corridor—Caltrain, Union Pacific, Amtrak, Amtrak California's *Capitol Corridor*, and the Altamont Commuter Express (ACE)—though the latter three only use the southernmost five miles and two stations. In 2012, the PCJPB and the CHSRA announced that the corridor, with minimal new construction of track-miles, would host both high-speed rail and commuter rail services when high-speed rail comes on-line in 2028.

3.3 The Southern Pacific Railroad Era

After absorbing the original San Francisco-San Jose Railroad Company in 1870, the Southern Pacific Railroad Company operated several long-distance and commuter passenger trains throughout the majority of the 20th century including *The Coast Daylight* to Los Angeles, *The Del Monte* to Monterey, and the *Peninsula Commute* to San Jose, the last of which being almost the exact same service that PCJPB's commuter rail service, branded as *Caltrain*, provides today. Southern Pacific discontinued long-distance passenger trains in 1971 as permitted by the 1970 Rail Passenger Service Act (PCJPB 2015a). Since the Peninsula Commute was a commuter and not long-distance train, Southern Pacific was required to continue operation. In 1979 Southern Pacific attempted to discontinue the service (even offering to buy the 9,000 daily riders 1,000 vans), but the California Public Utilities Commission ruled that SP could not (Miller 1987). The battle ended when San Francisco, San Mateo, and Santa Clara counties as well as the California Department of Transportation agreed to subsidize the Southern Pacific's operating losses (ibid).

3.4 The Peninsula Corridor Joint Powers Board

In 1987, the state released ownership and management of the stations to a newly-formed Peninsula Joint Corridors Powers Board (PCJPB). The PCJPB is made up of nine representatives from the San Mateo County Department of Transportation (SamTrans) and the Santa Clara Valley Transportation Authority (SCVTA), and three appointees from the City and County of San Francisco (PCJPB 1996). The joint powers board serves as the policy-making body for service on the corridor and it has the power to "enter into contracts, employ employees, acquire property, and incur debt in the provisioning of passenger service on the corridor." One of the early and arguably most important acts of the new PCJPB was the acquisition, for $220 million, of the right-of-way from Southern Pacific in 1991. As part of the sale, the Southern Pacific retained the right to operate freight service on the corridor (PCJPB 2015b). Union Pacific Railroad, which merged with Southern Pacific in 1996 to form the largest rail company in the United States, assumed these trackage rights and maintains service today (UP 2015). South of San Jose, the PCJPB operates three Caltrain round trips on Union Pacific trackage to Gilroy, California.

3.5 Union Pacific Railroad and Freight Service

Freight service still has a significant presence on the Peninsula Corridor. In addition to a short-line railroad providing service between the Port of San Francisco and the corridor since 2000, Union Pacific has invested in its own short branch lines that connect to the corridor, including a $2 million upgrade in Redwood City in 2009. In total, 26 shippers transport about two million tons of freight cargo (mostly aggregates, scrap metal, food, and recyclable waste) through the corridor annually (Greenway 2013). This amounts to 20,000-30,000 rail cars annually or about 70 cars per night on three weeknight trains (Tillier 2009). Additionally, the Union Pacific operates a small rail yard in South San Francisco.

Union Pacific freight rail service is a small, yet entrenched entity that will likely stay on the corridor for many more years and it is not just the railroad's prerogative. Together, the Ports of Redwood City and San Francisco are working with the local freight trade group, the Seaport Industrial Association to ensure continued freight rail service between San Francisco and San Jose (Port of Redwood City, 2009). The aforementioned 1991 agreement helps to secure this continued use: Union Pacific "reserves the perpetual right of access to and from and use" of the corridor (UP/PCJPB 1991). Furthermore, Union Pacific is guaranteed one 30-minute window between 10:00 A.M. and 3:00 P.M. each day to run freight trains on each of the northbound and southbound tracks. Between midnight and 5:00 A.M., one main track is reserved for freight use.

If and when high-speed rail arrives, this contract will likely need to be renegotiated and the trackage rights agreement does have some provisions for these negotiations. First, Section 8.3b of the agreement stipulates that if PCJPB would like to "commence construction of facilities on all or substantially all of the length" of the corridor that are "incompatible with the double mainline Freight Service," PCJPB must modify the remaining single track to be "reasonably suitable" for the volume and speed of Union Pacific's freight service. If the construction is incompatible with continuing freight service, according to Section 8.3c of the agreement, PCJPB must file permission to abandon freight service on the corridor and Union Pacific must "be allowed to participate in the abandonment proceedings." Regardless of whether high-speed rail and Caltrain find a way to co-exist with Union Pacific or not, the freight service on the Peninsula will play an important role in the future of the corridor.

3.6 The California Public Utilities Commission

As the agency that oversees California rail safety, the California Public Utilities Commission (CPUC) is an important rule-making body on the Peninsula Corridor. A few of the existing CPUC rules have the potential to significantly affect the future of the corridor.

3.6.1 Freight Train Lateral Clearance

In 1948, CPUC General Order 26-D (revising 26-C) declared that platforms between 8" and 48" above the top of the rail require have a side clearance of 7' 6" from the track centerline. As noted in Chapter 2, this rule is in effect to account for brakemen riding on the edges of freight cars or for moving potentially hazardous materials. Current platforms on the corridor are 8" above the top of the rail so a lateral gap of 4'8" required clearance applies; this 4'8" is measured from the track centerline, meaning about 2 feet beyond the rail (CPUC 1948). Any attempts to raise the height of the platform would mean that additional clearance would be required. However, as noted earlier, since ADA requirements mandate a narrow gap (<3") between the train and platform, these regulations conflict with level-boarding goals of passenger rail on the corridor.

3.6.2 High Voltage Power Regulatory Framework

As the agency overseeing rail safety, the CPUC is developing the regulatory framework for high-voltage 25kV AC systems necessary for high-speed rail—the 750V DC system for light rail in Los Angeles and Sacramento and 600V DC system in San Francisco and San Diego does not provide enough power for the accelerations required by high-speed rail[11]. However, at the behest of both the Union Pacific Railroad and the California High-Speed Rail Authority, CPUC explicitly defines the scope of this current framework as limited to "high-speed rail passenger system capable of operating at speeds of 150 mph or higher, located in dedicated rights-of-way with no public highway-rail at-grade crossings and in which freight operations do not occur" (CPUC 2014) As the Peninsula has not only numerous highway-rail grade crossings, but also freight operations, this framework does not apply to the corridor's electrification plans including the plans of CHSRA itself. The author speculates that Union Pacific is seeking additional concessions from Caltrain in order to operate with freight and high-speed rail.

3.7 The Corridor Today

3.7.1 Baby Bullet Service

In 2004, Caltrain debuted their express "Baby Bullet Service" which radically changed service patterns on the corridor and is widely attributed to partially fueling Caltrain's ridership growth in the last decade. By constructing two-mile long, four track sections at roughly 1/5 and 4/5 of the way along the corridor, the agency opened up the possibility for express train overtakes, with the promise of connecting San Jose and San Francisco in under an hour. These Baby Bullets, named because their top speed of 79 mph was a "baby" speed relative to existing "bullet trains," run in two patterns of service that serve only a handful of stations along the route. With this new express service, overnight, rail commuters could now outrun an automobile driving on an adjacent freeway (see Table 3-1)

Table 3-1: Travel Time Savings with Baby Bullet Service

Station Pair	Mileage	2001 Timetable	2012 Timetable	Travel Time Saved*	2012 Average Train Speed+
San Francisco - San Jose	47.5	90 minutes	54 minutes	40% less time	53 mph
San Francisco - Palo Alto	30.1	61	35	43%	52
San Francisco - Millbrae	13.7	24	16	33%	51
San Jose - Millbrae	33.8	66	39	41%	52
San Jose - Palo Alto	17.4	28	20	29%	52
Millbrae - Palo Alto	16.4	38	19	50%	52

**Travel Times are typical times at peak period accounting for schedule padding at San Francisco/San Jose termini*

+Maximum Allowable Speed on corridor is 79 mph

[11] Low voltage DC power also requires more traction power substations, which increases costs and wayside land requirements

Source: PCJPB, Calculations by Author

The make-all-stops, local rush-hour train disappeared with the new timetable: the only trains serving the corridor during the peak periods were the Baby Bullets and limited service trains that either ran as skip-stop or half-corridor local/half-corridor express (allowing for overtakes on the local portion). Of course, there were winners and losers in terms of stations with this new timetable format. Said former Caltrain engineer and service planner, Robert Doty: "We built a new schedule with new service patterns that optimizes crew and equipment utilization--without adding equipment or crews-and instituted what we call a 'pull model' that draws riders to where our trains need to be to execute the plan" (Johnston 2005). Caltrain also used parking availability as a metric for choosing the stations that the Baby Bullets would serve (ibid). In a region with poor last-mile connections to transit, cars are often a necessary tool to for many Peninsula residents to access train service; approximately 65% of Caltrain riders during the weekday peak had a car available for their trip according to a recent survey. (PCJPB 2013b). In the AM peak period, approximately half of passengers arrived at their station of origin via private automobile and 80% of those automobile drivers parked at their station. The implementation of the Baby Bullet service, by moving service away from local stations, has encouraged this "drive-and-ride" practice.

Figure 3-1: 2004 Caltrain map showing express "Baby Bullet" stations. Note that each "Baby Bullet" train only serves a portion of the express stops. Source: PCJPB

Because of the relative mobility of private automobiles on the Peninsula, Caltrain can afford to "pull" flows to Baby Bullet stations to improve travel times on the whole corridor. Drivers trade a slightly longer drive to a Baby Bullet station as a tradeoff for faster rides to their destination; the 10 Baby Bullet stations accounted for 83% of Caltrain ridership in 2014 despite representing approximately 1/3 of the stations served daily on the corridor (PCJPB 2014a). Again, it should be emphasized that Baby Bullet stations were chosen because of ample parking availability and because they were spaced adequate distances apart to allow for longer sustained high speeds; they are not necessarily places that enjoy high nearby populations or jobs.

3.8 Bicycle Usage

The other consequence of the skip-stop and Baby Bullet-style timetable is new behavior of bicyclists. Because the majority of Peninsula residents live on relatively flat terrain, bicycling is a viable last mile connection for many riders. Coupled with the fact that San Francisco's financial district is over a mile from the San Francisco terminal and that many tech campuses are several miles from the rail corridor, bikes are also a useful and often necessary form of transportation on both ends of many commutes (Wilson 2014). This has led to high bike ridership on Caltrain. The Baby Bullet timetable also creates increased bicycle demand because a bicycle affords cyclists the flexibility to take trains that do not necessarily stop nearest to their workplace or residence. Not surprisingly, the top 10 stations for cyclists, which account for 75 percent of cyclist-passenger volumes (San Francisco, 22nd Street, Millbrae, Hillsdale, San Mateo, Redwood City, Palo Alto, Mountain View, Sunnyvale and San Jose Diridon) are all Baby Bullet Stations (PCJPB 2014a). 14% of

Caltrain commuters commute via bicycle and over 90% of those bicyclists take their bike on board with them suggesting that bikes are used at both ends of the train commute.

Bike usage on Caltrain is a mixed blessing. During Caltrain's annual passenger count in the rainier month of February, nearly 6,000 commuters took their bikes on-board on an average weekday; in the relatively dry summer and fall, when biking conditions are ideal, the author speculates that the figure is probably even higher (PCJPB 2014a). Bicycle usage is growing faster than Caltrain's ridership as a whole and has tripled in the last ten years as both San Francisco and Peninsula cities have made investments in bicycle transportation facilities; the number of passengers bringing bicycles on Caltrain has grown four times faster than overall ridership since 2008 (Wilson 2014, Boone 2015). Caltrain itself has improved its bicycle accommodations: Baby Bullet trains typically have capacity for 48 bicycles and other limited-stop and local trains have capacity for 80. However, demand is outpacing supply; in the aforementioned survey, Caltrain "bumped" 50 bicyclists per week due to lack of space. The San Francisco Bike Coalition, a bicycle advocacy group supplied riders with the following satirical "tardy slip" as a form of pushing for more bike capacity.

Figure 3-2: San Francisco Bike Coalition "tardy slip" for bicycles.
Source: San Francisco Bicycle Coalition

Accommodating bikes however, comes at the expense of seats for riders. Caltrain claims that each bike displaces two seats (PCJPB 2014b). With an average of weekday peak trip length of 22.8 miles (or about 35 minutes), Caltrain must also cater to riders without bicycles who desire seats. Unlike subway or light rail systems, Caltrain cars lack dedicated standing areas, so standing room is limited and uncomfortable for standees (PCJPB 2014c). And as we see in the following section, a seat on a Caltrain commuter train is increasingly hard to come by.

3.8.1 Ridership Explosion and Balanced Commute

Since the Baby Bullet service plan debuted in 2004, Caltrain ridership has climbed dramatically. Average weekday ridership exceeded 60,000 passengers per day during the busier summer months in 2014 up from slightly under 24,000 passengers per day in 2004 (PCJPB 2014d). Increased ridership has resulted in higher levels of crowding on some AM and PM peak period trains. In 2014, 15 trains operated at above 95% of their seat capacity during Caltrain's annual survey period in February (again, a typically lighter travel month); this represents a 50% increase from 10 trains operating at over 95% of seated capacity in 2013.

One unique property of Caltrain service versus other commuter rail services in the nation is that commute patterns are relatively balanced; there is a 60/40 split between peak and reverse peak commuters (PCJPB 2014d). The reverse peak commuting behavior (living in San Francisco and commuting southbound) is partially attributable to tech campuses locating themselves in Silicon Valley where land is more abundant and more affordable than in dense San Francisco. Additionally, San Francisco appeals more to young millennials looking to live without owning an automobile; to accommodate the carless reverse peak commuters, dozens of employer-subsidized shuttles operate between Peninsula Caltrain stations and Silicon Valley employment centers.

3.9 Caltrain Modernization

Caltrain is currently undergoing two major capital projects with up to $800 million in funding support from the Proposition 1A's connectivity funding to improve the corridor and prepare for the entry of high speed rail in 2028 (PCJPB

2014e). These improvements should help Caltrain address some of the limited capacity issues it is starting to face as a system during peak hours.

3.9.1 Communication-Based Overlay Signal System (PTC)

In their compliance with the federal mandate of installing a positive train control system on corridors supporting operation of both passenger and freight rail, Caltrain is developing a unique system known as the Communication-Based Overlay Signal System or CBOSS. Caltrain claims this system will allow the railroad to reduce both separation between trains and "gate down" time at grade-crossings (PCJPB 2014e). With construction underway and the system scheduled to be operational by 2015, CBOSS will require specialized in-cab equipment which raises compatibility concerns with Union Pacific, Amtrak, and eventually the CHSRA operator.

3.9.1.1 Electrification

The electrification of the Caltrain corridor between San Francisco and San Jose (the Peninsula Corridor Electrification Project or PCEP) has the power to change the Caltrain timetable once again. At the time of writing, Caltrain plans to continue operating the Baby Bullet service using diesel locomotives, but add in more frequent service to all stations during the rush-hour commute. An electrified Caltrain will have a service frequency of six trains per hour per direction up from five, a change that it estimates will alone increase demand by 10% (2009 EIR). Because the electric multiple unit (EMU) trains that Caltrain plans on procuring are able to accelerate and decelerate faster than the diesel trainsets, the agency will be able to serve more stations on a single train run while preserving the travel time savings. Below are the anticipated station service levels according to the 2013 Final Environmental Impact Report for the PCEP.

Table 3-2: Anticipated change in service levels at Caltrain stations after electrification project

Station Name	Baby Bullet Service*	Daily Trains in 2014 Timetable	Daily Trains After Electrification Project	Percent Change in Daily Trains
SF-4th & King	Yes	92	114	+24%
SF-22nd Street	No	58	90	+55%
Bayshore	No	40	66	+65%
South SF	No	46	78	+70%
San Bruno	No	56	66	+18%
Millbrae	Yes	82	114	+39%
Broadway	No	0	54	N/A
Burlingame	No	58	66	+14%
San Mateo	Yes	70	96	+37%
Hayward Park	No	40	66	+65%
Hillsdale	Yes	74	102	+38%
Belmont	No	46	66	+43%
San Carlos	No	64	78	+22%
Redwood City	Yes	72	102	+42%
Atherton	No	0	54	N/A
Menlo Park	No	66	96	+45%
Palo Alto	Yes	86	108	+26%
California Ave	No	52	66	+27%
San Antonio	No	46	66	+43%
Mountain View	Yes	80	108	+35%
Sunnyvale	Yes	62	84	+35%
Lawrence	No	56	66	+18%
Santa Clara	No	58	66	+14%
College Park	No	4	4	0%
San Jose	Yes	92	114	+24%

Tamien	No	40	48	+20%

Change in Daily Train Summary

Baby Bullet Stations	**710**	**942 (+232 vs 2014)**	**+33%**
Other Stations	730	1096 (+366 vs 2014)	+50%

**Regular Commute Direction (Northbound in AM, Southbound in PM)*
Source: Fehr and Peers, Author's Calculations

Increased service to non-Baby Bullet stations like Lawrence, California Avenue, Belmont, Hayward Park or South San Francisco will reduce Peninsula commuters need to use their bicycle or automobile because three-mile connections between a train station and a workplace or home now become one-mile or less connections instead. Many residents who need a bike on both ends of their commute might find themselves only needing their bicycle on one end of their commute.

This new timetable could enable residents to store a bike at a station and not carry it onboard a train. This should relieve some of Caltrain's bicycle crowding challenges, but could also drive overall ridership demand. There may be bike-averse riders who have been driving their cars between San Francisco and their Peninsula destinations; with increased traffic on the region's highways and rail service that brings riders "close enough" to their origins or destinations, it is quite probable that a certain degree of mode-shift to Caltrain will occur. Also, to a lesser extent, if Caltrain can remove bike capacity in exchange for more seats, riders who place a high value on having a seat for their commute might return as well.

3.10 The Transbay Transit Center and Downtown Extension

If an electrified corridor draws commuters to Caltrain because of its more comprehensive coverage for last-mile commuters, then the Downtown Extension (DTX) to the Transbay Transit Center (TTC) project should draw even more riders to the rail system; in the words of a 1987 Metropolitan Transportation Commission report (the Bay Area's metropolitan planning organization), the extension is "the single most important improvement that can be made to [Caltrain]". Though the project is delayed due to a current funding gap, the most recent scheduled date of completion (scheduled as of 2014) was in 2024, well ahead of high-speed rail's expected arrival in 2029. In the words of Gillian Gillett, San Francisco current transportation policy director, "We have been trying to extend Caltrain into downtown, where the ridership is, since 1900 (Bialick 2013)." While the current Caltrain San Francisco terminal is well-connected to transit, it is over a mile from the city's downtown financial district. Whether riders make their terminal-to-workplace connections via transit, bicycle, or on foot, the transfer requires a non-trivial amount of time (Transbay 2001). The following table shows that the estimated commute time reduction on for commuters travelling between the Peninsula and San Francisco is on the order of 15 minutes or anywhere from 20% to 40% shorter depending on the origin and destination pairs; given this savings coupled with the political unwillingness and financial infeasibility of expanding freeways in San Francisco or San Mateo counties, this project will very likely bring amount a significant mode shift to from automobile to Caltrain.

3.10.1 The Transbay Transit Center: Location, location, location

How much better situated is the Transbay Transit Center to jobs in San Francisco? To examine this, we perform a GIS analysis to look at number of employees located near each Caltrain station. We use a mapping tool called "buffer" which draws an imaginary ring around each station and then use employment data from ArcGIS's business analyst tool to count 1) how many employees work inside each ring and 2) how many people lived inside each ring. We chose rings of ¼ and ½ of a mile to represent a reasonable walking distance from the train station.

Figure 3-3: Example of the buffer tool used on (from top) Transbay Transit Center, current Caltrain terminus at 4th and King Street, Caltrain stop at 22nd Street, and Bayshore Caltrain station

As we see in Table 3-4, the Transbay Transit Center has an extremely high proportion of Caltrain-friendly jobs nearby. It is worth mentioning that this is based on historic data from 2012; the Transbay Transit Center development itself is anticipating 4,400 units of housing, over six million square feet of office space, and over 100,000 square feet of retail (Transbay 2011) when construction is completed. Using assumptions of 2.63 persons per household (Census Bureau 2015), 225 square feet per person in offices, and 450 square feet per person in retail locations (U.S. Green Building Council, 2008), that is an additional 12,000 residents and 27,000 employees. Of course, other stations (e.g. Hayward Park and San Carlos) are planning Caltrain-oriented development projects of their own, but that will only increase reliance on Caltrain as a mode of transportation. In summary, the Transbay Transit Center will be a huge source of potential Caltrain riders.

Source: ESRI Business Analyst, 2013 dataset

Depending on service assumptions, the ultimate effect of the extension on Caltrain ridership is unclear. In a recent presentation, the Transbay Joint Powers Authority (TJPA), the entity responsible for coordinating financing, design, construction, operation, and maintenance of the terminal, estimated a demand increase of 50% for Caltrain. The 2001 environmental impact report, which was written when Caltrain had a daily ridership of 35,000 (Caltrain now sees in excess of 60,000 riders on some days) estimated a 2040 ridership of 64,000 riders, an increase of over 80%. A 2013 study by the San Francisco County Transportation Authority estimated that with the extension, Caltrain's ridership in and out of San Francisco would increase 201% by 2040; that represents a ridership into San Francisco increasing from approximately 12,000 in 2013 to 36,000 in 2040. Finally, transportation engineering consultancy, Fehr and Peers, responsible for the environmental impact report for the Peninsula Corridor Electrification Project (PCEP), estimates San Francisco ridership to increase from its 2013 estimate of 11,000 to 14,000 with electrification in 2020 to 25,000 with the extension in 2040 of which 8,500 would go to the Transbay Transit Center. That the ridership split between the Transit Center (16,500) and the current terminal at 4th and King Streets (8,500) is skewed in favor of the latter represents the impact of service planning with California High-speed Rail in which only two of every 6 Caltrain services per hour reach the downtown Transbay Transit Center.

Table 3-3: San Francisco Caltrain weekday ridership forecasts
Source: TJPA 2001, Fehr and Peers 2013, SFCTA 2013, and TJPA 2013

Planning Document	2020 PAX (no	2040 PAX***	Train Volume	Notes

	TTC)			
2001 Transbay Transit Center Downtown Extension EIS/EIR	12,950 (-5% vs. 2001)^	32,135 (+136% vs 2001)	132 per day	Drafted before Baby Bullet ridership growth spurt (2004), but 2040 accounts for electrification/high speed rail
2013 Fehr and Peers Final Environmental Impact Report for PCEP	13,692 (+27% vs. 2013)***	23,056 (+114% vs. 2013)	114 per day	Blended HSR service with 1 of 3 trains into Transbay Transit Center, 2 of 3 terminating at 4th and King
2013 SFCTA San Francisco Transportation Plan Update	Not available	36,309 (+170% vs. 2013)*	6 trains per hour at peak	Uses transportation SFCTA SF-CHAMP** activity model
2013 Downtown Extension Presentation at November Transbay JPA meeting	"increase Caltrain riders into SF by more than 50%" (Slide 9)"			Service assumptions and planning year unknown, but most likely operating under Blended Service HSR planning assumptions
***Assumes current split of passengers to other stations in San Francisco (i.e. 22nd Street Station)** ****San Francisco County Transportation Agency, San Francisco CHained Activity Modeling Process** *****Ridership into SF in 2013 according to the Fehr and Peers study was 12,063 passengers/day** **^Does not account for Baby Bullet ridership growth or electrification project being complete**				

The above table summarizes forecasts for San Francisco ridership on Caltrain. In comparing the 2013 SFCTA study which had full Caltrain service (6 trains per hour) into the Transbay Transit Center with the Fehr and Peers study which had two Caltrain trains per hour into the Transbay Transit Center, we see the impact and importance of the Transbay Transit Center to Caltrain. The ridership difference between partial (two trains per hour) and full (six trains per hour) Caltrain service into the Transbay Transit Center is nearly 13,000 riders per day. Since we are counting riders into San Francisco and not just the Transbay Transit Center, this means that these 13,000 riders are using means other than Caltrain (i.e. BART, private shuttle, bicycle, or automobile) to access San Francisco increasing their own disutility and increasing traffic congestion on adjacent freeways.

3.10.2 Concluding Thoughts on Caltrain's Ridership Boom

Caltrain has seen significant ridership growth in the last decade with a system that, aside from the inauguration of Baby Bullet service, has not significantly changed since 1907. However, electrification and the Transbay Transit Center have the potential to stimulate unprecedented levels of demand.

In Figure 3-7, we have a simplified model of a Peninsula resident deciding how to commute into the central business district (CBD) of San Francisco in the peak direction. Because the Baby Bullet service makes it difficult to use minor stations during rush hour periods, the resident needs to have the ability to either drive, take transit, or walk to a "major" station; that is, one with Baby Bullet service. Electrification eliminates "Decision #1" since an electrified Caltrain can serve all stations during the peak period (the answer will be "Yes" to Decision #1 for many more Peninsula residents.) The extension to the CBD at the Transbay Transit Center eliminates "Decision #2" since, assuming most Peninsula residents have the ability to reach a minor station via walking, transit or automobile, the connection to the CBD is complete. With the constraints of a necessary bike or transfer penalty as well as the constraint of having to reach a major station eliminated, ridership should grow even faster. Of course, this is an oversimplification: not all residents can reach a minor Caltrain station nor do all commuters into San Francisco work in the downtown core. However, the fact that neither of these improvements (electrification and the extension into downtown) have been realized yet and that Caltrain ridership has more than doubled since 2004 suggests that Caltrain's current capacity challenges are minor compared to what the system will face in the coming decades. While Caltrain's current ridership boom has a high correlation to a booming Bay Area economy (with some growth possibly attributable to the travel time savings from the Baby Bullet service and changing attitudes towards car ownership), the future growth is instead a result of not-yet-existing physical infrastructure and new planned transit-oriented development.

3.11 CHSRA and the Development of Blended Service on the Peninsula Corridor

We have now taken an in-depth look at Caltrain's upcoming capacity challenges. What we have ignored so far, for the most part, is Caltrain's financial and future operational partner on the Peninsula corridor: California High-Speed Rail.

When voters passed Proposition 1A authorizing bond money for the high-speed rail project in 2008, the plan called for a completely separate high-speed service between San Francisco and Los Angeles. On the Peninsula Corridor, high-speed rail was to run on tracks separate from Caltrain, but in the same right-of-way; that is, the corridor was to be changed from a mainly 2-track corridor to a 4-track one. The high-speed rail line, unlike the existing Peninsula corridor which has 40 grade crossings along its length, was to be fully grade-separated using a series of aerial structures, berms, trenches, and tunnels (LTK 2013a). It soon became clear that tunneling was the preferred option for Peninsula residents; for the CHSRA, tunneling was the most expensive option, and therefore, the least preferred. Cities along the Peninsula protested the Authority's plans; the city of Burlingame went so far as to erect two 59 foot poles with a net in the middle to "model the visual impact" of an aerial alignment and the *Palo Alto Voice* likened the project to a modern-day Berlin Wall cutting through the Peninsula.

Figure 3-4: Burlingame's HSR "mock up" aimed at building a consensus against CHSRA plans for a dedicated corridor
Source: San Mateo Daily Journal

In 2010, civic leaders representing residents of San Mateo and Santa Clara counties on the San Francisco Peninsula, the 4th and 5th strongest Proposition 1A supporters (60.8% and 60.3% respectively), sued the CHSRA over not exploring the negative impacts the project would have on local homeowners adjacent to the alignment (L.A. Times 2008, Durand 2013). Furthermore, cost estimates had ballooned from $45 billion to nearly $100 billion (Draft 2012 Business Plan). Lawmakers in Sacramento and Peninsula congressmen in Washington began pushing instead for what became known as a "blended system," in which high speed rail trains would "share use of an electrified/upgraded Caltrain (PCJPB-owned and operated) corridor between San Jose and the San Francisco Transbay Transit Center" (CHSRA 2012).

3.11.1 Strategic Shift in 2012

In May 2012, the CHSRA released an updated business plan calling for the "blended system," and revised their project cost estimate to $68.5 billion down from nearly $100 billion for the fully-separated system (CHSRA 2012). One month later, the California State Assembly, in an endorsement of the blended system and the revised business plan approved the sale of $4.7 billion in state bonds for construction of the first phase of the project (California Senate 2012). This included $706 million for the electrification of the Caltrain corridor and $42 million for upgrades to Caltrain's signaling system; the CHSRA describes these improvements as "bookends" to "strengthen and improve existing rail networks, while also connecting them with the future high-speed rail system" (PCJPB 2014). And finally, in early September 2013, Governor Jerry Brown signed into law a bill heavily restricting the possibility of a four-track alignment on the Peninsula (it would take a nine agency agreement to return to a four-track system), cementing the likelihood of a blended corridor (Durand 2013).

3.11.2 What is meant by "Blended Service"

The California High-Speed Rail Authority uses the term "blended service" and "blended system" when referring to the shared corridor with Caltrain. Essentially this means "maximizing the use of existing regional and commuter rail systems in urban areas" subject to the constraint of achievable travel time between Los Angeles and San Francisco of no more than 160 minutes and between San Francisco and San Jose of no more than 30 minutes. This "achievable travel time" has come to mean the ability of a high-speed train running in ideal conditions without stopping (CHSRA 2014, Thompson 2014). Additionally, the CHSRA defines "blended operations" as "at all phases of development, seek to use new and existing rail infrastructure more efficiently through coordinated delivery of services, including interlining of trains from one system to another, as well as integrated scheduling to create seamless connections" (CHSRA 2012). It remains to be seen what the Authority means by "interlining of trains," but it does leave room for the possibility of co-branded commuter/high-speed services (e.g. a Caltrain "high-speed" service). "Coordinated delivery of services" could mean as little as working with Caltrain to optimize a timetable or as much as using Caltrain to operate the last leg (San Jose-San Francisco) of a Los Angeles-San Francisco run. We will discuss possible coordination scenarios later in the thesis.

3.12 Capacity Needs of California HSR

What kind of capacity needs will the future California High-speed Rail line require? The text of the 2008 bond measure (Proposition 1A) that authorized nearly $10 billion for California High-speed Rail provides some insight into the capacity required. Section 2407.09(c) of the proposition requires an "achievable operating headway time" of five minutes or less. If we take the inverse of that statement, the system should be able to handle twelve trains per hour per direction.

Even if that is the case, it has become apparent that the CHSRA does not intend to operate 5-minute headways on the Peninsula as part of its blended service with Caltrain[12]. The latest service planning document issued by the Authority as part of its 2014 business plan shows headways of 15 minutes between San Francisco and San Jose (or four trains per

---

[12] The legality of the CHSRA's planned ultimate operating scheme has already been challenged in court and will not be discussed at length in this thesis

hour per direction). The blended operations analysis done by a consultant for Caltrain in 2012 (which we will review later in this thesis), also calls for four trains per hour per direction.

3.12.1 Revenue-Neutral Requirement

One of the cornerstones of both the campaign for Proposition 1A was that the California HSR line would be revenue-neutral; that is, operating revenues would meet operating costs. In fact, in the last two business plans, the Authority has claimed that once the initial construction segment is operational, private investment will finance the remaining sections with the expectation that operational profits will also cover part of the capital construction costs. The text of the bond measure itself reads "the planned passenger service by the authority in the corridor or useable segment thereof will not require a local, state or federal operating subsidy." While several of the CHSRA's ridership studies have supported this claim, if private enterprise cannot be convinced of ultimately profiting from operation, additional funding will have to come from the federal government. At the time of writing, the chances of voters passing another bond measure or California raising enough capital through the state budget process or the federal government funding the shortfall (about $50 billion) are slim.

In the next sections, we explore the importance of the San Francisco-San Jose segment and the Transbay Transit Center to the ultimate profitability of a California HSR line. We aim to show that any private operator needs a guaranteed level of access to the Transbay Transit Center and San Francisco to ensure profitability for the entire network.

3.13 High-speed Rail as a Short-Haul California Airline

High-speed rail works best for travelers travelling distances of 100-500 miles. For distances less than 100 miles, commuter rail services (and private automobile) offer superior service because of smaller "first-mile" and "last-mile connections" between stations (parking spaces) and ultimate origins or destinations. On distances longer than 500 miles, airplanes have an advantage due to their superior speed. In the middle range of 100 to 500 miles, the added time required for using airlines makes high-speed rail more competitive. Most of the city pairs in California fall into this 100-500 mile "sweet spot."

However, because there is no high-speed rail alternative in California today, competition between the San Francisco Bay Area and the Los Angeles Basin in the airline industry is intense. Between the Los Angeles and San Francisco regions alone, all four major airlines in addition to JetBlue and Virgin America offered a combined 174 daily flights in 2013 (Perkins 2013). Delta Airlines operates a special west coast "Delta Shuttle" service that features hourly departures between SFO and LAX; in the 1990s, United debuted "Shuttle by United" with departures every half-hour between the airports until delays in and out of San Francisco hampered operations and compelled United to cancel its "airline within an airline." And as recently as 2013, a new airline startup, "Surf Air" began offering subscription air service to smaller airports in the Bay Area and Los Angeles region.

Diverting passengers from air travel and capturing air travel spillover is going to be an important part of California HSR's ridership and revenue. A 1999 Charles River Associates (CRA) study for CHSRA anticipated a 56% diversion of local air traffic to high speed rail (CRA). A more recent 2010 Aviation Systems Consulting study that assumed HSR fares to be 83% of an equivalent airfare forecasted Bay Area airports to experience a diversion of 54.4%, 63.1%, and 53.4% to the San Joaquin Valley, North Los Angeles Basin, and South Los Angeles Basin respectively

(Association of Bay Area Governments 2010). Given the size of the intra-California air market and those diversion rates, it is safe to say that air diversion is a critical piece of the California HSR revenue model.

3.13.1 Value of Frequency to Business Travelers and HSR's competitive advantage

The ability of California HSR to attract business travelers and commuters will depend, not only on its travel speed, but its frequency. Ryan Westrom, in his thesis, finds that "The combination of high-speed and appropriate frequency will result in a service that is most advantageous, and thus competitive, for a trip of the right distance" (Westrom 2014). For high-speed rail service, much like the airline sector, frequency helps reduce schedule displacement between trains. Schedule displacement is the perceived "wait time"; for example, a traveler that ideally would depart at noon, but instead departs on a 12:45pm plane or train has a schedule displacement of 45 minutes.

Herein lies high-speed rail's competitive advantage. Because the marginal cost of adding rail service is much less than the marginal cost of an additional flight (rail's fixed costs are higher than those of airlines), rail can support much higher frequencies than air even if their railcars are not as full (in general, rail can run profitably with a lower load factor, that is, percent of seats occupied). Furthermore, in short-haul markets like California, frequency is more important than in long-haul markets because the ratio of wait time to travel time is much higher.

As noted before, the best frequency offered in the market today by a single airline (and only between SFO and LAX) is the Delta Shuttle with a single trip per hour. Markets like San Francisco-Merced, Los Angeles-Bakersfield, or Fresno-Palmdale see at best one or two trips per day from a single airline. Given its competitive conditions, California high-speed rail will divert business travelers from air and be successful only if it can provide reliable frequency into and out of its San Francisco and Los Angeles termini.

3.13.2 Value of Peak Trips to an Operator (HSR Revenue Management)

Peak trips, that is, those trips taking place during "rush hour" where temporal demand is highest, are also very important to a private operator's business model. Since consumer willingness to pay for peak trips is higher than off-peak, the operator could charge a fare premium maximizing yield (revenue per seat-mile). The operator could then lower fares in off-peak periods to capture additional passengers, simultaneously driving up load factor and offering an affordable option for families or leisure travelers. This practice grew out of the air industry, but has been adopted by rail operators worldwide including Amtrak (since 1991), SCNF (1993), VIA RAIL (1993), Deutshe Bahn (2002) and Great Northeastern Railway in the U.K. (2004). Later in this chapter, we will examine where exactly this peak might exist for San Francisco and how that peak compares to the peak demand time for Caltrain's commuter rail service.

3.13.3 Importance of San Francisco

On the Peninsula Corridor, San Jose may have a larger population, but it is less dense; and thus, the value of a station in the city's central business district is less important. San Francisco, on the other hand, should be a very lucrative hub for a high-speed rail operator. The Transbay Transit Center will be located adjacent to a dense downtown core with excellent transit connections to Oakland and the East Bay via BART as well as Marin County via frequent express bus service across the Golden Gate Bridge. In his thesis, Westrom claims that "any city, regardless of its distance from a metropolitan area, can move into the commutable realm of a central city if the travel time resulting from a HSR improvement moves to below one hour." Because of the Transbay Transit Center's central location within San Francisco, cities like Gilroy and Merced become reasonable commutes for commuters living and working in San Francisco.

3.13.4 The Transbay Transit Center Advantage versus San Francisco International Airport

One of the biggest advantages the Transbay Transit Center holds over San Francisco International Airport (SFO) is its proximity to San Francisco's Central Business District. SFO may be closer than the Transbay Transit Center to Silicon Valley, but HSR stations in Millbrae and San Jose shared with Caltrain service would give HSR an advantage in that regard. As opposed to a 5 minute walk from the Transbay Transit Center, a taxi trip from SFO to the CBD requires about 25 minutes and costs $45 and a similar BART trip costs $9 and takes about 5 minutes longer, at 31 minutes each way. Furthermore, San Francisco International Airport, who loses one its two runways during heavy fog levels (due to FAA minimum separation requirements), is also one of the most weather-delayed airports in the United States giving time-conscious travelers reason to avoid the airport if a reasonable alternative such as high-speed rail exists (Barba 2014).

3.13.5 Peaking problem at Transbay

The Transbay Transit Center is going to be an important piece of infrastructure for both Caltrain and HSR commuters. It is also going to be important at the same time of day. As one of the largest cities in both terms of population and jobs, San Francisco will no doubt be a key traffic generator for the California high-speed rail operator. If we are to use air travel demand as a proxy for HSR demand, we can see the importance of the San Francisco hub at the same time of the day as Caltrain demand into the Transbay Transit Center.

First, we look at Caltrain ridership based on the current schedule. We have taken a moving average to smooth the data between individual trains. The time axis shown is the time the train arrives (northbound) or departs (southbound) San Francisco. For northbound passengers, we adjust the arrival time 15 minutes later to account for the idea that if Peninsula commuters could save 15 minutes with a train to the Transbay Transit Center; they would depart from their Peninsula station 15 minutes later than they do today.

In examining the data, we look at the peak (commuting northbound in the morning) and reverse peak (commuting southbound in the evening). Again, it is worth mentioning that Caltrain's peak-reverse peak commute balance is approximately 60-40.

Figure 3-5: Northbound Caltrain ridership by time of day (adjusted 15 minutes earlier to account for transfer time. Peak travel periods selected empirically by author
Source: Caltrain Annual Passenger Counts

Figure 3-6: Southbound Caltrain ridership by time of day. Peak travel periods selected empirically by author
Source: Caltrain Annual Passenger Counts

Next, we look at 7000-plus seats available on airlines flying between San Francisco and airports along the high-speed rail route on an average January 2015 weekday. While SFO-LAX has the biggest share of the market (93.5% of seats), there was also service to Burbank (4%), Fresno (2%) and Bakersfield (0.5%). We round SFO departure times to the nearest 15 minute bin and take the 2nd moving average to smooth out peaks from individual flights (San Francisco International Airport 2015). Again, for northbound passengers, we adjust the arrival time to account for the fact that an air traveler would like to be in the central business district 45 minutes after his preferred arrival time (accounting for any deplaning process plus taxi or BART to San Francisco).

In looking at the above graph, we do see overlaps between Caltrain peak and reverse peak volumes and air traffic peak and reverse peak volumes suggesting that demand into the Transbay Transit Center will be temporally similar between Caltrain and California HSR.

A better metric than seats arriving and departing to compare air travel to potential rail travel would be "revenue per airliner" arriving and departing San Francisco. Unfortunately we cannot estimate average fare for these individual flights and airlines are not willing to share sensitive revenue data. However, Peter Belobaba, editor of *The Global Airline Industry* writes, "Peak departure times (early morning and late afternoon) are most attractive to a large proportion of travelers in many markets, as "time of day" distributions of desired departure times tend to be clustered around 8am to 9am in the morning and 5pm to 6pm in the afternoon." These times overlap precisely with Caltrain's regular and reverse peak (54).

Given the supply of seats at SFO and our knowledge of peak travel times of 8am to 9am and 5pm to 6pm, we can infer that the highest revenue intra-California flights into and out of San Francisco and by extension, the most valuable trains to a private high-speed rail operator in California, will be those arriving and departing San Francisco at similar times as Caltrain's peak.

3.14 Conclusions

In this chapter, we have described the upcoming challenges that Caltrain will face in terms of capacity. Without high-speed rail, the electrification project, or the downtown extension, Caltrain is already reaching its capacity limitations. Installation of the CBOSS positive train control system will allow Caltrain to run tighter headways of up to 6-trains per hour and a recent railcar investment will allow for 6-car trains instead of today's five car consists. However, the Peninsula Corridor Electrification Project will bring new demand as minor stations see peak hour service.

Fehr and Peers in their Final Environmental Impact Report for the project, claim that due to shorter passenger trips leading to higher turnover as well as riders taking trains at the edges of the peak, Caltrain will manage capacity better than it does today. Even so, Fehr and Peers' forecasts already appear out of date given Caltrain's astonishing ridership growth. Fehr and Peers predicted a no-project 2020 average weekday ridership of 57,047 passengers, up from a baseline of 47,066 in 2013. In 2014, just one year after Fehr and Peers report, ridership was recorded at 53,466, well past the halfway point between the 2013 baseline and the 2020 prediction (PCJPB 2014d). On multiple days in late 2014, average weekday ridership topped 60,000 riders (PCJPB 2015). Caltrain provides a valuable service on the Peninsula by keeping its patrons out of congestion and serving as a reliever for those to driving—out of choice or necessity—on Peninsula freeways. As its service patterns change and Peninsula land-use brings residents closer to stations, this ridership rise will only continue.

The Transbay Transit Center will bring even more demand into the Caltrain system. As we sketched out earlier, the project eliminates a missing link in the Peninsula-San Francisco commute by bringing commuters directly to the central business district. If any single project increases demand for peninsula rail services, this is the project. However, questions remain about Caltrain's service levels at the station, how the station will be operated, and the extent of crowding on the Caltrain system once tracks are built and service is extended there.

The Peninsula is a critical piece of high-speed rail too. CHSRA intends to sell the right to finance, build, operate, and maintain the California high-speed rail system to a private investor. As part of the Proposition 1A bond measure, the operating profits from high-speed rail should not only cover operating and maintenance costs, but also help recoup capital expenditures beyond the initial construction segment. In order for high-speed rail to be profitable in California, we claim that consistent, peak-hour service into San Francisco is vital. This is due to the business and leisure demand into the city as well as the Transbay Transit Center's ideal location in the resident-and job-rich urban core, one of the densest central business districts in the country.

In the next chapter, we look at the feasibility of operating two railroads that are "competing" not necessarily against each other (though the 2014 CHSRA business plan shows a bit of this behavior on the Peninsula), but rather competing for access to the same track at roughly the same time. We will review the degree present of institutional coordination as well as the analysis that has been completed supporting the feasibility of operations. We will discuss market-based alternatives to coordination as well as the implications of separating ownership of infrastructure from railroad operators as has been done in the Europe and could be done at the Transbay Transit Center. There are challenges in both approaches that might be solved with stronger, more service-based scenario planning and a higher degree of institutional coordination, not just on the Peninsula, but across the entire California passenger rail network.

4 The Blended System—Current Status and Implications

4.1 Introduction

In the last chapter, we completed an in-depth review of the future capacity challenges that we anticipate Caltrain and the California High-Speed Rail Authority will face. Caltrain is already, by some measures, at capacity. However, new transit-oriented development on the Peninsula, the electrification of the system and the extension to the Transbay Transit Center in the commercial heart of San Francisco will stimulate unprecedented levels of demand for the two-track rail system. When California HSR is added to the mix in the form of the "blended system", demand on the line will overlap spatially (in terms of San Francisco) and temporally (in terms of AM and PM peak travel). The blended system and its implications on the operation of the rest of the California rail network will be a theme throughout the rest of this chapter and this thesis.

In this chapter, we will start by discussing the level of analysis that been done regarding the blended system—we note that the extent of analysis is relatively minor given its importance to both Caltrain and high-speed rail. The two main sources of analysis are 1) memoranda of understanding between the California High-Speed Rail Authority and the Peninsula Corridor Joint Powers Board (PCJPB) and 2) a feasibility study conducted by LTK Engineering on behalf of the PCJPB. Next, we will review the Transbay Transit Center and Downtown Extension and its unique position as a third-party owned entity that will host both Caltrain and the California HSR operator. This will lead into a discussion of the financial ability of both operators to pay for entry into the Transbay Transit Center: would and should access charges lead to a market-based solution to the capacity crunch at the terminal? To that end, we introduce a model (developed previously in this author's Transportation Research Board paper) for understanding each operator's financial position and analyze the willingness to pay for access into the San Francisco terminal. Finally, we discuss the implications of our model's results and why coordination and integration of services might be a better solution for solving the challenge of limited line and terminal capacity on the Peninsula Corridor.

4.2 Current Extent of Coordination between Caltrain and the CHSRA

The history of coordination between the CHSRA and PCJPB (Caltrain) predates the 2008 HSR bond measure. The two agencies signed a memorandum of understanding (MOU) in 2004 to "set forth a framework for future cooperation" which included aims of sharing engineering and service plans for the sake of compatibility between the systems (PCJPB 2004). In 2009, after voters passed the bond measure, the agencies signed another MOU which established a "working group" to plan, design, and implement high-speed rail service on the Peninsula as a joint project between the agencies. The so-called Peninsula Rail Program, was staffed 50-50 by both agencies and existed until the politically-motivated shift from separated infrastructure to "blended service" in 2012.

In 2013, the agencies signed a third MOU that described the blended system in broad terms (PCJPB 2013). Before continuing, it is important to note that a memorandum of understanding is a non-binding document which describes the intent of both parties. While the following affirmations are important in understanding the institutional "mindsets," they are by no means guarantees of future conditions. In fact, a key recital in the 2013 MOU was to nullify the two previous MOUs described in the prior section. The 2013 MOU does provide some interesting insights into understanding the relationship between CHSRA and PCJPB: the corridor is very much still under local control—high-speed rail, at this point, is very much a "tenant railroad" and freight is not going to disappear easily. We will now review the 2013 MOU in more detail to distill some of the themes and institutional mindsets buried within the document.

4.2.1 Freight will remain on the corridor

The CHSRA and PCJPB agreed that freight will remain in operation on the corridor as per the existing trackage rights agreement with Union Pacific Railroad. This brings a host of challenges that were described in earlier chapters, including clearance issues, regulations regarding HSR equipment crashworthiness, and track design constraints. Furthermore, freight's presence on the corridor will tighten work windows during the construction phase, potentially delaying the project.

4.2.2 Utilize primarily the existing track configuration

This recital was likely inserted to mollify the critics of the original high-speed rail construction plan. While it is a non-binding document, it does reaffirm the commitment to blended service.

4.2.3 Caltrain service remains operational during Blended System construction

While the theme of this goal is "minimize disruption," the wording when combined with the non-binding nature of the MOU itself says very little. In addition to commuter rail service, this recital attempts to protect the operations of other passenger rail services (i.e. Altamont Corridor Express and Amtrak California) as well as freight rail services.

4.2.4 PCJPB retains ownership of improvements and right-of-way

This might be the most important recital in the MOU because it reaffirms the PCJPB's control of operations on the corridor. The fact the CHSRA is paying for a lion's share of the improvements in terms of positive train control and electrification does not give them any more control than if PCJPB had financed the improvements on their own. At a future date, the CHSRA will likely have to negotiate some form of minimum service guarantee for high-speed rail on the corridor. Otherwise, PCJPB, as corridor owners, would have the right to squeeze high-speed rail service to a point where it is no longer possible to operate profitably on the corridor or at all.

4.3 Caltrain EMU Vehicle Procurement

Caltrain will need to make a decision on platform height which will dictate the design of their new EMU vehicles. The latest plans call for high-speed rail to have a 50" boarding height and level boarding. In Caltrain's most recent board meeting, Marian Lee, the executive officer for PCJPB's Caltrain Modernization Program said that her staff would "focus on boarding height compatibility with high-speed rail vehicles" in the spring and summer of 2015. It appears unlikely that high-speed rail will select a platform height compatible with commuter rail agencies; however, it appears quite likely that the CHSRA will use level boarding on their platforms. Level boarding is very important to HSR reliability due to aforementioned dwell time variations with stairs and passengers with wheelchairs, bicycles, or excess baggage.

In their project update report in March 2015 to the California State Legislature, the CHSRA also stated that they are working with Caltrain to develop an RFP for Caltrain's new vehicles that will "provide boarding capabilities at both high and low levels. This configuration will allow Caltrain to operate at existing platform levels once electrified service in the Peninsula is underway, yet provide flexibility for ultimate conversion to a common high-level boarding solution" (CHSRA 2015a)

4.4 Blended Operations Analysis

In transitioning strategies from separate right-of-way to the “blended system” Caltrain hired LTK Engineering Services to prepare a “proof of concept” for the future operations on the corridor. This report was published in March 2012 (LTK 2012). LTK used a proprietary rail simulation tool to model the corridor as well as several new sections of passing track that could be constructed. LTK concluded that the line, at its present state plus electrification and a new signal system, could at most support six Caltrain trains and two high-speed rail trains per hour per direction. Off-peak high-speed rail service was assumed to be two high-speed rail trains per hour per direction. According to LTK’s analysis and simulations, improvements to the system such as mid-line passing tracks would allow for the six Caltrain trains and four high-speed rail trains. LTK’s model strangely did not include the Transbay Transit Center, though this was updated in 2013 (LTK 2013).

Below is a summary of the results from the operations simulations. Using the output string-line (time-space diagrams) in the report, we calculate the standard deviation of headways experienced by the Caltrain commuter between 7 a.m. and 9 a.m. at three major stations on the corridor: San Jose, Redwood City, and San Francisco (current terminal, not the Transbay Transit Center). We also calculate the headway range to show the longest wait by a customer at that station during the AM peak versus the shortest wait experienced by a customer at that station. This is another measure of the schedule irregularity that LTK must introduce to accommodate high-speed rail on the corridor. In all these simulations, HSR runs at evenly-spaced, uniform headways.

Figure 4-1: Evaluation of Blended Service Plans

**Headway Range = Maximum Headway - Minimum Headway*

Source: Author’s Calculations based on Blended Service Operations Analysis

The LTK analysis shows that Caltrain will not be able to run very uniform headways in a blended system. The standard deviation of headways in the “baseline” scenario is low (about one minute) meaning that Caltrain services are practically uniform without high-speed rail. The difference between the longest and shortest headway is also small (four minutes in San Francisco and three minutes at Redwood City and San Jose). As soon as a single HSR train is added to the blended system, though, the variation between headways versus the baseline triples at San Jose and nearly quintuples at Redwood City. At San Francisco 4th and King station, the shortest headway is four minutes and the longest headway is 19 minutes, resulting in a 15 minute maximum headway gap.

Caltrain does not use uniform headways now, but these gaps in service might be undesirable for several reasons. First, at six evenly-spaced trains per hour, Caltrain becomes a “walkup service” during the peak period at main stops meaning that riders do not need to consult timetables, a significant amenity, especially for casual users and commuters with unpredictable schedules (MBTA 2010). Second, uneven spaced train services will lead to crowding on trains immediately following a temporal gap in service (this is known in the transit industry as the “bunching” effect) (Daganzo 2009). Finally, even headways allow for users to commit timetables to memory (e.g. “the train leaves San Carlos at 5; 15; 25; 35; 45; and 55 minutes past the hour”). In Caltrain’s pre-blended system plan when high-speed rail was on its own separate tracks, the PCJPB envisioned uniform skip-stop service as high as 10 trains per hour per direction with four trains per hour running into the Transbay Transit Center (American Public Transportation Association 2006).

4.4.1 Blended Analysis: Stakeholder Considerations Report

The original 2012 analysis neither included the Downtown Extension nor acknowledged the presence of freight. At the request of local stakeholders, LTK issued another report in 2013 that addressed some of these missing analyses.

In the new 2013 report, LTK included a new simulated service plan with the Downtown Extension incorporated. The report suggested that only two Caltrain trains per hour could use the Transbay Transit Center during the AM peak. Furthermore, this schedule functioned with a long mid-track overtake requiring significant new construction[13]. LTK did not simulate the as-built infrastructure nor did it simulate the possibility of running additional (more than two trains per hour per direction) Caltrain trains to the Transbay Transit Center. Concludes LTK, “the DTX and the TTC support the blended system. However, they result in higher levels of signal delay and more added Caltrain station stops to support the service extension to downtown San Francisco.”

Regarding freight service, LTK admitted that the 30-minute window allotted to freight in the middle of the service day may need to be renegotiated to midnight-5 a.m. only. A 30-minute window guaranteed for freight limits passenger service (Caltrain or HSR) to only two trains per hour. However, LTK only made qualitative remarks regarding the service and did not say whether or not it would be temporally possible given Caltrain and high-speed rail service levels.

Upon reviewing the report, we conclude that there is a lot more service planning that needs to be done regarding the blended system. When stakeholders complained that the Transbay Transit Center was not included in the original blended service analysis, LTK’s response was a short analysis where the firm conceded that the blended system was possible, but with only two Caltrain trains per hour. On top of that, the simulation was not shown in a string-line chart so we cannot speak to the uniformity or desirability of the already-limited Caltrain headways. As we discussed in Chapter 3, two commuter services per hour per direction is, in the author’s opinion, probably not enough to satisfy the intense job and housing concentrations at the Transbay Transit Center.

4.5 Design of Transbay Transit Center and Platform Sharing

The current design of the Transbay Transit Center calls for two HSR platforms supporting four tracks and a single Caltrain platform supporting two tracks. These segregated platforms are due to differences in the planned rail vehicle types for HSR and Caltrain.

In Caltrain’s August 2014 board meeting, the agency claimed that Caltrain’s new electric multiple unit (EMU) train sets would be incompatible with those of California high-speed rail. Caltrain claimed that their floor threshold would be 25” above top of rail (ATOR) while CHSR’s would be 51” (PCJPB 2014b). This would make it impossible to share platforms; Caltrain reemphasized this point in an October 2013 board meeting (PCJPB 2013). In a November 2014 presentation to the board, Caltrain made their first indication that they would consider shared platforms with CHSR (PCJPB 2014). As noted in Chapter Two, the CPUC requirements pertaining to freight service that are in conflict with ADA requirements regarding minimum gaps stand as a major obstacle to level boarding, and by extension, shared platforms.

In December 2014, however, in a presentation to the Transbay Transit Center Citizen Advisory Committee, the Transbay Joint Powers Authority Board of Directors made their most serious commitment yet to platform sharing. The Board cited improved train storage, delay recovery, commuter rail capacity, and flexibility as the main reasons to pursue shared platforms. In January and February 2015, the Caltrain will make a “tradeoff assessment and plans to arrive at a policy decision by May 2015” though at the time of our writing, this decision has not been made (PCJPB 2015).

Caltrain’s equipment for Baby Bullet services are relatively young. As a result, Caltrain intends to keep using that equipment after electrification in what they describe as an “interim period” of mixed electric and diesel operations. Caltrain recently “doubled down” on this strategy by committing $15 million to purchase 16 surplus Bombardier rail cars

[13] Ironically, the mid-line overtake would necessitate four-tracks through the communities that rallied against the original planned four-track system in favor of the mostly two-track blended system.

(the same type as the Baby Bullet cars) from Metrolink. These cars will be used to extend trainsets on peak hour trains (PCJPB 2014f).

Level boarding presents a challenge for this interim period because of discrepancies in platform heights; if platform heights were raised to accommodate level boarding on the EMU trains, the diesel equipment could no longer function due to a lower door height that the level of the platform. Caltrain's latest proposal is to run the diesel trains at Baby Bullet stations and provide local service with EMUs (2014). Mini-high platforms could be used to serve EMU trains at the low-platform stop (MBTA 2015). In the next chapter, we will discuss the impact of platform heights and shared platforms on service plans not just on the Peninsula but throughout California.

4.6 Case Study: Metro-North Railroad and Acela on the Northeast Corridor

Metro North Railroad and Amtrak operate in a shared corridor on the New Haven line that is similar in some ways to how the blended system will function in California[14]. Metro North, the commuter line providing service upstate New York and Connecticut and New York City, owns the line (along with Connecticut Department of Transportation) between the New York City border and New Haven, Connecticut. This line is also part of the Northeast Corridor between New Rochelle, New York and New Haven; in addition, then to Metro-North's 125,000 weekday commuters, Amtrak operates its profitable *Northeast Regional* and *Acela* trains along the line was well.

Like Caltrain and future California HSR on the Peninsula, the Northeast Corridor portion of the line is congested. Amtrak has the right to operate trains that do not cause "undue interference" with Metro North trains and reimburses Metro North for costs relating to maintenance and electricity on a car-mile basis. Metro North does not profit from Amtrak services nor does it receive a premium for peak hour or express Amtrak services, even though Amtrak is able to capture additional revenue from those trains. While Metro North will accommodate additional Amtrak trains if the schedule permits (and will modify its schedule by a few minutes in either direction if necessary), Amtrak is not allowed to add trains that cause undue interference to Metro North operations. On the same token, even though they own the corridor, Metro North is not permitted to act unilaterally and add trains that disrupt Amtrak services. These train slots are held by the two railroads indefinitely, which makes significant timetable changes cumbersome. Mr. Mel Corbett of Metro North Railroad, said if he had a blank slate in California, he would encourage a more "business-like arrangement" that included some sort of capacity charge that reflects express services that displace other train services or peak-period services that are more valuable to both commuter and intercity railroads (and by extension, commuter and intercity passengers).

4.7 Separating Infrastructure from Operators

Given the new concept of blended service and the challenge of Caltrain and HSR competing for the same track space, there is an opportunity to reorganize the institutional structure of the corridor to allow for separation of infrastructure from operations. An important consideration that has not been fully addressed to date is how the Transbay Transit Center fits into the HSR-Caltrain puzzle from an institutional perspective. The ultimate form of this arrangement will affect the operational structure of the Peninsula Corridor, and to some extent, the rest of the California HSR network.

[14] We would like to thank Mr. Mel Corbett of Metro North Railroad for providing the information described in this section.

4.7.1 Transbay Transit Center: Funding and Ownership

The $1.9 billion Transbay Transit Center (Phase 1) and $3 billion Downtown Rail Extension (Phase 2) are using a variety of funding sources, the largest of which is $2.7 billion in revenue from a special taxes, and impact fees from developers building in a designated Transbay Redevelopment District surrounding the terminal (Transbay 2011). Additionally, the Transbay Joint Powers Authority (TJPA) received $400 million in funding from the FRA's High Speed Intercity Passenger Rail Program for Phase 1 of the project (CHSRA 2015). Bay Area toll revenues, federal New Starts funds, and contributions from the City of San Francisco are expected to make up the rest of the funding for the two phases (Cauthen and Lydon 2015). Two notable entities absent from this list are the California High-Speed Rail Authority and the Peninsula Corridor Joint Powers Board (Caltrain)—the ultimate users of the Downtown Rail Extension. Undoubtedly, this will offer the TJPA some independence from the two operators; however, the Peninsula Corridor Joint Powers Board does have a seat on the TJPA Board of Directors[15]. The rail extension project is currently slated to be a design-bid-build, but the TJPA is also investigating the feasibility of a public-private partnership model (Transbay 2014).

At the time of writing, according to the author's conversations with the authority, the ownership of the Transbay Transit Center and Downtown Railroad Extension is going to remain with the Transbay Joint Powers Authority (or the private partner should the project proceed as a public-private partnership). Separation of infrastructure ownership from the operating railroads could lead to a more transparent allocation of infrastructure capacity; that is, each operator would understand why or why not it is awarded a train slot. This transparency could allow a high-speed operator to understand the risks associated with running trains on the Peninsula Corridor. Since the Transbay Joint Powers is going to own the tracks of the Downtown Extension as well as the Transbay Transit Center itself, it might facilitate separation of infrastructure ownership from operations along the entire corridor.

4.7.2 The Transbay-Penn Station Analogy

The Transbay Transit Center has been hailed as the "Grand Central Station of the West," though from a transit-operations perspective, a more apt name would be the "Penn Station of the West" (Transbay 2011). Metro-North Railroad, for the time-being, enjoys the luxury of being the sole operator of inter-city rail services at Grand Central. Penn Station, on the other hand, hosts not two, but three different rail operators and is experiencing capacity constraints on its own. The author's research colleague, Rebecca Heywood, is taking an in-depth look at developments at Penn Station and—similar to this thesis—how capacity improvements have impacts at the urban, metropolitan, and mega-regional level. Ms. Heywood and the author have collaborated in the development of the following table to help compare the "Penn Station of the West" with its century-old neighbor:

Table 4-1: Comparison of Tranbay Transit Center and Penn Station

	Transbay Transit Center	Penn Station
Platforms/Tracks	3/6 (potentially shared-use)	11/21 (some shared-use, some exclusive)
Owner	Transbay Joint Powers Authority or Private Partner	Amtrak
Operating Railroads	California HSR, Caltrain	Amtrak, New Jersey Transit (NJT), Long Island Railroad (LIRR)
Operating	Caltrain and HSR will enter and	NJT enters from Hudson River Tunnels

15 The rest of the TJPA board of directors includes representatives from the San Francisco Board of Supervisors, the office of the Mayor of San Francisco, AC Transit (the main bus tenant in the terminal), and the San Francisco Municipal Transportation Agency

Configuration	exit on three tracks (with interlockings) from the south	(west), LIRR enters from the East River Tunnels (east), and Amtrak accesses from Empire Connection (west) as well as Hudson/East River Tunnels
Dispatching/Control	TBD—likely Caltrain since Caltrain already dispatches 51 miles of corridor south of terminal	Amtrak (for Amtrak and NJT) and LIRR (for LIRR trains)
Peak Hour Service Levels	6 HSR, 2 commuter (planned)	8 Amtrak, 39 NJT, 26 LIRR (current)
How do tenant railroads add additional train services?	TBD	Amtrak and LIRR control dispatching; permanent new service unclear, limited by tunnel capacity
Passenger Facility Charges	TDB—TJPA suggested a small surcharge on tickets for Caltrain and California High Speed Rail	Unknown
Planned Improvements	Tail tracks and train box extension to allow for storage of additional trains	• Gateway Tunnels to supplement Hudson River tunnels • Addition of Metro-North Railroad service (due to East Side Access in Grand Central Terminal and shift of some LIRR trains to Grand Central) • Two new concourses for passenger circulation
Key Capacity Concerns	Caltrain demand will exceed supply with only 2 trains per hour per direction; shared platforms	No additional capacity in Hudson River Tunnels, concerns about stability of tunnels due to Hurricane Sandy damage

Source: Author and Rebecca Heywood

One of the key differences between the Transbay Transit Center and Penn Station is the ownership structure. There is not a "Penn Station JPA" that might act as an unbiased arbiter of rail capacity in the terminal; Amtrak is the owner and exercises its ownership power at the expense of New Jersey Transit. Since the Hudson River Tunnels are at capacity during the peak hours, Amtrak does not currently accept new services from New Jersey Transit, forcing many commuters to transfer to the select New Jersey Transit trains that run direct to Penn Station on the other side of the Hudson River at Secaucus Junction or Hoboken Terminal. Because the tunnels are at capacity (at 24 trains per hour), it becomes a zero-sum game: if Amtrak wanted to add a train across the Hudson during the peak (e.g. an extra Northeast Regional, Acela, or any of the 8 other Amtrak routes that use the Hudson River tunnels), New Jersey Transit would have to remove one of their peak trains from the system.

Transbay will likely face capacity challenges of its own; but fortunately as of yet, the two operators still have time to find a better method for sharing control of capacity allocation in the system. If Caltrain has too much control, it will be difficult for the high-speed rail authority to find a private concessionaire to finance the first phase of the high-speed project. If the CHSRA has too much control, commuters on the Peninsula and in San Francisco will lose access to the most valuable (in terms of convenience to residents and jobs) station in the Caltrain system.

In the next section, we imagine how operators will respond to an infrastructure owner such as the Transbay Joint Powers Authority and how that response affects the level of service experienced by intercity travelers and commuters alike.

4.8 Analyzing the Financial Relationship between Railway Industry Players in Shared Railway Systems

The following section is adapted from TRB-2015-15-1697, a paper written by this author, Maite Pena-Alcaraz, and Alex Prodan, which examines the train operator's perspective in shared railway systems. The original paper used the Northeast Corridor as a case study (Levy et al. 2015). This thesis adapts the model to examine operations on the Peninsula and at the Transbay Transit Center. Parallels can be drawn between the Northeast Corridor (NEC) between Boston and Washington, D.C., and the Peninsula Corridor. Both corridors host (or will host) a mix of freight, commuter, and intercity passenger rail and both have multiple owners (assuming PCJPB does not take ownership of the Downtown Extension).

The federal 2008 Passenger Rail Investment and Improvement Act (PRIIA) addressed Amtrak's role as the majority owner and sole intercity passenger operator on the corridor (PRIIA 2008). Section 208 requires Amtrak to develop a plan to bring the corridor to a state of good repair. Section 212 of the law requires the establishment of the Northeast Corridor Infrastructure and Operations Advisory Commission. The role of this commission is to develop a plan for the future of the corridor, including a plan to charge infrastructure track-access charges (fees). Amtrak must not cross-subsidize commuter, intercity and freight services, and each service must pay the costs incurred by operating that service on the network (can be interpreted as "operating/marginal/direct cost recovery"), as well as proportionate costs that can be distributed to more than one service (can be interpreted as "fixed cost recovery"). The NEC Commission released their proposed policy framework that allocates costs based on a variety of metrics such as train moves, train miles, and gross-ton miles, in January 2015. The law, however, does not change the status quo of having Amtrak be the sole intercity passenger operator on the NEC.

4.9 Modelling the Train Operator's perspective

This model analyzes a corridor capable of supporting high-speed and commuter rail services along a corridor. This is not unlike the NEC or the future Peninsula corridor. Rail capacity is fixed in our medium-term time horizon; that is, there is no opportunity to make infrastructure improvements that will increase the maximum train throughput of the corridor. Train operators may be able to adjust their capacity to better serve the users in this time frame. For the purposes of this thesis, freight traffic is ignored, but we acknowledge the importance of both passenger operators reaching an agreement with Union Pacific on the corridor.

4.9.1 System Players

Five main players are being considered in this analysis: society, the government, the regulator, the IM, and train operators. Behind each entity's actions are differing motivations.

Society represents the view of the best interests of the entire population. On the Peninsula corridor, this represents intra-city California travelers as well as commuters.

The ***Government*** is usually the investor in the infrastructure and does not necessarily represent the same views as society. The government is responsible for creating laws and regulations that govern different aspects related to operating the railway, from financial relationships between different players to safety of operations. On the Peninsula Corridor, this might be the California State Legislature or the three counties—San Francisco, San Mateo, and Santa Clara—through which the corridor traverses.

The ***Regulator*** is responsible for enforcing existing laws and regulations. In the EU, each state's regulator is responsible for ensuring that the state's national legislation is followed by all other entities. In the US, the Federal Railroad Administration regulates safety on US railroads.

The ***Infrastructure Manager (IM)*** is the entity that, at a minimum, is responsible for managing and maintaining the infrastructure. As of today, the plan is for the Transbay JPA or its public-private partner to be the infrastructure manager for the Downtown Extension and the Transbay Transit Center. For the purposes of this thesis, we will also imagine an infrastructure manager for the entire Peninsula Corridor, a fictitious "Peninsula Infrastructure Management Organization" (PIMO).

Train Operators (TOs) are the entities providing passenger or freight services. They may or may not receive subsidies, but it is assumed that any entity (perhaps, the counties of San Francisco, San Mateo, and Santa Clara) subsidizing rail operators is separate from the IM; that is, the IM will be under no obligation to favor one operator over another.

In the United States, commuter train services are sometimes planned and offered for bid by railway agencies while railway operators provide operations staff, perform maintenance, and collect fares. By our definition, the Peninsula Corridor Joint Powers Board is an example of a railway agency, while Veolia Transportation or TransitAmerica Services are examples of railway operators. For the purposes of this paper, all references to TOs would refer to both players; the operator of commuter rail services on the Peninsula will be referred to as "Caltrain."

4.9.2 Understanding the Relationship between TJPA/PIMO, Caltrain, and the California HSR Operator

One of the key goals of this section is to understand how different players respond to capacity access charges that a third party might impose on Caltrain or CHSR. Therefore, the focus in our overall work will be primarily on the relationship between Train Operators and Infrastructure Manager (CHSRA and Caltrain operating and TJPA or PIMO managing). We will develop a model for the relationship between train operator and infrastructure manager and then apply it to both a high-speed rail operator (CHSR) and a commuter operator (Caltrain).

4.10 Modelling the Train Operator-Infrastructure Manager Relationship

This paper analyzes a corridor capable of supporting high-speed or long-distance passenger service and freight rail service as well as commuter rail service around large urban areas along the corridor. This is an accurate representation of the Peninsula Corridor, with the exception that high-speed rail will go beyond the corridor's limits In this model, capacity is fixed in our medium-term time horizon; that is, there is no opportunity to make infrastructure improvements that will increase the maximum train throughput of the corridor. However, train operators may be able to adjust their capacity (e.g. adding to service levels) to better serve the users (traveling public) in this time frame.

In order to analyze TO profits and cash flows, a simplified model that captures main revenue and cost streams is proposed for the medium-term time-horizon (PPIAF 2011). TOs are assumed to be rational entities, and will only operate if their cash flows (after recovering capital costs at an adequate rate of return) are positive in the medium term; for a subsidized entity such as Caltrain, government subsidies are included in these cash flows. TOs are driven by profit maximization. TOs' main decisions are 1) the number of services that they are willing to operate, 2) their willingness to pay to access the infrastructure, and 3) the fares that they will charge the final users. We acknowledge that passenger rail operators may have public service requirements that dictate minimum frequency levels, service spans, or fare ceilings; but nevertheless, profit maximization are the operators' objective.

While there is intermodal competition (e.g. commuter rail versus automobile traffic or bus transit service), it is assumed that there is no intra-modal competition: there is no direct competition for traffic in the corridor between operators. The only time at which that operators compete directly is when competing for available track capacity where they can run scheduled services. We also assume that the services, offered by different TOs, are not substitutes. In the case of California, this is mostly true, though the CHSRA predicts that a small share of ridership (about 1500 riders per day) will be concentrated in Northern California, and consequently, competing almost directly with Caltrain (CHSRA 2014).

4.10.1 Train Operator Profits and Cash Flow

TO profits and cash flow can be determined by analyzing TO revenues and costs for a given number of trains, n. A TO faces cost of accessing the tracks, $ac(n)$ or track-access charges, some fixed costs, fc, such as the cost of buildings and the purchase of trains, and variable costs of operating trains, $vc \cdot n$, such as fuel, personnel, train maintenance, and train lease, if trains are being leased.

The two main sources of revenue come from the government, s (subsidies), and from transporting users (passengers). The revenues obtained from transporting users can be determined by multiplying the fare (f) by the demand served. The demand served is limited by either the capacity (reduced by a reasonable average load factor) of the trains ($c \cdot n$) by user demand (d). According to literature, user transportation demand depends fundamentally on the fare (f), the frequency of the service (proportional to $\frac{1}{n}$), and the travel time (tt) (Bebiano 2014). While intercity passengers are typically more sensitive to the fare and the travel time, commuter passengers are typically more sensitive to the fare and the frequency.

Summarizing, the costs and revenues of a TO can be determined using the following formulas:

$$Cost(\boldsymbol{n}, \boldsymbol{ac}) = fc + vc \cdot \boldsymbol{n} + \boldsymbol{ac}(\boldsymbol{n}) \tag{4-1}$$

$$Revenues(\boldsymbol{n}, \boldsymbol{f}) = s + \boldsymbol{f} \cdot min(d(\boldsymbol{f}, \boldsymbol{n}, tt), \boldsymbol{n} \cdot c) \tag{4-2}$$

Here, bold letters are used to denote the main TOs' decision variables. Please note that some of these variables may be pre-determined or conditioned by regulations. For instance, the fare of commuter services is typically set by the government (or the operating agency which is accountable to the government). Likewise, access charges under cost-allocation and priority-rule mechanisms are fixed inputs for TOs.

The TO level of service and the fares, given the access charges (ac), can be determined maximizing profits:

$$\max_{\boldsymbol{n},\boldsymbol{f}} revenues(\boldsymbol{n}, \boldsymbol{f})\text{-}costs(\boldsymbol{n}) \tag{4-3}$$

$$\max_{\boldsymbol{n},\boldsymbol{f}} s + \boldsymbol{f} \cdot \min(d(\boldsymbol{f}, \boldsymbol{n}, tt), c \cdot \boldsymbol{n}) \text{-} fc\text{-}vc \cdot \boldsymbol{n}\text{-}ac(\boldsymbol{n}) \tag{4-4}$$

Equation (4-4) is equivalent to: $\max_{\boldsymbol{n},\boldsymbol{f}} \boldsymbol{f} \cdot \min(d(\boldsymbol{f}, \boldsymbol{n}, tt), c \cdot \boldsymbol{n}) \text{-}vc \cdot \boldsymbol{n}\text{-}ac(\boldsymbol{n})$.

The TO willingness to pay to access the infrastructure, given the level of service and the fare (*n*,*f*), can be determined ensuring that the resulting cash flow is positive:

$$revenues\text{-}costs(\boldsymbol{ac}) \geq 0 \tag{4-5}$$

Note that capital expenditures and financing costs are also required to compute cash flows. However, we will assume that TOs have almost no CAPEX and negligible financing costs.

$$s + f \cdot \min(d(f, n, tt), c \cdot n) \text{-} fc\text{-}vc \cdot n\text{-}\boldsymbol{ac}(n) \geq 0 \tag{4-6}$$

$$\boldsymbol{ac}(n) \leq s + f \cdot \min(d(f, n, tt), c \cdot n) \text{-} fc\text{-}vc \cdot n \tag{4-7}$$

4.11 Results and Implications: A California HSR operator’s perspective

The previous formulas can be further extended in different scenarios to understand the behavior of different types of TOs operating in a shared railway system. In this first scenario, we will use the these formulas to determine service level and fare when passengers’ demand is a linear function of the fare, with some elasticity e

In this case, the elasticity is defined as $e = \text{-}\frac{\Delta d / d_0}{\Delta f / f_0} = \text{-}\frac{(d\text{-}d_0)\cdot f_0}{(f\text{-}f_0)\cdot d_0}$, and the demand as a function of the fare can be determined using $d(f) = \text{-}e \cdot \frac{d_0}{f_0} \cdot f + (1+e) \cdot d_0$.

4.11.1 Calculations:

The optimal level of service and fare (n^*, f^*) to maximize profits can be determined separating the problem in two subcases:

Case 1: If $d(\boldsymbol{f}, \boldsymbol{n}, tt) \geq c \cdot \boldsymbol{n}$**,** i.e., if the passenger demand is constrained by the capacity of the trains scheduled (as is likely the case with future Caltrain scenarios) then we can start computing the optimal fare for a given level of service $f^*(n)$. In this case, obtaining the fare that maximizes profits is equivalent to obtaining the fare that maximizes revenues. That means that maximizing fare with the objective of ensuring a demand $(d(f) = \text{-}e \cdot \frac{d_0}{f_0} \cdot f + (1+e) \cdot d_0)$ still higher to or equal than the capacity $(c \cdot n)$. Doing the computation we obtain:

$$f^*(n) = arg \max_{f} f \cdot c \cdot n : d(f) \geq c \cdot n \leftrightarrow f^*(n) = \frac{(1+e)}{e} \cdot f_0 \text{-} \frac{c \cdot f_0}{e \cdot d_0} \cdot n \tag{4-8}$$

Given this, the optimal level of service can be obtained maximizing profits:

$$n^* = arg\max_n s + f^*(n) \cdot c \cdot n\text{-}fc\text{-}vc \cdot n\text{-}ac(n) \quad (4\text{-}9)$$

Assuming that track-access charges are linear, as they essentially are in the Northeast Corridor ($ac(n) = ac_f + ac_v \cdot n$), the optimal level of service is either:

$$n^* = 0, f^* = 0 \quad (4\text{-}10)$$

$$n^* = \frac{(1+e)\cdot d_0}{2\cdot c}\text{-}\frac{e\cdot d_0\cdot(vc+ac_v)}{2\cdot c^2\cdot f_0}, f^* = \frac{(vc+ac_v)}{2\cdot c} + \frac{(1+e)\cdot f_0}{2\cdot e} \quad (4\text{-}11)$$

Note that these computations assume that any level of service is possible. Slight adjustments should be made to obtain the optimal solutions considering that possible service levels are discrete (integer number of trains).

Case 2: if, conversely, $d(\boldsymbol{f}, \boldsymbol{n}, tt) < c \cdot \boldsymbol{n}$**,** i.e., if the passenger demand is less than the train capacity, we can still compute the optimal fare for each level of service $f^*(n)$. Again, maximizing profits is equivalent to maximizing revenues. That means to maximize the revenue with the objective of ensuring a demand ($d(f) = \text{-}e \cdot \frac{d_0}{f_0} \cdot f + (1+e) \cdot d_0$) lower than the capacity ($c \cdot n$). Doing the computation we obtain:

$$f^*(n) = arg\max_f f \cdot d(f) \colon d(f) \le c \cdot n \leftrightarrow f^*(n) = \frac{(1+e)}{2e} \cdot f_0 \quad (4\text{-}12)$$

Given this, the optimal level of service can be obtained maximizing profits:

$$n^* = arg\max_n s + f^*(n) \cdot d(f^*(n))\text{-}fc\text{-}vc \cdot n\text{-}ac(n) \quad (4\text{-}13)$$

Assuming again that track-access charges are linear ($ac(n) = ac_f + ac_v \cdot n$), we obtain that the optimal level of service is:

$$n^* = \left\lceil\frac{(1+e)\cdot d_0}{2\cdot c}\right\rceil, f^* = \frac{(1+e)\cdot f_0}{2\cdot e} \quad (4\text{-}14)$$

Summarizing, the optimal level of service and fare (n^*, f^*) to maximize profits are either:

$$n^* = \left\lceil\frac{(1+e)\cdot d_0}{2\cdot c}\right\rceil, f^* = \frac{(1+e)\cdot f_0}{2\cdot e},$$

$$n^* = \frac{(1+e)\cdot d_0}{2\cdot c}\text{-}\frac{e\cdot d_0\cdot(vc+ac_v)}{2\cdot c^2\cdot f_0}, f^* = \frac{(vc+ac_v)}{2\cdot c} + \frac{(1+e)\cdot f_0}{2\cdot e}, \text{ or} \quad (4\text{-}15)$$

$$n^* = 0, f^* = 0$$

4.11.2 Implications:

Despite the complicated mathematical expressions, these formulas can be distilled to obtain some implications:

1. When variable costs are small with respect to the fares that users can afford, the optimal solution is to maximize revenues and offer the minimum number of trains that allow serving all the demand for the optimal fare.
2. When variable costs are comparable to the fares that users can afford, the optimal solution is a trade-off between maximizing revenues and covering variable costs. In this case, the capacity provided by the TO should be optimized in such a way that most passenger demand is served without providing excess train capacity.
3. Finally, in those cases in which the users cannot viably accept a fare level that allows TOs to cover at least the variable costs, the TO should not operate any train.

We can illustrate these points with an example using data derived from CHSRA's operational cost estimates.

According to the CHSRA's 2012 business plan, it faces fixed operational (direct) costs of $fc = \$400.2\ million$ per year ($fc = \$1.124\ million$ per day) and variable operational costs of $vc = \$41.6\ million$ per train and per year ($vc = \$10{,}400$ per train and per day). The elasticity of the demand is estimated to be equal to $e = 0.67$ (Lago et al, 1981). The CHSRA predicts an average fare to be \$47.68 per rider (averaged across all riders, regardless of distance traveled) and the level of service to be $n = 219$ trains per day, with a realized demand of $d_0 = 76{,}400$passengers per year ($d_0 = 76{,}400$ passengers per day), and with an average train capacity of $c = 760$ passengers, with 80% load factor (Morrison 1990). We found the 80% load factor, a measure of train seat capacity utilization, to be a reasonable input given CHSRA's business plans and Amtrak's NEC 2015 financial plan and operating data (CHSRA 2014, Amtrak 2014). Again, for the purposes of our equations, since the CHSRA expects the train to be at worst revenue neutral, the operator does not receive operating subsidies, meaning $s = 0$.

Figure 4-2: California High-Speed Rail Operator Response to Access Charges

Figure 4-2 represents the CHSR operator's expected profits when the strategies presented in Equation 4-15 are used to determine the level of service and the fare. We then compare the profits and service-decisions obtained with these strategies to those obtained using planned levels of service and expected average fares.

While these numbers are based off of forecasts and omit many details regarding actual operational procedures (e.g., number of services is not a good indicator of the mix of services or whether or not the trains arrive at ideal times), it does reasonably model the impact of an access charge for part of the infrastructure. As the access charge increases, the profit-maximizing operator will decrease services and increase fares; in essence, the operator passes PIMO's "Peninsula tax" on to its customers. This would open up space at the Transbay Transit Center for commuter rail services at the expense of intercity California travelers. The model shows that the access charge quickly becomes unbearable for the projected operator's business plan. Even so, if San Francisco represented enough revenue, the operator may still choose to concentrate its services there.

One shortcoming of this model is that is does not take into effect the elasticity of frequency. Frequency is important up to a point. Again, we draw a parallel to the air travel market and refer to Belobaba's *Global Airline Industry* text:

> There exists a "saturation frequency" in each market, defined as the point at which additional frequency does not increase demand, even for business travel). For example, in the short-haul Boston–New York shuttle market, two competitors currently each offer non-stop flights every hour (one at the top of the hour and the other at 30 minutes past each hour), such that flights in this market depart every half hour. It would measurable positive impact on the total volume of demand in this market (100).

We believe that a profit-maximizing operator will survive longer by cutting some of its Peninsula service. The operator will do this because, even with poor frequency, the value of the direct-to-downtown service is attractive to the business traveler. Once the access charge crosses a certain threshold—in our model about $35,000 per train per day ($1028/train/day for the planned service levels)—the high-speed operator is no longer profitable and the Peninsula would stop seeing high-speed rail service.

This model also shows what might happen if there were some kind of CHSRA-imposed fare ceiling on the operator. Because our price elasticity is less than -1, the lower fares would stimulate more demand and the operator would need to offer more services to capture additional passengers and make up the revenue loss. This would, in turn, put more demand on the limited Peninsula infrastructure. Similar results are obtained for a broad range of fare elasticity values: lower elasticity representing business users willing to pay high fares to ride convenient CHSR services, and higher elasticity representing additional users that start to ride CHSR instead of other transportation alternatives.

4.12 Results and Implications: The Commuter Rail Operator's Perspective

Next, we will take a look at a slightly different scenario for a commuter rail operator's perspective. Here we assume fares are constant either because demand is very elastic to fares, i.e., almost all the demand is lost if a user's fare is above certain fare threshold, or fares are set by the government(f_0)) Instead of demand being dependent on fares, demand depends on level of service.

In this scenario, the elasticity to the level of service can be defined as $e_n = -\frac{\Delta d/d_0}{\Delta h/h_0}$ where h is the average headway between consecutive trains. Since the headway is proportional to $1/n$, the elasticity can also be computed as $e_n = -\frac{(d-d_0)\cdot n}{(n-n_0)\cdot d_0}$. Therefore, the demand can be determined by $d(n) = (1+e_n)\cdot d_0 - \frac{e_n \cdot d_0 \cdot n_0}{n}$.

4.12.1 Calculations:

The optimal level of service and fare (n^*, f^*) to maximize profits can be determined repeating the same type calculations carried out in the high-speed operator example. Assuming again that track-access charges are linear ($ac(n) = ac_f + ac_v \cdot n$), it is determined that the optimal level of service that a TO can operate would be either:

$$n^* = \sqrt{\frac{f\cdot e_n \cdot d_0 \cdot n_0}{vc + ac_v}},$$

$$n^* = \frac{(1+e_n)\cdot d_0 \cdot n_0}{2\cdot c} - \frac{\sqrt{(1+e_n)^2 \cdot d_0^2 \cdot n_0^2 - 4\cdot c \cdot e_n \cdot d_0 \cdot n_0}}{2c}, \text{ or} \qquad (4\text{-}16)$$

$$n^* = 0$$

The operator's choice of one level of service over the other would depend on how revenues and cost compare. If revenues obtained from fares are much higher than variable costs, then the optimal strategy to maximize profit would be to maximize revenues. If revenues are comparable to variable costs, the optimal strategy would be to ensure that there is no excess-capacity on the trains. Finally, if variable costs are higher than the revenues per train, the TO should not operate any trains.

Note that this level of service is independent of the level of subsides and the fixed costs (from operations and access-charges). These values would only affect to whether the TO's cash flow are positive and hence the TO can sustainably operate these level of service.

4.12.2 Implications

This scenario is representative of the situation of the commuter rail train operators, including an operator such as Caltrain. Caltrain faces fixed operational (direct) costs of $fc = \$142{,}614$ per day and variable operational costs of $vc = \$2{,}406$ per train and per day. Fixed and variable costs were determined by checking Caltrain historic service levels and plotting against operating costs from Caltrain operating budgets using the assumption that the relationship is linear as follows:

$$Total\ costs = fc + vc{*}trains/day$$

Given that this is an oversimplification (variable costs are also dependent on passenger demand, staffing levels, etc.), we adjust fixed costs so that Caltrain runs its services with a balanced operating budget (as it does today).

The elasticity of the demand with respect to the headway (the inverse of frequency) is estimated to be equal to $e = 0.41$ (Lago et al 1981). In 2014, Caltrain's average fare was \$4.62 (for simplicity, operating revenues divided by ridership), the level of weekday service averaged $n_0 = 92$ trains per day, with a realized demand (daily rideship) of $d_0 = 52{,}611$

passengers per day. Each train's seated capacity is $c = 650$. Again, we realize that Caltrain has the potential to carry much more than its seated capacity (several weekday trains today do in fact have loads as high as 120% of seated capacity and seats do "cycle" as average trip length on the 51-mile corridor is just over 20 miles), but we simplify to reach an interpretable model. Subsidies are $s = \$121\ thousand$ per day and are based on historical operating budgets.

Figure 4-3 compares current Caltrain operating profits with the expected profits when the profit maximizing strategy presented in equation 4-16 is used to determine the level of service. The results show that higher profits can be unlocked by reducing the number of services, especially when variable costs increase due to track-access charges. Please note that even under the profit maximizing strategy, despite the subsidy, Caltrain would not be able to operate if access charges exceed $575 per train per day, since the variable costs of operating the train would be higher than the revenues obtained. As a result, operating a train would only increase the cost burden for the system.

Figure 4-3: Caltrain Response to Access Charges

This model validates what we already know to be true: Caltrain operates a lot of trains that serve a public-utility benefit. Mid-day and late night trains are unlikely profitable, but Caltrain's mission is to meet the growing mobility needs of the San Francisco Bay Area region, so these trains are operated in spite of the loss to the agency.

Unlike the model for the high-speed operator, we hold fares constant and only assume an elasticity to travel time under the assumption that 1) the commuting public likely values this travel time, and, by extension, frequency (see Chapter 3) enough that fares have a minimal effect on ridership choices[16], and 2) significantly raising fares is politically challenging for a government agency due to intense public opposition to fare increases.

As the access charge to run trains increases, the "Caltrain-for-profit" strategy ceases to be sustainable either. At approximately $250 per train per day, the access charges force the already-subsidized operator into operating losses. While these values are approximate given the simplifications we make for the two models, we see an order-of-magnitude difference between the high-speed rail operator's and the commuter rail operator's willingness to pay. The California HSR operator has two orders of magnitude greater the ability to pay for an access charge to San Francisco or the Peninsula. How will an infrastructure manager react to this knowledge? We will now discuss the impact of this disparity from the infrastructure manager's point of view.

4.13 The Infrastructure Manager's Perspective

As noted earlier, the benefit of an independent infrastructure manager such as the fictitious "PIMO" or the planned Transbay Joint Powers Authority or the Transbay private-public partner is that independence lessens the likelihood of a Penn Station scenario where one operator gets preference over the other. However, even with a lack of institutional bias, any third-party operator would likely prefer to do business with an operator with a two orders of magnitude greater

[16] The *Silicon Valley Business Journal* reported that the average annual income of a Caltrain rider is $117,000 suggesting that time is likely a more important consideration than fare (Weinstein 2014)

willingness to pay than a competing operator. Before we continue, however, we will perform a quick estimation of the Transbay Joint Powers Authority's (or PIMO's) need in terms of access charge revenue.

We will use cost data from the CHSRA's 2012 business plan report and Caltrain's 2015 operating cost data to derive cost estimates for the infrastructure manager's costs:

Table 4-2: Operating Cost Estimate for Infrastructure Manager

Infrastructure Manager Cost Estimate			Transbay JPA		PIMO (SF Peninsula)	
		Cost/Unit	Count	Total	Count	Total
Shared Station	Each	$4.1M	1	$4,100,000	3	$12,300,000
Commuter Station	Each	$116,000	-	-	27	$3,132,000
Track Maintenance	Route Miles	$200,000	2	$400,000	110	$22,000,000
Traction Power	Train-Set Miles	$7.45	147,570	$1,099,393	5,789,265	$43,130,024
Administration & Support	8% of cost	NA	-	$447,951	-	$6,444,962
Contingency	5% of cost	NA	-	$279,970	-	$4,028,101
			Total Annual Cost	**$6,327,314**		**$91,035,087**

Sources: CHSRA Operation and Maintenance Methodology from 2012 Business Plan, PCJPB, Author's Calculations

Using this simple cost estimate, we can see how many trains per day the infrastructure owner needs to schedule given a certain access charge. Keep in mind that the values shown in the next table are only the access charges needed to recover enough to cover costs; that is, at these service levels, the infrastructure manager would not be able to make investments in infrastructure expansion. Also, this table assumes that each train is charged an equal price for using the infrastructure at certain times. In Europe, capacity is often priced more expensively at peak periods and we might expect the same to hold true at the Transbay Transit Center.

Table 4-3: Estimated Trains/Day for an Infrastructure Manager to Fully Recovery Operating Costs

Access Charge per train	Transbay	PIMO
$50	347 trains/day	4988 trains/day
$125	139	1995
$250	69	998
$500	35	499
$1,000	17	249
$2,500	7	100
$5,000	3	50
$10,000	2	25
$15,000	1	17
$20,000	1	12
$25,000	1	10
$30,000	1	8
$35,000	0	7
$50,000	0	5
$75,000	0	3
$100,000	0	2

Source: Author's Calculations

Assuming neither operator has a budget constraint and desires to run the services outlined in their respective future operating plans, Table 4-3 shows that an infrastructure manager at the Transbay Transit Center would need to charge an average per train access fee of approximately $125 per train to recoup annual infrastructure costs. These charges are in line with similar charges on a per-train-mile basis to those in Europe according to a 2014 UIC report (Prodan and Teixiera). An infrastructure manager for the entire Peninsula would need to charge a fee much closer to $2500. We determined earlier that $2,500/train is out of Caltrain's price range given current costs, but current costs would be much lower if Caltrain did not have to maintain the 51-miles of infrastructure and stations it does today[17].

Returning to the $125 value, from an order of magnitude standpoint, we find that that access charge is well-within the capabilities for a high-speed rail operator, though it would be a significant burden for Caltrain at current subsidy levels. Even if we were to assume, however, that Caltrain could receive additional subsidies so that it could compete with a high-speed rail operator, there would need to be some form of check on the Transbay JPA to keep charges reasonable. A Caltrain operator with seemingly bottomless subsidies and a high-speed operator with a large willingness to may lead to both operators paying higher access charges. In his thesis, Sakamoto expressed a similar concern with Trenitalia, the state-owned railway in Italy, which could conceivably cover any losses with increased subsidies (2012). This "unbalanced competition" as Sakamoto warns, will be felt by the passengers in the form of increased ticket prices.

4.14 Implications of our Train Operator and Infrastructure Manager Models

Since we have determined it would be difficult for the high-speed rail operator and Caltrain to compete on a level playing field for access to the Transbay Transit Center due to the fiscal strength of the high-speed rail operator, perhaps there are other methods for allocating space. Why is it important that Caltrain have access to the Transbay Transit Center? If the high-speed operator can afford it, should it not have as much access as it wants? This author's response to that premise is that Caltrain service generates many of positive externalities including travel time and savings for individuals driving on adjacent freeways, the environmental and public health benefits of less congestion, and the agglomerative benefits of connecting San Jose and Silicon Valley with San Francisco. In a pure-willingness to pay (monetary) access charging scheme, those externalities are difficult to measure. And while it is true that high-speed rail also offers some of these positive externalities, we assume that high-speed rail riders will have other choices (e.g. take Caltrain from San Jose as a "last leg") that Caltrain riders may not necessarily have.

As an alternative to a monetary access charge, instead, for example, capacity could be allocated on a points instead of monetary basis. This would mean giving each operator a set number of "points" and then allowing operators to place bids for specific train using those points. This would conceivably allow a commuter rail operator like Caltrain to have a better opportunity in securing train-paths into the Transbay Transit Center than if it is only based on a pay-for-access policy. This would require some form of auction to allocate train paths according to an operator's interest in accessing the Transbay Transit Center at a particular time. However, because of the interdependence of one train path on the train paths of the other operators, there is potential for "gaming" an auction system by protecting certain time windows with high bids. Also, as Maite-Pena Alcaraz points out in her thesis, there is an inherent uncertainty associated with auctions (2015). This uncertainty means that operators will not know their rolling stock or staffing needs until the timetable is set. And for the commuters, this uncertainty will invariably be passed on in the form of changing timetables and non-uniform headways.

In Chapter Two, we reviewed European capacity allocation policies. In general, most European states create their timetables by negotiating conflicts and then referring to some sort of priority rule scheme if the conflicts cannot be

---

[17] The author is not able to estimate Caltrain's costs if it were an operations-only entity. However, at an author-estimated revenue of about $4,000 for a peak hour train (average fare x ridership), $2,500 would not be an insignificant about for the agency.

resolved. Priority rules (such as HSR takes priority during peak times) would undoubtedly solve conflicts between PCJPB and CHSRA and provide the California HSR operator with a better sense of the revenue risks associated with the blended system, but the fact remains that capacity into the Transbay Transit Center and along the line is constrained no matter how trains are scheduled.

The author would like to emphasize that freight access is not even considered in this model. While freight on the Peninsula appears to not be concerned with rail access during the peak period, it does create certain constraints during the midday that will be detrimental to the service quality provided by either Caltrain or HSR unless addressed.

4.15 Conclusions

In this chapter, we took an in-depth look at the levels of coordination between the CHSRA and PCJPB in developing the operational aspects of the blended system on the Peninsula. In short, given how high-speed rail will radically and forever change rail service on the Peninsula, there remains much work to be done in terms of capacity coordination between the two agencies. Though we reviewed some of the physical challenges of sharing track in Chapter Two, aside from the brief section on platform sharing and its effect on capacity, this chapter purely addressed capacity along the line and into and out of the Transbay Transit Center. Aside from access to the corridor and the terminal, the two agencies need to coordinate on other issues such as positive train control, California Public Utilities Commission regulations on freight traffic on electrified corridors, additional safety precautions and maintenance required for 125mph track, maintenance windows for track repairs, etc. In its comments on the 2014 CHSRA business plan, the California HSR Peer Review Group noted the PUC regulations for electric catenary and the potential redundant positive train control system as two key non-capacity related issues. The Peer Review Group goes on to say that these represent "near term decisions" that could be "made by the parties acting separately that would ultimately compromise the performance of the system." In July 2013, PCJPB published a Caltrain/HSR Blended System Planning Process chart, which is depicted in Figure 4-4

The chart lists the three documents produced by LTK Engineering thus far, implying that the next step in the process is the Service Plan Options. Caltrain has not yet published service plan options, but is planning on issuing a request for proposals for trainsets in July 2015 and award a contract in the winter of 2015/2016. This suggests that the agency has skipped level two ("Service Plan Options") at this point and is proceeding with fleet procurement (step 3, "fleet need"). This step, as we will discuss later in this thesis, is critical to achieving a truly integrated system and a step both agencies skip at their own peril.

In March 2015, the PCJPB hired a new CEO after their previous CEO retired. Their choice, former Redwood City councilman, Jim Hartnett, served previously on the boards of both the CHSRA and the PCJPB. Hartnett helped implement the 2009 MOU between the two agencies; writes in the *Palo Alto Weekly:*

> "When high-speed rail officials attended a meeting in Mountain View in November 2011 to discuss their new vision for the rail line, it fell to Hartnett to make the case for what is now known as a 'blended system.' . . .Hartnett called the new approach a 'rethinking of the whole high-speed rail approach.'"

The hiring of Hartnett reflects the increased importance of the blended system in the eyes of the PCJPB board of directors and inspires hope for even higher levels of collaboration between the agencies.

In reviewing the blended service feasibility study, we conclude that the presence of high-speed rail, as planned, severely degrades the service quality experience by Caltrain riders. Users experience increased travel times and highly variable headways between trains that lead to gaps in service and crowding. Adequate commuter access to the Transbay Transit Center has not been considered (the author does not consider two trains per hour per direction "adequate" given the concentration of residences and jobs noted in Chapter Three).

This discussion brought us to the Transbay Transit Center, a vital station for both operators. Fortunately, unlike Penn Station in New York, neither operator is going to outright own the station—it will be owned and managed by a third-party, the Transbay Joint Powers Authority. However, it is quite possible that the TJPA eventually cedes control to one of the operators or at least, prioritizes one operator over another. We employ a train operator financial model to emphasize that the high-speed rail operator will likely have a much higher willingness to pay for access to the terminal than Caltrain. Given the planned third-party arrangement, there is also an opportunity for a transition to a third-party operator for the entire Peninsula corridor. There is a lot of uncertainty in the ultimate management arrangement of the terminal; and regardless of the ultimate structure, unless there are codified capacity access rules, this uncertainty will remain.

This uncertainty might be enough to drive away a potential private investor in the project. Because the CHSRA's plan is to sell the right to construct (and then operate for profit) to a concessionaire, there needs to be significant revenue potential. Without the opportunity to fully capitalize on the prime location of the San Francisco station, this private investor takes on a revenue risk. Without a clear plan for the Peninsula Corridor, the CHSRA will have to finance that revenue risk; and potentially, the list of qualified private investment interest will shrink.

Finally, any kind of capacity allocation mechanism has network ramifications regarding service patterns and service levels for the rest of the high-speed rail operator's network. Access to the Peninsula Corridor or Transbay Transit Center will dictate timetables customers experience on the rest of the network.

This thesis suggests that coordination and *integration*, in the spirit of the blended system as written in the 2012 CHSRA Business Plan, and not competition, is paramount to the success of the shared-use concept on the Peninsula. When CHSRA and Caltrain compete against each other for access to track, whether it is on the Peninsula or only at the Transbay Transit Center, both operators lose. We want to emphasize the importance of planning service before making large infrastructure investment decisions such as platform heights or shared facilities. Because of the high capacity utilization of the Peninsula Corridor, the blended service concept demands this approach. In the next chapter, we will examine how certain infrastructure and service decisions impact the long term service goals of Caltrain, the high-speed rail operator, and the rest of the operators (i.e. Amtrak, Metrolink, and COASTER) in the Golden State.

5 Southern California Blended Service—A (Relatively) Blank Slate

In the third and fourth chapters, we discussed the blended system in Northern California. We now turn to Southern California where the CHSRA also plans to blended services with local commuter rail agencies. While both system "bookends" have many of the same issues, the ultimate form of the blended system in Southern California is much less defined than its northern counterpart

5.1 Introduction: Southern California Rail Overview

Like Chicago, Southern California is a rail transportation hub for both freight and passenger services. Because of the strength of the Ports of Los Angeles and Long Beach, both BNSF and Union Pacific have large rail yards and intermodal facilities in the region.

Los Angeles Union Station just north of downtown serves as the nexus of passenger rail service in the region: Four Amtrak long distance trains and the majority of regional commuter rail lines originate or terminate in the historic station. Additionally Los Angeles is the main hub for Amtrak California's Pacific Surfliner, the busiest Amtrak route outside of the Northeast Corridor (Malouff 2013). The owner of the station, the Los Angeles County Metropolitan Transportation Agency, runs three rail transit lines through the facility as well. A beautiful California Mission-Revival rail terminal, Union Station hosts more rail services than any other station in California and is the second busiest Amtrak station outside of the Northeast Corridor.

In this chapter, we will review the rail network in Southern California. We will discuss challenges that the commuter rail players face in the present, as well as ones that are anticipated in the future. We will discuss the importance of Los Angeles Union Station to the CHSRA and regional railroads as well as the opportunity presented by the Southern California Regional Interconnector Project.

Southern California faces many of the same issues that we see on the Peninsula Corridor in Northern California. Freight and passenger rail share tracks in many corridors and the California High-Speed Rail Authority plans to operate "blended service" on approximately the same length of corridor. While no single passenger line experiences the demand seen by Caltrain in Northern California, long single-track segments and high-levels of freight constrain capacity. In 2008, a collision between a Metrolink commuter train and a Union Pacific freight train on a single track section in northwest Los Angeles County led Congress to require railroads to implement positive train control on shared corridors by the end of 2015; the Southern California Regional Rail Authority, appropriately, leads most of the nation on this front (Stagl 2013). Los Angeles, along with Chicago, should be considered a national hubs of freight-passenger rail interaction.

This chapter is shorter than the chapter looking at Northern California blended service for the simple reason that most of the blended service planning has yet to be completed at the time of writing. In fact, at the time of writing, not even the route is set: the CHSRA is currently studying a new route alternative between Burbank and Palmdale, California through the Angeles National Forest (Weikel 2015).

5.2 The Commuter Rail Players: SCRRA (Metrolink) and NCTD (COASTER)

The Southern California Regional Rail Authority (SCRRA), another California joint powers authority, is responsible for the planning, design, and operation of Metrolink, the largest commuter rail service in Southern California. Governed by representatives from five of the six counties in which it operates—Los Angeles, Orange, Riverside, San Bernardino, and Ventura, SCRRA operates Metrolink services on a 512-mile network[18]. However, Caltrain, in the north, continues to have higher annual ridership on its Peninsula Corridor, which is 1/10 of the size of the Metrolink network. Metrolink carries fewer passengers for several reasons. First, jobs and housing in the region are not as concentrated around stations as they are on the Peninsula; last-mile transit connections are a challenge for many riders. Secondly, Metrolink's schedules

---

[18] Even though SCRRA operates Metrolink service into San Diego County to Oceanside, San Diego County is not represented on the Board of Directors

provide less service frequency than Caltrain—its relatively minimal "reverse commute" service make it an unattractive service for commuters travelling away from downtown Los Angeles in the morning. Finally, because Metrolink runs a majority of its trains on a network owned by freight railroads (a significant portion of which is single-tracked), delays can create reliability and service challenges, though Metrolink has improved on-time performance—in part due to capacity improvements—in recent years (Gbenekama 2012).

In February 2014, Metrolink became the first commuter rail in the United States to use a positive train control system on its network (SCRRA 2014). The project uses Wabtec Corporation's Interoperable Electronic Train Management System (I-ETMS), a vendor agreed upon by the four Class One freight railroads, Amtrak (outside of the Northeast Corridor), and North County Transit District (NCTD) in San Diego, but interestingly, not the PCJPB (ibid). While the initial PTC implementation is on BNSF's track, Metrolink expects to add PTC capability in its own rail territory in 2015 (ibid)

In the south of the region, connecting with Metrolink in Oceanside, is the NCTD's COASTER. The other commuter rail service in the region runs between Oceanside and San Diego. The connections between Metrolink's Los Angeles-Oceanside line and Coaster's Oceanside-San Diego line at the Oceanside Transit Center are not very well coordinated. COASTER, while operating on its own corridor and one of comparable size to Caltrain's SF Peninsula Corridor, has lower ridership and service levels than Caltrain. As with Caltrain, the Union Pacific Railroad operates freight service along the corridor.

5.3 The San Luis Obispo-Los Angeles-San Diego Corridor: Amtrak as Commuter and Rail2Rail

The other commuter rail passenger service in the region is Amtrak California's *Pacific Surfliner* (or "Surfliner"). Running over 300 miles between San Luis Obispo, Los Angeles, and San Diego, the Surfliner provides the one-seat ride between Los Angeles and San Diego that Metrolink and Coaster do not provide on their own. The Pacific Surfliner is managed by the LOSSAN (Los Angeles-San Luis Obispo) Joint Powers Authority with nine counties sitting on the board and staffed through the Orange County Transportation Authority. This switch from state management to local governance occurred in 2012 with the supposition that local management would provide—to quote the bill's sponsor, state senator Alex Padilla, "greater administrative, procurement and operational efficiencies that come with integration" (Gabbard 2012).

5.3.1 The Rail2Rail Program

Integration has thus far been fairly minimal with the exception of the Rail2Rail Program and a joint LOSSAN-specific timetable. While many of the Surfliner trains skip minor stops, it acts as a supplementary commuter service for both local Metrolink and COASTER commuters as well as those traversing across the boundary of the two networks at Oceanside. Metrolink and Amtrak California have negotiated a program in which Metrolink monthly pass holders can ride the Amtrak service on the LOSSAN corridor at no charge. NCTD and Amtrak California have taken Rail2Rail a step further:

6 of the 24 daily Pacific Surfliner trains stop at all of the COASTER stations and any type of passenger (single ride or monthly) can pay COASTER fare to ride on Amtrak (NCTD 2015)[19].

5.3.2 The Missed Opportunity at Oceanside

SCRRA's Metrolink and NCTD's COASTER both end their service at Oceanside, a station in the northern part of San Diego County. The trains operate the same Bombardier bi-level equipment and share platforms at Oceanside, yet through-running is rare; only during horseracing season at Del Mar south of Oceanside do Metrolink trains cross into NCTD territory (Hymon 2011). On top of the lack of through running, it is very difficult to make connections between the two systems. Commuters wanting to transfer northbound from a COASTER train to a Metrolink train bound for Los Angeles find that only two Metrolink trains meet with one of the 11 daily COASTER trains that arrive at Oceanside from San Diego. Commuters wanting to transfer southbound from a Metrolink train to a COASTER train bound for San Diego find that only two COASTER trains meet with one of the 6 daily Metrolink trains that arrive at Oceanside from Los Angeles. Displayed in small text at the corner of the LOSSAN –specific timetable is the warning that even those few transfer opportunities between Metrolink, and COASTER "are not guaranteed" (LOSSAN 2015).

5.4 Freight Presence

Freight has a large presence in Southern California, and with the large port complex of Los Angeles and Long Beach coupled with the low-costs of long-distance rail freight transportation, rail freight volumes in the region continue to grow. The blended corridor in Southern California will run along the BNSF's San Bernardino Subdivision from Hobart Yard near Union Station to Fullerton. According to the Southern California Association of Governments (SCAG), this corridor saw 45 freight trains per day in 2010 and is projected to grow to 90 freight trains per day in 2035 (2012). The line from Burbank to Union Station is owned by SCRRA, but Union Pacific operates about 10 trains per day on that section as well via trackage rights as it is the link between Los Angeles and the Central California coast (Leachman 2011)

While the CHSRA's relationship with freight will to be important across the state, it is on these blended corridors that it is critical. The Peninsula Corridor sees low (but continuous) freight traffic and freight is a tenant on the rail line. In Southern California rail goods movement is a vital part of the regional economy and high-speed rail will operate on freight-owned infrastructure. Any decisions made regarding service on the blended corridor in Southern California will need to have Union Pacific and BNSF at the negotiating table.

5.5 Los Angeles Union Station

Los Angeles Union Station is a critical piece of passenger rail infrastructure in California. With 14 tracks used by three rail operators (Metrolink, Amtrak, and Amtrak California), it will be an important hub for a high-speed rail operator and its future demands careful consideration.

5.5.1 Transit Hub

One of the reasons Union Station is so critical to high-speed rail is that, like the Transbay Transit Center, it is a transit hub in the Los Angeles region. It is best connected to Los Angeles' population centers and will continue to serve as a regional

---

[19] NCTD pays Amtrak California $4.28 for each COASTER passenger that uses the service. SCRRA has a similar agreement in place for a single roundtrip each day on the Ventura county line as well as trips between Los Angeles and Bob-Hope Airport

transportation hub regardless of the ultimate success of high-speed rail. Metrorail, the Los Angeles subway, operates on 7-minute headways during the rush hour peak. A project is under construction to bring two more of Los Angeles' heavily-patronized light rail lines (in addition to the existing Gold Line) into the terminal as well (LACMTA 2015). For a high-speed rail passenger without a car, these connections are critical and will make the train much more competitive with the automobile.

5.5.2 Run-Through Tracks

One of the regional rail projects with the greatest potential to change service patterns in Southern California are run-through tracks at Union Station.

Union Station is currently a "stub-end" terminal, meaning that trains must leave and enter on the same tracks. This causes both congestion on the tracks leading into the station as well as operational delays as trains must either "back in" or "back out" of the station.

The run through tracks are known as the Southern California Regional Interconnector Project (SCRIP)[20] This $350 million effort will save through travelers 15-20 minutes of travel time and increase Union Station's rail capacity (definition of capacity unknown) by 40-50% according to Los Angeles Metro, the leading agency on the project (Wiekel 2014). Currently, all trains arriving and departing from and to San Diego to the south, San Luis Obispo and Santa Barbara to the north, and San Bernardino County to the east all use the same approach tracks to the north of the terminal. The four run-through tracks (some of which will be used by high-speed rail trains) will add a high degree of flexibility for express services and inter- and intra-agency transfers and coordination. Construction is scheduled to be completed in late 2019 or early 2020 (ibid).

5.6 Adding HSR to the Mix: The Southern California Blended System

As part of their new business plan in 2012, the CHSRA chose to pursue a blended system in Southern California conceptually similar to the blended system in Northern California. As in Northern California, freight rail will remain on the corridor. At the time of writing, this blended system runs approximately 50 miles from Burbank in the San Fernando Valley through Los Angeles Union Station and on through to Anaheim's newly-built Anaheim Regional Transportation Intermodal Center (ARTIC). Currently, electrification in Southern California is solely for high-speed rail. Metrolink and Amtrak California have no plans to electrify their systems. From 2022 to 2028, riders would transfer from high-speed rail at Burbank to diesel powered trains for the ride southward to Los Angeles and Anaheim. When the CHSRA completes Phase 1 in 2028, HSR trains would share tracks, but not stations, with Metrolink and the Pacific Surfliner south through Los Angeles and onwards towards ARTIC. Metrolink and the Surfliner, though, would continue conventional rail operation, though possibly at higher speeds than today.

Metrolink and Amtrak California also have no plans to change the specifications in terms of door height or floor height on their operating train sets. This means that high-speed rail will have to use separate platforms even if tracks are shared in between stations. This could limit the capacity and benefit of Los Angeles Union Station because the run-through tracks need to be assigned specifically to high-speed rail or conventional rail.

---

[20] The Southern California Regional Interconnector Project (SCRIP) is not to be confused with the Regional Connector project which is the aforementioned light-rail connection. Both projects will be transformational to Los Angeles' regional transit system.

A December 2014 *Los Angeles Times* article by Ralph Vartabedian discusses the fogginess of the blended system in Southern California. Writes Vartabedian, “Metrolink hasn't had to think about sharing rail because, unlike the Bay Area, the bullet train ultimately is supposed to have its own tracks all the way to Union Station” (2014). The plan to have trains terminate in Burbank for a six-year period (or more as some project followers fear) could, as RailPAC president Paul Dyson notes, “severely crimp ridership” (ibid) As one critic points out, the Burbank terminus would “require major investments in parking lots and other maintenance operations” that are normally built at a rail system’s endpoints” (ibid)

The blended service in Southern California is not as well defined as it is in Northern California; in the words of project spokesman Lisa Marie Alley: “We are looking at all of the options.” However, this means there are opportunities for creativity and integration.

5.7 Conclusions

In this chapter, we reintroduce the reader to the Southern California rail network and the issues it is facing in regards to capacity constraints. We discuss the CHSRA’s relatively nascent plans for a blended corridor in Southern California between Burbank and Anaheim, a distance of 50-miles, similar to the 51-mile Peninsula Corridor. The Authority will face many of the same challenges as it is facing on the Peninsula, but to varying degrees. For example, passenger rail demand is lower in the south than on the Peninsula, but freight demand is higher. The corridor is owned by multiple operators, but one of those operators is the financially powerful BNSF Railway. Both corridors host a hub station on the HSR network: San Francisco on the Peninsula and Los Angeles Union Station in the Southland. The level of coordination at this point is low, but there is lots of opportunity for creative solutions in bringing improved passenger rail service—both high-speed and commuter—to the region.

In the next chapter, we will merge our knowledge of North and South and discuss in broad terms a “wish-list” for the California rail network, a network composed of myriad commuter agencies, freight railroads, California HSR, and even a private high-speed rail venture. We will look at the local decisions made in Northern California and Southern California and evaluate their overall network impact.

6 Measuring the Statewide Impact of Local Decisions

We thank the reader for reading this thesis so far and we hope we have adequately presented the challenges that California will face in the implementation of its high-speed rail line on the blended corridors. Now we move onto an important part of the thesis in which we first describe a dream for the California rail network and then ask ourselves how design and implementation decisions made on the local level impact this dream.

6.1 Putting it all together: Inferring a "wish-list" for California

Based on our research of planning documents and board meeting minutes issued by the railroad agencies and metropolitan planning organizations in California, we compile a "wish list", or rather a summary of goals for commuter and intercity rail agencies across California. While this list is not exhaustive, we believe it represents a broad set of goals that will be impacted by decisions made regarding blended HSR-commuter service.

6.1.1 Level Boarding and Interoperability

Level boarding presents a large operational advantage for commuter rail agencies in that passengers do not need to ascend or descend steps to board a train. Level boarding reduces both station average dwell time length and dwell time variation. In systems with high bicycle ridership, dwell time length will be shortened as bicyclists can now "roll" on board the train. Wheelchair users require ramps to board trains at non-level platforms. If ramps are not present, time-intensive wheelchair lifts are required. Agencies need to budget for a worst-case scenario in terms of wheelchair delay lest these delays cause on-time performance to suffer. This dwell time "padding" gets added into the published timetable, increasing crew costs and making the service less attractive to customers.

Level boarding is relatively rare on U.S. commuter passenger rail systems outside of the Northeast Corridor. The new Eagle P3 project out of Denver, Colorado is one of the first modern commuter rail systems designed to have level-boarding at all doors. Caltrain has expressed a strong desire for level boarding as its high bicyclist volumes increase station dwell times significantly.

Metrolink, COASTER, and Amtrak California have not pursued level boarding. Metrolink and COASTER use cars with the floor 25" above the top of railroad rail; Amtrak California uses mostly cars that are 18" above the rail (Parsons Brinckerhoff 2012). As mentioned earlier, because of the California Public Utilities Commission General Order 26-D which requires certain horizontal clearance minimums, level boarding on shared freight infrastructure is legally impossible. The difference between Caltrain and other agencies in California is that Caltrain has identified GO 26D as a necessary hurdle while the other operators have cited it as a reason not to pursue level boarding. However, to the author's knowledge, neither agency has yet to take steps to seek a waiver for GO-26D.

Interoperability here means the ability of one system to serve passengers of another system. Passengers travelling on a corridor would not need to wait for a specific train operator, but could board the first train that arrives that serves their destination. This allows operators to complement each other's services instead of being competitors. Interoperability goes a step beyond commuter operators being simple "feeders" to the HSR system; instead, commuter operator trackage can be served by HSR trains to provide a "one-seat ride" and HSR trackage could be used by commuter rail to provide intra-regional "express service" The following operators have potential for interoperability:

- Amtrak California’s Surfliner and SCRRA’s Metrolink (already partially exists through Rail2Rail)
- Amtrak California’s Surfliner and NCTD’s COASTER (already partially exists through Rail2Rail)
- CHSRA and PCJPB’s Caltrain
- CHSRA and Metrolink (between Palmdale and Los Angeles and Anaheim)
- CHSRA and Amtrak California (between Burbank and Los Angeles and San Diego
- Metrolink and COASTER (between Los Angeles and San Diego)

If Metrolink and Pacific Surfliner service was fully interoperable with high-speed rail service, customers could make use of interlined ticketing and trains as well as timed transfers to make connections across Southern California. Shared platforms would allow ease of access between systems and reduce infrastructure costs. However, there are sections of the Surfliner’s and Metrolink’s networks that see very low levels of passenger demand and that operate on aging, single-track infrastructure; the benefits of electrification might outweigh the costs.

The PCJPB believes the Baby Bullet service was instrumental to its ridership growth in the last decade; to quote the Caltrain operations manager, “the Baby Bullet is part of our branding.” However, the heterogeneous service adds complexity to the mainly two-track system: train meets and passes have to be precisely scheduled to avoid delays. Caltrain’s future partner, the CHSRA, also needs to decide what kind of service they want to offer commuters on the Peninsula. Currently the CHSRA plans for a minimum of three Peninsula stations (San Jose, Millbrae/SFO, and San Francisco) with the option of a 4th at either Redwood City or Palo Alto. The typical Baby Bullet train stops at six or seven stations between San Jose and San Francisco (inclusive). Given the similarities of Baby Bullets with Peninsula high-speed rail service from a stop-pattern perspective, there is potential for a merging of the two services or provide timed transfers at San Jose or mid-corridor.

6.1.2 Ability to modify service levels as needed

As noted earlier in this thesis, one of HSR’s main competitive advantages versus the airlines is the ability adjust service levels to adjust to consumer travel demand (the marginal cost of an additional train is lower than that of an additional flight plus airlines face capacity constraints of their own). This is only true, however, when the infrastructure is available to add services. Frank Vacca, the program manager of the CHSRA, admitted at the 2015 ASME Joint Rail Conference that capacity enhancements would be needed if a high-speed operator determined it needed more service than the current plan of two high-speed trains per hour per direction on the Peninsula (four per hour per direction with a midline overtake section).

On the commuter rail side, agencies often provide additional services for special events. For example, Caltrain provides additional train service or additional stopping patterns for large events along the Peninsula such as the annual “Bay to Breakers” Race, San Francisco Giants games (and World Series victory parades as of late), or Stanford football games. In Southern California, Metrolink has offered similar additional trains for sporting events and county fairs. Additionally, as demand for commuter rail service grows, agencies would likely be able to add trains to the timetable to avoid crowding.

6.1.3 Shared corridor and success of Xpress West

The CHSRA is not the only entity planning a high-speed rail line in California. “Xpress West”, a private venture backed by Las Vegas hotel developer Tony Marnell, is a planned 185-mile rail line between Palmdale and Las Vegas. The line has been supported by LACMTA as well as the CHSRA, though those agencies do not plan to offer financial assistance (Xpress West 2014). Xpress West would operate in a dedicated right of way at speeds of up to 150mph providing service to Las Vegas at a minimum of every 20 minutes during operating hours.

Because Palmdale is in the far northeast corner of Los Angeles County, in the relatively-isolated Antelope Valley, the Xpress West project is going to rely on high-speed rail, Metrolink, and personal vehicles to draw ridership from the Los Angeles Basin. This would mean entail the creation of a Palmdale transfer hub. It is safe to assume that, ideally, the Xpress West line could be interoperable with the high-speed rail line so that instead of a transfer, travelers could enjoy a one-seat ride between Downtown Los Angeles (or San Francisco) and Las Vegas. Both San Francisco-Las Vegas (#17) and Los Angeles-Las Vegas (#10) are among today's top 20 airport-pair markets in terms of passengers served (Bureau of Transportation Statistics 2014). And on the surface side of the transportation network, the Interstate 15 corridor between the Los Angeles Basin and Las Vegas is a highly congested automobile corridor, especially at peak weekend travel periods (Brennan 2006). This interoperability could help foster what Xpress West has envisioned as a "Southwest HSR Network" connecting San Francisco, Los Angeles, Las Vegas and eventually, San Diego (via California HSR), Salt Lake City, Phoenix, and Denver (Xpress West 2014). While there is considerable doubt as to whether the Xpress West system will ever come to fruition, this link would represent a pivotal step in in bringing HSR across state lines.

6.1.4 Integrated Southern California Rail Network

While the Southern California rail network is impressive on a track-mile basis, the network is splintered and ridership has not reached expectations. While Caltrain in the north has seen unprecedented ridership growth, Metrolink's has lost nearly 600,000 annual riders since 2008. As discussed earlier in the section, there are missed opportunities not only for institutional cooperation between Amtrak California, SCRRA and NCTD, but also within agencies themselves. Increased integration has been discussed in a positive light at both LACMTA board meetings and in the LOSSAN JPA's 2012 strategic implementation plan board meetings, but no integration aside from Rail2Rail and the Del Mar train have been implemented (LACMTA 2012, LOSSAN 2012). In his thesis, Ulrich Leister hypothesizes why the current system in place in Southern California has a low market share:

> Most rail lines are targeted for the commuter market, with rail service only during peak hours. Usability of the rail system is low because departure times and stopping patterns are usually not standardized and therefore customers need good knowledge of "their" rail line. On top, infrequent service leaves passengers without flexibility and, in the worst case, forces customers to structure their day according to a train's timetable (Leister 2011).

In an ideal Southern California rail system, riders could count on reliable and smooth connections between systems and "one-seat rides" between hubs. A single ticketing system would reduce complexity for riders and encourage transfers to local transit systems. With an integrated system, true headways would decrease as train operators could complement each other's services. The arrival of HSR to the region promises lots of opportunity to take a serious look at the rail system California has and the system it wants and deserves.

Connecting San Diego and Los Angeles with a one-seat commuter rail ride is an oft-discussed goal, but an institutional challenge—both Metro and the LOSSAN JPA have stated it in their near-term ambitions. The biggest challenge largely institutional: there are two existing commuter railroad owners and operators in SCRRA and NCTD and a new Los Angeles-San Diego commuter service would directly compete with Amtrak California. Aside from slightly more comfortable seats, Wi-Fi, on-board snack purchases, and slightly faster travel time (due to less station stops), Amtrak is a direct substitute for commuter service. Today's LOSSAN provides users with a joint timetable that lists Metrolink, COASTER, and Amtrak services on the corridor, but higher levels of integration will provide a more seamless customer experience.

6.1.5 High-frequency and uniform HSR and commuter service

In their 2008 published preliminary operating plan, the CHSRA scheduled eight trains per hour on the Peninsula. Before the blended service concept surfaced, the idea of ten Caltrain trains per hour per direction during peak periods was proposed by Caltrain in their Caltrain 2025 draft service plan. This ten train timetable was also reported in the CHSRA's preliminary alternatives analysis document in 2010. These ten trains included four trains per hour terminating at the Transbay Transit Center. It is reasonable to assume that Caltrain's capacity needs have not decreased since the report was written and may likely have grown (based on ridership outpacing expectations). According to the 2013 State Rail Plan, Metrolink, Amtrak California, and COASTER all expect to increase service levels as well.

As discussed earlier in this thesis, uniform headways are advantageous to both intercity travelers as well as commuters. Uniform, "clockface" (trains leave at the same minute each hour) headways not only make it easier to for customers to memorize schedules, but also result in less crowding on individual trains (under the assumption of uniform passenger arrival rates at stations). Transit agencies around the world, operate on uniform headways and clockface timetables when practical. And again, before the blended service concept appeared, both the CHSRA and the Caltrain published uniform, "clockface" timetables in service planning documents.

6.1.6 Bond Measure Satisfied

Finally, it is important that the requirements of Proposition 1A, the bond measure that voters passed authorizing funds for high-speed rail, are met. Proposition 1A mandates a maximum designed trip time of 30 minutes from San Francisco to San Jose and 160 minutes from San Francisco to Los Angeles. The CHSRA's program management team issued a memo stating that even in the blended system those requirements could be reached if the Peninsula is upgraded to 125mph track (CHSRA 2014b). As noted in chapter 2, the FRA only permits 125mph operations if an "impenetrable barrier" blocks traffic at grade crossings. These impenetrable barriers will inevitably increase "gate-down time" and intersection delay along the Peninsula.

Additionally, the bond measure calls for achievable operating headways of 5 minutes or less. This currently runs contrary to the blended service analysis which stated that the Peninsula Corridor is capable of operating at most 10 trains per hour per direction (6 minute headways); and this headway is contingent upon significant new construction. In the 2014 Business Plan, the CHSRA showed an operating run time of 180 minutes between San Francisco and Los Angeles (as opposed to 160) and 50 minutes between San Francisco and San Jose (as opposed to 30).

It has become clear that the CHSRA would like to differentiate between what the system is capable of and what will happen once the system begins operating in revenue service. This distinction has already become the subject of a lawsuit between two farmers in Central California and the CHSRA. Frank Vacca, the CHSRA's program manager, in a declaration to the court argued the following, making clear the CHSRA's distinction:

> For the purposes of the Business Plan, the operating plan described that shows a travel time between San Francisco and Los Angeles of 180 minutes (or three hours) . . . was representative of the information provided for the ridership forecasting model to forecast ridership levels based on specific patterns and frequency of train

service. These service patterns were designed to achieve maximum commercial yield and were in no way tied to the ultimate performance capabilities for the travel time along the Phase 1 corridor (Vacca 2013).

For this thesis, we will agree with the CHSRA's approach and evaluate the ability of the Authority to meet "theoretical" design criteria outlined in Proposition 1A. The three main requirements that we will examine are 1) travel time, 2) headway, and 3) the ability of the project to be operationally self-sustaining (revenue-neutral at worst).

Furthermore, it is worth noting that there is no timeline required in the Proposition for satisfying the bond measure. One argument the CHSRA could make is that this is a phased project and that the bond measure requirements will ultimately be met at some unspecified date in the future.

6.1.7 Minimize Costs and Project Timeline across California

Californians are naturally suspicious of large-scale infrastructure projects. In 1989, the Loma Prieta Earthquake caused a section of the upper deck of the San Francisco-Oakland Bay Bridge to collapse onto the lower deck. An analysis done by the California Department of Transportation determined the bridge was seismically unsound (Mladjov 2011) and that a new eastern span would be necessary. It took the state over 12 years to begin construction of that span. When construction commenced in January 2002, the project was slated to be complete in 2007 with costs estimated at $1.4 billion (ibid). The bridge finally opened in 2013 at a cost of $6.4 billion (Associated Press 2013). Even after the bridge opened, there has been ongoing concern that certain anchor rods built installed within the bridge are deficient (ibid). If Caltrans cannot build a ½ mile span to replace a structurally unsound bridge without a 900% cost overrun and a 24-year timeline, Californians might justifiably wonder how realistic the estimates are for $68 billion, 520-mile rail line to be completed by 2028.

Cost minimization and construction expediency are critical for not just the CHSRA, but for the entire state. Caltrain is on the cusp of electrifying its corridor, a project that was first proposed in 1992 (Morrison Knudsen Corporation 1992). Freeway and airport congestion are nearing their limits today, and with the state projected to add 10 million people by 2050, adequate rail capacity will only become more critical. While there is an impetus to build projects fast, there is also a need to build projects right. Short-term costs need to be weighed against long-term benefits, not just for individual regions or agencies, but for the wellness of the Golden State as a whole.

6.2 Upcoming Local Design Decisions

We will now discuss four key design choices that are upcoming in the next few years involving the CHSRA and local rail agencies and operators. Several of these local decisions will not only impact the rest of the California rail network, but will also set a precedent for future decisions made along the HSR corridor. Complicated negotiations taking place in Northern California on the more-developed blended corridor will undoubtedly have an impact on negotiations that occur in Southern California. Furthermore, the Southern California-Northern California rivalry will continue to play a role in the state's politics. Politicians in both regions will be on the lookout for special benefits afforded to their counterpart region and not their own. The California High-Speed Train Project is the largest truly cross-regional project to take place in the Golden State and no local decision will be allowed to exist unnoticed in its regional bubble.

6.2.1 Decision A: Platform Height and Equipment Floor Height

In Chapter Four, we discussed the coordination between Caltrain and the CHSRA in procuring equipment with similar door heights in order to facilitate common-use platforms. Common-use platforms are possible without having matching door heights: as long as the platform is below the floor of the vehicle, mini-high platforms can be constructed to accommodate both wheelchair users and passengers now needing to make a large step-up to a higher vehicle floor. However, there is no known instance of mini-low platforms where a ramp down to a lower door exists. This means that low-floor vehicles would not be compatible with platforms built for level-boarding with high-floor vehicles (e.g. the CHSRA's trainsets).

The CHSRA has been adamant thus far about having a nominal 50" vehicle floor. A 2009 technical memorandum from project consultant Parsons Brinckerhoff (PB) advocates for a platform height between 45.47" and 51.18" based on trains in existing service (CHSRA 2009b). According to PB, only the Alstom AGV Duplex operates with low platforms today (12.36" floors); however, that train can only operate up to 200 mph which falls short of the 220 mph requirement of Proposition 1A. Having a lower platform height would undoubtedly facilitate level boarding; for example, CHSRA can adapt to the existing Caltrain system and not the other way around. While the CHSRA would like to use a "service-proven" HSR system that has been in operation for at least 5-years, it would be worth examining the price-premium manufacturers would ask for a low-floor, high-speed train that has better compatibility with the existing system. For now, however, we will assume that CHSRA will adopt a 50" platform.

Design Decision	Explanation
1—Caltrain acts alone	Caltrain selects an EMU car compatible with existing diesel fleet
2—CHSRA coordinates with Caltrain	Caltrain selects an EMU car compatible with HSR fleet
3—CHSRA coordinates with Caltrain and Southern California railroads	There is a standard vehicle door height across CHSRA, Caltrain, Metrolink, the Pacific Surfliner, and possibly COASTER
4—CHSRA coordinates with Caltrain, Southern California Railroads, and other California rail agencies	In addition to the standard described in Choice 3, other agencies such as the San Francisco Municipal Transit Authority, Bay Area Rapid Transit, Amtrak California (Capital Corridor and San Joaquin trains), or the Valley Transportation Authority standardize as well.

6.2.2 Decision B: Capacity Allocation Strategy

In Chapter Four, we discussed the gulf between the CHSRA's willingness and ability to pay for access to the Transbay Transit Center versus Caltrain. Given the importance the terminal station to both operators, a method of determining who gets access and at what time will be necessary. This could come in the form of integrated planning or some sort of capacity allocation mechanism.

Design Decision	Explanation
1—Do not develop a capacity allocation strategy	Build infrastructure first with general service plan assumptions and wait to negotiate particular train slots when all parties are ready
2—Create a codified capacity allocations strategy	Develop a formal set of rules to allocate capacity on the Peninsula and access into the Transbay Transit Center
3—Negotiate capacity	Agree on a service plan prior to putting HSR operating contract out for bid

6.2.3 Decision C: Southern California Electrification Timeline and Scope

Though Metrolink has come out against electrification, rail advocates such as the Rail Passengers Association of California and Nevada (RailPAC) have pointed out many potential benefits to electrification of the entire blended corridor in 2022 prior to the CHSRA's plan to electrify in 2028 (McCallon 2012).

Design Decision	Explanation
1—Conventional (Diesel Corridor)	Maintain Amtrak California and Metrolink service as diesel powered operations ; truncate HSR in Burbank from 2022-2028
2—Electrification	Convert the blended corridor to all-electric operation (aside from freight) and commence Los Angeles HSR service in 2022

6.2.4 Decision D: Two Track Corridor on the Peninsula

In Chapter Four we discuss LTK's Blended Operations analysis done on behalf of the PCJPB where they verified the performance of different track configurations on the blended system. We will examine the impact of three different design choices on the performance of the corridor and the HSR system.

Design Decision	Explanation
1—Keep corridor as is	Do not add additional passing tracks
2—Expand the corridor to include passing tracks	Build LTK's recommended "midline overtake" section on the corridor, adding about 10 miles of quadruple track
3—Revert to the four-track option	Build HSR on the Peninsula as originally planned with two dedicated tracks for HSR use

6.3 Measuring Network Impact

In this section, we are going to take a look at these four major upcoming decisions and discuss the impact of each design decision from the viewpoint of the California rail network.

6.3.1 Decision A: Platform Height and Equipment

Choice 1—Railroads Act Alone

If Caltrain acts alone on platform height, they will likely select a 25" floor height to maintain compatibility with their diesel-electric rolling stock (PCJPB 2013a). Metrolink and COASTER will continue to maintain their fleets at the 25" door height as well. The CHSRA will continue forward with its planned 50" floor height.

	Railroads Act Individually
Level Boarding and Interoperability	Door heights for Caltrain, Metrolink, COASTER, and Amtrak California will be below HSR platforms. HSR trains will require mini-high platforms to use commuter rail stations since there will be a minimum of 25" step.
Ability to modify service levels	HSR will be limited by its four-track terminal at Transbay Transit Center and Caltrain by its two track terminal at all times. In Southern California, run-through capacity for non-HSR passenger railroads may be limited at Los Angeles Union Station where high-speed rail would need two of the four run-through tracks.
Shared corridor with Xpress West	In order to have an opportunity for a shared system, Xpress will likely specify same vehicle floor height at HSR to ensure interoperability. However, Xpress West will not be able to stop at non-HSR stations without constructing mini-high platforms
Integrated Southern California Rail Network	Metrolink and Amtrak California are likely limited to two through-tracks at Union Station. Cross-platform, timed-transfers at Union Station between regional and high-speed rail are impossible.
High-frequency, uniform HSR and commuter service	According to LTK's blended service plan (which makes an assumption of separate facilities), Caltrain will have uneven headways and limited access to Transbay Transit Center. Unless politically-challenging passing tracks are constructed, high-speed rail will not function beyond two trains per hour per direction on the corridor which is a far-cry from HSR's original 12 trains per hour goal. Frequency and uniform headways for HSR and commuter service in Southern California will depend mainly on the degree of track sharing between HSR, freight, and other passenger rail agencies as well as the ability of all operating railroads to coordinate service planning.
Bond Measure Satisfied	Theoretical travel times and headways are satisfied with 125mph on the Peninsula, however satisfying the revenue-neutral requirement and securing a private operator will be challenging.

Minimize costs and build system quickly	Metrolink and Amtrak California save time and money by not changing vehicle floor heights or platforms for interoperability. However, CHSRA will need to construct new stations in its shared corridors.

Choice 2—California HSR and Caltrain Act Together

If Caltrain and HSR act together, Caltrain will likely procure a vehicle with two door heights—one compatible with its current diesel fleet and one that matches HSR floor heights. Amtrak California and Metrolink would likely keep their existing 25” and 18” floor heights respectively. This analysis assumes Caltrain will eventually convert its platforms to the same height as HSR since not doing so would defeat the purpose of buying the dual floor height cars.

	Caltrain and HSR Act Together
Level Boarding and Interoperability	Caltrain will be able to stop at any HSR station and vice versa. Even if stations are segregated, HSR trains could use Caltrain platforms (or vice versa) in the event of an emergency. HSR trains could make stops at commuter stations for special events.
Ability to modify service levels	HSR and Caltrain could use each other's spare platforms at the Transbay Transit Center, however, limited line capacity on the Peninsula would still be a challenge. Caltrain and HSR would have increased flexibility to create merged service plans (see fifth box in matrix). In Southern California, run-through capacity for non-HSR passenger railroads may be limited at Los Angeles Union Station where high-speed rail would need two of the four run-through tracks.
Shared corridor with Xpress West	Even though, it is currently only a Southern California system, Xpress West would have full access to Peninsula (and CHSRA would have access to Las Vegas), but this might be financially impractical since Xpress West would have to pay access charges to Caltrain and CHSRA—a timed transfer at Palmdale with CHSRA might be preferable. Xpress West will not be able to stop at Southern California commuter (non-HSR) stations without constructing mini-high platforms.
Integrated Southern California Rail Network	As with Choice 1, Metrolink and Amtrak California are likely limited to two through-tracks at Union Station. Cross-platform, timed-transfers at Union Station between regional and high-speed rail are impossible. However, a successful negotiation between Caltrain and HSR regarding platform heights might pave the way for a similar arrangement in Southern California.
High-frequency, uniform HSR and commuter service	With HSR and Caltrain having the same floor height, services on the Peninsula could be interchangeable; that is, HSR could run a Caltrain-like service plan with multiple stops or Caltrain could run a HSR-like service plan with few stops. Increased heterogeneity in service patterns during a single time period, however, will limit the practical capacity of the corridor. Frequency and uniform headways for HSR and commuter service in Southern California will depend mainly on the degree of track sharing between HSR, freight, and other passenger rail agencies as well as the ability of all operating railroads to coordinate service planning.
Bond Measure Satisfied	Theoretical travel times and headways are satisfied with 110mph on the Peninsula. If Caltrain and CHSRA interline services, the two agencies would need to negotiate a joint-venture revenue and cost-sharing agreement so a private investor in the system can expect to generate revenues on the

	corridor even if it is not operating the trains.
Minimize costs and build system quickly	Caltrain will have increased expense by ordering non-standard vehicles with two door heights. HSR may be able to avoid building separate facilities at its Peninsula stations, but will need to build its own facilities in Southern California. Construction time would be accelerated on the Peninsula because of reduced scope.

Choice 3—California HSR, Caltrain, and Southern California Rail Agencies Act Together

If Caltrain and HSR include Amtrak California's Pacific Surfliner, SCRRA (Metrolink), and possibly NCTD (COASTER) in the vehicle floor height decision, additional opportunities for level boarding and interlining emerge. Level boarding has the highest benefit for Amtrak California and Metrolink on routes shared with HSR[21]. All Metrolink routes besides the Inland Empire-Orange County Line are technically shared with HSR's blended system for a certain distance outside of Los Angeles Union Station, but in particular, the Ventura County Line and the Antelope Valley Line north of Los Angeles and the Orange County Line south of Los Angeles together cover the entire blended corridor between Burbank and Anaheim. The Pacific Surfliner covers the entire blended corridor from Burbank to Anaheim. This will facilitate platform and station sharing, and at a minimum, cross-platform transfers.

For the purposes of this thesis, let us assume that ultimately the entire Southern California rail network adopts a HSR-compatible floor height. This could mean building "mini-high" platforms at stations without high-platforms or purchasing railcars capable of using two different platform heights as Caltrain is currently exploring.

The benefits described in "Choice 2" will be extended to Southern California. Metrolink and Amtrak would provide higher-quality "feeder service" to the HSR trunk. Xpress West will have access to all stations on the shared corridor in Southern California. All four run-through tracks at Union Station could be used by any rail provider allowing for higher returns on investment. Additional negotiation will be required to guarantee that Metrolink and Amtrak services are not competing with high-speed rail for revenue; that is, in order to satisfy the revenue-neutral requirement of the bond measure, Amtrak and Metrolink need to be feeding the HSR network, not stealing passengers from it. Costs would increase in that Metrolink and Amtrak would need to purchase new equipment and raise their existing platforms, but costs might decrease in that stations could be shared and costly new HSR station infrastructure could be eliminated.

Choice 4 —Bringing the SFMTA and BART to the platform height discussion

In his TRB paper that we discussed in Chapter 2, Reinhard Clever presents a creative proposal in which Caltrain would use outbound tracks on the San Francisco Municipal Transportation Agency's Market Street Subway through downtown San Francisco. The Market Street Subway is a light-rail subway; in the inbound direction, five lines converge; in the outbound direction, Clever argues that a coupling operation at the end of the subway could increase the capacity and allow room for Caltrain. A one-mile single track construction along Seventh Street in San Francisco to complete a Caltrain loop

---

[21] Amtrak California's *Capitol Corridor* (as well as SJRRC's *Altamont Commuter Express* and Amtrak's *Coast* Starlight) share with Caltrain (and California HSR) a 3-mile piece of the Peninsula Corridor. However, at this part of the corridor there is a third track we expect that these services will operate on the additional track and not share tracks with HSR or Caltrain (though they sometimes share one of Caltrain's two mainline tracks today.

would allow for Caltrain trains to continue back onto the main Peninsula Corridor and serve the Peninsula. The two main advantages of this system are:

1. Caltrain would no longer need access to the Transbay Transit Center to serve downtown San Francisco while Peninsula commuters would still enjoy a one-seat ride to key San Francisco destinations
2. Light-rail lines and BART share several of these key downtown stations with Caltrain enabling easy connections for San Francisco residents commuting southbound and Peninsula residents commuting outside the San Francisco CBD

This system would free capacity in the Transbay Transit Center and set a helpful precedent for integration across the state. CHSRA could add services into Transbay Transit Center if necessary since Caltrain would have less requirements for track. The largest challenges to this plan are institutional in nature. SFMTA is expecting its new Central Subway to provide a light-rail connection between Caltrain and downtown and has used Caltrain transfer traffic as a large component of ridership (SFMTA 2008). There is the technical challenge of bringing a heavy rail system into a light rail system regarding horizontal and vertical clearances, platform heights (SFMTA's Market Street Subway platform height is about 36") and different power systems (600V DC in Market Street Subway vs 25 kV HSR/Caltrain), but Clever points to systems in Europe (e.g. Karlsruhe, Germany) where these challenges have been overcome and where the ridership gains have been on the order of 400% (Clever 2013).

Clever takes his idea of system "convergence" a step further by suggesting that the HSR system could run on underutilized BART right-of-way between the Peninsula Corridor and SFO Airport. BART operates on Indian Broad-gauge track which is 5 feet, 6" between rails and incompatible with the 4 feet, 8.5" standard gauge on the Peninsula Corridor and the future HSR network (ibid). The BART track between the BART-Caltrain transfer station at Millbrae is only operable after 8pm during the week and on weekends; writes Clever, "Demand was so low [on the Millbrae-SFO BART shuttle]--it was known as the 'ghost train!'" (5).

If the BART track were replaced with standard gauge track, HSR trains could provide a one-seat ride to SFO complete with checked baggage and through-ticketing from points across the Central Valley and San Jose. The BART SFO airport station is under-utilized: it has three tracks, one of which is never used; this third track could serve as a terminal for high-speed rail. Sending one HSR train out of every three would free up space north of Millbrae on the Peninsula. And though Clever does not go this far, Caltrain could reduce service on the Peninsula south of Millbrae as HSR could shuttle some commuters from San Jose to the Millbrae station where Caltrain could provide a timed-transfer for its short loop into and through San Francisco. While this depends on careful schedule planning, the returning southbound Caltrain could meet the airport HSR train returning southbound to provide a seamless connection to the southern half of the state.

Again, this would entail further institutional challenges and construction costs. BART would need to release its right-of-way and station platform at SFO and infrastructure costs would increase in the short term. For truly blended integration, participating agencies need to act as blended entities. Clever warns the reader, "[Without integration], the California HSR system will almost certainly not reach its full potential, and will instead degrade into a very fast connection between huge parking lots" (13).

6.3.2 Decision B: Capacity Allocation Strategy

Choice 1—Do Not Develop a Capacity Allocation Strategy

This choice means that infrastructure decisions will be made prior to service planning decisions. Agencies have a general idea of their capacity needs in terms of trains per hour per direction. CHSRA wants at least four trains per hour per direction. Caltrain plans on operating six (but has planned for up to 10) with two serving Transbay Transit Center. In the South, Metrolink, COASTER, Amtrak have expansion plans as well. According to the 2013 California State Rail Plan, non-HSR passenger rail volumes on the blended corridor sections are going to increase significantly between 2014 and 2030:

Table 6-1: Forecasted 2040 non-HSR train volume increases on Southern California blended corridor

Blended Section	Owner	2014 Daily Round Trips	2040 Daily Round Trips	Percent Increase
Palmdale-Burbank*	SCRRA	15	23	53.3%
Burbank-Los Angeles	SCRRA	36	53	47.2%
Los Angeles-Fullerton	BNSF	29	49	69.0%
Fullerton-Anaheim	SCRRA	29	39	34.5%

**Palmdale-Burbank will not likely be shared for its entire length*
Source: California State Rail Plan/Author's Calculations

	Do Not Develop a Capacity Allocation and Capacity Pricing Strategy
Level Boarding and Interoperability	While the lack of a capacity allocation scheme does not eliminate the possibility of level boarding and interoperability, interoperability is by definition, a capacity consideration, so it is unlikely that interoperability will occur in an environment that does not consider capacity before infrastructure decisions are made.
Ability to modify service levels	Absent a capacity allocation plan, adding frequency will be easy when capacity is abundant, but capacity will become congested more quickly than without centralized planning. A "buffet style" approach to track access in which operators take capacity without regard for other operators capacity needs will quickly congest available train slots. Adding services will require coordination among all affected operators and will be institutionally challenging and time-consuming.
Shared corridor with Xpress West	Xpress West will have no capacity challenges on their track since they are operating in a dedicated corridor. However, being a private operator trying to compete for track access with public agencies such as SCRRA or CHSRA, Xpress West will likely have last priority to access the Palmdale-Los Angeles section of the network.

Integrated Southern California Rail Network	Because of the joint powers authority setup of the SCRRA/Metrolink and the unidirectional nature of commute patterns (inbound from outlying counties to Downtown Los Angeles in A.M., outbound in P.M.), the "confederacy of counties" will be resistant to any changes that affect services in that county. Because service changes will have a rippling effect (positive or negative) across all lines in and out of Los Angeles Union Station, it will be difficult to coordinate and alter Metrolink service without inconveniencing members of the JPA.
High-frequency, uniform HSR and commuter service	Initially, operators can deploy high-frequency, uniform service, but as capacity gets consumed, it becomes difficult to continue that trend. For example, the LTK blended operations analysis assumes HSR has the first opportunity on the Peninsula for capacity and schedules trains with uniform headways; as a result, reasonable Caltrain frequencies to the Transbay Transit Center are not implemented and operating headways are highly variable. While the LTK analysis was conducted to verify broad feasibility of the blended system, it is apparent that service uniformity becomes an increasingly complex exercise under the assumption of not developing a capacity allocation policy.
Bond Measure Satisfied	Adherence to the bond measure requirements is dependent on whether or not the HSR operator has first access to train slots on parts of the corridor it does not own. Otherwise, HSR will find its trains "stuck" behind slower moving commuters or freight services and the time requirements may not be satisfied. If service quality suffers, revenue will suffer as well and the revenue-neutral requirement may not be met.
Minimize costs and build system quickly	If careful planning of capacity needs can minimize infrastructure costs, then not developing a capacity allocation and pricing strategy will result in unnecessary/redundant rail infrastructure. If capacity is allocated inefficiently with a first-come, first-serve mentality, then agencies could find themselves in a situation where it is politically easier (yet more expensive) to build more capacity than to renegotiate with multiple host railroads. The State of California will also receive less competitive bids from private operators due to the uncertainty related to track access.

Not developing a capacity allocation scheme or negotiating capacity prior to relatively precise service planning decisions is high-risk, especially to a private investor in the HSR system. The State will pay a risk-premium for not working out capacity agreements especially on the congested Peninsula Corridor and the BNSF section of the corridor between Los Angeles and Anaheim. In her dissertation Maite Pena-Alcaraz notes how lack of capacity planning has made modification of Northeast Corridor services an extremely difficult exercise (2015).

The National Cooperative High Research Program Report 773 (2014) describes the current method of adding service, which is the likely outcome of waiting to implement a capacity allocation policy: "On corridors they own, freight carriers fully control the technical assessment of the operations for proposed and existing shared-use territories even when the passenger rail sponsor underwrites the cost of such analysis" (14). It is unlikely that any private investor will be willing to risk potentially lucrative Los Angeles-Anaheim service to that level of capacity uncertainty.

Choice 2—Create a Codified Capacity Allocation Strategy

This choice requires agencies to develop a codified policy for allocating and pricing capacity. Again, this is only relevant to corridors where HSR is sharing track or station access (i.e. the blended corridors in Northern or Southern California). There are network challenges, however, with a capacity allocation and pricing mechanism. Take for example the Transbay Joint Powers Authority (TJPA). If the TJPA, through any allocation mechanism, awards an HSR train the right to depart at a certain time, Caltrain needs to allow that train to continue to move through the corridor and continue southward. When that train arrives in Metrolink territory, it needs to be accepted to move towards Los Angeles Union Station.

	Create a codified capacity allocation and pricing procedure
Level Boarding and Interoperability	If agencies work together to try to formulate an approach to allocating capacity, the benefits of level boarding and interoperability will be quickly realized as level boarding and interoperability minimize on-track competition and a capacity allocation mechanism emphasizes the trade –offs necessary with on-track competition.
Ability to modify service levels	While adding frequency may not be easy, a codified capacity allocation procedure will make it transparent so operators will know the necessary steps to add services.
Shared corridor with Xpress West	If California develops a capacity allocation mechanism, it will need to consider the potential of Xpress West or other private operators seeking access to the track. In this sense, a capacity allocation mechanism will be beneficial to Xpress West and any other private operators evaluating the commercial viability of a new rail enterprise.
Integrated Southern California Rail Network	The Peninsula Corridor is relatively simple compared to the network in Southern California. Agencies will have to decide the scope of any capacity allocation and pricing mechanisms. If the mechanism only applies to corridors with HSR, operations outside the corridor will be subject to decisions made within the corridor; for example, if Metrolink is limited to two trains per hour by the capacity allocation mechanism, should those two trains continue on to Ventura to the west, Oceanside to the south, or San Bernardino or Riverside to the east? Furthermore, unless a capacity allocation procedure takes into account the effects of network integration (e.g. connections between HSR and Metrolink), it is unlikely to lead to coordinated timetables.
High-frequency, uniform HSR and commuter service	As we discuss in Chapter Four, a high-speed operator has a much higher willingness to pay for access to infrastructure so a capacity allocation mechanism could put commuter rail agencies at a disadvantage in securing high frequencies and uniform headways. Also, because train paths are not independent of one another (a high-speed train could disrupt the paths of multiple commuters and vice versa), it is unlikely that a capacity allocation mechanism will result in uniform headways or produce high levels of track utilization.
Bond Measure Satisfied	A HSR operator with a higher ability to pay for infrastructure access might be able to guarantee itself the capacity needed to meet the headway and travel time goals defined in the bond measure.

	However, if there is significant expense in securing access to the track, this will impact an operator's bottom line and the revenue-neutral requirement of the bond measure will not be satisfied.
Minimize costs and build system quickly	With a capacity allocation mechanism in place, a private HSR operator will have a much better understanding of the risks associated with operating in a shared-use environment and the risk premium that the operator demands might be lessened.

Choice 3—Negotiate Service

With this option, railroad agencies work together to develop a timetable. Like the discussion on platform and vehicle floor height, the number of included parties increases the complexity of conversations but also could potentially yield a better outcome. The parties would agree on a preliminary timetable, build infrastructure to match said timetable, and then make changes at a later date if necessary. The timetable itself could be part of the concession contract for the private high-speed rail operator.

	Negotiate service
Level Boarding and Interoperability	Level boarding and interoperability will increase the ability for operators to mix and match services and remove constraints on the negotiating process. If service negotiations occur before agencies make design decisions affecting level boarding and interoperability, it is more likely the agencies will reach design decisions facilitating these goals.
Ability to modify service levels	Like Choice 1, no capacity allocation strategy, adding service will be easy when capacity is available, but difficult once capacity is saturated. At saturation, more and more entities will be affected by small changes in the timetable and adding services will become a challenging and time-consuming exercise. At some point, if demand patterns change enough, an entire re-evaluation of the timetable may be necessary.
Shared corridor with Xpress West	Xpress West will need to be involved in the timetable negotiation discussion if the CHSRA would like to integrate Xpress West service on the corridor. Again, however, the addition of Xpress West potentially puts five operators (CHSRA, Amtrak California, Metrolink, and BNSF) on the Burbank-Anaheim corridor and could make negotiations extremely complicated.
Integrated Southern California Rail Network	This choice perhaps leaves California with the best opportunity of integrating the Southern California rail network. Transfers could be carefully timed and freight rail demand could be managed in windows that avoid conflicts with passenger trains. Cuts on some lines could be justified with enhanced connections at transfer stations like Anaheim, Oceanside Burbank, Fullerton, or Los Angeles Union Station.
High-frequency, uniform HSR and commuter service	Much like the ability of negotiated service plan to create an integrated Southern California rail network, careful service planning could also lead to a high-frequency and even headway rail service. This will be achieved with well-timed transfers and removal of redundant services.
Bond Measure Satisfied	During the service negotiation period, the CHSRA could use the bond measure requirements as a hard or soft constraint. The CHSRA could perform a financial analysis of any proposed timetables and understand the revenue potential to ensure that the revenue-neutral requirement is satisfied. If capacity improvements are needed, the necessity of said improvements might very well be realized during this timetable negotiation process.

Minimize costs and build system quickly	Having a more definitive operating plan in place and knowing the infrastructure necessary to meet that operating plan will reduce the risk for a private HSR operator even more than a defined capacity allocation mechanism. This will minimize the risk costs borne by the State. Also, a better understanding of infrastructure needs will help avoid excessive construction and lead to lower costs and a shorter project timeline.

Careful service planning (negotiated service) might very well be welcomed by the freight railroads. Freight railroads are also capacity constrained and long-standing arrangements with passenger trains pre-date some existing capacity constraints. If freight operators are guaranteed slots to operate trains throughout the day, they may be more hospitable to increase passenger operations in the blended corridors. As long as it does not cause traffic delay, freight railroads have an interest in more passenger operators since additional train movements typically results in higher access fee revenues. Detailed discussions with freight railroads may reveal necessary capacity improvements as well.

Freight railroads operate on networks that cross state lines; the California network effects discussed in this thesis might pale in comparison to the network effects felt across a multi-state, multi-hub freight network. NCHRP 773 notes this fact and admits that host freight railroads are often unwilling to "bear the disruptions associated with an embargo of freight operations over a shorter section of track during, for example, commuter rush hour periods" (15) Because of the location of BNSF's large Hobart yard (several miles from Union Station and on the future blended Anaheim-Los Angeles corridor), the CHSRA might find that separate railroad infrastructure is a necessary requirement to ensure smooth, profitable operations for the private operator.

There are certain institutional challenges in moving towards a capacity allocation mechanism. For the multiple infrastructure managers—the TJPA, the PCJPB, CHSRA, SCRRA, and possibly the BNSF—to cede control of capacity and break current capacity contracts and switch to a formalized out-of-headquarters mechanism. All operators will have to perceive any capacity mechanism as unbiased and result in a net positive for their respective operation.

6.3.3 Decision C: Southern California Electrification Timeline and Scope

Choice 1—Maintain Diesel Rail Service

Here Metrolink and the Pacific Surfliner continue diesel operation on the corridor. Between 2022 and 2028, high-speed rail passengers alight at the Burbank Station and then transfer to a diesel train for the ride to Los Angeles or Anaheim.

The Transfer from Burbank to Los Angeles Union Station

Currently the Surfliner does not stop at the station, but assuming it did, Burbank-Los Angeles sees 38 passenger trains per day and the timetable is concentrated around the peaks (Amtrak, SCRRA 2015 timetables). The San Fernando station will see 50 trains per day or one train every 19 minutes (CHSRA 2014b). Again, these trips would be concentrated, albeit slightly less so, around the peak times—high speed rail is expected to have a minimum of 30-minute headways throughout the day; Metrolink has a 60-minute headway midday from the Burbank Station.

In his January 2015 presentation to the CHSRA, East Japan Railway Company's vice chairman, Masaki Ogata, stressed the importance of total trip time and the importance of shortening all pieces (door-to-door) of the HSR journey. The total trip time from Burbank in a scenario with conventional rail service is as follows:

$$Total\ Trip\ Time = Transfer\ Time\ at\ Burbank + Diesel\ Trip\ to\ Destination\ South\ of\ Burbank + Transfer\ Penalty \quad \text{6-1}$$

Though Mr. Ogata does not define it, the transfer penalty is an unknown, yet non-zero value of time (often expressed in minutes) that represents the customer's disutility of transferring. It is based on the following items:

- Quality of transfer station—how difficult is it for customers to transfer between services? Will bags be transferred automatically? Is there an enclosed waiting area for customers?
- Uncertainty in connections—is there any chance of missing connections? Will the customer need to purchase another ticket, and if so, does the customer have assurance that there will be adequate ticketing machines and staff available to avoid missing a connection? Is there enough space on the next train?
- Reliability of connections—What is the on-time performance of connecting trains?[22] Will connecting trains (HSR or commuter) wait for a customer's train if the customer's original train is late?

While it is difficult to find a value for this transfer penalty for high-speed rail services, Guo estimates it at 9.52 minutes for the MBTA (2008). For example, the Toronto Transportation Commission uses a coefficient of 10, suggesting that each transfer is 10 times more painful for a customer than time spent travelling in-vehicle (Wilson 2014)

Table 6-2: Travel time via conventional rail (Source: SCRRA, Author's Calculations)

Burbank to:	Distance	Best Travel Time	Avg. Speed	% trip time increase on top of a 158-minute San Jose-Burbank HSR trip[b]	% trip time increase on top of a 62-minute Fresno-Burbank HSR trip[b]
Los Angeles	13 miles	27 minutes	35mph	17%	44%
Anaheim	44 miles	65 minutes[a]	44mph	41%	105%

[a] Author used best time of Pacific Surfliner and Metrolink and subtracted lay-over time at Union Station under assumption that run-through tracks would reduce this layover to 1-2 minutes.

[b] Using travel times from all-stop HSR train in 2014 Business Plan, Author assumes five-minute transfer at Burbank and 100% on-time departures from Burbank

[c] In the 2014 Business Plan, the CHSRA schedules approximately 10 minutes travel time between Burbank and Los Angeles after Phase 1 is complete

Table 6-2 puts into numeric terms the inconvenience of a transfer at Burbank. An HSR passenger adds 12 minutes of in-vehicle travel time versus a one-seat HSR ride; the additional 13 miles (even assuming a 5-minute seamless transfer at Burbank) adds 17% to the total trip time of a San Jose-Burbank trip and 44% to a Fresno-Burbank trip. While it is expected that improvements will be made to increase speeds on the diesel portion, the train averages 35-45mph during the route compared with an average speed of 139 mph between San Jose and Fresno and 147 mph between San Jose and Burbank, a difference which might be psychologically frustrating for a potential HSR user.

[22] Metrolink's on-time performance (defined as arriving within 5 minutes of scheduled time) was 94.9% in February 2015 (SCRRA 2015). The Pacific Surfliner's on-time performance in all of 2014 was 80.5% (Amtrak 2015).

	Maintain conventional diesel service
Level Boarding and Interoperability	Since diesel power has no impact on the door height of the trailer cars, whether or not to maintain diesel service has no impact on level boarding. Between 2022 and 2028, high-speed rail commuters would have to transfer at Burbank for trips south to Los Angeles (24 minutes from Burbank) and Anaheim (67 minutes from Burbank). With diesel service maintained, it would be difficult for Antelope Metrolink trains to use the upgraded high-speed rail infrastructure between Palmdale to Burbank (which cuts travel time between Palmdale and Burbank from 80 minutes to 15).
Ability to modify service levels	Metrolink and Amtrak California would face the same issues they do today in adding frequency until 2028. Once HSR service is extended to Los Angeles, HSR trains will have to navigate around slower moving and slower-accelerating diesel passenger service and freight trains.
Shared corridor with Xpress West	Xpress West will likely have less competition (since Metrolink will likely remain on their own non-HSR tracks) on the Palmdale-Burbank section of a route into Los Angeles, but will face higher congestion between Burbank and Los Angeles for reasons stated above.
Integrated Southern California Rail Network	HSR would essentially be run as its own separate system outside of the blended corridor. However, outside of that, integration would still be possible with timed transfers at key stations. The distinct propulsion systems would give agencies less incentive, however, to pursue compatible platforms (just as a lack of compatible platforms would give agencies less incentive to adopt electric operation).
High-frequency, uniform HSR and commuter service	With diesel service, Metrolink and Amtrak California trains cannot easily intermix with HSR trains and the opportunity to offer supplemental HSR via commuter connections and overlapping service diminishes. Unlike Amtrak California today, HSR will be a premium service in Southern California and thus unlikely to be part of a “Rail2Rail”-type program. Run-through tracks at Union Station would need to be segregated (two for HSR, two for diesel service) unless all four were electrified. This could tighten capacity.

Bond Measure Satisfied	Sharing tracks with slower diesel service might slow down actual travel times, but theoretical travel times will not be affected. To meet the theoretical headway requirement, the signal system will need to be upgraded to accommodate high levels of mixed and variable traffic. In terms of meeting revenue-neutrality, having Metrolink and Amtrak California provide slow, diesel-powered connections (24 minutes + Transfer Time + Transfer Penalty) at Burbank to the hub at Union Station might cut into ridership and revenue for HSR.
Minimize costs and build system quickly	Keeping the Surfliner and Metrolink on diesel operation moves the expense of electric catenary infrastructure to a later date and saves the cost of buying new locomotive equipment. The timeline for a one seat ride from Northern California and the Central Valley remains as scheduled in 2028. There is the additional cost of hosting an HSR terminal in Burbank from 2022 to 2028 before the terminal moves south to Los Angeles Union Station.

Choice 2— Convert the Blended Corridor to All-Electric Passenger Operation

Upgrading Amtrak California and/or Metrolink to electrified propulsion will not be cheap. Also, there is an issue of electrification outside of the corridor—since all Metrolink and Amtrak California trains operate on segments outside the corridor as well as within the corridor, there is the question of whether to electrify only the shared HSR portion of the corridor, electrify the LOSSAN corridor, or electrify the entire Southern California Passenger Rail network. The last two options are likely cost-prohibitive and the utility of electrifying tracks that have relatively low passenger usage is minimal[23].

In Figure 6-7, Paul Dyson illustrates the Rail Passenger Association of California and Nevada's plan to electrify part of the blended corridor; Dyson refers to the project as "Electrolink" (2014). This takes advantage of the existing and under construction double-track sections of railroad that could sustain high levels of electric service. In this scenario, the Surfliner would maintain diesel operation, but conceivably could be run less frequently since the electric service provided by Metrolink would be more frequent.

Between Union Station and Fullerton, CA, the corridor is owned by the BNSF railway. This might make electrification more institutionally challenging. Unlike on the Peninsula, the freight railroads in Southern California operate double-stack cars that require high overhead catenary clearance.

Amtrak California or Metrolink could run with dual-mode locomotives paid for by connectivity funding from HSR. Alternatively, the entire Southern California passenger rail network could feed into a dedicated "electric shuttle" that runs exclusively on the blended corridor. This, however, would necessitate smooth timed transfers that would likely need to be facilitated by shared platforms.

---

[23] For example, Metrolink operates just six daily round trips on its Riverside line

	Convert Metrolink (and potentially the Pacific Surfliner) to blended service on the Burbank-Anaheim corridor by the time HSR reaches Burbank
Level Boarding and Interoperability	Metrolink trains can enjoy benefits of electrified system including reduced travel time from Palmdale to Los Angeles on the Antelope Valley Line. Metrolink trains could provide transfers from HSR to smaller stations with reduced travel times.
Ability to modify service levels	Adding frequency would be easier because of increased homogeneity of service characteristics (speed and acceleration) on the corridor. Metrolink would be able to run express trains that provide HSR-like travel times within the region.
Shared corridor with Xpress West	With faster Metrolink service from Palmdale to Downtown Los Angeles, Xpress West will have less incentive to seek through-service on to Los Angeles, but if they do, there are less “slow-moving trains” in their path to Union Station.
Integrated Southern California Rail Network	With similar train operating characteristics and infrastructure requirements, it is much simpler for HSR to share track with electrified Metrolink and Surfliners. This will enable better use of capacity at Union Station and smoother operations along the blended corridor. The electrified blended corridor will not improve the connection at Oceanside between Metrolink and COASTER.
High-frequency, uniform HSR and commuter service	With electric service, Metrolink and HSR trains can intermix and provide a more comprehensive service plan. A program similar to, but more expansive than, Rail2Rail will allow for passengers use both systems. Speed homogeneity will allow for better use of available track capacity. The Burbank-Union Station trunk may need capacity increases to accommodate higher levels of Metrolink and HSR services, especially when freight continues to operate on the line.
Bond Measure Satisfied	Metrolink and Surfliner electrification will not negatively impact the ability of the CHSRA to meet Proposition 1A requirements, but In terms of meeting revenue-neutrality, electrification of the corridor will allow for a one-seat ride from the Bay Area to Los Angeles and Anaheim as early as 2022. Metrolink service can act as a more effective feeder service for HSR by reducing total trip time.

Minimize costs and build system quickly	There might be savings in accelerating the construction of the catenary system between Burbank and Los Angeles and Anaheim. By extending to Los Angeles, the CHSRA avoids the prospect of a temporary terminal in Burbank. The largest additional expense will be converting Metrolink and the Surfliner to electric operation in the corridor. Institutional hurdles will need to be addressed, especially the issue of interacting with freight on both publicly-owned and freight rail-owned infrastructure.

6.3.4 Decision D: Tracks on Peninsula

Choices 1 and 2—Two Track Corridor or Two Track Corridor with Midline Overtake

We will compare the two-track corridor and the two-track corridor with a midline overtake in the same table. The two track corridor is the existing system (plus electrification, positive train control, and the downtown extension) which actually has two four-track areas meant for Baby Bullet overtakes. The midline overtake option includes building an additional four-track section on the middle of the corridor that allows for HSR overtakes of Caltrain services.

	Two Tracks or Two Tracks with midline overtake
Level Boarding and Interoperability	Two Tracks—Since the two-track corridor constrains capacity, this drives the CHSRA and Caltrain to pursue level boarding. The lack of capacity afforded to the high-speed rail on the Peninsula will mean that interlining HSR passengers onto trains might become a priority. This, in turn, will dictate the importance of a smooth transfer at San Jose and again encourages the concept of interoperability. Midline Overtake—Level boarding is less critical with the increased infrastructure, but the transfer at San Jose might be important if the demand for HSR services grows
Ability to modify service levels	Two Tracks—With a two-track corridor, it will be very difficult to add service. LTK estimates the most that the corridor can handle “as is” is 8 trains per hour with even 30- minute headways for HSR. If it becomes apparent that HSR needs more service to match demand, it will be very difficult and costly to build midline overtake tracks without disrupting service on the Peninsula. In 2004, when overtake tracks were built for Caltrain’s Baby Bullet project, the agency suspended weekend service—this would likely be financially unviable for high-speed rail and weekend travel demand. Midline Overtake --Slightly easier with more track infrastructure, but signal and run-time delays already surfacing in LTK analysis. Grade crossing delay will be exacerbated.
Shared corridor with Xpress West	Two Tracks—This design choice sets a precedent for the idea of “common infrastructure.” However, if the HSR operator starts to experience loss of revenues due to inadequate service on the Peninsula and into San Francisco, the likelihood of sharing revenues with Xpress West on the Palmdale-Los Angeles portion diminishes Midline Overtake —both effects are diminished in comparison to the two-track option
Integrated Southern	Two Tracks—Limited infrastructure requires careful planning of transfers in San Jose. This could create down-line scheduling conflicts in Southern California and require trains to be truncated at

California Rail Network	Burbank to avoid slot conflicts with other passenger rail or freight. Midline Overtake —the effect is diminished in comparison to the two-track option
High-frequency, uniform HSR and commuter service	Two Tracks—LTK analysis reveals that high-frequency, uniform-headway service very difficult with two-track system (see Chapter 4). Like the effect on level-boarding, this might encourage more serious discussion on interlining and cooperation between CHSRA and Caltrain. Baby Bullet service is probably not feasible because of timetable disruption it causes on other Caltrain services. Ten Caltrain trains per hour is not feasible. Again, in Southern California, HSR has to take precedence to ensure well-timed connections at San Jose. The limited ability to add HSR service possibly makes it easier for Caltrain to access the Transbay Transit Center. Midline Overtake —Baby Bullet service might be feasible (according to LTK analysis). Baby Bullet trains could be used as a superior substitute for HSR service than local Caltrain services making a transfer at San Jose more palatable for a time-sensitive HSR customer.
Bond Measure Satisfied	Both options can only theoretically satisfy the bond measure if the tracks are upgraded to 125mph operation. Two Track—Limited service HSR trains during peak hours could take away ridership from system. Given the importance of San Francisco to intra-California traffic, airlines will likely continue to provide capacity on the market and take away pricing power of an HSR operator. Midline Overtake —More HSR trains per hour would allow for better competition with airlines and allow for more ridership/revenue capture.
Minimize costs and build system quickly	Two track—Minimizes infrastructure cost on Peninsula and avoids likely political battles with expansion and decreases construction timeline. This is only true if HSR never determines that it needs to expand, because building the midline overtake tracks after operation begins will be expensive, time-consuming, and politically challenging. Since Baby Bullet service is infeasible, Caltrain now has more incentive to switch to an all-EMU fleet (Caltrain has planned to use diesel trains to continue Baby Bullet service). This might increase costs in the short-run since the diesel fleet is retired earlier than the expected useful

	lifetime, but long-run benefits of single fleet type might outweigh costs. Midline Overtake —political opposition to adding tracks will prolong timeline and there is extra cost associated with more tracks.

Choice 3—Four Track Corridor

The four-track option is very unlikely, but should be examined in order to contrast the results with the two-track options. Legislators and their constituents on the Peninsula, despite overwhelmingly being in favor of Proposition 1A, made their voice heard in Sacramento. The four-track option, with its trenches and viaducts cutting through the neighborhoods along the Peninsula Corridor, is a threat to land values on the Peninsula and the least preferable option.

	Four Tracks
Level Boarding and Interoperability	Level boarding for Caltrain is unlikely because HSR will have dedicated tracks and no incentive to integrate and cooperate. Caltrain would lose incentive to procure compatible trains. Caltrain loses benefit of shared platforms at Transbay Transit Center Level boarding in Southern California unlikely because Northern California would be getting dedicated stations (sets precedent).
Ability to modify service levels	Adding service is easiest for HSR on the Peninsula. Caltrain is still capacity constrained and might have difficulty accessing HSR tracks for passing (Baby Bullet overtakes) In Southern California, there is more flexibility to add frequency in the capacity constrained environment since schedule choices in the South do not need to be coordinated as carefully with timetables on the Peninsula Corridor.
Shared corridor with Xpress West	Sharing corridor is probably easier because of less northern timetable constraints, but CHSRA will lack framework for sharing infrastructure in Northern California. Higher San Francisco revenues for CHSR might encourage political leaders to advocate for Xpress West service through to Los Angeles (since the competition would not interfere as much with California HSR operator revenues)
Integrated Southern California Rail	There is a lack of a framework for sharing infrastructure and a dedicated corridor in Northern California might encourage a similar dedicated corridor in Southern California. This separate HSR system will leave little incentive for integration among other Southern California passenger rail

Network	providers.
High-frequency, uniform HSR and commuter service	For HSR service, high-frequencies and uniform headways are much easier due to lack of Caltrain interference. Caltrain will likely loses opportunity to run as many Baby Bullets as HSR track will be "reserved" for HSR
Bond Measure Satisfied	Best opportunity for bond measure to be satisfied since there is new infrastructure to allow for upgraded signals, fast speeds, and revenue-optimal HSR service levels in Northern California.
Minimize costs and build system quickly	Costs will escalate well-above $68B since this requires two additional tracks (likely above grade or in trench) in very dense urban corridor Politically challenging as the reversal to a four-track system requires reversal of nine-party MOU explicitly calling for a predominantly two-track corridor

6.4 Effect of Decisions on Other Local Decisions

Not one of these design decisions is completely independent of other choices that will be made in the next few years. In this section, we will assume a certain design choice gets made first and then examine its impact on the rest of the design decisions. In each of the following matrices, we look at one design decision and speculate as to the likely effect that decision has on other design choices.

6.4.1 Decision A—Platform Height and Equipment Coordination

<table>
<tr><th></th><th>Capacity Allocation Process</th><th>Southern California Electrification</th><th>Tracks on the Peninsula</th></tr>
<tr><td>1—Railroads acts alone</td><td>A lack of coordination between railroads regarding platform height might imply that there will be a similar lack of coordination regarding service planning until absolutely necessary</td><td>Since there is no chance of interoperability with distinct platform heights, It is very unlikely that the blended corridor will be electrified earlier than 2028 or that Metrolink would ever electrify its trains</td><td>Since separate platforms for Caltrain and HSR creates a distinct HSR service pattern, a midline overtake will likely be necessary</td></tr>
<tr><td>2—CHSRA coordinates with Caltrain</td><td rowspan="3">With separate entities behaving as “one railroad system,” it becomes increasingly likely that precise service planning and capacity negotiations can take place before construction</td><td>Caltrain coordination with CHSRA might set a precedent for Southern California to attempt to coordinate with CHSRA regarding roll-out of electrification.</td><td rowspan="2">Depending on the level of interlining and integration between HSR and Caltrain, a midline overtake may not be useful</td></tr>
<tr><td>3—CHSRA coordinates with Caltrain and Southern California railroads</td><td rowspan="2">The benefits of interoperability would be fully realized if the entire corridor is electrified as soon as possible</td></tr>
<tr><td>4—CHSRA coordinates with Caltrain, Southern California Railroads, and other California rail agencies</td><td>With such high levels of integration, keeping the corridor on two tracks is much more feasible</td></tr>
</table>

6.4.2 Decision B—Capacity Allocation Strategy

	Participants in Platform Height Discussion	**Southern California Electrification**	**Tracks on the Peninsula**
1—Do not develop a capacity allocation strategy	Without a capacity allocation strategy, railroads may not realize their true capacity constraints and will be less likely to pursue coordination on platform height and interoperability	Combining HSR with an electrified corridor through to Union Station would require careful capacity planning, so it would be easier to avoid interlining the railroads and initially truncating HSR service at Burbank	The exact needs of HSR will not be clear so it is less likely that Caltrain and the CHSRA will try and justify a mid-line overtake expansion
2—Create a codified capacity allocations strategy	A codified capacity allocation mechanism reveals the challenges of segregating some infrastructure and blending other infrastructure, so this might facilitate increased coordination between all passenger railroads.	Because Metrolink would need to provide higher service levels to facilitate a Burbank transfer, capacity would need to be completely re-negotiated or re-allocated twice (once in 2022 and once in 2028 when Phase 1 is complete) with freight railroads. It would be politically easier to integrate (i.e. electrify) immediately and negotiate only one time.	With a better understanding of 1) which operator gets access to the track at what time and 2) what kind of integration there will be between CHSRA and Caltrain, the specific needs of the Peninsula in terms of infrastructure will be realized
3—Negotiate capacity	Careful service planning will reveal the benefits of high levels of integration		

6.4.3 Decision C—Southern California Electrification

<table>
<tr><th></th><th>Participants in Platform Height Discussion</th><th>Capacity Allocation Process</th><th>Tracks on the Peninsula</th></tr>
<tr><td>1—Conventional (Diesel Corridor)</td><td>With slower diesel service, the only advantage of a similar platform height is a quicker HSR-Metrolink transfer, decreasing the incentive to collaborate</td><td>A capacity allocation process is less likely because Metrolink, UP, Amtrak, and BNSF already have capacity agreements in place and HSR initially stays out of this shared corridor</td><td rowspan="2">On one hand, a slower trip (i.e. transferring) into Los Angeles could justify a slower trip into San Francisco (fewer tracks), but at the same time , the delay experienced in Southern California could encourage better performance on the Peninsula (more tracks)</td></tr>
<tr><td>2—Electrification</td><td>This large investment in Metrolink will encourage CHSRA and Metrolink to seriously consider converting the Metrolink system so it is HSR-compatible</td><td>The investment required in electrifying early will encourage the CHSRA and Metrolink try and maximize utility of the infrastructure. This will encourage new capacity allocation agreements.</td></tr>
</table>

6.4.4 Decision D—Tracks on the Peninsula

	Participants in Platform Height Discussion	Capacity Allocation Process	Southern California Electrification
1—Keep corridor as is	With a two-track corridor, Caltrain and HSR will practically be forced into collaborating on platform height since capacity is so limited. This might encourage similar collaboration in Southern California	With increasing capacity on the Peninsula (from constrained two track corridor moving towards dedicated HSR capacity), there is less incentive to create a capacity allocation procedure or carefully plan service	Since not all HSR trains will be able to serve San Francisco due to line capacity constraints, it is likely that the same strategy could be applied in Southern California and Los Angeles, reducing the need for early electrification
2—Expand the corridor to include passing tracks	Increased capacity on the Peninsula removes pressure for platform sharing and integration		Increased one-seat rides in San Francisco increase pressure for one-seat rides into Los Angeles, making the case for early electrification
3—Revert to the four-track option	Platform sharing unlikely with 4-track option because the only benefit is increased flexibility at the Transbay Transit Center		Because platform sharing is unlikely, it is also unlikely to electrify Southern California early because the utility of an electrified Metrolink decreases

6.5 Conclusions

In this chapter, we take a step back and ask what needs to happen to bring California the rail network it needs to sustain its future economic prosperity. We identify seven “wish list” items ranging from interoperability to a train to Las Vegas to high-frequency, uniform-headway commuter rail service. We then review four key design decisions that are going to be made by the CHSRA and its commuter rail agency partners in the next few years of project development. For each of these decisions, we evaluate how a certain design decisions affects the probability of one of these wish list aspirations to be realized.

Agreement on platform height, as mundane as the topic appears at the surface, unlocks much capacity in terms of integration between commuter and high-speed rail systems. However, this requires careful planning of revenue splits and service patterns across the system. Service levels cannot be planned one system at a time as has been done in the past—capacity in California is too precious and high-speed rail needs adequate rail capacity to be successful. Without a guaranteed access plan, a high-speed rail investor will demand a risk premium to operate the service. Revenue neutrality will be put at risk, a risk ultimately absorbed by the taxpayers of California.

In Southern California, we examine the impact of a temporary terminal in Burbank. We agree with the Rail Passenger Association of California and Nevada that there are significant ridership risks to forcing passengers to transfer to conventional diesel service to reach Los Angeles and Anaheim. However, there is also a risk in converting Metrolink (and possibly the Pacific Surfliner) to electrified service. One possibility is simply accelerating construction phasing so that a one-seat ride to Los Angeles on electrified track comes on-line in 2022 along with Burbank, but in this case, high-speed rail might lose an important feeder system in Metrolink.

The elephant in the room in Southern California is the freight presence in the blended corridor. Unlike in Northern California, where freight is a tenant, in Southern California, between Los Angeles and Anaheim, the freight operator—specifically BNSF—is an owner. Furthermore, this subdivision is vital to BNSF as it is a link between the Ports of Los Angeles/Long Beach and its sprawling Hobart Intermodal Yard. Before any kind of blended service expansion gets implemented in Southern California, the BNSF Railway's needs will need to be addressed.

On the Peninsula, in their blended service analysis, LTK Engineering stressed the importance of a long, mid-line overtake track. Leaving the corridor as is has implications that will be felt across the state. Interregional travelers will have limited service into Los Angeles and capacity negotiations will be strained. While the limited capacity on the line will undoubtedly push the Caltrain and CHSRA to collaborate on level boarding and integration. However, without level boarding and with the corridor remaining as is, the service quality experienced by both the HSR traveler and the Caltrain commuter will be hamstrung.

Having the same vehicle floor height is only the first step. For instance, Caltrain and high-speed rail could have the same platform height and share platforms in the Transbay Transit Center which would reduce the impact of delays or mechanical failures in either system. However, the two operators could go further. High-speed rail could offer customers the opportunity to transfer to a Caltrain train with a simple cross platform timed transfer at San Jose to a Caltrain express. As long as baggage could be quickly transferred, this would practically be a pure substitute for a high-speed rail ride into the Transbay Transit Center. Reinhard Clever challenges California to take integration a step further. His proposal for Caltrain to share tracks and platforms with the San Francisco Municipal Transportation Agency and share stations BART is bold and creative yet requires significant institutional coordination and acceptance of one standard over another and ceding of institutional jurisdiction, a non-trivial hurdle for agencies with clear definitions of scope.

In the last section of this chapter, we examine design decisions on other design decisions—in this complex rail network, no one single decision can be made without having some impact on other choices to be made in the future.

In 2009, the CHSRA, in coordination with Parsons Brinckerhoff, developed a comprehensive set of technical standards that define how high-speed rail will function. This includes everything from platform height to cabin layout to station security. Of course, 2009 is prior to 2012 when the blended system philosophy was laid out in the CHSRA

Business Plan. It would be worthwhile for the CHSRA to revisit these standards now that integration is more important than ever to a successful California network.

We thank the reader for continuing with us through this exploration of the challenges and opportunities that California's present and future rail network faces. In our concluding chapter, we will review these challenges and opportunities and offer a path for California to proceed, both on the corridor and statewide level.

7 Conclusion

As we arrive at our conclusion, we thank the reader for their continued attention and hope that they have found this to be a valuable reading thus far. Supposing that the reader is just joining us or is reading this thesis intermittently, we will briefly review the topics we covered in the first five chapters. Finally, we will discuss our findings, conclusions, and recommendations, and review the contributions we have made to the story of high-speed rail implementation in the context of California.

Without the California High-Speed Rail Authority's radical departure from the dedicated high-speed system to a blended line shared with the region's commuter railroads, a decision that was urged by the governor, this thesis would not exist. Legislators determined that the dedicated line was no longer feasible due to costs and local opposition to a high-speed rail line slicing through communities. Moreover, the blended system creates many new opportunities for the State to take a new look its rail network beyond the high-speed line.

7.1 Brief Review of First Five Chapters

7.1.1 Chapter One: Introduction to Railroad Capacity, HSR, and California

In this chapter, we discuss the importance of rail capacity management as the demand for passenger and freight rail grows. In the second half of the chapter, we introduce California and its rail geography. We end the chapter by explaining the motivation behind this research: to improve capacity management, and to understand how local decisions regarding capacity affect the California rail network's ability to move both intraregional and interregional travelers as well as freight.

Importance of Capacity Management

The railroad is anything but an antiquated 19th century technology. America's freight rail network is prevailingly relevant as it carries 40% of the nation's freight (as measured by ton-miles carried). Not only does rail provide a "green" alternative to carrying freight by truck, it also reduces congestion on our nation's highways. Ton-miles carried per mile of track have more than tripled since 1980--meaning our track networks are busier than ever. Simultaneously, passenger rail traffic has increased by all measures. These two growth trends have led to a proliferation of shared track corridors where freight lines and passenger lines share rail infrastructure. It is important to study capacity on shared corridors because the congestion and delay costs affect both freight operators and passenger rail agencies. Furthermore, we are finding it difficult to increase capacity through adding infrastructure because 1) urbanization of the country has created a scarcity of adequate right of way for railroads, 2) finding funding for these megaprojects has become politically and financially infeasible.

California is finding itself as "Exhibit A" in representing the trends of passenger and freight rail growth. In Southern California, the Ports of Los Angeles and Long Beach are a massive origin and destination of freight rail traffic for the entire country. The BNSF and Union Pacific Railroads share tracks with local commuter rail and Amtrak service as they move goods across the Los Angeles Basin. In Northern California, Caltrain finds itself in a capacity crunch on the Peninsula, as it confronts the mixed blessing of standing-room only trains and year-over-year ridership increases. California, the world's 8th largest economy, is experiencing transportation capacity challenges on all fronts. The Los

Angeles-San Francisco short-haul air market is both one of the busiest and most delay-prone in the U.S. Expanding runway capacity is both politically and financially infeasible. Because of the urbanization of the state, adding freeway lanes is also a costly task; the recent expansion of the I-405 freeway in Los Angeles cost the state $1.1 billion for 20 lane-miles of highway.

High-speed Rail and California

California, along with some other states, views high-speed rail as one of the most effective alternatives to meet rising transportation demand. We discuss why California and its voters believed that an investment in high-speed rail was the right choice for the state in 2008 and the advantage that high-speed rail has over air transportation, its main source of ridership.

Introduction to California's Rail Geography

We conclude Chapter 1 with an introduction to California and its current rail geography. California hosts a wide array of passenger rail services as well as being a freight transportation hub. Amtrak California operates three intercity routes, of which the *Pacific Surfliner*, running in the San Luis Obispo-Los Angeles-San Diego (LOSSAN) corridor, enjoys the highest ridership of any Amtrak line outside the Northeast Corridor. In Northern California, two commuter rail lines (PCJPB's *Caltrain* and SJRRC's *Altamont Commuter Express)* carry passengers into and within Silicon Valley and the San Francisco Peninsula. In Southern California, the Southern California Regional Rail Authority (SCRRA) operates a web of Metrolink commuter lines reaching across the region from its Los Angeles Union Station hub, and the North County Transportation District's *COASTER* provides coastal commuters service into San Diego's Santa Fe Depot. Finally, Amtrak operates four long-distance trains across the state, connecting points as far away as Chicago.

A main theme of California's rail network, is "one line, multiple owners." Though California commuter rail agencies own nearly 50% of the track on which they operate their networks, virtually 100% of the network is shared with freight railroads that operate freight traffic 365 days per year. This creates scheduling challenges for both freight and passenger railroads now and into the future. Because the California High-Speed Rail Authority has elected to share track and corridor with these commuter railroads, the future high-speed rail operator will not escape the complex negotiation process with California's freight railroads.

7.1.2 Chapter Two: California HSR and Sharing Capacity—Literature and Practice

In this chapter, we review literature regarding sharing capacity, capacity allocation mechanisms, and the California high-speed train project. We then discuss the practical challenges of sharing capacity whether between high-speed rail and conventional passenger rail or between conventional passenger rail and freight. We end the chapter with an overview of European shared corridors, where separation of infrastructure, ownership, and train operations has been mandated by the European Commission.

Capacity Allocation and Pricing

In this section of the literature review, we review different forms of capacity allocation discussing Stephen Gibson's three distinct mechanisms in practice today—administered, cost-based, and market-based—as well as Patricia Perennes' discussion of rail-specific challenges in managing capacity. We look at novel proposals for capacity allocation mechanisms and how they might be applied on all, or part of the California system.

Integration and Institutions

We then introduce the concept of integration, using Reinhard Clever's TRB paper on the "six levels of integration." This theme of integration is important throughout the remainder of this thesis. We also introduce the concept of competition versus collaboration in public agencies through a paper written by Meyer et al. While the CHSRA is not, per se, competing with any of the other passenger rail agencies in the state, it will inadvertently, by nature of rail capacity, be competing for access on shared corridors.

California HSR Research

There is a great deal of literature previously written on the California high-speed train project, though not that much research has been performed since 2012 when the CHSRA radically changed their project delivery approach from fully-dedicated line to blended system. We categorize the current research into three groups: 1) studies regarding the project's feasibility, revenue, and ridership, 2) the relationship between California HSR and the state's air transportation network, and 3) novel approaches to high-speed rail service in the state.

7.1.2.1 Shared Corridor Challenges

There are numerous challenges to sharing track, which we separate into physical, operational, and institutional challenges. Physical challenges include rail geometry constraints, and clearance issues for platforms and overhead catenary. Operational challenges include the difficulty of accommodating heterogeneous train speeds, implementing positive train control among different railroad operators, and finding time to perform on-track maintenance on busy corridors. Lastly, institutional challenges include priority rules for operators and the competition versus collaboration for railroads providing passenger service.

7.1.2.2 Vertical Separation versus Integration

In final section of the literature review, we review the multitude of research looking at the effects of unbundling (vertical separation) versus traditional vertically integrated railroads. Both sides have advocates and detractors, with the general consensus being that "the jury is still out" on the optimal institutional structure. What is certain, however, is that the European Union is performing a grand experiment on unbundling with its railway directives.

7.1.2.3 European Union Case Studies

The European Union's directive to separate railway services from railway infrastructure ownership with minimal further instructions to member-states has led to a variety of institutional arrangements between infrastructure managers and operators across the European Union. In this section, we examine the "network statements" of different infrastructure managers which govern access to the railway network. In particular, we look at managers in Belgium, Germany, the Netherlands, Portugal, Spain, Switzerland, and the United Kingdom. Much of the competition thus far, has been "off-track," but in Italy, we discuss on-track competition between high-speed rail operators. We discuss the relevance of the European Union experience to California and conclude that, while current institutional arrangements preclude vertical separation in California, European models could serve as models for new capacity allocation mechanisms.

7.1.3 Chapter Three: Northern California Blended Service

In this chapter, we discuss the past, present, and future of the Peninsula Corridor in northern California. The blended service on the Peninsula is going to be capacity constrained and the San Francisco northern terminus of the high-speed rail line is critical for both Peninsula commuters and intrastate high-speed rail travelers.

History and Today's Peninsula Corridor Joint Powers Board

In this section, we provide a brief history of the San Francisco-San Jose rail line and introduce the institution that has managed the rail service—branded as *Caltrain*— on the Peninsula for nearly 30 years: the Peninsula Corridor Joint Powers Board. The PCJPB, made up of the three counties served on the 87-mile network, owns the northernmost 51-miles of track between San Francisco 4th and King Station and San Jose Rod Diridon Station; Union Pacific uses the corridor and maintains perpetual trackage rights. The PCJPB and its relationship with Union Pacific, the CHSRA, and the private operator on the corridor is already and will continue to be instrumental in the ultimate form of high-speed rail in California.

Current and Future Ridership Growth (Electrification and Transbay Transit Center)

We then discuss a "state of the system" for Caltrain. The railroad is enjoying huge ridership growth due, in part, to the thriving economic conditions in Silicon Valley and San Francisco and the millennial generations' desire to leave the car at home or not own a car at all and be productive on the daily commute. The Baby Bullet express service has provided reasonable travel times, but has also drawn many commuters to bring their bikes on-board—Caltrain sees more bicyclists on-board than any other transit agency in the nation.

We discuss two projects that have the potential to change commute patterns and drive ridership even higher. First, an electrification of the corridor will allow for more local services--making the train a more convenient option. Second, the extension of Caltrain to the Transbay Transit Center (the Downtown Extension) will bring Caltrain riders much closer, via a one-seat ride to San Francisco's job center.

"The Blended System"

After a review of Caltrain and its ridership growth, we introduce the implications of welcoming high-speed rail to the corridor as part of a "blended system." In this blended system, high-speed trains will share tracks with commuter rail agencies on the north and south "bookends" of the line; aside from track speed upgrades and potential construction of passing sections, the Peninsula Corridor will remain very much the same as it is today. The change from separated to shared infrastructure was made in response to ballooning costs and public outcry from residents in affluent Peninsula communities who felt the construction of a dedicated high-speed line would destroy property values. The switch has implications for some of the requirements of the bond measure: that a high-speed train can travel from San Francisco to San Jose in 30 minutes, that it can operate at 5-minute headways, and that the operation will require no public subsidy.

The Importance of the Transbay Transit Center and Peninsula to HSR

In the last section of the chapter, we discuss criticality of the Transbay Transit Center to Caltrain as a ridership source and destination and to the high-speed operator as a revenue generator. The two operators will compete both spatially for the six-track terminal and temporally during the rush hour peak. Currently, the plan is to have two dedicated Caltrain tracks and four dedicated HSR tracks, but the Transbay Joint Powers Authority is urging the two agencies to consider adopting standard vehicle height to allow for platform sharing.

7.1.4 Chapter Four: Current Levels of Coordination and Train Operator Model

In Chapter Four, we review publically available documents regarding the partnership between the California High-Speed Rail Authority and the Peninsula Corridor Joint Powers Board. We then turn to the Transbay Transit Center, which will be owned by the Transbay Joint Powers Authority, a third-party entity with no formalized allegiance to either of the two rail operators. We apply the train operator model previously developed for the Northeast Corridor and apply it to the Transbay Transit Center and Downtown Extension. We conclude that a high-speed rail operator would have a much higher willingness than a commuter rail operator to pay for access to the important San Francisco station.

LTK Blended Operations Analysis and Memorandum of Understanding

We review two significant documents regarding blended operation: the memorandum of understanding between the PCJPB and the CHSRA and the blended operations analysis performed by LTK Engineering. The Memorandum of Understanding highlights that 1) freight will remain operating on the blended corridor, 2) the track configuration, aside from a few passing tracks, will remain the same, 3) Caltrain service will remain operational during the construction of the blended system, and 4) the PCJPB will retain ownership of the corridor assets (aside from to-be-construction Downtown Extension and Transbay Transit Center). LTK's blended service analysis shows that headways for Caltrain become very uneven with the addition of just one HSR train on the corridor, and that more than two HSR trains per hour will require the construction of a midline overtake track.

The Transbay Transit Center—Comparisons to Penn Station

In this section, we compare the Transbay Transit Center to Penn Station, an existing terminal with multiple operators. The Hudson River tunnels that bring New Jersey Transit and Amtrak trains into the station are at capacity, making peak hour schedule changes difficult for both operators. One key difference between Penn Station and the planned Transbay Transit Center is that, while Amtrak owns Penn Station and operates trains in and out of the terminal, the Transbay Joint Powers Authority will not operate any trains. We also discuss the challenges of scheduling commuter rail with high-speed rail on the Northeast Corridor and review the institutional arrangement between Metro North Railroad and Amtrak.

Application of TRB Paper Model

Using this knowledge that the northernmost 1.3 miles of the high-speed rail line will be a vertically separated structure, we apply the vertically-separated train operator model developed in our TRB paper. We determine that it would be difficult for the high-speed rail operator and Caltrain to compete on a level playing field for access to the Transbay Transit Center due to the fiscal strength of the high-speed rail operator. We conclude that it is much more likely that capacity will need to be negotiated between all parties, rather than auctioned to the highest bidder.

7.1.5 Chapter Five: The Southern California Network and Blended Plans

In Chapter Five we introduce the Southern California rail network and the planning that is underway for the arrival of high-speed rail. We discuss the Los Angeles Union Station terminal and the upcoming Southern California Regional Interconnector Project that will add four run-through tracks for HSR and commuter rail. As in Northern California, high-speed rail operations will be blended with commuter rail operations. However, in Southern California, there are more operators, freight-owned right-of-way, and no plans to electrify non-HSR operations. The development of blended service plans is nascent in Southern California, even more so than the Northern California system.

7.1.6 Chapter Six: Measuring the Statewide Impact of Local Decisions

In Chapter Six, we amalgamate our research on shared system challenges and the northern and southern ends of the California HSR network and construct a "wish-list" for the future California rail network. We then identify four upcoming decisions that are likely to be made on the local level—the platform height standard, the decision whether to electrify Metrolink and phase the project in Southern California, whether or not to adopt a formal capacity allocation mechanism (at the Transbay Transit Center or system wide), and whether to construct additional tracks on the Peninsula Corridor. We then evaluate the impact of each choice on the "wish-list" as well as the impact of each choice on the other choices. We find that items as mundane as platform height can have powerful impacts on the ultimate performance of the California passenger rail network.

This concludes our brief review of this thesis. We now move into the conclusions section of this chapter where we present the results of our analysis of the California rail system and provide recommendations to passenger rail agencies across the state.

7.2 Conclusions

7.2.1 Conflicting institutional priorities stand in the way of a unified California rail network

The California rail network is currently fragmented and there is much opportunity for interagency coordination. The Rail2Rail program in Southern California is a positive first step, but that program could be improved through expansion. The Southern California Regional Rail Authority (SCRRA) could work together with the North County Transit District (NCTD) to ensure timed transfers at Oceanside, or better yet, through-run trains to provide commuters options that are more affordable than buying an Amtrak California ticket. Agencies are often protective of their own assets and there is no overseeing agency that implores agencies to work together; as Clever notes in his paper on integration, "By dividing up project planning into separate professional disciplines studying engineering/capital costs, ridership/operating costs, and environmental impacts, sight of the system as a whole is lost. System wide ridership studies are completed without knowing the exact station locations or the level of integration with other modes" (Clever 12). This individualistic agency mindset will not work in a blended system in the future.

7.2.2 Northern California is further along in the planning process than Southern California

While the CHSRA announced the change to a blended system for both Northern and Southern California at the same time, the southern blended system lags the northern counterpart in definition. Granted, the Southern California network is much more complex than its northern counterpart; but since the Los Angeles Basin will see high-speed rail service six years before the Bay Area does, it is surprising that the north leads the south in this regard. Perhaps this is due to higher levels of political pressure in Northern California, but the fact remains that there is no blended operations analysis or significant memorandum of understanding signed between SCRRA, the CHSRA, Amtrak California, or the Class I freight railroads. In Northern California, this early planning has been beneficial because it has brought to the forefront questions such as platform compatibility and ultimate track layout. However, the current fuzziness of the Southern California system is also an opportunity, since all options are still available for consideration.

7.2.3 Shared corridor challenges most important to California

Congestion and Delays

The blended system will put high demands on the existing infrastructure. Commuter rail or freight delays can propagate to the HSR system. As the ratio of service volume to capacity tends towards 1, the system loses stability and on-time performance suffers. With degraded on-time performance and uncertainty regarding arrival times, schedule padding becomes necessary and makes rail as a mode less attractive to time-sensitive consumers.

The Caltrain corridor today experiences delays due to two main factors: aging equipment malfunctions and on-track suicides[24]. While some of the oldest equipment will be retired once the electric equipment arrives in 2020, Caltrain plans to continue operating some diesel-electric equipment concurrently with the electric vehicles. Aside from the two different fleet types making maintenance more time-consuming, this could lead to delays when aging equipment breaks down and a specific fleet type needs to be substituted. And since grade crossings will not be eliminated, on-track suicides will continue to cause significant delays, though ongoing prevention programs could help minimize this impact.

CPUC Requirements

There are two major California Public Utility commission regulations that stand in the way of an integrated rail system. The first is a regulatory framework for high-voltage operation on shared corridors. The CHSRA developed a framework for high-voltage operations on its dedicated corridors, but operations in blended corridors with freight and grade crossings have yet to be addressed. If this fails to be resolved, the CHSRA would be forced to either build dedicated right-of-way as originally planned or truncate its operations in Burbank and San Jose and forced travelers to transfer to commuter rail systems.

Overhead catenary wire presents clearance issues for freight railroads in Southern California. Freight railroads cannot operate in certain tunnels with double-stack containers if overhead catenary wires are not high enough to provide adequate clearance. It will also be challenging to construct the improvements necessary for the high-speed rail system on the BNSF-owned right of way in Southern California as freight railroads demand high levels of track availability.

The second CPUC requirement is the freight train lateral clearance requirement. This standard requires adequate lateral clearance for freight trains at platforms. The platforms of the SPRINTER light rail system in San Diego County are evidence of this requirement in action. The SPRINTER shares track with freight railroads but also uses high platforms. To meet the ADA maximum 3" gap requirements and satisfy the CPUC minimum gap requirements, the SPRINTER employs mechanical platform "gangways" that descend during light-rail operation during the day, and retract at night to allow for freight trains to pass. The CHSRA has opted for separate stations to avoid this CPUC requirement. Using the same stations as the commuter rail lines would mean that the CHSRA must find a compatible platform height and urge for CPUC rule changes. Freight railroads will likely contest rule changes because it will limit their operational flexibility to haul wide loads. However, the cost savings of not having to build HSR-specific stations are substantial and the rule changes are something the author believes are worth working hard to achieve.

Track Ownership and Priority

The blended system dictates that the CHSRA will operate on a multi-owner network, much like Amtrak does today. Instead of having sole control of its infrastructure, the CHSRA will have to work with the TJPA, PCJPB, SCRRA, and BNSF Railway to ensure smooth operation. As a result, the CHSRA will face many of the issues Amtrak faces today regarding train priority. The ability for the CHSRA to operate in a reliable fashion will depend on the priority rules that the CHSRA can negotiate with its host railroads on the blended corridors.

---

[24] Though the suicide rate fluctuates from year-to-year, Caltrain claims it has the highest suicide rate in the nation (Brotherhood of Locomotive Engineers and Trainmen 2003)

7.2.4 The blended service on the Peninsula as it stands today is infeasible

While neither agency would describe their relationship as a competitive one, the HSR operator and Caltrain will be competing for track access and access to the downtown San Francisco terminal. The blended operations analysis, though PCJPB emphasizes that it is a feasibility study, not a service plan, shows that Caltrain will likely have limited access to its most important station and that the HSR operations create uneven headways on the corridor. Furthermore, without passing track construction on the mid-peninsula (through the very same communities that fought against the four-track corridor), CHSRA can only operate two trains per hour, per direction into the terminal. The separate stations for high-speed rail up and down the Peninsula will make it difficult for passengers to "interline" or use both systems as envisioned in the 2012 CHSRA Business Plan describing blended service. A renewed dedication to blended operations is beginning to form with the current discussion on Caltrain's new electric equipment; the two agencies need to keep this momentum while moving forward towards a true shared system.

7.2.5 The blended system as planned does not match the former aspirations of Caltrain or HSR

The blended system was a significant scale-back of the high-speed rail system sold to voters in 2008, and it is evident that both the service plans of Caltrain and the CHSRA are nowhere nearly as customer-friendly as they were before 2012. Caltrain has shared aspirations of 10 trains per hour, per direction on the corridor, while the high-speed rail authority has shown service frequency as high as eight trains per hour per direction The LTK engineering study reported that six Caltrain trains per hour was feasible and four HSR trains was possible only with significant construction. Due to the importance of Caltrain to everyday commuters, significant construction of passing tracks is exceptionally challenging: the Ponderosa Project to build the passing tracks for Baby Bullet service required weekend shutdowns of the Caltrain system for two years This means that once a high-speed operator is running revenue service along with increased Caltrain service, performing track construction to expand capacity on the line or in the Transbay Transit Center will be more challenging than ever before.

The CHSRA asserts that they are still meeting the bond measure requirements that were approved by voters in 2008. However, the language of the measure suggests a very different system than what is on the table today. When the voter read the text "Achievable operating headway (time between successive trains) shall be five minutes or less" it would have been reasonable for the voter to assume that service levels would be somewhat higher than two high-speed trains per hour (30 minute headway) into San Francisco.

7.2.6 The impact of the phased approach has consequences that go beyond the phase

Truncating the initial operating segment in the San Fernando Valley for six years is a hugely important decision, especially if the CHSRA chooses to operate in separate station facilities than Metrolink. Metrolink service will need to match high-speed rail trains with much higher off-peak service levels than today to ensure a timed transfer. Even so, a smooth transfer will be very difficult because Metrolink will be at a separate platform, meaning that without costly facilities, baggage transfer will be difficult. Because transit connections are poor in the San Fernando Valley, the high-speed operator will demand high levels of parking to enable access for residents across the Southland. Building these parking facilities for a temporary terminal will impact land-use around the station site; the positive impacts of transit-oriented development, a key driver in ridership, will not be realized. This could affect long term ridership and the ability of the high-speed rail operator to generate revenues to cover operating costs.

7.2.7 Freight Impacts

It could be argued that a strong rail freight network is just as important to California's economy and congestion mitigation as a strong passenger one. The freight railroads provide a key role in transcontinental goods movement and one freight train removes hundreds of trucks from our nation's highways. It is for this reason that the freight railroads cannot be "pushed aside" for passenger service. On the Peninsula Corridor, freight will seek to continue operation on the corridor; however, increased corridor use from high-speed rail will narrow operating windows for freight service and finding time for track maintenance will be more difficult than it is today. In Southern California, BNSF will likely require high-speed rail to expand capacity on BSNF's San Bernardino subdivision between Los Angeles and Fullerton, CA. Negotiations for capacity with freight railroads can become contentious when freight railroads feel that they are losing flexibility to run trains efficiently through their network when necessary. However, if passenger rail can use freight infrastructure in a way that does not impede freight trains from operating efficiently, freight railroads would stand to benefit financially from increased access charges.

7.3 Contributions

This thesis contributes in two main ways to the development of shared corridors in California. First, it presents a way forward for the California rail network that will satisfy passenger rail agencies, freight operators, and both intraregional and interregional travelers. Second, it creates a framework for reviewing the system impacts of seemingly local choices; these local decision decisions ultimately impact the realization of this future California rail network.

California's Rail "Wish-List"

- ***Level Boarding and Interoperability***
- ***Ability to modify service levels***
- ***Xpress West integration with California HSR***
- ***Integrated Southern California network***
- ***High-frequency, uniform-headway HSR and commuter rail***
- ***Satisfy 2008 bond measure requirements***
- ***Minimize costs and build the network quickly***

It would be unfair to write that California has not done significant planning for its rail future. The 2013 California State Rail Plan, a requirement of the 2008 Passenger Rail Investment and Improvement Act (PRIIA) acts as a vision statement for the future of rail transportation in the Golden State. The plan provides a broad overview of potential challenges and the expected capacity requirements of future passenger and freight rail service. The 400-page plan emphasized integration and coordination across passenger rail agencies and between passenger rail agencies and freight rail operators. The plan lists proposed service levels as well as goals for new passenger rail routes such as, an Amtrak *Coast Daylight* operating between San Francisco and Los Angeles or new intercity service connecting the Coachella Valley with Los Angeles. While this thesis does not aim to validate or invalidate the feasibility of individual service level goals, it does draw broad conclusions regarding goals for the California rail network as a whole.

Based on planning documents and board meeting minutes of those individual passenger rail agencies, the seven "wish-list" items we identify, like the State Rail Plan, reflect the importance of coordination and integration among railway operators. These seven "wish-list" items, while not self-evident, would generally be agreed upon by most passenger rail agencies in the state, though some items are more important to certain agencies than others. We believe that if all seven "wish-list" items were realized, California could boast both a politically popular and cost-effective statewide rail service. Some of these "wishes" are correlated with one another; for example, if passenger rail agencies pursue interoperability, an integrated Southern California network is much more likely. On the other hand, some of these goals conflict with one another to a certain extent as well: high-frequency, uniform-headway HSR and commuter rail will likely cost more due to increased infrastructure costs. The aim of this thesis is not to state whether or not one goal is superior to another, but rather to understand the impacts of design decisions

The second contribution of this thesis is a developing a process for looking at local impacts of design decisions. It is difficult to quantify flexibility, nor would any quantification have very much inherent value—in the end, there will be one operating plan and to have more infrastructure than necessary would be wasteful (just as having less infrastructure than necessary will be costlier in the future). However, we do observe how each decision affects future decisions and whether or not the decision takes certain rail goals "off the table." California agencies could work together to refine their priorities instead of using our inferred priorities and could apply other larger-scale design decisions to this process as well.

7.4 Recommendations

7.4.1 Agency-specific

California High-Speed Rail Authority

Procure low-floor vehicles if manufacturers can supply them

The CHSRA is currently committed to purchasing service-proven technology. There are no high-speed trains currently in service that are 1) capable of 220mph operation and 2) a low-floor vehicle. If the commuter rail agencies are to pursue level boarding and compatibility with high-speed rail, having a low-floor vehicle has the potential to reduce cost across the network. This is because Caltrain and Metrolink currently operate the same Bombardier equipment and could simply raise platforms to accommodate their existing equipment[25][26]. If funding for level boarding falls short, it could be implemented at individual stations that would stand to benefit the most from its implementation. Even if Caltrain and Metrolink never raise platforms, high-speed rail could operate at commuter rail stations (without level boarding). This will save Caltrain the expense of procuring specialized electric vehicles with two sets of doors that match both existing and future platform heights.

There are clear risks for the CHSRA to try and conform to existing California vehicle floor heights as opposed to existing HSR vehicle floor heights. Vehicle costs of a new technology could be much higher due to reduced competition of vehicle suppliers. Lack of service-proven technology means that California would be the "guinea pig;" that is, they would be the first to experience any issues with reliability and safety of new equipment.

The Parsons Brinckerhoff technical memorandum advocating for high platforms for high-speed rail was released in 2009 when the plan for HSR service was to operate in a completely dedicated right-of-way. Since that time, the blended system demands a fresh approach and vision for the entire system and how it will integrate with existing rail services. By adopting a low-floor vehicle specification for its trainsets, the CHSRA will show its agency partners that it is committed to the success of blended operations and encourage them to follow suit and develop a similar attitude. Some risks like new low-floor high-speed rail vehicles would be worth taking.

Avoid terminating service temporarily in Burbank

The CHRSA's plan of terminating service from 2022 to 2028 before extending service to Anaheim and Los Angeles poses a risk. We show in Chapter Five that the transferring to conventional diesel service from Burbank to Los Angeles and Anaheim will add a non-trivial amount of time in addition to a perceived "transfer penalty." There are also associated costs with operating a temporary terminal station. Maintenance and baggage transfer facilities will be necessary; and because a Burbank/San Fernando Valley station has relatively minimal transit connectivity, the station will require a lot of additional parking for high-speed rail customers. The relative lack of access from the southern portion of the Los Angeles

[25] Metrolink's new Hyundai Rotem equipment that is replacing the Bombardier equipment has the same 25" floor height
[26] The *FrontRunner* commuter rail service in the Salt Lake City, Utah area currently operates the same Bombardier coaches and has level boarding.

Basin might initially crimp ridership and revenue. Either a private investor or the CHSRA (i.e. the State) itself will have to fund the cost of additional time for ridership and revenue to materialize after the section to Anaheim and Los Angeles is finished (presumably in 2028).

The CHSRA will have project acceleration costs if it chooses to build the initial operating segment system to downtown Los Angeles. The additional cost will have to be borne by the Authority and not a private investor, since the Authority claims that investment is contingent upon successful revenue service. At the time of writing, the Authority has only $12.5 billion in committed funding (in addition to carbon tax revenues, which were $250 million last year). This means the CHSRA only currently has 40% of the Phase 1 cost, estimated at $31 billion. Pursuing more public funding on top of the $31 billion to finish the railroad to Los Angeles might be political infeasible. However, if the CHSRA determines it is possible, it should prioritize completing the connection into Los Angeles Union Station.

Formulate a coherent blended service strategy at both ends of the line

In order to reduce the risk for the private operator that will ultimately run the system, the CHSRA needs to understand exactly what blended will entail. Will there be a Metrolink train meeting every HSR train in Burbank and will that train continue on to Anaheim from Los Angeles? Will Caltrain customers be able to use HSR trains instead of Caltrain trains between San Jose and San Francisco? How many HSR trains per hour will the high-speed operator be allowed to run during the peak hour into downtown San Francisco? The CHSRA needs to resolve these questions before making infrastructure decisions such as passing tracks in Northern California or a temporary Burbank Terminal, and definitely before it tries to seek a private operator to invest in the capital construction costs.

Peninsula Corridor Joint Powers Authority

Consider the impacts of competing for access with a statewide HSR operator

The Peninsula Corridor Joint Powers Board owns the right-of-way and track between San Francisco's 4th and King Station and San Jose and the Transbay Joint Powers Authority will own 1.3-mile Downtown Extension into the Transbay Transit Center. This seemingly puts the PCJPB/Caltrain on "home turf" when it comes to access negotiations since they represent San Francisco's 150-year old commuter railroad. However, the CHSRA serves a much larger, statewide constituency, than the three-counties that enjoy Caltrain service. Furthermore, the PCJPB structure is a weak because it requires voluntary contributions from the three counties; it has no dedicated funding source. The PCJPB and CHSRA currently have a healthy partnership, but ultimately, Caltrain will be working in partnership with the private HSR operator, not the CHSRA. Since the operator will be facing revenue and ridership pressure, and since Caltrain operates at their peak service levels at the most valuable times for high-speed rail, this relationship could potentially become more adversarial. In order to avoid this, the PCJPB and CHSRA need to carefully negotiate access, or at a minimum, agree on a method for objectively allocating and pricing capacity.

Grow stronger working relationship with Southern California

Whether or not Metrolink or Amtrak California decide to electrify operations on the blended corridor in Southern California, they will encounter many of the same issues faced by the Caltrain in the north. For example, there will be negotiations regarding service levels and access to railroad assets such as the run-though tracks at Los Angeles Union Station. As owners of tracks and right-of-way that will see frequent use from high-speed rail, both the Caltrain and Metrolink stand to benefit by sharing effort in developing the ultimate form of the blended system. Both agencies have overlapping interests: they want to protect rail service for the everyday commuter, they want to maximize the benefit of HSR connectivity funding, and they want to keep positive relationships with the freight railroads and other passenger railroads with whom they share infrastructure. Man-hours could be saved by working through HSR compatibility challenges together.

Service plan as soon as possible (try examples of better integration) and develop metrics to evaluate output timetable

In their blended service planning document, Caltrain writes that their next step is to look at service plan potions prior to fleet need. However, Caltrain is already beginning the procurement process for their vehicles. Service planning needs to come first and Caltrain needs to develop these service plans with the CHSRA. Planning should include multiple levels of integration, from shared platforms all the way to shared trainsets so the two agencies can understand the value of integration investments. For example, some HSR trains could replicate Caltrain's Baby Bullet service and stop a five or six intermediate stations between San Francisco and San Jose instead of two. Even though HSR trains acting as "Baby Bullets" making limited stops would hurt the utility of the high-speed service, it is the superior alternative to truncating service in San Jose and forcing HSR travelers to transfer to commuter rail. As LTK concluded, the capacity is simply not there for greater than two HSR trains per hour and Caltrain Baby Bullet service.

To evaluate the timetable from a customer perspective, the CHSRA and Caltrain should develop metrics which factor customer travel time, headway uniformity, and passenger residential and job data among other measures. Caltrain's origin-destination data is not reflective of consumers' preferred stations; rather, it is reflective of a Baby Bullet service which pulled riders to high-service stations. Timetable metrics would allow the two agencies to objectively compare one service plan to another and quantify the benefits of one timetable that may require the construction of passing tracks versus another.

Address the freight issues—CPUC horizontal clearance and operating windows

The freight issues are not going to disappear any time soon—the PCJPB has affirmed through its statements and in the memorandum of understanding with the high-speed rail authority that freight is going to remain on the Peninsula Corridor. By not addressing the freight issues, the PCJPB is implicitly encouraging the CHSRA to use separate facilities since the waiver is a necessary condition for Caltrain to have level boarding, but not one for HSR if it believes that separate facilities are sufficient for its own operation.

Currently Union Pacific is guaranteed one 30-minute window between 10:00 A.M. and 3 P.M. each day to run freight trains on each of the northbound and southbound tracks of the Peninsula Corridor. Between midnight and 5:00

A.M., one main track is reserved for freight use. With increased service due to HSR, these window capacity allocations may need to be revisited. As owner of the right of way, it is the PCJPB's responsibility to manage these operating window constraints before these issues affect Caltrain and HSR's service commitments.

Transbay Joint Powers Authority

Consider impact of reduced Caltrain service to Transbay Transit Center

Limited Caltrain service to the Transbay Transit Center will be as painful to the developers and the Transbay Transit Center as it will to the commuters themselves. Being able to boast a frequent one-seat ride to the points south on the Peninsula will raise rents for the property owners and developers as well as revenues for retail spaces, and as a result, revenue for the City of San Francisco. It is in the best interest of the developers at Transbay Transit Center to ensure that Caltrain has adequate access to the terminal.

Ensure there is a path for improvements to the Downtown Extension

One of the primary concerns raised in the environmental impact report for the Transbay Transit Center was the small size of the six track terminal. For a terminal station, six tracks is minimal—Los Angeles Union Station has 14 tracks and Penn Station in New York City has 21. In the 2002 Environmental Impact Report/Environmental Impact Statement, the Transbay Joint Powers Authority wrote that "additional turnaround or "tail" tracks will greatly assist in relieving congestion at the platform tracks." In 2008, the Transbay JPA announced that the "tail tracks would be deferred until operationally required." This means that the TJPA anticipates a need for track capacity, but will seek funding sources later. This raises interesting questions about who will be responsible for increasing capacity and who will fund that additional capacity. Will one operator be forced to choose between a large capital expenditure and service reductions, and if so, which operator? Because the TJPA is a separate entity, neither the CHSRA nor Caltrain has a "right" to access the vital terminal. If the CHSRA is required to fund capacity improvements, should the cost fall on the private operator or the public? The Transbay Transit Center should answer these questions sooner rather than later to help both the Caltrain and the CHSRA plan for service and future expansions of service.

Rail Agencies in Southern California

Quantify to the greatest extent possible, the benefits and costs of level boarding, platform sharing, electrification, and increased integration in terms of increased ridership and connectivity

The Southern California Regional Rail Authority, Amtrak California (the LOSSAN JPA), and even NCTD should, along with the CHSRA, evaluate the benefits and costs of level boarding, platform sharing, electrification, and increased integration across the greater Los Angeles area. Level boarding and platform sharing clearly provide a benefit for transfers to-and-from high-speed rail and allow for maximum use of limited capacity at Union Station, but the benefits of level boarding outside the blended corridor are markedly less. Increased integration could result in benefits such as a one-seat ride between Los Angeles and San Diego or Las Vegas, and could help commuter rail serve as a better first-mile and last-mile connection for high-speed intercity travelers; that is, both modes will see ridership gains with an integrated system. If "Blended Service" means integration between HSR and commuter rail operators, can Southern California operators integrate non-HSR service as well? Perhaps cooperation with the high-speed rail authority can help foster increased cooperation between Metrolink, Amtrak California, and COASTER.

Protect the run-through tracks at Los Angeles Union Station

In the author's opinion, the run-through tracks at Union Station will be a hugely important asset. In terms of cost per minute of travel time saved for through trains, the Southern California Interregional Connector Project is one of the most cost-effective rail projects under construction in the state. Unfortunately, only four tracks of the 14 at Union Station are going to "run-though" and it is likely that high-speed rail will need to use at least two of them. These tracks are going to be more valuable for regional commuter services as well: overnight, Metrolink will gain much more flexibility with service planning and timed-transfers and Amtrak California can now offer customers travelling from the northern half of the LOSSAN corridor to the southern half nearly 15 minutes off of their trip. It is important that the return on this $350 million capacity investment is maximized; CHSRA should demonstrate their plan to achieve high levels of track utilization before LACMTA (the owner of Union Station) allows high-speed rail to monopolize half of the run-though track improvements. However, it would be ideal to maintain all four tracks for general use through a consistent low-floor platform height.

7.4.2 General Recommendations

Integrated operations have a potential to bring HSR a revenue source; like regional airlines feed into international hubs, so too can commuter rail services feed into interregional high-speed rail services. To that end, service planning should drive infrastructure decisions; in an era where infrastructure such as new HSR stations or electrification are expensive and the public is leery of megaprojects, California can set an example with a well-conceived (i.e. well service planned), integrated rail system.

This integration is important from a risk management perspective as well. The CHSRA is fortunate to have their lead cheerleader in the statehouse: Governor Jerry Brown is the by far the most important political ally that the Authority has at its disposal. Unfortunately for the CHSRA, Governor Brown concludes his term in 2018 and it is unlikely that his successor—whether Democrat or Republican—will be such an ardent supporter of the project. At the time of writing, there are court cases that threaten the CHSRA's ability to use Proposition 1A funding on the grounds that the current system is not what was promised to voters. If the funding were to evaporate, it would be much better to be left with a useable system than a standalone section of high-speed track in the Central Value with little independent utility. The CHSRA was right to start construction in the Valley as they are filling key a rail gap between the Central Valley and the Los Angeles Basin; however, if this piece does not fit into the rest of the California network; this will truly be a "train to nowhere" as project critics attest.

7.5 Future academic research

In this section we will identify future academic research that would benefit both the rail stakeholders in California, but also future high-speed rail projects across the United States

7.5.1 Combine with research on stakeholder analysis to look at feasibility of certain institutional cooperation in California (and possibly generalize to other corridors)

In this thesis we have identified areas where coordination and integration would have great benefits, but we did not evaluate how likely it is that certain entities will cooperate with one another. Research that evaluated the feasibility and likelihood of cooperation based on characteristics of each entity in California would be beneficial as it would identify less-than-obvious institutional synergies. There is current research in M.I.T.'s Regional Transportation/High-speed Rail group that is looking at stakeholder analysis on the Northeast Corridor. This research, could of course, be expanded into a broad evaluation tool for stakeholder willingness to cooperate.

7.5.2 Revenue sharing and Integration between public and private sectors

One of the challenges with a blended system is how to share passenger fare revenues, especially between a private operator and a public agency. For example, if Caltrain transports a high-speed rail passenger who came from Fresno between San Jose and San Francisco, what share of that total fare revenue should Caltrain receive? Should Metrolink collect an additional fare (e.g. baggage fees) from a connecting HSR passenger between Los Angeles and Burbank in addition to the basic one-way fare it already collects? Though the idea of offering multi-agency tickets exists today in the U.S., the degree of complexity that the CHSRA is discussing in terms of interlining and through-ticketing does not yet exist in the U.S. rail market. Airlines already practice some form of fare proration between operators in offering code share flights, but this is between two private companies. Research that identifies some of the institutional and regulatory challenges of cooperating with a private for-profit operator would be useful and valuable for both the CHSRA and its public agency partners.

Outside of the CHSRA, high-speed rail is an attractive concept that has garnered private sector interest. The Texas Central Railway Company is currently planning on constructing a fully-private high-speed line between Houston and Dallas. Xpress West is a private company, yet is hoping to receive some federal loans to complete construction of a Palmdale-Las Vegas HSR link. We noted earlier how Las Vegas and Los Angeles would both stand to benefit from a one-seat ride between the two cities, however this would require the CHSRA to allow Xpress West to access tracks in between Palmdale and Los Angeles. CHSRA may be interested in running trains on Xpress West's infrastructure between Palmdale and Las Vegas as well. This type of private HSR/public HSR coordination is unique and will require a new institutional paradigm. Research examining how this relationship might work best from a political, legal and customer-satisfaction standpoint would be valuable and applicable across the country.

7.5.3 California Rail Authority

Japan is frequently cited as a leader in high-speed rail technology. Unlike California, however, Japan's institutional structure is one that is fully integrated. Each of the four "JR" companies that operate high-speed service, operate unprofitable commuter rail service as well. Timed transfers are commonplace and on-time performance is high, and operations on the system level are profitable. One potential avenue of research would be to see how well the Japanese model could be exported to California. This would reduce the complexity of capacity allocation as well as challenges associated with revenue sharing interlining and platform compatibility. However, it would require local authorities to cede control. The Peninsula Corridor Joint Powers Board, which lacks a dedicated source of funding might be amenable to being folded into the CHSRA; however, some politicians and constituents would be wary of ceding local control to a state agency[27].

7.6 Final Thoughts

We thank the reader for joining us by reading this thesis. We hope it has been a fulfilling experience and one that has sparked interest in the evolving California high-speed rail project. We hope that public officials in leadership positions at the passenger rail agencies we discuss in this thesis have the opportunity to read and understand the gravity and impact of

[27] The PCJPB depends on contributions from the three counties in which it operates service.

their decisions on the future of the state of California. The blended system was a radical change, but also one that presents a Golden State opportunity. Instead of building the California high-speed rail *line*, we have the opportunity to optimizing the California rail *network*.

California needs to take a "top-down" approach to high-speed rail, asking "what do we want as a state and how do we get there?" rather than a bottom-up approach which asks "how does this part of the system need to work for high-speed rail to be successful?" Taking this top-down approach has a huge impact on the system the state ultimately delivers. Optimizing locally constrains the system optimal solution; and as California agencies make decisions regarding vehicle fleets and track investments with Proposition 1A connectivity funding, we can see that the local optimizing is already beginning. Decisions made on the Peninsula Corridor can create capacity bottlenecks that affect HSR trains on an interregional level. And since these blended issues are being addressed presently on the Peninsula Corridor with the concurrent electrification of Caltrain, a precedent is being set for their southern neighbors. The Peninsula Corridor, therefore, is an important proving ground for whether or not blended operations can work.

The high-speed line from San Francisco to Los Angeles has a price tag of $68 billion. The CHSRA's business model depends on over half of that money—$37 billion—coming from a private operator bidding for a concession contract after the initial operating section from Merced to Burbank via Fresno and Bakersfield operates profitably. However, much of this blended uncertainty creates risk that will cost the state when the time comes to seek investors. The risks include not gaining access to key track stations, political opposition to new construction, and poor regional rail feeder connectivity. The CHSRA should focus on reducing the risk when it comes time to find a private operator. Investments made by the CHSRA and partner agencies in integration will pay dividends as the high-speed rail project moves forward. California is a land of dreamers, but in reality, it will be small and seemingly mundane decisions, like platform height and shared facility use, that will bring this project to a successful conclusion.

Works Cited

$2 Million Track Upgrade Assists Port Businesses. *Port of Redwood City*. N.p., Nov. 2009. Web. 6 May 2015. <http://www.redwoodcityport.com/p7iq/html/TrackImprove_pressrelease.htm>.

"2004 Memorandum of Understanding between the California High Speed Rail Authority (CHRSA) and the Peninsula Corridor Joint Powers Board (PCJPB)." Memo. N.d. *Caltrain*. Web. 8 May 2015. <http://www.caltrain.com/Assets/Peninsula+Rail+Program/2004_MOU_Between_CHSRA_and_PCJPB.PDF>.

2010 Caltrain On-Board Bike Count and Dwell Time Summary Report. San Francisco: Corey, Canapary and Galanis Research, 2010. *PCJPB*. Web. 13 May 2015. <http://www.caltrain.com/Assets/__Agendas+and+Minutes/BAC/Materials/2011/Bike+Count+$!26+Dwell+Time +Report.pdf>.

2014 Service Planning Methodology. San Francisco: Parsons Brinckerhoff, 2014. *CHSRA*. Web. 14 May 2015. <http://www.hsr.ca.gov/docs/about/business_plans/ BPlan_2014drft_Service_Planning.pdf>.

"ACEforward." *Altamont Corridor Express*. San Joaquin Joint Powers Authority, Mar. 2013. Web. 13 May 2015. <http://www.acerail.com/About/Public-Projects/ACEforward>.

"Airline Origin and Destination Survey (DB1B)." *RITeA*. Bureau of Tranportation Statistics, Sept. 2014. Web. 14 May 2015.

"Amtrak Pacific Surfliner On-Time Performance." *Amtrak On-Time Performance*. Amtrak, Mar. 2015. Web. 14 May 2015. <http://www.amtrak.com/ pacific-surfliner-train&mode=perf&overrideDefaultTemplate=OTPPageVerticalRouteOverview>.

ADIF (2013). 2013 Network Statement. Spain. Web accessed on March 20, 2014.

Allocation of Passenger Rail Liability for Inicidents on Shared Use Corridors. Washington, D.C.: AR010 Intercity Passenger Rail Subcommittee, 2010. *AR010 Intercity Passenger Rail Subcommittee*. Web. 13 May 2015. <http://ar010.york.cuny.edu/sub/research/research-problem-statements/ Allocation%20of%20Passenger%20Rail%20Liability.pdf>.

Amtrak. *FY2014 Budeget and Business Plan. FY 2015 Budget Request Justification. FY2014-2018 Five Year Financial Plan*. Washington, DC, 2014.

"An Introduction to Positive Train Control." *Metrolink*. Southern California Regional Rail Authority, Feb. 2014. Web. 13 May 2015. <http://www.metrolinktrains.com/agency/page/title/ptc>.

A.M. Lago, P. Mayworm, J.M. McEnroe. Transit Service Elasticities: Evidence from Demonstrations and Demand Models. Journal of Transport Economics and Policy, Vol. 15, No. 2, pp.99-119. May, 1981

ARTIC Level-Boarding Accessibility Compliance Assessment. Anaheim: Parsons Brinckerhoff, 2012. *Orange County Transportation Authority*. Web. 13 May 2015. <http://www.octa.net/uploadedFiles/MainSite/Content/About/Public_Committees/SNAC/20120828ARTICAssessment.pdf>.

Associated Press. "New San Francisco-Oakland Bay Bridge opens at last to replace the one damaged in 1989 earthquake after 12 YEARS of construction." *Associated Press* [San Francisco] 3 Sept. 2013: *The Daily Mail*. Web. 14 May 2015. <http://www.dailymail.co.uk/news/article-2409952/New-San-Francisco-Oakland-Bay-Bridge-opens-12-YEARS-construction.html>.

Baker, Jim. *Positive Train Control*. San Francisco: The Joint Council on Transit Wireless, 2012. *Transit Wireless*. Web. 13 May 2015. <http://transitwireless.org/wp-content/uploads/2012/05/PTC_whitepaper_may2012_ver2.pdf>.

Ball, Jeffrey. "The Proportion of Young Americans Who Drive Has Plummeted—And No One Knows Why." *The New Republic* 12 Mar. 2014: n. pag. *The New Republic*. Web. 13 May 2015. <http://www.newrepublic.com/article/116993/millennials-are-abandoning-cars-bikes-carshare-will-it-stick>.

Barba, Michael. "SFO Ranks Worst Airport When it Comes to Weather Delays." *SF Weekly* 21 Nov. 2014: n. pag. *SF Weekly*. Web. 7 May 2015. <http://www.sfweekly.com/thesnitch/2014/11/21/ sfo-ranks-worst-airport-when-it-comes-to-weather-delays>.

Baritaud, Manuel, and Dana Stefanescu. "Infrastructure Access Pricing for Subsidised Train Operators." *Think-up (Thematic Network to understand mobility prediction), Workshop" Setting the policy context*. 2000.

Belobaba, Peter, Amedeo Odoni, and Cynthia Barnhart, eds. *The Global Airline Industry*. Vol. 23. John Wiley & Sons, 2009.

Bialick, Aaron. "A Straighter Extension of Caltrain/HSR into Downtown SF: Is It Worth It?" *StreetsBlog SF*. N.p., 27 June 2013. Web. 6 May 2015. <http://sf.streetsblog.org/2013/06/ 27/a-straighter-extension-of-caltrainhsr-into-

downtown-sf-is-it-worth-it/>.

"BNSF Corporate Fact Sheet." *BNSF Railway Company*. N.p., Feb. 2015. Web. 13 May 2015. <http://www.bnsf.com/about-bnsf/pdf/fact_sheet.pdf>.

Borndorfer, Ralf, et al. "Auctioning Approach to Railway Slot Allocation, An."*Competition & Reg. Network Indus.* 7 (2006): 163.

Brennan, Peter J. "Traffic to Las Vegas Keeps Los Angeles Gamblers Stuck in Desert." *Bloomberg* 3 Apr. 2006: n. pag. *Bloomberg*. Web. 14 May 2015. <http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aye_ZS2Q1oYk>.

Brownstone, David, Mark Hansen, and Samer M. Madanat. *Review of "Bay Area/California High-Speed Rail Ridership and Revenue Forecasting Study"*. No. UCB-ITS-RR-2010-1. 2010.

"Building Area per Employee by Business Type." *U.S. Green Building Council*. N.p., 13 May 2008. Web. 7 May 2015. <http://www.usgbc.org/Docs/Archive/General/Docs4111.pdf>.

C.A. Nash, S. Coulthard, B. Matthews. Rail track charges in Great Britain – the issue of charging for capacity. *Transportation Policy*, Vol. 11, pp. 315-327, 2004.

"California election results." *Los Angeles Times* 4 Nov. 2008: n. pag. *Los Angeles Times*. Web. 13 May 2015. <http://www.latimes.com/news/local/la-2008election-california-results,0,3304898.htmlstory>.

"California High-Speed Rail Big Picture." *California High-Speed Rail Authority*. CHSRA, Aug. 2014. Web. 13 May 2015. <http://www.hsr.ca.gov/docs/newsroom/fact%20sheets/Big_Picture_FINAL_082214.pdf>.

"California Topographic Map." Map. *World Atlas*. Graphic Maps, 15 Nov. 2013. Web. 13 May 2015. <http://www.worldatlas.com/webimage/countrys/namerica/usstates/ca.htm>.

California High-Speed Rail Business Plan. 2012. hsr.ca.gov. Web. 13 May 2015

California. California Public Utilities Commission. *Order Instituting Rulemaking regarding Whether to Adopt, Amend, or Repeal Regulations Governing Safety Standards for the Use of 25 kV Electric Lines to Power High Speed Trains*. 2014. *California Public Utilities Commission*. Web. 6 May 2015. <http://docs.cpuc.ca.gov/PublishedDocs/Efile/G000/M131/K150/131150332.PDF>.

California. California Public Utilities Commission. *Regulations Governing Clearances on Railroads and Street Railroads with Reference to Side and Overhead Structures, Parallel Tracks, Crossings of Public Roads, Highways, and Streets*. 1948. *California Public Utilities Commisson*. Web. 6 May 2015. <http://www.cpuc.ca.gov/gos/GO%2026-D.pdf>.

Caltrans. *2013 California State Rail Plan*. 2013 Leg. Sacramento: California State Transportation Agency, 2013. *California Department of Transportation*. Web. 13 May 2015. <http://californiastaterailplan.dot.ca.gov/docs/Final_Copy_2013_CSRP.pdf>.

California. *Joint Powers Agreement Peninsula Corridor Project*. 1996 Leg. *Caltrain*. Web. 2 May 2015. <http://www.caltrain.com/Assets/Public/JPA_Agreement_and_Amendment_10-03-1996.pdf>.

California. Leg. Assembly. *Proposition 1a*. 2008 Leg. A. 1a. *California Secretary of State Alex Padilla*. Web. 15 May 2015. <http://vig.cdn.sos.ca.gov/2008/general/pdf-guide/ suppl-complete-guide.pdf>.

California. Leg. Senate. 2012 Leg. S. 1029. *Official California Legislative Information*. Web. 7 May 2015. <http://www.leginfo.ca.gov/pub/11-12/bill/sen/sb_1001-1050/sb_1029_bill_20120718_chaptered.pdf>.

"Caltrain Conducts Annual Count." *Caltrain*. N.p., n.d. Web. 7 May 2015. <http://www.caltrain.com/about/news/Caltrain_Conducts_Annual_Count.html>.

"Caltrain Level Boarding Discussion." *Caltrain Board of Directors*. PCJPB, 3 Oct. 2013. Web. 14 May 2015. <http://www.caltrain.com/Assets/__Agendas+and+Minutes/JPB/Board+of+Directors/Presentations/2013/10-3-13+Level+Boarding+Update.pdf>.

"Caltrain Modernization Program." *PCJPB 2014e*. N.p., n.d. Web. 15 May 2015. <http://www.caltrain.com/projectsplans/CaltrainModernization/Modernization/ CBOSS-PTC-Project.html>.

"Caltrain Moves Forward with New Rail Car Purchase." *Caltrain*. N.p., 4 Dec. 2014. Web. 15 May 2015 .<http://www.caltrain.com/about/news/Caltrain_Moves_Forward_With_New_Rail_Car_Purchase.html>.

"Caltrain searches for answers to suicide rate." Brotherhood of Locomotive Engineers and Trainmen. N.p., 10 Jan. 2010. Web. 18 May 2015. <http://www.ble-t.org/pr/news/headline.asp?id=5460>.

"Caltrain's Project 2025: Moving Beyond the Baby Bullet." *American Public Transportation Association*. N.p., 1 Nov. 2006. Web. 13 May 2015. <http://www.apta.com/ passengertransport/Documents/archive_3297.htm>.

Caltrain / HSR Blended Service Plan Operations Considerations Analysis. San Francisco: LTK Engineering Services, 2013. *Caltrain Modernization Program*. Web. 13 May 2015. <http://www.caltrain.com/Assets/Caltrain+Modernization+Program/Blended+System/ Caltrain-HSR+Blended+Service+Plan+Ops-Con-Report.pdf>.

Caltrain / HSR Blended Service Plan Operations Considerations Analysis. San Francisco: LTK Engineering Services, 2013. *Caltrain Modernization Program*. Web. 13 May 2015. <http://www.caltrain.com/Assets/Caltrain+Modernization+Program/Blended+System/ Caltrain-HSR+Blended+Service+Plan+Ops-Con-Report.pdf>.

Caltrain Electrification. San Francisco: Morrison Knudsen, 1992. *Sonic.net*. Web. 14 May 2015. http://sonic.net/~mly/Caltrain-Electrification/1992-report-75dpi/e75-095.html

Caltrain Triennial Customer Survey. N.p.: n.p., 2013. *Caltrain*. Web. 6 May 2015. <http://www.caltrain.com/Assets/_Marketing/pdf/ 2013+Caltrain+Triennial+Customer+Survey+-+Report.pdf>.

Caltrain_Electrification_EA-FEIR_Vol-I_July_2009-WEB.pdf>."DTX Update Transbay Transit Center." *TransbayCenter.org*. TJPA, 14 Nov. 2013. Web. 16 May 2015. <http://transbaycenter.org/uploads/2013/11/Item7_Final-DTX-P3-presentation-11-14-13.pdf>.

Cauthen, Jerry, and Peter Lydon. "The Downtown Caltrain Extension: Vital to the Future of the Bay Area." *Yodeler* 27 Jan. 2015: n. pag. *The Yodeler*. Web. 15 May 2015. <http://theyodeler.org/p=10072>.

Central Subway EIR/EIS--Transportation Analysis. San Francisco: SFMTA, 2008. *Central Subway*. Web. 14 May 2015. <http://centralsubwaysf.com/FSEIS-SEIR>.

Cesarini, Andrea. "Current Challenges of Rail Regulation in Europe: The European Regulators' View." 7th Florence Rail Forum. Florence. Florence. 29 Nov. 2013. Print.

Chen, Cynthia, Don Varley, and Jason Chen. "What affects transit ridership? A dynamic analysis involving multiple factors, lags and asymmetric behaviour."*Urban Studies* 48.9 (2011): 1893-1908.

Christner, Paul J., and Ronald A. Mauri. *Sharing of Track by Transit and Freight Railroads: Liability and Insurance Issues*. Technical rept. no. FTA-05-08. Cambridge: U.S. DOT Volpe Center, 2005. *Federal Transit Association*. Web. 13 May 2015. <http://www.fta.dot.gov/ documents/Shared_Track.pdf>.

CHSRA. *Alignment Design Standards for High-Speed Train Operation*. Technical rept. no. TM 2.1.2. San Francisco: Parsons Brinckerhoff, 2009. *Caltrain-High-Speed Rail Compatibility Blog*. Web. 14 May 2015.

<http://www.tillier.net/stuff/hsr/TM%202%201%202%20Alignment%20Design%20Standards%20R0%20090326%20TM%20Excerpt%20A.pdf>.

CHSRA. *Trainset Configuration Analysis*. Technical rept. no. TM 6.3. San Francisco: Parsons Brinckerhoff, 2009. *Caltrain-High-Speed Rail Compatibility Blog*. Web. 14 May 2015. <http://www.tillier.net/stuff/hsr/TM-6.3-Trainset-Configuration-Analysis-R0-090923.pdf>.

Clever, Reinhard, and Mark Merrill Hansen. *The Competitive Advantage of High Speed Rail: Airport and Station Accessibility as a Determinant of Mode Choice*. Saarbrücken, Germany: Verlag Dr. Müller, 2008. Print.

Clever, Reinhard. "Falling Through the Cracks: The Last Mile Applying Best Practices to the San Francisco to San Jose Section of the California High-Speed Rail System." *Transportation Research Board 92nd Annual Meeting*. No. 13-1790. 2013.

Clewlow, Regina RL, Joseph M. Sussman, and Hamsa Balakrishnan. "Interaction of high-speed rail and aviation." *Transportation Research Record: Journal of the Transportation Research Board* 2266.1 (2012): 1-10.

Cohen, Michael, ed. *Summary Population Projections by Race/Ethnicity and by Major Age Groups*. Rept. no. P-1. Sacramento: California DOF, 2014. *California Department of Finance*. Web. 15 May 2015. <http://www.dof.ca.gov/research/demographic/reports/projections/P-1/>

"Commuter Rail Platform Types." *Massachusetts Bay Transportation Authority*. N.p., n.d. Web. 15 May 2015. <http://www.mbta.com/riding_the_t/accessible_services/default.asp?id=21541>.

"Competition in the Spanish Passenger Market Approved." Railway Gazette. N.p., 13 June 2014. Web. 11 Nov. 2014. <http://www.railwaygazette.com/news/passenger/single- view/view/spanish-passenger-market-competition-approved.html>.

"Construction Update." *Perris Valley Line*. Riverside County Transportation Commission, 12 June 2014. Web. 13 May 2015. <http://www.perrisvalleyline.info/uploads/media_items/ rctc-pvl-universityneighborhoodbriefing-6-12-14-rev5-final.original.pdf>.

Competition Report. Berlin: Deutsche Bahn AG, 2012. Print.

Connell, Rich. "Super-long 'monster train' rolls through Southern California." *Los Angeles Times* [Los Angeles] 13 Jan. 2010: n. pag. *Los Angeles Times*. Web. 13 May 2015. http://articles.latimes.com/2010/jan/13/local/la-me-monster-train13-2010jan13/2

Daganzo, Carlos F. "A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons." *Transportation Research Part B: Methodological* 43.10 (2009): 913-921.

Davis, Mike, Nicole Parra. California Assembly. *Safe, Reliable High-Speed Passenger Train Bond Act for the 21st Century*. 2008 Assembly A. 3034. *California Legislative Info*. Web. 13 May 2015. <http://www.leginfo.ca.gov/pub/07-08/bill/asm/ab_3001-3050/ab_3034_bill_20080826_chaptered.html>.

Day, Michael. "High-speed Dream Turns Sour on Milan to Rome Rail Route." The Independent 4 Sept. 2014: Print.

DB Netz AG (2013). Network Stament 2013. Germany. Web accessed on March 20, 2014.

Donato-Weinstein, Nathan. "Younger and wealthier, Caltrain riders opt out of traffic." *Silicon Valley Business Journal* [San Jose] 28 May 2014: n. pag. *Silicon Valley Business Journal*. Web.15 May 2015. <http://www.bizjournals.com/sanjose/news/2014/05/28/younger-and-wealthier-caltrain-riders-opt-out-of.html?page=all>.

Drew and Nash 2011-- Drew, J. and Nash, C.A. (2011). Vertical separation of railway infrastructure - does it always make sense? Institute for Transport Studies, University of Leeds, Working Paper 594

Dyson, Paul, ed. "RailPAC Issues." *Rail Passenger Association of California and Nevada*. RailPAC, 9 Dec. 2014. Web. 15 May 2015. <http://www.railpac.org/issues/>.

Eschenbach, John. "California Doubles Down on Commuter Rail." *Mass Transit Magazine* 22 Sept. 2014: n. pag. *Mass Transit*. Web. 13 May 2015. <http://www.masstransitmag.com/article/11249130/california-doubles-down-on-commuter-rail>.

"Early Milestones." *Caltrain*. N.p., n.d. Web. 30 Apr. 2015. <http://www.caltrain.com/about/ History/Early_Milestones.html>.

European Conference of Ministers of Transport. *Railway Reform & Charges for the Use of Infrastructure* Paris, France, 2005.

Faiola, Anthony. "Competition heats up on Italy's newly privatised railways." *The Guardian* [London] 12 Nov. 2013: n. pag. *The Guardian*. Web. 13 May 2015. <http://www.theguardian.com/world/2013/nov/12/italy-italo-trains-trenitalia-rail-ntv>.

Final Environmental Impact Report--Peninsula Corridor Electrification Project. Rept. no. Volume 1. N.p.: PCJPB, 2009. *Caltrain*. Web. 16 May 2015. <http://www.caltrain.com/Assets/ Peninsula+Rail+Program/Electrification+2025/

Finger, Matthias, and Andrea Rosa. Governance of Competition in the Swiss and European

Fleming, Susan. United States. GAO. *High-Speed Passenger Rail: Preliminary Assessment of California's Cost Estimates and Other Challenges*. Rept. GAO-13-163T. Washington: GPO, 2012. *Government Accountability Office*. Web. 13 May 2015. <http://www.gao.gov/assets/660/650608.pdf>.

"Flight Schedules." *Fly SFO*. SFO, n.d. Web. 15 May 2015. <http://timetables.oag.com/itt/sfo/>.

Fox, Justin, et al. *Capacity Modeling Guidebook for Shared-Use Passenger and Freight Rail Operations*. No. Project 08-86. 2014.

FRA. Federal. *Freight Railroads Background*. By Stephanie Lawrence. Washington, D.C.: FRA, 2015. *Federal Railroad Administration*. Web. 13 May 2015. <https://www.fra.dot.gov/eLib/Details/L03011>.

Gabbard, Dana. "Gabbard: History of the Surfliner, LOSSAN and a Look at Pending Legislation." *Streetsblog Los Angeles*. Open Plans, n.d. Web. 13 May 2015. http://la.streetsblog.org/ gabbard-history-of-the-surfliner-lossan-and-a-look-at-pending-legislation/

Gbenekama, Delana G. Metrolink 20th Anniversary Report. Los Angeles: SCRRA, 2012. Metrolink. Web. 18 May 2015. <http://www.metrolinktrains.com/pdfs/20thAnniversaryReport/document.pdf>.

Gibson, Stephen. "Allocation of capacity in the rail industry." *Utilities Policy* 11.1 (2003): 39-42.

Greenway, Greg. "Planning for Compatibility: Freight on the Caltrain Corridor." American Railway Development Association. 3 June 2013. *American Railway Development Association*. N.p., n.d. Web. 4 May 2015. <http://www.amraildev.com/files/Presentations/AM2013/AM2013_Session02_Greenway.pdf>.

Grenzenback, Lance R., David T. Hunt, and Daniel F. Beagan. *National Rail Freight Infrastructure Capacity and Investment Study*. Cambridge, MA: Cambridge Systematics, 2007. *Cambridge Systematics*. Web. 13 May 2015. **Error! Hyperlink reference not valid.**

Grunwald, Michael. "Back on Tracks: Freight rail's renaissance is powering the U.S. economy even without Uncle Sam's help." *Time Magazine*. Time, 9 July 2012. Web. 13 May 2015. <http://business.time.com/2012/07/09/us-freight-railroads/>.

Guo, Zhan. *Transfers and path choice in urban public transport systems*. Diss. Massachusetts Institute of Technology, 2008.

Haire A., and Machemehl R., 2007, “Impact of Rising Fuel Prices on U.S. Transit Ridership,” Transportation Research Record of the Transportation Research Board, 1992(1), pp. 11–19.

"High-speed Rail at Grade Crossings—R & D." *Federal Railroad Administration*. FRA, 2015 Web. 13 May 2015. <http://www.fra.dot.gov/Page/P0103>.

"Highway-Rail Grade Crossing & Safety Initiative." *Federal Railroad Administration*. FRA, 2015 Web. 13 May 2015. <http://www.fra.dot.gov/Page/P0096>.

Holvad, Torben. "Review of railway policy reforms in Europe." *Built Environment* 35.1 (2009): 24-42.

Hunt, David. "Creating Network Capacity Through Operational Changes." Transportation Research Board, Rail Capacity Joint Subcommittee. Hyatt Regency, Washington, D.C. 14 Jan. 2014. Speech and PowerPoint Presentation.

Hymon, Steve. "Take Metrolink to see the races at Del Mar for $28 round-trip." *The Source*. LACMTA, 22 July 2011. Web. 13 May 2015. <http://thesource.metro.net/2011/07/22/take-metrolink-to-see-the-races-at-del-mar-for-28-round-trip/>.

Independent Ridership and Passenger Revenue Projections for High Speed Rail Alternatives in California. Technical rept. no. 1680-04. Boston: Charles River Associates, 2000. *California High Speed Rail Authority*. Web. 13 May 2015. http://www.hsr.ca.gov/docs/about/business_plans/BPlan_2000_TS_RiderRevenue.pdf

Infrabel (2013). Network Statement 2014. Belgium. Web accessed on March 8, 2014.

Johnson, Daniel, and Chris Nash. "Charging for scarce rail capacity in Britain: a case study." *Review of Network Economics* 7.1 (2008).

Johnston, Bob. "Caltrain's Commuter Choreography." *Trains* Feb. 2005: n. pag. *Proquest*. Web. 6 May 2015. <http://search.proquest.com/docview/206636369/fulltextPDF/AC40DDE12F1F4352PQ/ 1?accountid=12492>.

Kasuya, Shuichi. *High-speed rail commuting in the United States: a case study in California*. Diss. Massachusetts Institute of Technology, 2005.

Keller, Matt. "Caltrain Adds 6ths car to 4 Bombardier Trains." *ABC7 News*. ABC, 11 May 2015. Web. 15 May 2015. <http://abc7news.com/traffic/caltrain-adds-6th-car-to-4-bombardier-trains/713574/>.

Kille, Leighton Walter. "Harvard’s José Gómez-Ibáñez: Key insights on the future of transportation and infrastructure." *Journalist's Resource*. Harvard Kennedy School, 5 Sept. 2014. Web. 13 May 2015.

<http://journalistsresource.org/studies/environment/transportation/ harvard-jose-gomez-ibanez-transportation-infrastructure-research-chat#>.

Knight, Rob, ed. "Union Pacific Investor Factbook." *up.com*. Union Pacific Railroad Company, Jan. 2014. Web. 13 May 2015. <http://www.up.com/investors/attachments/factbooks/2013/fact_book.pdf>.

Kopp, Quentin. California. Leg. Assembly. *High Speed Rail Act*. 1996 Leg. A. 1420. Sec. 796. 1996. *California Legislative Info*. Web. 13 May 2015. http://leginfo.ca.gov/pub/95-96/bill/sen/sb_1401-1450/sb_1420_bill_960924_chaptered.html

L. Thompson. *Charges for the Use of Rail Infrastructure* International Transport Forum. Paris, France, 2008.

Leachman, Robert C. *Regional Rail Simulation Findings Technical Appendix*. Los Angeles: Southern California Association of Governments, 2011. *Berkeley Industrial Engineering and Operations Research*. Web. 13 May 2015. <http://www.ieor.berkeley.edu/People/Faculty/leachman-pubs/FinalUpdate_RailTechAppendix_Nov2011.pdf>.

"Level Boarding Progress Report." *Transbay Center*. N.p., n.d. Web. 15 May 2015. <http://transbaycenter.org/uploads/2014/12/Level-Boarding-CAC-12-09-14.pdf>.

Levy, S., Peña-Alcaraz, M., Prodan, A., Sussman, J.M. Presentation: Transportation Research Board 94th Annual Meeting, Session 362 Intercity Passenger Rail Transportation Research (January 12, 2015) Paper: Included in conference compendium as TRB 15-1697; Accepted for publication in Transportation Research Record (issue pending)

Lewis, Heather A. "Math Mistakes That Make the News." *PRIMUS* 25.2 (2015): 181-192.

Li, Janice. "Caltrain Needs More Bike Capacity Now." *San Francisco Bike Coalition*. SFBC, 5 Aug. 2014. Web. 15 May 2015. <http://Caltrain Needs More Bike Capacity Now>.

Liberalisation and Competition in the European Regional Rail Market. N.p.: n.p., 2013. Print.

Liu, Rongfang Rachel, Fei Yang, and Mei Chen. "Understanding the shared operation of commuter rail transit and freight railroads." *Journal of the Transportation Research Forum*. Vol. 44. No. 1. 2010.

LOSSAN Corridorwide Strategic Implementation Plan. San Deigo: LOSSAN, 2012. *Orange County Transportation Agency*. Web. 14 May 2015. <http://www.octa.net/pdf/projectid_260_14371.pdf>.

Lovett, Alexander H., C. Tyler Dick, and Christopher PL Barkan. "Determining Freight Train Delay Costs on Railroad Lines in North America." *Proceedings of RailTokyo 2015* (2015).

M. Bebiano Brando, O. Petrik, F. Moura, R.F. Allard, J. de Abreu e Silva. *Simulating the impact of the high-speed network on the performance of other transport operatros: the HSR on the Lisbon-Oporto link*. In: Proceedings of the 93rd Annual Meeting of the Transportation Research Board. Washington, DC, January 2014.

M. Herrero. *The Social Cost of Rail Capacity Constraint*. In: Proceedings of the ITEA Annual Conference. Toulouse, 2014

Malouff, Dan. "Top 10 busiest Amtrak routes." *Greater Greater Washington*. N.p., 16 Aug. 2013. Web. 13 May 2015. <http://greatergreaterwashington.org/post/19882/ top-10-busiest-amtrak-routes/>.

Maxwell, Ross R. "Intercity rail fixed-interval, timed-transfer, multihub system: Applicability of the Integraler Taktfahrplan Strategy to North America."*Transportation Research Record: Journal of the Transportation Research Board*1691.1 (1999): 1-11

McCallon, Larry. "High Speed Rail Integration--Southern California." Memo to Paul Dyson. 24 Nov. 2014. *Caltrain-HSR Compatibility Blog*. Web. 13 May 2015. <http://www.tillier.net/stuff/hsr/Metrolink_Response_p1.png>.

McClellan, James. "Railroad capacity issues." *background paper for Research to Enhance Rail Network Performance: A Workshop, Washington, DC: Transportation Research Board*. 2006.

Memo. N.d. *Caltrain*. Web. 8 May 2015. <http://www.caltrain.com/Assets/ Caltrain+Modernization+Program/Documents/Executed+CHSR-JPB+2013+Agreement.pdf>.

"Metrolink Initiates Demonstration of Life Saving Positive Train Control In Commuter Service." *Metrolink*. SCRRA, 20 Feb. 2014. Web. 13 May 2015. <http://www.metrolinktrains.com/ news/news_item/news_id/901.html>.

Meyer, Michael D., et al. "Collaboration: The key to success in transportation."*Transportation Research Record: Journal of the Transportation Research Board* 1924.1 (2005): 153-162.

"Milestones-1990s." *Caltrain*. N.p., n.d. Web. 2 May 2015. <http://www.caltrain.com/about/ Caltrain150/Milestones/Milestones_1990s.html>.

Miller, Steve. "The Peninsula Commute Story." *Askmar*. N.p., n.d. Web. 30 Apr. 2015. <http://www.askmar.com/Railroad/Peninsula%201987%20Commute.pdf>.

Mizutani, Fumitoshi, and Shuji Uranishi. "Does vertical separation reduce cost? An empirical analysis of the rail industry in European and East Asian OECD Countries." *Journal of Regulatory Economics* 43.1 (2013): 31-59.

Mladjov, Roumen V., S.E. "The Most Expensive Bridge in the World." *Modern Steel Construction* Sept. 2004: *DocStoc*. Web. 14 May 2015. <http://www.docstoc.com/docs/88884329/The-Most-Expensive-Bridge-in-the-World>.

Moser, Whet. "Chicago's Terrible Rail Bottleneck." *Chicago Magazine* 14 Feb. 2012: n. pag. Web. 13 May 2015. <http://www.chicagomag.com/Chicago-Magazine/The-312/February-2012/Chicagos-Terrible-Rail-Bottleneck/>.

Nash, Andrew. *Best practices in shared-use high-speed rail systems*. No. FHWA/CA/OR-2002/26. 2003.

"National Transit Database--Data." *National Transit Database*. FTA, Jan. 2014. Web. 16 May 2015. <http://www.ntdprogram.gov/ntdprogram/data.htm>.

Neff, John, and Matthew Dickens. *2012 Public Transportation Factbook: Appendix A*. Washington, D.C.: APTA, n.d. *American Public Transportation Association*. Web. 13 May 2015. http://www.apta.com/resources/statistics/Documents/FactBook/2012-Fact-Book-Appendix-A.pdf

Neff, John, and Matthew Dickens. *2013 Public Transportation Factbook*. Washington, D.C.: APTA, n.d. *American Public Transportation Association*. Web. 13 May 2015. <http://www.apta.com/resources/statistics/Documents/FactBook/2013-APTA-Fact-Book.pdf>

Network Rail (2012). Network Code. United Kingdom. Web accessed on March 8, 2014.

Network Rail (2013). The 2015 Network Statement. United Kingdom. Web accessed on March 8, 2014.

Neumann, Lance, and Kimon Proussaloglou. *Response of the Review of "Bay Area/California High-Speed Rail Ridership and Revenue Forecasting Study."* Cambridge: Cambridge Systematics, 2010. Web. 13 May 2015. http://www.hsr.ca.gov/docs/about/ridership/ridership_forcasting_FinalUCB_ITS.pdf

Nilsson, Jan-Eric. "Towards a welfare enhancing process to manage railway infrastructure access." *Transportation Research Part A: Policy and Practice*36.5 (2002): 419-436.

Nunez, Jennifer. "FRA Committee OKs HSR crashworthiness standards." Railway Age Magazine 17 June 2013: n. pag. Railway Age. Web. 18 May 2015. <http://www.railwayage.com/index.php/safety/fra-committee-oks-hsr-crashworthiness-standards.html?channel=60>.

Ogata, Masaki. *Crucial Concepts of High Speed Rail* Tokyo: East Japan Railway Company, 2015. *CHSRA*. Web. 13 May 2015. <http://www.hsr.ca.gov/docs/brdmeetings/2014/brdmtg_091614_Item3_XpressWest_Project_RevF.pdf>.

On the Move: Southern California Delivers the Goods. Los Angeles: Southern California Association of Governments, 2012. *Freight Works*. Web. 13 May 2015. http://www.freightworks.org/ Document Library/CRGMPIS_Summary_Report_Final.pdf

On-rail Competition: Consultation on Options for Change in Open Access. N.p.: n.p.,

P. Teixeira, A. Prodan. *Railway Infrastructure Pricing in Europe for High-Speed and Intercity Services: State of the 4 Practice and Recent Evolution.* Proceedings of the 93rd Annual Meeting of the Transportation Research Board. No. 14-4203, Washington, DC, January 2014.

Palley, Joel. *Impact of the Staggers Rail Act of 1980*. Washington, D.C.: FRA, 2011. *Ferderal Railroad Administration*. Web. 13 May 2015. https://www.fra.dot.gov/eLib/Details/L03012

Peña-Alcaraz, Maite, Joseph Sussman, and Sam Levy. "Implementation Challenges for Shared Railway Systems: Case Studies In California And The Northeast Corridor."

Pena-Alcaraz, Maite. *Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems*. Diss. Massachusetts Institute of Technology, 2015.

Perennes, Patricia. "Use of Combinatorial Auctions in the Railroad Industry: Can The “Invisible Hand” Draw The Railroad Timetable?." *5th Annual Competition and Regulation in Network Industries Conference*. 2013.

Perkins, Ed. "The New California Air-Travel Market." *Smarter Travel*. N.p., 5 Aug. 2013. Web. 7 May 2015. <http://www.smartertravel.com/blogs/today-in-travel/the-new-california-air-travel-market.html?id=15720224>.

PPIAF. Railway Reform: Toolkit for Improving Rail Sector Performance. Chapter 3, Railway economics and pricing. World Bank Report, Washington, DC, 2011.

Project Update Report to the California State Legislature. N.p.: n.p., 2015. *California High-Speed Rail Authoirty*. Web. 8 May 2015. <http://www.hsr.ca.gov/docs/about/legislative_affairs/ SB1029_Project_Update_Report_030115.pdf>.

ProRail (2013). Network Statement 2014. Netherlands. Web accessed on March 8, 2014.

RAIL COMPETITION. N.p.: n.p., 2012. Print.

Railway Sector. N.p.: n.p., 2012. Print.

"Recovery Act High Speed Rail Awards." *Whitehouse.gov*. The White House, Feb. 2010. Web. 13 May 2015. https://www.whitehouse.gov/sites/default/files/rss_viewer/hsr_awards_summary_public.pdf

REFER (2012). 2014 Network Statement. Portugal. Web accessed on March 8, 2014.

"Regional Airport Planning Committee Meeting Notice." Regional Airport Planning Committee Meeting. 26 Feb. 2010. *Association of Bay Area Governments*. N.p., n.d. Web. 15 May 2015. <http://www.abag.ca.gov/abag/events/agendas/n022610a_source.pdf>.

"Regional Connector Transit Project." *Metro*. LACMTA, Jan. 2015. Web. 13 May 2015. <http://www.metro.net/projects/connector/>.

"Regional Data: GDP and Personal Income." *Bureau of Economic Analysis*. N.p.: n.p., n.d. *U.S. Department of Commerce Bureau of Economic Analysis*. Web. 15 May 2015. <http://www.bea.gov/ iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=4#reqid=70&step=30&isuri=1&7022=21&7023=0&7024=non-industry&7033=-1&7025=0&7026=06000&7027=2014&7001=421&7028=-1&7031=0&7040=-1&7083=levels&7029=21&7090=70>.

"Regional Rail Update." *Metro Board Meeting Minutes*. LACMTA, 17 Oct. 2012. Web. 14 May 2015. <http://media.metro.net/board/Items/2012/10_october/20121017P&PItem26.pdf>.

"Ride 6 Amtrak Pacific Surfliner Trains." *North County Transit District*. NCTD, Jan. 2015. Web. 13 May 2015. <http://www.gonctd.com/amtrak>.

Revised 2012 Business Plan. N.p.: n.p., 2012. *California High-Speed Rail Authority*. Web. 7 May 2015. <http://www.hsr.ca.gov/docs/about/business_plans/BPlan_2012ExecSum.pdf>.

RFI (2013). 2014 Network Statement. Italy. Web accessed on November 5, 2014.

Rosenberg, Mike. "California high-speed-rail groundbreaking pushed back another few months." *San Jose Mercury News* 17 Sept. 2013: n. pag. *San Jose Mercury News*. Web. 13 May 2015. <http://www.mercurynews.com/california-high-speed-rail/ci_24117524/ california-high-speed-rail-groundbreaking-pushed-back-another>.

S.A. Morrison. The Value of Amtrak. *Journal of Law and Economics*, Vol. 33, No. 2, pp. 361-382, 1990.

Sakamoto, Ryusuke. *High speed railway productivity: how does organizational restructuring contribute to HSR productivity growth?*. Diss. Massachusetts Institute of Technology, 2012.

Service Delivery Policy. Boston: MBTA, 2010. *Massachusetts Bay Transportation Authority*. Web. 16 May 2015. <https://www.mbta.com/uploadedfiles/About_the_T/T_Projects/T_Projects_List/ 2010ServiceDeliveryPolicy.pdf>.

"SF Transportation Plan Update." San Francisco County Transportation Authority. 30 Jan. 2013. *San Francisco County Transportation Authority*. N.p., n.d. Web. 7 May 2015. <http://www.sfcta.org/sites/default/files/content/Planning/SFTP2/SFTPCACSlides-send.pdf>.

Sheyner, Gennady. "Jim Hartnett makes a return trip to Caltrain — now as CEO." *Palo Alto Weekly* 13 Mar. 2015: n. pag. *Palo Alto Online*. Web. 16 May 2015. <http://www.paloaltoonline.com/news/ 2015/03/13/jim-hartnett-makes-a-return-trip-to-caltrain-151-now-as-ceo>.

Socorro, M. Pilar, and M. Fernanda Viecens. "The effects of airline and high speed train integration." *Transportation Research Part A: Policy and Practice*49 (2013): 160-177.

Southern Pacific Railroad. and Peninsula Corridor Joint Powers Board "Trackage Rights Agreement---Peninsula Main Line and Santa Clara/Lick Line." Contract. 20 Dec. 1991. TS.

Staff Report for June 12 2014. Rept. no. 10. San Francisco: Transbay Joint Powers Authority, 2014. *Transbay Center*. Web. 15 May 2015. <http://transbaycenter.org/uploads/2014/06/Item10_PTG-Term-3-Amendment-1.pdf>.

State and County Quick Facts. N.p.: n.p., n.d. *United States Census Bureau*. Web. 7 May 2015. <http://quickfacts.census.gov/qfd/states/00000.html>.

Strategic Plan FY2015-2024. N.p.: n.p., 2014. *Caltrain*. Web. 15 May 2015. <http://www.caltrain.com/Assets/Caltrain+Modernization+Program/Documents/Strategic+Plan+Final+Doc.pdf>.

Swiss Federal Railways (2013). List of Infrastructure Services 2014. Switzerland. Web accessed on March 8, 2014.

Talebian, Ahmadreza, and Bo Zou. "Capacity Allocation in Vertically Integrated Railway Systems: A Rubinstein Sequential Bargaining Game Approach."*Transportation Research Board 94th Annual Meeting*. No. 15-5160. 2015.

Temple, Simon. "Open Access Long Distance Passenger Rail Services in the United Kingdom: The Grand Central Experience." *European Transport Conference 2014*. 2014.

"The many obstacles freight rail faces in NYC and Long Island." *Rail New York*. East-of-Hudson Freight Rail Alliance, Jan. 2013. Web. 13 May 2015. <http://www.railnewyork.com/issues.html>.

"The Program." *Transbay Center*. N.p., n.d. Web. 15 May 2015. <http://transbaycenter.org/project/program-overview>.

"The Pulse of Metrolink." Metrolink Board of Directors Meetings. One Gateway Plaza. 27 Feb. 2015. Speech and PowerPoint Presentation.

The Transit Center District Plan. N.p.: n.p., 2011. *Transbay Center*. Web. 7 May 2015. <http://transbaycenter.org/uploads/2011/10/Transit-Center-District-Plan-Plng-Dept-flyer-final.pdf>.

Thompson, Lou, ed. *Comments on the 2014 CHSRA Business Plan*. Sacramento: California High-Speed Rail Peer Review Group, 2014. Print.

Thompson, Louis S. "The Benefits of Separating Rail Infrastructure from Operations." *Public Policy for the Private Sector* Dec. 1997: n. pag. *World Bank*. Web. 13 May 2015. http://siteresources.worldbank.org/EXTFINANCIALSECTOR/Resources/282884-1303327122200/135thomp.pdf

Tillier, Clem. "Freight on the Peninsula." *Caltrain HSR Compatibility Blog*. N.p., 8 Mar. 2009. Web. 4 May 2015.<http://caltrain-hsr.blogspot.com/2009/03/freight-on-peninsula.html>.

Transbay Center. N.p., n.d. Web. 6 May 2015. <http://transbaycenter.org/uploads/2009/10/ Chapter_3.pdf>. (REVISIT)

Transportation Analysis--Caltrain Peninsula Corridor Electrification Project. San Jose: Fehr and Peers, 2014. *Caltrain*. Web. 16 May 2015. <http://www.caltrain.com/Assets/ Caltrain+Modernization+Program/DEIR/Appendix+D+TIA-Part+1.pdf>.

"UP-Southern Pacific Merger Creates Largest U.S. Railroad." *Union Pacific*. N.p., n.d. Web. 2 May 2015. <http://www.up.com/timeline/index.cfm/up-sp-merger>.

United States. Census Bureau. Population Division. "California Population Statistics." Web. 15 May. 2015. http://www.census.gov/population.hmtl

Van Ark, Roelof. *Review of "Bay/Area California High-Speed Rail Ridership and Revenue Forecasting Study."* Sacramento: CHSRA, 2010. *California High Speed Rail Authority*. Web. 13 May 2015. <http://www.hsr.ca.gov/docs/about/ridership/ridership_forcasting_resp_UCB_ITS.pdf>.

Vartabedian, Ralph. "Bullet train's eventual link to LA rail system far from clear-cut." *Los Angeles Times* 13 Dec. 2014: n. pag. *Los Angeles Times*. Web. 14 May 2015. http://www.latimes.com/local/politics/la-me-adv-bullet-metrolink-20141214-story.html#page=1

Weikel, Dan. "2,000 rally against plan to put high-speed rail in Angeles forest." *Los Angeles Times* 14 Jan. 2015: n. pag. *Los Angeles Times*. Web. 13 May 2015. <http://2,000 rally against plan to put high-speed rail in Angeles forest>.

"West Coast Main Line, United Kingdom." *Railway Technology*. Kable Intelligence Limited, Jan. 2010. Web. 13 May 2015. <http://www.railway-technology.com/projects/virgin/>.

Westrom, Ryan Jeremy. *The agglomerative role of transportation investment: a comparative analysis of Portuguese and American high-speed rail proposals*. Diss. Massachusetts Institute of Technology, 2014.

"Who Rides Caltrain? Passenger Count & Survey Important Source of Info." *Caltrain*. N.p., n.d. Web. 6 May 2015. <http://www.caltrain.com/about/news/Who_Rides_Caltrain__Passenger_Count___Survey_Important_Source_of_Info.html>

Wiekel, Dan. "Union Station to get $350 million in track upgrades." *Los Angeles Times* 27 May 2014: n. pag. *Los Angeles Times*. Web. 13 May 2015. <http://www.latimes.com/local/la-me-union-station-tracks-20140528-story.html>.

Wilner, Frank N. "Amtrak's Sisyphean struggle to run on time." *Railway Age* 5 July 2013: *Railway Age*. Web. 13 May 2015. <http://www.railwayage.com/index.php/blogs/frank-n-wilner/ amtraks-sisyphean-struggle-to-run-on-time.html>.

Wilson, Carolina. "Tired of Being Bumped, Caltrain Bike Commuters Push for More Space." *SFGate* 10 Dec. 2014: n. pag. *SFGate*. Web. 6 May 2015. <http://blog.sfgate.com/inthepeninsula/2014/12/ 10/caltrain-bike-commuters-space/>.

Wilson, Nigel. "Current Short-Range Transit Planning Practice." *Google Drive*. MIT, Web. 14 May 2015. <https://drive.google.com/file/d/0B7jf5xRaJfUBSGhtQmIzQnQ5Q1k/view?usp=sharing>.

Xpress West Project Update for CHSRA Board. Las Vegas: Xpress West, 2014. *California High-Speed Rail Authority*. Web. 13 May 2015. <http://www.hsr.ca.gov/docs/brdmeetings/2014/brdmtg_091614_Item3_XpressWest_Project_RevF.pdf>.

Zarembski, Allan M., James Blaze, and Pradeep Patel. "Shared Corridors, Shared Interests." *2011 Joint Rail Conference*. American Society of Mechanical Engineers, 2011.

Additional Publications

Capacity Challenges on the California High-Speed Rail Shared Corridors: How Local Decisions Have Statewide Impacts (attached)

Samuel Levy and Joseph M. Sussman

Capacity Challenges on the California High-Speed Rail Shared Corridors: How Local Decisions Have Statewide Impacts

Sam Levy and Joseph M. Sussman
Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA

In 2012, as a cost-control measure and in response to local opposition in the San Francisco Bay Area, the California High-Speed Rail Authority (CHSRA) adopted a "blended system" at the north and south bookends of the planned first phase of its high-speed rail line. In this blended operation, the high-speed rail line will share track and other infrastructure with commuter rail, intercity rail, and freight on the 50-mile Peninsula Corridor in Northern California and on 50 miles of right-of-way between Burbank, Los Angeles, and Anaheim in Southern California.

In Northern California, the Peninsula Corridor Joint Powers Board's (PCJPB) *Caltrain* commuter rail service between San Francisco and San Jose is experiencing record levels of ridership. Ridership will be further stimulated by the electrification of the line and its extension into San Francisco's central business district. With the California High-Speed Rail Authority competing spatially and temporally with Caltrain for access to high-revenue and high-cost infrastructure, we review different strategies for coordination and integration between the two agencies.

In Southern California, the final form of the blended system is more nebulous than its northern counterpart. For the first few years of high-speed rail service, the *Metrolink* service operated by the Southern California Regional Rail Authority (SCRRA) is expected to complement the high-speed rail system. However, since Metrolink operates on congested rail infrastructure, some of it owned by capacity-conscious freight railroads, there will exist the challenge of providing quality service and transfer opportunities for time-sensitive high-speed rail customers.

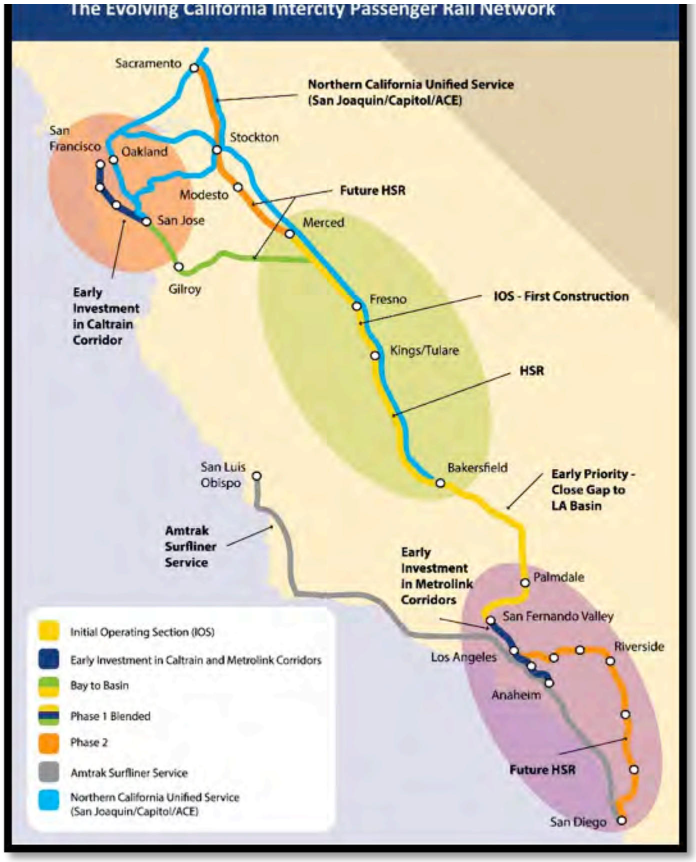
The change to a blended system was a dramatic change of direction for the CHSRA; as a result, a new paradigm is needed for implementation of the system over the next 15 years. The decisions made on the local blended corridor level will affect both the financial viability of the overall project and the quality of service experienced by customers across the entire California rail system.

8 Introduction

California's Proposition 1A, passed by voters in 2008, authorized $9.95 billion from a state bonds issue that provide partial funding for construction of an HSR line capable of transporting passengers between Los Angeles and San Francisco in no more than 160 minutes with trains capable of running at least 200 miles per hour. Since 2008, the CHSRA has won funding (totaling approximately $15 billion) from President Obama's American Recovery and Reinvestment Act and the FRA's High Speed Intercity Passenger Rail (HSIPR) program, as well as from California's cap-and-trade carbon emissions reduction program. With the remaining funding expected to come from private investment, the $68 billion Phase 1 of the system between San Francisco, Merced, Los Angeles, and Anaheim (with future phases to Sacramento and San Diego) is planned for completion in 2028 (See Figure 1).

The CHSRA's shift to a "blended system" from dedicated line reflects the reality that the costs and challenges associated with constructing new, dedicated rail infrastructure are enormous, especially in urban areas. Shared rail corridors represent the possibility of more efficient use, that is, higher utilization, of precious rail infrastructure. Multiple railroads can share the burden of track maintenance and traffic control, both of which require high fixed costs. Sharing track, when done properly, is an attractive option for both passenger rail agencies and freight railroads and increasingly common in the United States. However, sharing track comes with challenges for all participating railroad operators as well.

Figure 1. Map of current California HSR implementation plan. Blended service areas are shown in dark blue. Source: CHSRA



Sharing track requires coordination and more often than not, it is among non-homogenous rail traffic. In California, rail traffic is non-homogenous in both operating characteristics such as stopping patterns and running speed and in physical characteristics such as necessary lateral and vertical clearances and station platform and vehicle door heights.

The overarching goal of this research is to offer recommendations on how California's rail capacity might be managed better going forward and to develop a methodology to approach design decisions that affect capacity. As travel patterns and freight demands move across rail corridors with different owners, capacity planning becomes increasingly important.

9 Northern California Blended Service

The Peninsula Corridor Joint Powers Board (PCJPB) made up of the three counties served on the 87-mile network, owns the northernmost 51-miles of track between San Francisco 4th and King Station and San Jose Station and operates commuter service between the cities, branded as Caltrain. Union Pacific uses the corridor for freight service in and out of San Francisco and maintains perpetual trackage rights. The PCJPB and its relationship with Union Pacific, the CHSRA, and the private HSR operator on the corridor is already and will continue to be instrumental in the ultimate form of high-speed rail in California.

Caltrain is enjoying huge ridership growth due, in part, to the thriving economic conditions in Silicon Valley and San Francisco and the millennial generations' desire to leave the car at home or not own a car at all and be productive on the daily commute. The "Baby Bullet" express service has reduced travel times, but has also drawn many commuters to bring their bikes on-board—Caltrain sees more bicyclists on-board than any other transit agency in the nation.

Caltrain has two major upcoming capital projects that will spark additional ridership growth. First, an electrification of the corridor will allow for more local services--making the train a more convenient option. Second, the extension of Caltrain to the Transbay Transit Center will bring Caltrain riders much closer, and via a one-seat ride, to San Francisco's job center.

The Transbay Transit Center, which will be owned by the third-party Transbay Joint Powers Authority (TJPA), will be located near more jobs and residents than any other existing Caltrain station. For a private, for profit high-speed operator it will serve as a revenue generator in part because San Francisco is best equipped to accommodate the car-less HSR traveler. The two operators will compete both spatially for the six-track terminal and temporally during the rush hour peaks (See Figure 2). Currently, the plan is to have two dedicated Caltrain tracks and four dedicated HSR tracks, but the Transbay Joint Powers Authority is urging the two agencies to consider adopting standard vehicle height to allow for platform sharing. Without platform sharing and high levels of commuter-HSR integration, the local constraints will affect service quality not just on the Peninsula, but across the state.

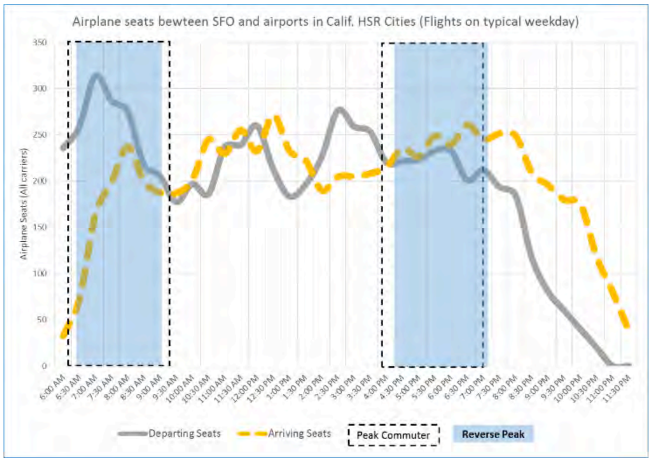


Figure 2. Caltrain peak periods (represented by blue and dashed boxes overlap temporally with air seats out of San Francisco (a proxy for HSR travel)
Source: Caltrain ridership, SF International Airport schedule data

Using this knowledge that the northernmost 1.3 miles of the high-speed rail line will be a vertically separated structure, we apply the vertically-separated train operator model developed by Levy, Sussman, Pena-Alcaraz, and Prodan in TRB 15-1697. This revealed that it will be difficult for the high-speed rail operator and Caltrain to compete on a level playing field for access to the Transbay Transit Center due to the fiscal strength of the high-speed rail operator, not to mention the state-wide political constituency behind HSR service.

10 Southern California Blended Service

As in Northern California, high-speed rail operations will be blended with commuter rail operations. However, in Southern California, there are more operators, sections of freight-owned right-of-way, and no plans to electrify non-HSR operations. Blended service planning is nascent in Southern California, even more so than the Northern California system.

Los Angeles Union Station to Southern California is analogous to the Transbay Transit Center in the Bay Area. It is better connected to Los Angeles' population centers and will continue to serve as a regional transportation hub regardless of the ultimate success of high-speed rail. With 14 tracks currently used by three rail operators (Metrolink, Amtrak, and Amtrak California's *Pacific Surfliner*), it will be an important hub for the high-speed rail operator and its future warrants careful consideration.

One of the key rail capital projects in Southern California are the Union Station run-through tracks, known as the Southern California Regional Interconnector Project (SCRIP) Currently, all trains arriving and departing from and to San Diego to the south, San Luis Obispo and Santa Barbara to the north, and San Bernardino County to the east all use the same approach tracks to the north of the "stub-end" terminal. The four run-through tracks (some of which will be used by high-speed rail trains) will add a high degree of flexibility for express services and inter- and intra-agency transfers and coordination. Construction is scheduled to be completed in late 2019 or early 2020 (ibid).

At the time of writing, this blended system runs approximately 50 miles from Burbank in the San Fernando Valley through Los Angeles Union Station and on through to Anaheim's newly-built Anaheim Regional Transportation Intermodal Center (ARTIC). Currently, electrification in Southern California is solely for high-speed rail. Metrolink and Amtrak California have no plans to electrify their systems. From 2022 to 2028, riders would transfer from high-speed rail at Burbank to slower, diesel powered trains for the ride southward to Los Angeles and Anaheim. When the CHSRA completes Phase 1 in 2028, HSR trains would share tracks, but not stations, with Metrolink and the Pacific Surfliner south through Los Angeles and onwards towards ARTIC. Metrolink and the Surfliner, though, would continue conventional rail operation, though possibly at higher speeds than today.

Unlike Caltrain in Northern California, Metrolink and Amtrak California also have no plans to change the specifications in terms of door height or floor height on their operating train sets. This means that high-speed rail will have to use separate platforms even if tracks are shared in between stations. This could limit the capacity and benefit of Los Angeles Union Station because the run-through tracks need to be assigned specifically to high-speed rail or conventional rail.

11 Conclusion

The blended system of shared corridors will put high demands on the existing infrastructure. Commuter rail or freight delays can propagate to the HSR system. As the ratio of service volume to capacity tends towards 1, the system loses stability and on-time performance suffers. With degraded on-time performance and uncertainty regarding arrival times, schedule padding becomes necessary and makes rail as a mode less attractive to time-sensitive consumers.

California's Rail "Wish-List"

- ***Level Boarding and Interoperability***
- ***Ability to modify service levels***
- ***Xpress West integration with California HSR***
- ***Integrated Southern California network***
- ***High-frequency, uniform-headway HSR and commuter rail***
- ***Satisfy 2008 bond measure requirements***
- ***Minimize costs and build the network***

quickly

Figure 3: Inferred "wish-list" of an integrated, blended CHSRA network

The blended system dictates that the CHSRA will operate on a multi-owner network, in many ways like Amtrak does today. Instead of having sole control of its infrastructure, the CHSRA will have to work with the TJPA, PCJPB, SCRRA, and BNSF Railway to ensure smooth operation. As a result, the CHSRA will face many of the issues Amtrak faces today regarding train priority. The ability for the CHSRA to operate in a reliable fashion will depend on the priority rules that the CHSRA can negotiate with its host railroads on the blended corridors. Optimizing locally constrains the overall system optimal solution; and as agencies make decisions regarding vehicle fleets and track investments, we can see that the local optimizing has already begun. Decisions made on the Peninsula Corridor can create capacity bottlenecks affecting HSR trains inter-regionally. And since these blended issues are being addressed presently on the Peninsula Corridor with the concurrent electrification of Caltrain, a precedent is being set for their southern neighbors. The Peninsula Corridor, therefore, is an important proving ground for whether or not blended operations can work.

Integrated operations have a potential to bring HSR a revenue source; like regional airlines feed into international hubs, so too can commuter rail services feed into interregional high-speed rail services. To that end, service planning should drive infrastructure decisions; in an era where infrastructure such as new HSR stations or electrification are expensive and the public is leery of megaprojects, California can set an example with a well-conceived (i.e. well service- planned), integrated rail system.

12 References

California High-Speed Rail 2014 Business Plan. California High-Speed Rail

"Caltrain Modernization Program" Peninsula Corridor Joint Powers Board. May 2015

Caltrans. *2013 California State Rail Plan*. 2013 Leg. Sacramento: California State Transportation Agency, 2013. *California Department of Transportation*.

The authors would thank the National University Rail (NURail) Center, a US DOT-OST Tier 1 University Transportation Center for sponsoring this study.

National University Rail Center - NURail
US DOT OST-R Tier 1 University Transportation Center

NURail Project ID:

NURail2012-MIT-R01 High-Speed Rail as a Complex Sociotechnical System

Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems

By

Maite (Maria Teresa) Pena-Alcaraz

Supervised by Professor Joseph M. Sussman

10/1/2015

Grant Number: DTRT12-G-UTC18

DISCLAIMER
The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

National University Rail Center - NURail
US DOT OST-R Tier 1 University Transportation Center

TECHNICAL SUMMARY

Title

Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems

Author: Maite (Maria Teresa) Pena-Alcaraz

Introduction

In the last 15 years, the use of rail infrastructure by different train operating companies (shared railway system) has been proposed as a way to improve infrastructure utilization and to increase efficiency in the railway industry. Shared use requires coordination between the infrastructure manager and multiple train operators. Such coordination requires capacity planning mechanisms that determine which trains can access the infrastructure at each time, *capacity allocation*, and the access charges they have to pay, *capacity pricing*.

The objective of this thesis is to contribute to the field of shared railway systems coordination by 1) developing a framework to analyze the performance of shared railway systems under alternative capacity pricing and allocation mechanisms, and 2) using this framework to understand the implications of representative capacity pricing and allocation mechanisms in representative shared railway systems.

Approach and Methodology

This research uses quantitative techniques from the fields of operations research and economics to solve the capacity allocation and pricing problem from the viewpoints of 1) the train operators and 2) the infrastructure manager and then seeks to find the appropriate balance between them from a public policy viewpoint

Two cases studies are performed. The first deals with the Northeast Corridor of the U.S. where competition for track capacity among intercity, commuter and freight traffic is intense. The second deals with Tanzania where the competition is less intense but the question of how to operate the rail system in the public interest is front and center.

Findings

The fundamental finding of this research is the demonstration that the **effective** implementation of shared railway systems requires the design and implementation of capacity allocation and pricing mechanisms. These mechanisms are the rules needed for coordinating the multiple agents that share the infrastructure. In this research, we analyze the performance of shared railway systems under alternative mechanisms to price and allocate railway capacity and are able tom teach several important conclusions summarized below.

Conclusions

The first conclusion of this research is that the implications of capacity pricing and allocation mechanisms for shared railway systems are still unclear. While this thesis tries to offer clarity in this area, there is still much work to be done. In that sense, we join (Drew and Nash, 2011; Nash, 2010) in recommending to academics that they invest in this research topic. Any progress in research that contributes to a better understanding of the implications of alternative mechanisms to price and allocate capacity could immensely help practitioners and policy makers. This is particularly important in a context in which several countries are currently restructuring their railway sector to allow shared use.

The second conclusion of this research is that sharing railway infrastructure capacity is not straightforward. In the railway industry, as compared to other network industries, there are very strong interactions between capacity planning and infrastructure operations. Chapter 5 shows that capturing this interactions is critical to implement capacity-based mechanisms and to understand the implications of price-based mechanisms in the railway industry. Despite these differences, regulators and policy makers rely on the lessons learned from other network industries. Although these lessons are useful and can serve as guiding principles to design mechanisms to price and allocate railway capacity, Chapter 2 shows that they often do not work in the railway industry. We thus recommend that policy makers are cautious and question the validity of assumptions based on other industries. We also recommend to academics that they reach other communities beyond their domains doing research in these topics. A better understanding of what works and what does not work across network industries and why would also be very valuable for practitioners and policy makers.

The third conclusion, on a more positive note, is that the implementation of adequate capacity pricing and allocation mechanisms can mitigate the coordination problems of shared railway systems while maintaining the benefits of shared infrastructure use in the railway industry. Chapter 2 shows that the introduction of TO competition enabled by shared use may have similar effects to the introduction of regulation to ensure that TOs behave as even-handed integrated railway companies. Chapter 4 shows that shared use may allow the IM to recover more infrastructure costs than those recovered in dedicated corridors by enabling the entrance of new TOs that offer profitable services that the current TO does not provide. In a context in which the

NEC and many other systems are moving ahead with the implementation of new capacity pricing and allocation mechanisms, we conclude with three more recommendations. This research shows important trade-offs among alternative mechanisms to price and allocate railway capacity. We recommend the use of the framework developed in this thesis to identify *personalized* mechanisms to price and allocate capacity, aimed at the specific characteristics of the systems. At the same time, we recommend that practitioners and policy makers consider alternative mechanisms to price and allocate railway capacity before locking their system into one of them. Even if those cases where stakeholders have to make a decision soon, we recommend that they allow for some flexibility to adapt the mechanism implemented as they gather more information and better understand the implications of alternative mechanisms for their systems. Finally, we recommend railway companies and regulators that they measure the performance of their systems using a wide variety of performance metrics and to share information and best practices with other railway systems. The data form these experiments will contribute to the improved understanding and management of shared railway systems.

Recommendations

Our recommendations are embedded in the conclusions above.

In addition we recommend some continuations of the research reported on herein.We identify four lines of future research that we find particularly relevant to better understand the implications of capacity pricing and allocation mechanisms in shared railway systems. These lines include 1) additional validation of the framework developed, 2) the extension of the models and algorithms proposed, 3) the utilization of the framework developed to answer other related and relevant shared railway systems research questions, and 4) the development of a broader understanding of capacity pricing and allocation across network industries.

Publications

Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems: Lessons for the Northeast Corridor (attached)

Maite Pena-Alcaraz and Joseph M. Sussman

Papers

1) Pena-Alcaraz, Perez-Arriaga, Sussman, (2014). Capacity pricing schemes to implement open-access rail in Tanzania. ***Network Industry Quarterly*, Vol. 16, No. 2, pp. 18-22.**

2) Pena-Alcaraz, (2015). Capacity pricing and allocation in shared railway systems. ***Intercity Rail Passenger Systems Update,* No. 23, pp.13-15.**

3) Levy, Pena-Alcaraz, Prodan , Sussman, (2015). Analyzing the financial relationship between railway industry players in shared railway systems: the train operator's perspective. ***Transportation Research Record*, in press.**

4) Pena-Alcaraz, Webster, Ramos, (2015). Train timetable design for shared railway systems using a linear programming approach to approximate dynamic programming. ***Transportation Science, review & resubmit*.**

5) Pena-Alcaraz, Sussman, Webster, Perez-Arriaga, (2015). Comparison of price-based and capacity-based mechanisms for pricing and allocating railway capacity in the Northeast Corridor. ***Transportation Research Part A: Policy and Practice, review & resubmit.***

Primary Contact

Professor Joseph M. Sussman
sussman@mit.edu
(617) 253-4430

Principal Investigators

Name: Maite Pena-Alcaraz, Joseph M. Sussman

Title: **Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems**

Department: Civil and Environmental Engineering

University: Massachusetts Institute of Technology

Phone number: (617) 803-2182

Maite Pena-Alcaraz
maitepa@alum.mit.edu
(617) 253-4430

Professor Joseph M. Sussman
sussman@mit.edu
(617) 253-4430

Other Faculty and Students Involved

Name: Joseph M. Sussman

Title: JR East Professor of Civil and Environmental Engineering and Engineering Systems

Department: Civil and Environmental Engineering

University: Massachusetts Institute of Technology

sussman@mit.edu

617.253.4430

NURail Center
217-244-4999
nurail@illinois.edu
http://www.nurailcenter.org/

Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems

by

Maite (Maria Teresa) Pena-Alcaraz

M.S., Engineering Systems Modeling, Comillas Pontifical University, Madrid (Spain), 2010

B.S., Industrial Engineering, Comillas Pontifical University, Madrid (Spain), 2009

B.S., Mathematics, Complutense University of Madrid (Spain), 2012

Submitted to the Engineering Systems Division

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Signature of Author

…………………………………………………………………………..............................

Engineering Systems Division

May 4, 2015

Certified by ……………………………………………………………......Joseph M. Sussman

JR East Professor of Civil and Environmental Engineering and Engineering Systems

Thesis Committee Chair

Certified by ……………………………………………………………......Mort D. Webster

Associate Professor of Energy Engineering, Pennsylvania State University

Thesis Committee Member

Certified by ……………………………………………………………......Andres Ramos

Professor, Comillas Pontifical University

Thesis Committee Member

Certified by ……………………………………………………………......Jose Antonio Gomez-Ibanez

Derek C. Bok Professor of Urban Planning and Public Policy, GSD and HKS, Harvard University

Thesis Committee Member

Certified by ……………………………………………………………......Ignacio Pérez-Arriaga

Visiting Professor at the Center for Energy and Environmental Policy Research, MIT

Professor and Director of the BP Chair on Energy and Sustainability, Comillas Pontifical University

Thesis Committee Member

Accepted by ………………………………………………………………......Munther A. Dahleh

William A. Coolidge Professor of Electrical Engineering and Computer Science

Director, Institute for Data, Systems and Society

To my parents, for teaching us the value of effort and education

To my siblings, Ali and Manuel, for reminding me the importance of challenging assumptions

To my husband, Noel, for supporting me on this journey

Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems

by

Maite Pena-Alcaraz

Submitted to the Engineering Systems Division

on May 4, 2015 in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Engineering Systems

Abstract

In the last 15 years, the use of rail infrastructure by different train operating companies (shared railway system) has been proposed as a way to improve infrastructure utilization and to increase efficiency in the railway industry. Shared use requires coordination between the infrastructure manager and multiple train operators. Such coordination requires capacity planning mechanisms that determine which trains can access the infrastructure at each time, *capacity allocation*, and the access charges they have to pay, *capacity pricing*.

The objective of this thesis is to contribute to the field of shared railway systems coordination by 1) developing a framework to analyze the performance of shared railway systems under alternative capacity pricing and allocation mechanisms, and 2) using this framework to understand the implications of representative capacity pricing and allocation mechanisms in representative shared railway systems.

There are strong interactions between capacity planning and infrastructure operations in the railway industry; the operations on the infrastructure determine the available capacity in the system. As a consequence, the framework developed in this thesis to evaluate the performance of shared railway systems under alternative capacity pricing and allocation consists of two models: 1) a train operator model and 2) an infrastructure manager model. The train operator model is a financial model that anticipates how train operators would respond to the capacity pricing and allocation mechanisms and determine their demand for infrastructure use. The infrastructure manager model is a network optimization model that determines the optimal train timetable (infrastructure manager's decisions) that accommodates the train operators' demands for scheduling trains, considering the topology of the system, safety constraints, and other technical aspects of the infrastructure for shared railway systems. To be able to solve the train timetabling optimization problem in meaningful instances, this thesis develops a novel approximate dynamic programming algorithm based on linear programming that extends previous algorithms proposed in the literature to effectively solve large network optimization problems.

This thesis then uses the train operator model to compare the operational decisions of train operators in shared railway systems with the operational decisions of even-handed integrated railway companies. We show that train operators in shared railway systems make the same operational decisions as an integrated railway company when variable access charges reflect variable infrastructure manager's costs to operate trains on the infrastructure. We also identify two cases in which the train operators may have incentives to deviate from the integrated railway systems' operational decisions: 1) when the infrastructure manager needs to recover part of the infrastructure management fixed costs, or 2) when the railway system is congested. This motivates the choice of the two case studies of this thesis, one based on the Central Corridor in Tanzania, and the other one based on the Northeast Corridor in the US.

We then show how to use the framework proposed in this thesis to analyze the trade-offs associated with the use of alternative mechanisms in these two cases. To our knowledge, this is the first effort to compare alternative mechanisms to price and allocate capacity in the same shared railway system. The results of this thesis show that there are important trade-offs associated with each mechanism and none of them is superior to the other on all dimensions. We thus recommend that system stakeholders carefully analyze the implications of alternative capacity pricing and allocation mechanisms before locking the system into one of them. This is particularly important today since several countries are currently restructuring their railway sector to allow shared use. We claim that the improved understanding of the system performance gained with the framework proposed in this thesis is important to be able to design adequate capacity pricing and allocation mechanisms that can mitigate the coordination problems of shared railway systems while maintaining the benefits of shared infrastructure in the railway industry.

Thesis Chair: Joseph M. Sussman

Title: JR East Professor of Civil and Environmental Engineering and Engineering Systems, Massachusetts Institute of Technology

Thesis Co-supervisor: Mort D. Webster

Title: Associate Professor of Energy Engineering, John and Willie Leone Family Department of Energy and Mineral Engineering, College of Earth and Mineral Sciences, Penn-State University

Acknowledgements

This PhD has been a life-changing journey. I want to briefly thank many people and institutions that have supported me during this journey and contributed to make it unforgettable, as well as those who helped me to "catch this train".

I want to start thanking my thesis committee. I have been lucky to have five extraordinary supervisors and mentors who have guided me during these years: professors Joseph Sussman, Mort Webster, Andres Ramos, Jose Antonio Gomez-Ibanez, and Ignacio Perez-Arriaga. Joe and Mort, thank you for your guidance, patience, example, and support. Andres, Tony, and Ignacio, thank you for sharing your knowledge and experience, and for always being happy to have a discussion about this research. I am indebted for all I have learned from the five of you, not only from a professional standpoint, but also from a personal standpoint.

I also want to acknowledge the support from two excellent scholarships. The research of this thesis was generously supported by La Caixa Foundation and Rafael del Pino Foundation. I want to thank Emilia, Pilar, Amadeo, and Vicente, who made sure that everything was going well during these years. I also want to thank other sources of funding that have supported this PhD. I want to acknowledge the support from the National University Rail Center (NURail), the Japan East Railway Company (JR East), the Japan International Transportation Institute (JITI), and the National Science Foundation (NSF). I was also fortunate to work with professors Richard Larson and Mort Webster as a teaching assistant in two quantitative courses they taught. Dick and Mort, thank you, not only for the funding support, but also for the opportunity to teach and learn from you.

I am very grateful to have been able to carry out my research at MIT. My work has tremendously benefited from many of the classes and excellent professors: John Tsitsiklis, David Autor, Nancy Rose, Richard de Neufville, Chris Magee, Fred Salvucci, Mikel Murga, and Eva Ponce, among others. This research has also benefited from many discussions with students in Professors Sussman and Webster's research groups. I want to thank my JITI project collaborators: Andres, Joel and Naomi, my JR East team, Joanna and Patton, as well as Sam, Alex, Bruno, Iori, Ryusuke, Tatsu, and Tolu for the very fruitful discussions about the NEC and about the role of rail infrastructure in the society. The nurturing research environment provided by my department, the Engineering Systems Division, and by my office micro-cosmos has also contributed immensely to this research. Thank you Abby, Alex, Berk, Fernando, Hector, Tommy, Vivek, and Claudia for all the research (and non-research) conversations. I have learned a lot form all of you. Thank you Beth and all the ESS leaders for making sure that we always had everything we needed to organize student seminars or any other useful (professional or social) activity.

This research has also benefited from internships and collaborations with the World Bank. Thank you Henry, Martin, Martha, Bernard, Sevara, and Jose Luis for giving me the opportunity to work with you and learn about the railway system in Tanzania, India, Kenya, Rwanda, and Burundi. I also want to thank the Institute for Research in Technology (IIT), my "home" before MIT, where I learned so much about railways and optimization. Thank you Carlos Batlle, Alexis, and many others for being unconditional friends and advisers. I also want to thank Comillas Pontifical University and Portaceli for all they taught me over the years.

I also want to thank Enrique, Marta, Joel, Carles, Toni, Carlos, Xabi, Paula, Ferran, David, Carmen, Helena, Anna, Quique, Roberto, Paula, Rafa, Claire, David, Santiago, Chema, Lucinio, Carlos, Patty, Adam, Asun, and many other Spanish and non-Spanish friends who have made these years memorable. Thank you for all the barbeques, dinners, games, hikes, trips to the beach, and your support at many times. Thanks also to Coque, Nacho, Bea, Alberto, Javi, Mario, Chus, Diego, Ale, Richi, Luis, Elisa, Ceci, Gonzalo, Rocio and David, my undergraduate, mathematical competitions, and high-school friends, to keep calling me and writing to me even when I was busy and took ages to answer.

Last but not least, I want to thank my family for all their support. My grandparents, my aunts and uncles, my cousins, my parents, my siblings and my husband have been supporting me all this time. Mom and Dad, thank you for convincing me that I could take any train I want if I work hard enough. Ali and Manuel, thank you for being unconditional friends and allies on every journey. Noel, thank you for your love, for your support, for your company, for preparing the best meals and the best tuppers, and for smiling and encouraging me to continue when I was tired. This would have not been possible without you.

I know that trying to keep this brief, there are many of you that I did not named explicitly, although you have greatly contributed to this experience. Thank you.

With the desire of keep riding many new trains with you,

Cambridge, Massachusetts

May 2015

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADP	Approximate Dynamic Programming
ALP	Approximate Linear Programming
ARLP	Adaptive Relaxed Linear Programming
DP	Dynamic Programming
FRA	Federal Railroad Administration, US
HSR	High Speed Rail
IM	Infrastructure Manager
IRR	Internal Rate of Return
LP	Linear Programming
NEC	Northeast Corridor, US
NPV	Net Present Value
MILP	Mixed-Integer Linear Programming
OD	Origin-Destination (pair)
QARLP	Q-factor Adaptive Relaxed Linear Programming (algorithm)
RLP	Relaxed Linear Programming
TO	Train Operator

Chapter 1 - Introduction

"The more we share, the more we have" – L. Nimoy (1995)

Infrastructures create the opportunity for essential services that should be put in place to enable economic activity, economic growth, and development (Munnell, 1992; World Bank, 2008). However, investment in infrastructure is extremely expensive, so high demand and strong business cases are typically necessary to justify such investments. In the past, large infrastructure systems were vertically integrated, i.e., the same entity was in charge of both managing and operating the system. The lack of competition and efficiency incentives of those entities erodes the performance of the system and has motivated over the years the introduction of new institutional structures. Shared infrastructure systems are widely proposed today as to 1) minimize the amount of infrastructure required to serve societal needs, and 2) achieve high levels of utilization and help recover the costs of the very high capital-intensive infrastructure in service systems such as the electric power sector, telecommunications, or transportation.

The greatest challenge for shared infrastructures is managing and coordinating access of competitive agents to the infrastructure (Gomez-Ibanez, 2003). The key question is determining an appropriate mechanism for deciding which agents can access the infrastructure at each time (*capacity allocation*) and the access price that each agent should pay (*capacity pricing*). The benefits of sharing infrastructure are particularly substantial in railway systems, where the infrastructure represents around 40-60% of the total final service cost (Gomez-Ibanez, 2003), creating the potential for sizable savings. However, the analysis and comparison of alternative capacity pricing and allocation mechanisms is particularly complex in railway systems, because there are strong interactions between capacity planning and infrastructure operations (Krueger et al., 1999; Pouryousef and Lautala, 2015).

The objective of this thesis is to analyze and compare alternative mechanisms for capacity pricing and allocation in shared railway systems according to several performance metrics. The main contribution of this thesis is the development of a framework to systematically analyze and compare alternative capacity pricing and allocation mechanisms. The results of this research are expected to be valuable for railway regulators, but also to infrastructure managers (IMs) and train operators (TOs), allowing them to better understand the implications of these mechanisms at the system level and to better plan shared-use railway systems by the implementation of appropriate capacity pricing and allocation mechanisms. This understanding will be valuable for other shared infrastructure systems as well.

1.1 Shared Railway Systems: Promises and Challenges

Recently, governments have started promoting the use of shared railway systems. Up until 1988, all major railways both managed the infrastructure and operated the trains, i.e., they were vertically integrated (Drew, 2006). In contrast, in shared railway systems, multiple TOs utilize the same infrastructure, i.e., there is some level of vertical separation between infrastructure management and train operations. Examples of shared railway systems are the Northeast Corridor (NEC) in the US and the railway system in most European countries. Several countries in Asia and Africa are also opening access to their railway systems.

Proponents of shared railway systems stress that the use of shared railway systems allows 1) a more efficient use of expensive railway infrastructure and 2) the introduction of competition. Achieving a more efficient use of current infrastructure is positive not only in cases when resources to invest in infrastructure are limited, but also in cases where additional deployment of infrastructure is simply not possible. Recovery of infrastructure investment is one of the main reasons behind the implementation of open-access in Tanzania (Pena-Alcaraz et al., 2014; World Bank, 2014). The difficulties in adding additional capacity, especially near the densely populated area of Penn Station in New York City, are one of the main reasons why multiple TOs share existing railway infrastructure in the NEC (Gardner, 2013). According to (Gomez-Ibanez, 2003), rail infrastructure is a natural monopoly but the train operations business is not. As a result, with shared use and open access, new competitors would be able to enter the train operations business with its consequent benefits for the end users. This is the main rationale behind the European Union railway packages (Perennes, 2014).

As mentioned above, however, shared railway systems can only provide these benefits when there is a strong coordination between the IM and the TOs (Gomez-Ibanez, 2003). Such coordination, in turn, requires capacity planning mechanisms that determines which trains can access the infrastructure at each time, capacity allocation, and the access price they need to pay, capacity pricing (Pena-Alcaraz, 2015). It is important to maintain transparency when the IM is also one of the TOs.

There is a broad literature that has explored various capacity pricing and allocation mechanisms for railways (Affuso, 2003; Crozet, 2004; Gibson, 2003; Nash, 2005; Perennes, 2014). In general, different countries have promoted different mechanisms for capacity pricing and allocation, with differing objectives. Studies have tended to focus on one mechanism and evaluated it according to performance metrics unique to that mechanism, making the comparison across different mechanisms to price and allocate railway capacity quite difficult and the implications for other systems ambiguous.

The objective of this research is to analyze and compare alternative mechanisms for capacity pricing and allocation in shared railway systems according to several performance metrics. We consider multiple criteria to analyze the performance of capacity pricing and allocation mechanisms from the perspective of the IM (cost recovery, capacity utilization), the

TOs (access charges, trains scheduled), and the end users (number of services, passenger fares or freight shipping rates). This thesis hypothesizes that alternative capacity pricing and allocation mechanisms would perform well for some metrics, but there is no silver-bullet mechanism that would perform well in all the metrics for every shared railway system. A better understanding of these trade-offs in performance is of particular importance today (Drew and Nash, 2011; Nash, 2010), since several countries are restructuring their railway sector to allow shared use. This understanding would allow the regulators of each country to design the most appropriate capacity pricing and allocation mechanisms to unlock the benefits of shared use in their railway system.

1.2 Literature Review

As noted earlier, the objective of this thesis is to analyze and compare alternative mechanisms for capacity pricing and allocation in shared railway systems. This section presents an overview of the main capacity pricing and allocation mechanisms proposed in the literature, and discusses the experiences in shared railway systems for different countries. This section also summarizes some lessons from other network industries that are then used as guiding principles in this research.

1.2.1 Capacity Pricing and Allocation Mechanisms

There are three main types of mechanisms to price and allocate capacity: negotiation-based, administrative-based, and market-based mechanisms. Under negotiation-based mechanisms, the TOs and the IM negotiate to determine which trains can access the infrastructure and at what price. The main drawback of negotiations is that they can be very complex and time consuming (Nash, 2003). In addition, they often result in non-transparent bilateral contracts that prevent adaptation to future needs or create barriers to new operators. Under administrative-based mechanisms, the regulator establishes access rules and oversees the capacity pricing and allocation process. The regulator punishes (e.g. fines) any deviation from the rules. The use of these types of mechanisms relies on the ability of the regulator to gather information from the TOs and the IM to eliminate information asymmetries. These mechanisms are also slow to adapt to new system needs.

The shortcomings of negotiation-based and administrative-based mechanisms, together with the need for transparency and non-discriminatory access have motivated the introduction of market-based mechanisms for capacity pricing and allocation. In the NEC, for instance, with current bilateral infrastructure access contracts, 1) the price that each TO pays to access the infrastructure depends on the time at which the contract was signed (companies who signed their agreements when there was still plenty of excess capacity are paying much less than other companies to operate the same type of services), 2) access charges and slots are rigid (none of the companies want to lose their current slots and it is difficult to make room for new trains because multiple contracts would need to be renegotiated), and 3) the IM is not able to raise enough revenues to afford basic maintenance of the lines (this has contributed to the current backlog in maintenance in the NEC) (Gardner, 2013). As a result, the Federal Railroad Administration (FRA) requires all the railroads to agree on a market-based mechanism for pricing and allocating capacity to substitute for the current negotiation-based mechanism (PRIIA, 2008). According to (Gibson, 2003), there are two main types of market-based mechanisms for capacity pricing and allocation: 1) price-based and 2) capacity-based.

Price-based mechanisms are those that determine the price at which capacity will be offered, and let TOs decide whether they are willing to access the infrastructure or not. Price-based mechanisms are typically complemented with priority rules that allow the IM to decide which train to schedule when there are conflicts (multiple TOs willing to pay the predetermined access charges). An example of a price-based mechanism would be a cost-allocation mechanism that assigns infrastructure-related cost proportionally to the volume of infrastructure use (Crozet, 2004; Nash, 2005; Lopez-Pita, 2014; Texeira and Prodan, 2014). The access charge could also be adjusted considering the TOs' demand for scheduling trains (e.g. introducing congestion prices). These charges could also be adjusted with a base tariff that allows the IM to recover infrastructure costs that are fixed in nature.

Capacity-based mechanisms are those that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity, e.g. an auction (Affuso, 2003; McDaniel, 2003; Newbury, 2003; Perennes, 2014; Stern and Turvey, 2003). There are multiple types of auctions: simple auctions in which TOs bid to get some predefined slots (either in a segment of the infrastructure or for the full path) or submit their desired timetable when they bid, and combinatorial auctions where the TOs' bid depends on the result of the auction. Capacity-based mechanisms have been widely studied in the literature but have not yet been implemented on the railway system in any country.

1.2.2 International Context

As Table 1-1 shows, shared railway systems are not an isolated phenomenon in the NEC. Starting in 1991 several countries have started opening access to their railway systems. However, different countries have adopted very different mechanisms to price and allocate railway capacity. The US uses negotiation-based, Australia and India administrative-based and European Union countries market-based mechanisms to allocate capacity.

Capacity pricing also varies from country to country. Although most IMs charge TOs the marginal cost of operating the train on the infrastructure (Nash, 2003), the calculation of the marginal cost of operating one more train on the infrastructure is based on several assumptions. As a result, (Nash, 2005; Lopez-Pita, 2014; Texeira and Prodan, 2014) conclude that charging mechanisms in shared railway corridors are getting more heterogeneous.

Furthermore, different countries design capacity planning mechanisms with different objectives and evaluate those using different metrics. As a result, the comparative performance of different mechanisms is still unclear (Drew and Nash, 2011; Nash, 2010). According to Nash

(2003), “it is important to recognize that the concept of multiple operators may be relatively new for railroads: This means that the institutional framework has not been developed, and the intellectual understanding may not be in place, to facilitate planning and operating the shared-use system.” The authors warn against moving ahead quickly with the design of pricing and allocation mechanisms before understanding the implications of such mechanisms for all stakeholders.

Table 1-1 Railway systems international organization around the world (Source: author, based on (ADB, 2014; Agosta, 2015; Gomez-Ibanez and de Rus, 2006; Levy, 2015; Olievschi, 2013; Pozzo di Borgo, 2005; Sakamoto, 2012; Texeira and Prodan, 2014; The Economist, 2015))

Area	Railway System Characteristics
Africa	Seven African countries present vertically-integrated railway systems (Algeria, Botswana, Egypt, Morocco, Namibia, South Africa, and Tunisia). In the rest of Africa, most railway systems were concessioned between mid-1990s and 2010 (e.g. Burkina Faso, Cameroon, Ivory Coast, Mozambique, Senegal, Tanzania, and Togo). Starting in 2010, there has been a promotion of open-access and shared corridors, especially in those countries were concessions failed (e.g. Guinea and Tanzania)
North and South America	Canada and the US present vertical integration in their freight railway system. In both countries, private freight operators have to accommodate passenger operators on the tracks. Other countries like Cuba and Honduras also have a vertically-integrated railway system. Argentina has recently announced that the railway system will be nationalized and vertically integrated by the end of 2015. The railway system is currently vertically separated and concessioned in most countries in Latin America (e.g. Argentina, Brazil, Chile, Colombia, and Mexico). In the US there is an important shared railway system, the NEC. A similar shared system has been proposed now in California (blended system) to accommodate high-speed rail (HSR) and commuter services.
Asia	The railway system is vertically integrated in countries like China, India, Indonesia, Japan, Malaysia, etc. Starting in 2013, Indian railways changed the regulation to allow for open access and shared use in the new dedicated freight corridor. There is also open-access in Russia and Kazakhstan (freight cars only). Mongolia and Uzbekistan are also implementing open-access railway policies.

Europe	Most European countries have implemented open-access and shared railway use for freight following the EU first railway package of 1991. The passenger railway system in Europe is also vertically separated in most countries, also moving to open-access and shared use. By 2014 only Italy already had competition on the tracks, with two competing HSR companies offering services, although other countries like Romania, Spain, and Ukraine are also moving towards the introduction of competition.
Oceania	The railway system in New Zealand was renationalized and vertically integrated in 2008. The railway system in Australia is vertically separated and operates with open-access and shared use policies.

The power sector, the telecommunication sector, and the aviation industry have made significant progress sharing infrastructure as compared to the railway industry. The experience in these networks show that: 1) marginal access pricing have significant advantages over other capacity pricing mechanisms in most circumstances, 2) capacity pricing and allocation are complementary problems, and 3) price-based and capacity-based mechanisms often yield equivalent results. First, marginal access pricing ensures that vertically-separated agents' operational decisions match the decisions of an integrated company. In addition, (Perez-Arriaga and Olmos, 2009; Perez-Arriaga, 2013; Rubio, 1999) show that marginal access pricing allows the IM to recover fixed infrastructure costs when there is no lumpiness in infrastructure investment, no uncertainty, no information asymmetry, and when the operators do not have market power. Note that those conditions never occur in reality. However, the economic results are still useful when there are small deviations from these conditions (e.g. operators do not have strong market power). Second, capacity pricing and capacity allocation are two sides of the same coin in shared corridors. For instance, if access charges are low, the demand may exceed the available capacity, as occurred in US airports after the Airline Deregulation Act of 1978 (Vaze, 2011). In these cases, a capacity allocation mechanism is needed. Conversely, the number and type of slots available affect the operators' willingness to pay to access the infrastructure (Laffont and Tirole, 1993; Laffont and Tirole, 2000; Vazquez, 2003). Third, there is no difference between price-based and capacity-based mechanisms to price and allocate capacity if there is perfect information and no uncertainty according to (Weitzman, 1974). However, under imperfect information and uncertainty the specific design influences the performance of the system (Czerny, 2010).

As mentioned before, however, there is an important difference between the railway industry and other network industries. In shared railway systems, capacity planning at the strategic level is tightly coupled with infrastructure operations at the tactical level (Krueger et al., 1999; Pouryousef and Lautala, 2015). The understanding of the implications of pricing and allocation mechanisms thus requires 1) determining the TOs' demand for scheduling trains and 2) designing the optimal train timetable to determine whether the TOs' demand for scheduling trains can be accommodated on the infrastructure. Consequently, it is difficult to predict whether the conclusions from other network industries will apply in this case. This thesis addresses these questions and uses the lessons from other network industries as guiding principle for the research, when appropriate.

1.3 Thesis Contributions

The objective of this thesis is to contribute to the field of shared railway systems coordination by 1) developing a framework to analyze the performance of shared railway systems under alternative capacity pricing and allocation mechanisms, and 2) using this framework to understand the implications of representative capacity pricing and allocation mechanisms in representative shared railway systems.

There are strong interactions between capacity planning and infrastructure operations in the railway industry; the operations on the infrastructure determine the available capacity in the system. The framework developed in this thesis to evaluate the performance of shared railway systems under alternative capacity pricing and allocation consists of two models: 1) a financial model that anticipates how TOs would respond to the capacity pricing and allocation mechanisms and determine their demand for infrastructure use, and 2) a network optimization model that determines the optimal train timetable (IM's decisions) that accommodates the TO demand for scheduling trains, considering the topology of the line, safety constraints, and other technical aspects of the infrastructure. We use this framework to analyze the trade-offs associated with the use of alternative mechanisms in the context of the national rail system for Tanzania, which is implementing a new open-access model, and in the context of the NEC in the US, where the FRA is now trying to develop a new capacity pricing and allocation mechanism.

The main contributions of this thesis are:

1. Formulation of a TO Model to anticipate the response of TOs to capacity pricing.
2. Formulation of the train timetabling problem (IM Model) for capacity allocation in shared railway systems to model interactions between capacity planning and infrastructure operations.
3. Development of an approximate dynamic programming algorithm based on linear programming to be able to solve the train timetabling problem in relevant instances.
4. Identification of cases in which traditional capacity pricing mechanisms result in sub-optimal operational decisions, with an illustration of this problem in case studies based on the Central Corridor in Tanzania and the Northeast Corridor in the US.
5. Comparison of two representative capacity pricing and allocation mechanisms in those systems, concluding that there are important trade-offs between them.

The first contribution of this thesis is the formulation of a financial model to capture how rational TOs would behave under a capacity pricing and allocation mechanism, given the institutional and regulatory framework and the technical characteristics of the specific type of railway service that they provide (see Chapter 2). The model analyzes three main operational decisions: 1) the TO's demand for scheduling trains, 2) the TO's willingness to pay to access the infrastructure, and 3) the passenger fare or freight shipping rate they would charge to end users. The first two operational decisions are then used as inputs for the IM Model. The model proposed is simple by design. The main objective is to allow the regulator and the IM to anticipate the response of the TOs to alternative capacity pricing and allocation mechanisms. More detailed models would rely on extensive information about the TOs that is not readily available to the regulator or the IM (Levy et al., 2015).

The results obtained in Chapter 2 show that while the estimates of the fares charged to end users are sensitive to the demand curve and elasticity assumptions; the TOs' willingness to pay for access estimates are robust to model assumptions. This suggests that the level of detail of the model is adequate to capture the interactions between the TOs and the IM. This model analyzes each TO independently of other TOs. Once we know all of the TOs' demands for scheduling trains, we need to determine if there is capacity available to schedule all the services.

The second contribution of this thesis is the formulation of a train timetabling model for shared railway systems. The train timetabling problem has been widely studied in the literature (Cacchiani et al., 2010; Caimi et al., 2009; Caimi et al., 2011; Caprara et al., 2002; Caprara et al., 2011; Castillo et al., 2009; Cordeau et al., 1998; Ghoseiri et al., 2004; Liebchen, 2008; Liebchen and Peeters, 2009; Pena-Alcaraz et al., 2011; Zhou and Zhong, 2005). However, we argue in Chapter 3 that traditional train timetabling models cannot be used to analyze capacity planning mechanisms in shared railway systems for three reasons. First, most models assume a fixed number of trains to be scheduled on the infrastructure. However, the number of trains to schedule is the main decision variable of the capacity allocation problem in shared railway systems. Second, most approaches also assume a single TO that tries to schedule trains. This TO could iteratively solve the train timetabling problem, introducing small modifications in each train's desired timetable until the resulting timetable meets its needs. In shared railway systems, in contrast, multiple TOs request access to the infrastructure and the TO is informed afterwards whether its train can be scheduled or not. We argue that consequently, TOs have incentives to be flexible in shared railway systems to ensure that most of their trains get scheduled in the first iteration. Third, most of these models assume that all trains follow the same path. Again, this assumption does not hold when the nature of the services operated in the shared railway system is different. For example, commuter services are typically scheduled around metropolitan areas, whereas intercity and freight operators offer services between cities.

Chapter 3 of this thesis presents a train timetabling problem formulation for shared railway systems that explicitly considers a variable number of trains, with large flexibility margins, traveling along different paths. This approach 1) introduces a discrete variable that indicates whether a train can be scheduled or not, 2) uses flexibility margins to ease conflicts, making fast trains travel slowly when there are slow trains ahead and making slow trains wait at sidetracks when fast trains overtake them, and 3) specifies safety constraints (spacing of the trains) for each train path.

These three additional considerations make the train timetabling problem for shared railway corridors very difficult to solve. From a computational standpoint, the size of the model increases rapidly (more than linear) with the number of stations and the number of trains to schedule. As a result, commercial solvers are only able to solve the problem for a relatively small number of trains. Furthermore, most of the techniques developed in the train timetabling literature are designed for traditional single-operator train timetabling problems and cannot be used in this case. Most classical decomposition approaches do not work because of the large number of discrete variables needed to specify which trains are scheduled and the order in which trains go through each station. Any technique that exogenously fixes train order cannot be used here because of the large flexibility margins and because train spacing constraints are specific to each individual train.

As a result, this thesis proposes an alternative class of solution algorithms using approximate dynamic programming techniques (Bertsekas and Tsitsiklis, 1996; Bertsekas, 2006; Powell, 2007) to be able to solve the problem in meaningful instances. The third contribution of this thesis is the development of a novel Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm that extends previous algorithms developed by (Farias and Van Roy, 2003; Farias and Van Roy, 2004). This algorithm allows us to decompose and solve large problems that are intractable with MILP commercial solvers while still converging to a solution within an optimality gap (Pena-Alcaraz et al., 2015a). This is the main contribution of this thesis because this algorithm makes it possible to analyze price-based and capacity-based mechanisms considering interactions between infrastructure operations and capacity available.

The fourth contribution of this thesis spans Chapters 2, 4, and 5. Chapter 2 compares the operational decisions of an integrated railway company with the operational decisions of vertically-separated TOs under different institutional and regulatory environments. It shows that both vertically-separated perfectly-competitive and vertically-separated regulated monopolistic TOs would make the same operational decisions as an integrated railway company when variable access charges reflect variable IM costs to operate trains on the infrastructure. Not surprisingly, the capacity pricing literature (Lopez-Pita, 2014; Nash, 2003; Texeira and Prodan, 2014) recommends the use of access charges equal to the infrastructure marginal costs of operating the train. However, Chapter 2 also shows that this approach cannot be used in two cases: 1) when the IM needs to recover part of the infrastructure management fixed costs, or 2) when the railway system is congested (overcrowded or overloaded with traffic). Most railway systems fall into at least one of these two categories.

We illustrate the first case in Chapter 4 in the case of the Central Corridor in Tanzania. The Central Corridor goes from the port (Dar es Salaam) to an inland container terminal (Isaka) that serves as a dry port for Rwanda, Burundi, Uganda, and the Eastern portion of Democratic

Republic of Congo. The infrastructure is owned by RAHCO, a publicly owned company. TRL is the only current TO; it operates around six trains per week. Although the corridor is single track, there is plenty of spare capacity that could be used by multiple private companies that have expressed interest in starting operating new services between Dar es Salaam and Isaka (Pena-Alcaraz et al., 2014; World Bank, 2014). Due to the low number of trains that operate in the system, the infrastructure maintenance costs do not increase (for all practical purposes) when more trains are operated. As a result, maintenances costs are assumed fixed.

If access charges are set following the traditional approach (access charges equal to the variable costs of managing and maintaining infrastructure), TOs would pay zero (0) access charges to access the infrastructure and the IM would not collect any revenues. However, it is critical to ensure that the IM is able to raise revenues to maintain the infrastructure and keep the system operational. As a result, the IM has to allocate infrastructure fixed costs among TOs through the access charges. Chapter 4 shows that the introduction of (non-zero, in this case) variable access charges distorts the operational decision of TOs. Chapter 4 then discusses how to avoid this problem with other price-based mechanisms such as the introduction of fixed access charges and how to allocate infrastructure costs among different types of TOs. Chapter 4 also analyzes the potentials of capacity-based mechanisms and shows that they would not allow the IM to recover infrastructure costs. Because the Central Corridor is currently not congested, capacity pricing and allocation can be solved independently. In other words, it is easy to analyze each TO independently because there is enough spare capacity to schedule all the trains.

Chapter 5 discusses how to analyze alternative capacity pricing and allocation mechanisms in the context of the NEC, a very congested shared railway system. In this case, we need both the TO Model to anticipate how each TO will respond to the mechanism and the IM Model to determine the final allocation of infrastructure capacity. The main spine of the NEC stretches from Boston, MA to Washington, DC. This segment is shared by several passenger and freight TOs: an intercity passenger TO that operates around 150 trains per day, eight commuter TOs that operate over 2,000 trains per day, and several freight TOs that operate around 70 trains per day. Until now, capacity pricing and allocation in the corridor has been managed via bi-lateral contracts negotiated between the IM and the TOs. However, the limitations of this negotiation-based mechanism motivated the FRA's request to Amtrak and the rest of the commuters and freight railway companies to agree on a new capacity pricing and allocation mechanism by the end of 2015 (PRIIA, 2008). Chapter 5 analyzes the performance of the system under two proposed capacity pricing and allocation mechanisms: a price-based (cost-allocation and priority-rule) mechanism proposed by Amtrak (Gardner, 2013; NEC Commission, 2014) and a capacity-based (auction) mechanism (Affuso, 2003; Perennes, 2014).

The results of Chapter 5 show that the capacity-based mechanism considered could result in almost 20% more IM cost recovery and 20% more trains scheduled as compared to the price-based mechanism considered in the NEC. The price-based mechanism, on the other hand, ensures higher profits for the TOs, making them more resilient to uncertainty in end users' transportation demand. However, this mechanism is not very resilient to uncertainty in infrastructure capacity availability. Under the capacity-based mechanism, NEC intercity TOs are in a better position than commuter TOs to access the tracks with current levels of service. The priority level of each TO is a design choice in price-based mechanisms. This choice, however, has important implications for NEC commuter and intercity passengers and TOs, especially if the IM does not have access to sophisticated methods to solve the train timetabling problem. To our knowledge, this is the first effort to compare a price-based mechanism and a capacity-based mechanism in the same shared railway system. The results show that there are very important trade-offs among them and that none of them is superior to the other in all dimensions.

1.4 Thesis Organization

The main body of this thesis consists of five chapters: Chapter 2 presents the TO Model used 1) to compare the operational decisions of an integrated railway company with the operational decisions of vertically-separated TOs considering the institutional and regulatory environments and the technical characteristics of the service, and 2) to anticipate the TO response to alternative capacity pricing and allocation mechanisms.

Chapter 3 describes the IM Model. It presents first the formulation used to determine the optimal train timetable in shared railway systems. This model considers TOs' demand for scheduling trains and determines the optimal train timetable; i.e., the optimal IM decision regarding which trains to schedule and their timetable. Chapter 3 then describes the algorithm proposed to solve the train timetabling problem in large-scale systems and analyzes the results obtained in different cases.

Chapter 4 uses the TO Model to analyze alternative capacity pricing mechanisms in the context of the Central Corridor in Tanzania. This chapter discusses the main policy implications of this research for shared railway systems where the IM has to assign fixed costs among different types of railway services.

Chapter 5 uses the TO and the IM Models to analyze and compare alternative capacity pricing and allocation mechanisms in the context of the NEC. This chapter discusses the main policy implications of this research for railway systems in congested shared railway systems.

Chapter 6 concludes this work by summarizing our finding and the main conclusions of our research. It also discusses the main policy implications and future directions of this research.

We now begin with a discussion of the TO Model and the TO response to capacity pricing and allocation mechanisms within different institutional and regulatory environments.

Chapter 2 - The Train Operator Problem: Determining Train Operator Response to Alternative Capacity Pricing and Allocation Mechanisms

"All models are wrong, but some are useful" – G.E.P. Box (1987)

The introduction of shared railway systems requires the design and implementation of capacity pricing and allocation mechanisms. Chapter 1 shows that there are several alternative mechanisms to price and allocate railway capacity. The objective of this thesis is to analyze and compare such mechanisms. The improved understanding of the implications of different capacity pricing and allocation mechanisms would allow the stakeholders of shared railway systems to design and choose the mechanism that best fits their system needs and overarching goals. This chapter argues that while there is an extensive literature focused on the relation between capacity pricing and infrastructure costs, the response of train operators (TOs) to capacity pricing and allocation mechanisms is still unclear. The first step to analyze capacity pricing and allocation mechanisms is thus to anticipate how TOs respond to capacity pricing.

This chapter proposes the use of a TO financial model (TO Model) to capture the interrelation between different TO's operational decisions. The two main contributions of this chapter are 1) to demonstrate that the TO Model responses are robust to a broad set of model inputs; and 2) to compare the operational decisions of an integrated railway company with the operational decisions of vertically-separated TOs considering the institutional and regulatory environments and the technical characteristics of the service. The operational decisions of integrated railway companies are then used as a benchmark to identify necessary conditions under which profit maximizing TOs would make operational decisions that also maximizes social welfare. An early version of this work is accepted for publication in Transportation Research Record (Levy, Pena-Alcaraz, Prodan, and Sussman, 2015).

The rest of the chapter is structured as follows: Section 2.1 reviews the main studies that analyze how capacity pricing and allocation affects the performance of shared railway systems, summarizes the contributions of this chapter, and discusses the modeling assumptions. Section 2.2 presents the TO Model and introduces the cases in which the model will be discussed. Section 2.3 analyzes how the TOs' estimated response changes with any changes in the TO Model inputs. Section 2.4 then uses the TO Model to compare the responses of TOs in different cases. All the results of this chapter are illustrated with examples based on the Northeast Corridor (NEC) in the US. Section 2.5 concludes with some highlights and recommendations.

2.1 Capacity Pricing and Allocation in Shared Railway Systems

This section summarizes the literature on capacity pricing and allocation in shared railway systems, presents the main contributions of this chapter, and finishes presenting a discussion of the main modeling assumption.

2.1.1 Literature Review

This section summarizes the main findings of two different bodies of literature: one that studies capacity pricing in the railway industry from the perspective of the infrastructure manager (IM) and other that studies the financial performance of TOs.

The capacity pricing literature in the railway industry focuses on the potentials for infrastructure cost recovery (Nash, 2005; Texeira and Lopez-Pita, 2012; Lopez-Pita, 2014; Texeira and Prodan, 2014). However, there are two main challenges to being able to relate infrastructure costs with capacity utilization in the railway industry: the nature of railway infrastructure costs and the nature of railway capacity. First, (NEC Commission, 2014) shows that different types of infrastructure costs vary with different operational variables (trains, trains-miles, gross-ton-miles, frequency of service, etc.). Second, the available capacity in the railway industry depends on infrastructure operations (Krueger et al., 2009; Pouryousef and Lautala, 2015). To understand infrastructure operations we need to be able to anticipate TOs operational decisions. According to (Nash et al., 2004; Drew and Nash, 2011), the impact of capacity pricing and allocation mechanisms on the TO operational decisions are still inconclusive (Nash et al., 2004; Drew and Nash, 2011).

There is also a broad literature that describes the TO revenues and costs for different operational decisions (Belli et al., 2001; Martland, 2011; PPIAF et al., 2011). These financial models pay little attention to access charges, since the need to establish capacity pricing and allocation mechanisms on the railway systems is still relatively new (Nash, 2003). Moreover, these models are usually descriptive, and they are rarely used to estimate TO operational decisions. For these reasons, TO financial models have not yet been used to anticipate how TOs respond to different pricing mechanisms. This literature gap prevents us from understanding the implications of capacity pricing and allocation mechanisms for shared railway systems.

2.1.2 Chapter Contributions

The objective of this chapter is to help fill the identified literature gap by 1) developing a TO financial model (TO Model) that explicitly models the relation between TO operational decisions and access charges; 2) analyzing the sensitivity of the results of the model to changes in the model inputs; and 3) proposing the use of the model to compare the behavior of TOs within different institutional and regulatory environments.

First of all, this chapter proposes a TO Model that discusses the TOs' response to alternative capacity pricing and allocation mechanisms as a function of the regulatory environment, the competitive landscape, and the characteristics of the type of railway service that each TO offers. The TOs' response is characterized by: 1) the passenger fare or freight shipping rate charged to end users; 2) the number of trains operated; and 3) the access charges paid to the infrastructure manager to access the infrastructure. We call these variables *operational decisions*.

Second, the objective of the TO Model is to allow regulators and IMs to anticipate the TO demand for scheduling trains and their ability to pay to access the infrastructure. In that sense, the TO Model is designed to rely only on public information about the TOs that is already available to the regulators and IMs. We need to be sure, though, that the estimates obtained without detailed TO information are accurate. This chapter carries out sensitivity analyses and studies the results obtained with different model inputs. We show that the TO demand for scheduling trains and the TO ability to pay to access the infrastructure are very robust to model inputs. In other words, the TO demand for infrastructure use does not change much with small changes in the inputs of the model (cost and demand estimates). This suggests that the level of detail of the model is adequate to capture the interactions between the TOs and the IM.

Third, we use the TO Model to analyze the operational decisions of integrated railway companies. We use these results to compare those obtained within different institutional and regulatory environments. The results show that the operational decisions of a profit-maximizing TO match the decisions of an even-handed integrated railway company when variable access charges reflect variable IM costs to operate trains on the infrastructure. This finding justifies the use of traditional mechanisms adopted in most countries to price and allocate capacity. However, the results also show that there are two cases in which these mechanisms cannot be used: 1) when the IM needs to recover part of the infrastructure management fixed costs; and 2) when the railway system is congested. Unfortunately, most railway systems fall into at least one of these two cases. This motivates the choice of the two cases studied later in this thesis: Chapter 4 studies the case of Tanzania where the IM needs to recover infrastructure costs and Chapter 5 studies the NEC, the most congested railway system in the US.

2.1.3 Model Assumptions

The TO Model proposed in this chapter assumes that: 1) TOs make operational decisions with the objective of maximizing profits; 2) each train serves a single origin-destination (OD) pair; and 3) different types of services are not substitutable. This section explains why we make these assumptions, how we use them, and how we expect these assumptions to affect the results.

The first assumption is necessary to determine TO operational decisions given the main revenue and cost streams. This is a standard assumption to replicate the decision process followed by private companies and is consistent with the privatization of TOs that has followed the implementation of many shared railway systems. This assumption allows us to determine the most likely TO operational decisions given revenues and costs.

The second assumption is used to compute the demand transported, relating the end users' demand for transport with the train capacity. We assume a single OD pair because the data that TOs publish in their annual financial reports typically aggregates all the revenues obtained in the same corridor. As a result, we only have information about the fares and distance of an average trip. The impact of this assumption in the results depends on the nature of the services. In the NEC, where most trips occur between Boston and New York City, and New York City and Washington DC, this assumption leads us to underestimate the capacity of the trains. If most trips share a segment (as it does with trips from New York City to Philadelphia and New York City to Washington DC) this assumption would lead to overestimating train capacity. This bias could be corrected by (in order of increased complexity): adjusting the final number of trains in cases in which the capacity of the train is binding, adjusting the train capacity used as an input of the model, or including information of all OD pairs.

The third assumption is used to be able to solve the TO Model independently for different types of services. If the TOs provides partially substitutable services (e.g. the case Amtrak's high-speed service Acela and Amtrak's regional service in the NEC that serve the same OD pairs with different speeds) then the access charges, number of services, and fares of these services should be determined at the same time considering all the costs and revenues related to these services in the TO Model. Otherwise, the end user's demand may be overestimated.

Note again that we use the TO Model to anticipate the TOs' response to capacity pricing and allocation. In that sense, we are mostly interested in estimating the TO demand for scheduling trains and the TO ability to pay for access that we use as inputs of the IM Model in Chapter 3. Section 2.3 shows that these variables are very robust to model inputs. This finding suggests that the level of detail of the TO Model is adequate for the purpose of this thesis. More sophisticated TO Models are necessary to address other research questions that require higher level of detail in the understanding of the TO operations or the TO–end user interactions.

2.2 Train Operator Model

The objective of this section is to discuss the operational decisions of rational TOs under a capacity pricing and allocation mechanism, given the institutional and regulatory framework and the technical characteristics of the specific type of railway service that they provide. As discussed above, there are three main operational decisions that TOs initially control: 1) the passenger fare or freight shipping rate charged to end users; 2) the number of trains operated; and 3) the access charges paid to the infrastructure manager to access the infrastructure.

As was mentioned before, we assume that rational TOs are profit maximizing. In other words, they make operational decisions with the objective of maximizing profits. We also analyze whether TOs are medium term sustainable agents, i.e., whether they are able to ensure positive cash flows in the medium term. TO profits and cash flow can be determined by analyzing TO revenues and costs for a given number of trains. In the rest of the chapter we use capital letters to denote operational decisions (decision variables) and lower-case letters to denote model inputs (parameters):

N number of trains services that the TO would like to schedule.

There are three main types of costs that TOs face:

AC the cost of accessing the tracks or access charges if the TO schedules any trains. This cost often has a fixed and a variable component: $AC(N \neq 0) = AC_f + AC_v \cdot N$.

fc fixed costs, such as the cost of buildings and the cost of purchasing cars and locomotives in the medium term.

vc variable costs of operating a train, such as fuel, personnel, train maintenance, and train lease, if trains are being leased.

TOs face fixed costs independently of any operational decision. These costs do not vary over the medium term. Variable costs depend on the number of trains operated. TOs know variable cost per train before they make operational decisions. Finally, TOs face access charges if they decide to schedule trains. The exact value of the access charges depend on the level of service. In general, TOs also know how much they will be charge as a function of their demand. This does not happen in the Netherlands, where the IM determines access charges once it knows all the trains scheduled (Texeira and Lopez-Pita, 2012). Although determining access charges is a good way to ensure that all infrastructure costs are recovered; this practice is not recommended neither implemented in most countries because it increases the uncertainty faced by TOs.

The two main sources of revenue come from transporting users (cargo or passenger) and from the government (subsidies). The revenues obtained from transporting users can be determined by multiplying the passenger fare or freight shipping rate by the demand transported. The demand transported is limited by either the capacity (reduced by a reasonable average loading factor) of the trains or by end users' demand. According to literature, end users' transportation demand depends fundamentally on the fare, the frequency of the service, and the travel time (Bebiano et al., 2014). While intercity passengers are typically more sensitive to the fare and the travel time, commuter passengers are typically more sensitive to the fare and the frequency, and freight users tend to be sensitive to the fare. The frequency of the service is inversely proportional to the number of services when we assume a single OD pair and uniform services. Government subsidies depend in general on the demand transported too.

F passenger fare or freight shipping rate.

s government subsidies.

c capacity of the trains (maximum number of passengers or net tons).

d end users' demand for transportation.

tt travel time.

Summarizing, the costs and revenues of a TO can be determined using the following formulas:

$$cost(N, AC) = fc + vc \cdot N + AC(N) \tag{2.1}$$

$$revenues(F, N) = s(F, N) + F \cdot min(d(F, N, tt), N \cdot c) \tag{2.2}$$

Note that some of these variables may be pre-determined or conditioned by regulations. For instance, the fare of commuter services is typically set by the government. Likewise, access charges under price-based mechanisms are fixed inputs for TOs.

In order to characterize how TOs operate, we will use equations (2.1) and (2.2) to relate the TO operational decisions. For instance, given the access charges (AC), the TO demand for scheduling trains and the fares charged to the end users can be determined maximizing profits, $\max_{N,F}[revenues(F,N) - costs(N)]$:

$$\max_{N,F}[s(F,N) + F \cdot \min(d(F,N,tt), c \cdot N) - fc - vc \cdot N - ac(N)] \tag{2.3}$$

Fixed costs do not depend directly on the number of trains operated or the fare. If the subsidy does not depend directly on the number of trains then equation (2.3) is equivalent to: $\max_{N,F}[F \cdot \min(d(F,N,tt), c \cdot N) - vc \cdot N - ac(N)]$.

Since the number of trains and the fares depend on the access charges, the access charges can be determined implicitly using sensitivity analysis. In general, the TO maximum willingness to pay to access the infrastructure can be determined considering that all variable costs should not exceed variable revenues (to ensure that the TO is interested in operating trains). The function that equation (2.3) maximizes depends on the number of trains. We know that if the TO decides not to operate any trains, the TO will not have any variable costs and it will not have to pay to access the infrastructure. Similarly, it will not receive any revenues from operations. However, in the medium term, the TO faces fixed costs fc independently of the decision of how many trains N to operate and may sometimes receive a subsidy $s(0)$ (typically $s(0) = 0$). As a result, the TO faces a profit of $s(0) - fc$ when it does not operate any train. As a result, the TO would never operate a number of trains that results on smaller profits than $s(0) - fc$, because it would be better off simply not operating any trains. In other words, the maximum of equation (2.3) can never be lower than $s(0) - fc$. As a result, the TO maximum willingness to pay to access the infrastructure as a function of F, N is presented in equation (2.4).

$$\begin{gathered} AC_v \cdot N \leq {}^{\partial s(F,N)}\!/_{\partial N} \cdot N + F \cdot \min(d(F,N,tt), c \cdot N) - vc \cdot N \\ AC_f \leq \max(0, s(F,N) + F \cdot \min(d(F,N,tt), c \cdot N) - vc \cdot N - AC_v \cdot N) \end{gathered} \tag{2.4}$$

The maximum willingness to pay to access the infrastructure for which the TO is sustainable in the medium term can be determined ensuring that the resulting cash flow is positive, $revenues - costs(AC) \geq 0$. Although capital expenditures (CAPEX) and financing costs are also required to compute cash flows, we will initially assume that TOs have almost no CAPEX and negligible financing costs. As a result, the TOs willingness to pay to access the infrastructure as a function of the number of trains and the fares can be calculated using equation (2.5):

$$AC(n) \leq s(f,n) + f \cdot \min(d(f,n,tt), c \cdot n) - fc - vc \cdot n \quad (2.5)$$

The implications of equations (2.3), (2.4), and (2.5) depend on the context in which TOs operate. The context is determined by both the institutional and regulatory environment, and the technical characteristics of the type of railway service that the TO provides.

2.2.1 Institutional and Regulatory Environment

There are three main institutional factors that we have to consider to study equations (2.3), (2.4), and (2.5): the vertical structure of the system, the regulation of the TOs, and the competitive landscape.

The vertical structure of the system determines whether the railway system is vertically integrated or vertically separated. Although capacity pricing and allocation mechanisms are critical when there is some level of vertical separation between the TO and the IM, we will also study the behavior of vertically-integrated railway systems as a basis of comparison for vertically-separated railway systems.

The regulatory environment determines how much control the TOs have over the three operational decisions: 1) the passenger fare or freight shipping rate charged to end users; 2) the number of trains operated; and 3) the access charges paid to the infrastructure manager to access the infrastructure. This chapter distinguishes unregulated and regulated systems. Within unregulated markets, the TO fully controls the first two decisions. The TO level of control over the access charges is affected by the capacity pricing and allocation mechanism in place. In regulated environments, a central planner controls the fares that end users pay or the rents that the TO extracts.

Since the capacity pricing and allocation mechanism also impacts the interactions between the TOs and the IM, the mechanism itself informs us about the most relevant interactions to study. For instance, determining the TO willingness to pay to access infrastructure as a function of the number of trains scheduled is particularly relevant when capacity-based mechanisms are used to allocate and price capacity. Likewise, determining the number of trains that the TO is willing to provide given the access charges is particularly relevant to design price-based mechanisms.

Finally, the competitive landscape also determines the response of the TOs to capacity pricing and allocation mechanisms. TOs will behave substantially different if they operate in monopolistic, oligopolistic, or a perfectly competitive market.

2.2.2 Technical Characteristics

The technical characteristics of the railway system determine the parameters of the model (cost, capacity of the trains). They also determine the nature of the demand. As was mentioned before, the end users' transportation demand depends fundamentally on the fare, the frequency of the service (inversely proportional to the number of trains), and the travel time (Bebiano et al., 2014). While intercity passengers are typically more sensitive to the fare and the travel time, commuter passengers are typically more sensitive to the fare and the frequency, and freight users tend to be sensitive to the fare. The literature proposes three functional forms to capture the dependency of the demand on these factors: linear demand function, isoelastic demand function, and bounded isoelastic demand function. In the three cases, the elasticity determines the relation between changes in demand as other factor (x) changes: $e = \frac{\Delta d / d_0}{\Delta x / x_0}, e = \frac{\partial d}{\partial x}$.

Figure 2-1 shows a comparison of these functions in a case in which the demand depends only on the fare, for an elasticity value $e = -0.67$, and an initial demand $d_0 = 31{,}250$ for an initial fare $f_0 = \$96.5$. An isoelastic curve with these parameters indicates implies that there is an unlimited demand for the transportation service as the fares decrease. Similarly, it assumes that there are few passengers willing to pay extremely large fares. With these parameters, the isoelastic demand function indicates that there is at least one passenger willing to pay up to $\$481$ million to travel by train. In practice, isoelastic demand functions approximate the end users' demand well for intermediate values of fares; but it is unrealistic to assume that there are no end users' demand or fare willingness to pay bounds. This thesis thus analyzes only linear and bounded isoelastic curves to capture the nature of end users' transportation demand.

Considering the combinations of institutional and regulatory, and technical factors we propose four main cases of study:

1. Vertically-separated unregulated monopolistic train operator
2. Vertically-separated regulated monopolistic train operator
3. Vertically-separated perfectly competitive train operator
4. Vertically-integrated railway company

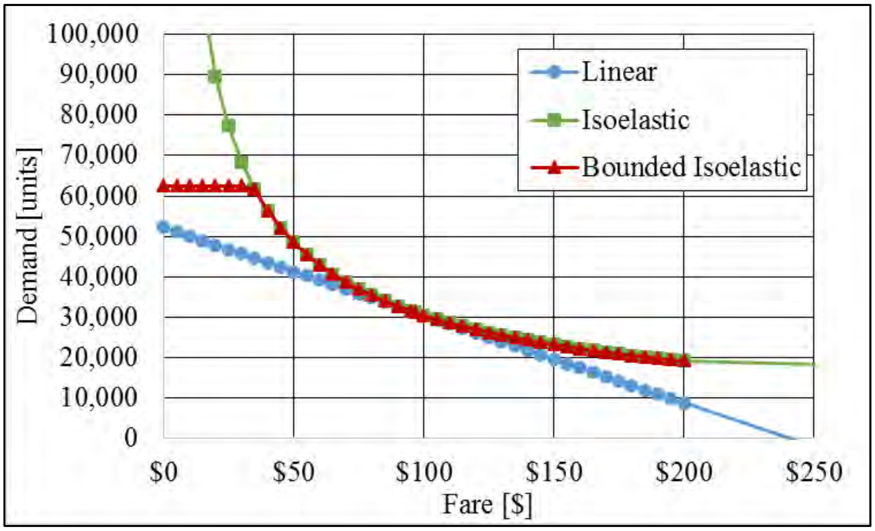


Figure 2-1 Comparison of linear, isoelastic, and bounded isoelastic demand functions (assuming demand only depends on fare) (Source: author)

2.3 Model Sensitivity to Inputs

This section analyzes the behavior of vertically-separated unregulated monopolistic intercity passenger TOs in the context of the NEC for different model inputs to understand the robustness of the results. We first compare the results obtained for different end users' demand functional forms and for different values of the elasticity. We then analyze the results obtained for different cost values.

We use data published by the TOs in their annual financial plan to determine the inputs of the model. According to (Amtrak, 2014) a TO like Amtrak faces fixed operational (direct) costs of $fc = \$102.5m$ per year ($fc = \$281k$ per day) and variable operational costs of $vc = \$1.25m$ per train and per year ($vc = \$3{,}425$ per train and per day). In 2013, Amtrak's average fare were equal to $f_0 = \$96.5$, the number of trains averaged $n = 150$ trains per day, with a realized demand of $d_0 = 11.4m$ passengers per year ($d_0 = 31{,}250$ passengers per day), and with an average train capacity of $c = 210$ passengers assuming a physical capacity of 250 seats with 85% load factor (Amtrak, 2014). No subsidies are required for the operations of intercity services in the NEC (Amtrak, 2010; Amtrak, 2012). We assume that end users' demand for traveling on a specific type of intercity service depends mostly on the fare (and not that much on the frequency or small variations on travel time). According to (Morrison, 1990) the elasticity of the demand of intercity passengers in the NEC to the fare is equal to $e = -0.67$. We also assume that the access charges depend linearly on the number of trains: $ac(n) = ac_f + ac_v \cdot n$.

2.3.1 Bounded Isoelastic Demand Function

In this case we assume that the end users' demand for rail services is a bounded: isoelastic curve that depends on the fare charged by the TO: $d(f) = \min(\bar{d}, kf^e), f \leq \bar{f}$. The value of k is equal to $k = d_0 f_0^{-e}$ for $f < \bar{f}$. $\bar{f}, \bar{d}$ are, respectively, the maximum fare and the maximum expected demand.

Equation (2.3) can be used to determine the number of trains that the TOs would like to operate and the fare charged to end users given the access charges. From an analytic standpoint, it is easier to determine those numbers assuming that the number of trains that the TOs would like to operate is continuous (equal to $n(f) = {d(f)}/{c}$) and that the demand is isoelastic. In that case, if $-1 < e \leq 0$, then the optimal fare f^{*c} is unbounded (the optimal solution is to charge end users as much as possible), and if $e \leq -1$, then the optimal fare is $f^{*c} = \frac{vc+ac_v}{c}\frac{e}{e+1}$.

In practice, though, the TOs can only operate an integer number of trains and the demand is bounded by $\bar{d}$. Considering this, the optimal number of trains that the TOs could operate is:

$$\begin{array}{lll} n^* = min\left(\left\lceil\frac{d(f^{*C})}{c}\right\rceil, \left\lceil\frac{\bar{d}}{c}\right\rceil\right) & if & \frac{vc+ac_v}{fc} \leq \frac{d(f^{*C})}{c} - \left\lfloor\frac{d(f^{*C})}{c}\right\rfloor \\ n^* = min\left(\left\lfloor\frac{d(f^{*C})}{c}\right\rfloor, \left\lceil\frac{\bar{d}}{c}\right\rceil\right) & if & \frac{d(f^{*C})}{c} - \left\lfloor\frac{d(f^{*C})}{c}\right\rfloor < \frac{vc+ac_v}{fc} \leq 1 \\ n^* = 0 & if & 1 < \frac{vc+ac_v}{fc} \end{array} \tag{2.6}$$

Note that $\lceil x \rceil$ and $\lfloor x \rfloor$ represent the closest integer number over or under x respectively. The TO would decide whether to have some excess capacity or some unmet demand depending on the amount of exceeding capacity or unmet demand and the relation between variable costs and revenues. The TO will not operate any trains if variable costs are larger than the revenues. Considering this, the optimal fare (f^*) would be:

$$f^* = \min(d_0{}^{-\frac{1}{e}} f_0, \bar{f}) \quad if \quad -1 < e \leq 0$$
$$f^* = \min(\left(\frac{n^* c}{k}\right)^{\frac{1}{e}}, \bar{f}) \quad if \qquad e \leq -1 \tag{2.7}$$

Equation (2.7) shows that for elasticity values between -1 and 0 the TOs are better-off increasing the fare as much as they can, i.e. up to $\bar{f}$ in this case.

Figure 2-2 shows the fare and number of trains that a TO would schedule as a function of the elasticity assuming $\bar{f} = \$200, \bar{d} = 62{,}500$ (double the current demand). In this case, the fares range from \$97 to \$200, and the number of trains from 68 to 129 trains per day when the IM does not charge any variable access charge per train. Note that both the fares and the number of trains to schedule are fairly robust when elasticity changes around $e = -0.67$: the fares change less than 1% and the number of trains change less than 10% for changes in elasticity of $\pm 20\%$.

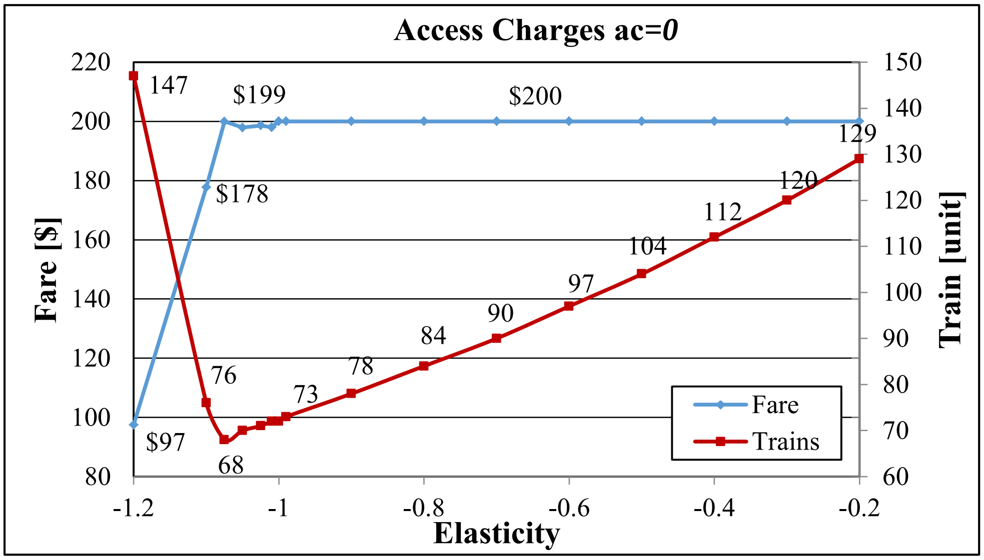


Figure 2-2 Fare and number of trains to be scheduled by a TO as a function of the elasticity assuming 0 access charges, maximum fares of \$200, and a cost and revenue structure similar to Amtrak (Source: author)

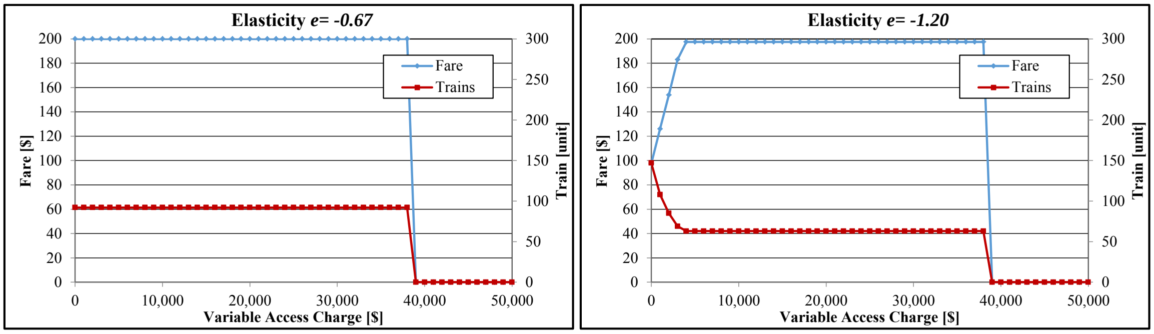


Figure 2-3 Fare and number of trains to be scheduled by a TO as a function of the variable access charges for different elasticity values assuming a cost and revenue structure similar to Amtrak

(Source: author)

As Figure 2-3 shows, the number of trains that the TO operates is robust for a large range of access-charge values too. When the elasticity values $e \in (-1,0)$, the number of trains that the TO would operate (92 trains in this case for $e = -0.67$) does not change when access charges increase (unless the TO has no incentive to operate trains and would then operate 0 trains). For elasticity values $e < -1$, the number of trains would decrease as the variable access charges increase from \$0 to \$4,000 and does not change afterwards (again, unless the TO has no incentive to operate trains and would then operate 0 trains).

We can use equations (2.4) and (2.5) to determine the maximum variable charge that a TO would be able to pay to access the infrastructure. This maximum access charge is equal to $ac_v(n = 0) = \$39{,}000$ independently of the elasticity. If access charges increase over that value the TO would not operate any trains (see Figures 2-3, 2-4, and 2-5). The maximum variable charges that TOs can sustainably pay in the medium term, i.e., the maximum variable access charges for which the TO has no profits neither losses after paying for capital at an adequate rate of return are $ac_v(\pi = 0) = ac_v(n = 0) - {}^{fc}/_{n^*}$, i.e., up to \$35,000 for an elasticity value of $e = -0.67$ (Figure 2-4) and up to \$34,000 for an elasticity value of $e = -1.2$ (Figure 2-5).

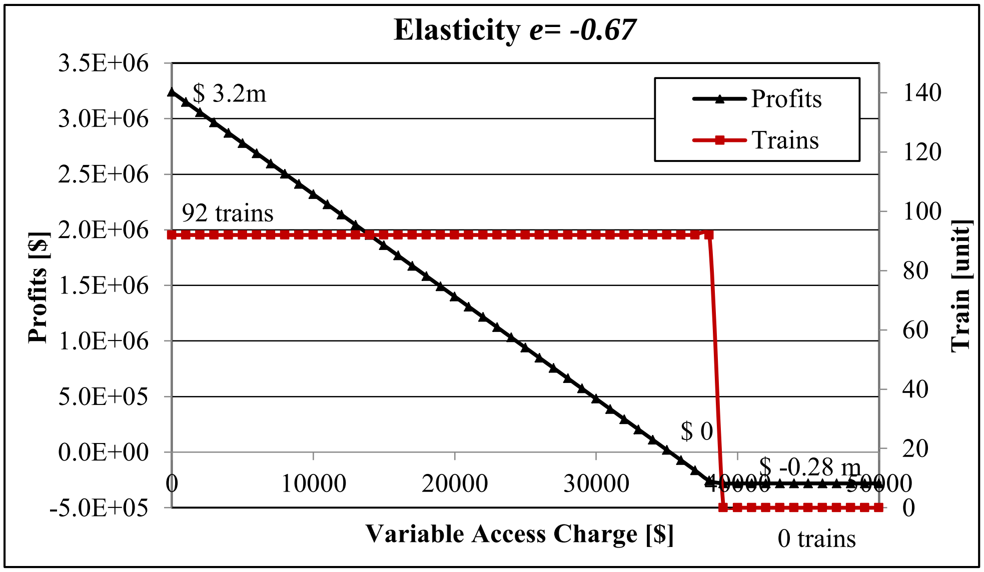


Figure 2-4 Profits and number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -0.67 assuming a cost and revenue structure similar to Amtrak

(Source: author)

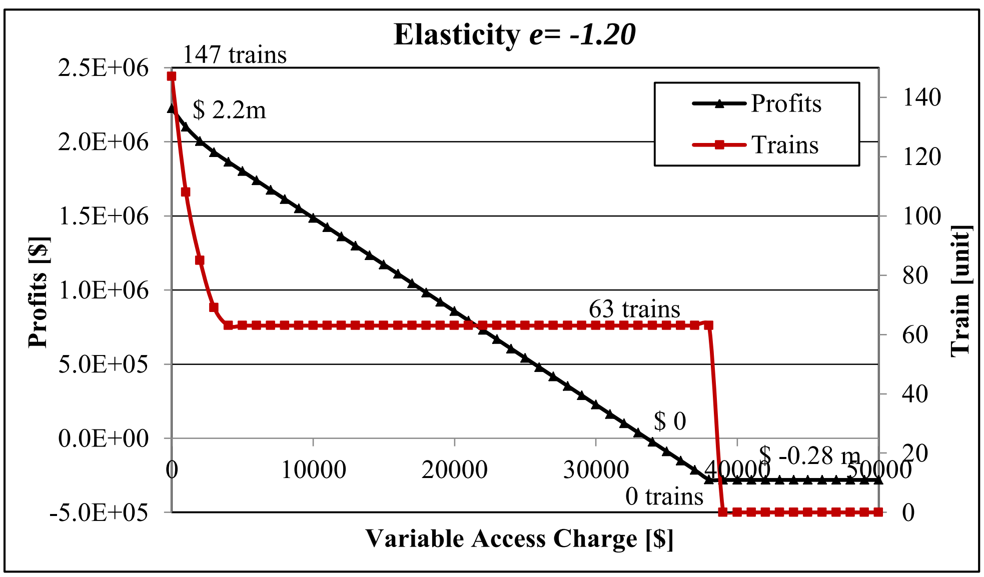


Figure 2-5 Profits and number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -1.2 assuming a cost and revenue structure similar to Amtrak

(Source: author)

The first point labeled in Figures 2-4 and 2-5, represents the total profits obtained when the TO does not pay any access charges. This point represents the maximum possible profits that the TO could ever achieve in the market: $\$3.2m$ or $\$2.2m$ (depending on the elasticity value). This point also allows us to determine the maximum fixed access charges that the TO can afford. We know that a TO would never accept losses over $\$281{,}000 \approx \$0.3m$ (its fixed costs). As a result, the maximum fixed access charges that the TO can pay are respectively $\$3.5m$ ($\$3.2m + \$0.3m$) or $\$2.5m$ ($\$2.2m + \$0.3m$). Note that the TO can only be sustainable in the medium term if total fixed access charges are smaller than $\$3.2m$ or $\$2.2m$ respectively (so it can have $\$0$ profits).

The second point labeled corresponds to the maximum variable access charges that the TO can sustainably afford: $\$35{,}000$ per train per day or $\$34{,}000$ per train per day respectively. At that point, the TO would operate 92 or 63 trains (depending on the elasticity value assumed). The final point corresponds to the variable access charges ($\$39{,}000$ per train per day in both cases) for which the TO would not have any incentives to operate trains (since operating a train would increase the burden of the debt).

If the demand increases year by year, the number of trains that operators want to schedule would increase by the same rate. However, the fares would not change if the elasticity of the demand to the fare is lower than -1.

2.3.2 Linear Demand Function

This case analyzes how the results presented in the previous cases would change when we assume that the demand is a linear function of the fare: $d(f) = e \cdot \frac{d_0}{f_0} \cdot f + (1-e) \cdot d_0$ (to ensure that elasticity is $e = \frac{\Delta d / d_0}{\Delta f / f_0}$).

The optimal number of trains and fare (n^*, f^*) to maximize profits can also be determined using equation (2.3). The results show that the number of trains and the fares to maximize profits are either:

$$n^* = \left\lceil \frac{(1-e)d_0}{2c} \right\rceil, f^* = \frac{(e-1)f_0}{2e},$$

$$n^* = \frac{(1-e)d_0}{2c} + \frac{ed_0(vc+ac_v)}{2c^2 f_0}, f^* = \frac{vc+ac_v}{2c} + \frac{(e-1)f_0}{2e}, \text{ or} \quad (2.8)$$

$$n^* = 0, f^* = 0$$

Equation (2.8) implies that when variable costs are small with respect to the fares that end users can afford, the optimal solution is to maximize revenues and offer the minimum number of trains that allow serving all the demand for the optimal fare. However, when variable costs are comparable to the fares that end users can afford, the optimal solution is a trade-off between maximizing revenues and covering variable costs. In this case, the capacity should be optimized in such a way that most demand is served without providing excess train capacity. Finally, in those cases in which the end users cannot viably accept a fare level that allows TOs to cover at least the variable costs, the TO should not operate any train.

We also illustrate these results with an example inspired in the Amtrak intercity services of the NEC. Figure 2-6 shows the fare and number of trains that a TO would schedule as a function of the elasticity. In this case, the fares range from \$97 to \$298, and the number of trains from 87 to 149 trains per day when the IM does not charge any variable access charge per train. Again, for changes in elasticity of $\pm 15\%$, the number of trains changes less than 5%, and the fares change less than 10%. When we compare these numbers with the ones obtained in Section 2.3.1 we see that the numbers of trains in this case differ less than 20% with respect to the bounded isoelastic demand. The fares however vary up to 50% with respect to the ones obtained in the bounded isoelastic demand case.

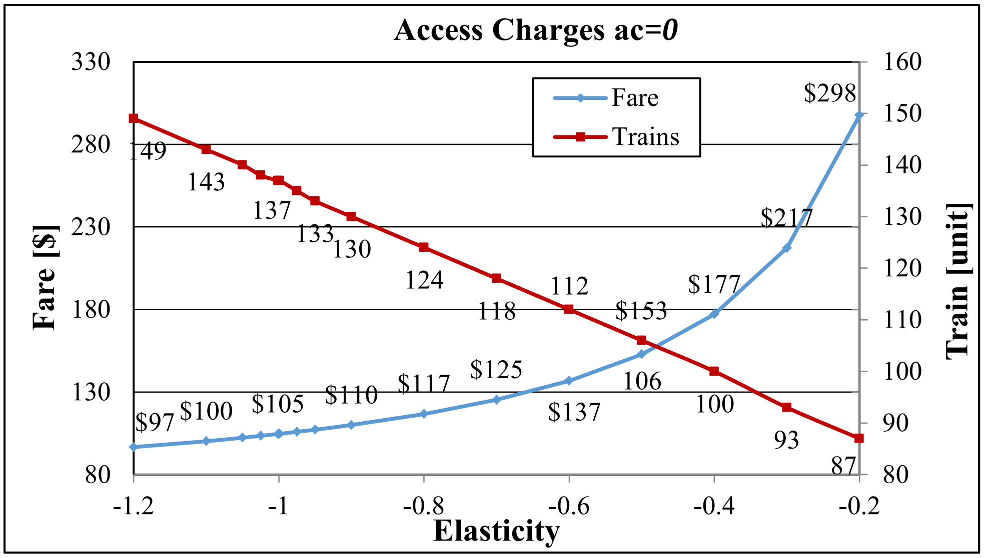


Figure 2-6 Fare and number of trains to be scheduled by a TO as a function of the elasticity assuming 0 access charges and a cost and revenue structure similar to Amtrak (Source: author)

Figure 2-7 represents the expected profits of the TO and optimal number of trains for different values of the variable access charges. In this case, the optimal operational decision is to operate 116 trains per day with fares on the order of $128 if access charges are low. If the access charges increase over $32,000 per train then the profit maximizing strategy suggest operating between 37 and 0 trains. The TO would not operate any trains if the access charges are over $47,000 per train. The access charges willingness to pay is within 20% of the value obtained in Section 2.3.1. The number of trains is also within 20% of the values obtained in Section 2.3.1 for most access charges. The fares, however, vary considerably (over 35%) with respect to the bounded isoelastic demand case.

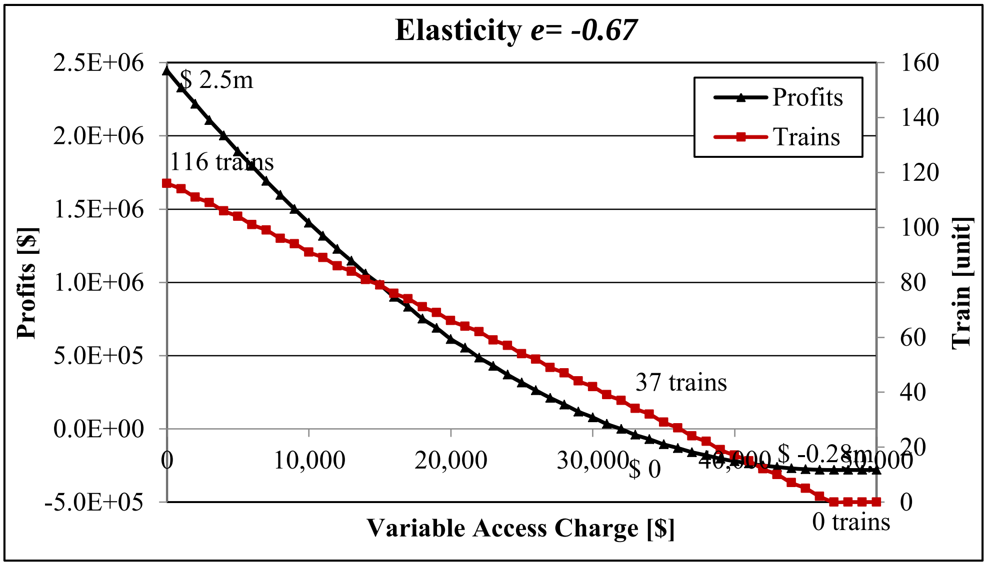


Figure 2-7 Profits and number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -0.67 assuming a cost and revenue structure similar to Amtrak

(Source: author)

2.3.3 Cost Sensitivity Analysis

This case analyzes how the results presented in the previous cases would change for different fixed-cost and variable-cost values. The first finding in this section is that the number of trains and the fares do not change when fixed costs change. This result makes sense for two reasons. First, once the TO incurs in the fixed cost, fixed costs are perceived as sunk costs and although they affect TO profit, they do not affect TO operational decisions. Second, and related to the previous point, fixed costs do not appear in equations (2.6), (2.7), and (2.8).

Our second finding is that the results are also very robust to changes in the variable access charges. As mentioned above, current Amtrak's variable costs are equal to $vc = \$3{,}425$ per train and per day (Amtrak, 2014). Figure 2-8 shows that the fares and the number of trains do not vary for different variable access charges unless access charges increase 1,000%. In that case, variable costs would be so high that the TO would prefer not to operate trains.

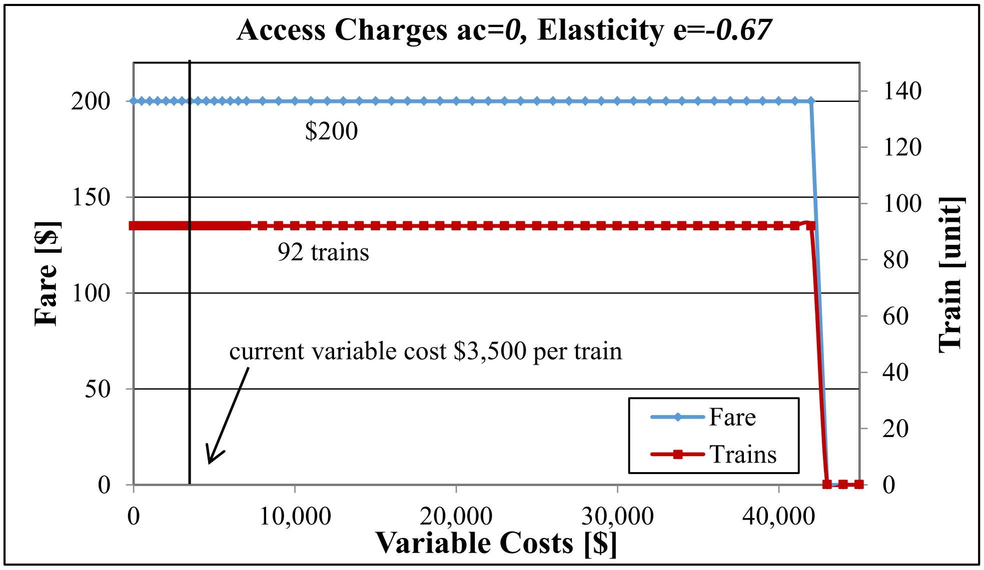


Figure 2-8 Fare and number of trains to be scheduled by a TO as a function of the variable costs assuming 0 access charges and a cost and revenue structure similar to Amtrak (Source: author)

Figures 2-9 shows that the fares vary less than 8% and the number of trains vary less than 7% when variable costs change ±100%.

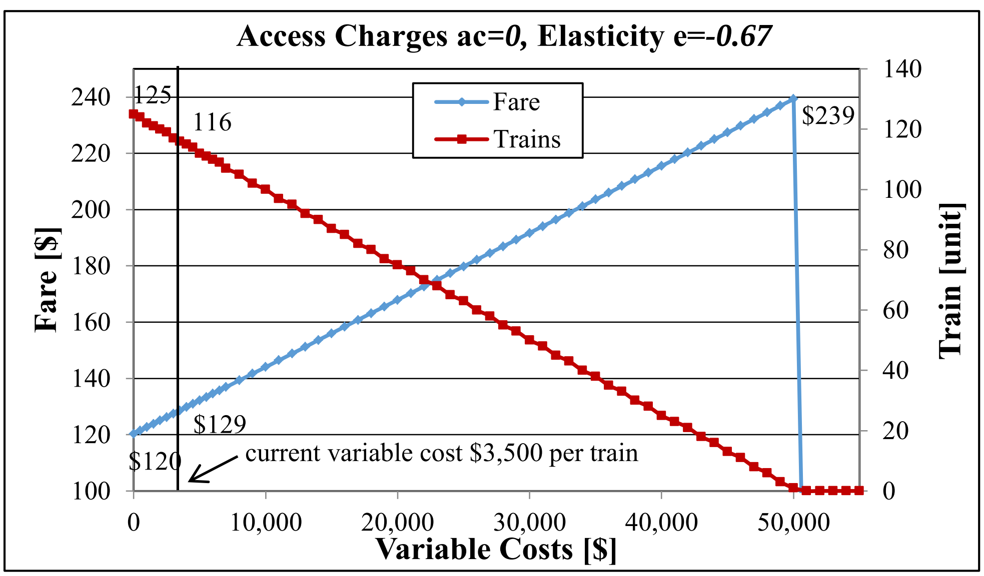


Figure 2-9 Fare and number of trains to be scheduled by a TO as a function of the variable costs assuming 0 access charges and a cost and revenue structure similar to Amtrak (Source: author)

In summary, these results show that the TO demand for scheduling trains and the TO ability to pay to access the infrastructure are very robust to model inputs. In other words, the TO demand for infrastructure use does not change much with small changes in the inputs of the model (cost and demand estimates). This suggests that the level of detail of the model is adequate to capture the interactions between the TOs and the IM. The results in Chapter 5 predicts differences in the number of services scheduled under alternative mechanisms that are over 20%. The robustness results shown in this chapter demonstrate that the differences obtained are significant (higher than the variances obtained in the sensitivity analysis).

The passenger fares or freight shipping rates charged to the end users are not as robust. As a result, more detailed information may be required to study in detail how TOs set fares and shipping rates. Although we use the fares to compute TO profits, this thesis does not focus on the interactions between TOs and end users.

2.4 Model Results and Implications

Section 2.3 shows that the TO Model estimates are robust to a broad set of model inputs. In this section, we use the TO Model to the behavior of TOs in different cases. We propose the study of TOs in both vertically-separated and vertically-integrated railway systems, within unregulated and regulated markets. Although each TO competes with other TOs to access the infrastructure, we distinguish whether each TO competes or not with other TOs to offer the same railway services to the end users (e.g., high-speed rail). Note that within the same vertically-separated railway system, different TOs providing different services may face different regulatory and competitive schemes. For instance, in shared systems like the NEC, one could argue that a private intercity passenger TO is best modeled as an unregulated monopoly, the commuter TOs are best modeled as regulated monopolies, and the freight TOs are best modeled as competitive TOs.

2.4.1 Vertically-Separated Unregulated Monopolistic Train Operator

This case assumes vertical separation between TO and the IM. It also assumes that although the TO may compete with other types of TOs to access the infrastructure, it is the only TO who provides a specific type of railway service to the end users. That is, the TO is a monopoly in its railway service market and there are no substitute services offered by other TOs. Unregulated refers to the fact that except for the constraints imposed by the capacity pricing and allocation mechanism in place, the TO has full control over the fares and the number of trains. Following the discussion that leads to equation (2.3), TOs in this case would determine its operational decisions with the objective of maximizing its profits given the access charges. Their willingness to pay to access the infrastructure is given by equations (2.4) and (2.5).

The two cases in this section analyze how an intercity passenger TO would respond to capacity pricing and allocation mechanisms. We assume here that end users' demand for traveling on a specific type of intercity service depends mostly on the fare (and not that much on the frequency or small variations on travel time). We refer the reader to Sections 2.3.1 and 2.3.2 for a discussion of the TO behavior when the demand is modeled as a bounded isoelastic function and a linear function respectively. Section 2.4.1.1 discusses the results obtained when the TO implement a revenue management mechanism. This case assumes that that the end users' demand is a bounded isoelastic function. Section 2.4.1.2 compares the results obtained in the three vertically-separated unregulated monopolistic intercity TO. We assume again that there are no operations subsidies and that the access charges depend linearly on the number of trains: $ac(n) = ac_f + ac_v \cdot n$.

2.4.1.1 Intercity Passenger TO using Revenue Management Mechanisms with Bounded Isoelastic Demand Function

This case proposes to study the behavior of TO assuming that the demand as a function of a fare behaves as a bounded isoelastic curve, that is, $d(f) = \min(\bar{d}, kf^e)$, for $f < \bar{f}$. This time we assume that the TO has a perfect revenue management mechanism in place that allows it to charge each end user the maximum fare that they are willing to pay.

In this case, the optimal number of trains (n^*) to maximize profits can be determined using equation (2.3). The results obtained show the number of trains to maximize profits is (depending on how the maximum fare compares to the variable costs):

$$\begin{aligned} &n^* = k \cdot (vc + ac_v)^e \cdot c^{-1-e} \quad if\ \overline{f} \cdot c > vc + ac_v \\ &n^* = k \cdot \overline{f}^e c^{-1}\ or\ n^* = 0 \qquad otherwise \end{aligned} \tag{2.9}$$

If the result is not an integer, it should be rounded to the immediate lower or upper integer (the one that maximizes profits). Figures 2-10 and 2-11 show the number of trains and the TO profits for different elasticity values with no access charges, and the number of trains and the TO profits for different variable access charges and elasticity value equal to -0.67 using the same Amtrak-inspired example used in the previous cases.

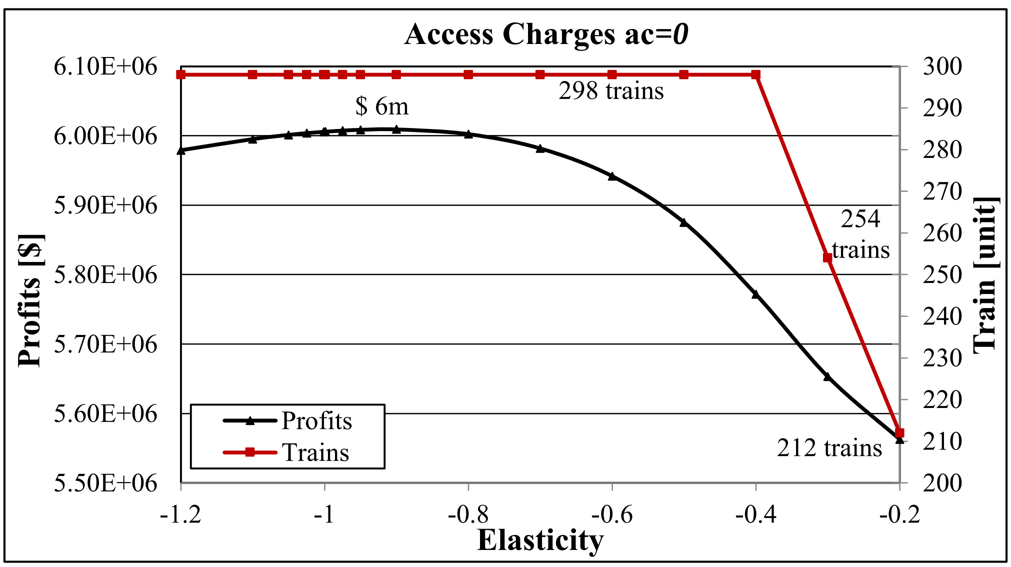


Figure 2-10 TO profits and number of trains to be scheduled by a TO as a function of the elasticity assuming 0 access charges and a cost and revenue structure similar to Amtrak (Source: author)

In this case, for elasticity values over -0.4, the maximum number of trains that the TOs are willing to operate when there are no access charges is driven by the maximum demand that TOs expect.

When the access charges increase, the operators have incentives to operate fewer trains. This case shows two important results: the maximum variable access charges that a TO like Amtrak (with elasticity value -0.67) would be able to pay are below $\$39k$ per train. Variable access charges should not be above $\$35k$ to ensure sustainable TO operations in the medium term. With these level of variable access charges TOs would still operate 92 to 95 trains.

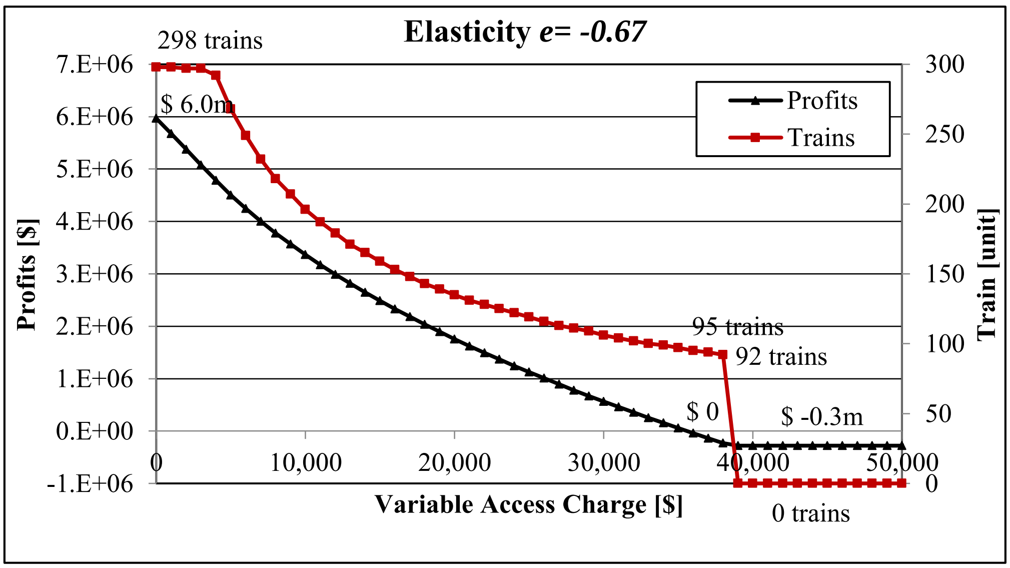


Figure 2-11 Profits and number of trains to be scheduled by a TO as a function of the variable access charges for elasticity equal to -0.67 assuming a cost and revenue structure similar to Amtrak

(Source: author)

2.4.1.2 Result Comparison

In this section we summarize and compare the main results obtained for the different cases of vertically-separated, unregulated, monopolistic intercity passenger TOs.

Figures 2-12 and 2-13 compare the number of trains that a TO like Amtrak would operate as a function of the variable access charges for elasticity values of -0.67 (very inelastic demand;

as Morrison, 1990 characterized the NEC intercity passenger demand), and -1.2 (elastic demand). We also compare the optimal number of trains obtained with this TO model with the current number of trains operated by Amtrak in the corridor.

First, the results suggest that even if the TOs have a perfect revenue management mechanism in place, the number of trains operated would decrease with respect to the current number of trains operated if variable access charges exceed $\$10{,}000$ to $\$15{,}000$ per train.

Second, the results show that for variable access charges above $\$35{,}000$ per train, the number of trains would importantly decrease over 85% with respect to current number of trains independently of the demand functional form assumed and fare mechanism in place.

Third, the TO always operates more trains than in the case with revenue management as compared to the cases without revenue management. This makes sense since the utilization of revenue management mechanisms allows the TOs to charge different fares to end users. As a result, the TO has no need to reduce the number of trains to be able to increase fares.

Finally, note again that the TO ability to pay to access the infrastructure is very robust to different model assumptions. The value of the maximum access charges over which the TO would not operate any trains varies less than 20% across these cases. This is very important, because the operations of a TO that implements revenue management mechanisms is very different that the operations of a TO that does not.

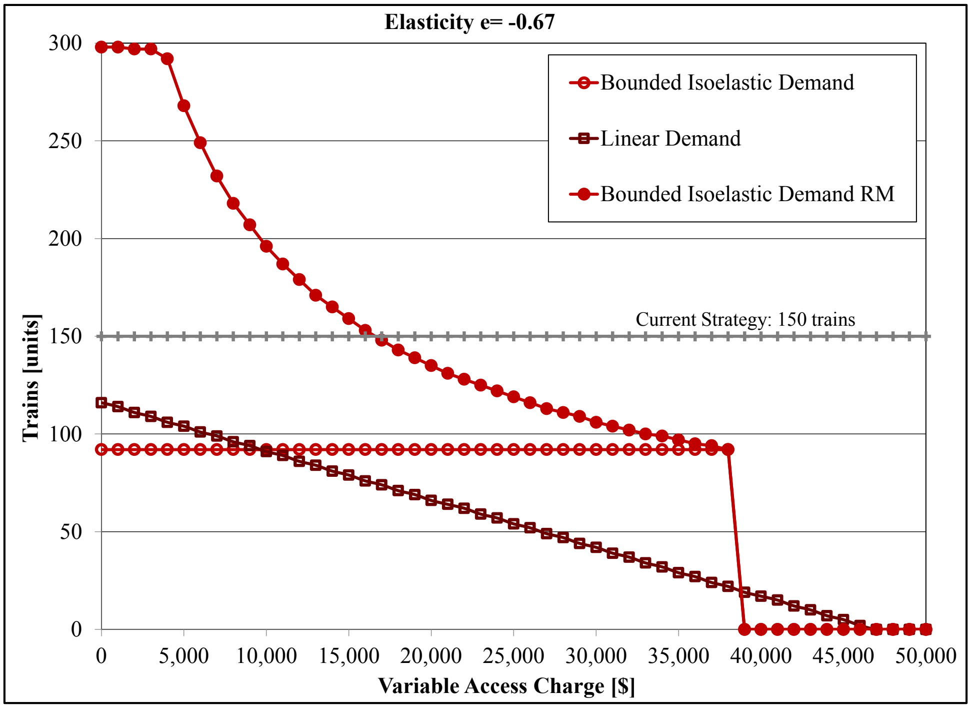


Figure 2-12 Number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -0.67 for different demand functions and fare mechanisms (Source: author)

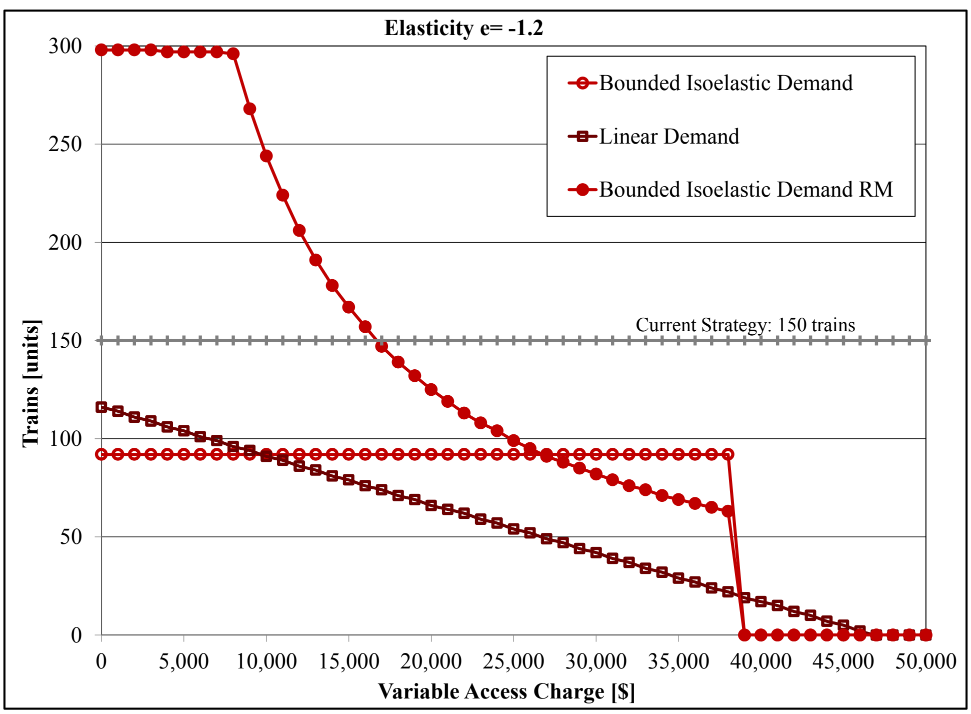


Figure 2-13 Number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -1.2 for different demand functions and fare mechanisms (Source: author)

2.4.2 Vertically-Separated Regulated Monopolistic Train Operator

These cases assume again vertical separation between TO and the IM. We assume again that the TO is a monopoly in its railway service market and there are no substitute services offered by other TOs. Unlike the previous cases, these cases assume that TO operations are regulated to ensure that the TOs do not charge excessive fares to the end users. We analyze two main cases, one representative of an intercity passenger TO and one representative of a commuter TO.

2.4.2.1 Intercity Passenger TO

Section 2.3.1 analyzes the behavior of intercity passenger monopolistic unregulated TOs that determine the fares and number of trains with the objective of maximizing their profits (see equation (2.3)). In this case however we assume that there is an even-handed regulator that ensures that the TO optimizes service without extracting excessive rents from the end users. That is, we assume that the regulator tries to ensure that the TOs have zero profits (after reimbursing capital at an adequate rate of return). Although this regulation prevents the TOs from making large profits (as in the cases presented in Sections 2.3 and 2.4.1), it also ensures that the access charges that the TO pays to the IM are not excessive (i.e., profits will never be negative in this case, unlike what could happen in the cases presented in Sections 2.3 and 2.4.1). We assume from now on, that intercity passenger demand is a bounded isoelastic function on the fare. As a result, equation (2.3) has to be adjusted in this case.

$$\max_{N,F} N,$$

$$\text{s.t. } s(F,N) + F \cdot \min(d(F), c \cdot N) - fc - vc \cdot N - AC(N) = 0 \qquad (2.10)$$

Assuming again that there are no subsidies and that access charges are linear, we first see that the average fare charged to the end users in this case has to be equal to the total costs divided by the total number of travelers. The optimization problem can then be solved by finding the maximum number of trains (n^*) for which the TO can recover costs.

Figures 2-14 and 2-15 show the solution of the optimization problem for different values of variable access charges and for two demand elasticity values (-0.67 and -1.2) assuming that no revenue management mechanism is in place. Figure 2-14 shows the number of trains that a regulated and an unregulated TO would operate and Figure 2-15 shows the fares that they would charge to the end users.

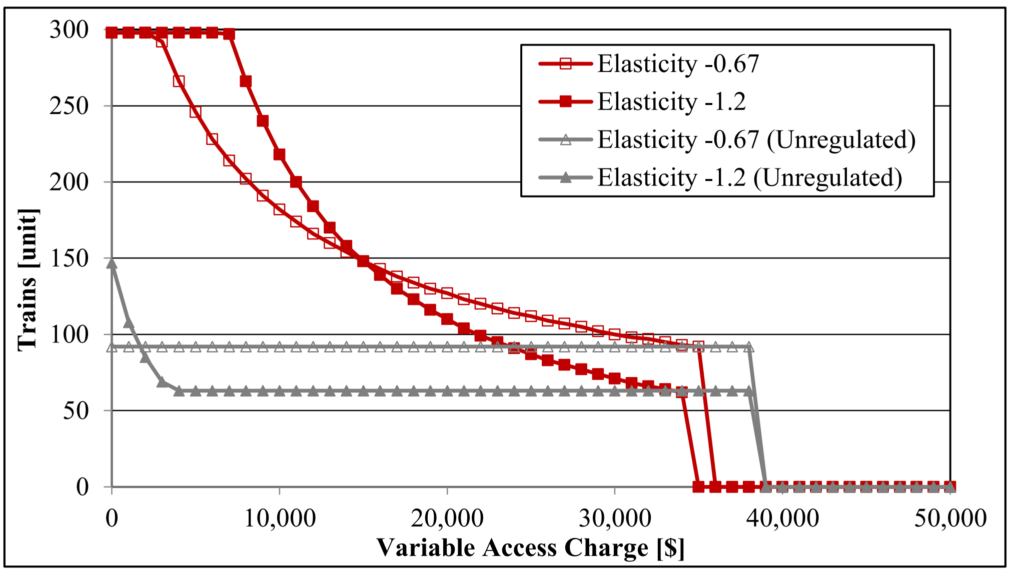


Figure 2-14 Number of trains to be scheduled by a (regulated vs. unregulated) TO as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)

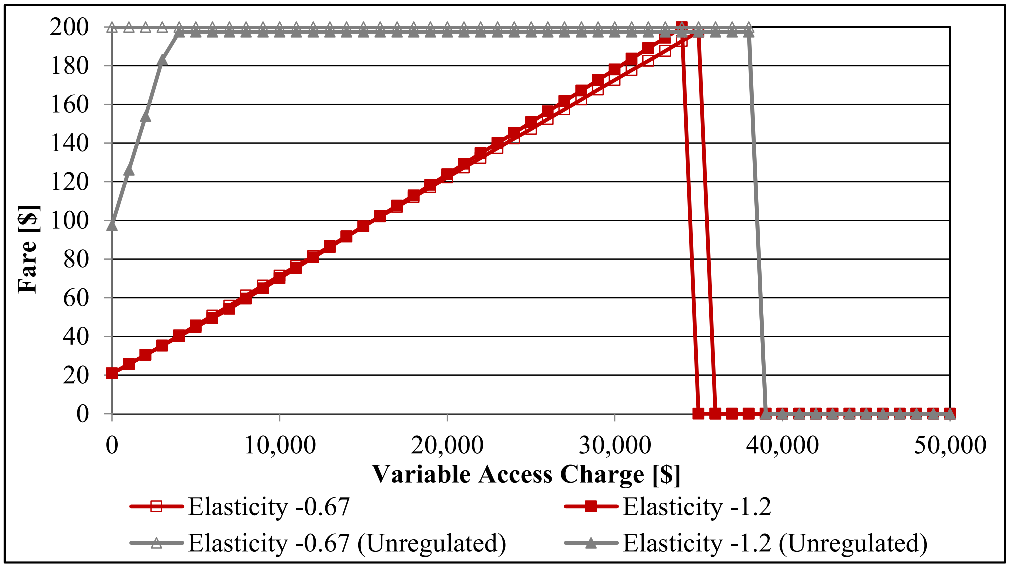


Figure 2-15 Fares charged to the end users by a (regulated vs. unregulated) TO as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)

There are two important take-away messages from these figures. First of all, as we may expect, regulated train operators operate more services and charge lower fares to the end users. The only exception comes when the variable access charges are so high that the TO is not able to sustainably operate any longer. Second, the demand for scheduling trains of a regulated intercity passenger TO in the NEC is higher than the current number of trains for variable access charges up to $15,000 per train.

2.4.2.2 Commuter TO

This case is inspired by the commuter TOs in the NEC and is different from the previous cases in two senses. First, it assumes that the TOs are regulated by imposing a fare limit. In other words, the regulator does not allow the TOs to change the fares they charge to the end users. Second, it assumes that the demand depends on the frequency of the service. According to (Lago et al., 1981), the demand for commuter services increases when the frequency of service increases and vice-versa.

Since the fare is fixed, the elasticity of the demand to the frequency can be defined as $e_n = \frac{\Delta d / d_0}{\Delta h / h_0}$ where h is the average headway between consecutive trains. Since the headway is proportional to $1/n$, the elasticity can also be computed as $e_n = -\frac{(d-d_0)\cdot n}{(n-n_0)\cdot d_0}$. Therefore, assuming a linear demand function on the frequency, the demand can be determined given the number of trains using $d(n) = (1-e_n)\cdot d_0 + \frac{e_n \cdot d_0 \cdot n_0}{n}$. We can also write $d(n) = \max\left(0, (1-e_n)\cdot d_0 + \frac{e_n \cdot d_0 \cdot n_0}{n}\right)$ to avoid negative demands for low number of services.

The optimal number of trains (n^*) to maximize profits can be determined using equation (2.3) considering that the fare is fixed in this case. Assuming again that access charges are linear ($ac(n) = ac_f + ac_v \cdot n$) and that the subsidy is a lump sum (paid by the commuter agency to the TO), we can determine that the optimal number of trains would be either:

$$n^* = \sqrt{-\frac{f\cdot e_n \cdot d_0 \cdot n_0}{vc + ac_v}},$$

$$n^* = \frac{(1-e_n)\cdot d_0}{2\cdot c} - \frac{\sqrt{(1-e_n)^2 \cdot d_0^2 - 4\cdot c \cdot e_n \cdot d_0 \cdot n_0}}{2c}, \text{ or} \tag{2.11}$$

$$n^* = 0,\ n^* = \frac{(1-e_n)\cdot d_0}{c}$$

The choice of one number of trains over other would depend on how revenues and cost compare. If revenues obtained from fares are much higher than variable costs, then the optimal strategy to maximize profit would be to maximize revenues. If revenues are comparable to variable costs, the optimal strategy would be to ensure that there is no excess-capacity on the trains. Finally, if variable costs are much higher than the revenues per train, the TO should not operate any train.

Note that this number of trains is independent of the level of subsides and the fixed costs (from operations and access-charges). These values would only affect whether the TO can sustainably operate in the medium term.

As we mentioned before, this case is representative of the situation of the commuter rail TOs in the NEC. According to (MBTA, 2013a; MBTA, 2013b) a TO like the MBTA, the commuter operator in the Boston area, faces fixed operational (direct) costs of $fc = \$435.1k$ per day and variable operational costs of $vc = \$1{,}666$ per train and per day. The elasticity of the demand with respect to the headway (frequency) is estimated by (Lago et al., 1981) to be equal to $e = -0.41$. As expected, the elasticity is negative because demand increases when headway decreases (i.e., the number of services increases). In 2014, MBTA's average fare ranged from $f_0 = \$7 - \$25$ (we assume an average fare of $f_0 = \$13$), the number of trains averaged $n_0 = 485$ trains per day, with a realized demand of $d_0 = 130.6k$ passengers per day. The train average capacity considered is $c = 350$ passengers, with 80% + load factor. Subsidies $s = \$234k$ per day are considered following (MBTA, 2013a).

Figure 2-16 compares current MBTA profits with the expected profits when the profit maximizing strategy presented in equation 16 is used to determine the number of trains. The results show that higher profits can be unlocked by reducing the number of services, especially when variable costs increase due to access charges. Note that even under the profit maximizing strategy, the TO would not be able to operate if access charges exceed \$2,500 per train per day, since the variable costs of operating the train would be higher than the revenues obtained. In that case, operating a train would only increase the cost burden for the system.

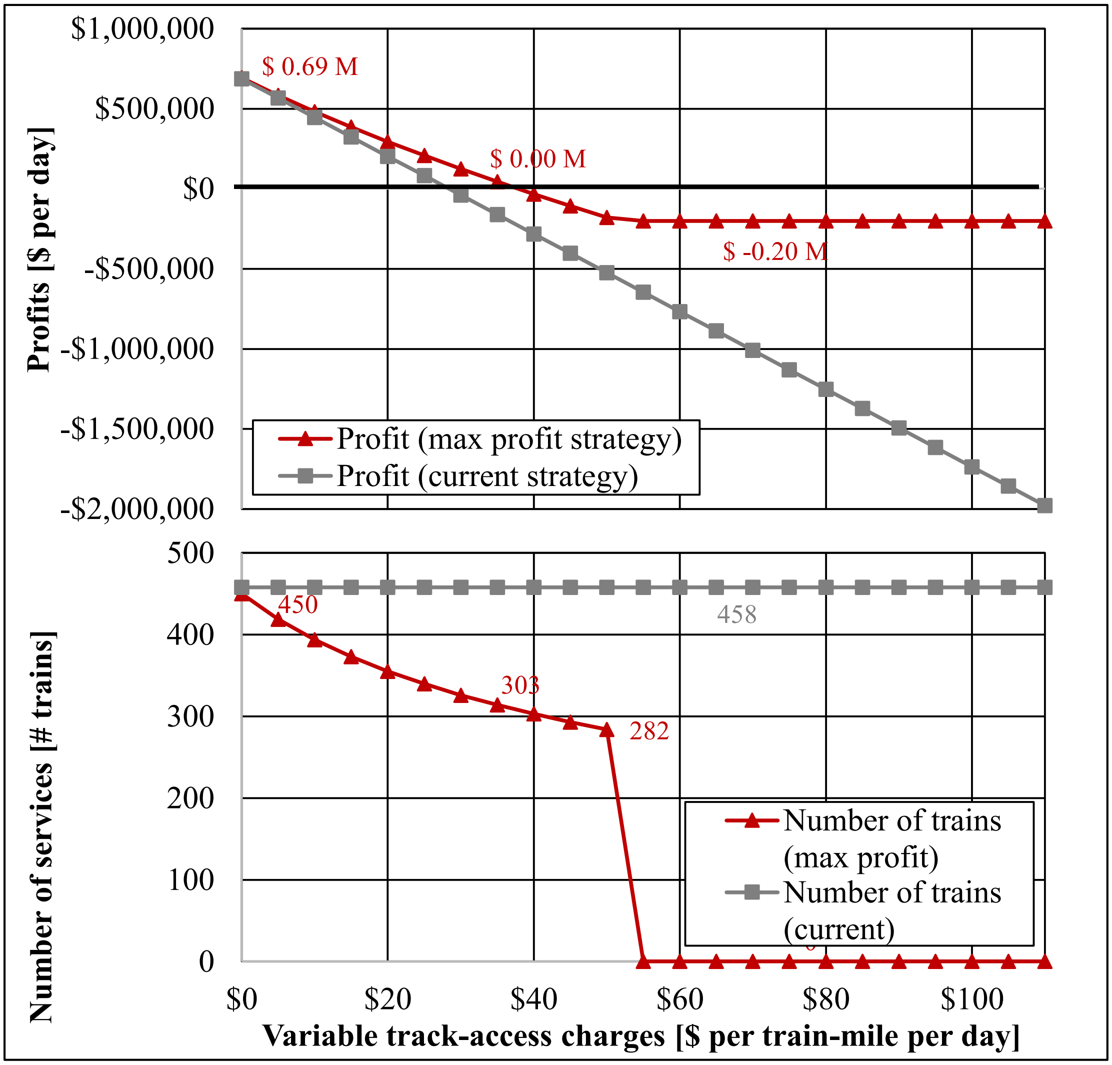


Figure 2-16 MBTA's expected profits and number of trains for different variable access charges

(Source: author)

This analysis shows that under fare limits, a commuter TO similar to the MBTA would be able to make positive profits (after adding the operations subsidy) when variable access charges are low. As a result, a regulation that ensures that the TO always has zero profits would be preferable from the end users' standpoint. However, regulating the end users' fare is much easier for the regulator than ensuring that the TOs have no profits; what justifies the use of these types of measures when the room to gain profits by the TOs are low.

We also want to compare the variable access charges that a commuter TO could afford with the variable access charges than an intercity passenger TO could afford in the context of the NEC. The previous analysis show that the maximum variable access charges that a commuter TO could afford range between $1,500 and $2,800. We have also shown that the maximum variable access charges that an intercity passenger TO could afford up to $30,000 to $40,000. Our analysis shows that with variable access of $15,000 to $20,000, the number of trains offered by an intercity passenger TO would not be dramatically affected. So these results suggest that intercity TOs in the NEC are able to pay around 10 times higher variable access charges than commuter TOs.

2.4.3 Vertically-Separated Perfectly Competitive Train Operator

These cases assume again vertical separation between TO and the IM. This section assumes that the TOs operate in a perfectly competitive railway service market. An important discussion in these cases is whether TOs may face fixed costs or not. From a theoretical standpoint, if the TOs face any type of fixed costs, it would be more efficient to have a single TO than multiple TOs; and as a result, the market would not be perfectly competitive. Consequently these cases assume no operational fixed costs and no fixed access charges.

The rest of this section presents the analysis for a perfectly competitive intercity passenger service market. Assuming again that the demand is a bounded isoelastic function on the fares, the number of trains and the fares (n^*, f^*) that the TOs would offer in equilibrium and perfect competition can be determined considering that: 1) the optimal number of trains in equilibrium would be the one at which no TO could be able to schedule an additional train without losing money; and 2) the optimal fares in equilibrium would be the ones at which the TOs would be indifferent between offering the services or not.

Conditions 1) and 2) can be formalized in equation (2.12):

$$\max_{N,F} N,$$

$$\text{s.t. } s(F,N) + F \cdot \min(d(F), c \cdot N) - vc \cdot N - AC_v \cdot N = 0 \tag{2.12}$$

Assuming no subsidies and a bounded isoelastic demand curve, we can analytically solve the problem stated in equation (2.12) to determine that n^*, f^* are either:

$$\begin{array}{ll} f^* = \frac{vc+ac_v}{c}, n^* = \left\lfloor \min\left(\frac{k \cdot (vc+ac_v)^e}{c^{e+1}}, \frac{\overline{d}}{c}\right) \right\rfloor & if\ vc + ac_v \leq c \cdot \overline{f} \\ f^* = 0, n^* = 0 & otherwise \end{array} \tag{2.13}$$

Figures 2-17 and 2-18 show the solution of the optimization problem in the Amtrak-inspired example for different values of variable access charges and for two demand elasticity values (-0.67 and -1.2). Figure 2-17 shows the number of trains that perfectly competitive TOs would operate and Figure 2-18 shows the fares that they would charge to the end users. These results are compared with the number of trains and fares of regulated and unregulated monopolistic TOs. The monopolistic TO's number of trains and fares are slightly different from the ones presented in Sections 2.3.1 and 2.4.2.1, because they assume no fixed costs or access charges for comparability purposes. In particular, the number of trains that a regulated monopolistic TO would operate is slightly higher than the one in Sections 2.3.1 and 2.4.2.1, and the fares charged are slightly lower (since there is no fixed cost to recover).

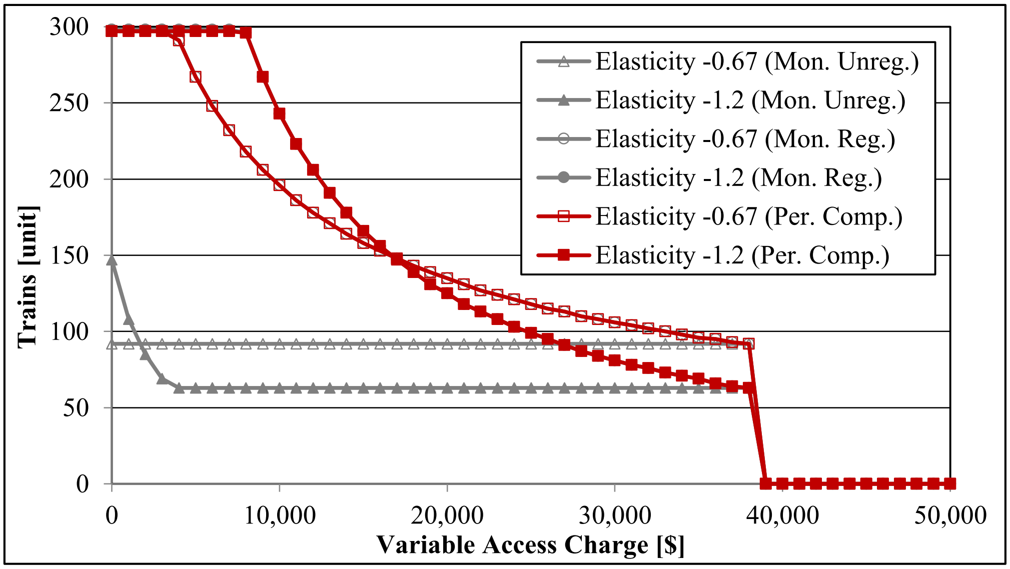


Figure 2-17 Number of trains to be scheduled by (perfectly competitive vs. monopolistic) TOs as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)

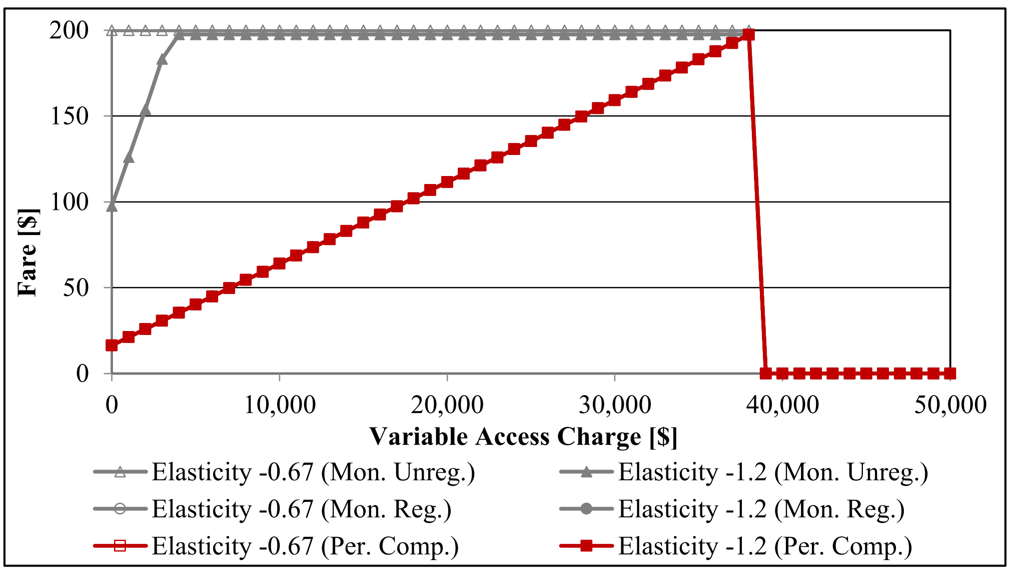


Figure 2-18 Fares charged to the end users by (perfectly competitive vs. monopolistic) TOs as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)

These results show that perfectly competitive TOs would perform very similarly to monopolistic regulated TOs, as we may expect. The levels of services and fares for these two cases completely overlap in Figures 2-17 and 2-18. The only differences between these cases come from the discrete nature of the number of trains. In some instances a regulated monopolistic

TO would be able to offer one more train than a perfectly competitive market with slightly higher fares and lower train utilization rates. Note however, that an unregulated operator would always operate fewer trains and charge higher fares to the end users. As we showed in Section 2.4.2., the demand for scheduling trains of both regulated intercity passenger TOs and perfectly competitive intercity TOs in the NEC is higher the current number of trains when variable access charges are lower than or equal to $15,000 per train per day. In these cases, the TOs have incentives to operate trains as long as variable access charges are smaller than $38,550 per train.

2.4.4 Vertically-Integrated Railway Company

These cases analyze a vertically-integrated railway company. We use the results of this section as a benchmark for the previous cases. Equations (2.3), (2.4), and (2.5) have to be adjusted to analyze vertically-integrated systems because 1) a vertically-integrated railway company also faces the infrastructure management costs; and 2) track access charges in this case are not necessary (transfer between the TO and the IM that cancels out in a vertically-integrated system).

This section considers two main cases: one where the railway company offers only one type of railway service and another one where the railway company offers two types of (non-substitutable) railway services.

2.4.4.1 Single Type of Service: Intercity Passenger Service

We assume again that the intercity passenger service face a bounded isoelastic demand as a function of the fare charged to the end users. Apart from the pure operations-related costs discussed in Section 2.2, an integrated railway company also faces infrastructure-related costs. At a high level, the infrastructure costs of a railway company can be aggregated in fixed and variable costs. For the context of this research, we consider fixed costs (fc_{IM}) all costs that do not change in the medium term with the number of trains operated on the infrastructure. We consider variable costs (vc_{IM}) all costs that depend on the number of trains operated on the infrastructure in the short or medium term horizon. As a first order approximation, we will assume that variable costs depend linearly on the number of trains operated.

Equation (2.3) could be adapted to this case by including these costs and eliminating the access charges (internal transfer between the TO and the IM):

$$\max_{N,F}[s(F,N) + F \cdot \min(d(F,N,tt), c \cdot N) - fc - vc \cdot N - fc_{IM} - vc_{IM} \cdot N] \tag{2.14}$$

Initially, the only difference between equations (2.3) and (2.14) is the fact that the access charges appear in place of the infrastructure related costs. This makes sense, because the objective of access charges is to pass the infrastructure costs on to the TO. As a result, if the access charges scheme just replicated the infrastructure cost scheme (i.e., $AC_f = fc_{IM}, AC_v = vc_{IM}$), there would be no differences between the operational decisions of a vertically-separated TO and the operational decisions of an integrated company. In other words, the vertical separation of the system would introduce no distortion in the operational incentives. Consequently, depending on the regulatory and competitive environment, the integrated railway company would exhibit the same type of behavior than the vertically-separated TOs discussed in Sections 2.3, 2.4.2, and 2.4.3. This finding is consistent with the findings of other network industries (Gomez-Ibanez, 2003; Laffont and Tirole, 1993; Laffont and Tirole, 2000; Perez-Arriaga, 2013). Not surprisingly, many countries have adopted infrastructure marginal costs to price infrastructure capacity (Texeira and Lopez-Pita, 2012; Texeira and Prodan, 2014).

If there are some differences though on the variable component (i.e., $AC_v \neq vc_{IM}$), then a vertically-separated TO would have incentives to operate more trains (if $AC_v < vc_{IM}$) or fewer trains (if $AC_v > vc_{IM}$) than an integrated railway company. Figure 2-19 shows how these differences would result in total utility losses.

Figures 2-19 and 2-20 show the social utility, and the end users, TO, and IM utilities respectively for an integrated railway system, and compared them to those of vertically-separated systems as a function of the variable access charges. We assume $fc_{IM} = \$1m, vc_{IM} = \$10{,}000$.

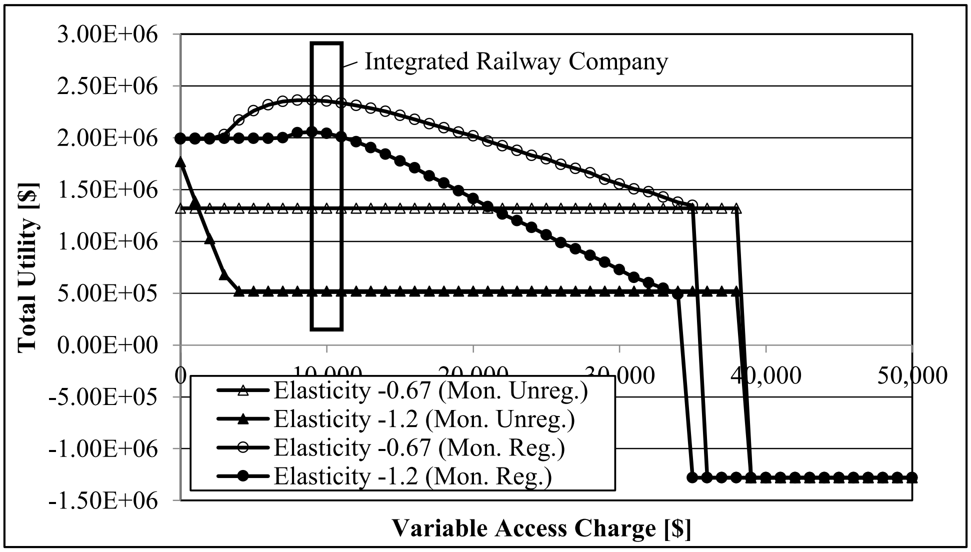


Figure 2-19 Total social utility associated with integrated and vertically-separated monopolistic (unregulated and regulated) TOs as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)

Note that the differences between variable infrastructure costs and variable access charges do not only result on different utility distributions between the TO, and the IM; but also in utility losses for the society at large. The regulatory framework and the competitive scheme also affect the utility. As Figures 2-19 and 2-20 show, the greediness of unregulated TOs does not only change the utility distribution between end users and TOs, but also imposes losses in the total social utility driven by the lost demand from reduced number of trains.

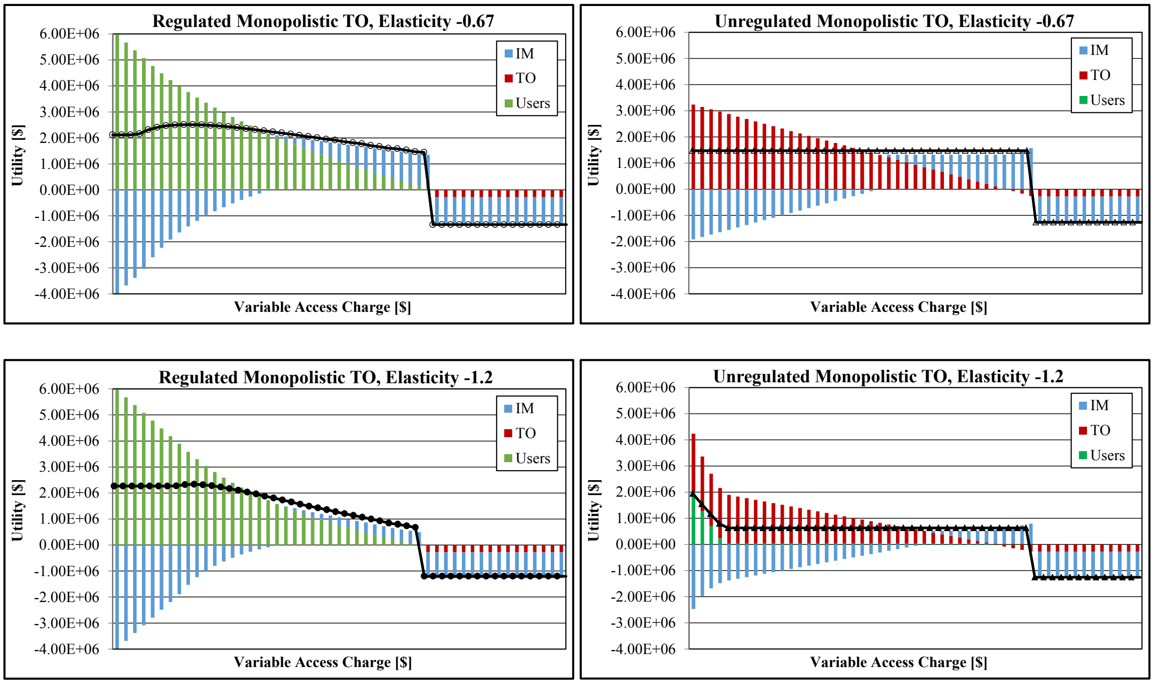


Figure 2-20 End users, TO, and IM's utilities for vertically-separated monopolistic (unregulated and regulated) system as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)

Any differences in the fixed cost component (i.e., $AC_f \neq fc_{IM}$) would mostly affect the distribution of utility between the TO and the IM (welfare transfers). These differences would not affect the operations of an unregulated monopolistic TO unless they drive the TO profits below their fixed costs. If this is the case, the unregulated monopolistic TO would no longer have incentives to offer any train service (and there would be losses on total social utility). These differences though would also affect the number of trains and fares that a regulated monopolistic TO would operate, slightly reducing social utility. These results are further discussed in the context of Tanzania in Chapter 4.

2.4.4.2 Several Types of Services

This case considers not only the infrastructure-related costs (fc_{IM}, vc_{IM}), but also the operations of different types of services i (such as intercity passenger train services, commuter services, or different types of freight services). We assume that in general, the subsidies, fares, number of trains, infrastructure and operations related variable costs of different types of services may be different (and we indicate that by adding the subscript i). Equation (2.3) could be adapted to this case:

$$\max_{N,F}[\sum_i (s_i(F_i, N_i) + F_i \cdot \min(d(F_i, N_i, tt_i), c_i \cdot N_i) - vc_i \cdot N_i - vc_{IMi} \cdot N_i) - fc - fc_{IM}] \tag{2.15}$$

Initially, equation (2.15) can be solved independently for each i if: 1) we are able to ensure that there are no interdependencies between the best number of trains in different markets; and 2) we design a mechanism to allocate the fixed cost between the different services. In some instances we can guarantee these two conditions. For example, the optimal levels of service and fares of the various types of services are independent if there are no infrastructure capacity limitations and the services are not substitutes. Furthermore, we have shown that in unregulated monopolistic markets and in perfectly competitive market, the level of operations (provided that the TOs have incentives to operate trains) do not depend on the fixed costs. In this case, the results obtained match the results obtained in the previous sections.

This discussion is further extended in the rest of the dissertation to discuss how the results change 1) when the IM needs to assign fixed costs among the different types of services and 2) when there are infrastructure capacity limitations. Unfortunately, most railway systems fall into those categories. The issues around infrastructure capacity limitations are addressed by integrating the results of this chapter with the results of the infrastructure manager model from Chapter 3. We illustrate these issues in the context of the NEC (Chapter 5), where intercity passenger TOs, commuter TOs, and freight TOs compete to get access to the infrastructure. The issues around the allocation of infrastructure capacity fixed costs are in the context of the Central Corridor in Tanzania (Chapter 4) where different types of freight service TOs (general cargo and container TOs) share the same infrastructure. While (Laffont and Tirole, 1993; Laffont and

Tirole, 200; Perez-Arriaga, 2013) show that the use of marginal infrastructure costs to price capacity often allows for infrastructure cost recovery in other network industries, these results show that infrastructure cost recovery is not possible with marginal infrastructure cost pricing in most railway systems.

2.5 Conclusions

This chapter presents a simple TO Model based on standard TO financial models to discuss how TOs respond to alternative capacity pricing and allocation mechanisms as a function of the institutional and technical context in which the TOs operate. The TOs' response is captured by analyzing three main operational decisions: 1) the passenger fare or freight shipping rate charged to end users; 2) the number of trains operated or number of trains; and 3) the access charges paid to the infrastructure manager to access the infrastructure. Understanding the TOs' response to different access charges is an important step to analyze and compare alternative capacity pricing and allocation mechanisms. The model proposed allows regulators to robustly infer the TO demand for scheduling trains and their ability to pay to access the infrastructure with little information about the TO cost structure and the end users' demand.

There are four main take-away messages from this analysis (in reversed order). First, if the access charges reflect the infrastructure-related costs associated with operations, a vertically-separated TO would make the same operational decisions as an integrated railway system. However, if the access charges do not reflect the costs in which the IM incurs as a result of the operations of trains in the system, the TO would have incentives to provide different levels of service. This often translates into a loss in the total welfare. These implications also stand when different TOs share the same infrastructure. In other words, the use of marginal infrastructure costs to price capacity ensures that TOs make the same decisions than an integrated railway company. However, these mechanisms cannot be used in all cases. Chapters 4 and 5 discuss how to analyze these issues in cases in which there is a need to recover infrastructure costs or when infrastructure capacity is limited.

Second, the introduction of any type of TO regulation or the introduction of competition in the operations result in higher levels of service, lower fares for the end users, and higher levels of total welfare as compared to the operations of unregulated monopolistic operators. The operations of TOs in perfectly competitive markets and the operations of regulated monopolistic TOs are very similar. In some instances though, regulated TOs operate one more train because of the discrete nature of the number of trains.

Third, the number of trains estimate produced by our models depends on the functional form of the demand assumed, on the elasticity, and on the existence of any type of revenue management mechanisms. Although the results obtained are pretty robust to model inputs, a good characterization of the users' demand is important to accurately estimate the TOs operations. The evidence presented in the literature (Morrison, 1990) and the comparison of the results with current levels of service operated suggest that the demand of the NEC intercity passenger operators used as an example to illustrate the results in the different cases, is best characterized as a bounded isoelastic function of the fare. Nonetheless, the robustness of the results also justifies the approximation of the demand function by a linear function if this approximation simplifies the calculations.

Finally, this research analyzes the maximum access charges that different types of TOs would be able to pay to access the infrastructure. These results are very robust across the different cases studied. The model also anticipates the TOs' response (number of trains) to access charges. This information is used as inputs of the IM Model presented in Chapter 3, allowing the IM to anticipate and understand the operational goals and infrastructure needs of operators on their network. This information is also valuable for regulators, enabling them to understand the performance of the system under alternative mechanisms to price and allocate railway capacity.

Chapter 3 - The Infrastructure Manager Problem: Determining Train Timetable and Infrastructure Access Charges under Alternative Capacity Pricing and Allocation Mechanisms

"You never change things by fighting the existing reality. To change something, build a new model that makes the existing model obsolete." – B. Fuller (1981)

The train operator (TO) Model proposed in Chapter 2 can be used to anticipate the demand of individual TOs to access the infrastructure under alternative capacity pricing and allocation mechanisms. In congested shared railway systems, there are often conflicts between the services that different TOs would like to operate within the existing infrastructure. The next step to evaluate capacity pricing and allocation mechanisms is thus to analyze which trains can be scheduled within available infrastructure capacity.

However, capacity availability in the railway industry cannot be known in the absence and understanding of infrastructure operations. Therefore assessment of capacity requires the determination of the train timetable, which eliminates any potential conflicts between the TOs' requests to use infrastructure capacity. Although there is a broad literature that proposes train timetabling methods for dedicated railway systems, there are few models that can be used for shared competitive railway systems.

This chapter proposes a train timetabling model for shared railway systems explicitly considering a variable number of trains, with large flexibility margins (TOs' willingness to deviate from their desired timetable), and a variety of train services traveling along different paths. The TOs' demand for scheduling trains is assumed to be exogenous (input from Chapter 2). The model is formulated and solved both as a mixed integer linear programming (MILP) problem (using a commercial solver) and as a dynamic programming (DP) problem. We solve the DP formulation with a novel algorithm based on a linear programming (LP) approach to approximate dynamic programming (ADP) that can solve much larger problems than commercial MILP solvers.

This model can be used to evaluate the best possible train timetable under alternative capacity pricing and allocation mechanism. We use the results to understand the interactions between capacity planning and capacity operations in shared railway systems. Understanding these interactions is important to be able to design effective capacity pricing and allocation mechanisms. Part of this work has been submitted for publication (Pena-Alcaraz, Webster, and Ramos, 2015a).

The rest of the chapter is structured as follows: Section 3.1 reviews the train timetabling literature in shared railway systems and summarizes the main contributions of this chapter. Section 3.2 describes and formulates the train timetabling problem in shared railway systems, and motivates the assumptions of the chapter. Section 3.3 presents a DP formulation of the problem, and describes the LP-based ADP solution algorithm. Section 3.4 compares the computational performance of the ADP algorithm with the performance of commercial MILP solvers and illustrates the insights obtained using the algorithm to design shared railway systems' timetables for several cases with traffic patterns similar to the traffic of the Northeast Corridor (NEC) in the US. Section 3.5 summarizes the main implications of the results obtained with the model and concludes.

3.1 Train Timetabling Problem for Shared Railway Systems

As mentioned above, the design, assessment, and implementation of capacity pricing and allocation mechanisms at the strategic level are tightly coupled with the infrastructure operations at the tactical level. In other words, capacity utilization cannot be determined in the absence of infrastructure operations because available railway capacity depends on how the infrastructure is operated (Krueger et al., 1999; Pouryousef and Lautala, 2015). The operations in the railway industry are defined by the train timetable that determines the arrival and departure time at every station of all trains scheduled. As a result, the design of the train timetable is a critical step in any capacity planning mechanism. The timetable specifies how the competing demands for infrastructure access are coordinated to meet the infrastructure manager (IM)'s objectives and constraints.

The train timetabling problem has been widely studied in the literature. There are two main approaches to design the best train timetable that meets a set of operational constraints, both based on MILP formulations. (Castillo et al., 2009; Ghoseiri et al., 2004; Liebchen, 2008; Liebchen and Peeters, 2009; Pena-Alcaraz et al., 2011; Zhou and Zhong, 2005) present formulations to compute the train arrival and departure times. Traditionally, these models have been called multi-mode resource constrained project scheduling models. (Cacchiani et al., 2010; Caimi et al., 2009; Caimi et al., 2011; Caprara et al., 2002; Caprara et al., 2011; Cordeau et al., 1998) present formulations that represent the final timetable as a collection of nodes and arcs. Each arc represents possible train arrival and departure times at stations. Infrastructure and operational constraints are imposed by determining subsets of compatible and incompatible arcs. Traditionally, these models have been called multi-commodity flow models.

We argue, however, that these models cannot be used to analyze capacity planning mechanisms in shared railway systems for three reasons. First, with the exception of (Caprara et al., 2011), the approaches above assume a fixed number of trains to be scheduled on the infrastructure. However, the number of trains to schedule is the main decision variable of the capacity allocation problem in shared railway systems. Second, with the exception of (Caprara et al., 2011) again, the approaches above assume a single TO that tries to schedule trains. This TO could iteratively solve the train timetabling problem, introducing small modifications in each train desired timetable until the resulting timetable meets its needs. In shared railway systems, however, multiple TOs request access to the infrastructure. To ensure that the TOs reveal the value to themselves of each train to be scheduled and to avoid strategic behavior, the IM accepts inputs from the TOs only at specified time-points. As a result, TOs have incentives to provide large flexibility margins around the desired train timetables requested to ensure that the trains they value are scheduled even when there are small conflicts with other trains. The flexibility margin determines how much time TOs are willing to deviate from the desired timetable. Third, most of these models assume that all trains follow the same path. Again, this assumption does not hold when the nature of the services operated in the shared railway system is different. For example, commuter services are typically scheduled around the metropolitan areas, whereas intercity and freight TOs offer services between cities.

This chapter presents a multi-mode resource constrained project schedule formulation for shared railway systems that explicitly considers a variable number of trains, with large flexibility margins, traveling along different paths. This approach 1) introduces a discrete variable that indicates whether a train can be scheduled or not, 2) uses flexibility margins to ease conflicts, making fast trains travel slowly when there are slow trains ahead and making slow trains wait at sidetracks when fast trains overtake them, and 3) specifies safety constraints (spacing of the trains) for each train path.

These additional considerations make the problem very difficult to solve. From a computational standpoint, the size of the model increases exponentially with the number of stations and the number of trains to schedule. As a result, commercial solvers are only able to solve the problem for small number of trains. Furthermore, most of the techniques developed in the train timetabling literature are designed for traditional single-operator train timetabling problems and cannot be used in this case. Most classical decomposition approaches do not work because of the large number of discrete variables needed to specify which trains are scheduled and to pinpoint the order in which trains go through each station. Any technique that exogenously fixes train order cannot be used here because of the large flexibility margins and because train spacing constraints are specific to each individual train.

To be able to solve the problem in meaningful instances, we propose an alternative class of solution algorithms using ADP techniques (Bertsekas and Tsitsiklis, 1996; Bertsekas, 2006; Powell, 2007). This research develops a novel Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm that extends previous algorithms developed by (Farias and Van Roy, 2003; Farias and Van Roy, 2004). This algorithm allows us to decompose and solve large problems that are intractable with MILP commercial solvers while still converging to a solution within an optimality gap.

In summary, the introduction of shared railway systems requires the design, assessment, and implementation of capacity planning mechanisms to coordinate multiple TOs and the IM. The use of this novel algorithm allow us to solve the train timetabling problem in shared railway systems considering a large number of trains (100 to 150 trains). As a result, we are able to determine the optimal capacity allocation plan given the TO's demand for capacity under alternative capacity pricing and allocation mechanisms. We can use these results to anticipate the answers to relevant policy-type questions such as: how much should intercity TOs pay to be able to schedule services that conflict with commuter train services; whether freight TOs would be able to schedule any trains on the infrastructure, etc. The answers to these questions are central to our ability to design effective capacity pricing and allocation mechanisms.

This research makes both methodological and railway systems-specific contributions. From a methodological standpoint, we present a model that explicitly considers the relevant characteristics of shared railway systems, and offers a novel ADP algorithm for solving this complex train timetable problem for large system sizes that are computationally intractable using commercial software. From a transportation standpoint, the modeling framework and the algorithm developed enable us to simulate optimal decisions by an IM for shared railway systems. These results can be used to answer relevant policy-type questions to design appropriate pricing and allocation mechanisms and to understand the implications of infrastructure shared use.

3.2 Mixed-Integer Programming Formulation

In this section, we formulate the train timetabling problem for shared railway systems under capacity pricing and allocation mechanisms. As we discussed in Chapter 1, there are two main types of market-based mechanisms for capacity pricing and allocation: 1) mechanisms that determine the price at which capacity will be offered, and let TOs decide whether they are willing to access the infrastructure or not (price-based mechanisms); and 2) mechanisms that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity (capacity-based mechanisms or auctions) (Gibson, 2003). Price-based mechanisms are typically complemented with priority rules that allow the IM to decide which train to schedule when there are conflicts (multiple TOs willing to pay the predetermined access charges).

The model presented here determines the optimal set of trains that the IM can accommodate, assuming that an auction mechanism is implemented. Under an auction, at some predetermined frequency, the TOs will have the opportunity to submit bids. Each bid will consist of a list of the trains that the TO wants to schedule on the infrastructure, the desired timetable for each train, and the access charges they are willing to pay to schedule each train. The IM will then determine the set of trains that can actually be scheduled, their timetable, and the access charges that the TOs will pay. We assume that the IM's objective is to maximize revenue and cannot restrict access to the infrastructure beyond the infrastructure constraints (e.g., safety, infrastructure maintenance plans). This thesis assumes that the IM is government owned and not for profit, or in other words, that it does not uses it market power to restrict the access to the infrastructure to the TOs.

We also discuss below how to modify the model to determine the optimal set of trains that the IM can accommodate under alternative price-based mechanisms. The differences between the IM models for each mechanism affect mainly the definition of the parameters and the choice of the objective function. The constraints however are related to the physical characteristics of the infrastructure and remain unchanged across mechanisms. The model formulation is discussed below.

3.2.1 Sets

$i, i \in \{1, \dots, I\}$ train services proposed by the TOs in the bidding process.

$j, j \in \{1, \dots, J\}$ railway system stations.

3.2.2 Parameters

We use again lower-case letters to denote parameters. The information that the TOs provide in the bidding process for every train i is:

ini_{ij} a Boolean matrix that indicates the initial station j from which train i departs.

fin_{ij} a Boolean matrix that indicates the final destination (station j) of train i.

a_i the maximum access price (access charge) that the TO is willing to pay if train i is scheduled. For price-based mechanisms the access price will be predetermined (using, for example, a model to allocate infrastructure-related costs proportionally to infrastructure use) and fixed by the IM depending on the characteristics of the service. It is important to note that the TO will only operate a train if that price is less than or equal to its willingness to pay determined in Chapter 2.

$t_{ij}^{arr}, t_{ij}^{dep}$ the desired arrival and departure time of train i at every station j in the path of train i.

$\Delta\, td_i, p_i^{\Delta\, td}$ maximum acceptable translation, defined as the maximum acceptable difference between the desired timetable and the actual timetable at the initial station (see Figure 3-1) of train i and penalty imposed by the TO if the IM translates the train over the desired timetable. The penalty specifies the reduced access price that the TO is willing to pay.

$\Delta\, tr_i, p_i^{\Delta\, tr}$ maximum acceptable change in train i total travel time (see Figure 3-1) and penalty imposed by the TO if the IM increases the travel time of train i at any station with respect to the desired timetable.

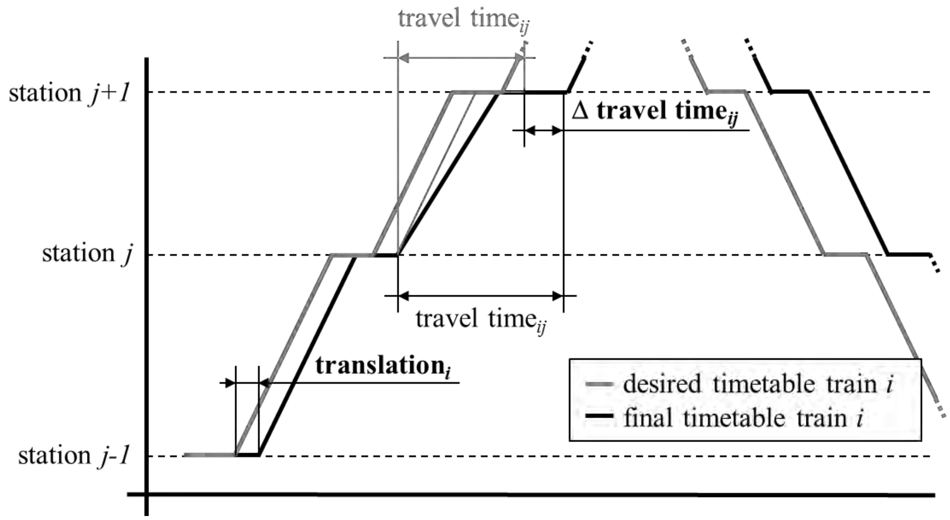


Figure 3-1 Time-space diagram representation of possible changes with respect to the desired timetable (Source: author)

The information about the topology of the line and the type of service is represented by the following two matrices:

$stat_{ij}$ a Boolean matrix that indicates whether train i travels through station j or not.

$next_{ijj'}$ a Boolean matrix that indicates for each train i the station j' that train i will visit immediately after station j. Train i may not stop at station j'.

In addition, the topology of the tracks and the signaling system will determine the minimum safe headway (time elapsed) between consecutive maneuvers at every station:

h_j^{arr}, h_j^{dep} minimum headway between consecutive arrivals/departures to/from station j.

In some cases the minimum safe headway depends also on the type of service and on the characteristics of the rolling stock. If that is the case, the former parameters will have different values for each train pair. The IM can set larger minimum headway to ensure the reliability of the timetable (including time-slack to recover delays in the system).

3.2.3 Variables

We use capital letters for variables. The endogenous decision variables of this problem are:

S_i binary variable that indicates whether train i is scheduled.

$T_{ij}^{arr}, T_{ij}^{dep}$ final arrival and departure time (timetable) of every train i scheduled at every station j in the path of the train.

ΔTD_i, ΔTR_{ij} final train i translation and increment of travel time per station j. Note that these variables can be determined knowing $T_{ij}^{arr}, T_{ij}^{dep}$ and vice versa. This research assumes $\Delta TR_{ij} \geq 0$ to ensure that the resulting train timetable is feasible. ΔTD_i can either be positive or negative; so we define the auxiliary positive variable ΔTD_i^{+} as the absolute value of ΔTD_i.

$O_{ii'j}$ binary disjunctive variable with value 1 if train i departs before train i' at station j and value 0 otherwise.

3.2.4 Objective Function

As discussed before, the objective of the problem is to determine which trains should be scheduled and when, in order to maximize the IM's revenue:

$$max\left[\sum_i a_i S_i - p_i^{\Delta td} \Delta TD_i^{+} - p_i^{\Delta tr} \sum_j \Delta TR_{ij}\right] \tag{3.1}$$

Alternative objective functions could be defined for different capacity pricing and allocation mechanisms. For example, the functions:

$$max\left[\sum_i S_i\right] \tag{3.2}$$

$$max\left[\sum_i pr_i S_i\right] \tag{3.3}$$

could be used to maximize the number of trains scheduled or the number of priority trains scheduled respectively under price-based mechanisms. In this case pr_i would be a parameter that indicates the priority level of each train i. This priority level can, for example, be proportional to the number of passengers times the miles of the service.

3.2.5 Constraints

The first set of constraints establishes the relation between the desired timetable and the final timetable of every train scheduled:

The departure time from the first station can be determined as:

$$T_{ij}^{dep} = t_{ij}^{dep} + \Delta TD_i, \forall i, j: ini_{ij} \tag{3.4}$$

The travel time between intermediate stations can be determined as:

$$T_{ij'}^{dep} - T_{ij}^{dep} = t_{ij'}^{dep} - t_{ij}^{dep} + \Delta TR_{ij}, \forall i, j, j': next_{ijj'}, fin_{ij'} = 0 \tag{3.5}$$

At the final station, the travel time can be determined using:

$$T_{ij'}^{arr} - T_{ij}^{dep} = t_{ij'}^{arr} - t_{ij}^{dep} + \Delta TR_{ij}, \forall i, j, j': next_{ijj'}, fin_{ij'} \tag{3.6}$$

Note that the arrival time at the initial station is not defined in the timetable, nor is the departure time from the last station.

To ensure that the timetable is feasible, the scheduled stopping and travel time at each station must be greater than or equal to the stopping and travel time in the desired timetable:

$$T_{ij}^{dep} - T_{ij}^{arr} \geq t_{ij}^{dep} - t_{ij}^{arr}, \forall i, j: stat_{ij}, ini_{ij} + fin_{ij} = 0 \tag{3.7}$$

$$T_{ij'}^{arr} - T_{ij}^{dep} \geq t_{ij'}^{arr} - t_{ij}^{dep}, \forall i, j: next_{ij'j}, fin_{ij} = 0 \tag{3.8}$$

The maximum translation and increment of travel time for which the TO receives a discount is constrained for each train scheduled. The allowable translation of a train is bounded by a maximum translation defined by the TO:

$$-\Delta td_i \leq \Delta TD_i \leq \Delta td_i, \forall i \tag{3.9}$$

In addition, the absolute value of the translation ($\Delta TD_i^+ = |\Delta TD_i|$) is determined using the following linear constraints:

$$\Delta TD_i^+ \geq \Delta TD_i, \Delta TD_i^+ \geq -\Delta TD_i \ \forall i \tag{3.10}$$

The maximum change on travel time is bounded by the maximum increment on travel time specified by the TO:

$$\sum_{j:stat_{ij}} \Delta TR_{ij} \leq \Delta tr_i, \forall i \tag{3.11}$$

The TO may impose additional conditions within the bid to define the acceptable changes with respect to the desired timetable. That happens when the TO is not interested in operating the train if the departure from or the arrival at one major station is changed. In this case, additional constraints are included to ensure that the timetable respects the TO's requests if the train is scheduled.

The final set of constraints ensures that the timetable proposed by the IM can be accommodated by the existing infrastructure. The IM must ensure first that the difference between the departure times of every pair of trains scheduled is greater than or equal to the minimum safe headway, so at least one of the following equations must hold:

$$T_{ij}^{dep} - T_{i'j}^{dep} \geq h_j^{dep} \tag{3.12}$$

$$T_{i'j}^{dep} - T_{ij}^{dep} \geq h_j^{dep} \tag{3.13}$$

These conditions can be expressed using the following disjunctive constraints:

$$T_{ij}^{dep} - T_{i'j}^{dep} \geq h_j^{dep} - M_{ii'j}\left(O_{ii'j} + 2 - S_i - S_{i'}\right), \forall i, i', j: i < i', stat_{ij}, stat_{i'j} \tag{3.14}$$

$$T_{i'j}^{dep} - T_{ij}^{dep} \geq h_j^{dep} - M_{ii'j}\left(3 - O_{ii'j} - S_i - S_{i'}\right), \forall i, i', j: i < i', stat_{ij}, stat_{i'j} \tag{3.15}$$

In these equations $M_{ii'j}$ is a "big enough" number to ensure that one and only one of the equations (3.12) and (3.13) holds. In this formulation we use $M_{ii'j} = h_j^{dep} + t_{i'j}^{dep} -$

$t_{ij}^{dep} + \Delta td_{i'} + \Delta td_i + \max(\Delta tr_i, \Delta tr_{i'})$, which is the smallest possible $M_{ii'j}$ that can be chosen for this problem. The binary disjunctive variable $O_{ii'j}$ is used to automatically activate only one of the constraints depending on the value of the other variables. $O_{ii'j}$ has value 1 if train i departs before train i' at station j. This problem has on the order of $O(I^2J)$ binary variables and is very difficult to solve for large I (number of trains) or J (number of stations) due to a large integrality gap.

Similar constraints are included for inter-arrival times to ensure that the order of the trains is preserved between stations.

$$T_{ij+1}^{arr} - T_{i'j+1}^{arr} \geq h_{j+1}^{arr} - M_{ii'j}\left(O_{ii'j} + 2 - S_i - S_{i'}\right), \forall i, i', j: i < i', stat_{ij+1}, stat_{i'j+1} \quad (3.16)$$

$$T_{i'j+1}^{arr} - T_{ij+1}^{arr} \geq h_{j+1}^{arr} - M_{ii'j}\left(3 - O_{ii'j} - S_i - S_{i'}\right), \forall i, i', j: i < i', stat_{ij+1}, stat_{i'j+1} \quad (3.17)$$

For these constraints, a value of $M_{ii'j} = h_{j+1}^{arr} + t_{i'j}^{arr} - t_{ij}^{arr} + \Delta td_{i'} + \Delta td_i + \max\left(\Delta tr_i, \Delta tr_{i'}\right)$ is used.

We emphasize that this formulation for shared railway systems differs in three aspects from traditional train timetabling problem formulations. First, it introduces the discrete variable S_i that indicates whether train i can be scheduled or not, which adds to the complexity of the problem. In contrast, the timetabling problem for a vertically-integrated railway will assume that all trains will be scheduled. Second, it uses flexibility margins $\Delta td_i, \Delta tr_i$ to alleviate conflicts. This is necessary because when different TOs are requesting trains, these conflicts are more likely to occur. Large flexibility margins result on high values of $M_{ii'j}$, making the problem hard to solve. Third, this formulation specifies safety constraints (spacing of the trains) for each train path, requiring the definition of the matrices $stat_{ij}, ini_{ij}, fin_{ij}, next_{ij'j}$.

This model is generalizable to other shared railway systems. The same equations will apply, with different parameter values to capture the system-specific information about the topology of the infrastructure, the path of the trains, the safe headways imposed by the signaling system, etc.

3.3 Linear Programming Approach for Approximate Dynamic Programming

As discussed above, the size of the MILP model proposed in Section 3.2 increases rapidly as a function of the number of stations and trains to schedule. We propose a novel solution algorithm using ADP techniques (Bertsekas and Tsitsiklis, 1996; Bertsekas, 2006; Powell, 2007) to tractably solve large timetabling problems in shared railway systems.

Specifically, we propose a Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm that extends the Approximate Linear Programming (ALP) and the Relaxed Linear Programming (RLP) algorithms developed by (Farias and Van Roy, 2003; Farias and Van Roy, 2004). QARLP introduces three main innovations with respect to ALP and RLP algorithms: 1) it incorporates the possibility of learning from previous solutions, allowing the algorithm to improve the solution obtained by refining the sampling strategy in subsequent iterations, 2) it formulates the Bellman equation using Q-factors, and 3) it implicitly samples through the state-action space, enabling the indirect identification of promising areas in the solution space, which is very difficult for large multidimensional problems. This approach decreases the solution time compared to a MILP commercial solver while still ensuring convergence to the optimal solution within a specified optimality gap.

3.3.1 Dynamic Programming Formulation

The problem defined in Section 3.2 can be reformulated as follows.

3.3.1.1 Stages

There are $i = 1, \dots, I$ decision stages (one for each train proposed to be scheduled), and a terminal stage $i = I + 1$.

3.3.1.2 State

The Markovian state variable is the timetable of the trains scheduled so far; that is, a matrix with the departure and the arrival times from/to the stations of all the trains scheduled so far:

$$x_i = \{timetable_{i-1}\}, \forall i \tag{3.18}$$

The timetable is defined as $timetable_{i-1} = \left[T_{i_1 j}^{arr}, T_{i_1 j}^{dep}; T_{i_2 j}^{arr}, T_{i_2 j}^{dep}; \dots\right]$, $\forall j$, $i_1, i_2, \dots : S_{i_k} = 1, i_k < i$.

3.3.1.3 Control

At every stage, the control variable indicates whether the IM decides to schedule train i or not, and, if scheduled, the specific timetable of train i at all stations j in the path.

$$u_i = \left\{S_i, T_{ij}^{arr}, T_{ij}^{dep}\right\} \in U(x_i), \forall i, j : i\ I, stat_{ij} \tag{3.19}$$

A train can only be scheduled if it does not present any conflict with the trains already scheduled. As we discuss in Section 3.4.1., the present and future value of scheduling each train ensures that the order in which the trains are visited (stages) does not affect the solution obtained.

3.3.1.4 State Transition Function

Given the state and the control at one decision stage, the state in the following decision stage can be computed, which incorporates the timetable of the new train if it is scheduled.

$$x_{i+1} = f(x_i, u_i) = \begin{cases} timetable_{i-1} & if\ S_i = 0 \\ \left[timetable_{i-1}; T_{ij}^{arr}, T_{ij}^{dep}\right] & if\ S_i = 1 \end{cases}, \forall i \leq I \tag{3.20}$$

3.3.1.5 Cost Function

The cost associated with a state-control pair is the sum of the penalties minus the revenue obtained if train i is finally scheduled. The sign of the cost function has been chosen to formulate a minimization problem. The cost associated with each state-action pair is evaluated using:

$$g(x_i, u_i) = g(u_i) = -a_i S_i + |\,\Delta\, TD_i| p_i^{\Delta\, td} + \sum_j \Delta\, TR_{ij}\, p_i^{\Delta\, tr}, \forall i \leq I \tag{3.21}$$

3.3.1.6 Bellman Equation

The policy that minimizes the sum of current and future costs at every decision stage can be determined by solving the Bellman equation and calculating the cost-to-go or value function:

$$J_i^*(x_i) = \min_{u_i} g(u_i) + J_{i+1}^*\big(f(x_i, u_i)\big), \forall i \leq I \tag{3.22}$$

$$J_{I+1}^*(x_{I+1}) = 0 \tag{3.23}$$

This equation can be reformulated using Q-factors, which represent the cost-to-go for every feasible state-control pair:

$$Q_i^*(x_i, u_i) = g(u_i) + \min_{u_{i+1}} Q_{i+1}^*(f(x_i, u_i), u_{i+1}), \forall i \leq I \tag{3.24}$$

$$Q_{I+1}^*(x_{I+1}, u_{I+1}) = 0 \tag{3.25}$$

The relation between the cost-to-go function and the Q-factor is:

$$J_i^*(x_i) = \min_{u_i} Q_i^*(x_i, u_i), \forall i \tag{3.26}$$

The optimal policy (timetable) can be determined solving the Bellman equation or the Q-factor Bellman equation using backward induction. However, when the dimension of the state space and/or the dimension of the control space increase, the solution of the exact

DP program becomes impracticable because the size of the problem grows exponentially. The benefit of reformulating the MILP model as a DP problem is that we can apply efficient solution algorithms such as the one proposed later in this chapter.

3.3.2 Linear Programming Algorithm

(Borkar, 1988; De Ghellinck, 1960; Manne, 1960) show that solving the Bellman equation (3.22) is equivalent to solving the LP problem proposed in equation (3.27) for any positive vector c because the inequality $J \leq J^*$ holds for every feasible solution J of the problem. The vector c is called the state-relevance weight vector.

$$\max c\mathrm{J}, \text{ s.t. } g_i(u_i) + \alpha \mathrm{J}_{i+1}\big(f_i(x_i, u_i)\big) \geq \mathrm{J}_i(x_i), \forall i \leq I, x_i, u_i \in U(x_i) \tag{3.27}$$

Note that the original problem does not have any discount factor, so we will use $\alpha = 1$ from now on. This LP problem has as many variables as possible states (value of the cost-to-go function at each state) and as many constraints as possible state-control pairs. When the state and control space of the problem are large, this results in a very large number of variables and constraints.

(Schweitzer and Seidman, 1985; Farias and Van Roy, 2003) proposed a modification of the previous formulation called the Approximate Linear Problem (ALP):

$$\max c\,\Phi\, r, \text{ s.t. } g_i(u_i) + \Phi_{i+1}\big(f_i(x_i, u_i)\big) r_{i+1} \geq \Phi_i(x_i) r_i, \forall i \leq I, x_i, u_i \in U(x_i) \tag{3.28}$$

where the real value function J_i^* is approximated by a linear combination of basis functions $J_i(x_i) \cong \sum_{k=1,\ldots,R} \Phi_{ki}(x_i) r_{ki} = \Phi_{\mathrm{i}}(x_i) r_i$. In this approximation, there are only $R \cdot I$ variables (number of basis functions and number of stages). However, the number of constraints remains the same as in equation (3.27) (one constraint for each state-control pair).

To reduce the number of constraints in this problem, (Farias and Van Roy, 2004) proposed a Relaxed Linear Problem (RLP) formulation. RLP proposes a strategy which samples constraints from the ALP formulation. Farias and Van Roy showed that for an appropriate probability distribution function Ψ over the set of state-control pairs, the number of constraints that must be sampled does not depend on the number of state-control pairs. In particular, to obtain a solution close enough to the optimal solution obtained using the ALP formulation with $1-\delta$ confidence level ($\Pr\{|\|J^* - \Phi r_{ALP}\|_{1,c} - \|J^* - \Phi r_{RLP}\|_{1,c}| < \epsilon\} \geq 1-\delta$), the number of samples required is on the order of a polynomial in the number of state variables, $1/\epsilon$, and $\log 1/\delta$. Note that these convergence results are computed over the basis-function approximation, that is, the RLP formulation converges to the best approximation over the basis functions chosen with confidence level $1-\delta$ within a number of samples that does not depend on the number of state-action pairs. The RLP formulation is:

$$\max c\Phi r, \text{ s.t. } g_i(u_i) + \Phi_{i+1}\big(f_i(x_i,u_i)\big)r_{i+1} \geq \Phi_i(x_i)r_i, \forall i \leq I, (x_i,u_i) \in X \tag{3.29}$$

where X is the set of state-action pairs sampled.

The main drawback of the RLP formulation presented in equation (3.29) is that the convergence results proved in (Farias and Van Roy, 2004) are based on an idealized choice of the probability distribution used to sample the constraints. In particular, the choice assumes knowledge of an optimal policy. Although it is unrealistic to assume that the optimal policy is known a priori, it is possible to obtain a reasonable approximation of the optimal policy by solving the RLP. Applying this idea, we propose the following Adaptive Relaxed Linear Programming (ARLP) algorithm:

Step 0: Set $t=0$, and sample X_0, giving each state-control pair equal probability to be sampled (Ψ_0 uniform distribution).

Step 1: Solve the problem $\max c\Phi r$,

---

s.t. $g_i(u_i) + \Phi_{i+1}\big(f_i(x_i, u_i)\big)r_{i+1} \geq \Phi_i(x_i)r_i, \forall i \leq I, (x_i, u_i) \in X_t$

Step 2: Set $t = t + 1$. Determine the optimal policy $\pi_{it}, \forall i$ according to the last problem solved ($\pi_{it}(x_i) = \text{argmin}\, g_i\big(\pi_t(x_i)\big) + \Phi_{i+1}\Big(f_i\big(x_i, \pi_{it}(x_i)\big)\Big) r_{i+1}, \forall i \leq I, x_i$). Choose the next set of constraints sampled using a probability distribution function Ψ_t that assigns higher probabilities to promising solutions considering the last iteration (i.e., the probability to sample (x_i, u_i) increases as $\Phi_{i+1}\big(f_i(x_i, u_i)\big)r_{i+1} / \Phi_{i+1}\Big(f_i\big(x_i, \pi_{it}(x_i)\big)\Big) r_{i+1}$ increases). In general, the variance of the probability distribution Ψ_t will decrease with t.

Step 3: If $t > T_0$ or the difference between the objective function is smaller than ϵ, stop. Otherwise go to Step 1.

Algorithm 3-1 Adaptive Relaxed Linear Programming (ARLP) algorithm

The ARLP algorithm iteratively solves a sequence of RLP problems, each with a manageable number of variables and constraints. This approach takes advantage of the reduced dimensionality of the RLP formulation while incorporating a mechanism to refine the sampling strategy Ψ_t using the best approximation of the optimal solution obtained so far. As a consequence, the convergence of the algorithm would not require the knowledge of the appropriate probability distribution function Ψ a priori.

However, because the basis function approximation used reduces the dimensionality of the problem, finding the state-control pair (e.g., timetable) that corresponds to known basis function values becomes a challenge. In other words, although it is very easy to determine the value of each basis function for a given state-control pair, solving the inverse problem (determining a state-control pair associated with a given basis function value) is extremely difficult in these cases. This is because the basis functions are a projection from the higher-dimensional state-action space to a lower-dimensional space, and the mapping from the low-dimensional projection back to the higher dimensional space is underdetermined. Therefore, there is no straightforward way to define Ψ_t based on low-cost regions in the basis function space, and to sample state-action pairs from it.

We solve this problem by 1) reformulating the algorithm using Q-factors instead of the cost-to-go function, and 2) determining Ψ_t implicitly using a Metropolis Hasting algorithm (Rubinstein and Kroese 2008) that accepts or rejects a state-control sample based on how promising the sample is according to the Q-factor best guess.

To do that, we sample uniformly across the state-action space to determine a candidate state-action pair (x_i', u_i'), we compute the value of its associated Q-factor $(\Phi_i(x_i', u_i')r_i)$, and compute the ratio of this Q-factor with the range of possible Q-factor values in the latest iteration $(\underline{\Phi_i(x_i, u_i)r_i}, \overline{\Phi_i(x_i, u_i)r_i})$:

$$\overline{\phi} = \frac{\Phi_i(x_i', u_i')r_i - \underline{\Phi_i(x_i, u_i)r_i}}{\overline{\Phi_i(x_i, u_i)r_i} - \underline{\Phi_i(x_i, u_i)r_i}} \tag{3.30}$$

We then draw a sample $\overline{\xi}$ from a probability distribution Ξ_t. We accept the state action pair if $\overline{\phi} \geq \overline{\xi}$. To ensure convergence to the optimal solution, the variance of the probability distribution Ξ_t must decrease with t, the probability of accepting any sample must be strictly positive, and the probability of accepting any sample with associated $\overline{\phi} \geq 1$ must be 1. To ensure that these conditions hold, we determine the probability of accepting a proposed sample based on a sample drawn from $\Xi_t \sim U(-\frac{1}{t}, 1)$. In other words, we accept a sample with probability $\alpha = \min\left(\max(\overline{\phi}, 0) + \frac{1}{t}, 1\right)$. The performance of the algorithms improves when the state-control pairs associated with binding constraints in the previous iteration are retained in future iterations.

We call this algorithm a Q-factor Adaptive Relaxed Linear Problem (QARLP) algorithm.

Step 0: Set $t = 0$, and sample X_0, giving each state-control pair equal probability to be sampled (Ψ_0 uniform distribution).

Step 1: Solve the problem $\max c\Phi r$,

s.t. $g_i(u_i) + \Phi_{i+1}(f_i(x_i, u_i), u_{i+1})r_{i+1} \geq \Phi_i(x_i, u_i)r_i, \forall i \leq I, (x_i, u_i, u_{i+1}) \in X_t$

Step 2: Set $t = t + 1$. Determine the optimal policy according to the last problem solved. Repeatedly sample state-control pairs until one pair is accepted if the ratio of its associated Q-factor ($\Phi_i(x_i, u_i)r_i$) and the range of possible Q-factor values in the latest iteration ($\overline{\phi}$) is greater than or equal to a draw of a probability distribution Ξ_t ($\overline{\xi}$).

Step 3: If $t > T_0$ or the difference between the objective function is smaller than ϵ, stop. Otherwise save binding constraints from previous iterations and go to Step 1.

Algorithm 3-2 Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm

Note that in this algorithm c, Φ, r have slightly different meanings than in Algorithm 3-1: c is a positive constant for every state-control pair at every decision stage and Φ, r are functions of the state-control pair (not only of the state: $\Phi, r = \Phi, r_i(x_i, u_i)$).

3.3.2.1 Capturing the problem structure: choosing basis functions

The choice of basis functions that capture relevant information about the state and the action while at the same time decreasing the amount redundant information (and hence the dimensionality of the problem) is a critical design choice of these types of ADP algorithms. In this research, we use basis functions that capture: 1) the total number of trains scheduled, as well as the total changes in the TO's desired timetable (state variable), 2) whether train i is scheduled or not, and the total changes in its desired timetable (control), 3) the number of conflicts of the trains scheduled so far with the following trains to be scheduled, and 4) and a constant. This reduces the dimensionality of the approximate cost-to-function to $R = 8$. That is,

$$\Phi_i(x_i, u_i) = (\sum_{i'<i} S_{i'}, \sum_{i'<i} \Delta TD_{i'}, \sum_{i'<i} \Delta TR_{i'}, S_i, \Delta TD_i, \Delta TR_i, C_i, 1) \quad (3.31)$$

where C_i is the number of conflicts of the trains scheduled so far with the following trains: $C_i = \sum_{i' > i}(Y_{i'})$. The variable $Y_{i'}$ has value 1 if the desired timetable of train i' conflicts with the timetable of any train scheduled so far and 0 otherwise. These basis functions capture the most relevant features of the problem, and therefore enable us to achieve a low approximation error around the optimal solution and to differentiate promising solutions from solutions that are less promising while reducing the dimensionality.

3.4 Results

In this section, we present the results for the train timetabling problem from both the MILP and the ADP formulations. Section 3.4.1 presents the computational results of the chapter, comparing the solution times between the commercial MILP solver and the QARLP algorithm. Section 3.4.2 illustrates the insights gained by using the model to design different timetables for a shared railway system.

3.4.1 Computational Results

We begin by presenting the results obtained from solving the timetable problem for a railway system with the infrastructure represented in Figure 3-2. It consists of a double-track corridor with 12 stations. Stations 1 and 7 are terminal stations at both ends of the line. Stations 2-12, 3-11, 4-10, 5-9 and 6-8 represent five stations along the corridor. We use a different station number for each traffic direction. Traffic moves in the direction of increasing station numbers in a dedicated track per direction. As a result, traffic traveling in different directions only interacts at the stations. The system presented includes the critical characteristics required to represent a corridor such as the NEC, for which the FRA is currently developing a new capacity pricing and allocation mechanism to foster rail efficiency (Gardner, 2013).

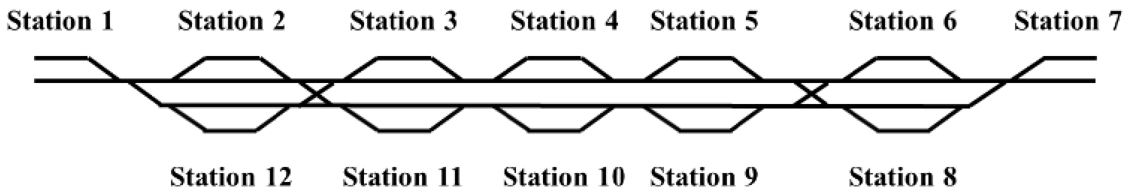


Figure 3-2 Detailed corridor infrastructure (Source: author)

Stations 1, 2 and 12 represent main stations in the same metropolitan area (e.g., Boston), stations 3, 4, 5, 9, 10 and 11 are all in another metropolitan area (e.g., New York), and stations 6, 7 and 8 are in yet a third distinct metropolitan area (e.g., Washington DC). Five types of services are considered: Boston commuter trains traveling around the Boston metropolitan area (stations 1, 2, and 12); New York commuter trains; DC commuter trains; and intercity and freight trains traveling between Boston and Washington DC. Intercity and freight trains may not stop at every station. Freight trains travel the line at speeds much lower than commuter and intercity trains. Intercity trains travel at higher speeds than commuter trains.

At present, around 2,000 commuter trains, 150 intercity trains and 70 freight trains travel around the NEC every day (Amtrak, 2010; Gardner, 2013). In practice, most of the conflicts to schedule trains occur around peak hours; where the IM would have to control for conflicts within sets of around 100-250 trains to make changes in the timetable.

We assume that the commuter TOs (one in each metropolitan area) request scheduling commuter trains every 30 minutes, and that one intercity TO requests operating a train every hour. The number of trains requested by TOs depends on the total time horizon considered.

The MILP formulation from Section 3.2 is implemented in GAMS 24.1.2 and solved using CPLEX 12.5 on a PC at 2.40 GHz, 4GB, intel core i7, under Microsoft Windows 7 64 bits. To reduce the size of the problem, when the desired arrival and departure times of two trains are very far apart, the value of the binary variable $O_{ii'\ j}$ is fixed a priori (since the relative order in which they pass through the station cannot change). We run CPLEX with options CHEAT $= 0.05$, RINSHeur $= 50$, Threads $= -1$ for faster solution times, and use a 5% optimality gap. A smaller optimality gap may be required if the timetable problem has multiple near-optimal solutions with very different implications for different TOs in terms of which trains are scheduled to ensure that the IM's choice of the trains to schedule is not arbitrary. In practice, for the cases solved for this research, the difference in the objective function between scheduling one additional train or not is large. As a consequence any solution within a 5% optimality gap of the optimal solution ensures that the set of trains scheduled is the same as the set of trains scheduled in the optimal solution unless there are twin trains (TOs willing to pay the same to operate trains with the exact same timetable). In that case neither CPLEX nor the QARLP algorithm would be able to distinguish those trains in the solution and the choice of one solution over other would be random.

We then solve the identical problem using the QARLP algorithm proposed in the previous section. Although theoretically the relative order of the trains does not change the solutions obtained or the convergence speed of the algorithm to the optimal solution, in practice the relative order of the trains may speed up or slow down the process of finding the optimal solution. The results presented in this chapter correspond to cases in which the relative train order (trains considered at each stage) was randomly assigned.

Table 3-1 shows the number of equations, variables, and discrete variables for problems with several different numbers of requested commuter and intercity trains. Figure 3-3 presents the execution time and the number of iterations required to convergence (within 5% integrality gap) for the MILP and QARLP algorithms. We compute the QARLP algorithm integrality gap using the "best possible" solution bound generated by the MILP commercial software.

Table 3-1 Train timetabling problem size for traffic patterns representative of the NEC Traffic (Source: author)

Number of Trains	Equations	Number of Variables	Discrete Variables
15	970	510	91
30	3,715	1,607	292
60	14,533	5,565	919
120	57,481	20,537	3,145

Note that the QARLP execution time increases as a polynomial function of the number of trains to schedule. In contrast, the MILP solution times increase exponentially. In fact, solving the MILP problem with CPLEX for 75, 90, or 120 trains within a 5% convergence gap is *computationally intractable*. Extrapolating from a regression estimate (Figure 3-3), the solution time using CPLEX for 120 trains would be approximately 46 days. The solutions obtained with the QARLP algorithm for 90 and 120 trains in approximately 20 minutes are better than those obtained for the MILP formulation with CPLEX after 20 hours and 35 hours respectively. In the cases with 15, 30, and 60 trains the solutions obtained with both methods are almost identical.

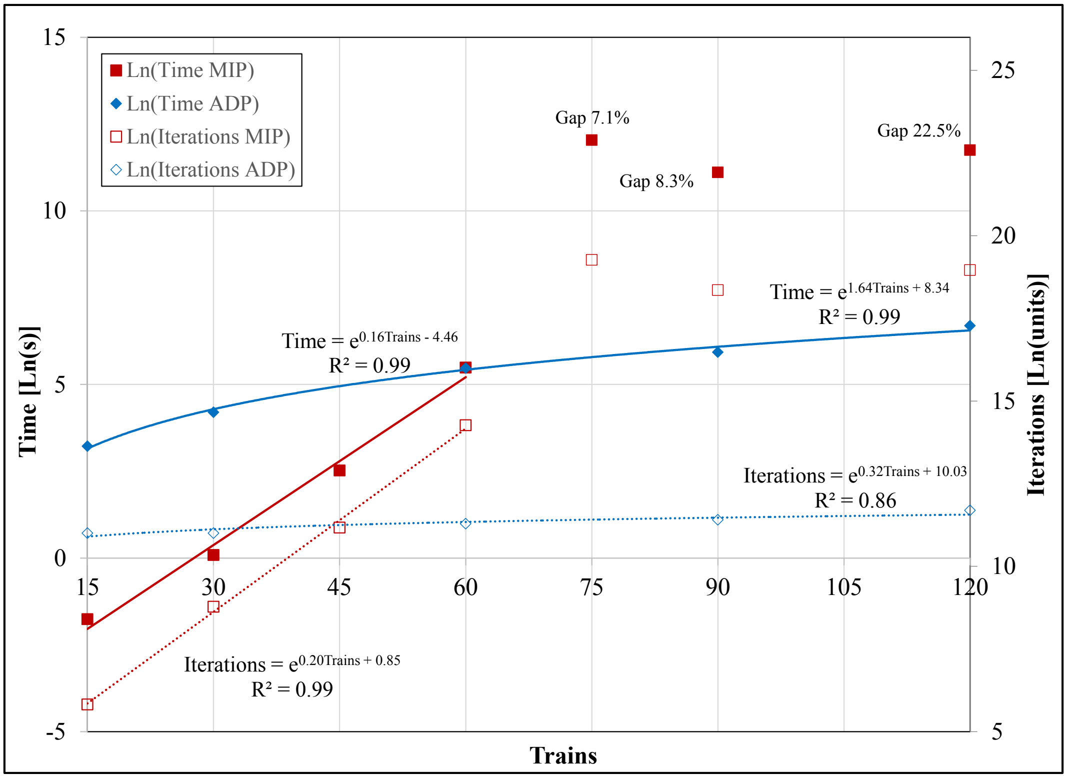


Figure 3-3 Comparison of the execution time and the iterations required to convergence (within 5% integrality gap) of the MILP approach and the QARLP algorithm (Source: author)

3.4.2 Design of Timetables for Systems with Traffic Patterns similar to the US Northeast Corridor's Traffic

In this section, we are interested in analyzing the timetables designed for relevant cases that illustrate trade-offs involved in the capacity planning process in shared railway systems. We analyze the optimal capacity allocation plan (train timetable) to determine how to coordinate different TOs' conflicting demand for scheduling trains. The ability to solve this allocation problem is critical for designing effective capacity pricing and allocation mechanisms. Figures 3-4, 3-5, 3-7, and 3-8 show time-space diagrams for timetables designed by the IM model for cases with different demands for accessing the infrastructure. The y-axes represent distance in miles from station 1 and the x-axes represent time in minutes at which different trains are scheduled to pass through each point of the line (vs. desired scheduled in dashed line). The horizontal segments represent the stopping times at stations. We assume no interaction between trains traveling in different directions.

Figures 3-4 and 3-5 show the timetable for a case with demand for scheduling an intercity train in the system when commuter trains around the three metropolitan areas operate every 30 minutes. Figure 3-7 shows a case in which two competing intercity TOs request scheduling intercity trains when commuter trains around the three metropolitan areas operate every 30 minutes. These two cases provide information about how much intercity TOs will have to pay to be able to schedule services that conflict with commuter train services. Figure 3-8 shows a case with demand for scheduling a freight train in the system when commuter trains around the three metropolitan areas operate every 1 hour. This case is designed to analyze whether freight TOs would be able to schedule *any* trains on the infrastructure. The IM model proposes the final timetable analyzing the trade-off between eliminating trains and readjusting the desired schedules, according to the objective function in (3.1).

For this example, we assume that each commuter TO pays 1 unit to schedule a commuter service and gets a 3% discount from the original access charge for every minute that one of their train schedules is changed. To analyze the first case, we need to solve a train timetabling problem with 115 commuter trains and 1 intercity train. Figure 3-4 shows the timetable of all the trains scheduled. For clarity purposes, only the schedules of conflicting trains are shown in Figures 3-5, 3-7, and 3-8.

Note that when the IM tries to schedule the intercity train, it will initially conflict with 14 commuter trains (see Figure 3-5). Rescheduling the commuter trains to accommodate the intercity service requires that the commuter TOs receive a discount of 2.1 units on their total access charges. As a result, the IM would only schedule the intercity train if it represents more than 2.1 units of revenue.

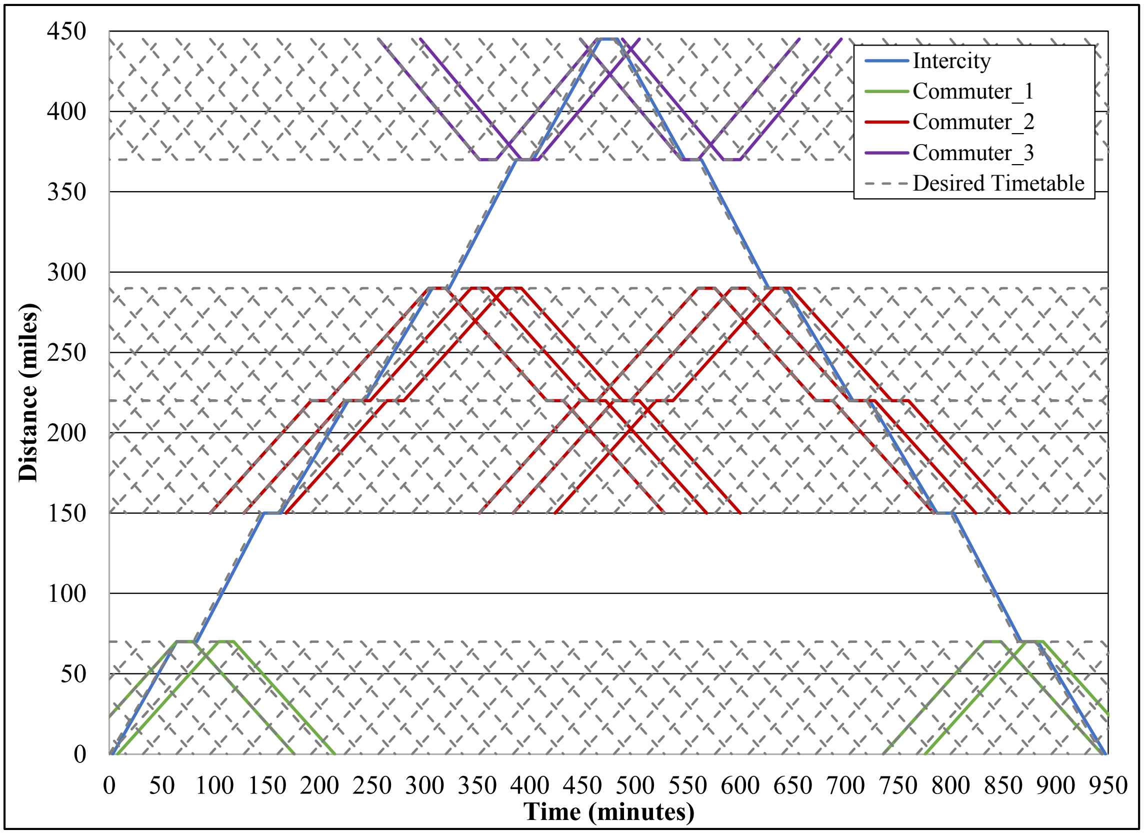


Figure 3-4 Timetable proposed by IM to schedule an intercity train in a system with commuter trains operating every 30 minutes – including conflicting and non-conflicting trains (Source: author)

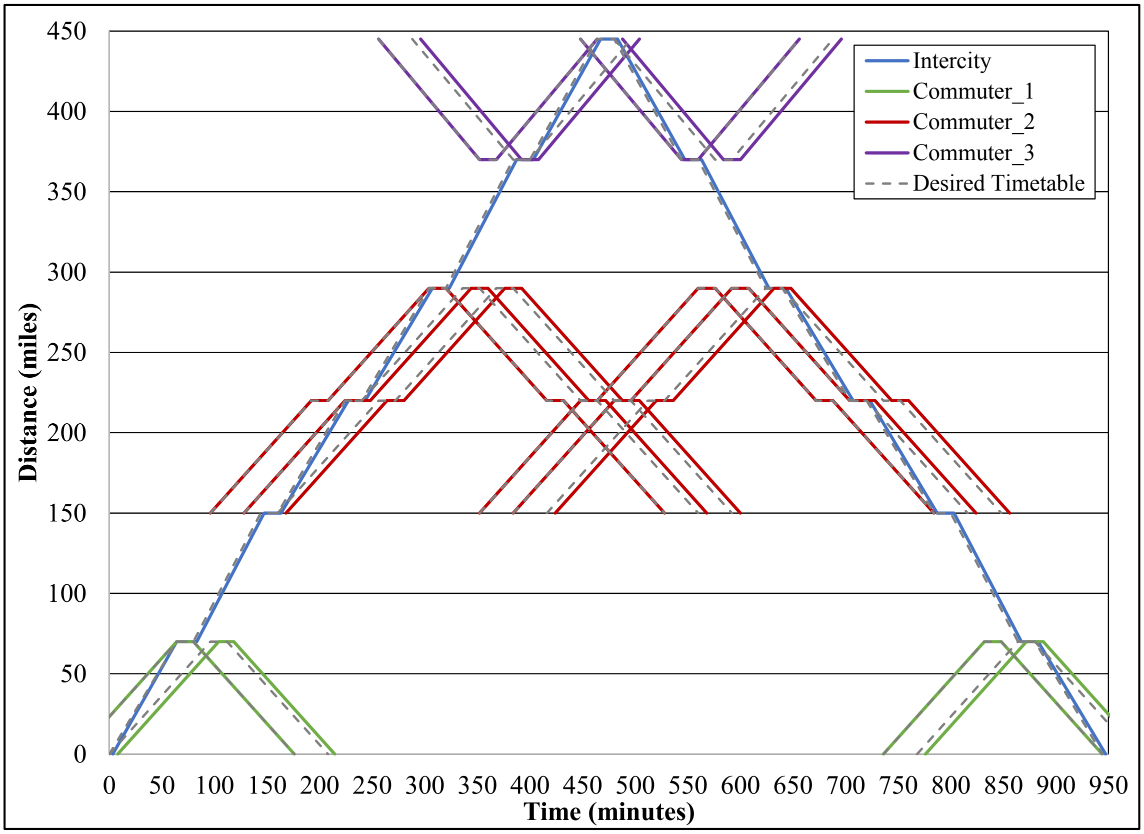


Figure 3-5 Timetable proposed by IM to schedule an intercity train in a system with commuter trains operating every 30 minutes (Source: author)

If the frequency of commuter trains increases, for example to one commuter train every 15 minutes instead of every 30 minutes, the intercity train will initially conflict with 22 commuter trains and will only be scheduled if it represents more than 3.6 units of revenue for the IM (i.e., if the intercity bid is higher than 3.6 units). Conversely, if the frequency of commuter trains decreases to one train every 60 minutes, the intercity train will be scheduled if it represents at least 1.5 units of revenue for the IM. The model can be used to quantify the trade-off between commuter and intercity trains for any other frequency of service (see Figure 3-6). The exact value of the trade-off for low frequencies of commuter services depends on whether there are conflicts among the desired timetables of the trains or not. The results show that the price that an intercity or freight TO will have to bid to be able to schedule a train (minimum access charge) can vary considerably as a function of the congestion of the line. This minimum intercity access charge reflects the congestion rent. The results show that greater cost recovery is expected in congested infrastructure.

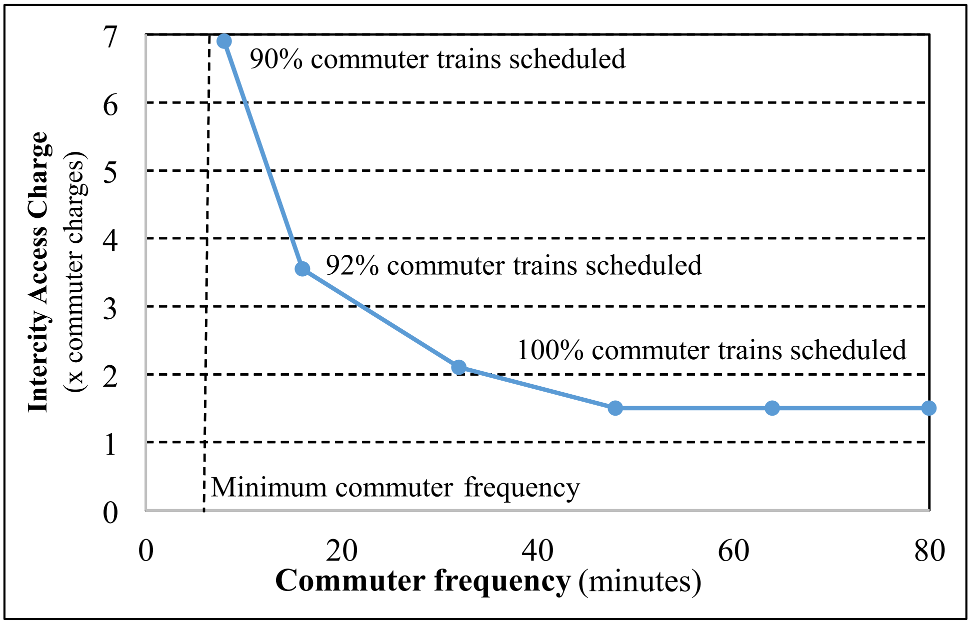


Figure 3-6 Intercity to commuter access charge ratio as a function of the commuter frequency
(Source: author)

We can use the same logic to analyze what would happen if two intercity TOs want to schedule intercity trains at the same time. This case also assumes that commuter TOs try to schedule commuter trains every 30 minutes. We assume again that each commuter TO pays 1 unit to schedule a commuter service and gets a 3% discount from the original access charge for every minute that one of their train schedules is changed. In simple auctions (like the one considered in this chapter), TOs reveal how much access charges they are willing to pay to schedule a train. This case is important because the TO's willingness to pay to access the tracks may change if a competing service is scheduled right before. According to Figure 3-7, scheduling two intercity trains would require changing the desired schedule of 14 commuter trains again. However, the changes in the commuter timetable are much larger (because commuters are overtaken by two intercity trains) as compared to the previous case. As a result, the IM would only schedule the two intercity trains if they represent more than 4.0 units of revenue. If the revenue from scheduling the intercity trains represent between 2.1 and 4.0 units, at most one of the intercity trains would be scheduled.

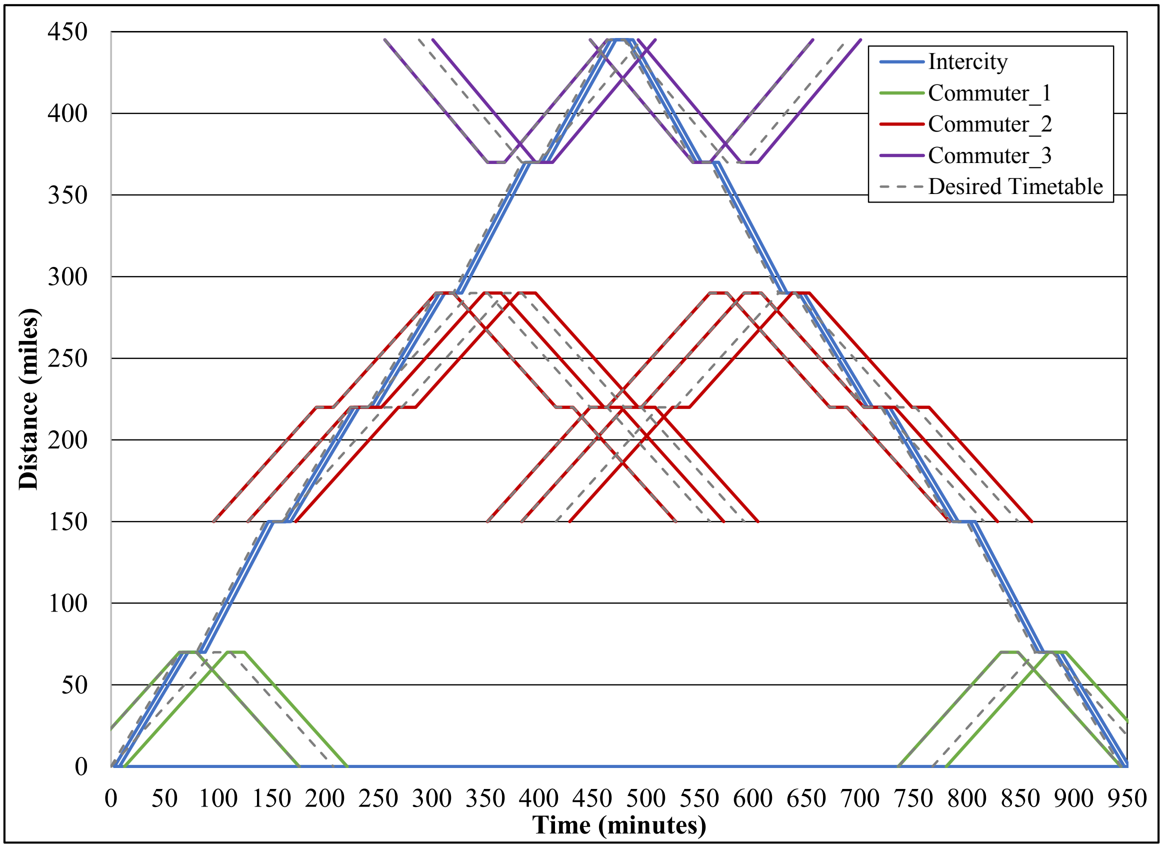


Figure 3-7 Timetable proposed by IM to schedule two intercity trains in a system with commuter trains operating every 30 minutes (Source: author)

Furthermore, note that although both trains would like to depart station 1 at minute 0, one of them will depart at minute 3 and the other one at minute 8. In some cases, none of the TOs may be interested in operating a second intercity service just 5 minutes after another one. These results suggest that intercity TOs may avoid getting their train scheduled just after other intercity train in simple auctions by controlling 1) how flexible their schedule is, 2) how much discount in the access charge they request if the schedule of the train is changed, and 3) how much they are willing to pay to access the infrastructure. This is important because the railway literature assumes that the value of scheduling trains for TOs can only be captured by combinatorial auctions (Perennes, 2014). However, these results demonstrate that TOs can also avoid getting their trains scheduled right after other competing service in congested systems even when using simple auctions.

The third case considered in this section analyzes whether a TO would be able to schedule trains in the system if it cannot afford to pay high access charges. This discussion can be particularly relevant to understand if freight TOs with low access charges willingness to pay may be able to access the infrastructure in shared railway systems. Figure 3-8 shows that a freight train could be scheduled paying the same access charges as commuter trains (with frequency one train pair hour) if the freight TO is very flexible (in terms of the total allowed translation and increment of travel time it accepts).

The minimum access charge that a freight TO must pay when the line is more congested will depend on how many trains have to be rescheduled to eliminate conflicts. If the commuter TO wants to increase the frequency of commuter service from one train per hour to one train every 30 minutes, the freight train will only be scheduled if the net access charge that the freight TO is willing to pay represents more than 3 units of revenue for the IM (since three commuter services could not be operated). In general, the relative speeds among different types of services have a major impact on the capacity utilization of the system.

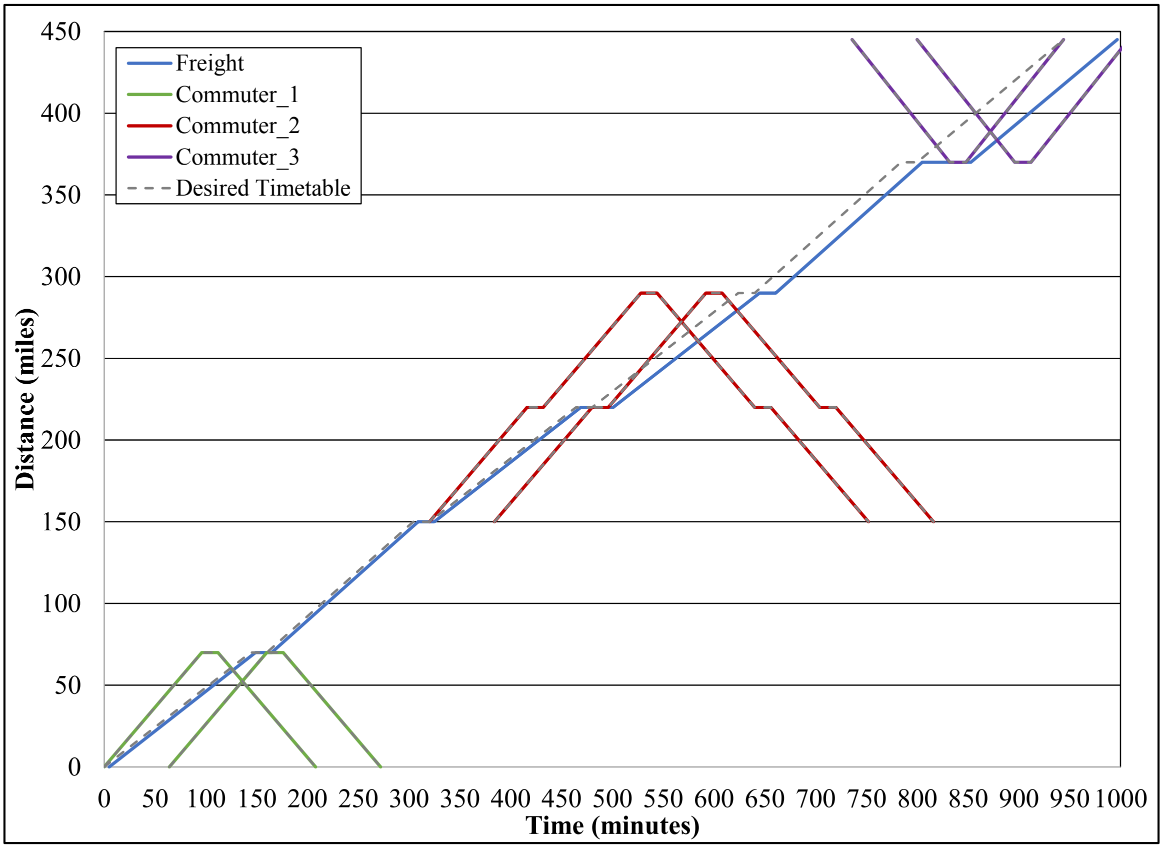


Figure 3-8 Timetable proposed by IM to schedule a freight train in a system with commuter trains operating every 60 minutes (Source: author)

The results in these cases show that intercity and freight services would have to either pay higher access charges or to be more flexible to compete with commuter trains. The exact amount (access charges or flexibility needed) depends on the frequency of commuter services. This strong dependency on the operational variables demonstrates that we need to consider the interactions between capacity planning and operations to evaluate capacity pricing and allocation mechanisms.

3.5 Conclusions

This chapter proposes a train timetabling model for shared railway system that explicitly considers a variable number of trains, with large flexibility margins, traveling along different paths. The model is formulated as both a MILP problem and as a DP problem. The MILP is solved using commercial software and the DP is solved using a novel algorithm for ADP. The timetables designed with the model are used to evaluate how capacity pricing and allocation may impact different railway system stakeholders. As a result, the contributions of this chapter are both methodological and domain specific.

On the methodological side, the main contributions of the chapter include:

1) The formulation of a train timetabling model for shared railway systems that would allow regulators and decision makers to determine the optimal use of railway infrastructure capacity. This model explicitly considers a variable number of trains, with large flexibility margins, traveling along different paths to analyze the interdependencies between operations and available infrastructure capacity and how they affect the coordination between the TOs and the IM.

2) The development of a novel algorithm for rapidly solving the train timetable problem in shared railway systems, ensuring convergence to the optimal solution within a specified optimality gap. We obtain solutions within 5% of the optimal solution for problem sizes that cannot be solved within a 5% convergence gap using commercial MILP software.

3) The algorithm developed, a Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm, extends the Approximate Linear Programming (ALP) and the Relaxed Linear Programming (RLP) algorithms developed by (Farias and Van Roy, 2003; Farias and Van Roy, 2004). QARLP introduces three main innovations with respect to ALP and RLP algorithms: 1) it incorporates the possibility of learning from previous solutions, allowing the algorithm to improve the solution obtained by refining the sampling strategy in subsequent iterations, 2) it formulates the Bellman equation using Q-factors, and 3) it implicitly samples through the state-action space, enabling the

indirect identification of promising areas in the solution space, which is very difficult for large multidimensional problems. These ideas can be generalized to efficiently solve other large-scale network optimization problems.

Moreover, the results of the train timetabling model can be used to simulate and evaluate the best possible behavior of the IM in shared railway systems under different capacity pricing and allocation mechanisms. The domain-specific contributions of this chapter are:

4) The modeling framework and the algorithm developed in this chapter enable us to simulate optimal decisions by an IM for shared railway systems. These results can be used to answer relevant policy-type questions to understand the implications of infrastructure shared use.

5) This chapter also shows that the implications of capacity planning mechanisms depend on the characteristics of the system and the TO demand for accessing the infrastructure. We propose the use of this model as a tool to allow regulators and decision makers to better understand the interactions between capacity planning and operations under alternative capacity pricing and allocation mechanisms.

This chapter considers the TOs' infrastructure access demand (characterized both as the demand for scheduling trains and the revealed willingness to pay to access the infrastructure) as exogenous to the problem. However, the TO's infrastructure access demand depends on the capacity pricing and allocation mechanism. Section 5.1 (Chapter 5) discusses the integration of the IM model proposed in this chapter with the model of TO bidding behavior developed in Chapter 2 to better quantify the trade-offs between utilization and level of service on the one hand, and infrastructure cost recovered under different capacity pricing and allocation mechanisms. The rest of Chapter 5 uses that framework to analyze alternative capacity pricing and allocation mechanisms in the context of the NEC. These results are valuable to design and evaluate alternative capacity pricing and allocation mechanisms to effectively coordinate the TOs and the IM in shared railway system.

Chapter 4 - Policy Implications for the Central Corridor in Tanzania and Other Shared Railway Systems with Infrastructure Cost Recovery Constraints

"If you want to travel fast, travel alone. If you want to travel far, travel together." – African Proverb

In 2013, Tanzania's government committed to the implementation of one of the first shared railway systems in Africa (Big Results Now, 2013) as a way to ensure adequate level of rail service by 1) allowing efficient train operators (TOs) to access the infrastructure and operate train services through an open-access policy, and 2) providing sustainable resources through access charges to maintain the infrastructure and keep the system operative in the future. These objectives are critical to prevent future railway systems failures such as the 2001 and 2006 Tanzanian railway system concessions failures (Olievschi, 2013) that resulted in a major underinvestment in rail transportation in the country (Railistics, 2013). This underinvestment critically impacted the operating capacity and the reliability of the railway system, essential to improving accessibility to the East African landlocked countries: Rwanda, Burundi, Uganda, and Western Democratic Republic of Congo (AICD, 2008; Amjadi and Yeats, 1995; Arvis et al., 2010; Raballand and Macchi, 2009).

The implementation of a shared railway system requires new railway regulations that clarify the roles and responsibilities of railway institutions (Railistics, 2013; World Bank, 2014) as well as the design and implementation of a new mechanism to price and allocate railway capacity. This chapter analyzes how alternative capacity pricing and allocation mechanism for freight TOs would affect the performance of the Central Corridor in Tanzania. We are particularly interested in the number of trains operated in the system and the revenues collected to maintain the infrastructure and recover capital costs. Chapter 2 shows that traditional approaches to price and allocate railway capacity may not work in two cases: 1) when the infrastructure manager (IM) needs to recover part of the infrastructure management fixed costs or 2) when the railway system is congested. Tanzania's Central Corridor falls into the first case. As mentioned above, one of the main purposes of the introduction of shared use in Tanzania is to ensure that the IM is able to raise revenues to maintain the infrastructure and keep the system operational.

An important characteristic of the Central Corridor is that it is not congested. The only current TO, TRL, operates around six trains per week, leaving plenty of spare capacity that could be used by other TOs (World Bank, 2014). As a result, allocating capacity is fairly easy and we can solve the capacity pricing problem independently of the capacity allocation one. In other words, we can use the TO Model to determine the TOs' demand for scheduling trains on the infrastructure. This process can be done for each type of service independently, since there is enough infrastructure capacity to accommodate the demand of the all TOs (Pena-Alcaraz et al., 2014).

The results of this chapter show that the introduction of variable access charges distorts the operational decision of TOs, as predicted in Chapter 2. We then discuss how to avoid this problem with other pricing mechanisms such as the introduction of fixed access charges. We also discuss how to allocate fixed access charges among multiple types of freight TOs and show the need for price discrimination in this context. The results also show that it is not possible to recover infrastructure costs from dedicated container or general freight traffic in the context of the Central Corridor. However, the shared use of the infrastructure by container and general freight TOs allows the IM to fully recover infrastructure costs. This is one of the benefits of shared railway systems. This work is published in Network Industry Quarterly (Pena-Alcaraz, Perez-Arriaga, and Sussman, 2014).

The rest of the chapter is structured as follows: Section 4.1 presents the main types of capacity pricing mechanisms and discusses how the TO Model presented in Chapter 2 can be used to determine the behavior of TOs under each mechanism. Section 4.2 presents the resulting number of trains that container and general cargo freight TOs would operate under alternative capacity pricing mechanisms. Section 4.3 concludes with some recommendations for capacity pricing mechanisms in shared railway systems with infrastructure cost recovery constraints.

4.1 Capacity pricing mechanisms for shared railway system

The Central Corridor goes from the port (Dar es Salaam) to an inland container terminal (Isaka) that serves as a dry port for Rwanda, Burundi, Uganda, and the Eastern portion of Democratic Republic of Congo (see Figure 4-1). The infrastructure is owned by RAHCO, a publicly owned company. TRL is the only current TO; it operates around six trains per week. Although the corridor is single track, there is plenty of spare capacity that could be used by multiple private companies that have expressed interest in starting operating new services between Dar es Salaam and Isaka (Pena-Alcaraz et al., 2014; World Bank, 2014).

As we discussed in Chapter 1, the implementation of a shared railway systems requires some level of vertical separation between the TOs that operate the trains in the system and collect the revenues selling transportation services to the final customers and the IM that maintains and manages the infrastructure. Vertical separation requires the definition of a capacity pricing mechanism that determines the access charges that TOs pay to the IM to access and use the infrastructure (Gomez-Ibanez, 2003). The IM uses these revenues to cover infrastructure costs. The use of the state national budget to cover shortfalls is the last resort.

Figure 4-1 Central Corridor (Source: United Nations, 2006, Map No. 3667 from (World Bank, 2014))

The railway literature proposes capacity-based and price-based (also called cost-based) mechanisms to price railway capacity (Gibson, 2003). Capacity-based mechanisms are those that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity. However, in cases like the Central Corridor with plenty of excess capacity, the TOs would be able to access the infrastructure paying very low access-charges. As a result, capacity-based mechanisms are not an option to recover infrastructure costs in non-congested systems. There are three cost-based capacity pricing mechanisms designed to allow the IM to recover maximum infrastructure costs: variable access charges, two-part tariffs (variable access charges plus a fixed access charge), and fixed access charges (Gibson, 2003). Under variable access charges, TOs pay some amount per train operated; the charge is in general a function of the type of train, distance, and tonnage. Under fixed access charges, each TO pays an annual lump sum to have a license to operate, regardless of the number of trains the TO operates during the year.

The practice and the broad economic literature in the field recommend the use of variable access charges based on marginal cost plus mark-ups (DB, 2009; Lopez-Pita, 2014; Nash, 2005; World Bank, 2014). However, from an engineering standpoint, infrastructure-related costs in Tanzania are mostly independent of the number of trains. Due to the low number of trains that operate in the system, the infrastructure maintenance costs do not increase (for all practical purposes) when more trains are operated. As a result, maintenances costs are assumed fixed. In other words, the short-term and long-term infrastructure marginal costs are very low and high mark-ups are required to recover infrastructure costs. This research analyzes the implications of resulting alternative pricing mechanisms for the system.

For this analysis, we compare the behavior of vertically-separated TOs with the behavior of an integrated railway company (social planner). We assume that both the vertically-separated TOs and the integrated railway company are rational agents, i.e. they determine the number of services per direction per week to operate by maximizing the annual operating margin (operating profits). A vertically-separated TO would only be interested in operating trains if the average annual net cash flow is positive after remunerating any invested capital at an adequate rate of return (no operations subsidies).

We use the financial TO Model developed in Chapter 2 to determine the integrated railway company, vertically-separated TO, and vertically-separated IM's operating margin and cash flow for a representative year under different levels of service. See Chapter 2 for more details about the TO Model and (Pena-Alcaraz et al., 2014; PPIAF et al., 2011; World Bank, 2014) for detailed model assumptions. The integrated railway company faces capital costs associated with the investments in railway infrastructure, variable costs of operating trains (train lease, personnel, fuel), and obtains revenues from transporting freight. The vertically-separated case is similar: the TO faces cost of accessing the tracks (access charges), variable costs of operating trains, and obtains revenues from transporting freight. The IM faces investment costs in railway infrastructure, maintenance costs, and obtains revenues from access charges.

Investment in railway infrastructure includes $300 million investment required to rehabilitate the current Tanzanian railway system (CPCS, 2013; World Bank, 2014) plus periodic investment in maintenance. The revenues of the TOs are determined multiplying the cargo transported (minimum between the capacity of the trains operated and the demand) by the freight shipping rate. Due to the strong competition from trucks that offer door-to-door transportation services, railway companies have an upper limit on the freight shipping rate they may charge and they have low control over the demand that would likely shift to rail. The state should facilitate strong intermodal integration with the port and with truck companies that provide last mile transportation to/from the terminal rail station to make rail transportation more attractive and increase the utilization of the highly underused railway capacity. All the financial information used in this analysis is publicly available (World Bank, 2014).

4.2 Discussion of the Results

In this section, we discuss the main results obtained for alternative capacity pricing mechanisms designed to recover maintenance and financial infrastructure costs and to ensure that TOs can viably operate (positive profits) in Tanzania in two scenarios: 1) considering only container TOs, and 2) considering both container and general cargo (non-containerized freight) TOs.

4.2.1 Container Traffic

Figure 4-2 shows the annual operating margin and the cash flow for a vertically-separated container TO, for the IM, and for an integrated railway company in Tanzania under variable and fixed access charges when no other type of TO operate in the line. Both access charges have been calculated to recover as much of the infrastructure costs as possible, while ensuring that the operating margin and the net cash flow of the vertically-separated TO are positive. Note that it is not possible to recover all the infrastructure cost ($22.9 million per year in Tanzania) only with container services. The maximum charges that a vertically-separated TO could viably pay are $0.035 per ton-km (variable, assuming the TO would operate four trains) or $19.1 million per year (fixed). We compute these numbers estimating the TO maximum revenues, the variable and fixed costs, and therefore the maximum fixed and variable access charges that the TO can viably pay to achieve an annual net cash flow equal to zero.

The results also show that under variable access charges only, a rational vertically-separated TO would only operate two trains per direction per week while the social planner would operate four. This mismatch happens because when the social planner tries to maximize its operating margin, it operates a train when the additional revenues produced are higher than the additional variable costs (train lease, personnel and fuel). For the social planner, most infrastructure investment cost is a sunk cost: it is already made and it is independent of the number of trains operated in the system. Under variable access charges in contrast, the infrastructure costs are charged as variable costs for TOs. Therefore, a rational TO would only operate a train if the additional revenues produced are higher than the true variable costs plus a share of the infrastructure cost that appears now as an artificial variable cost (the variable access charge).

Under fixed access charges, the infrastructure costs are charged as a fixed cost for TOs. Therefore, this cost will also be a sunk cost for the TO. Consequently, the TO will operate a train when additional revenues produced are higher than the true variable operational costs and there is no mismatch with the number of trains operated by the social planner.

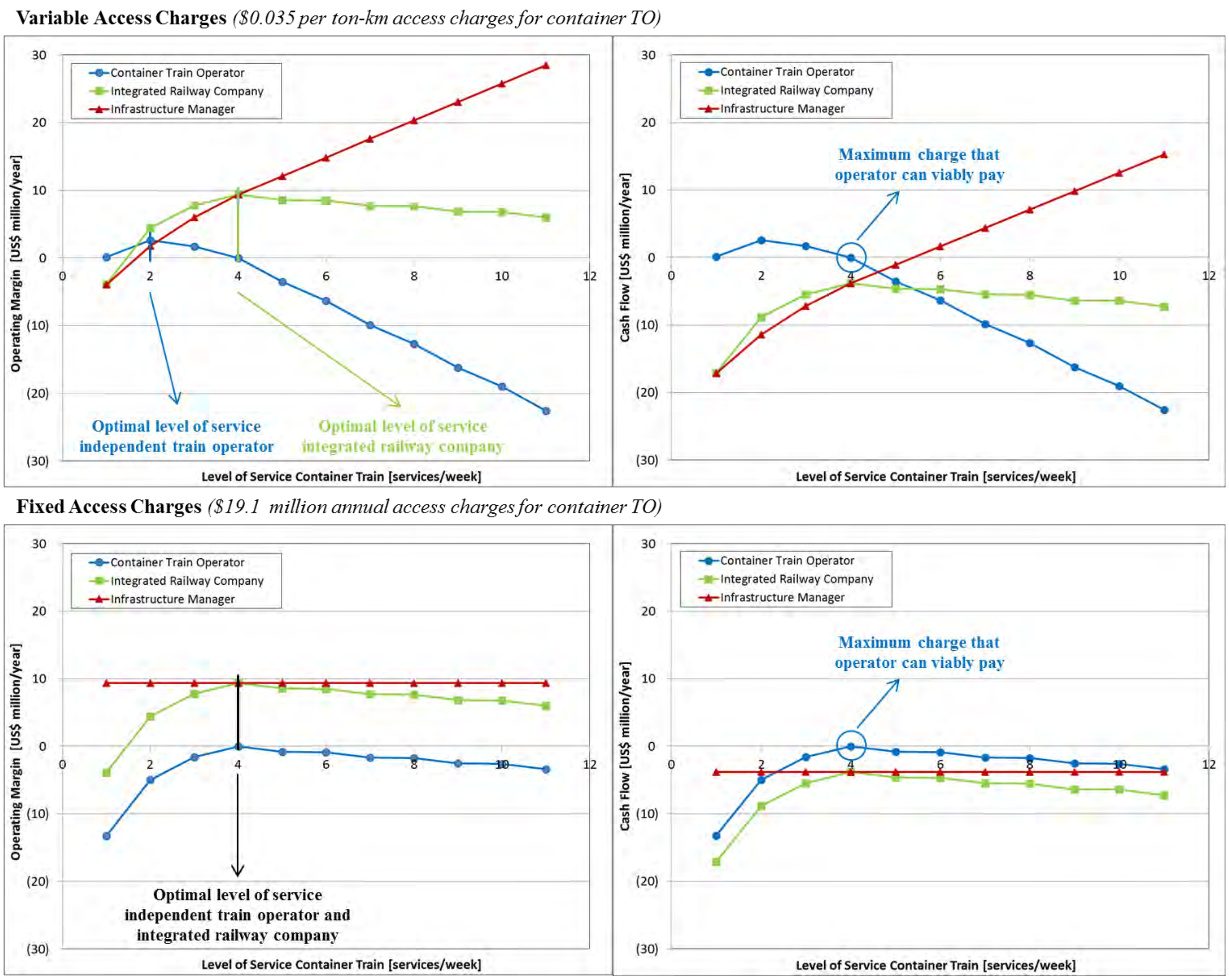


Figure 4-2 Operating margin and cash flow for different levels of service with variable and fixed access charges (Source: author)

4.2.2 Container and General Cargo Traffic

The previous section considers container traffic because container shippers have high willingness to pay to ship containers. Nonetheless, there is plenty unused capacity in the Tanzanian railway system and there are other types of customers interested in transporting non-containerized freight (general cargo) along the corridor. We carried out a similar analysis of costs and revenues for general cargo services (World Bank, 2014) 0.010 per ton-kilometer (variable) or \$10.5 million per year (fixed). In both cases an integrated railway company and a vertically-separated TO would operate ten services per week.

Considering these numbers, the IM would need to charge a variable access charge of \$0.023 per ton-kilometer (variable, assuming the TO would operate four services) or \$12.4 million per year (fixed) to the container TO to recover all infrastructure costs. Note that if the container TO was charged only \$10.5 million per year or \$0.010 per ton-km it would not be able to recover infrastructure costs (only \$21.0 and \$15.9 million per year respectively, assuming that a TO would operate four trains with access charges of \$0.010 per ton-km). This shows, first of all, that discriminate pricing would be needed to recover infrastructure costs. Although a general cargo TO cannot viably pay as much as a container TO per ton to access the infrastructure, allowing access to the infrastructure to general cargo TOs 1) allows the IM to recover infrastructure costs (not possible only with container TOs), 2) allows container TOs to pay lower charges to access the infrastructure, and 3) improves welfare (for general cargo TOs and general cargo shippers) from a state point of view.

Although these charges are consistent with the industry benchmark (World Bank, 2014), a regulator needs considerable information (operational costs, demand estimates) to determine the maximum access charges that each TO is able to pay. Lower charges would not allow the IM to recover infrastructure costs; higher charges (particularly for general cargo in this case) would not allow TOs to viably operate trains in the system.

With a variable access charge of \$0.023 per ton-kilometer, a vertically-separated container TO would only operate three (note that the variable charges are now lower than in Section 4.3.1) train services per direction per week (instead of the four that a social planner would operate). Under fixed access charges, the number of trains operated by vertically-separated TOs in equilibrium matches the number of trains that an integrated railway company would operate.

The main challenge to implement fixed access charges in this case consists of determining the share of infrastructure costs ($22.9 million per year) that each TO should pay. Nonetheless, our computation shows that the number of trains operated by the TOs is robust when the distribution of fixed access charges change: the container and the general cargo TO would be able to pay up to $19.1 million and $10.5 million per year respectively while still being profitable. Any choice such that the annual fixed access charge for the container TO is lower than or equal to $19.1 million, for the general cargo TO is lower than or equal to $10.5 million, and the sum of both charges is $22.9 million would improve number of trains with respect to variable charges while enabling infrastructure cost recovery. This result has important implications: 1) it relaxes the constraint on how much information the regulator needs to determine fixed access charges, and 2) it allows the regulator to design the fixed charge level for TOs with different objectives: such as ensure equity, ensure efficiency, ensure general cargo services.

Under fixed access charges with no variable charges per train, states could implement different mechanisms to allocate operating licenses among potential TOs. First, the regulator could determine a fee (fixed access charge) that a container and a general cargo TO would have to pay to get the license to recover infrastructure costs ($22.9 million per year). If the charges allow the operators to viably operate, they would apply for the license and retain the additional profits ($19.1 or $10.5 million per year minus access charge for each type of TO). Second, when there are several companies willing to operate trains, the state could implement an auction to allocate the license to operate in each market. If the license is awarded to the TO with higher willingness to pay at each market, the most efficient container and general cargo TOs would bid $19.1 and $10.5 million respectively. In this case, the publicly owned IM would obtain $29.6 million per year (instead of $22.9). The IM can either use the additional revenues to invest in infrastructure in the future or transfer them to the government. If the license is awarded to the TOs that offer best freight shipping rate to customers provided that the IM can recover infrastructure costs, the IM would recover \$22.9 million per year, the TOs would recover their costs with some return, and the customers would benefit from a discount in their shipping rate of \$6.7 million per year (\$29.6 minus \$22.9). Table 4-1 summarizes these results. Further options can be explored when more than one license per market (container and general cargo) are allocated.

Table 4-1 Number of container and general cargo trains operated by TOs and revenues raised by IM for different variable and fixed access charges. These numbers are compared with the reference number of trains (operated by a social planner) and the IM costs (Source: author)

variable access charges							
container TO			general cargo TO			IM	
ac_v [\$ ton-km]	n [trains]	n reference [trains]	ac_v [\$ ton-km]	n [trains]	n reference [trains]	revenues [\$m]	costs [\$m]
0.000	4	4	0.000	10	10	0.0	22.9
0.010	3	4	0.010	10	10	14.5	22.9
0.023	3	4	0.010	10	10	19.8	22.9
0.030	2	4	0.010	10	10	18.6	22.9
0.046	1	4	0.010	10	10	16.7	22.9
0.092	1	4	0.010	10	10	22.9	22.9
fixed access charges							
container TO			general cargo TO			IM	
ac_f [\$m]	n [trains]	n reference [trains]	ac_f [\$m]	n [trains]	n reference [trains]	revenues [\$m]	costs [\$m]
up to 19.1	4	4	up to 10.5	10	10	22.9	22.9

4.3 Conclusions

In this chapter, we analyze different capacity pricing mechanisms designed to recover infrastructure costs (periodic maintenance and financial costs) and to ensure that TOs can viably operate (positive net cash flow) in Tanzania. The insights derived from this case are useful to design capacity pricing mechanisms for shared railway systems with infrastructure cost recovery objectives in other countries.

First of all, we show that the adoption of variable access charges widely used in the railway industry may create incentives for rational TOs to operate fewer trains than an integrated railway company (social planner). This is consistent with the results of Chapter 2. We show that the use of fixed access charges aligns the behavior of vertically-separated firms with the behavior of an integrated railway company. This result is important in the railway industry because IMs faced important fixed costs, i.e., for the most part infrastructure costs do not vary with the number of trains operated in the system as is generally assumed.

The results obtained also show that discriminate pricing may be needed to be able to recover infrastructure costs when different types of TOs face very different levels of cost and revenues. The results also show the benefits of sharing the infrastructure among different types of TOs: shared use allows for infrastructure cost recovery.

This case also illustrates that regulators need considerable information about the sector (demand and cost) to determine adequate access charge levels that TOs can viably pay. A benefit of introducing fixed access charges is that the number of trains operated by TOs is robust for a wide range of fixed access fees. This relaxes the regulator information need. The ability to achieve a good level of service with a wide range of fixed access charges also allows the regulators and IMs to design effective pricing mechanisms with very different objectives and with very different implications in terms of the welfare distribution among stakeholders. This chapter also discusses why the use of capacity-based pricing mechanisms would not allow the IM to recover infrastructure costs in non-congested railway systems.

Future work should analyze further how to implement fixed access charges effectively, especially in cases with competing TOs in the same market to avoid barriers to entry. Future research should also determine how these conclusions change with demand uncertainty, elasticity in the demand, and imperfect information.

In this chapter we are able to analyze the capacity pricing problem independently of the capacity allocation problem because the Central Corridor in Tanzania is not congested. Chapter 5 explores the performance of shared railway systems under alternative capacity pricing and allocation mechanisms in instances in which infrastructure capacity is limited and there are important interactions between capacity planning and infrastructure operations.

Chapter 5 - Policy Implications for the Northeast Corridor in the US and Other Congested Shared Railway Systems

"Traffic congestion is caused by vehicles, not by people in themselves."–
J. Jacobs (2002), The Death and Life of Great American Cities

This chapter focuses on the main spine of the Northeast Corridor (NEC) that stretches from Boston, MA to Washington, DC. With over 2,000 trains per day, the NEC is one of the most congested railway corridors in the US. Until now, capacity pricing and allocation in the corridor is managed via bi-lateral contracts. The price that each train operator (TO) pays to access the infrastructure and schedule their trains depends mostly on how much capacity was available when the contract was signed (Gardner, 2013). This imposes two challenges in today's operations: 1) the revenues collected by the infrastructure manager (IM) represent a very small percentage (10%) of the costs in which the IM would need to incur to bring the infrastructure to a state-of-good-repair, and 2) the introduction of new services is extremely complicated. Even if the timetable of some train could be shifted to make room to schedule new trains, rescheduling those trains would require the renegotiation of the contracts. As a result, the FRA required Amtrak and the rest of the NEC commuters and freight railway companies to agree on a new capacity pricing and allocation mechanism (PRIIA, 2008).

NEC stakeholders face two important questions: Which mechanism to price and allocate railway capacity should they implement? What are the implications of such mechanism for each of them and for the overall performance of the system? Chapter 2 points out that in a congested railway systems like the NEC, traditional mechanisms that charge marginal infrastructure costs to

TOs and impose simple priority rules to overcome conflicts may not work. Chapter 1 shows that the implications for the system of alternative mechanisms to price and allocate capacity are still unclear. Furthermore, the interactions between infrastructure planning at the strategic level and infrastructure operations at the tactical level are particularly strong in congested railway systems.

In this chapter, we utilize the framework developed in this thesis to evaluate the performance of the NEC considering both planning and operational aspects. As described above, this framework consists of two models: a TO Model presented in Chapter 2 and an IM Model presented in Chapter 3. The TO Model simulates the behavior of the TOs to determine their demand for scheduling trains on the infrastructure and their willingness to pay for access. The IM Model determines whether that demand can be scheduled in the existing infrastructure. The results obtained are the demand for scheduling trains, the access charges (capacity pricing), and the optimal train timetable (capacity allocation: set of trains scheduled and their timetable).

We then use this information to analyze and compare the performance of a case based on the NEC under two alternative capacity pricing and allocation mechanisms: a price-based cost-allocation and priority-rule mechanism proposed by Amtrak (Crozet, 2004; Gardner, 2013; Nash, 2005; NEC Commission, 2014; Lopez-Pita, 2014; Texeira and Prodan, 2014) and an auction mechanism widely proposed in the railway economic literature (Affuso, 2003; McDaniel, 2003; Newbury, 2003; Perennes, 2014). To understand the implications for various stakeholders, we measure performance from the perspective of the IM (cost recovery), the TOs (access charges, trains scheduled), and the end users (number of services, fares). This chapter focuses on the interactions between intercity and commuter TOs that are responsible for most of the traffic during the very congested peak hours.

The results of this chapter show that there are important trade-offs associated with each these two mechanisms and none of them is superior to the other on all dimensions. We argue that the trade-offs observed cannot be explained solely by the simplifications of this case study. As a result, we recommend NEC stakeholders that they analyze the implications of alternative pricing and allocation mechanisms in detail before locking the system into one of them. Part of this work has been submitted for publication (Pena-Alcaraz, Sussman, Webster, and Perez-Arriaga, 2015b).

The rest of the chapter is structured as follows: Section 5.1 describes the framework inputs and outputs, and discusses how to integrate the TO and the IM Models. Section 5.2 and 5.3 present the results obtained using that framework to evaluate the performance of the NEC under both price-based and capacity-based mechanisms. Section 5.4 compares both mechanisms to price and allocate railway capacity in the context of the NEC. Section 5.5 summarizes the main conclusions of the chapter and identifies lines of future research.

5.1 Shared Railway System Performance Evaluation Framework: Inputs, Outputs, and Model Integration

The NEC (see Figure 5-1) is one of the railway corridors most widely studied in the literature (Archila, 2013; Clewlow, 2012; Kawakami, 2014; Pena-Alcaraz et al., 2013; Sussman et al., 2012; Sussman et al., 2015). However, the implications of new mechanisms to price and allocate railway capacity in this system are still unclear. This section describes the inputs and outputs needed to evaluate the performance of shared railway systems using the framework proposed in this thesis. The main inputs required include both the mechanisms to price and allocate capacity, and inputs of the TO Model and the IM Model. We then discuss how the inputs and outputs of the TO Model and the IM Model relate and how we integrate both models.

Figure 5-1 Northeast Corridor (Source: NEC Infrastructure Master Plan Working Group 2010, from (Sussman et al., 2012))

5.1.1 Mechanisms Selection

As we discuss in Chapter 1, there are three main types of mechanisms to price and allocate capacity: negotiation-based, administrative-based, and market-based mechanisms. The use of market-based mechanisms for capacity pricing and allocation is preferred in systems like the NEC characterized by capacity scarcity (congestion) and conflicting demand (Perennes, 2014; PRIIA, 2008).

According to (Gibson, 2003), the two main types of market-based mechanisms for capacity pricing and allocation for shared railway systems are 1) price-based and 2) capacity based. Price-based mechanisms are those that determine the price at which capacity will be offered, and let TOs decide whether they are willing to access the infrastructure or not. Price-based mechanisms are typically complemented with priority rules that allow the IM to decide which trains to schedule when there are conflicts (multiple TOs willing to pay the predetermined access charges to schedule conflicting services). This chapter studies a traditional price-based cost-allocation mechanism that assigns infrastructure-related costs proportionally to infrastructure use (Crozet, 2004; Nash, 2005; Lopez-Pita, 2014; Texeira and Prodan, 2014) complemented with priority rules for capacity allocation purposes. This mechanism was proposed by Amtrak and is currently being considered for implementation in the NEC (Gardner, 2013; NEC Commission, 2014).

Capacity-based mechanisms are those that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity, e.g. an auction (Affuso, 2003; McDaniel, 2003; Newbury, 2003; Perennes, 2014; Stern and Turvey, 2003). Auction mechanisms have been widely discussed in the literature but have not yet been implemented in any railway system. There are multiple ways to auction the access to the infrastructure (Vazquez, 2003). We could allow TOs to bid for the access to a segment of the infrastructure or for the access to different slots. Auctioning the access to segment presents two problems: 1) the value for a TO of accessing a segment of the infrastructure is contingent on the TO's ability to get access to the rest of the train path, and 2) TOs may engage in strategies to deliberately overbook some segments to restrict access to the system to other TOs, as occurred in France in 2008 (Barrow, 2012). We also know that different types of trains in shared railway systems need different types of slots: commuter trains travel only around urban areas whereas intercity and freight trains travel between cities, and slow trains cannot use fast trains' slots (and vice-versa). Auctioning slots thus requires the IM to predetermine how many types of slots to allocate to each type of service before receiving the TOs' bids. In this chapter we thus assume that TOs bid for the desired timetable (along the whole train path) of the trains they would like to schedule over the next period of time (typically six months), for the access charges they are willing to pay, and for their flexibility to modify the desired timetable to accommodate other conflicting services. We assume that the TOs can only bid once per period (one-round auction) and they would pay the access charges they bid minus any compensation if the desired timetable is modified (first-price auction), i.e., we consider a complex one-round first-price auction. Having second bidding round would allow the TOs to use railway capacity still available after the first bidding round. This chapter studies a first round auction because in such an auction, TOs would have more incentives to reveal their willingness to pay to access the infrastructure. The literature proposes the use of second-price auctions to ensure that the auction is strategy-proof and the TOs reveal their willingness to pay. However, the bids in this auction are complex and combinatorial (as mentioned above, each TO bids for a combination of timetable, access charges, and flexibility). As a result we cannot guarantee that we have enough information from the submitted bids to determine the second best price from a similar service for each train scheduled.

The objective of this chapter is to identify trade-offs in the choice among these two alternative capacity pricing and allocation mechanisms for shared railway systems in the context of congested railway systems, and in particular, in the NEC. This chapter focuses on how the introduction of alternative pricing and allocation mechanisms impacts the ability of intercity and commuter TOs to compete for the access to infrastructure capacity. With over 2,000 commuter trains and 150 intercity trains scheduled in the NEC per day (Gardner, 2013), the ability of commuter and intercity TOs to access the infrastructure has a direct impact on the NEC passengers.

5.1.2 Models' Inputs and Outputs

To use the framework proposed we need information about the system to be able to use the TO and the IM Models. The information required for the models can be collected from the annual TOs' financial reports and the IM's network report. As mentioned before, this is a design choice of both models. A model that allows regulators to anticipate the system reaction to a capacity pricing and allocation mechanism should not require extensive information about the railway system that only the TOs and the IM possess.

The main inputs of the TO Model are the TOs' cost and revenue structure. In terms of the costs, the TO Model aggregates all cost sources into fixed and variable costs. In terms of the revenues, the TO Model uses information about subsidies (if any) and end user's demand. According to (Gardner, 2013), there are currently one intercity TO (Amtrak), eight commuter TOs, and four active freight TOs sharing the infrastructure in the NEC. This chapter focuses on the relation between intercity and commuter services during peak hours. We use data from Amtrak's financial report to model the intercity TO's profit structure. We assume that the profit structures of different commuter TOs are similar, so we use data from MBTA's financial report as a proxy to understand the commuter TOs' cost and revenue streams.

As we mentioned in Chapter 2, an intercity TO like Amtrak operating in the NEC faces fixed operational (direct) costs of $fc = \$281k$ per day and variable operational costs of $vc = \$3{,}425$ per train and day according to (Amtrak, 2014). In 2013, Amtrak's average fare was equal to $f_0 = \$96.5$, the number of trains was $n = 150$ trains per day in average, with a realized demand of $d_0 = 31{,}250$ passengers per day. The average train capacity was $c = 210$ passengers assuming a physical capacity of 250 seats with 85% load factor (Amtrak, 2014). No subsidies are required for the operations of intercity services in the NEC (Amtrak, 2010; Amtrak, 2012). We assume that the demand of intercity trains depends linearly on the fare, with an elasticity of -0.67 (Morrison, 1990).

According to (MBTA, 2013a; MBTA, 2013b) a TO like the MBTA, the commuter TO in the Boston area, faces fixed operational (direct) costs of $fc = \$435k$ per day and variable operational costs of $vc = \$1{,}666$ per train and per day. The elasticity of the demand with respect to the headway (frequency) is estimated by (Lago et al., 1981) to be equal to -0.41. In 2013, MBTA's average fare ranged from $f_0 = \$7 - \$25$ (average fare of $f_0 = \$13$ are considered), the number of trains averaged $n_0 = 485$ trains per day, with a realized demand of $d_0 = 130{,}600$ passengers per day. The train average capacity considered is $c = 350$ passengers, with $80\% +$ load factor. Subsidies $s = \$234k$ per day are considered following (MBTA, 2013a). Commuter TOs in the NEC are subjected to fare regulation, i.e. they cannot change the fares charged to the end users.

The main inputs of the IM Model are the information about the infrastructure and the TOs' demand for scheduling trains. To capture the main infrastructure characteristics of the NEC, we consider the system presented in Figure 5-2. It consists of a double-track corridor with 12 stations. Stations 1 and 7 represent terminal stations at both ends of the line (Boston and Washington DC respectively). Stations 2-12, 3-11, 4-10, 5-9 and 6-8 represent five stations along the corridor. We use a different station number for each traffic direction. Traffic moves in the direction of increasing station numbers with a dedicated track per direction. As a result, traffic traveling in different directions only interacts at the stations. Stations 1, 2 and 12 represent main stations in the Boston metropolitan area, stations 3, 4, 5, 9, 10 and 11 are all in the New York metropolitan area, and stations 6, 7 and 8 are in the Washington DC metropolitan area. Five types of services can be considered: Boston commuter trains traveling around the Boston metropolitan area (stations 1, 2, and 12); New York commuter trains; Washington DC commuter trains; and intercity and freight trains traveling between Boston and Washington DC. Intercity and freight trains may not stop at every station. Intercity trains travel at higher speeds than commuter trains. The distance between Boston and Washington DC is approximately 450 miles, and the distance traveled by commuter TOs around each city ranges from 40 to 70 miles per direction. Note again that although the infrastructure considered is simple and does not include many intermediate stations such as Philadelphia, New Haven, etc., it contains all the important elements to capture the dynamics of the interaction between commuter and intercity traffic under both capacity pricing and allocation mechanisms.

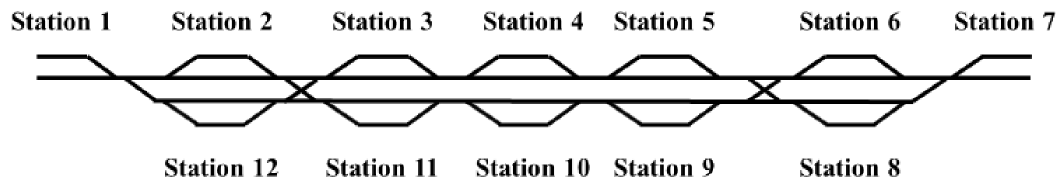


Figure 5-2 Detailed corridor infrastructure (Source: author, Figure 3-2)

The TO Model has two main outputs: the demand for scheduling trains (number of trains that each TO would like to schedule and the access charges that each TO is willing to pay to schedule them) and the fares the TOs would charge to the end users. The IM Model has two main outputs: the final access charges that each TO would pay and the final train timetable (set of trains scheduled and their timetable).

5.1.3 Integration of the Train Operator and the Infrastructure Manager Model

In Section 5.1.2 we mention that the demand for scheduling trains is both an output of the TO Model and an input of the IM Model. There are important differences however between them. Specifically, the demand for scheduling trains considered as an input of the IM Model includes five pieces of information: the number of trains that the TOs ask to schedule, the desired timetable of each train to schedule, the access charges that the TO would pay to schedule each of the trains, information about whether they are flexible to modify the timetable of the train (and by how much) in case of conflict with other train, and how much the IM should compensate them to modify the desired timetable. However the demand for scheduling trains that we get as an output of the TO Model only includes information about the maximum number of trains that each TO would like to schedule and the maximum access charges that each TO is willing to pay to schedule each of the trains.

There are three important observations. First, the outputs of the TO Model are independent of the IM decisions. The TO Model characterizes the maximum number of trains and access charges that the TO would accept. The IM Model would never propose a solution in which the TOs schedule more trains or to pay more than that. In that sense, all the solutions of the IM Model are feasible from the TO perspective.

Second, TOs have incentives to disclose their demand for scheduling trains and their maximum willingness to pay for access to the IM when they do not have market power. In other words, we can use the outputs of the TO Model as inputs of the IM Model. A very interesting line of future research should consider how TOs behave when they have market power and how this behavior will affect the performance of the system following the discussion of Section 5.1.1.

Third, there are three additional IM Model inputs that we cannot get directly from the TO Model: the desired timetable of the trains, how much TOs are willing to change that desired timetable in case of conflict, and how much compensation they should receive for their flexibility to do that. In this chapter we assume that passenger TOs in the NEC are willing to reschedule each train a maximum of 15 minutes and they require a reduction in the access charges of that train equal to 3% per minute modified. These numbers are based on the standards to define passenger train timetables "Introducing a timetable that is easy to remember on the most important lines […] long-distance trains that stops at the main stations only should arrive every 15 minutes" (Kroon et al., 2009), and on the definition of punctuality in in Europe and the US (FRA, 2009; Renfe, 2015). According to (FRA, 2009), an Acela (HSR) train is considered *on-time* if it arrives its endpoint terminal within 10 minutes of the scheduled arrival time. On the other hand, a Northeast Regional (intercity) train is considered *on-time* if it arrives within 10 minutes for trips less than 250 miles, 15 minutes for trips between 251 and 350 miles, and 20 minutes for trips between 351 and 450 miles.

According to (Renfe, 2015), Renfe reimburses 50% of the ticket to HSR passengers that experience a delay of more than 15 minutes.

In terms of the desired timetable, once we know the departure time from the first station, we could use public information from the current timetable to determine the arrival and departure time to all the other stations. In Section 5.3.1 we show that in the NEC if 1) the frequency of commuter trains is higher than or equal to two trains per hour, and 2) TOs prefer to schedule trains uniformly, then the exact desired timetable does not affect the number of trains scheduled nor the access charges that each TO has to pay. This result is based on the fact that a perturbation on the desired timetable of the intercity trains does not affect how many commuter trains would conflict with the intercity train (see Figure 5-4) when the schedule is dense enough (i.e., when the frequency of the commuter is at least 2 commuters per hour). This is a very important result of this research. It suggests that the exact TO's desired timetable affects the final timetable, but it does not affect the number of trains scheduled neither the access charges in many congested shared railway systems. In other words, the exact TO's desired timetable will not be necessary to understand the implications of capacity pricing and allocation in many congested shared railway systems.

5.2 Evaluation of a Price-Based Mechanism: Cost-Allocation and Priority-Rules

In this section we analyze the implications of the price-based cost-allocation and priority-rule mechanism defined in Section 5.1.1 for a shared railway system like the NEC. We first use the TO Model to anticipate the number of trains that a TO would like to operate for different values of variable access charges and determine the resulting TO profits and IM revenues. We then use the IM Model to analyze if the TOs' demand can be scheduled on the infrastructure.

5.2.1 Train Operator Model Results

Figure 5-3 shows the demand for scheduling train services of an intercity TO in the NEC with Amtrak's cost structure and its resulting profits for different variable access charges. Figure 5-4 shows the demand for scheduling train services of a commuter TO in the NEC with MBTA's cost structure and its resulting profits for different variable access charges. For clarity purposes, we do not show the fares in Figures 5-3 and 5-4 although we use them to compute the profits of the TOs. For the intercity TO, the fares increase from $128 to $237 as access charges increase. For the commuter TO, the fares do not change because they are regulated. Figures 5-3 and 5-4 use a distance of 50 miles for commuter trains (Boston, MA to Cranston, RI) and a distance of 450 miles for intercity trains (Boston, MA to Washington, DC) to compute the variable access charges per mile (Gardner, 2013). We first see that the number of trains that the TO would like to schedule decreases as variable access charges increase.

Figures 5-3 and 5-4 show the maximum variable access charges that both intercity and commuter TOs are able to pay. Table 5-1 summarizes this information. In particular, the maximum access-charge that an intercity TO like Amtrak would be able to pay is $102 per train-mile per day, which is equivalent to $46,000 per train per day for Boston to Washington. The maximum variable access charge that a commuter TO like MBTA would be able to pay is $52 per train-mile per day, which is equivalent to $2,578 per train per day. With higher variable access charges the TOs would be better-off not operating any trains. That means that an intercity TO is able to pay two (2) times as much as a commuter TO in a per mile basis, or almost eighteen (18) times as much as a commuter TO in a per train basis. We also determine the maximum sustainable access charges (access charges for which the TOs would have 0 profits after reimbursing capital at an adequate rate of return). That means that their finances allow them to continue operations over the medium term without additional operational subsidies.

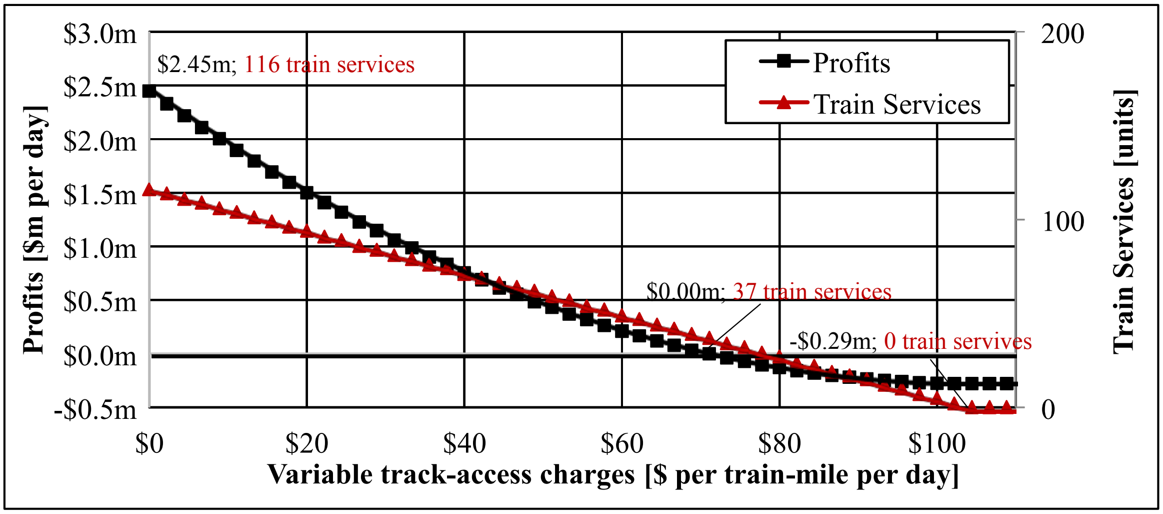


Figure 5-3 NEC intercity TOs expected profits per day and number of trains per day for different variable access charges (Source: author)

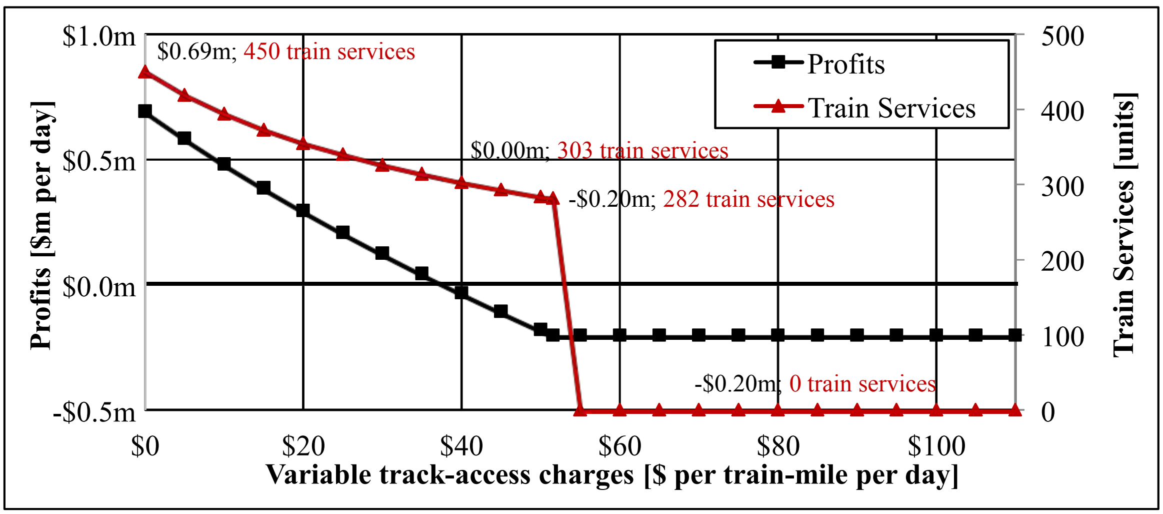


Figure 5-4 NEC commuter TOs expected profits and number of trains per day for different variable access charges (Source: author)

Table 5-1 TOs' expected profits and number of trains to schedule per day for different variable access charges (Source: author)

	reference point	no access charges			maximum access charges			maximum sustainable access charges		
	profits (n=0) [$m]	**ac$_v$** [$ tr-mi]	**N** [trains]	**profits** [$m]	**ac$_v$** [$ tr-mi]	**N** [trains]	**profits** [$m]	**ac$_v$** [$ tr-mi]	**n** [trains]	**profits** [$m]
Intercity	-0.28	0.00	116	2.47	102	1	-0.28	71	37	0.00
Commuter	-0.20	0.00	450	0.69	52	282	-0.20	40	303	0.00

Although both figures only show the response of TOs to variable access charges (per train), we can use them to determine the response of the TOs to fixed access charges too. As discussed in Section 2.2 and equation (2.5), a fixed access charge in addition to the variable access charge would not change the TO operational decisions (number of trains and fares) as long as the resulting TO profits are greater than $s(0) - fc$, (the profits the TO would obtain operating 0 trains). If the resulting profits were lower than the profits operating no trains, a profit maximizing TO would be better off operating no trains (and paying no access charges). Otherwise, the TO would see the fixed access charge as a fixed lump sum that would not change the optimality conditions in the profit equation and as a result, would not change its operational decisions either.

The maximum fixed access charges that the intercity TO would be willing to pay would be $2.75 million per day ($2.47 million + $0.28 million). With this access charge, it would still operate 116 trains, and its profits would be –$0.28 million. If the access charge increases beyond that point, the TO would not operate any trains. Likewise, the maximum fixed access charges that the Boston and Washington DC commuter TOs would be willing to pay are $0.99 million per day ($0.69 million + $0.43 million – $0.23 million). The New York City commuter TO would be willing to pay up to $1.98 million per day (because the commuter trains around New York City travel double the distance than the commuter trains in Boston and Washington DC in the NEC, part in the New York City – Boston and part in the New York City –Washington DC segment). According to (Gardner, 2013), the NEC should invest $51.9 billion (uninflated) from 2010 to 2030 to bring the system to a state of good repair. That means that the IM would need to invest $7.10 million per day. That means that the maximum possible recovery with any mechanisms is $6.71 million per day ($2.75 million + 2×$0.99 million + 1.98 million), i.e., the maximum possible recovery considering intercity traffic and commuter traffic around Boston, New York City, and Washington DC is 95%.

The main advantages of including fixed access charges are that they allow for maximum IM revenue collection while they do not affect the TO's operational decisions (Pena-Alcaraz et al., 2014). However, the use of fixed access charges may create barriers of entry for new competitors. Determining the right fixed access charges for different TOs is also challenging. The rest of this chapter thus assumes that there are no fixed access charges. Note again thought that as shown in Chapters 2 and 4, the use of variable access charges higher than the variable infrastructure maintenance costs would result in operational levels that are suboptimal from a social welfare standpoint.

Table 5-2 shows these same results from the perspective of the IM. According to (Gardner, 2013), a cost-allocation model would allow the IM to charge TOs for the use of the infrastructure. The access charges would depend on the segment in which the trains are scheduled and on their infrastructure needs. The resulting access charges for intercity and commuter TOs would be comparable (as opposed to those of freight trains that do not use passenger stations). According to (Texeira and Prodan, 2014) variable access charges in other countries vary from $0 per train mile (e.g. Estonia and Norway) to $50 - $100 per train mile (in France and Netherlands).

Table 5-2 IM expected revenues for different variable access charges and resulting TOs' demand for scheduling trains per day assuming three commuter TOs (Source: author)

ac_v [$ train-mi-day]	**revenues – IM** [$m]	**n – intercity TO** [trains]	**n – commuter TO** [trains]
currently	0.76	153	458 (x3)
0	0.00	116	450 (x3)
25	2.69	88	340 (x3)
50	4.19	60	284 (x3)
52	4.28	59	282 (x3)
75	1.08	32	0
100	0.23	5	0
125	0.00	0	0

We can see that the revenues that the IM obtains do not always increase when access charges increase, since the TOs' demand for scheduling trains drop. These results suggest that the maximum revenues that the IM would be able to collect from these intercity and commuter TOs is $4.28 million per day when it charges $52 per train-mile. At this point, the intercity TO would be contributing $1.37 million, the Boston and Washington DC commuter TOs would be contributing $0.73 million each, and the New York commuter TO would be contributing $1.45 million (because New York commuter trains travel double the distance than Boston and Washington DC ones as mentioned above). Note that with the current pricing and allocation contracts, the IM revenues are equal to $0.76 million per day, so the IM raises revenues equal to only 10% of the basic infrastructure costs required to bring the system to a state of good repair. The results also show that this price-based mechanism would allow the IM to recover up to 60% of infrastructure costs ($4.28 million / $7.10 million per day).

5.2.2 Infrastructure Manager Model Results

The TO Model anticipates the response of the TOs when each of them optimizes their operational decisions on their own. As a result, the revenues and profit presented in Figures 5-3 and 5-4 and Tables 5-1 and 5-2 assume that the TOs are able to schedule all the trains on the infrastructure. This may not be the case in a congested corridor like the NEC. In this case, however, Table 5-2 shows that TOs' demand for scheduling trains under this mechanism is always lower than the current level of operations (presented in the first row of Table 5-2). In other words, we know that there is enough capacity in the corridor to schedule all the trains that TOs would like to schedule under this price-based mechanism. Consequently, when we use the IM Model in this case, we see that the IM should be able to schedule all the trains in the system if TOs are willing to adjust their desired schedule to accommodate conflicting trains. Note however that TOs with priority to access the infrastructure may not have incentives to be flexible on their trains scheduling preferences. In other words, when there are conflicts since they have priority to access the infrastructure, priority TOs know that their trains will be scheduled independently of their flexibility, whereas the trains of other TOs may only be scheduled if all TOs are flexible. This will thus have a direct impact on the other TOs ability to schedule trains in the system, their profits, and on the total revenues collected by the IM. The IM Model can thus be used to understand the implications of different priority rules for the system. This is an important consideration for this mechanism, because the priority rules grant some TOs priority over others.

5.3 Evaluation of a Capacity-Based Mechanism: Auction

In this section we analyze the implications of the auction mechanism defined in Section 5.1.1 for a shared railway system like the NEC. In this case we use first the IM Model to determine the minimum access charges that an intercity TO have to pay (as a function of the commuter TO access charges) to be able to schedule an intercity train as a function of the commuter frequency. We then use that information and the results from the TO Model to estimate the number of trains that a TO would operate and the access charges it would pay. We use the results of both models to estimate the final TO profits and IM revenues.

5.3.1 Infrastructure Manager Model Results

We start analyzing the optimal capacity allocation plan (train timetable) to determine how to coordinate different TOs' conflicting demands for scheduling trains. Figure 5-5 shows the time-space diagram for the optimal timetable designed using the IM Model in a case in which an intercity TO tries to schedule one train and three commuter TOs try to schedule commuter trains around Boston, New York, and Washington DC every 30 minutes. The y-axes represent distance in miles from station 1 and the x-axes represent time in minutes at which different trains are scheduled to pass through each point of the line (vs. the desired timetable in dashed line). There are no interactions between trains traveling in different directions. The IM Model proposes the final timetable analyzing the trade-off between eliminating trains and readjusting the desired timetable, according to the objective function in equation (3.1). We can use this information to determine how much intercity TOs will have to pay to be able to schedule services that conflict with commuter services. For this example, we need to solve a train timetabling problem with 115 commuter trains and 1 intercity train. For clarity purposes, only the schedules of conflicting trains are shown in Figure 5-5.

As Figure 5-5 shows, the intercity train would initially conflict with 14 commuter trains. Rescheduling the commuter trains to accommodate the intercity service requires that the commuter TOs receive a total discount equivalent to the access charges of 2.1

commuter trains. As a result, the IM would only schedule the intercity train if the intercity TO pays access charges higher than 2.1 commuter trains, i.e., if its bid is higher than 0.33 times that of the commuter TOs per train-mile (considering the miles traveled by intercity and commuter trains). This number does not change with the desired timetable of the intercity train because the intercity train still conflict with 14 commuter trains. The same results are obtained when the frequency of commuter trains is higher than 2 trains per hour. For these frequencies, the number of conflicting trains does not depend on the intercity train desired timetable, nor the total discount that the commuter TOs should receive because the timetable is dense and uniform. That means that we do not need to know the exact desired timetable to determine the relationship between intercity and commuter train bids in the NEC if the frequency of commuters is greater than 2 trains per hour.

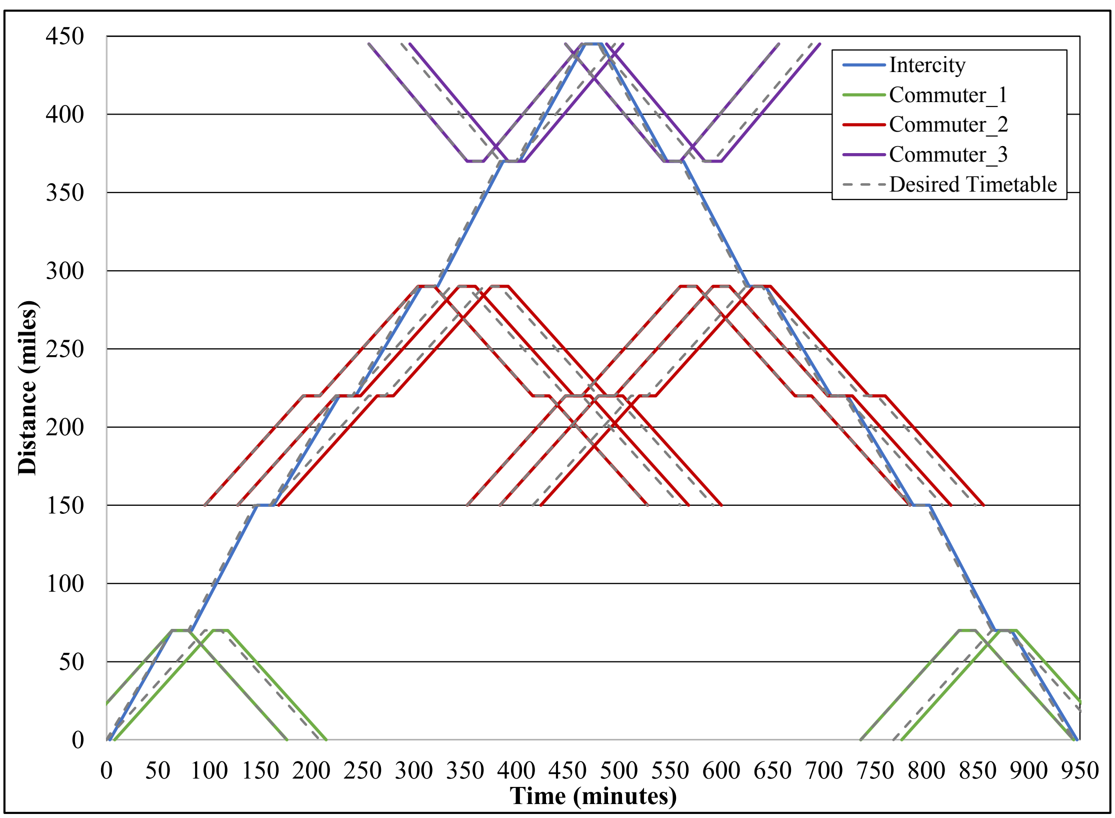


Figure 5-5 Timetable proposed by IM to schedule an intercity train in a system with commuter trains operating every 30 minutes (Source: author, Figure 3-5)

If the frequency of commuter trains increases, for example to one commuter train every 15 minutes instead of every 30 minutes, the intercity train will initially conflict with 22 commuter trains and will only be scheduled if the intercity TO pays access charges equivalent to the access charges of 5.3 commuter trains (i.e., if the intercity TO bid is higher than 0.82 times that of the commuter TOs' per train-mile). Conversely, if the frequency of commuter trains decreases to one train every 60 minutes, the intercity train will be scheduled almost always (the IM would need to compensate the commuter TOs with a total compensation that ranges between 0 and 0.4 times the access charges paid by a commuter trains depending on the desired timetable, what translates in bids higher than 0.00 or 0.06 times that of the commuter TOs per train mile). The model can be used to quantify the trade-off between commuter and intercity trains for any other frequency of service (see Table 5-3).

Table 5-3 Minimum intercity to commuter access-charge per train-mile bid ratio to ensure that their train is scheduled as a function of the commuter frequency (Source: author)

commuter frequency [minute]	**ratio** [per train-mi]	**commuter trains scheduled** [%]
5	5.86	73%
10	0.84	100%
15	0.82	100%
30	0.33	100%
60	0.00-0.06	100%

Table 5-3 shows that, when the system is congested, an intercity TO may have to pay more than a commuter TO to schedule a train: the intercity TO has to pay between 0.82 and

5.86 times the access charges of commuter TOs per train-mile or between 5.27 and 37.67 times the commuter TOs' access charges per train. This minimum intercity access charge reflects the congestion rents. The results show that greater cost recovery is expected in congested infrastructure. The frequency of commuter services in the NEC today ranges between 10 and 30 minutes. A first analysis of the results suggests that, with current levels of traffic, intercity TOs have an advantage over commuter TOs to access the infrastructure under this capacity-based mechanism.

5.3.2 Train Operator Model Results

How can we know whether commuter TOs would be able to compete to access the infrastructure with the intercity TO? Fortunately, we can use the TO model (equation (2.4)) to determine and compare the maximum access charges that each TO would be willing to pay as a function of the number of trains they want to schedule. Figure 5-6 summarizes these results, the maximum variable access charges that an intercity and a commuter TO with the cost structure of Amtrak and MBTA respectively could bid as a function of the number of trains to schedule. Similarly to the cost-based approach, the results show that the intercity TO ability to pay to access the infrastructure is almost double that of the commuter TO counterparts. Note that the TOs' willingness to pay to access the infrastructure decreases with the number of train services with the exception of the commuter TOs' willingness to pay when they schedule between 0 and 280 train services. This happens because end users' demand for commuter services is elastic to the frequency and increases substantially when more trains are scheduled (or decreases substantially when the frequency of commuter trains is very low). As a result, commuter TOs have incentives to ensure a minimum service frequency.

We need to make one adjustment before we use these results as inputs of the IM Model. The TO Model assumes that all the trains have the same OD pair. However, the 150 intercity services that (Gardner, 2013) mentions, include for instance Boston to New York services and New York to Washington DC services that we count in the IM Model as a single service. They also average the number of trains during the day without considering differences between peak and off-peak hours. We use the following equivalences between frequency and number of trains: 118 intercity services in the TO Model are equivalent to 1 train per hour in the IM Model and 450 commuter services in the TO Model are equivalent to 6 commuter trains per hour in the IM Model (Amtrak, 2014; Amtrak, 2015; MBTA, 2013a; MBTA, 2015).

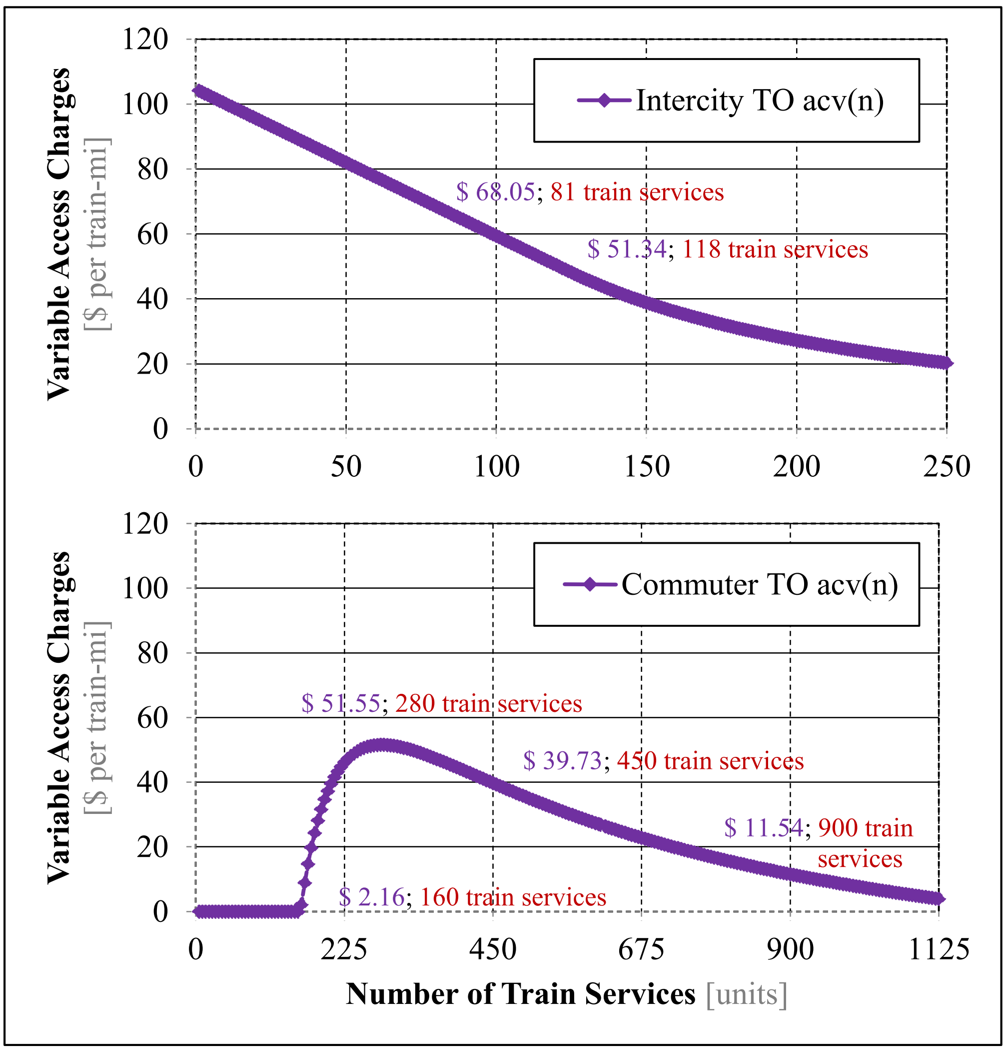


Figure 5-6 NEC intercity and commuter TOs maximum willingness to pay for access (maximum variable access charges) as a function of the number of trains to schedule (Source: author)

Table 5-4 shows the result of combining the information in Table 5-3 and Figure 5-6. The first three columns show the bids of each commuter TO as a function of the desired frequency (number of trains to schedule). The next three columns show the bid of an intercity TO that tries to schedule 1 train per hour. The last two columns determine how many trains of each type can be scheduled and compute the resulting revenues for the IM (again, assuming three commuter TOs, one in the Boston area, one in the New York City area, and one in the Washington DC area).

Table 5-4 TOs' demand for scheduling trains for different variable access charges and resulting IM expected revenues per day assuming three commuter TOs (Source: author)

commuter frequency [minute]	**n commuter TO** [trains]	**ac_v commuter TO** [$ train-mi]	**n intercity TO** [trains]	**ac_v intercity TO** [$ train-mi]	**n commuter n intercity** [trains]	**revenues IM** [$m]
5	900	12	81	68	657, 81	4.00
10	450	40	118	51	450, 118	6.30
15	300	51	118	51	300, 118	5.80
16	280	52	118	51	280, 118	5.61
28	160	2	118	51	160, 118	2.80
30	150	0	118	51	0, 118	2.73

Considering that the NEC infrastructure needs in the NEC amount to $ 7.10 million per day from 2010 to 2030, the results for this mechanism show that the IM would be able to recover up to 89% ($6.30 million / $7.10 million) of infrastructure costs with intercity and commuter train services around Boston, New York City and Washington DC. This is a substantial recovery ration considering 1) that currently the IM only recovers 10% of infrastructure costs and 2) that the maximum potential for recovery with these services amounts to 95% as mentioned in Section 5.2.1.

Note that the intercity to commuter TO bid ratio exceed the ratio presented in Table 5-3 when the commuters frequency (headway) is bigger than 5 minutes. In other words, the intercity TO is almost always able to schedule all the intercity services. These results confirm that the intercity TO in the NEC is usually in better position to access the tracks than the commuter TO under an auction mechanism with current levels of service. If the frequency of commuter trains were to increase by 85%, with 5 minutes headways, the intercity TO would not be able to schedule trains if it bid less than $68 per train-mile (5.86 x $12 per train-mile). Using the TO Model, we can determine that the intercity TO is able to bid over $68 per train-mile if its demand for scheduling trains is equal to 81 trains or less.

In this case, the commuter TO would only be able to schedule 657 trains (73% of 900). This equilibrium is stable because none of the TOs would want to schedule more trains. As Figure 5-6 shows, the commuter TO would be willing to pay higher access charges for 657 trains than for 900 trains. As a result, scheduling only 657 trains at the 900-train access charge level translates to extra profit for the TO. Similar results are obtained for the intercity TO bids and for all other commuter TO bids with more than 280 trains. Between 160 and 280 commuter trains the equilibrium is not stable because the demand for commuter services would significantly decrease due to the amount of service reduction, and also the commuter TO profits when not all trains are scheduled.

5.4 Comparison of Price-Based and Capacity-Based Mechanisms in the Northeast Corridor

The previous sections discuss the operational decisions of intercity and commuter TOs under a price-based (cost-allocation and priority-rule) mechanism and a capacity-based (auction) mechanism in the NEC. The results are summarized in Tables 5-2 and 5-4 respectively.

Table 5-5 shows the number of trains that each TO scheduled and its profits for different access charges, together with the revenues raised by the IM under both mechanisms. Although there is not a one-to-one comparison between both mechanisms, we can compare both sides of the table.

Table 5-5 Distribution of revenues (rev., in \$m), profits (π, in \$m) and number of trains scheduled (n) per day comparison for different access charges (ac_v, in \$ per train-mile) (Source: author)

Price-Based Mechanism							Capacity-Based Mechanism						
intercity TO			commuter TO			IM	intercity TO			commuter TO			IM
ac_v	n	π	ac_v	n	π	rev.	ac_v	n	π	ac_v	n	π	rev.
0.0	116	2.5	0.0	450	0.7	0.0	68.1	81	-0.3	11.6	657	0.2	4.0
25.0	88	1.3	25.0	340	0.5	2.7	51.3	118	-0.3	25.0	640	-0.2	5.9
50.0	60	0.5	50.0	284	-0.2	4.2	51.3	118	-0.3	50.0	325	-0.2	6.0
51.3	59	0.4	51.3	282	-0.2	4.3	51.3	118	-0.3	51.3	297	-0.2	5.8
51.6	59	0.4	51.6	282	-0.2	4.3	51.3	118	-0.3	51.6	280	-0.2	5.6
53.3	57	0.4	53.3	0	-0.2	1.4	51.3	118	-0.3	51.6	280	-0.2	5.6

These results show that the revenues collected by the IM under the capacity-based mechanism studied are higher than the revenues collected when a price-based mechanism with similar charges for intercity and commuter TOs is implemented. The capacity-based mechanism does not only allow the IM to collect higher revenues (around 20% more than using the price-based mechanism), but also results in higher number of trains scheduled (20% more) and higher total welfare (also 20% higher) as compared to the price-based mechanism. Note that these differences are significant considering the robustness of the TO Model (discussed in Chapter 2) and the precision of the IM Model (discussed in Chapter 3). These advantages, however, have a cost for the TOs, who receive much lower profit (losses in most cases). Consequently, the results of the auction mechanism are not resilient to uncertainty in transportation demand, since the TOs would prefer not to operate trains if the profits decrease even further. This auction mechanism may thus require the design of a procedure to redistribute revenues to ensure that TOs can sustainably operate over the medium term.

5.4.1 Response to Uncertainty

It is important to note that both the TO Model and the IM Model considered in this research are deterministic. They intend to capture the essence of a normal day of operations. However, the TOs and the IM face several sources of uncertainty when making their operational decisions. The two most important sources of uncertainty in the NEC in the medium term are 1) the condition of the infrastructure, 2) the end users' demand for transportation services.

The first source of uncertainty is particularly critical until the NEC reaches a state of good repair that ensures that the infrastructure is reliable. With today's backlog of maintenance work, the need of last-minute maintenance and interventions has a direct impact on the capacity on the corridor and on the TOs ability to schedule trains. This uncertainty gets amplified under price-based mechanisms. Any problems with infrastructure availability under price-based mechanisms would reduce the number of services operated by TOs that often operate the minimum number of services that allows them to be profitable. This lack of infrastructure capacity unevenly affects those TOs with lower priority assigned. Under capacity-based mechanisms, the TOs will still make profits even if some trains are not scheduled due to infrastructure availability problems. The IM under the auction mechanism studied would have important incentives to avoid uncertainty on the infrastructure capacity availability, since fewer trains scheduled would lower its ability to recover infrastructure costs as compared to the deterministic case.

The uncertainty in the demand for transportation has a major impact on the expected revenues that the TOs would collect. The fact that the price-based mechanism ensures high TO profits makes this mechanism more resilient to demand uncertainty than the auction mechanism, where the TOs operational decision will probably change if the TOs expect a very uncertain demand. Note that there is also uncertainty in the demand distribution. Passengers do not arrive homogeneously during the day; they instead concentrate around some particular times. As a result we may expect TOs scheduling some more trains than the ones that the model indicates. For example, while the commuter TO Model shows that the optimal number of trains to schedule is 397, MBTA currently runs 485 trains on the infrastructure. Although the model already considers a load factor of 80% to accommodate part of this demand, scheduling 485 commuter trains would result in an average load factor of 65% (industry benchmark for commuter services). This would result in a higher operational cost to accommodate the same demand. The need to offer 485 vs. 397 trains would depend on the exact distribution of the demand. The load factor of the model can be adjusted to consider this uncertainty. This uncertainty will propagate to the expected profits. As a result, we may expect to see a lower TOs' willingness to pay to access the infrastructure even for the same number of trains to schedule, and hence a lower IM ability to recover infrastructure costs than in the deterministic case.

5.4.2 Mechanism Implementation

In terms of implementation, price-based mechanisms are easier to implement than capacity-based mechanisms. As mentioned above the priority rules have important implications for NEC commuter and intercity passengers and TOs, especially if the IM does not have access to sophisticated methods to solve the train timetabling problem. Nonetheless, determining which trains to schedule under a price-based mechanism, once the priority rules and the demand for scheduling trains are known, is easy to understand for all stakeholders.

However, the implementation of an auction mechanism *requires* the IM to be able to solve the train timetable problem proposed in Chapter 3 to ensure transparency in the capacity allocation process. As we discussed in Chapter 3, solving the train timetabling problem is difficult in railway systems with large numbers of trains and stations. Note also that understanding the implications for infrastructure utilization of capacity rules also *requires* the IM to solve the optimal train timetable given the operators demand for scheduling trains. In other words, being able to solve the train timetabling problem is critical to evaluate both price-base and capacity-based mechanisms.

5.4.3 Other Considerations: Gaming the Mechanisms

As we mentioned before, these results also assume that TOs do not have market power or do not take advantage of their power to game the mechanism to their interest and disclose their willingness to pay to access the infrastructure. However, TOs often have market power in the railway industry. The results of this chapter show, for instance that commuter TOs' ability to pay for access is much lower than the one of their intercity counterpart. This is particularly important under auction mechanisms where the TOs have incentives to keep lowering their bids while their trains get scheduled to maximize their profits. In particular, intercity TOs could use the framework proposed in this thesis using only publicly available data to replicate the results of Table 5-3 to understand that they could get their trains scheduled paying much less than their maximum willingness to pay. These aspects are beyond the research of this thesis. Note in any case, that the framework proposed in this chapter would allow the regulator to anticipate the results of the auction and to investigate any variation with respect to these numbers.

5.5 Conclusions

This chapter shows how to use the framework developed in Chapters 2 and 3 of this thesis to evaluate the performance of a congested shared railway system based on the NEC under two alternative capacity pricing and allocation mechanisms considering both planning and operational aspects. The two alternative capacity pricing and allocation mechanisms evaluated are a cost-allocation and priority-rule mechanism, which was proposed by Amtrak and is currently being considered by the NEC Commission and the different TOs in the NEC, and a capacity based (auction) mechanism. Section 5.1.3 discusses how to integrate both the TO Model and the IM Model to be able to anticipate the demand for scheduling trains, set the access charges (capacity pricing), and set the final train timetable (capacity allocation: set of trains scheduled and their timetable).

The results of the chapter show that there are important trade-offs between the mechanisms analyzed. The capacity-based mechanism studied results in almost 20% more IM cost recovery and 20% more trains scheduled as compared to the price-based mechanism studied in the NEC. However, it also results in lower profits for the TOs. The price-based mechanism, on the other hand, is easy to implement (does not requires the IM's ability to solve a sophisticated the train timetabling problem) and ensures higher profits for the TOs, making the TOs more resilient to uncertainty in end users' transportation demand. Note again that this comes with a cost to end users (who will have fewer trains) and to the IM (who will obtain fewer revenues from access charges). The price-based mechanism is not very resilient to uncertainty in infrastructure capacity availability. Under the capacity-based mechanism, intercity TOs are in better position than commuter TOs to access the tracks with current NEC levels of service. The priority level of each TO is a design choice in price-based mechanisms, but this choice has important implications for NEC commuter and intercity passengers and TOs that need to be benchmarked against the optimal capacity allocation determined with the IM Model.

This analysis also benchmarks the IM cost recovery of the price-based mechanism and the capacity-based mechanism studied, 60% and 89% respectively, with the maximum IM cost recovery achievable with any capacity pricing and allocation mechanism and the current IM cost recover, 95% and 10% respective. This shows that 1) both mechanisms allow for much greater IM cost recovery than the current capacity pricing and allocation mechanism implemented, and 2) the capacity-based mechanism designed IM cost recovery level (89%) is very close to the maximum IM cost recovery achievable in the NEC (95%) considering intercity services and commuter services around Boston, New York City and Washington DC. Slightly higher IM cost recovery numbers could be achieved considering other commuter services in the NEC (Philadelphia, Connecticut) and freight services too.

To our knowledge, this is the first research that compares the performance of price-based and capacity-based mechanisms in the same railway system. The results show that neither of these two mechanisms is superior to the other on all dimensions. A better understanding of these trade-offs is necessary to design effectively coordinate shared railway systems. We believe that the stakeholders in the NEC should carefully analyze the implications of alternative pricing and allocation mechanisms before locking the system into one of them.

Although this chapter focuses on the interactions between intercity and commuter TOs, the framework proposed is valid to analyze other questions such as the implications of the mechanisms for freight TOs, for the end users, or for the whole system. Future research should consider the variety of services operated in the NEC (services with different speeds and stops, serving different OD pairs), the distribution of the revenues collected by the IM among the different NEC infrastructure owners, and the effects of TO's market power to refine the understanding of the trade-offs among alternative pricing and allocation mechanisms. Future research should also analyze how these results change in the context of other congested and non-congested shared railway systems.

Chapter 6 - Conclusions

"If we knew what we were doing, it would not be called research." – attributed to A. Einstein

In this thesis we developed a framework to analyze the performance of shared railway systems under alternative capacity pricing and allocation mechanisms. We then use this framework to understand the implications of representative capacity pricing and allocation mechanisms in the Central Corridor in Tanzania and the Northeast Corridor in the US. In this chapter Section 6.1 summarizes the work presented in this thesis and the main findings of our research, Section 6.2 presents the main conclusions and recommendations, and Section 6.3 describes possible directions for extensions and further research.

6.1 Summary of Thesis

Recently, governments have started promoting the use of shared railway systems. Shared use allows for a more efficient utilization of existing railway infrastructure but requires a strong coordination between the infrastructure manager (IM) and the train operators (TOs). Such coordination, in turn, requires capacity planning mechanisms that determine which trains can access the infrastructure at each time, capacity allocation, and the access price they need to pay, known as capacity pricing. This is particularly challenging in the railway industry because there are strong interactions between capacity planning and infrastructure operations.

We start in Chapter 1 by presenting a literature review of alternative capacity pricing and allocation mechanisms. We draw two important conclusions in relation to the existing research in shared railway systems at a macroscopic level. First, we conclude that capacity pricing and allocation mechanisms used for coordination purposes in shared railway corridors are getting more heterogeneous. Second, we observe that different mechanisms are evaluated using different metrics. As a result, the comparative performance of different mechanisms to price and allocate railway capacity is still unclear. This thesis aims to help fill this literature gap by 1) developing a framework to evaluate the performance of shared railway systems under alternative capacity pricing and allocation mechanisms (Chapters 2 and 3), and 2) using this framework to understand the implications of representative capacity pricing and allocation mechanisms in the Central Corridor in Tanzania (Chapter 4) and the Northeast Corridor (NEC) in the US (Chapter 5). This thesis does not answer the question of which capacity pricing and allocation is best, because there is no unambiguous answer. The selection of the most appropriate mechanism to price and allocate railway capacity depends on the shared railway systems goals.

Chapter 2 presents a TO Model that anticipates TOs' demands for infrastructure use under alternative capacity pricing and allocation mechanisms. The focus of this thesis is in the interactions between railway infrastructure operations and available infrastructure capacity. As a result, the TO Model proposed is simple by design. The main objective is to allow the regulator and the IM to anticipate the response of the TOs relying only on information that is readily available to them. Nonetheless, the structure of the model is consistent with the standard financial models of TOs used to analyze investments in the railway industry. The results obtained in Chapter 2 show that the TO demand for infrastructure use estimates are robust to model inputs. In other words, the TOs' demand for using the infrastructure does not change much with small changes in the inputs of the model (cost and demand estimates). This suggests that the level of detail of the model is adequate to capture the interactions between the TOs and the IM. This model analyzes each TO independently of other TOs. Once we know TOs' demand for scheduling trains, we need to determine if there is capacity available to schedule all the services. We deal with this question in Chapter 3.

Chapter 3 presents train timetabling problem for shared railway systems that determines which of these trains can be scheduled on the tracks considering the topology of the line, safety constraints, and other technical aspects of the infrastructure. This model explicitly captures the interrelation between infrastructure operations and available infrastructure capacity. From a computational standpoint, the size of the IM Model increases rapidly with the number of stations and the number of trains to schedule. This thesis proposes a novel solution algorithm based on a linear programming approach to approximate dynamic programming (QARLP algorithm) to be able to solve the problem in meaningful instances. This algorithm allows us to decompose and solve large problems that are intractable with MILP commercial solvers while still converging to a solution within an optimality gap.

The economic literature have long suggested the use of traditional mechanisms that price capacity with the marginal infrastructure costs and use simple priority rules to allocate capacity in case of conflict. In Chapter 2 we compare the operational decisions of vertically-integrated railway systems with those of a vertically-separated profit-maximizing TO. We show that the operational decisions of a profit-maximizing TO match the decisions of an even-handed railway industry regulator when variable access charges reflect variable IM costs to operate trains on the infrastructure. These results justify the use of traditional mechanisms to price and allocate capacity that have been adopted in most countries.

However, Chapter 2 also shows that there are two cases in which these mechanisms cannot be used: 1) when the IM needs to recover part of the infrastructure management fixed costs as it happens in Tanzania, or 2) when the railway system is congested as it happens in many US corridors. In fact, most railway systems fall into at least one of these two categories. This motivates the choice of the two case studies of this thesis and the use of the framework developed to understand the trade-offs associated with the use of alternative mechanisms in these cases. To our knowledge, this is the first effort to compare alternative mechanisms to price and allocate capacity in the same shared railway system.

We illustrate the first case in Chapter 4 in the case of the Central Corridor in Tanzania. Due to the low number of trains operated in the system today, infrastructure maintenance costs do not increase (for all practical purposes) when more trains are operated. Therefore, maintenances costs are assumed fixed. If access charges are set following the traditional approach, operators would not have to pay to access the infrastructure. However, it is critical to ensure that the IM is able to raise revenues to maintain the infrastructure and keep the system operational. As a result, the IM has to allocate infrastructure fixed costs among TOs through the access charges. Chapter 4 first shows that the introduction of variable access charges distorts the operational decision of TOs and then discusses how to avoid this problem with other price-based mechanisms such as the introduction of fixed access charges. We also discuss how to allocate infrastructure costs among different types of TOs and conclude that charging different access charges to different types of TOs is beneficial for all the stakeholders. Chapter 4 also discusses why the introduction of capacity-based mechanisms in non-congested shared railway systems does not allow the IM to recover costs.

Chapter 5 then analyzes alternative capacity pricing and allocation mechanisms in the context of the very congested NEC, in the US. In this case, we need both the TO Model to anticipate how each TO will respond to capacity pricing and allocation mechanisms and the IM Model to determine the final allocation of infrastructure capacity. Until now, capacity pricing and allocation in the corridor has been managed via bi-lateral contracts negotiated between the IM and the TOs. However, the limitations of this negotiation-based mechanism motivated the FRA's request to Amtrak and the rest of the commuters and freight railway companies to agree on a new capacity pricing and allocation mechanism by the end of 2015. Chapter 5 considers two representative mechanisms to price and allocate railway capacity: a price-based (cost-allocation and priority-rule) mechanism proposed by Amtrak and a capacity-based (auction) mechanism. The results of Chapter 5 show that there are important trade-offs associated with each mechanism and none of them is superior to the other on all dimensions. NEC stakeholders should carefully analyze the implications of alternative capacity pricing and allocation mechanisms before locking the system into one of them.

6.2 Conclusions and Recommendations

At the beginning of this thesis we mentioned that the implementation of shared railway systems requires the design and implementation of capacity pricing mechanisms. These mechanisms are the rules needed for coordinating the multiple agents that share the infrastructure. In this thesis, we analyze the performance of shared railway systems under alternative mechanisms to price and allocate railway capacity. This section summarizes three main conclusions of this work and recommends some courses of action bases on these conclusions.

The first conclusion of this research is that the implications of capacity pricing and allocation mechanisms for shared railway systems are still unclear. While this thesis tries to offer clarity in this area, there is still much work to be done. In that sense, we join (Drew and Nash, 2011; Nash, 2010) in recommending to academics that they invest in this research topic. Any progress in research that contributes to a better understanding of the implications of alternative mechanisms to price and allocate capacity could immensely help practitioners and policy makers. This is particularly important in a context in which several countries are currently restructuring their railway sector to allow shared use.

The second conclusion of this research is that sharing railway infrastructure capacity is not straightforward. In the railway industry, as compared to other network industries, there are very strong interactions between capacity planning and infrastructure operations. Chapter 5 shows that capturing this interactions is critical to implement capacity-based mechanisms and to understand the implications of price-based mechanisms in the railway industry. Despite these differences, regulators and policy makers rely on the lessons learned from other network industries. Although these lessons are useful and can serve as guiding principles to design mechanisms to price and allocate railway capacity, Chapter 2 shows that they often do not work in the railway industry. We thus recommend that policy makers are cautious and question the validity of assumptions based on other industries. We also recommend to academics that they reach other communities beyond their domains doing research in these topics. A better understanding of what works and what does not work across network industries and why would also be very valuable for practitioners and policy makers.

The third conclusion, on a more positive note, is that the implementation of adequate capacity pricing and allocation mechanisms can mitigate the coordination problems of shared railway systems while maintaining the benefits of shared infrastructure use in the railway industry. Chapter 2 shows that the introduction of TO competition enabled by shared use may have similar effects to the introduction of regulation to ensure that TOs behave as even-handed integrated railway companies. Chapter 4 shows that shared use may allow the IM to recover more infrastructure costs than those recovered in dedicated corridors by enabling the entrance of new TOs that offer profitable services that the current TO does not provide. In a context in which the NEC and many other systems are moving ahead with the implementation of new capacity pricing and allocation mechanisms, we conclude with three more recommendations. This research shows important trade-offs among alternative mechanisms to price and allocate railway capacity. We recommend the use of the framework developed in this thesis to identify *personalized* mechanisms to price and allocate capacity, aimed at the specific characteristics of the systems. At the same time, we recommend that practitioners and policy makers consider alternative mechanisms to price and allocate railway capacity before locking their system into one of them. Even if those cases where stakeholders have to make a decision soon, we recommend that they allow for some flexibility to adapt the mechanism implemented as they gather more information and better understand the implications of alternative mechanisms for their systems. Finally, we recommend railway companies and regulators that they measure the performance of their systems using a wide variety of performance metrics and to share information and best practices with other railway systems. The data form these experiments will contribute to the improved understanding and management of shared railway systems.

6.3 Future Research

In this section we identify four lines of future research that we find particularly relevant to better understand the implications of capacity pricing and allocation mechanisms in shared railway systems. These lines include 1) additional validation of the framework developed, 2) the extension of the models and algorithms proposed, 3) the utilization of the framework developed to answer other related and relevant shared railway systems research questions, and 4) the development of a broader understanding of capacity pricing and allocation across network industries.

There are three different ways to validate the framework developed in this thesis. First we could validate the models with the results obtained. This thesis uses current operational data to calibrate the models. This prevents us from using that same data to validate the models. Nonetheless, this thesis extensively discusses the results obtained, compares it with industry benchmarks, and carries out multiple sensitivity analysis to understand the sensitivity of the results to the models' inputs. Second, we could validate the models comparing them to other models already validated in the field. In that sense, the TO Model presented is based on a TO financial model used extensively to analyze investments on the railway system. The IM Model presented is also based on train timetable models widely adopted in the railway industry. Third, we could further validate the models using them to analyze additional capacity pricing and allocation mechanisms in other shared railway systems. In this direction, the author collaborates with other students in the MIT Regional Transportation and HSR Research group that are using a similar framework to analyze the performance of shared railway systems in California and in Europe. California is working towards the implementation of a *blended* HSR system in which commuters and high-speed trains will share the infrastructure in San Francisco (Levy, 2015). Many European countries have experienced changes in the mechanisms to price and allocate railway capacity in the last 10 years (Prodan, 2015) and thus represent excellent natural experiments to calibrate and further validate the framework proposed in this thesis. The author has also participated in studies to implement shared railway systems in Tanzania and India in collaboration with the World Bank. Future data from any of these projects would be useful to further validate this framework and to improve our understanding of the implications of capacity pricing and allocation mechanisms for shared railway systems.

The results of the thesis show that there are important trade-offs between alternative mechanisms. Consequently, we recommend shared railway systems' stakeholders that they carefully analyze the implications of alternative capacity pricing and allocation mechanisms before locking their system into one of them. We discuss three possible directions to extend the models and the algorithm developed in this thesis to further analyze these issues, together with the main challenges that we envision:

- The TO Model proposed can be extended to capture more institutional, technical, and operational details of the shared railway systems. This extensions could consider, for instance, the effects of modeling multiple OD pairs, substitutable services, different time periods during the day, more detailed demand models, minimum number of service constraints imposed by Public Service Obligations, maximum number of services due to rolling stock limitations, funding availability, etc. As we discuss in Chapter 2, these extensions would require more information about TOs' costs and revenues and about the end users' transportation demand. However, they would also allow us to anticipate TOs' operational decisions with more accurate estimates of the train capacity and the interactions between services. The extensions discussed would not impose a challenge to solve the underlying TO Model optimization problem and determine the TO operational decisions. This thesis solves this problem analytically. Most of the extensions discussed would facilitate the determination of the operational decisions by imposing bounds on the feasible space. Others may require the use of numerical optimization methods, but there is ample room to add complexity to this model from a computational standpoint. This thesis focuses on the TO-IM relation to capture the interactions between infrastructure operations and available infrastructure capacity. These extensions would connect this work with a broad field of research that studies the details of the end users-TO relation (Bebiano et al., 2014; Ben-Akiva and Lerman, 1985; Lago et al., 1981; Morrison, 1990).

- While the IM Model proposed is fairly comprehensive, this thesis uses it to analyze the allocation of capacity in simple instances (the infrastructure details are aggregated and only consider a few stations). The IM Model can be easily parameterized to consider the topology of the infrastructure in more detail. However, the size of the model increases rapidly with the number of stations and services. The instances analyzed in this thesis are at the limit of what current MILP commercial solvers can handle. The two main options to handle this increased dimensionality are 1) to use the timetables in simple instances as a starting point to develop more detailed timetables and 2) to use the novel algorithm proposed in Chapter 3 to solve the resulting optimization problem.

- Finally, in this direction, the algorithm proposed in Chapter 3 has proven promising to solve large scale network optimization problems. Chapter 3 shows that choice of the basis functions takes advantage of the problem structure and is thus problem specific. Further research to study this algorithm, to develop principles to choose the basis functions, and to incorporate the algorithm in commercial packages that could be used by practitioners would be extremely useful to efficiently solve the train timetabling problem and other large network optimization problems.

This thesis presents a framework that allows regulators, IMs, and TOs to analyze and compare alternative capacity pricing and allocation mechanisms. This framework could be used in the future to analyze two other related and relevant aspects of shared railway systems. From a prescriptive standpoint, the ability to analyze and understand alternative mechanisms to price and allocate railway capacity is critical to design effective mechanisms to coordinate shared railway systems. From a descriptive standpoint, the extension of the results of this thesis considering that TOs in railway systems generally have market power is also critical to analyze, compare, and design capacity pricing and allocation mechanisms.

There is a broad field of research that analyzes the design of mechanisms to price and allocate capacity in network industries (Affuso, 2003; Greve and Pollitt, 2013; McDaniel, 2003; Newbury, 2003; Parkes, 2001; Perennes, 2014; Stern and Turvey, 2003; Vazquez, 2003). The desirable properties of a good mechanism to price and allocate railway capacity are: strategy proofness, allocative efficiency, individual rationality, budget balance (Greve and Pollitt, 2013), and transparency (Vazquez, 2003). This thesis shows that the important interactions between infrastructure operations and available infrastructure capacity affect the performance of shared railway system under alternative capacity pricing and allocation mechanisms. As a result, it is important to design mechanisms that consider these interactions and respond to the overarching goals of the system's stakeholders (Perennes, 2014), and to use the framework proposed in this thesis together with the expertise of the mechanism design field to the implications of such mechanisms for the performance of shared railway systems.

With respect to the second question, the integration of the models developed in this thesis assumes that TOs reveal the IM their real demand for scheduling trains and their willingness to pay for access. As discussed in Chapter 5, this assumption is not valid when TOs and IMs exercise market power. The framework proposed in this thesis could be complemented with game theory and industrial organization concepts to analyze the performance of shared railway systems under mechanisms to price and allocate capacity when TOs and IMs exercise their market power. This line of research would rely on behavioral economic models of the TOs and IMs to capture their response. These models would determine how the outputs of the TO Model relate with the inputs of the IM Model in equilibrium. The comparison of these results with the results obtained in this thesis are important to understand the impact of the market power of railway system stakeholders in the performance of shared railway systems. These findings will also have important implications for the design of efficient mechanisms to price and allocate railway capacity.

Finally, the work of this thesis is part of a larger research field that analyzes infrastructure planning, management, and operations in different network industries (Gomez-Ibanez, 2003; Jacquillat, 2015; Laffont and Tirole, 1993; Laffont and Tirole, 200; Perez-Arriaga, 2013; Sussman, 2000; Vaze, 2011). Although the practical experiences of regulators in one network industry are used in practice when regulating other network industries; the research bodies on different network industries are mostly disconnected. Any efforts 1) to identify lessons that are transferable among industries and 2) to understand how differences in the institutions and in the characteristics of the physical systems among industries translate into modeling differences would be very useful to further connect these bodies of research. We envision that these efforts could be both deductive, using conceptual models of different network industries to analyze the responses of such industries to alternative capacity pricing and allocation mechanisms; and inductive, analyzing the response to alternative capacity pricing and allocation mechanisms in different case studies across network industries.

In a context in which shared-use is becoming more relevant in our economies, this thesis analyzes the prospects to effectively share railway infrastructure. We discuss that while sharing railway infrastructure is not straightforward; understanding alternative rules to coordinate agents is a first step to being able to design adequate rules that unlock the potential benefits of shared use. We thank readers for their attention and hope that some parts of this thesis can help in the design and analysis of effective rules to share infrastructure in network industries.

Appendix A - The Infrastructure Manager Problem: Detailed Train Timetable

This appendix includes the detailed timetable of the first case study presented in Chapter 3 and represented in Figures 3-4 and 3-5.

Intercity train timetable

Table A-1 Intercity train timetable. 12:00pm corresponds to minute 0 in Figures 3-4 and 3-5 (Source: author)

Intercity	
Station	**Train 1**
Station 1	12:00 PM
Station 2	12:25 PM
Station 3	12:50 PM
Station 4	1:15 PM
Station 5	1:40 PM
Station 6	2:05 PM
Station 7	2:30 PM
Station 8	2:55 PM
Station 9	3:20 PM
Station 10	3:45 PM
Station 11	4:10 PM
Station 12	4:35 PM
Station 1	4:55 PM

Commuter train timetable

Table A-2 Commuter train timetable. 12:00pm corresponds to minute 0 in Figures 3-4 and 3-5 (Source: author)

Commuter_1

Station	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6	Train 7	Train 8	Train 9	Train 10	Train 11	Train 12	Train 13	Train 14	Train 15	Train 16	Train 17	Train 18
Station 1	8:48 AM	9:20 AM	9:52 AM	10:24 AM	10:56 AM	11:28 AM	12:08 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:48 PM	5:20 PM	5:52 PM
Station 2	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:44 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:48 PM	5:20 PM	5:52 PM	6:24 PM	6:56 PM	7:28 PM
Station 12	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:58 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:44 PM
Station 1	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:34 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:44 PM	8:16 PM	8:48 PM	9:20 PM
Station	**Train 19**	**Train 20**	**Train 21**	**Train 22**	**Train 23**	**Train 24**	**Train 25**	**Train 26**	**Train 27**	**Train 28**	**Train 29**	**Train 30**	**Train 31**	**Train 32**	**Train 33**	**Train 34**	**Train 35**	**Train 36**
Station 1	6:24 PM	6:56 PM	7:28 PM	8:00 PM	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:56 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM
Station 2	8:00 PM	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:32 AM	2:56 AM	3:28 AM	4:00 AM	4:32 AM	5:04 AM
Station 12	8:16 PM	8:48 PM	9:20 PM	9:52 PM	10:24 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:48 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM
Station 1	9:52 PM	10:24 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:24 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM

Commuter_2

Station	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6	Train 7	Train 8	Train 9	Train 10	Train 11	Train 12	Train 13	Train 14	Train 15	Train 16	Train 17	Train 18	Train 19	Train 20	Train 21	Train 22
Station 3	5:04 AM	5:36 AM	6:08 AM	6:40 AM	7:12 AM	7:44 AM	8:16 AM	8:48 AM	9:20 AM	9:52 AM	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:48 PM	3:12 PM	3:44 PM	4:16 PM
Station 4	6:56 AM	7:28 AM	8:00 AM	8:32 AM	9:04 AM	9:36 AM	10:08 AM	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:08 PM	4:40 PM	5:04 PM	5:36 PM	6:08 PM
Station 5	8:32 AM	9:04 AM	9:36 AM	10:08 AM	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:44 PM	6:16 PM	6:40 PM	7:12 PM	7:44 PM
Station 9	8:48 AM	9:20 AM	9:52 AM	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:48 PM	5:20 PM	6:00 PM	6:32 PM	6:56 PM	7:28 PM	8:00 PM
Station 10	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:52 PM	8:24 PM	8:48 PM	9:20 PM	9:52 PM
Station 11	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:44 PM	8:16 PM	8:48 PM	9:28 PM	10:00 PM	10:24 PM	10:56 PM	11:28 PM

Station	Train 23	Train 24	Train 25	Train 26	Train 27	Train 28	Train 29	Train 30	Train 31	Train 32	Train 33	Train 34	Train 35	Train 36	Train 37	Train 38	Train 39	Train 40	Train 41	Train 42	Train 43
Station 3	4:48 PM	5:20 PM	5:52 PM	6:24 PM	7:04 PM	7:28 PM	8:00 PM	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM
Station 4	6:40 PM	7:12 PM	7:44 PM	8:16 PM	8:56 PM	9:20 PM	9:52 PM	10:24 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM
Station 5	8:16 PM	8:48 PM	9:20 PM	9:52 PM	10:32 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM
Station 9	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:48 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM	4:00 AM	4:32 AM	5:04 AM	5:36 AM	6:08 AM	6:40 AM	7:12 AM
Station 10	10:24 PM	10:56 PM	11:28 PM	12:08 AM	12:40 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM	7:28 AM	8:00 AM	8:32 AM	9:04 AM
Station 11	12:00 AM	12:32 AM	1:04 AM	1:44 AM	2:16 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM	7:28 AM	8:00 AM	8:32 AM	9:04 AM	9:36 AM	10:08 AM	10:40 AM

Commuter_3

Station	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6	Train 7	Train 8	Train 9	Train 10	Train 11	Train 12	Train 13	Train 14	Train 15	Train 16	Train 17	Train 18
Station 7	8:48 AM	9:20 AM	9:52 AM	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:56 PM	5:20 PM	5:52 PM
Station 8	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:48 PM	5:20 PM	5:52 PM	6:32 PM	6:56 PM	7:28 PM
Station 6	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:48 PM	7:12 PM	7:44 PM
Station 7	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:44 PM	8:24 PM	8:48 PM	9:20 PM
Station	**Train 19**	**Train 20**	**Train 21**	**Train 22**	**Train 23**	**Train 24**	**Train 25**	**Train 26**	**Train 27**	**Train 28**	**Train 29**	**Train 30**	**Train 31**	**Train 32**	**Train 33**	**Train 34**	**Train 35**	**Train 36**
Station 7	6:24 PM	6:56 PM	7:28 PM	8:08 PM	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM
Station 8	8:00 PM	8:32 PM	9:04 PM	9:44 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM	4:00 AM	4:32 AM	5:04 AM
Station 6	8:16 PM	8:48 PM	9:20 PM	10:00 PM	10:24 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM
Station 7	9:52 PM	10:24 PM	10:56 PM	11:36 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM

References

ADB, (2014). *Managing for development results: Rail infrastructure tariffs – enabling private development in Mongolia's railway sector*. Asian Development Bank.

Affuso L., (2003). Auctions of rail capacity? *Utilities Policy*, Vol. 11, pp. 43-46.

Agosta B., (2015). *Development of a methodology for evaluating investments in infrastructure for the sustainable exploitation of shale energy in Argentina*. MIT Master Thesis.

AICD, (2008). *Transport prices and costs in Africa: A review of the main international corridors*. World Bank Working Paper 14.

Amjadi, A., A.J. Yeats, (1995). *Have transport costs contributed to the relative decline of Sub-Saharan African exports?* World Bank Policy Research Working Paper 1559.

Amtrak, (2010). *The Northeast Corridor infrastructure master plan*. NEC Master Plan Working Group: Washington, DC.

Amtrak, (2012). *The Amtrak vision for the Northeast Corridor: 2012 update report*. Washington, DC.

Amtrak, (2014). *FY2014 budget and business plan. FY 2015 budget request justification. FY2014-2018 five year financial plan*. Washington, DC.

Amtrak, (2015). *Amtrak system timetable*. Washington, DC.

Archila A.F., (2013). *Intercity Passenger Rail Productivity in the Northeast Corridor: Implications for the Future of HSR*. MIT Master Thesis.

Arvis J.F., G. Raballand, J.F. Marteau, (2010). *The cost of being landlocked: logistics costs and supply chain reliability*, World Bank.

Barrow K., (2012). SNCF fined 60.9m euro for anti-competitive behavior. *International Railway Journal*, December 18, 2012.

Bebiano B.M., O. Petrik, F. Moura, R.F. Allard, J. de Abreu e Silva, (2014). *Simulating the impact of the high-speed network on the performance of other transport* *operators: the HSR on the Lisbon-Oporto link*. In: Proceedings of the 93rd Annual Meeting of the Transportation Research Board. Washington, DC.

Belli P., J. Anderson, H. Barnum, J. Dixon, J.P. Tan, (2001). *Economic analysis of investment operations.* World Bank Institute, Washington, DC.

Ben-Akiva, M.E., S.R. Lerman, (1985). *Discrete choice analysis: theory and application to travel demand*. (Vol. 9). MIT Press.

Bertsekas D., (2006). *Dynamic programming and stochastic control*. Ed. Athena Scientific, Belmont, MA.

Bertsekas D., J.N. Tsitsiklis, (1996). *Neuro dynamic programming*. Ed. Athena Scientific, Belmont, MA.

Big Results Now, (2013). *Executive summary of the transport lab national key results area report*, 2013. Big Results Now, Tanzania Development, Vision 2025, Tanzania.

Borkar V., (1988). A convex analytic approach to Markov decision processes. *Probability Theory Related Fields,* Vol. 78, pp. 583-602.

Cacchiani V., A. Caprara, P. Toth, (2010). Non-cyclic train timetabling and comparability graphs. *Operations Research Letters*, Vol. 38, No. 3, pp. 179-184.

Caimi G., D. Burkolter, T. Herrmann, F. Chudak, M. Laumanns, (2009). Design of a railway scheduling model for dense services. *Networks and Spatial Economics,* Vol. 9, pp. 25-46.

Caimi G., F. Chudak, M. Fuchsberger, M. Laumanns, R. Zenklusen, (2011). A new resource-constrained multicommodity flow model for conflict-free train routing and scheduling. *Transportation Science*, Vol. 45, No. 2, pp. 212-227.

Caprara A., M. Fischetti, P. Toth, (2002). Modeling and solving the train timetabling problem. *Operations Research*, Vol. 50, No. 5, pp. 851-861.

Caprara A., L. Kroon, P. Toth, (2011). Optimization problems in passenger railway systems. *Wiley Encyclopedia of Operations Research and Management Science*.

Castillo E., I. Gallego, J. Ureña, J. Coronado, (2009). Timetabling optimization of a single railway track line with sensitivity analysis. *TOP,* Vol. 17, No. 2, pp. 256-287.

Clewlow, R.R.L., (2012). *The climate impacts of HSR and air transportation in the Northeast Corridor*. MIT Ph.D. Thesis.

Cordeau J.F., P. Toth, D. Vigo, (1998). A survey of optimization models for train routing and scheduling. *Transportation Science*, Vol. 32, No. 4, pp. 380-404.

CPCS, (2013). *Financial and economic analysis of Tanzania Intermodal and Rail Development Project (TIRP),* Report prepared for RAHCO and The World Bank, CPCS Report 12124, Canada.

Crozet Y., (2004). European railway infrastructure: towards a convergence of infrastructure charging? *International Journal of Transport Management*, Vol. 2, pp. 5-15.

Czerny A.I., (2010). Airport congestion management under uncertainty. *Transportation Research Part B*, Vol. 44, pp. 371-380.

DB, (2009). *Feasibility study for the Isaka-Kigali/Keza-Gitega-Musongati railway project.* Study Summary Report prepared for Multinational Tanzania, Rwanda, Burundi, DB, Germany.

De Ghellinck G., (1960). Les problemes de decisions sequentielles. *Cahiers du Centre d'Etudes de Recherche Operationnelle,* Vol. 2, pp. 161-179 (In French).

Drew J., (2006). *Rail freight: The benefits and cost of vertical separation and open access.* Ms. Thesis, Institute for Transport Studies, University of Leeds.

Drew J., C.A. Nash, (2011). *Vertical separation of railway infrastructure - does it always make sense?* Institute for Transport Studies, University of Leeds, Working Paper 594.

Farias D.P., B. Van Roy, (2003). The linear programming approach to approximate dynamic programming. *Operations Research,* Vol. 51, No. 6, pp. 850-865.

Farias D.P., B. Van Roy, (2004). On constraint sampling in the linear programming approach to approximate dynamic programming, *Mathematics of Operations Research*, Vol. 29, No. 3, pp. 462-478.

FRA, (2009). *Metrics and standards for intercity passenger rail service*. Federal Railroad Administration, Department of Transportation, FRA-2009-0016.

Gardner S., (2013). Creating capacity for growth – new approaches for managing the future of the NEC. Socioeconomic and Financial Aspects of Intercity Passenger Rail Subcommittee. *2013 Annual Meeting of the Transportation Research Board*, Washington, D.C.

Ghoseiri K., F. Szidarovszky, M.J. Asgharpour, (2004). A multi-objective train scheduling model and solution. *Transportation Research, Part B: Methodological,* Vol. 38, No. 10, pp. 927-952.

Gibson S., (2003). Allocation of capacity in the railway industry. *Utilities Policy,* Vol. 11, pp 39-42.

Gomez-Ibanez J.A., (2003). *Regulating infrastructure: monopoly, contracts, and discretion.* Harvard University Press, Cambridge, MA, USA.

Gomez-Ibanez J.A., G. de Rus, (2006). *Competition in the railway industry: an international comparative analysis*. Edward Elgar Editors, Northampton, MA, USA.

Greve T., M.G. Pollitt, (2013). Network Procurement Auctions. Cambridge Working Papers in Economics.

Jacquillat, A., (2015). *Integrated allocation and utilization of airport capacity to mitigate air traffic congestion*. MIT Ph.D. Thesis.

Kawakami, S., (2014). *Application of a Systems-Theoretic Approach to Risk Analysis of High-Speed Rail Project Management in the US*. MIT Master Thesis.

Kroon L., D. Huisman, E. Abbink, P.J. Fioole, M. Fischetti, G. Maroti, A. Schrijver, A. Steeenbeek, R. Ybema, (2009). The new Dutch timetable: The OR revolution. *Interfaces*, Vol. 39, No. 1, pp. 6-17.

Krueger H., P.A. Farrington, H. Black Nembhard, D.T. Sturrock, G.W. Evans, (1999). *Parametric modeling in rail capacity planning*. WSC'99. 1999 Winter Simulation Conference Proceedings. `Simulation - A Bridge to the Future', 2.

Laffont J.J., J. Tirole, (1993). *A theory of the incentives in regulation and procurement.* Cambridge: MIT Press.

Laffont J.J., J. Tirole, (2000). *Competition in telecommunications.* Cambridge: MIT Press.

Lago A.M., P. Mayworm, J.M. McEnroe, (1981). Transit service elasticities: Evidence from demonstrations and demand models. *Journal of Transport Economics and Policy*, Vol. 15, No. 2, pp.99-119.

Levy S., M. Pena-Alcaraz, A. Prodan, J.M. Sussman, (2015). Analyzing the financial relationship between railway industry players in shared railway systems: the train operator's perspective. *Transportation Research Record,* in press.

Levy S., (2015). *Capacity Challenges on the San Francisco Peninsula Corridor--How Local Decisions Have Statewide Impacts*. MIT Master Thesis.

Liebchen C., (2008). The first optimized railway timetable in practice. *Transportation Science*, Vol. 42, No. 4, pp. 420-435.

Liebchen C., L. Peeters, (2009). Integral cycle bases for cyclic timetabling. *Discrete Optimization*, Vol. 6, No. 1, pp. 98-109.

Lopez-Pita A., (2014). *UIC study of European railway infrastructure charges for freight*. UIC Report.

Manne A.S., (1960). Linear programming and sequential decisions. *Management Science,* Vol. 6, No. 3, pp. 259-267.

Martland C.D., (2011). *Toward more sustainable infrastructure: Project evaluation for planners and engineers*. Ed. John Wiley & Sons, Boston, MA.

MBTA, (2013a). *Financial statements, required supplementary information and supplementary information*. Boston, MA.

MBTA, (2013b). *Commuter rail executive summary*. Boston, MA.

MBTA, (2015). *MBTA commuter rail schedules and maps*. Boston, MA.

McDaniel T., (2003). Auctioning access to networks: evidence and expectations. *Utilities Policy*, Vol. 11, pp. 33-38.

Morrison S.A., (1990). The value of Amtrak. *Journal of Law and Economics*, Vol. 33, No. 2, pp. 361-382.

Munnell A., (1992). Policy watch: Infrastructure investment and economic growth. *The Journal of Economic Perspectives*, Vol.6, No. 4, pp. 189-198.

Nash C.A., (2003). *Best practices in shared-use high-speed rail systems*. Mineta Transport Institute Report.

Nash C.A., (2005). Rail infrastructure charges in Europe. *Journal of Transport Economics and Policy*, Vol. 39, No. 3, pp. 259-278.

Nash C.A., (2010). European rail reform and passenger services – the next steps. *Research in Transportation Economics*, Vol. 29, No. 1, pp. 204-211.

NEC Commission, (2014). *Northeast Corridor commuter and intercity rail cost allocation policy*. Passenger Rail Investment and Improvement Act of 2008, Section 212(c), Amended on December 17, 2014.

Newbury D.M., (2003). Network capacity auctions: promise and problems. *Utilities Policy*, Vol. 11, pp. 27-32.

Olievschi V.N., (2013). *Rail transport: Framework for improving railway sector performance in Sub-Saharan Africa*. World Bank, Sub-Saharan Africa Transport Policy Program (SSATP), Working Paper No. 94.

Parkes D.C., (2001). *Iterative combinatorial auctions: Achieving economic and computational efficiency*. UPenn Ph.D. Thesis.

Pena-Alcaraz M., A. Fernandez, A.P. Cucala, A. Ramos, R.R. Pecharroman, (2011). Optimal underground timetable design based on power flow for maximizing the use of regenerative-braking energy. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, Vol. 226, No. 4, pp. 397-408.

Pena-Alcaraz M., S.J. Carlson, A.F. Archila, N. Stein, J.M. Sussman, (2013). *Analysis of High-Speed Rail Implementation Alternatives in the Northeast Corridor: the Role of Institutional and Technological Flexibility*. Proceedings of the 92nd Annual Meeting of the Transportation Research Board. Washington, DC.

Pena-Alcaraz M., J.M. Sussman, I. Perez-Arriaga, (2014). Capacity pricing schemes to implement open-access rail in Tanzania. *Network Industry Quarterly*, Vol. 16, No. 2, pp. 18-22.

Pena-Alcaraz M., (2015). Capacity pricing and allocation in shared railway systems. *Intercity Rail Passenger Systems Update*, TRB, No. 23, pp.13-15.

Pena-Alcaraz M., M.D. Webster, A. Ramos, (2015a). *Train timetable design for shared railway systems using a linear programming approach to approximate dynamic programming*. Engineering Systems Division Working Paper Series. Submitted for publication.

Pena-Alcaraz M., J.M. Sussman, M.D. Webster, J.I. Perez-Arriaga, (2015b). *Comparison of price-based and capacity-based mechanisms for pricing and allocating railway capacity in the Northeast Corridor*. Engineering Systems Division Working Paper Series. Submitted for publication.

Perennes P., (2014). Use of combinatorial auctions in the railway industry: Can the “invisible hand” draw the railway timetable? *Transportation Research Part A: Policy and Practice*, Vol. 67, pp. 175-187

Perez-Arriaga I.J., L. Olmos, (2009). *A comprehensive approach for computation and implementation of efficient electricity transmission network charges*. MIT Center for Energy and Environmental Policy Research, Vol. 7.

Perez-Arriaga I.J., (2013). *Regulation of the power sector*. Chapters 4, 5, and 6. Edited by Perez-Arriaga, Springer Verlag.

Pouryousef H., P. Lautala, (2015). *Capacity evaluation of directional and non-directional operational scenarios along a multiple-track U.S. corridor*. Proceedings of the 94th Annual Meeting of the Transportation Research Board. Washington, DC.

Powell W., (2007). *Approximate dynamic programming: Solving the curses of dimensionality*, Ed. Willey, NJ.

Pozzo di Borgo P., (2005). *Railway concession experiences in Sub-Saharan Africa*. World Bank.

PPIAF, World Bank, TRS, (2011). *The Railway Reform Toolkit*. World Bank.

PRIIA, (2008). *Passenger Rail Investment and Improvement Act of 2008*. Federal Railroad Administration.

Prodan A., (2015). *Railway infrastructure pricing and interaction of infrastructure managers and train operators*. Working paper.

Raballand G., P. Macchi, (2009). *Transport prices and costs: The need to revisit donor's policies in transport in Africa*. World Bank.

Railistics, (2013). *Institutional assessment of Tanzania railway system report*. Report prepared for RAHCO and The World Bank, Railistics, Tanzania.

Renfe, (2015). *Renfe's punctuality commitment*. Renfe's website www.renfe.com Accessed on March 26, 2015.

Rubinstein R.Y., D.P. Kroese, (2008). *Simulation and the Monte Carlo Method*, Second Edition. John Wiley & Sons, Hoboken, N.J., chapter 6.

Rubio F.J., (1999). *Metodologia de asignacion de costes de la red de transporte en un contexto de regulacion abierta a la competencia*. Ph.D. Thesis, Comillas Pontifical University (in Spanish).

Sakamoto R., (2012). *High speed railway productivity: How does organizational restructuring contribute to HSR productivity growth?* MIT Master Thesis.

Stern J., R. Turvey, (2003). Auctions of capacity in network industries. *Utilities Policy*, Vol. 11, pp. 1-8.

Sussman J.M., (2000). *Introduction to transportation systems*. Artech House, Boston, MA.

Sussman J.M., A.F. Archila, S.J. Carlson, M. Pena-Alcaraz, N. Stein, (2012). *Transportation in the Northeast Corridor of the U.S.: A multimodal and intermodal conceptual framework*. Report for JITI.

Sussman J.M., T.P. Doyle, J.C. Moody, M. Pena-Alcaraz, (2015). *The evaluation of international HSR markets for JR East: A systems approach*. Report for Japan Railways East.

The Economist, (2015). Africa rail and roads: all aboard. *The Economist*, February 28, 2015, Middle East and Africa Edition.

Texeira P.F., A. Lopez-Pita, (2012). *UIC study on railway infrastructure charges in Europe, high-speed*. UIC Report.

Teixeira P.F., A. Prodan, (2014). Railway infrastructure pricing in Europe for high-speed and intercity services: State of the practice and recent evolution. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2448, Railroads, pp. 1-10.

Vaze V., (2011). *Competition and congestion in the National Aviation System: multi-agent, multi-stakeholder approaches for evaluation and mitigation*. MIT Ph.D. Thesis.

Vazquez C., (2003). *Modelos de casacion de ofertas en mercados electricos*. Ph.D. Thesis, Comillas Pontifical University (in Spanish).

Weitzman M.L., (1974). Price vs. quantities. *Review of Economic Studies*, Vol. 41, No. 4, pp. 477-491.

World Bank, (2008). *Infrastructure and growth in developing countries: Recent advances and research challenges*. Development Research Department Report, World Bank.

World Bank, (2014). *Project appraisal document for Tanzania intermodal and rail development project*. World Bank.

Zhou X., M. Zhong, (2005). Bicriteria train scheduling for high-speed passenger railroad planning applications. European Journal of Operational Research, Vol. 167, No. 3, pp. 752-771.

Additional Publications

Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems: Lessons for the Northeast Corridor (attached)

Maite Pena-Alcaraz and Joseph Sussman

Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems: Lessons for the Northeast Corridor

Maite Pena-Alcaraz & Joseph Sussman
Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA

In the last 15 years, the use of rail infrastructure by different train operating companies (shared railway system) has been proposed as a way to improve infrastructure utilization and to increase efficiency in the railway industry. Shared use requires coordination between the infrastructure manager and multiple train operators (Gomez-Ibanez-2003). Such coordination requires capacity planning mechanisms that determine which trains can access the infrastructure at each time, capacity allocation, and the access charges they have to pay, capacity pricing.

This research contributes to the field of shared railway systems coordination by 1) developing a framework to analyze the performance of shared railway systems under alternative capacity pricing and allocation mechanisms, and 2) using this framework to understand the implications of representative capacity pricing and allocation mechanisms in the Northeast Corridor in the US, one of the busiest shared railway systems worldwide.

NEC Background

The Northeast Corridor (NEC) is a 457-mile stretch of fully electrified railway line between Boston and Washington DC. Figure 1 shows the NEC infrastructure ownership. As mentioned above, the NEC is one of the busiest shared railway systems worldwide, with over 2,000 commuter trains, 150 intercity trains and 70 freight trains per day.

However, with current bilateral infrastructure access contracts, 1) the price that each TO pays to access the infrastructure depends on the time at which the contract was signed (companies who signed their agreements when there was still plenty of excess capacity are paying much less than other companies to operate the same type of services), 2) access charges and slots are rigid (none of the companies want to lose their current slots and it is difficult to make room for new trains because multiple contracts would need to be renegotiated), and 3) the IM is not able to raise enough revenues to afford basic maintenance of the lines (this has contributed to the current backlog in maintenance in the NEC) (Gardner, 2013).

As a result, the Federal Railroad Administration (FRA) requires all the railroads to agree on a market-based mechanism for pricing and allocating capacity to substitute for the current negotiation-based mechanism (PRIIA, 2008).

Although the NEC is one of the railway corridors most widely studied in the literature, the implications of new mechanisms to price and allocate railway capacity in this system are still unclear.

Figure 1. NEC Infrastructure Ownership (Source: NEC Infrastructure Master Plan Working Group 2010)

Capacity Pricing and Allocation Mechanisms

According to (Gibson, 2003), there are two main types of market-based mechanisms for capacity pricing and allocation: 1) price-based and 2) capacity-based.

Price-based mechanisms are those that determine the price at which capacity will be offered, and let TOs decide whether they are willing to access the infrastructure or not. Price-based mechanisms are typically complemented with priority rules that allow the IM to decide which train to schedule when there are conflicts (multiple TOs willing to pay the predetermined access charges). An example of a price-based mechanism would be a cost-allocation mechanism that assigns infrastructure-related cost proportionally to the volume of infrastructure use (Crozet, 2004; Nash, 2005; Lopez-Pita, 2014; Texeira and Prodan, 2014). The access charge could also be adjusted considering the TOs' demand for scheduling trains (e.g. introducing congestion prices). These charges could also be adjusted with a base tariff that allows the IM to recover infrastructure costs that are fixed in nature.

Capacity-based mechanisms are those that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity, e.g. an auction (Affuso, 2003; McDaniel, 2003; Newbury, 2003; Perennes, 2014; Stern and Turvey, 2003). There are multiple types of auctions: simple auctions in which TOs bid to get some predefined slots (either in a segment of the infrastructure or for the full path) or submit their desired timetable when they bid, and combinatorial auctions where the TOs' bid depends on the result of the auction. Capacity-based mechanisms have been widely studied in the literature but have not yet been implemented on the railway system in any country.

Methodology

There are strong interactions between capacity planning and infrastructure operations in the railway industry; the operations on the infrastructure determine the available capacity in the system. As a consequence, the framework developed in this research to evaluate the performance of shared railway systems under alternative capacity pricing and allocation consists of two models: 1) a train operator model and 2) an infrastructure manager model. The train operator model is a financial model that anticipates how train operators would respond to the capacity pricing and allocation mechanisms and determine their demand for infrastructure use. The infrastructure manager model is a network optimization model that determines the optimal train timetable (infrastructure manager's decisions) that accommodates the train operators' demands for scheduling trains, considering the topology of the system, safety constraints, and other technical aspects of the infrastructure for shared railway systems. To be able to solve the train timetabling optimization problem in meaningful instances, this research develops a novel approximate dynamic programming algorithm based on linear programming that extends previous algorithms proposed in the literature to effectively solve large network optimization problems.

This research then uses the train operator model to compare the operational decisions of train operators in shared railway systems with the operational decisions of even-handed integrated railway companies. We show that train operators in shared railway systems make the same operational decisions as an integrated railway company when variable access charges reflect variable infrastructure manager's costs to operate trains on the infrastructure. We also identify two cases in which the train operators may have incentives to deviate from the integrated railway systems' operational decisions: 1) when the infrastructure manager needs to recover part of the infrastructure management fixed costs, or 2) when the railway system is congested.

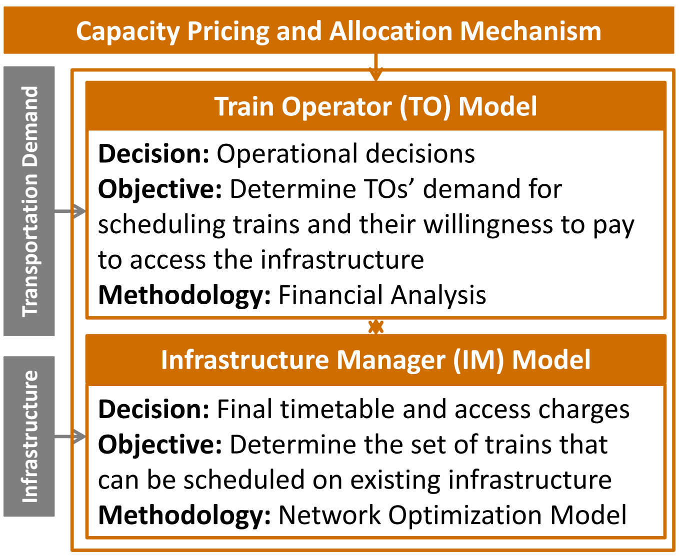


Figure 2. Framework to analyze capacity pricing and allocation mechanisms

Results Discussion

This section summarizes the main results obtained when we use the framework proposed in this research to analyze the trade-offs associated with the use of alternative mechanisms in the Northeast Corridor. To our knowledge, this is the first effort to compare alternative mechanisms to price and allocate capacity in the same shared railway system. The results of this research show that there are important trade-offs associated with each mechanism and none of them is superior to the other on all dimensions. We thus recommend that system stakeholders carefully analyze the implications of alternative capacity pricing and allocation mechanisms before locking the system into one of them. This is particularly important today since several countries are currently restructuring their railway sector to allow shared use.

Conclusions and Recommendations

The first conclusion of this research is that the implications of capacity pricing and allocation mechanisms for shared railway systems are still unclear. While this research tries to offer clarity in this area, there is still much work to be done. In that sense, we recommend to academics that they invest in this research topic. Any progress in research that contributes to a better understanding of the implications of alternative mechanisms to price and allocate capacity could immensely help practitioners and policy makers. This is particularly important in a context in which several countries are currently restructuring their railway sector to allow shared use.

The second conclusion of this research is that sharing railway infrastructure capacity is not straightforward. In the railway industry, as compared to other network industries, there are very strong interactions between capacity planning and infrastructure operations.

The third conclusion is that the implementation of adequate capacity pricing and allocation mechanisms can mitigate the coordination problems of shared railway systems while maintaining the benefits of shared infrastructure use in the railway industry. We argue that the improved understanding of the system performance gained with the framework proposed in this research enables stakeholders to design adequate capacity pricing and allocation mechanisms.

The authors would like to thank NURail, the U.S. Department of Transportation University Transportations Centers Program, La Caixa Foundation and Rafael del Pino Foundation for sponsoring this research.

References

[1] Gomez-Ibanez J.A., (2003). Regulating infrastructure: monopoly, contracts, and discretion. Harvard University Press, Cambridge, MA, USA.

[2] Gardner S., (2013). Creating capacity for growth – new approaches for managing the future of the NEC. Socioeconomic and Financial Aspects of Intercity Passenger Rail Subcommittee. 2013 Annual Meeting of the Transportation Research Board, Washington, D.C.

[3] PRIIA, (2008). Passenger Rail Investment and Improvement Act of 2008. Federal Railroad Administration.

[4] Gibson S., (2003). Allocation of capacity in the railway industry. Utilities Policy, Vol. 11, pp 39-42.

[5] Crozet Y., (2004). European railway infrastructure: towards a convergence of infrastructure charging? International Journal of Transport Management, Vol. 2, pp. 5-15.

[6] Nash C.A., (2005). Rail infrastructure charges in Europe. Journal of Transport Economics and Policy, Vol. 39, No. 3, pp. 259-278.

[7] Lopez-Pita A., (2014). UIC study of European railway infrastructure charges for freight. UIC Report.

[8] Teixeira P.F., A. Prodan, (2014). Railway infrastructure pricing in Europe for high-speed and intercity services: State of the practice and recent evolution. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2448, Railroads, pp. 1-10.

[9] Affuso L., (2003). Auctions of rail capacity? Utilities Policy, Vol. 11, pp. 43-46.

[10] McDaniel T., (2003). Auctioning access to networks: evidence and expectations. Utilities Policy, Vol. 11, pp. 33-38.

[11] Newbury D.M., (2003). Network capacity auctions: promise and problems. Utilities Policy, Vol. 11, pp. 27-32.

[12] Perennes P., (2014). Use of combinatorial auctions in the railway industry: Can the “invisible hand” draw the railway timetable? Transportation Research Part A: Policy and Practice, Vol. 67, pp. 175-187

[13] Stern J., R. Turvey, (2003). Auctions of capacity in network industries. Utilities Policy, Vol. 11, pp. 1-8.

[14] Pena-Alcaraz M., (2015). *Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems*. MIT PhD Thesis.