MINETA NATIONAL TRANSIT RESEARCH CONSORTIUM

Modeling Taxi Demand with GPS Data from Taxis and Transit

MNTRC Report 12-16

MNTRC MINETA NATIONAL TRANSIT RESEARCH CONSORTIUM

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

MINETA TRANSPORTATION INSTITUTE
LEAD UNIVERSITY OF MNTRC

The Norman Y. Mineta International Institute for Surface Transportation Policy Studies was established by Congress in the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA). The Institute's Board of Trustees revised the name to Mineta Transportation Institute (MTI) in 1996. Reauthorized in 1998, MTI was selected by the U.S. Department of Transportation through a competitive process in 2002 as a national "Center of Excellence." The Institute is funded by Congress through the United States Department of Transportation's Research and Innovative Technology Administration, the California Legislature through the Department of Transportation (Caltrans), and by private grants and donations.

The Institute receives oversight from an internationally respected Board of Trustees whose members represent all major surface transportation modes. MTI's focus on policy and management resulted from a Board assessment of the industry's unmet needs and led directly to the choice of the San José State University College of Business as the Institute's home. The Board provides policy direction, assists with needs assessment, and connects the Institute and its programs with the international transportation community.

MTI's transportation policy work is centered on three primary responsibilities:

Research

MTI works to provide policy-oriented research for all levels of government and the private sector to foster the development of optimum surface transportation systems. Research areas include: transportation security; planning and policy development; interrelationships among transportation, land use, and the environment; transportation finance; and collaborative labor-management relations. Certified Research Associates conduct the research. Certification requires an advanced degree, generally a Ph.D., a record of academic publications, and professional references. Research projects culminate in a peer-reviewed publication, available both in hardcopy and on TransWeb, the MTI website (http://transweb.sjsu.edu).

Education

The educational goal of the Institute is to provide graduate-level education to students seeking a career in the development and operation of surface transportation programs. MTI, through San José State University, offers an AACSB-accredited Master of Science in Transportation Management and a graduate Certificate in Transportation Management that serve to prepare the nation's transportation managers for the 21st century. The master's degree is the highest conferred by the California State University system. With the active assistance of the California Department of Transportation, MTI delivers its classes over a state-of-the-art videoconference network throughout the state of California and via webcasting beyond, allowing working transportation professionals to pursue an advanced degree regardless of their location. To meet the needs of employers seeking a diverse workforce, MTI's education program promotes enrollment to under-represented groups.

Information and Technology Transfer

MTI promotes the availability of completed research to professional organizations and journals and works to integrate the research findings into the graduate education program. In addition to publishing the studies, the Institute also sponsors symposia to disseminate research results to transportation professionals and encourages Research Associates to present their findings at conferences. The World in Motion, MTI's quarterly newsletter, covers innovation in the Institute's research and education programs. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

REPORT 12-16

MODELING TAXI DEMAND WITH GPS DATA FROM TAXIS AND TRANSIT

Eric J. Gonzales, Ph.D.
Ci (Jesse) Yang, M.S.
Ender Faruk Morgul, M.S.
Kaan Ozbay, Ph.D.

July 2014

A publication of

Mineta National Transit Research Consortium

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. CA-MNTRC-14-1141	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Modeling Taxi Demand with GPS Data from Taxis and Transit		**5. Report Date** July 2014 **6. Performing Organization Code**
7. Authors Eric J. Gonzales, Ph.D., Ci (Jesse) Yang, M.S., Ender Faruk Morgul, M.S., and Kaan Ozbay, Ph.D.		**8. Performing Organization Report** MNTRC Report 12-16
9. Performing Organization Name and Address Mineta National Transit Research Consortium College of Business San José State University San José, CA 95192-0219		**10. Work Unit No.** **11. Contract or Grant No.** DTRT12-G-UTC21
12. Sponsoring Agency Name and Address U.S. Department of Transportation Research & Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590	Edward J. Bloustein School of Planning and Public Policy Rutgers, The State University of New Jersey 33 Livingston Avenue New Brunswick, NJ 08901	**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code**
15. Supplemental Notes		

16. Abstract

Identifying factors that influence taxi demand is very important for understanding where and when people use taxis and how taxi demand relates to the availability and quality of transit service. This study used a large set of global positioning system (GPS) data from taxis in New York City, along with demographic, socioeconomic, and employment data to identify the factors that drive taxi demand. A technique was developed to measure and map transit accessibility based on the time required to access a transit vehicle from a specific location and time of day. Taxi data were categorized by pickups and drop-offs, and a hybrid cross-classification and regression model was developed to estimate the taxi demand across space and time. The study identified transit accessibility, population, age, education, income, and the number of jobs in each census tract as the factors with strongest explanatory power for predicting taxi demand. The study also includes a comparison of the cost of travel by taxi and transit for specific trips between Penn Station and each of the three major New York area airports. The model and analysis results show how the number of passengers traveling together in a group and the value they place on their time affect the likelihood of choosing taxi or transit for an airport access trip. A number of findings are presented in this report that are specific to New York City. However, the methods developed in this study and demonstrated in this report can be applied generally to cities around the United States and the world where similar GPS data from taxis and schedule information from transit are available.

17. Key Words Level of service; Transit accessibility; Trip generation; Choice models; Taxi travel demand	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 68	**22. Price** $15.00

Form DOT F 1700.7 (8-72)

Library of Congress Catalog Card Number:
2014945605

To order this publication, please contact:

Mineta National Transit Research Consortium
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/mntrc/index.html

072514

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Transportation's University Transportation Centers Program under Grant Number DTRT12-G-UTC21.

The authors would like to thank the Mineta Transportation Research Institute for its support. The authors also thank MTI staff, including Executive Director Karen Philbrick, Ph.D.; Director of Communications and Technology Transfer Donna Maurillo; Research Support Manager Joseph Mercado; and Webmaster Frances Cherman, who also provided editorial and publication support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Taxis provide an alternative to conventional public transit services in many cities, and understanding the demand for taxis requires consideration of the role that taxis serve in the greater transportation system. This report presents the results of a study to model taxi demand across time and space, explicitly accounting for the presence and quality of transit service. The primary objective of the study was to identify the factors that drive taxi demand and to understand how this varies by location and time of day. This was accomplished by developing demand models for taxi trip generation and mode choice that explicitly account for the characteristics of transit service in the neighborhoods where trips are made. The resulting insights are useful for making regulatory, planning, and engineering decisions about how to manage taxi markets, accounting for their role in the transportation system.

A secondary objective of the study was to demonstrate how emerging “big data” from taxis and transit systems can be integrated with demographic, socioeconomic, and employment information to develop useful demand models. In particular, large sets of data that include records referencing specific times and locations provide a wealth of information that can be used to model and understand how demand varies across time and space. Ultimately, developing methods to systematically analyze and extract meaningful information from these large data sources will help improve the way transportation systems are monitored and managed.

The dataset includes records of every taxi trip in New York City over a 10-month period. The data was tracked by automatically operating Global Positioning System (GPS) receivers installed in each licensed taxi. Additional data sources included detailed transit schedule and routing information from transit agencies available online in the Google Transit Feed Specification (GTFS) format. Demographic, socioeconomic, and employment data were obtained from the U.S. Census Bureau at the spatial resolution of census tracts. By properly processing the data and integrating the various types of information in a Geographic Information System (GIS), it was possible to develop models that provide insights into the factors that determine the number of trips made by taxi. This study demonstrates the model for NYC, but the methods are general and can be applied to cities around the world where similar data is collected and available.

The study was conducted in two parts. First, a trip generation model was developed to identify location characteristics that determine the number of taxi trip origins (pickups) and taxi trip destinations (drop-offs) that are generated during each hour of the day. Second, a mode choice model was developed and analyzed to determine how the competitive appeal of taxi travel versus transit changes by time of day as the travel cost for each mode varies.

A trip generation model was developed as a hybrid cross-classification and regression model. Taxi demand and transit accessibility data were classified by hour of day. A separate regression model was then developed to estimate the number of taxi pickups and drop-offs in a census tract for each hour of the day. In order to fit these models, the different data sources had to be aggregated to the same spatial and temporal resolution. Since demographic, socioeconomic, and employment data is available at the level of census tracts, that was the spatial unit used for the analysis.

In order to account for the spatial and temporal variation of transit accessibility across the city, a method was developed to measure Transit Access Time (TAT) based on transit schedules. The TAT at a specific point and time represents the time that it would take a person to access the nearest transit departure, which includes walking to a transit station and waiting for the next departing vehicle. Detailed transit schedule information was extracted from a database of transit route and schedule records that are available from transit agencies in the GTFS format. Then a clustering algorithm was used to identify the minimum access time from a location, which may not be at the nearest station because a further station with more frequent service may provide better transit accessibility. The result is a quantitative measurement of transit accessibility that can be mapped to show variation across different locations in the city and used as an explanatory variable in the trip generation model.

The trip generation models that have been developed from these data reveal that there are six characteristics of a census tract that have the greatest explanatory power for estimating taxi demand:

- Transit accessibility
- Population size
- Median age
- Percent of population educated beyond bachelor's degree
- Median income per capita
- Number of job opportunities (irrespective of residence)

An additional, detailed investigation of taxi demand within Manhattan shows that there are certain types of employment opportunities that are more correlated with taxi trips than others. The number of employees working in retail, accommodation and food service, and healthcare are the strongest determinants of the number of taxi trips. The magnitude of their influence also changes with the time of day, so patterns are revealed about how activities in NYC vary over the course of the day and which activities are most associated with taxi use. Although it is not possible to know the precise trip purpose without a traveler survey, these findings support the notion that people are more likely to use taxis when traveling to and from stores, hotels, restaurants, and hospitals. It appears that taxis and transit sometimes operate in competition and at other times are complements because both modes follow and influence the levels of activity in neighborhoods across the city.

The second part of the study focused on the costs of taxi and transit trips for a few specific origin-destination (OD) pairs. Using the data for NYC, the analysis looked at trips between Penn Station and each of the three major airports in the area: John F. Kennedy International Airport (JFK), LaGuardia Airport (LGA), and Newark Liberty International Airport (EWR). The taxi data reveals how travel time and fare varies for each OD pair by time of day. This was compared against the main non-driving alternative, public transit. A script was used to extract detailed, time-specific transit routes and access locations from the Google Maps API Transit Directions Service. These were used determine the waiting time, travel time, and fare for making the same trips, at the same times of day, as the taxi trips.

In the context of a mode choice analysis, the comparison of trip costs by taxi and transit show how the likelihood of travelers choosing one mode or the other changes over the course of the day. A sensitivity analysis is particularly useful in showing the tipping points at which the number of passengers traveling together in a group, or the value they place on their time, makes the additional expenditure for a taxi worthwhile. Typically transit is more competitive during the day when the frequency of service is high, especially during the morning and evening peaks when traffic congestion also slows taxis. Taxis are more competitive in the evening hours when traffic moves quickly and less frequent transit service imposes longer waiting times on travelers. In two cases, there was no trade-off observed: Transit is both faster and cheaper than taxis for trips to JFK during the afternoon peak and trips from JFK during the morning peak, when traffic congestion eliminates the competitive advantage of taxi speed.

The models and findings presented in this report are specific to NYC, because of the data sources used. The methods can be generally applied to any city in which similarly detailed data on taxi use and transit schedules are available. The models provide insights about both the spatial and temporal variation of taxi demand across the city. These models and insights are useful for designing taxi regulations and transit schedule improvements. The models also show how characteristics of a neighborhood and competing transit service affect the number of trips made by taxi.

I. INTRODUCTION

Transportation systems in large cities are inherently multimodal. Of the modes available for use by the general public, taxis have received relatively less attention in the modeling literature. In the largest cities in the United States and around the world, taxis play a major role in the urban transportation system and present an alternative mode of travel to conventional public transportation systems. For example, 13,363 licensed taxis in New York City (NYC) served 172 million trips in 2006, accounting for 11 percent of all travel in the city.[1]

There are concerns from transportation planners and citizens of New York City about how taxi service is distributed across the city in space and time because taxis tend to congregate and serve trips in parts of Manhattan and the city's airports. In order to effectively regulate the taxi industry and plan for its effective integration into the citywide transportation system, it is necessary to understand the demand for taxis. Relevant information includes where taxis are used, when taxis are used, and which factors tend to drive people to use taxis as opposed to other modes, such as transit.

In recent years, technologies have allowed for the collection of detailed transportation data in much larger quantities than was previously available. For example, the Taxi and Limousine Commission (TLC) in NYC logs Global Positioning System (GPS) data for every taxi trip in the city, including the time and location of pickups and drop-offs. A complete set of trip data is used for this study from the 10-month period from February 1, 2010, to November 28, 2010, consisting of 147 million observed trips. Due to its large scale, this source is an example of "big data." Using big data to develop useful models for taxi demand requires developing procedures to clean and process the information so that it can be organized in a useful way.

The primary objective of this study is to identify the factors that drive demand for taxis, accounting for the effect of transit service availability and quality of service. The approach was to develop demand models that acknowledge the distribution of taxi demand in space and time so that it is possible to identify how taxi demand varies from neighborhood to neighborhood and how the demand evolves over the course of a day.

MODELING APPROACH

The analysis in this study was conducted in two parts. First, a trip generation model was developed to identify the factors that determine the number of taxi pickups and drop-offs generated in a given neighborhood. The model followed a hybrid approach, classifying taxi records by hour of day and then using regression to model within each hour the number of pickups and drop-offs within each census tract. The result was a model (actually a set of models) that provides predictive capability and makes distinctions between location and time of day.

The goal of the trip generation model was to identify the factors that affect the demand for taxi trips at the level of a census tract. Thus, extensive data from the U.S. Census Bureau, including characteristics related to population, age, education, income, and employment by industry sector, were considered. An additional goal was to identify what effect, if any, accessibility of transit has on demand for taxis in a neighborhood, so transit schedules were also considered for this part of the analysis.

A second analysis was conducted to investigate the competition between taxi and transit for specific origin-destination pairs. The goal was to determine how mode choice is likely to change over the course of a day. Traffic congestion affects the speed and price of taxis, and changes in transit service headways affect the amount of time that travelers can expect to wait for service. For this part of the analysis, consideration was given to trips between Penn Station and each of the three major NYC airports: John F. Kennedy International Airport (JFK), LaGuardia Airport (LGA), and Newark Liberty International Airport (EWR). Trips to and from airports constitute an important market for taxis and provide a case for demonstrating a general modeling methodology that could be extended across other origin-destination pairs. The results provide insights about how the operating characteristics of taxis and transit, as well as traveler preferences, such as value of travel time, affect the tendency of people to choose one mode or the other.

ORGANIZATION OF THE REPORT

This report is organized in four main sections as follows. The first section is a review of the existing literature on models for taxi markets, trip generation, and airport trips. The literature review serves to demonstrate that there is an unmet need for research to model taxi demand, and there is an opportunity to use complete taxi data from GPS to conduct this type of research. The second section provides a detailed description of the data sources used for each analysis. This includes a description of the GPS data, as well as the transit schedules, demographic, socioeconomic, and employment data associated with each census tract in NYC. The third section describes the trip generation modeling procedure and results. The fourth section describes the mode modeling procedure and findings. Finally, a conclusion links the findings from the modeling and analysis described throughout the report to show how the objectives of the study have been addressed.

II. LITERATURE REVIEW

The literature on taxis and public transit systems is diverse in the breadth of topics covered and the level of detail of analysis. Many studies have addressed trip generation and mode choice related to public transit systems; far fewer have sought to model demand for taxis. A review of the literature on trip generation and mode choice for these modes reveals a need to use modern data sources to answer questions about where and when taxis are most likely to be used for travel, and by whom.

MODELING TAXI MARKETS

Taxis are an important transport mode in urban areas, including airport ground access, first, because they provide passengers with convenient, comfortable and prompt trips[2] and, second, they can be used as to complement or substitute for the mass transit system.[3] Three types of taxi markets are generally discussed in the literature: the dispatch (telephone order) market, cab stand market (also known as rank place market), and the hail market in which cabs cruise streets in search of customers.[4]

Most of the literature on taxi markets focuses on issues related to regulation and the aggregated effects of policies on the supply of taxis in a market. Studies dating back to the 1960's raise debates about how taxi markets should be regulated, if at all, on the basis of economic analysis. Arguments in favor of regulation are based on the notion that regulated taxi markets limit the number of private vehicles on the road,[5] protect public transit systems,[6] and provide public safety and consumer protection.[7] Others argue that regulations on the taxi industry impose costly economic inefficiencies.[8]

The importance of understanding taxi markets in order to better manage them has driven more recent empirical research. Schaller[9] presents an analysis of the number of taxicabs in 118 U.S. cities using multiple linear regression models. The factors influencing the size of a city's taxi fleet include population, employment, use of complements to taxi cabs (e.g., transit), cost of taxis, and taxi service quality. However, the model predicts the quantity of taxicabs instead of the number of trips generated. Schaller shows that the number of workers commuting by subway, the number of households with no vehicles available, and the number of airport taxi trips have significant explanatory power for the number of cabs in operation. This work, and most others that address the size and characteristics of taxi markets, are based on aggregated citywide data.[10]

In recent years, technologies, such as GPS, installed in each cab have allowed the collection and analysis of much richer sets of taxi data.[11] This data has been used in some research for calibrating models of traffic conditions.[12] There are limited examples of large taxi GPS datasets to model taxi demand.[13] The availability of big data from GPS for taxi trips now presents an opportunity for more quantitative analysis of the factors that drive the demand for taxis.

MODELING TRIP GENERATION

Although little work has been done to model taxi trip generation explicitly, trip generation models for other modes, such as public transit, have been much more thoroughly developed. Since both taxis and transit provide transportation service to the general public, insights from transit trip generation models may provide a starting point for the development of taxi trip generation models. Furthermore, a goal of the present study was to determine if transit service has an effect on taxi use, so the interconnections between these modes were important to consider.

Trip generation models were used to predict the total number of trips that originate or terminate in a Transportation Analysis Zone (TAZ), and this constitutes the first step of a travel demand forecast. These models relate the total number of trips produced in a TAZ to a variety of factors related to the TAZ and transportation modes available:[14]

- Level Of Service (LOS) of the mode;
- accessibility of the mode;
- demographics of the TAZ (e.g., population, race);
- socioeconomics of the TAZ (e.g., income, education);
- other characteristics of the TAZ (e.g., as land area);
- land use in the TAZ.

TAZs are geographic units of analysis that may vary in size depending on the model and analysis objectives. The models that are presented and analyzed in this report use census tracts as the geographic units for the TAZs because the relevant socioeconomic data is already collected and aggregated at that level.

Three methods are commonly used to model trip production: rate method,[15] cross-classification,[16] and regression.[17] Regression is a widely used statistical method for exploring the relationship between response variables and explanatory variables with various approaches for validating the model. If enough information is available, trip generation based on regression models can be very useful to forecast travel demands in each TAZ of an urban transportation system.[18] A large dataset with sufficiently detailed information about travel and TAZ characteristics is necessary to model trip generation across a large geographic area using regression.

Characteristics of the trip (e.g., travel purpose) and characteristics of the traveler (e.g., age and income) have been identified as influential factors that affect the trips generated by different travel modes.[19] Trips to residential areas and non-residential areas[20] and trips for business and non-business purposes[21] are analyzed separately in most studies. A number of studies have been conducted about the generation of airport trips[22] and travel to schools.[23] Researchers have also studied trips generated by elderly people, because their needs and behavior have some distinct difference from other population groups.[24]

MODELING MODE CHOICE WITH TRANSIT

In addition to developing models of travel demand to forecast the number of trips that will be made, other models have been developed to specifically analyze how users choose between modes. Mode choice models are used to compare among various transportation options, for example public transportation services and taxis.[25] The most common method is to estimate the probability that a person will choose a specific mode using a logit model based on the utility of completing the trip by each of the modes available. There are many different types of logit models, such as a binary logit model,[26] nested logit model,[27] ordered logit model,[28] and box-cox logit model.[29]

Racca and Ratledge[30] present a comprehensive list of factors that are used for modeling mode choices involving transit including transit LOS, accessibility, land use, demographics, and characteristics of the trips. That study shows that high transit service is focused at locations with high employment and population densities in the city of Wilmington, Delaware. The analysis on mode split versus mean age and time of day indicate that these variables affect the modes that people choose, and this means that they may also relate to taxi trip generation. Corpuz[31] shows that socioeconomic characteristics and time of day have influenced people's choices between private vehicles and public transportation. Workers and households with higher incomes are more likely to use cars over public transit in that time-of-day analysis. The train and the bus are more likely to be picked during morning and late afternoon peaks, because people want to avoid the time and the cost of driving in congestion.

Most of the research that seriously considers taxis as a mode choice focuses on trips to and from airports. Harvey[32] was one of the first studies to demonstrate the factors influencing the airport access mode choice of departing airline passengers based on a travel survey in the San Francisco Bay Area. The analysis, using a multinomial logit model, shows that travel time and travel cost are two strong explanatory variables. Business travelers are found to be more sensitive to airport access travel time than leisure travelers, and values of time for most individuals are estimated to be at least as high as the average wage. Extra luggage, which is defined as more than one piece per person, deters passengers from choosing transit. Psaraki and Abacoumkin[33] analyzed the mode split for Athens International Airport in Greece to predict future mode shares and found that international passengers are more likely to use taxis or be dropped off by private cars. Pels et al.[34] also studied mode choice in the San Francisco Bay Area and reported that business travelers have higher value of time and higher access time elasticity compared to leisure travelers. The authors reported that access time has a larger influence on mode and airport choice compared to the dollar cost. However, they calculated travel times for each alternative mode as follows: public transit travel time estimates are drawn from train and bus schedules, and taxi travel times are based on distances from the center of the origin zip code to the airport. Therefore their estimations do not account for walking and transit transfer times for public transit or the effect of traffic congestion or road network circuity for taxis or private vehicles.

Some studies have investigated the factors that influence airport ground access mode choice including demographics, trip cost, travel time, travel time reliability, and accessibility. Gupta et al.[35] developed a ground access mode choice model for NYC using an air passenger survey and a nested logit model. They prepared transit LOS data using online schedules, and waiting times are also taken into account. Demographic characteristics, trip cost, travel time, and trip purpose are shown to be the most significant variables in passengers' mode and airport choice, which is consistent with previous studies. Tam et al.[36] investigated how travel time reliability affects mode choice using a combined dataset from revealed and stated preference surveys. The authors state that increasing reliability can attract more passengers to use bus services. Luken and Garrow[37] studied the airport choice problem in NYC. Their analysis, based on online ticketing data, showed that the accessibility of the airport significantly affects the airport choice. They acknowledge that their model can be improved by explicitly considering peak and off-peak driving times.

GAPS IN THE LITERATURE

The existing literature on travel demand models is extensive, especially for trip generation and mode choice. Although there are many studies that address trip generation for public transit systems and the choices that people make between travel by transit and private car, taxis have not received as much attention. In large cities, taxis provide a substantial portion of all travel; e.g., roughly 11 percent of all trips in New York.[38] However, most of the literature on taxis is focused on economic analysis of taxi markets from a regulatory and supply-side perspective. There is a need for research to identify the factors that drive taxi demand. Based on existing research for transit and other modes, the important factors for taxis are likely to include demographics, land use characteristics, and properties of other modes, like transit LOS.

The availability of high-resolution GPS data for taxis in NYC provides an opportunity to address these existing gaps in the literature with a detailed study of taxi demand across space and time. Not only can demand models be developed for taxis, but they can be developed with consideration for the variations in activity patterns that occur through the course of a day.

III. DATA SOURCES AND PREPARATION

In order to conduct the proposed analysis on trip generation and mode choice for taxis, the relevant data must be collected and organized. There are three main bodies of data used for this study. The first is a complete collection of GPS taxi data for every taxi trip made in NYC within a 10-month period. Second, detailed transit schedule for the same geographic region is acquired using Google Transit. Finally, these transportation data are supplemented with demographic, employment, and land use data, which are expected to include key characteristics of the locations that are associated with the highest rates of taxi use. The following subsections describe these data sources and their limitations in more detail.

TAXI DATA

The database of taxi trips contains information of 147 million taxi trips made between February 1, 2010 and November 28, 2010. Each record includes information about where and when a trip was made, the distance traveled, and the fare paid. Specifically, the dataset includes the following data fields for each record:

- Taxi Medallion Number, Shift Number, and Driver Name;
- Pickup Location (latitude and longitude), Date, and Time;
- Drop-off Location (latitude and longitude), Date, and Time;
- Distance Traveled from Pickup to Drop-Off;
- Number of Passengers;
- Fare Paid, including breakdown by Fare, Tolls, Tips;
- Method of Payment (e.g., cash, credit card).

These data are collected by the Taxi and Limousine Commission (TLC) using the GPS and meter devices that are installed in every licensed taxi in the city.

Although there are advantages to working with big data such as this, the dataset has some limitations. For example, none of the intermediate locations that a taxi passes along a route from a passenger pickup to a drop-off are logged, so it is not possible to know for certain which routes individual taxis follow to serve each trip. For the purposes of modeling trip generation, however, it is sufficient to look at the number of trips starting (pickups) and ending (drop-offs) across space and at different times of day in the city. For mode choice analysis, trips connecting specific origin-destination pairs can be extracted from this larger dataset in order to compare the properties of taxi travel with public transit.

Preparing Taxi Data for Trip Generation Modeling

The raw taxi data requires some filtering in order to remove errors in the dataset. The deficiencies in the GPS data are mostly due to satellite errors, receiver noise errors, coordinate transformation errors, and errors made by the driver.[39] The taxi GPS data were processed to minimize the influence of outliers. Some false records were eliminated, for example, records with a total fare amount equal to zero or a travel distance less than the straight-line distance between the origin and destination. Sometimes more than two criteria were used to determine whether to remove a data point (e.g., fare amount, distance, and travel time). Ultimately, less than 2 percent of the original taxi records were eliminated through this filtering process.

The goal of the trip generation analysis was to identify which demographic, employment, and land use factors have the strongest effect on the number of taxi trips made. An additional goal of this study was to identify whether the availability and accessibility of public transit is related to the use of taxis in a neighborhood when controlling for these other factors. In order to conduct this analysis, the raw taxi data must be processed into a format that is compatible with the other data sources. Since the spatial resolution of much of the demographic data is at the level of census tracts, this is the same level of spatial resolution that should be used for aggregating the taxi data. For the taxi trip generation models developed in this study, census tracts were used as the geographic unit for the Transportation Analysis Zones (TAZs).

In addition to the spatial aggregation, the taxi data was also aggregated by hour of day so that the trip generation model could account for temporal variations in demand. The process of aggregating pickup and drop-off records for this study was similar to the process used by Yazici et al.[40] in that taxis were used as probes to monitor traffic conditions by time of day. The distribution of pickups (origins) and drop-offs (destinations) were considered separately because they are clustered differently in time and space. Thus, separate models were developed to understand these two trip ends.

This aggregated data can be visualized on maps of New York City in which each census tract is shaded based on the number of taxi trips that were observed starting or ending. Although the dataset is split into 24 hours of the day, illustrative examples are shown for the afternoon peak at 5:00 p.m. (Figure 1) and late at night at 12:00 a.m. (Figure 2). The figures provide a visualization of where pickups and drop-offs are located at different times of day. All maps were constructed with the same scale so that they can be compared directly with one another. The figures show that the locations where demand is concentrated are mostly in Manhattan and downtown Brooklyn, but a more complete statistical analysis is necessary to quantify how this demand relates to characteristics of each census tract and the transit service available at each location.

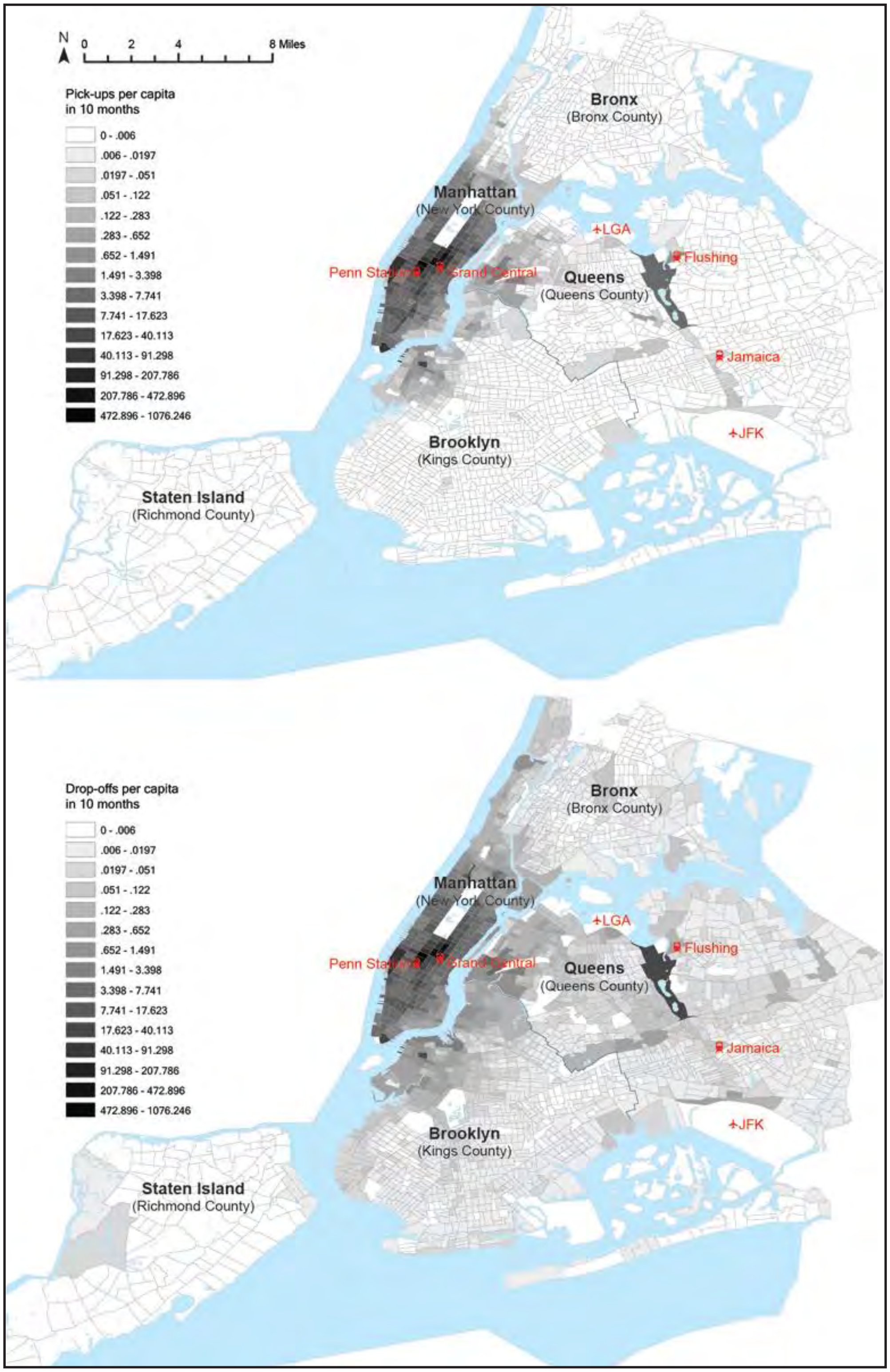

Figure 1. Taxi Pickups and Drop-Offs from 5:00 p.m. – 6:00 p.m.

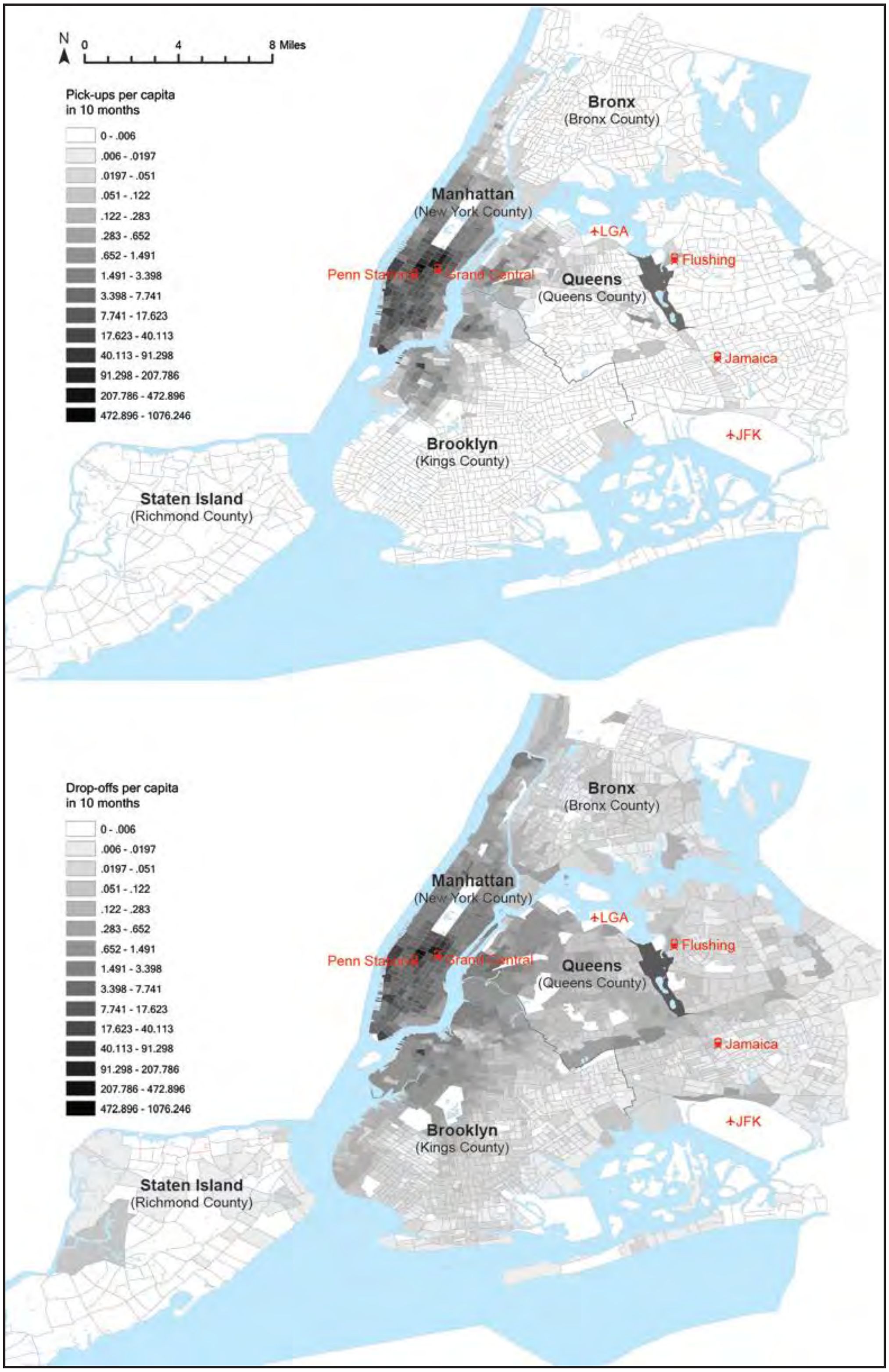

Figure 2. Taxi Pickups and Drop-Offs from 12:00 a.m. – 1:00 a.m.

Preparing Taxi Data for Mode Choice Analysis

In order to evaluate mode choice, that analysis focused on a few specific origin-destination pairs involving travel between New York Penn Station and the region's three main airports. The relevant taxi trips were extracted from the larger dataset by identifying only those that have one trip end near Penn Station and the other trip end near an airport. All airport trips within a 500-foot radius of the center of Penn Station, within a 1-mile radius of the center of EWR, and within the census tract of JFK (census tract ID: 36081071600), and LGA (census tract ID: 36081033100) were considered. As NYC taxis are not allowed to pick up passengers from EWR, the cabs must return empty, so only trips from Penn Station to EWR were included in this study. JFK and LGA have taxi trips in both directions, so a total of five OD pairs were considered. The relative locations of Penn Station and each of the region's airports is shown in Figure 3.

Figure 3. Locations of New York Penn Station and Three Major Airports: John F. Kennedy International (JFK), LaGuardia (LGA), and Newark Liberty International (EWR)

The total fare is a flat rate between most of Manhattan and JFK or EWR airports, plus any tolls, tips, and surcharges. The fare between Manhattan and LGA has some variability because trips are charged the normal metered rate. The travel time has largest variability among all four variables due to the variability of traffic conditions. The trip distance is mostly stable for all airport trips, but the slight variability indicates that alternative routes might be taken for the same origin-destination pairs.

TRANSIT DATA

Data on transit service over the same geographic area as the taxi data is necessary for determining how transit service affects taxi trip generation and how travelers choose between the two modes. The Metropolitan Transit Authority (MTA) operates extensive bus, subway, and commuter railroad services in New York City. Additionally, New Jersey Transit operates commuter rail services from Penn Station into New Jersey, including a connection to EWR. AirTrain services at JFK and EWR connect the airport terminals with the regional rail network.

In addition to published route and schedule information provided by the operating agencies, several web-based services make route and schedule information available in electronic format. Examples include Google Transit, Bing Maps, and MapQuest. For this study, data from Google was utilized in two ways. First, station and schedule information was extracted in an electronic format for analysis of transit accessibility. Second, a specialized program was developed to make use of the trip-planning functionality of Google Maps API Transit Directions Service.

Preparing Transit Data for Modeling Taxi Trip Generation

One of the goals of this study was to identify what role that transit accessibility plays in determining the number of taxi trips that start or end in a census tract. Therefore, information is needed about the locations of transit stations and the frequency of service at these stations. This information is available in a standardized electronic format called the Google Transit Feed Specification (GTFS). Transit agencies submit their route and schedule information in GTFS to Google so that users of Google Maps can search for directions online. The GTFS data is a series of files that describe the locations of transit stations, the sequence of stations served by each route, and sets of times when vehicles depart from each station. Together, these make up a comprehensive description of all routes and schedules operated by an agency.

The raw data from GTFS can be directly combined to determine the number of scheduled vehicle departures per hour from each transit station. For example, Figure 4 shows a map of subway stations in NYC with each station shaded to represent the number of subway trains serving the station in the 5:00 p.m. to 6:00 p.m. hour on weekday afternoons. Additional analysis and calculations are required to convert the information on this map into a measure of transit accessibility for each census tract in the city. That process is part of the methodology for this study, and the details are described in the section about trip generation modeling. Although rail schedules are complete throughout the region, some data for bus routes is missing from Google’s databases (notably, bus routes in Queens are not currently available in the GTFS format). Due to this limitation of the data, the analysis in this study is based only on measuring subway accessibility in New York.

Figure 4. Frequency of Subway Service at Each Station in New York City Between 5:00 p.m. – 6:00 p.m.

Preparing Transit Data for Mode Choice Modeling

Whereas trip generation models are constructed only based on the characteristics of the locations at the beginning and end of a trip, mode choice models require information about the complete trip itself. In order to compare taxi and transit service between Penn Station and the airports in the NYC region, specific information about travel times and fares are needed for each of the relevant origin-destination pairs. The transit service to each airport is summarized as follows:

- JFK, located in Queens, NY, is accessible from AirTrain, buses, and car/taxis. AirTrain JFK connects to the Long Island Rail Road (LIRR) and the NYC subway and bus system at Jamaica and Howard Beach.

- LGA, located in Queens, NY, is four miles from Manhattan and can be accessed by car or taxi. LGA does not have a direct rail link, but bus services do connect to the LIRR and subway.

- EWR is located in Newark, NJ, and is accessed from Manhattan via the Holland and Lincoln Tunnels by car/taxi. AirTrain Newark provides access to New Jersey Transit trains into NYC Penn Station.

The travel time and monetary costs of each airport trip was summarized by hour of the day in the taxi dataset, so comparable data was required for the competing transit trip. Without live data collection of travel times for passengers, the next best data source for assessing transit is to look to the schedules. The travel time and fare for using transit depends on the specific route selected (e.g., whether only subway and bus are used, or if commuter rail is used).

Google Maps API Transit Directions Service, which offers free transit route guidance with a daily request limit, was used to obtain transit data. The information was gathered in XML format using a web-based JavaScript code. An application was developed that extracts one week of travel time and route information (including weekdays and weekends) based on schedules for the five origin-destination pairs every 5 minutes throughout the day. The routing information provided by Google was assumed to be the optimal transit option for the requested time and origin-destination pair since the web-based routing service compiles all available scheduling information for different transit modes and routes. The fare was estimated based on the optimal route. Data for approximately 2,016 transit trips have been collected for each origin-destination pair over a 10-month period. The transit travel duration of each trip includes waiting time, transfer time, and in-vehicle travel time. Each data point was also associated with an estimated fare based on the service utilized to complete the trip.

DEMOGRAPHIC, EMPLOYMENT, AND LAND USE DATA

In addition to characteristics of the taxi and transit modes themselves, there are characteristics of the places that where taxi trips start and end that are likely to have an effect on the magnitude of taxi demand. The literature on trip generation models shows that population, demographic characteristics of the population, employment, land use, and other characteristics of a transportation analysis zone can all have important explanatory power for predicting the number of trips that each zone generates. Much of this data is collected and made available by government entities such as the United States Census Bureau. Since so many of the relevant population characteristics are aggregated at the level of census tracts, this is a logical scale for analyzing taxi demand.

The sources of data for the explanatory factors considered in this study include:

- demographic data for each census tract is available from the U.S. Census 2010, including total population, population categorized by age, and population categorized by race;

- socioeconomic data is available from the American Community Survey 5-year (2007-2011) estimate of education and income;

- employment data by census tract, including categorization by age, earnings, type, race, ethnicity, educational attainment, and sex is available for NYC from 2010 Workplace Area Characteristic (WAC) data available from the U.S. Census Bureau;

- geographic data including relevant shapefiles (e.g., rivers, roads, county, census tract), and land area.

The population density and employment density in 2010 was calculated for all 2,167 census tracts in NYC. Figure 5 shows that the population density and employment density are concentrated in Manhattan. Some census tracts consisting of cemeteries, parks, or islands do not have employment associated with them, so the WAC employment data covers 2,143 census tracts. Census tracts with variables that are lacking certain required information are excluded from the linear model analysis; e.g., where the population or employment are zero. Ultimately, 116 out of 2,167 census tracts (5 percent) were omitted from the analysis, because there was insufficient population or employment in those few regions to create a useful data point. With all of the demographic, employment, and land use data aggregated by census tract, the dataset was prepared with a large set of variables that can used to develop models for taxi trip generation.

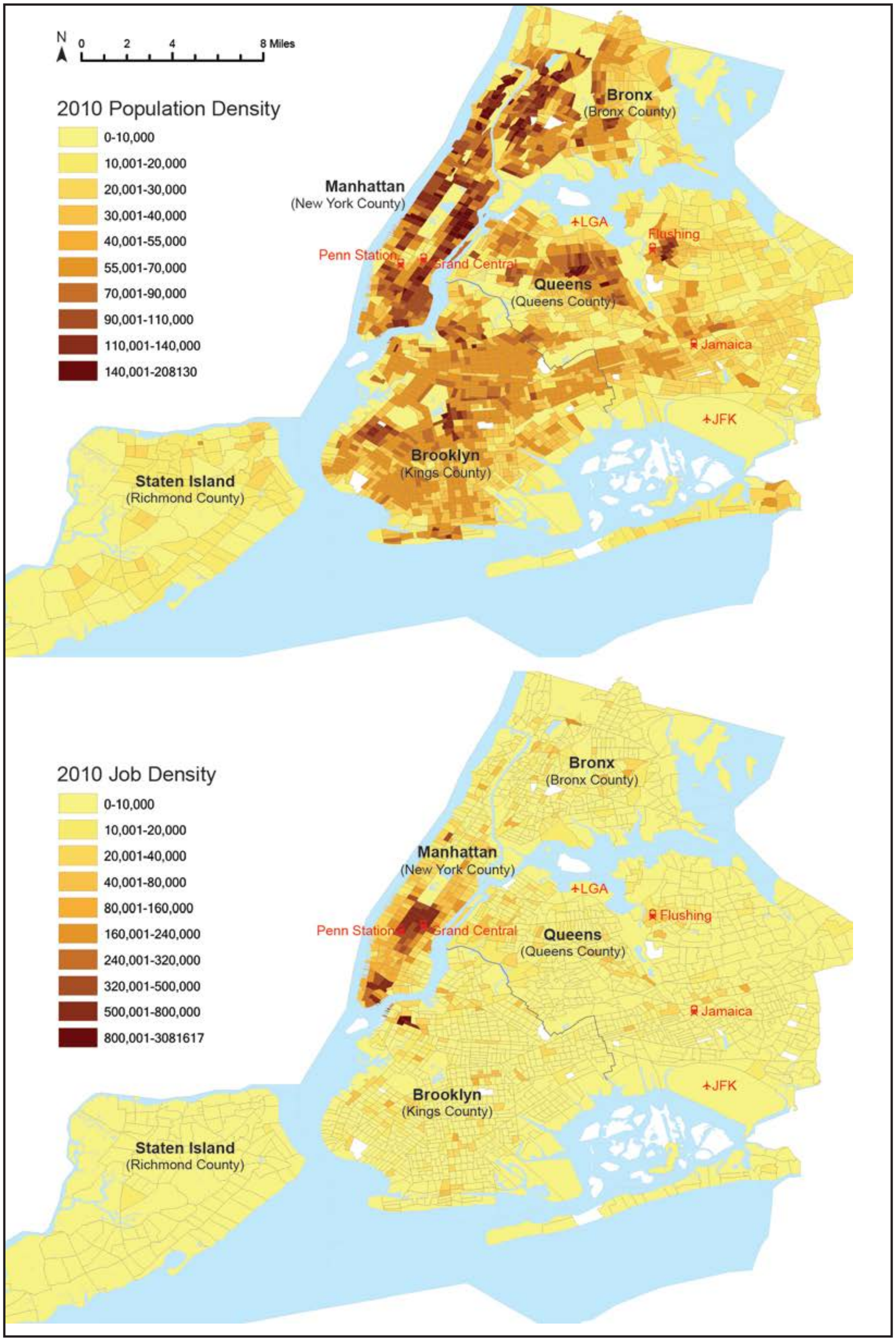

Figure 5. 2010 Population and Job Density (per square mile)

IV. ANALYSIS 1: TAXI TRIP GENERATION

The first analysis conducted in this study was the development of a trip generation model for taxi demand that accounts for the role that transit plays in determining taxi demand. There are two important methodological contributions of this trip generation study. The first was the development of a novel transit accessibility measure based on the time to access and wait for transit. This requires processing raw transit schedule information to determine how much time it takes person at a specific location and time of day to access the public transit system. The second contribution was the development of a hybrid cross-classification and regression model for estimating taxi trip generation. The taxi data was cross-classified by pickup and drop-off and aggregated by hour of the day. Within each classification, a multiple linear regression model was estimated to identify the factors that influence taxi demand.

DEFINING TRANSIT ACCESS TIME (TAT)

Transit LOS and accessibility must be quantified in order to be used as an explanatory variable to model taxi. A new measure was developed that combines the estimated walking time a person must spend to access the nearest station (transit accessibility) and the estimated time that person will wait for transit service (transit LOS).

This measure is the Transit Access Time (TAT), and it represents the minimum expected time for a person at a specific location and time of day to walk to, wait for, and board a transit vehicle. For a walking speed of 3.1 mph,[41] the transit access time in minutes is:

$$\mathrm{TAT} = \frac{60D}{v_w} + \frac{60}{f} \qquad (1)$$

where f is the frequency of subway dispatches per hour at the nearest station, D is the distance to the nearest station (mi), and v_w is the walking speed (mph).

The minimum TAT was calculated at each location by the following steps. First, the transit schedule in GTFS provides the number of transit departures (i.e., frequency) in each hour at each station. The waiting time depends on the frequency based on the second term of Equation 1, and it was calculated separately for each hour of the day to account for variations in the schedule. Then, a fine grid was imposed on the study area with cells measuring 250 meters square, which is small enough that the walking time to cross each cell is less than one minute. Each cell was characterized by the location of its centroid, and a TAT was calculated for each cell. A modified K-nearest-neighbor algorithm was implemented by calculating the minimum TAT from the K nearest transit stations by screening distance and waiting time to all transit stations from the centroid.

People are assumed to be well informed about transit schedules and to choose the nearby station that minimizes the sum of their walking and waiting time. Thus, the TAT is a metric of transit accessibility that is independent of specific origin-destination demand patterns. For simplicity, the method looks only at the closest access from each location (cell centroid) to the nearest subway departure, in space and time, anywhere in the system. The minimum

TAT was calculated for each cell in NYC at each hour of the day, and this was used to quantify transit accessibility in the city with spatial resolution of 250 meters and temporal resolution of an hour.

The minimum TAT for each census tract was determined by averaging the values across all of the 250-meter square cells included within the census tract. This provides a better TAT measure than simply calculating from census tract centroids, because a large census tract may have a centroid near a transit station but extensive peripheral land that has relatively low accessibility. The TAT was calculated for different times of day for each census tract using only the subway data in this study, because the GTFS bus schedule data that is available from the MTA is incomplete (e.g., bus data for Queens are not available).

VISUALIZING TRANSIT ACCESS TIME AND DEMAND

In order to develop an intuitive understanding of what that TAT represents, it is useful to look at a visualization of how it varies over space and time. Figure 6 shows the TAT for subways during 5:00 p.m. – 6:00 p.m. (afternoon) and 12:00 a.m. – 1:00 a.m. (midnight). The map of TAT shows that there is greater transit accessibility in Manhattan and along the subway routes than in other parts of the city, which is expected, based on the spatial coverage of the subway network. The transit accessibility is also generally greater during 5:00 p.m. – 6:00 p.m. than at 12:00 a.m. – 1:00 a.m., because services operate more frequently during the peak hours than late at night.

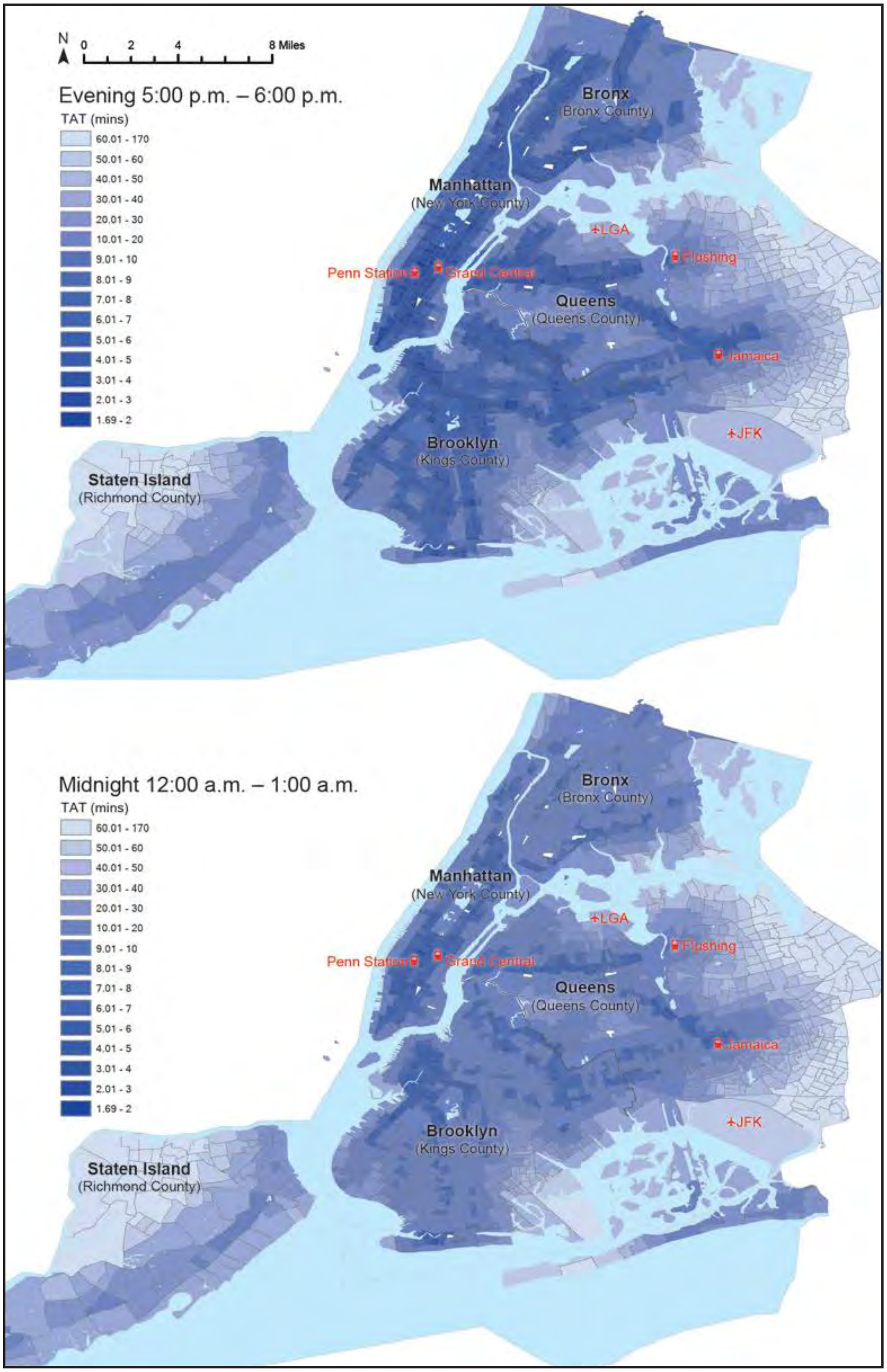

Figure 6. Transit Access Time (minutes) During Evening Rush (5:00 p.m. – 6:00 p.m.) and Midnight (12:00 a.m. – 1:00 a.m.)

The visualization of TAT can be compared with the mapping of taxi pickups and drop-offs during the evening and midnight hours shown in Figure 1 and Figure 2, respectively. These figures suggest that the pickups and drop-offs per capita are higher where the TAT is lower (i.e., transit is more accessible), which is a negative correlation between TAT and taxi use. The mapping of TAT and taxi demand provide a visualization of their relationship and help provide intuition about why such a relationship exists.

These visualizations also show that it is necessary to split the dataset up by hour of the day, because the distribution of activities in NYC changes with time. There are also differences between the rates of taxi pickups per capita at 5:00 p.m. and at 12:00 a.m. For example, there are more taxi pickups at Jamaica at 5:00 p.m. than that at 12:00 a.m., which could result from people getting off the subway at Jamaica and then taking a taxi complete a trip home from work. In some areas of lower Manhattan there are more pickups at 12:00 a.m. than at 5:00 p.m., which indicates concentrations of nightlife.

The drop-offs per capita show big differences between 5:00 p.m. and 12:00 a.m. as well. For example, there are more drop-offs per capita at some popular locations such as Penn Station, Grand Central Station, and Flushing at 5:00 p.m. than at 12:00 a.m., which is consistent with the fact that these are busy transit hubs that used by commuters. Although the total amount of travel activity in the city is lower at midnight than at 5:00 p.m., many areas of the outer boroughs actually see a greater rate of drop-offs in the late night hours. This suggests that people use taxis more often to travel to outlying neighborhoods when it is dark and transit services are less frequent. There appears to be a consistent trend at all times of day that pickups are more concentrated around transit hubs and central areas, whereas drop-offs are more dispersed around the city. Clearly, trip-making behavior by taxis is asymmetric in that trip origins are more concentrated than trip destinations.

With the hourly data for TAT, taxi pickups, taxi drop-offs, and all other demographic and socioeconomic information, visual inspection of the maps is interesting but insufficient for determining the quantitative relationship between the explanatory variables and the taxi demand. A multiple linear regression model is introduced in the next section in order to achieve this objective.

METHOD FOR MODELING TAXI DEMAND

Linear models have been broadly applied to trip generation.[42] The idea behind multiple linear regression modeling is to explore the relationship between the dependent variable and independent variables with the assumption that this relationship is linear as follows:

$$Y = \sum_{i=0}^{n} \beta_i X_i + \varepsilon \qquad (2)$$

where Y is the number of taxi trips generated in a TAZ (response variable), X_i is one of n independent variables, X_0 is the intercept, β_i is the coefficient corresponding to X_i, and $\mathcal{E}$ is the error representing the difference between the modeled and observed number of taxi trips.

Using least squares estimation (i.e., maximum likelihood estimation) coefficients were estimated for each explanatory variable by minimizing the mean squared error between the modeled Y and observed Y. The goal was to select a set of explanatory variables that results in low model error and in which each explanatory variable has a statistically significant coefficient. There are many methodological and statistical criteria for selecting important variables. For example, stepwise selection and best subset regression are two methods for comparing model specifications in order to identify the best set of explanatory variables to include in the final model. The following steps describe several procedures used to select important variables in this study.

Step 1: Check Correlation Coefficients

An analysis of the correlation coefficients among the response variable and all explanatory variables shows how closely each pair of variables vary with each other. A correlation coefficient that is greater than 0.5 or less than -0.5 was considered strong in this analysis. The strong correlation between an explanatory variable and the response variable could indicate the explanatory variable is important. Strong correlation among explanatory variables leads to multicollinearity in the model, because it is not possible to identify which factor has the more significant statistical relationship with the response variable. The Variance Inflation Factor (VIF) quantifies the severity of multicollinearity in an Ordinary Least Squares regression by measuring how much the variance of an estimated regression coefficient increases due of multicollinearity.[43] Each indicator has a VIF value to indicate the degree of multicollinearity, and a large value indicates that a variable needs to be either removed or replaced. A common rule of thumb is that if the VIF of each factor is larger than five, then multicollinearity is high.

Step 2: Stepwise Selection

Stepwise selection (or forward and backward selection) is a method of selecting variables by adding or eliminating one at a time. The best model was chosen by seeking the model with the lowest value of Akaike Information Criterion (AIC) and smaller Residual Sum of Squares (RSS). AIC is a measure of the complexity of the model, and it is a function of maximum likelihood and the number of parameters included in the model. A smaller AIC value indicates a better goodness of fit.[44] The AIC value is especially useful when comparing models with a large number of explanatory variables. The stepwise method involves ranking the importance of each factor by listing the AIC values would result from removing it. Then, the least relevant factors can be eliminated one by one until a suitable model is specified.

Step 3: Best Subsets Regression

Best subsets regression (a.k.a. complete subset regression) is a method to select the best subset of predictors among all the possible combinations of predictors (2^k combinations if there are k predictors in the initial model).[45] There are several metrics for comparing model performance:

- R squared (R^2) is the coefficient of determination that quantifies the variance in the model error, and it is also an indicator of how well the model fits the data points.

- Adjusted R squared ($AdjR^2$) is similar to R^2 but incorporates a penalty for the number of extra explanatory variables added to the model; a higher $AdjR^2$ is better.

- Bayesian Information Criterion (BIC), which is similar to AIC, is a function of the maximized value of the likelihood function and the number of variables included in the model. The difference from AIC is that the penalty term for the number of variables included in the model is larger in BIC than in AIC.[46] For both metrics, lower values are an indication of a better model.

- Mallows C_p assesses overfitting of the model, and a desirable model has C_p close to the number of explanatory variables, p.[47]

The best subset method works well to refine the selection of explanatory variables from the important factors that are already identified. It is very useful for modeling the same major pickups and drop-offs at different times of day based on the same set explanatory variables, because trips at different times of day could be associated with different explanatory factors.

It is very difficult to achieve an R^2 greater than 0.8 in most trip generation studies, because there are many things affecting the response variable, and we sought the simplest possible model to gain insights for transportation planning. There have been some studies by transportation planners on regional growth[48] and trip generation[49] using linear regression and achieving very low R^2 or adjusted Adj R^2 (much less than 0.5 sometimes less than 0.1), however the value of these models is not in the final estimate of the response variable but in identifying statistically significant explanatory variables that help us understand what drives demand. The goal of this study was to identify the relationships between taxi demand and important socioeconomic and land use factors at different times of day and at different locations. Therefore the models were developed based not only on R^2 but also on other criteria used to select an appropriate model. In order to use the fewest number of variables for the model, the most statistically significant explanatory variables were identified by the t-statistic or p-value (p-value<0.05 is significant at 95% confidence level).

TAXI TRIP GENERATION MODEL AND DISCUSSION

Using the methodology presented in the previous section, an initial model was developed and refined to predict the number of pickups or drop-offs generated in each census tract by hour of the day from the 10-month taxi GPS data in NYC. The full list of the explanatory variables that were considered in the initial model are listed in Table 1.

Table 1. List of Explanatory Variables in Each Model

Factor group	Factors or factor category	No. of variables	Initial model	All NYC factors	Manhattan model
TAT	TAT at specific hour	1	√	√	√
Population	Total population (Pop)	1	√	√	√
	Population by race	8	√		
	Population by age	14	√		
Age	Medium age (MedAge)	1	√	√	√
Education	Percentage education higher than high school	1	√		
	Percentage education higher than Bachelor (EduBac)	1	√	√	√
Income	Median household income	1	√		
	Mean household income	1	√		
	Median family income	1	√		
	Mean family income	1	√		
	Per capita income (CapInc)	1	√	√	√
Employment	Total jobs (TotJob)	1	√	√	
	Jobs by age	3	√		
	Jobs by earnings	3	√		
	Jobs by types	20	√		√
	Jobs by race	6	√		
	Jobs by ethnicity	2	√		
	Jobs by education attainment	4	√		
	Jobs by sex	2	√		
Total No. of variables		70	70	6	25

Several influential factors from the initial full model have been identified using stepwise selection based on AIC values and RSS: TAT, total population, median age, educational attainment, income, total jobs, and jobs by type. The correlation coefficient was checked to remove factors that are too closely related to each other in selecting major factors for the second model. Since median household income, mean family income, and per capita income are highly correlated, only one should be included in each model to avoid multicollinearity. Due to better performance of the model with per capita income and the higher correlation coefficient with the response variables, per capita income was selected. Similarly, jobs by type or jobs by sex are closely related to total jobs. In this case, total jobs was chosen for the second model (selected factors are listed in Table 1). To prevent multicollinearity, only one factor or category among two or more correlated factors was included.

Models with and without the intercept were estimated for pickups and drop-offs for each hour of day in NYC. In most of the models the intercept was not significant, and it is intuitive that if a census tract has no population and no jobs, then there are likely to be no trips as well. The coefficients of the other explanatory variables were very similar whether or not the intercept was included in the model. Therefore, the intercept was removed from the models formulated in this study. The results, including the six major variables for each time of the day, are presented in Table 2 for taxi pickups and Table 3 for taxi drop-offs.

Table 2. Model Fit Statistics and Coefficients for Generation of Pickups in All of New York City

Hour	Model Fit Statistics				Coefficients of Explanatory Variables					
	R^2	AdjR^2	C_p	BIC	Pop	MedAge	EduBac	CapInc	TAT	TotJob
12 a.m.	0.47	0.46	5.87	-1248.26	0.48	-198.18		0.26	-36.38	0.32
1 a.m.	0.38	0.38	4.00	-943.85	0.38	-142.05		0.19	-30.63	0.21
2 a.m.	0.30	0.30	4.90	-704.99	0.31	-103.55		0.14	-23.54	0.13
3 a.m.	0.26	0.26	5.67	-577.45	0.23	-72.69		0.10	-17.65	0.09
4 a.m.	0.32	0.32	5.74	-753.91	0.17	-52.83		0.07	-11.97	0.07
5 a.m.	0.52	0.52	5.14	-1464.34	0.16	-49.18		0.06	-7.56	0.06
6 a.m.	0.48	0.48	6.00	-1303.97	0.35	-112.34	-25.36	0.15	-13.72	0.14
7 a.m.	0.56	0.56	6.00	-1621.84	0.64	-210.90	-64.23	0.31	-23.79	0.24
8 a.m.	0.61	0.61	6.00	-1872.85	0.70	-255.40	-99.03	0.41	-32.77	0.36
9 a.m.	0.61	0.61	6.00	-1886.77	0.62	-249.00	-106.00	0.42	-37.68	0.43
10 a.m.	0.62	0.62	6.00	-1959.95	0.57	-228.94	-103.30	0.39	-37.49	0.43
11 a.m.	0.63	0.63	6.00	-1990.18	0.47	-216.05	-114.07	0.40	-39.79	0.49
12 p.m.	0.63	0.63	6.00	-2002.51	0.44	-221.88	-125.07	0.43	-43.76	0.54
1 p.m.	0.63	0.63	6.00	-2019.18	0.41	-216.54	-125.18	0.43	-44.67	0.53
2 p.m.	0.63	0.63	6.00	-2013.17	0.40	-219.94	-132.22	0.44	-46.60	0.55
3 p.m.	0.64	0.64	6.00	-2048.37	0.41	-213.26	-121.48	0.42	-43.62	0.49
4 p.m.	0.64	0.64	6.00	-2059.71	0.39	-190.62	-101.24	0.36	-37.57	0.42
5 p.m.	0.65	0.64	6.00	-2082.61	0.49	-237.27	-123.22	0.44	-44.38	0.50
6 p.m.	0.64	0.64	6.00	-2034.77	0.57	-285.92	-155.81	0.54	-54.72	0.63
7 p.m.	0.63	0.63	6.00	-1981.57	0.60	-300.80	-154.44	0.56	-56.44	0.67
8 p.m.	0.62	0.61	6.00	-1915.26	0.55	-277.92	-129.10	0.50	-52.44	0.63
9 p.m.	0.59	0.59	6.00	-1790.40	0.56	-266.61	-111.59	0.47	-52.15	0.60
10 p.m.	0.56	0.56	6.00	-1642.60	0.56	-257.86	-92.14	0.44	-50.00	0.56
11 p.m.	0.53	0.53	6.00	-1504.76	0.55	-236.71	-54.38	0.37	-44.68	0.45

Table 3. Model Fit Statistics and Coefficients for Generation of Drop-Offs in All of New York City

Hour	Model Fit Statistics				Coefficients of Explanatory Variables					
	R^2	AdjR^2	C_p	BIC	Pop	MedAge	EduBac	CapInc	TAT	TotJob
12 a.m.	0.60	0.59	6.00	-1809.31	0.67	-190.57	19.37[a]	0.22	-36.05	0.22
1 a.m.	0.60	0.59	6.00	-1812.34	0.52	-136.56	25.23	0.14	-27.86	0.16
2 a.m.	0.59	0.59	6.00	-1796.86	0.41	-100.30	24.64	0.10	-20.09	0.12
3 a.m.	0.61	0.61	6.00	-1900.70	0.30	-68.96	18.14	0.06	-14.69	0.09
4 a.m.	0.59	0.59	6.00	-1782.06	0.18	-40.04	10.99	0.04	-10.13	0.07
5 a.m.	0.45	0.45	6.00	-1169.35	0.06	-22.93	-7.29	0.04	-7.71	0.11
6 a.m.	0.43	0.43	4.02	-1120.11		-49.12	-54.92	0.14	-16.61	0.39
7 a.m.	0.47	0.47	4.16	-1262.46		-95.27	-109.00	0.27	-32.15	0.71
8 a.m.	0.53	0.53	4.00	-1528.36		-132.37	-129.14	0.36	-41.17	0.83
9 a.m.	0.57	0.57	4.24	-1706.02		-140.65	-131.07	0.37	-44.19	0.76
10 a.m.	0.60	0.60	6.00	-1840.99	0.17	-156.26	-114.32	0.36	-42.18	0.60
11 a.m.	0.61	0.61	6.00	-1876.01	0.23	-168.98	-117.39[a]	0.37	-43.79	0.56
12 p.m.	0.62	0.62	6.00	-1952.91	0.31	-190.60	-122.02	0.40	-46.73	0.56
1 p.m.	0.62	0.62	6.00	-1957.41	0.33	-192.48	-116.00	0.40	-45.16	0.55
2 p.m.	0.62	0.62	6.00	-1923.38	0.39	-204.54	-116.39	0.41	-45.51	0.54
3 p.m.	0.62	0.62	6.00	-1943.27	0.44	-207.17	-109.44	0.39	-43.12	0.47
4 p.m.	0.62	0.62	6.00	-1958.00	0.45	-192.80	-90.60	0.35	-37.40	0.38
5 p.m.	0.64	0.64	6.00	-2052.05	0.65	-251.19	-96.90	0.42	-42.86	0.42
6 p.m.	0.65	0.65	6.00	-2122.94	0.86	-317.45	-106.75	0.50	-51.41	0.44
7 p.m.	0.63	0.63	6.00	-2019.56	0.95	-338.33	-92.14	0.51	-52.89	0.43
8 p.m.	0.64	0.64	6.00	-2047.76	0.98	-327.00	-56.46	0.45	-49.35	0.34
9 p.m.	0.66	0.66	6.00	-2163.16	0.97	-312.84	-42.50	0.42	-48.29	0.33
10 p.m.	0.66	0.66	6.00	-2182.20	0.95	-297.63	-26.59[a]	0.38	-45.89	0.33
11 p.m.	0.63	0.63	4.00	-2026.33	0.84	-254.24		0.31	-42.39	0.29

[a] Indicates non-significance of the coefficient, p-value>0.05, otherwise, it is significant.

The interpretation of the trip generation results for both pickups and drop-offs is useful for transportation planning and regulation of taxi services. The magnitude and sign of the coefficient for each explanatory variable indicates how much taxi demand will increase (for positive coefficients) or decrease (for negative coefficients) as the explanatory variables increase by one unit. For example, the coefficient of 'TotJob' is 0.32 for pickups at 12:00 a.m. in NYC (Table 2) indicating that an increase of one job in a census tract is associated with an average increase of 0.32 taxi pickups in the 12:00 a.m. hour over a 10-month period. Similarly, there is an average decrease of 36 taxi pickups at the same hour over a 10-month period as TAT increases by one minute, which provides insight about how dramatically taxi demand changes with the availability and accessibility of transit service.

The errors of the trip generation model (i.e., difference between observed and modeled taxi demand) provide information on when and where taxi demand is underestimated or overestimated. This gives some idea of where and when more taxi use would be expected than actually occurs, based on citywide trends, so the information can be useful for planning locations of taxi stands or providing incentives for cab drivers to operate over certain times of day and in certain parts of the city. At locations where the model estimates higher taxi use than is actually realized, it is possible that there is a latent demand that goes underserved because there are simply not enough taxis circulating at the specific location and time to carry as many passengers as would like to use taxis.

The results show that population, education, income, and total jobs positively influence both taxi pickups and drop-offs in NYC. This is expected, because high total population and high total number of jobs are indicators of places with high human activity and where people are more likely to be traveling by any mode, including taxi. However, median age and TAT negatively affect the trip-making by taxis. This shows that younger people are more likely to take taxis. The results also show that taxi demand is high where transit is more accessible (TAT is small). It is not clear from the available data whether the relationship between taxis and transit is competitive or complementary. Thus, it cannot be concluded whether the convenience of transit service in an area causes high taxi demand because people use taxis to complement transit or if the large number of taxi trips are associated with high levels of activity that also happen to be where high levels of transit service are provided. The reality is likely that taxis and transit are sometimes operating in competition and other times as complements, because both modes follow and influence the levels of activity in neighborhoods across the city.

The distribution of coefficients at different times of day also sheds light on how those factors influence the number of taxi pickups and drop-offs. For example, the total number of jobs has a higher influence on taxi demand from 7:00 a.m. to 6:00 p.m., which indicates that extra taxi demand during this period in NYC is likely caused by people going to and from work or work-related activities. The coefficients for TAT values from 8:00 a.m. to 11:00 p.m. show increased taxi trips associated with good transit accessibility (short TAT) during all but the late and overnight hours, so it is possible that many of the trips are being made to or from transit facilities, enabling taxis to complement transit service. It is also possible that the places that have good transit service are also desirable for taxi use for other reasons. For example, it might be easier to hail a cab on busy streets in Manhattan under which the busiest subway lines also run.

Another interesting observation from the stepwise modeling is that some of the variables in the category of jobs-by-type are very influential in the linear model performance, especially for the pickups and drop-offs in Manhattan, as listed in Table 4 and Table 5. TAT loses its influence for drop-off trips in Manhattan, compared to when “total jobs” was used. Factors related to job types seem to play key roles in generating the taxi trips in Manhattan. Some of the influential industry sectors are retail, accommodation and food service, and health care.

From 11:00 p.m. to 8:00 a.m., it appears the drop-off taxi demand is not significantly related to income, while from 9:00 a.m. to 10:00 p.m. it is. This indicates that people are taking taxis in the evening regardless of income; however, in the daytime, wealthy people are more likely to take taxis, perhaps because those in the lower income brackets have access to more competitive, affordable travel modes during the day. Similar situations were also observed for taxi pickup demand except that the time period is slightly earlier.

People tend to take taxis to places with retail activities from 8:00 a.m. to 4:00 p.m. (Table 3) and taxi trips away from these places from 12:00 p.m. to 11:00 p.m. These retail-related activities could be commuting to jobs in retail sales, shopping to purchase goods, or meeting with other people. Unfortunately, without data about individual trip purposes, it is not possible to know precisely which activities each traveler in the census tract engaged in, but the high correlation with retail activities shows the importance of retail land use and employment in determining taxi demand.

Accommodation and food service jobs, which are an indication of hotel and restaurant activity, are located all over Manhattan. It is not surprising to see, from pickup and drop-off trip generation coefficients, that they are influential almost all day, but they have relatively higher influence at breakfast time (7:00 a.m. – 9:00 a.m.), lunch time (1:00 p.m. – 2:00 p.m.), and dinner time (5:00 p.m. – 11:00 p.m.). These results provide us with a thorough understanding of the relationship between taxi demand and popular activities in Manhattan. If combined with other information, such as population, income, and TAT, the models provide predictions of taxi demand across time and space.

Table 4. Model Fit Statistics and Coefficients for Generation of Pickups in Manhattan

Hour	Model Fit Statistics				Coefficients of Explanatory Variables													
	R^2	$AdjR^2$	C_p	BIC	Pop	MedAge	EduBac	CapInc	TAT	JobCon	JobRet	JobTrW	JobFin	JobRea	JobPro	JobHea	JobEnt	JobFod
12 a.m.	0.78	0.77	11.68	-339.20			230.85		-431.29	15.88			-1.01	-13.25				20.26
1 a.m.	0.69	0.68	11.71	-255.41			177.18		-285.36	13.91			-1.01	-14.24				16.47
2 a.m.	0.61	0.60	9.49	-202.36			97.03			13.48			-1.11	-14.34			-2.49	15.23
3 a.m.	0.57	0.56	7.09	-176.59			70.29			10.51			-0.88	-11.85			-2.42	11.85
4 a.m.	0.66	0.65	17.14	-231.99			48.85			7.25			-0.56	-7.30			-1.30	7.81
5 a.m.	0.76	0.75	25.55	-321.28	0.29		50.97		-229.78			3.41				0.31		2.26
6 a.m.	0.65	0.65	13.51	-231.69	0.82	-177.08		0.13							1.53	0.80	2.41	
7 a.m.	0.74	0.73	15.21	-302.89	1.50	-348.46		0.25							2.76	1.27	3.87	
8 a.m.	0.81	0.80	29.08	-378.73	1.52	-395.80		0.32							4.26	1.46	4.80	
9 a.m.	0.83	0.82	43.97	-404.27	1.26	-360.67		0.29							3.07	1.66		8.30
10 a.m.	0.86	0.86	57.19	-454.41	1.01	-308.34		0.24						15.05		1.63		8.87
11 a.m.	0.89	0.88	61.13	-508.46				0.17						11.95	2.04	1.16	3.95	6.55
12 p.m.	0.90	0.90	67.32	-538.85				0.17			7.23				3.46	1.02	3.54	7.39
1 p.m.	0.91	0.91	66.95	-564.52		-141.02		0.23			9.17				2.61	1.26		9.15
2 p.m.	0.91	0.91	63.60	-580.34				0.16			8.98				3.00	0.96	3.68	8.35
3 p.m.	0.91	0.91	64.45	-567.34		-154.67		0.22			8.03			10.51		1.37		9.42
4 p.m.	0.91	0.90	64.95	-554.60		-93.66		0.20			6.89				1.87		3.21	6.96
5 p.m.	0.91	0.90	63.58	-555.30	0.98	-288.01		0.24			8.46				2.31			10.73
6 p.m.	0.92	0.91	58.17	-583.47				0.22			6.59		-1.11		4.10		4.94	13.84
7 p.m.	0.92	0.91	55.86	-584.17				0.22			5.94		-1.24		4.04		5.30	16.35
8 p.m.	0.91	0.90	54.69	-557.68		-250.61	335.81				7.85		-1.19		3.32			18.86
9 p.m.	0.90	0.89	35.05	-528.50	0.93	-382.70	310.79				7.11				1.72			19.76
10 p.m.	0.88	0.88	21.45	-495.67	0.84	-363.06	289.95				5.46					0.84		22.69
11 p.m.	0.84	0.84	19.30	-428.32	0.77	-332.23	285.40				4.16		-0.93					21.11

Notes: ‘Column data represents the number of jobs by industry. Column headers represent the following industries: JobCon’: Construction; ‘JobRet’ Retail Trade; ‘JobTrW’: Transportation and Warehousing; ‘JobFin’: Finance and Insurance; ‘JobRea’: Real Estate and Rental and Enterprises; ‘JobPro’: Professional, Scientific, and Enterprises; ‘JobHea’: Health Care and Social Assistance; ‘JobEnt’: Arts, Entertainment, and Recreation; ‘JobFod’: Accommodation and Food Services.

Table 5. Model Fit Statistics and Coefficients for Generation of Drop-Offs in Manhattan

Hour	Model Fit Statistics				Coefficients of Explanatory Variables													
	R^2	$AdjR^2$	C_p	BIC	Pop	MedAge	EduBac	CapInc	TAT	JobCon	JobRet	JobTrW	JobFin	JobRea	JobPro	JobHea	JobEnt	JobFod
12 a.m.	0.81	0.81	24.64	-384.88	1.19	-273.30	221.63			10.95			-0.86					10.66
1 a.m.	0.81	0.81	27.16	-380.93	0.98	-212.91	160.73			8.49			-0.56					7.07
2 a.m.	0.81	0.80	30.52	-376.12	0.79	-160.57	122.33			8.41				-4.91				5.56
3 a.m.	0.83	0.82	31.02	-401.68	0.58	-112.09	82.79			6.32				-2.97				3.59
4 a.m.	0.81	0.80	36.56	-379.22	0.33	-55.91	42.52					3.39				0.29		1.98
5 a.m.	0.73	0.72	5.36	-293.50	0.11						0.89				0.79	0.42	0.70	1.49
6 a.m.	0.78	0.77	2.85	-348.22									1.25	4.27	0.81	1.16		6.74
7 a.m.	0.85	0.84	8.74	-437.84									1.80	15.24	1.96	1.69		10.28
8 a.m.	0.89	0.89	24.82	-529.64							5.35			16.13	4.66	1.50		11.27
9 a.m.	0.92	0.92	35.50	-599.30				0.07			8.27				5.84	1.14		10.49
10 a.m.	0.92	0.91	48.88	-587.50				0.10			9.15				4.00	1.23		8.19
11 a.m.	0.91	0.91	54.87	-579.39				0.11			9.95				3.08	1.14		9.06
12 p.m.	0.92	0.92	58.89	-604.07				0.14			9.50				2.88		3.22	9.24
1 p.m.	0.91	0.91	51.34	-583.66				0.13			8.14				2.48	1.21		11.17
2 p.m.	0.89	0.89	55.47	-523.75				0.15			8.50				2.69	1.35		9.97
3 p.m.	0.87	0.87	62.04	-485.79				0.17			7.77				2.70		3.79	6.42
4 p.m.	0.86	0.86	62.77	-467.86				0.16			5.59				2.12		3.46	5.94
5 p.m.	0.87	0.87	68.76	-480.04				0.23					-1.18		2.93		4.23	10.35
6 p.m.	0.88	0.88	49.48	-510.84	1.47	-369.24		0.33								1.56		17.11
7 p.m.	0.88	0.88	34.37	-498.22	1.66	-367.88		0.33					-1.20					21.48
8 p.m.	0.85	0.85	37.19	-443.27	1.71	-476.52	252.02	0.17										15.09
9 p.m.	0.86	0.86	33.25	-467.69	1.78	-477.56	250.63	0.15										14.02
10 p.m.	0.87	0.87	33.91	-476.50	1.75	-457.24	258.87	0.12										13.62
11 p.m.	0.85	0.85	30.83	-442.12	1.51	-377.97	314.88					9.76						12.38

V. ANALYSIS 2: MODE COST AND CHOICE MODELING

The second analysis in this study is a detailed look at what is at stake when choosing between taxi and public transit as a transportation mode. Typically, mode choice models are based on comparing the utility associated with the generalized cost of travel by each mode available. Although taxi data is available for trips across all of New York City, it is useful to narrow the scope and focus on comparing data for specific sets of trips. Trips to and from airports are particularly important for non-driving modes, so these were selected as a focus for the mode choice analysis.

The objective of this part of the study was to develop a data-oriented method to compare the generalized cost for different non-driving modes for airport access and to understand whether transit or taxi yields a better utility at different times of the day. While the results of this study may be useful to individuals making travel choices, the method proposed in this study can also help policymakers understand the factors that affect mode choice in order to plan airport ground access.

GENERALIZED COST, UTILITY, MODE CHOICE

As web services and information technology become more advanced, it is easier for people to acquire complete information about travel by transit and taxi. Assuming that passengers make travel decisions based on monetary costs and travel time, the relative attractiveness of one mode over the other may change as transit schedules, fares, and taxi travel times vary for different times of day and days of the week. The relevant information can be obtained from Google Transit and a large set of taxi GPS data.

The total generalized cost for an individual trip in units of dollars can be computed for each mode i at time is denoted by TC_{ij} and calculated as follows:[50]

$$TC_{ij} = \alpha \times T_{ij} + \frac{F_{ij}}{n} \qquad (3)$$

where α is the passenger's value of time ($/hr), T_{ij} is the average travel time for the trip (hours), F_{ij} is the average fare paid for the trip ($), and n is the number of passengers sharing a taxi cab; for transit $n = 1$.The total generalized cost can also be expressed in units of hours by dividing the TC_{ij} by α.

The utility of each travel by each mode in hour j is based on the generalized cost:

$$U_{tran,j} = b - \beta \times TC_{tran,j} = b - \alpha\beta \times T_{tran,j} - \beta \times F_{tran,j} \qquad (4)$$

$$U_{taxi,j} = b - \beta \times TC_{taxi,j} = b - \alpha\beta \times T_{taxi,j} - \frac{\beta \times F_{taxi,j}}{n} \qquad (5)$$

where b is the benefit for each individual of completing a trip to or from the airport, and β is the equivalent utility of a dollar. For an airport trip, the benefit is assumed to be the same for both choices as long as the OD pair is fixed.

The choice between two modes, such as transit and taxi, is typically modeled with a binary logit model based on the difference of utilities between the choices.[51] The probability that an individual will choose transit over taxi is:

$$P_{trans,j} = \frac{e^{U_{trans,j}}}{e^{U_{trans,j}} + e^{U_{taxi,j}}} = \frac{e^{(-\beta \times TC_{tran,j})}}{e^{(-\beta \times TC_{tran,j})} + e^{(-\beta \times TC_{taxi,j})}} \qquad (6)$$

and the probability of choosing taxi over transit is:

$$P_{taxi,j} = 1 - P_{trans,j} \qquad (7)$$

The number of passengers using choosing mode i is the product of $P_{i,j}$ and the total travel demand. In the following sections, transit and taxi trips in NYC are compared based on their travel time, total cost and the corresponding choice probability.

COMPARING GENERALIZED COST OF TAXI AND TRANSIT

The two main types of public transportation services for airport access in NYC—transit (including train, AirTrain, subway, and bus) and taxi—were compared for trips between Penn Station and the three main airports in NYC: JFK, LGA, and EWR. These constitute the largest airport system in the United States. Penn Station was selected as the non-airport trip destination of interest for this study because it is a major hub of transit and taxi activity. Approximately 18 percent of taxi trips from Penn Station that leave the city are to EWR, and approximately one percent of all taxi trips to/from Penn Station are from/to LGA and JFK.

In order to make a consistent comparison between taxi and transit for each of the airport trips considered, the following assumptions about trip characteristics were made for the remaining parts of this analysis:

- All fares are calculated per passenger. On transit, this is the way that fares are always charged. For taxi, the number of passengers sharing the ride divides the total fare paid.

- Travel time for walking at Penn Station and at the airports is omitted, because the distance is similar whether passengers travel by taxi or transit, so it is unaffected by the mode choice.

- All transit passengers pay regular pre-paid fares. No discounts for senior citizens or weekly pass holders and no surcharges for on-board ticket purchase are included.

The data for airport trips by taxi and transit are summarized in Table 6 for each of the origin-destination pairs linking Penn Station and the region's airports. The taxi data is extracted from the entire 10-month dataset, and includes the number of passengers per trip, the total amount paid, the trip time, and the distance. The total fare is a flat rate between most of Manhattan and JFK or EWR airports, plus any tolls, tips, and surcharges. The fare between Manhattan and LGA has some variability because trips are charged the normal metered rate. The travel time has largest variability among all four variables, which indicates variability of traffic conditions. The trip distance is mostly stable for all airport trips, but the slight variability indicates that various alternative routes might be taken for the same OD pairs.

Table 6. The Taxi Trips Extracted from 10 Months of Taxi GPS Data and 1 Week of Transit Trips

		Taxi				Transit	
		Passenger No.	Total Amount ($)	Trip Time (min)	Trip Distance (mi)	Trip Time (min)	Fare ($)
OD pair	**No. of Obs**	Mean *(Std. Dev.)*	Mean *(Std. Dev.)*	Mean *(Std. Dev.)*	Mean *(Std. Dev.)*	Mean *(Std. Dev.)*	Mean *(Std. Dev.)*
Penn-JFK	5624	1.81	52.34	47.79	17.24	52.51	12.47
		(1.29)	*(5.26)*	*(17.57)*	*(1.89)*	*(10.34)*	*(1.69)*
JFK-Penn	2691	1.87	51.69	45.11	17.72	60.96	12.65
		(1.27)	*(4.33)*	*(12.74)*	*(1.79)*	*(10.03)*	*(1.37)*
Penn-LGA	9697	1.63	34.85	31.43	10.23	60.85	4.92
		(1.21)	*(5.64)*	*(10.30)*	*(1.38)*	*(6.32)*	*(3.57)*
LGA-Penn	3630	1.65	35.07	32.06	10.21	61.48	6.91
		(1.22)	*(5.58)*	*(9.35)*	*(1.71)*	*(7.57)*	*(3.23)*
Penn-EWR	1445	1.80	67.30	32.09	17.08	58.87	17.25
		(1.28)	*(10.48)*	*(9.49)*	*(2.05)*	*(19.90)*	*(9.54)*

The transit data is based on scheduled travel times and fares as opposed to actual realized travel times because the transit schedules are publicly available, and this is the basis for many travelers' decisions. The variation in travel times reflects the difference in expected waiting time, depending on what time the passenger starts their trip. For example, if trains depart for the airport every 30 minutes, the amount of waiting time that travelers experience is expected to be uniformly distributed between zero and 30 minutes. This is reflected in the expected travel times provided by Google Maps API Transit Directions Service. The variations in fare are due to the various combinations of modes that may make up the shortest travel time. A trip made entirely by subway and local bus will cost $2.50, but commuter trains charge substantially higher fares, and AirTrain also charges an additional fare ($5.00 at JFK and $5.50 at EWR).

A more detailed distribution of travel times by the day of the week is shown in Figure 7. Travel times for transit on weekdays are consistent, which could result from the fact that the transit schedule is similar for all weekdays, but it is necessary to analyze Saturday and Sunday separately. The average travel times for transit on weekends is higher because service headways are longer, and average travel times for taxis on weekends is lower because traffic is less congested. The day of the week affects the costs that travelers face on each mode.

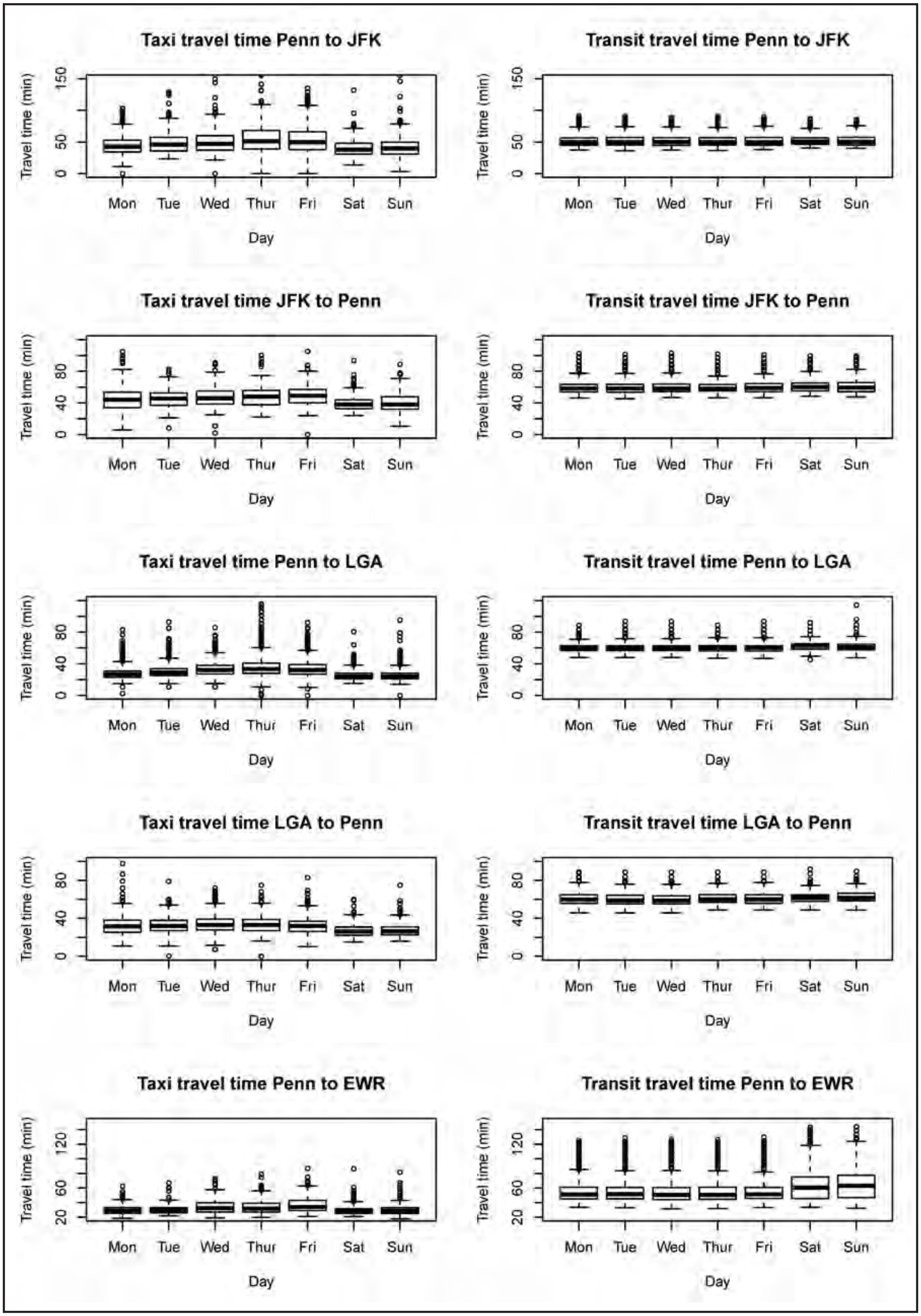

Figure 7. Boxplot of Transit Travel Time by Day of the Week

MODELING MODE CHOICE

In order to model mode choice based on observations of travel time and fare, these costs must be combined into an estimate of generalized cost and then converted to equivalent units of utility. This requires the calibration of parameters of the utility function based on the rate that passengers value their time. Normally, a mode choice model would be estimated based on extensive survey data of individual travelers' decisions, but, without a survey, an aggregated approach was adopted to fit model parameters so that the total number of transit riders was in agreement with observed ridership on the AirTrain.

Determination of Value of Time, $\boldsymbol{\alpha}$

The value of time for airport trips varies considerably from person to person, and it can be considered as a continuous random variable that is distributed across the user population.[52] The value of time for business trips can be higher than for leisure trips.[53] The distribution of values of time for airport trips is also likely to differ from that of other trip purposes, which makes estimation of this value difficult.

The UK Department for Transport suggests £47.95 (equivalent to $76) as the value of time for a taxi/minicab passenger and £39.65 ($63) for a rail passenger in 2010.[54] For trips to/from Penn Station, the income in Manhattan was used as a reference value.[55] The 5-year (2007-2011) American Community Survey estimate of per capita income in Manhattan is $61,290,[56] which is $29.5/hr if working a full-time job of 40 hours per week. Gupta et al.[57] considers a higher value of time for airport trips because travelers may be willing to pay more to avoid missing their flight. The authors suggested $42/hr for leisure trips and $63/hr for business trips. Since we do not know the value of time for passengers that made airport trips in NYC, a preliminary value of time $40/hr was used to represent everyone and anytime based on above references. In the sensitivity analysis that follows, wide ranging of values of time were considered.

Calibration of $\boldsymbol{\beta}$ Coefficient

Equation 4 and Equation 5 describe the relationship between cost and utility, and the β coefficient plays an important role in determining the outcome of the binary logit model. The β values are the marginal utility of total cost. In order to estimate mode choice, it is necessary to determine β, which was done by comparing the total number of transit and taxi trips. JFK AirTrain ridership information and taxi GPS data were used to estimate a single β. One value of β was used for all three airport trips because it is likely that the average marginal utility of total cost is similar for passengers using each of the airports. Furthermore, data are not available to estimate specific β values for LGA and EWR, because transit ridership between Penn Station and these airports is not available.

Paid ridership of the JFK AirTrain was 5.3 million passengers in 2010,[58] which accounts for nearly all of the transit trips to and from JFK. In the same time period, there were 3.386 million taxi trips to and from JFK, extrapolated from the complete 10-month records of taxi GPS data. Based on the trips counts above, 39% of non-driving trips were made by taxi and 61% we made by transit to get to and from JFK.

Without more detailed transit ridership data, the overall mode share for all trips to and from JFK during 2010 was considered to be the same as mode share for trips between Penn Station and JFK. The logit model was calibrated by selecting the β value that makes the model estimates over the course of the day match this observed mode share. Figure 8 shows the relationship between β and the aggregated probability of taking a taxi, based on the total generalized cost in each hour j between Penn Station and JFK (including both directions: Penn-JFK and JFK-Penn), assuming $n = 1$ passenger per taxi and passengers value their time at $\alpha = \$40$/hour. At each hour, the relationship between number of taxi trips ($n_{taxi,j}$) and the estimated number of transit users ($\hat{n}_{transit,j}$) is:

$$\frac{n_{taxi,j}}{P_{taxi,j}} - n_{taxi,j} = \hat{n}_{transit,j} \qquad (8)$$

$$P_{taxi,j} = \qquad (9)$$

when β, the expected probability of people choosing taxi to the airport is 0.39, which matches the data from 2010. This value of β was applied to the cost data for all the airport trips in order to estimate the mode share by taxi and public transit.

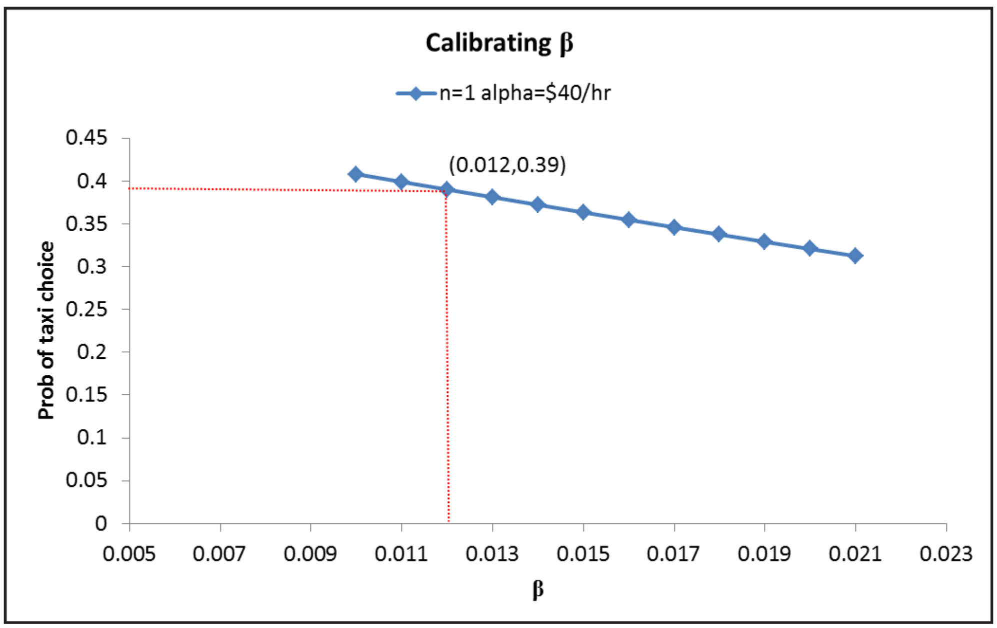

Figure 8. Calibrating β Using Taxi and Transit Data

Mode Choice by Time of Day

On weekdays, the average travel time for taxis is less than that of transit at all times. When considering both the time and money spent on the trip, the total cost indicates that even if the passenger is traveling alone, taxi has a cost advantage only in middle of the night (12 a.m. to 6 a.m.). The taxi travel times vary significantly, and the longest travel times are usually observed during morning peak (6 a.m. to 10 a.m.) and afternoon peak (2 p.m. to 6 p.m.) as shown in Figure 9.

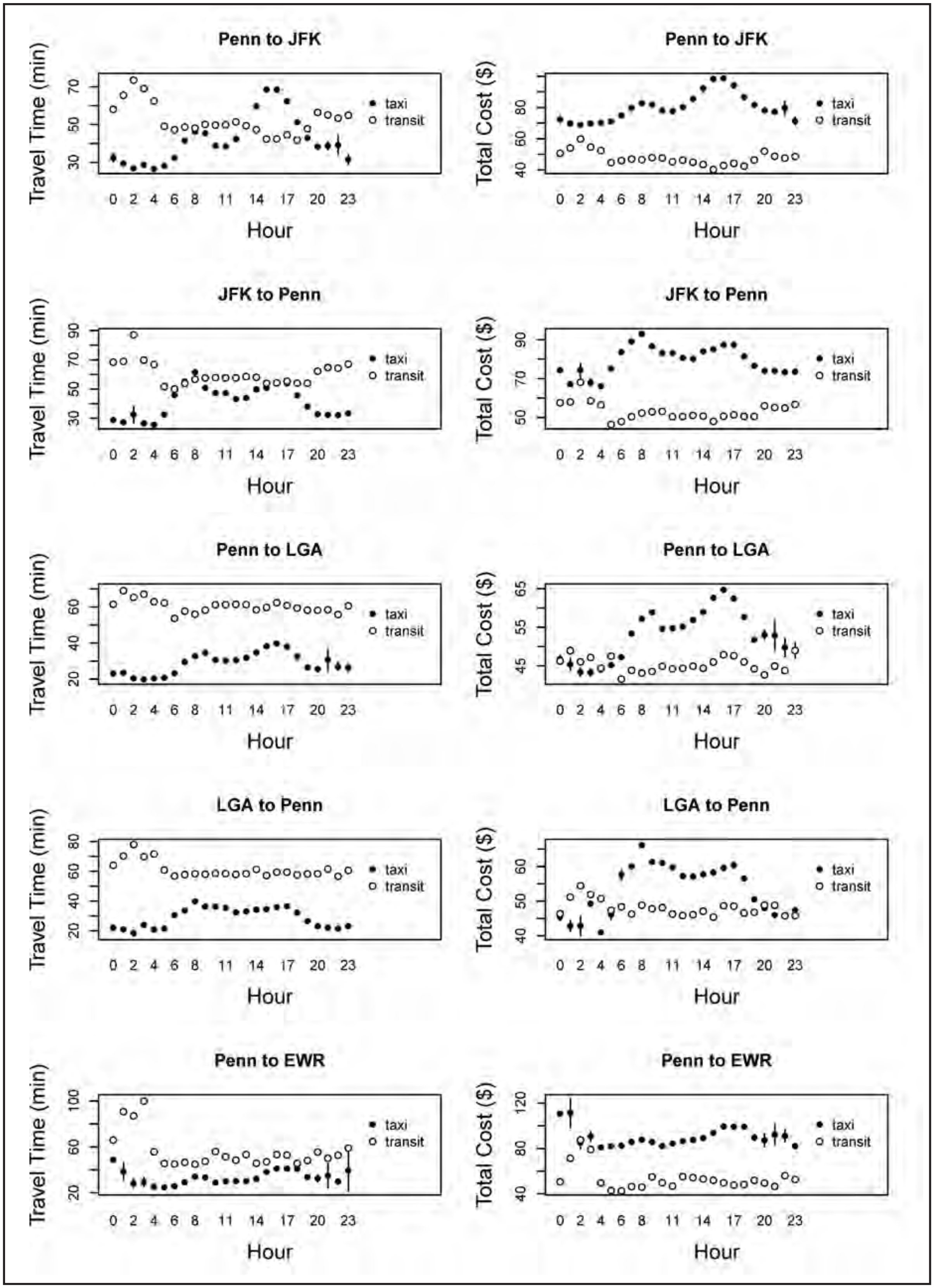

Figure 9. Comparison of Travel Time and Total Generalized Cost per Trip for Taxi and Transit by Weekday Hour (± standard error)

In order to consider the variability of travel costs, the standard error (SE) was calculated for the total cost of trips (Figure 9). There tend to be fewer observations in the taxi data at midnight, which results in relatively higher SE and a wider 95% confidence interval for the mean values at each hour (approximately equal to mean ± 1.96×SE). On the other hand, the transit data show relatively small variance within each of the 24 hours. The main difference in transit travel times arises from the waiting times and transfer times for the next available train or bus. The transit travel time and cost are less variable than taxi travel time and cost, which depend on the traffic condition at different times of day and from day to day.[59]

This analysis is limited to trips between Penn Station and the three airports. The probability of choosing taxi at each time of day was calculated using the binary logit model, and the results are plotted in Figure 10. Based on the difference between total cost for taxi and transit, the mode share can be expected to change for different times of the day. For example, transit tends to be more competitive during rush hours when traffic congestion makes taxi trips slower. On the other hand, taxis are more competitive in late night hours when transit headways are long.

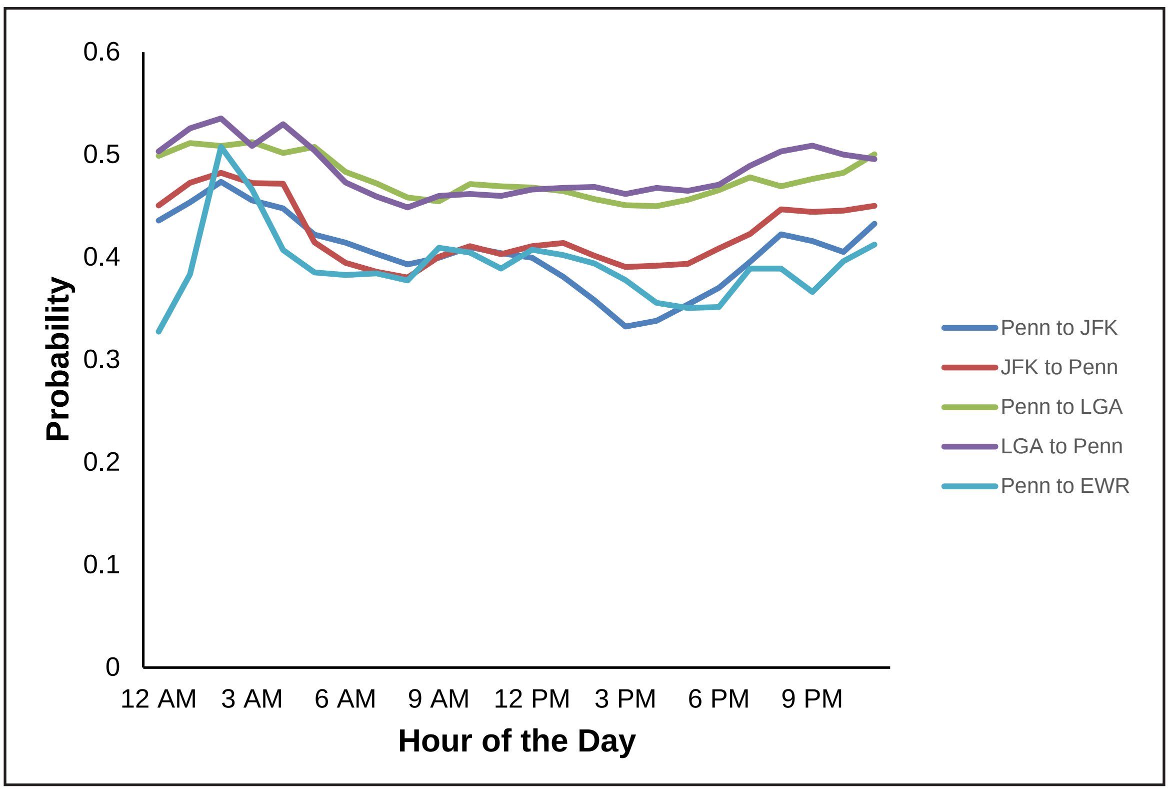

Figure 10. Probability of Choosing Taxi for Each Airport Trip by Time of Day

Since these OD pairs are just a partial set of all trips that go to and from the airports, the analysis reflects only the costs at those locations. Different locations may have a totally different trend based on the travel time and cost. Time and money are not the only factors people consider when making travel choices, but the literature suggests that these are the most important. It is possible that some people use transit for all without even considering taxi, or others take taxis to the airport without ever considering using transit.

Some factors that likely influence mode choice other than travel time and fare are that taxis provide a more personalized door-to-door service with additional benefits, such as assistance with luggage. Some of this value is captured in the tips that are included in the taxi data and the total fare paid, which includes tip. In reality, people may experience an additional penalty for using transit because they need to walk a certain distance to get transit service. These additional benefits or penalties are omitted from the analysis in order to focus on the effects of money cost and travel time on the competitiveness of each mode.

Sensitivity Analysis

The total cost is also influenced by the value of time and the number of passengers traveling together. Table 6 shows that on average there are 1.6 to 1.8 passengers taking taxis together to go to or from each airport. A sensitivity analysis was performed to investigate the effects of both the value of time and the number of passengers in the group on the probability of an individual's travel mode choice in detail.

The average travel time and fare from all records for taxi and transit are considered travel time and fare for each OD pair. This sensitivity analysis considers variation of the value of time, α (ranging from $10/hr to $70/hr), and passenger count, n (in the range 1 to 3), to see how much influence these factors have on total cost as shown in Figure 11a-e. If the value of time is fixed, changing the number of passengers affects only the taxi fare per person because the transit fare per person is fixed. The slope in Figure 11 for each mode is the travel time, and the intercept is the fare per person according to Equation 3. Intersections of taxi cost and transit cost are found for all OD pairs except trips from Penn to JFK (Figure 11a). The intersection indicates a value of time when the cost of taxi and transit are the same. This value of time at the intersection is a tipping point above which passengers should be willing to pay extra fare for the faster mode. For example, the transit cost for JFK-Penn intersects with taxi cost at $57/hr for $n = 2$, indicating that the total cost of taxi is higher when value of time is less than $57/hr, because the slope for transit exceeds the slope for taxi (Figure 11b). This means that if two passengers are traveling as a group, it is better to choose transit if the value of time is lower than $57/hr, otherwise it is more cost-effective to share a ride in a taxi.

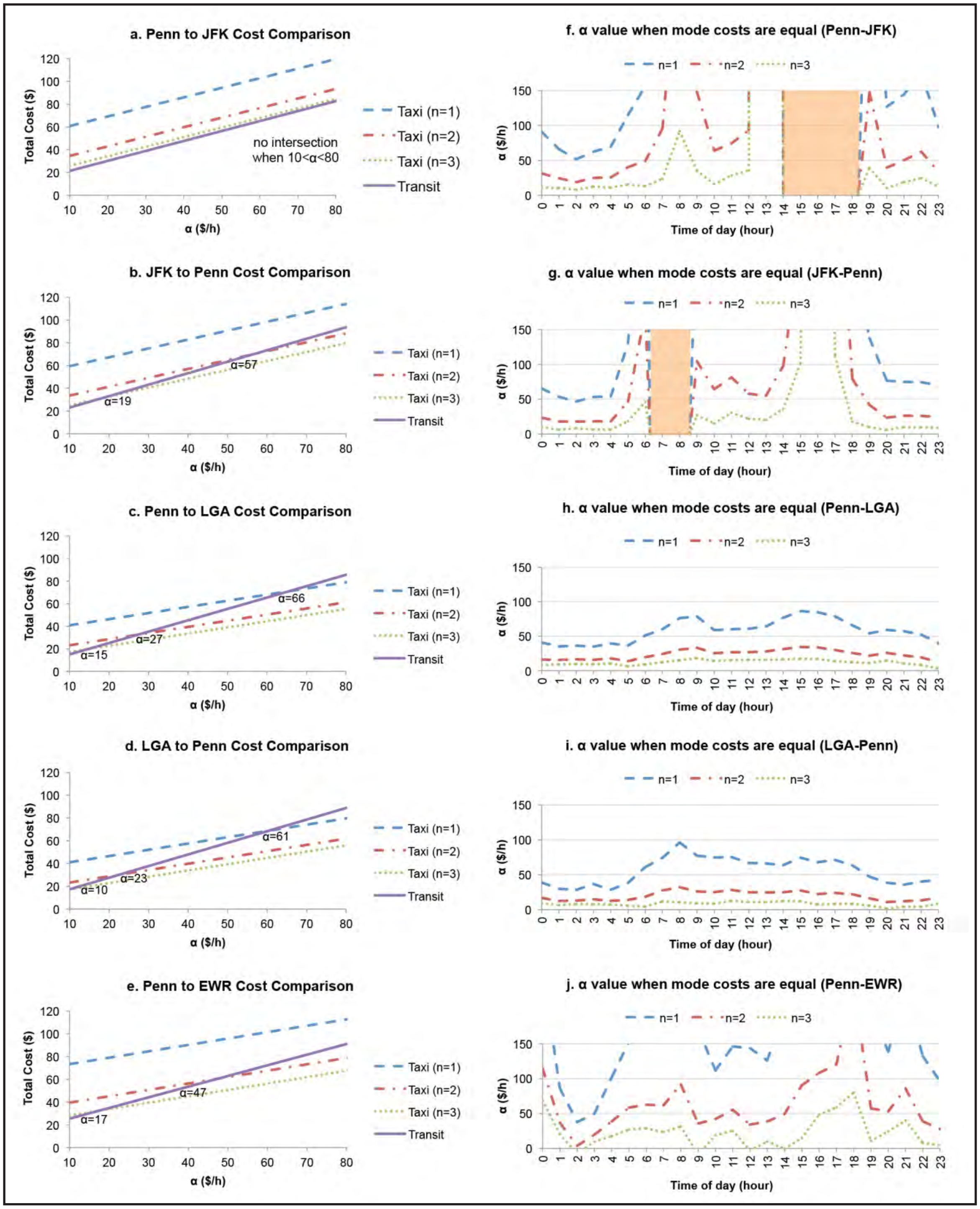

Figure 11. Sensitivity Analysis on the Value of Time (α) and Number of Passengers (n)

On average, there is no way that a trip from Penn Station to JFK will be less costly by taxi in the assumed range of values of time and number of passengers (Figure 11a). For the reverse direction, JFK-Penn, taxis do become competitive for sufficiently high values of time and passenger occupancies (Figure 11b).

For trips to and from LGA (Figure 11c-d), if traveling alone, the taxi costs are higher than transit costs when the value of time is less than $66/hr (Penn-LGA) or $61/hr (LGA-Penn);

however, if traveling with more than two people, the threshold is $27/hr. This relatively low value is reasonable because there is no direct transit service between Penn Station and LGA. Since the distance is shorter than any of the other airports, taxis are more competitive.

The trip cost from Penn to EWR seems similar to JFK-Penn, except that the intersection points differ slightly. Transit costs more if the value of time exceeds $47/hr ($n = 2$) or $17/hr ($n = 3$). Considering $63/hr as the value of time for business trips in NYC and $42/h as the value of time for leisure trips,[60] it is likely that a person traveling on business will use a taxi for JFK-Penn or Penn-EWR trips if traveling with more than two people, but a person traveling for leisure will use a taxi only if traveling with at least three people.

In order to account for the effect on the variation of travel time throughout a day, the threshold value of time within each hour at which passengers will switch their preferred mode was plotted (Figure 11f-j). For most cases, travel times are longer by transit than by taxi (i.e., the slope for transit exceeds the slope for taxi), so values of time greater than the threshold are associated with more cost-effective taxi service, and values of time less than the threshold are associated with more cost-effective transit service. For a couple of time periods, transit is actually faster than taxi because traffic congestion has such a severe effect on taxi travel times, and the interpretation switches, so in the shaded areas of Figure 11f-g, all trips served more cost-effectively by transit, regardless of the value of time. During these times, transit is faster and cheaper than taxi.

The relatively low tipping point values for LGA compared to EWR and JFK show that taxi is more competitive than transit for that airport, appealing to a wider range of values of time. There is also a pattern at all airports that taxi is more competitive in the early hours of the morning (around 2 a.m.) when transit service is also less frequent. These results have policy implications, because they show how airports differ in the competitiveness of ground access modes, and how this changes by time of day.

Results for EWR (Figure 11j) suggest that transit is more competitive from Penn Station to EWR, but for midnight trips taxis have a lower total cost than transit since the frequency of service is lower, which results in longer waiting times. However, because of the relatively long distance between Penn and EWR, it is possible that taxi is more likely to be chosen based on the factors like convenience and comfort, which are not considered in this study.

DISCUSSION

This part of the analysis presents a methodology to compare the total cost for two modes of transportation (transit and taxi) using taxi GPS data and high-resolution transit schedule information. Trips between NYC Penn Station and three New York area airports (JFK, LGA and EWR) at different times of day were used to illustrate the methods. As shown in the analysis of total cost and mode choice, transit is more cost-effective than taxi for most times of the day if passengers are traveling alone and value time at $40/hour, except during some midnight periods when transit service has long headways that contributes a significant amount of time to waiting or transfers.

The sensitivity analysis suggests that people are more likely to choose taxi to travel from Penn Station for airport trips if: 1) they place a high value on time, 2) they are traveling with a large group of people, or 3) they are traveling during late-night hours. It is also found that if people are traveling for business trips, taxis become a less costly choice for airport access. Due to the long distances between Penn Station and JFK or EWR, taxi fares for those trips are very high, making transit a more competitive mode most of the time (especially when $n = 1$), even though taxis offer an advantage in travel time. LGA airport, however, is closer to Penn Station, and the relatively low taxi fare and low travel time make taxi a more competitive choice for that OD pair.

The results show that the total cost or travel time for taxis always has a morning peak (between 6 a.m. to 10 a.m.) and an afternoon peak (between 1 p.m. to 6 p.m.). The taxi data provides an indication of traffic conditions in NYC,[61] so the use of this data to calculate the travel cost could incorporate both the temporal and spatial effects of traffic congestion in the city. However, this study is limited to the temporal analysis of the five most popular OD pairs for the airports, which all include trips to and from Penn Station. This can create bias if used to estimate total costs for the entire city. Future applications could be expanded to consider the spatial dimension as well by include multiple OD pairs distributed all over the city.

For trips to and from the airport, mode choice is also affected by other factors—such as convenience, comfort, and safety—that were not considered in this study because they cannot be easily measured and quantified. With additional data on the number of passengers using each mode by time of day, it may be possible to gain some insights into the effect of these less tangible factors by comparing the expected mode shares from the utility functions in this analysis with the observed mode shares.

The presented analysis could be used as an example of a practical method to estimate the travel cost including both time and money. As information and resources like travel time and fare are increasingly accessible, it should be possible to design a smartphone app or a small computer program for the transit ticket vending machine to estimate total cost using this methodology. This information along with the choice model can be used to understand the factors that affect the aggregate mode choice decisions of the public. This will be useful for transportation for planners and policy makers to improve the quality of travel options available to people traveling to and from airports.

VI. CONCLUSION

The study presented in this report makes use of big data from GPS records of taxi trips and from comprehensive transit schedule information in GTFS format. By connecting this data with additional sources of information about neighborhood demographics, socioeconomic characteristics, and employment, models were developed to explain the spatial and temporal variation of travel demand for taxis. A specific application was illustrated for NYC. The first part of the analysis resulted in the development of trip generation models that show how various factors (e.g., transit accessibility, demographics, and employment) affect the rate of taxi trip pickups and drop-offs in NYC. The second part of the analysis focused on trips between Penn Station and the three major airports in the NYC region to show how mode choice is affected by the size of the traveling group, the travelers' valuation of time, and the time of day.

In order to analyze taxi demand by using a set of GPS data for every trip made within a 10-month period, the data had to be processed to eliminate errors. Then the observations were aggregated by time of day and census tract to allow for systematic analysis and comparison explanatory variables that are aggregated at the same resolution. The result was a hybrid cross-classification and regression model that classifies trips by time of day, and then, within each hour, fits a regression model to explain the number of taxi trips generated. The model reveals that there are six factors that have the greatest explanatory power for determining the number of taxi pickups and drop-offs:

- Transit access time (TAT)
- Population size
- Median age
- Percent of population educated beyond bachelor's degree
- Income per capita
- Number of employment opportunities

Furthermore, the model shows that there are specific industrial sectors in which the number of employment opportunities within a census tract has the greatest effect on the rate of taxi trips. The number of employees working in retail, accommodation and food service, and health care are all significant determinants of the amount of taxi demand in a census tract. The magnitude of employment in these sectors provides an indication of the amount of related activity occurring, so people are more likely to use taxis to travel to and from places that have lots of stores, hotels, restaurants, and medical facilities. Although the correlation does not prove trip purpose, this observation supports the notion that these people are more likely to use taxis to travel to and from the listed activities.

To account for the effect of transit service on taxi demand, a method was developed to quantify the transit accessibility at any place and time in the form of a TAT. The TAT is a measure of the shortest time that it takes to board a transit vehicle, accounting for walking time and waiting from a specific location. This value was calculated using the complete schedule of subway services for the city, which is publicly available online in GTFS format. For the study of taxi demand in NYC, the TAT was evaluated for the subway system across each of the 2,167 census tracts in the city and for each hour of the day. The result can be visualized as a series of maps of transit accessibility, and the quantified value of transit accessibility was used as an explanatory variable in the taxi trip generation models that were developed.

In the second part of the study, the total cost of travel by taxi and by public transit was compared systematically across each hour of the day for five origin-destination pairs (between Penn Station and each of the New York area airports, not including the return trip from Newark to NYC). The relative competitiveness of transit and taxis depends on the number of passengers traveling in a group, the travelers' value of time, and the time of day. Traffic congestion changes the travel time by taxi at different times of day, and changes in schedule frequency affect the amount of time that passengers can expect to wait for transit at different times of day. At a value of time of $40/hour, transit is almost always more competitive for passengers traveling alone for all airport trips, except for a few hours in the middle of the night transit when headways are quite long.

A sensitivity analysis of the mode choice results was conducted to address the fact that some data is lacking from the existing dataset that would normally be used to calibrate a mode choice model. Many times there is a trade-off between faster, more expensive modes and slower, less expensive mode. A useful outcome of the analysis is the identification of tipping point values of time above which one mode becomes more competitive than the other for a specific origin-destination pair and time of day. For example, business travelers with higher values of time are more likely to choose taxi in many situations. An exception exists for trips to JFK in the afternoon peak and from JFK in the morning peak when traffic congestion makes taxi trips so slow that transit is both faster and cheaper.

This study provides useful insights in the ways that large data sources can be processed and integrated to improve our understanding of the way people use the transportation system. Some types of data still require individual surveys in order to estimate effectively. For example, estimating passengers' value of time should include assessing the number of passengers who travel between Penn Station and each of the airports by public transit; however much of this data is unavailable in the large-scale, aggregated data sources. Subway and bus users swipe a fare card only when entering the system, so their movements are not tracked. However, complete observations of the number of trip origins and destinations over several months does provide an opportunity to build models to understand taxi demand that are not susceptible to errors from sampling or undercounting.

Continuing work in this area includes improving the model specification to better represent trip counts and acknowledge the problems associated with correlations between adjacent census tracts. The nature of the taxi dataset, which not only tracks individual trips but also links them to shifts operated by individual drivers, provides the potential to look at the spatial and temporal distribution of taxi supply as well. These are refinements that can improve modeling capabilities.

Ultimately, the usefulness of these models of trip generation and mode choice is that they provide planners, engineers, and decision makers with information about how people use the transportation system. In this case, by identifying the factors that drive taxi demand, forecasts can be made about how this demand can be expected to grow and change as neighborhoods evolve. As decisions are made regarding the regulation of the taxi industry, the provision of transit service, and urban development, these models are useful for forming a complete and holistic vision of how travel patterns and use of modes can be expected to respond.

ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
EWR	Newark Liberty International Airport
GIS	Geographic Information System
GPS	Global Positioning System
GTFS	Google Transit Feed Specification
JFK	John F. Kennedy International Airport
LGA	Laguardia Airport
LIRR	Long Island Railroad
LOS	Level of Service
MTA	Metropolitan Transit Authority
NYC	New York City
OD	Origin-Destination (In the context of origin to destination pair)
RSS	Residual Sum of Squares
SE	Standard Error
TAT	Transit Access Time
TAZ	Transportation Analysis Zone
TLC	Taxi and Limousine Commission
VIF	Variance Inflation Factor
WAC	Workplace Area Characteristic

ENDNOTES

1. Schaller Consulting, "The New York City Taxicab Fact Book," (March 2006) www.schallerconsult.com/taxi/taxifb.pdf (accessed November 1, 2013).

2. Yeqian Lin, Wenquan Li, Feng Qiu, and He Xu, "Research on Optimization of Vehicle Routing Problem for Ride-sharing Taxi," *Procedia—Social and Behavioral Science* 43(2012): 494–502.

3. Drew Austin and P. Christopher Zegras, "Taxicabs as Public Transportation in Boston, Massachusetts," *Transportation Research Record* 2277 (2012): 65–74.

4. Paul Stephen Dempsey, "Taxi Industry Regulation, Deregulation, and Reregulation: The Paradox of Market Failure," *Transportation Law Journal* 24 (1996): 73–120; Mark W. Frankena and Paul A. Pautler, *An Economic Analysis of Taxicab Regulation*, Bureau of Economics Staff Report (Washington, D.C.: Federal Trade Commission, 1984); Bruce Schaller, "Entry Controls in Taxi Regulation: Implications of US and Canadian Experience for Taxi Regulation and Deregulation," *Transport Policy* 14 (2007): 490–506; Josep Maria Salanova, Miquel Estrada, Georgia Aifadopoulou, and Evangelos Mitsakis, "A Review of the Modeling of Taxi Services," *Procedia—Social and Behavioral Sciences* 20 (2011): 150–161.

5. Ralph Turvey, "Some Economic Features of the London Cab Trade," *The Economic Journal* 71 (1961): 79–92.

6. Chanoch Schreiber, "The Economic Reasons for Price and Entry Regulation of Taxicabs, *Journal of Transport Economics and Policy* 9 (1975): 268–279.

7. Richard B. Coffman and Chanoch Schreiber, "The Economic Reasons for Price and Entry Regulation of Taxicabs (Comment and Rejoinder)," *Journal of Transport Economics and Policy* 11 (1977): 288–304.

8. Michael E. Beesley, "Regulation of Taxis," *The Economic Journal* 83(1973): 150–172; Clive Gaunt and Terry Black, "The Economic Cost of Taxicab Regulation: The Case of Brisbane," *Economic Analysis and Policy* 26 (1996): 45–58.

9. Bruce Schaller, "A Regression Model of the Number of Taxicabs in U.S. Cities," *Journal of Public Transportation* 8 (2005): 63–78.

10. Paul Stephen Dempsey, "Taxi Industry Regulation, Deregulation, and Reregulation: The Paradox of Market Failure," *Transportation Law Journal* 24 (1996): 73–120; Jeremy P. Toner, "The Welfare Effects of Taxicab Regulation in English Towns," *Economic Analysis and Policy* 40 (2010): 299–312; Maya Bacache-Beauvallet and Lionel Janin, "Taxicab License Value and Market Regulation," *Transport Policy* 19 (2012): 57–62.

11. Fabien Girardin and Josep Blat, "The Co-evolution of Taxi Drivers and their In-car Navigation Systems," *Pervasive and Mobile Computing* 6 (2010): 424–434; Ruibin Bai, Jiawei Li, Jason A. D. Atkin, and Graham Kendall, "A Novel Approach to Independent

Taxi Scheduling Problem Based on Stable Matching," *Journal of the Operational Research Society* (2013) doi:10.1057/jors.2013.96.

12. Ryan Herring, Aude Hofleitner, Pieter Abbeel, and Alexandre Bayen, "Estimating Arterial Traffic Conditions Using Sparse Probe Data," (paper presented at the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), Funchal, Madeira, September 19–22, 2010).

13. Drew Austin and P. Christopher Zegras, "Taxicabs as Public Transportation in Boston, Massachusetts," *Transportation Research Record* 2277 (2012): 65–74.

14. Wende A. O'Neill and Eugene Brown, "Long-Distance Trip Generation Modeling Using ATS," *Transportation Research E-circular Number E-C026* (Washington, D.C.: Transportation Research Board, 2001); Johnnie Ben-Edigbe and Rosnita Rahman, "Multivariate School Travel Demand Regression Based on Trip Attraction" (paper presented at ICURPT 2010: International Conference on Urban, Regional Planning and Transportation, Paris, France, June 28–30, 2010); David P. Racca and Edward C. Ratledge, "Project Report for 'Factors That Affect and/or Can Alter Mode Choice,'" Project Report (Newark, DE: Center for Applied Demography & Survey Research, University of Delaware, 2004); Ajay Kumar and David Levinson, "Estimating, and Validating a New Trip Generation Model: A Case Study of Montgomery County, Maryland," *Transportation Research Record* 1413 (1992): 107–113; Bruce Schaller, "A Regression Model of the Number of Taxicabs in U.S. Cities," *Journal of Public Transportation* 8 (2005): 63–78.

15. *Trip Generation Manual*, 9th Ed. (Washington, D.C.: Institute of Transportation Engineers, 2012).

16. Wende A. O'Neill and Eugene Brown, "Long-Distance Trip Generation Modeling Using ATS," *Transportation Research E-circular Number E-C026* (Washington, D.C.: Transportation Research Board, 2001); Ajay Kumar and David Levinson, "Estimating, and Validating a New Trip Generation Model: A Case Study of Montgomery County, Maryland," *Transportation Research Record* 1413 (1992): 107–113.

17. Johnnie Ben-Edigbe and Rosnita Rahman, "Multivariate School Travel Demand Regression Based on Trip Attraction" (paper presented at ICURPT 2010: International Conference on Urban, Regional Planning and Transportation, Paris, France, June 28–30, 2010); Amir Mousavi, Jonathan Bunker, and Brian Lee, "A New Approach for Trip Generation Estimation for Traffic Impact Assessments" (paper presented at the 25th ARRB Conference, Shaping the future: linking policy research and outcomes, Perth, Australia, September 23–26, 2012).

18. Wende A. O'Neill and Eugene Brown, "Long-Distance Trip Generation Modeling Using ATS," *Transportation Research E-circular Number E-C026* (Washington, D.C.: Transportation Research Board, 2001); Johnnie Ben-Edigbe and Rosnita Rahman, "Multivariate School Travel Demand Regression Based on Trip Attraction" (paper presented at ICURPT 2010: International Conference on Urban, Regional Planning and Transportation, Paris, France, June 28–30, 2010);

19. Wende A. O'Neill and Eugene Brown, "Long-Distance Trip Generation Modeling Using ATS," *Transportation Research E-circular Number E-C026* (Washington, D.C.: Transportation Research Board, 2001); Ajay Kumar and David Levinson, "Estimating, and Validating a New Trip Generation Model: A Case Study of Montgomery County, Maryland," *Transportation Research Record* 1413 (1992): 107–113; Grace Corpuz, "Public Transport or Private Vehicle: Factors that Impact on Mode Choice" (paper presentat at the 30th Australasian Transportation Research Forum, Melbourne, Australia, 2007); Yu-Chun Chang, "Factors Affecting Airport Access Mode Choice for Elderly Air Passengers," *Transportation Research Part E* 57 (2013): 105–112.

20. Tim Schwanen and Patricia L. Mokhtarian, "What Affects Commute Mode Choice: Neighborhood Physical Structure or Preferences toward Neighborhoods?" *Journal of Transport Geography* 13 (2005): 83–99.

21. Johnnie Ben-Edigbe and Rosnita Rahman, "Multivariate School Travel Demand Regression Based on Trip Attraction" (paper presented at ICURPT 2010: International Conference on Urban, Regional Planning and Transportation, Paris, France, June 28–30, 2010).

22. Yu-Chun Chang, "Factors Affecting Airport Access Mode Choice for Elderly Air Passengers," *Transportation Research Part E* 57 (2013): 105–112.

23. Johnnie Ben-Edigbe and Rosnita Rahman, "Multivariate School Travel Demand Regression Based on Trip Attraction" (paper presented at ICURPT 2010: International Conference on Urban, Regional Planning and Transportation, Paris, France, June 28–30, 2010); Reid Ewing, William Schroeer, and William Greene, "School Location and Student Travel," *Transportation Research Record* 1895 (2004) 55–63.

24. Yu-Chun Chang, "Factors Affecting Airport Access Mode Choice for Elderly Air Passengers," *Transportation Research Part E* 57 (2013): 105–112; Jan-Dirk Schmöcker, Mohammed A. Quddus, Robert B. Noland, and Michael G. H. Bell, "Estimating Trip Generation of Elderly and Disabled People," *Transportation Research Record* 1924 (2005) 9–18.

25. Mintesnot Gebeyehu and Shin-ei Takano, "Diagnostic Evaluation of Public Transportation Mode Choice in Addis Ababa," *Journal of Public Transportation* 10 (2007): 27–50.

26. Kenneth E. Train, *Discrete Choice Methods with Simulation*, 2nd Ed., Cambridge: Cambridge University Press, 2009.

27. Hojong Baik, Antonio A. Trani, Nicolas Hinze, Howard Swingle, Senanu Ashiabor, and Anand Seshadri, "Forecasting Model for Air Taxi, Commercial Airline, and Automobile Demand in the United States," *Transportation Research Record* 2052 (2008): 9–20.

28. Mintesnot Gebeyehu and Shin-ei Takano, "Modeling the Relationship Between Travelers' Level of Satisfaction and Their Mode Choice Using Ordinal Models," *Journal of the Transportation Research Forum* 47 (2008): 103–118.

29. Benedikt Mandel, Marc Gaudry, and Werner Rothengatter, “A Disaggregate Box-Cox Logit Mode Choice Model of Intercity Passenger Travel in Germany and its Implications for High Speed Rail Demand Forecasts,” *The Annals of Regional Science* 31 (1997): 99–120.

30. David P. Racca and Edward C. Ratledge, “Project Report for ‘Factors That Affect and/or Can Alter Mode Choice,’” Project Report (Newark, DE: Center for Applied Demography & Survey Research, University of Delaware, 2004).

31. Grace Corpuz, “Public Transport or Private Vehicle: Factors that Impact on Mode Choice” (paper presentat at the 30th Australasian Transportation Research Forum, Melbourne, Australia, 2007).

32. Greig Harvey, “Study of Airport Access Mode Choice,” *Journal of Transportation Engineering* 112 (1986): 525–545.

33. Voula Psaraki and Costas Abacoumkin, “Access Mode Choice for Relocated Airports: the New Athens International Airport,” *Journal of Air Transport Management* 8 (2002): 89–98.

34. Eric Pels, Peter Nijkamp, and Piet Rietveld, “Access to and Competition Between Airports: a Case Study for the San Francisco Bay Area,” *Transportation Research Part A* 37 (2003): 71–83.

35. Surabhi Gupta, Peter Vovsha, and Robert Donnelly, “Air Passenger Preference for Choice of Airport and Ground Access Mode in the New York Metropolitan Region,” *Transportation Research Record* 2042 (2008): 3–11.

36. Mei Ling Tam, William H. K. Lam, and Hing Po Lo, “The Impact of Travel Time Reliability and Perceived Service Quality on Airport Ground Access Mode Choice,” *Journal of Choice Modelling* 4 (2011): 49–69.

37. Brittany L. Luken and Laurie A. Garrow, “Multiairport Choice Models for the New York Metropolitan Area: Application Based on Ticketing Data,” *Transportation Research Record* 2206 (2011): 24–31.

38. Schaller Consulting, “The New York City Taxicab Fact Book,” (March 2006) www.schallerconsult.com/taxi/taxifb.pdf (accessed November 1, 2013).

39. R. Zito, G. D’Este, and M. A. P. Taylor, “Global Positioning Systems in the Time Domain: How Useful a Tool for Intelligent Vehicle-Highway Systems?” *Transportation Research Part C* 3 (1995): 193–209.

40. M. Anil Yazici, Camille Kamga, and Kyriacos C. Mouskos, “Analysis of Travel Time Variability in New York City Based on Day-of-Week and Time-of-Day Periods,” *Transportation Research Record* 2308 (2012): 83–95.

41. Raymond C. Browning, Emily A. Baker, Jessica A. Herron, and Rodger Kram, "Effects of Obesity and Sex on the Energetic Cost and Preferred Speed of Walking," *Journal of Applied Physiology* 100 (2005): 390–398.

42. Johnnie Ben-Edigbe and Rosnita Rahman, "Multivariate School Travel Demand Regression Based on Trip Attraction" (paper presented at ICURPT 2010: International Conference on Urban, Regional Planning and Transportation, Paris, France, June 28–30, 2010); Amir Mousavi, Jonathan Bunker, and Brian Lee, "A New Approach for Trip Generation Estimation for Traffic Impact Assessments" (paper presented at the 25th ARRB Conference, Shaping the future: linking policy research and outcomes, Perth, Australia, September 23–26, 2012).

43. Nicolaus Loos, *Value Creation in Leverages Buyouts: Analysis of Factors Driving Private Equity Investment Performance* (Wiesbaden, Germany: Deutscher Uniersitäts-Verlag, 2006); Xiao Chen, Philip B. Ender, Michael Mitchell, and Christine Wells, "Regression Diagnostics," in *Regression with Stata* (UCLA Institute for Digital Research and Education), http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter2/statareg2.htm (accessed November 1, 2013).

44. Xin Yan and Xiao Gang Su, *Linear Regression Analysis: Theory and Computing* (Singapore: World Scientific Publishing Co., Pte., Ltd., 2009); John Fox, *Applied Regression and Generalized Linear Models*, 2nd Ed. (Thousand Oaks, CA: SAGE Publications, Inc., 2008).

45. Anthony Davies, "Exhaustive Regression: An Exploration of Regression-Based Data Mining Techniques Using Super Computation," RPF Working Paper No. 2008-008 (Washington, D.C.: Center of Economic Research, The George Washington University, 2008); A. I. McLeod and C. Xu, "bestglm: Best Subset GLM," http://cran.r-project.org/web/packages/bestglm/vignettes/bestglm.pdf (accessed July 26, 2013); Graham Elliott, Antonio Gargano, and Allan Timmermann, "Complete Subset Regressions," *Journal of Econometrics* 177 (2013): 357–373.

46. Xin Yan and Xiao Gang Su, *Linear Regression Analysis: Theory and Computing* (Singapore: World Scientific Publishing Co., Pte., Ltd., 2009); John Fox, *Applied Regression and Generalized Linear Models*, 2nd Ed. (Thousand Oaks, CA: SAGE Publications, Inc., 2008).

47. C. L. Mallows, "Some Comments on C_p," *Technometrics* 15 (1973): 661–675.

48. Richard G. Funderburg, Hilary Nixon, Marlon G. Boarnet, and Gavin Ferguson, "New Highways and Land Use Change: Results from a Quasi-Experimental Research Design," *Transportation Research Part A* 44 (2010): 76–98.

49. Daniel G. Chatman, "Does TOD Need the T?" *Journal of the American Planning Association* 79 (2013): 17–31.

50. Guohui Zhang, Yinhai Wang, Heng Wei, and Ping Yi, "A Feedback-Based Dynamic Tolling Algorithm for High-Occupancy Toll Lane Operations," *Transportation Research*

Record 2065 (2008): 54–63; Ender Faruk Morgul and Kaan Ozbay, "Simulation-Based Evaluation of a Feedback-Based Dynamic Congestion Pricing Strategy on Alternate Facilities," No. 11-3535 (paper presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C., January 23–27, 2011).

51. Kenneth E. Train, *Discrete Choice Methods with Simulation*, 2nd Ed., Cambridge: Cambridge University Press, 2009.

52. Lan Jiang and Hani S. Mahmassani, "Toll Pricing: Computational Tests for Capturing Heterogeneity of User Preferences," *Transportation Research Record* 2343 (2013): 105–115.

53. Surabhi Gupta, Peter Vovsha, and Robert Donnelly, "Air Passenger Preference for Choice of Airport and Ground Access Mode in the New York Metropolitan Region," *Transportation Research Record* 2042 (2008): 3–11; Greig Harvey, "Study of Airport Access Mode Choice," *Journal of Transportation Engineering* 112 (1986): 525–545.

54. "TAG UNIT 3.5.6: Values of Time and Vehicle Operating Costs," (London: Department for Transport, October 2012).

55. Greig Harvey, "Study of Airport Access Mode Choice," *Journal of Transportation Engineering* 112 (1986): 525–545.

56. "2007–2011 American Community Survey (ACS) 5-year Estimates DP03: Selected Economic Characteristics," (Washington, D.C.; United States Census Bureau, 2011).

57. Surabhi Gupta, Peter Vovsha, and Robert Donnelly, "Air Passenger Preference for Choice of Airport and Ground Access Mode in the New York Metropolitan Region," *Transportation Research Record* 2042 (2008): 3–11.

58. "2010 Air Traffic Report," (New York, NY: Port Authority of New York and New Jersey) www.panynj.gov/airports/pdf-traffic/ATR2010.pdf (accessed July 22, 2013).

59. Mei Ling Tam, William H. K. Lam, and Hing Po Lo, "The Impact of Travel Time Reliability and Perceived Service Quality on Airport Ground Access Mode Choice," Journal of Choice Modelling 4 (2011): 49–69.

60. Surabhi Gupta, Peter Vovsha, and Robert Donnelly, "Air Passenger Preference for Choice of Airport and Ground Access Mode in the New York Metropolitan Region," *Transportation Research Record* 2042 (2008): 3–11.

61. M. Anil Yazici, Camille Kamga, and Kyriacos C. Mouskos, "Analysis of Travel Time Variability in New York City Based on Day-of-Week and Time-of-Day Periods," *Transportation Research Record* 2308 (2012): 83–95.

BIBLIOGRAPHY

"2007–2011 American Community Survey (ACS) 5-year Estimates DP03: Selected Economic Characteristics." Washington, D.C.; United States Census Bureau, 2011.

"2010 Air Traffic Report." New York, NY: Port Authority of New York and New Jersey. www.panynj.gov/airports/pdf-traffic/ATR2010.pdf (accessed July 22, 2013).

Austin, Drew, and P. Christopher Zegras. "Taxicabs as Public Transportation in Boston, Massachusetts." *Transportation Research Record* 2277 (2012): 65–74.

Bacache-Beauvallet, Maya, and Lionel Janin. "Taxicab License Value and Market Regulation." *Transport Policy* 19 (2012): 57–62.

Bai, Ruibin, Jiawei Li, Jason A. D. Atkin, and Graham Kendall. "A Novel Approach to Independent Taxi Scheduling Problem Based on Stable Matching." *Journal of the Operational Research Society*. doi:10.1057/jors.2013.96, 2013.

Baik, Hojong, Antonio A. Trani, Nicolas Hinze, Howard Swingle, Senanu Ashiabor, and Anand Seshadri. "Forecasting Model for Air Taxi, Commercial Airline, and Automobile Demand in the United States." *Transportation Research Record* 2052 (2008): 9–20.

Beesley, Michael E. "Regulation of Taxis." *The Economic Journal* 83 (1973): 150–172.

Ben-Edigbe, Johnnie, and Rosnita Rahman. "Multivariate School Travel Demand Regression Based on Trip Attraction." Paper presented at ICURPT 2010: International Conference on Urban, Regional Planning and Transportation, Paris, France, June 28–30, 2010.

Browning, Raymond C., Emily A. Baker, Jessica A. Herron, and Rodger Kram. "Effects of Obesity and Sex on the Energetic Cost and Preferred Speed of Walking." *Journal of Applied Physiology* 100 (2005): 390–398.

Chang, Yu-Chun. "Factors Affecting Airport Access Mode Choice for Elderly Air Passengers." *Transportation Research Part E* 57 (2013): 105–112.

Chatman, Daniel G. "Does TOD Need the T?" *Journal of the American Planning Association* 79 (2013): 17–31.

Chen, Xiao, Philip B. Ender, Michael Mitchell, and Christine Wells. "Regression Diagnostics," in *Regression with Stata*. UCLA Institute for Digital Research and Education. http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter2/statareg2.htm (accessed November 1, 2013).

Coffman, Richard B., and Chanoch Schreiber. "The Economic Reasons for Price and Entry Regulation of Taxicabs (Comment and Rejoinder)." *Journal of Transport Economics and Policy* 11 (1977): 288–304.

Corpuz, Grace. "Public Transport or Private Vehicle: Factors that Impact on Mode Choice." Paper presented at the 30th Australasian Transportation Research Forum, Melbourne, Australia, 2007.

Davies, Anthony. "Exhaustive Regression: An Exploration of Regression-Based Data Mining Techniques Using Super Computation." RPF Working Paper No. 2008-008. Washington, D.C.: Center of Economic Research, The George Washington University, 2008.

Dempsey, Paul Stephen. "Taxi Industry Regulation, Deregulation, and Reregulation: The Paradox of Market Failure." *Transportation Law Journal* 24 (1996): 73–120.

Elliott, Graham, Antonio Gargano, and Allan Timmermann. "Complete Subset Regressions." *Journal of Econometrics* 177 (2013): 357–373.

Ewing, Reid, William Schroeer, and William Greene. "School Location and Student Travel." *Transportation Research Record* 1895 (2004): 55–63.

Fox, John. *Applied Regression and Generalized Linear Models*, 2nd Ed. Thousand Oaks, CA: SAGE Publications, Inc., 2008.

Frankena, Mark W., and Paul A. Pautler. *An Economic Analysis of Taxicab Regulation*. Bureau of Economics Staff Report, Washington, D.C.: Federal Trade Commission, 1984.

Funderburg, Richard G., Hilary Nixon, Marlon G. Boarnet, and Gavin Ferguson. "New Highways and Land Use Change: Results from a Quasi-Experimental Research Design." *Transportation Research Part A* 44 (2010): 76–98.

Gaunt, Clive, and Terry Black. "The Economic Cost of Taxicab Regulation: The Case of Brisbane." *Economic Analysis and Policy* 26 (1996): 45–58.

Gebeyehu, Mintesnot, and Shin-ei Takano. "Diagnostic Evaluation of Public Transportation Mode Choice in Addis Ababa." *Journal of Public Transportation* 10 (2007): 27–50.

Girardin, Fabien, and Josep Blat. "The Co-Evolution of Taxi Drivers and their In-Car Navigation Systems." *Pervasive and Mobile Computing* 6 (2010): 424–434.

Gupta, Surabhi, Peter Vovsha, and Robert Donnelly. "Air Passenger Preference for Choice of Airport and Ground Access Mode in the New York Metropolitan Region." *Transportation Research Record* 2042 (2008): 3–11.

Harvey, Greig. “Study of Airport Access Mode Choice.” *Journal of Transportation Engineering* 112 (1986): 525–545.

Herring, Ryan, Aude Hofleitner, Pieter Abbeel, and Alexandre Bayen. “Estimating Arterial Traffic Conditions Using Sparse Probe Data.” Paper presented at the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), Funchal, Madeira, September 19–22, 2010.

Jiang, Lan, and Hani S. Mahmassani. “Toll Pricing: Computational Tests for Capturing Heterogeneity of User Preferences.” *Transportation Research Record* 2343 (2013): 105–115.

Kumar, Ajay, and David Levinson. “Estimating, and Validating a New Trip Generation Model: A Case Study of Montgomery County, Maryland.” *Transportation Research Record* 1413 (1992): 107–113.

Lin, Yeqian, Wenquan Li, Feng Qiu, and He Xu. “Research on Optimization of Vehicle Routing Problem for Ride-sharing Taxi.” *Procedia—Social and Behavioral Science* 43 (2012): 494–502.

Loos, Nicolaus. *Value Creation in Leverages Buyouts: Analysis of Factors Driving Private Equity Investment Performance*. Wiesbaden, Germany: Deutscher Uniersitäts-Verlag, 2006.

Luken, Brittany L., and Laurie A. Garrow. “Multiairport Choice Models for the New York Metropolitan Area: Application Based on Ticketing Data.” *Transportation Research Record* 2206 (2011): 24–31.

Mallows, C. L., “Some Comments on C_p.” *Technometrics* 15 (1973): 661–675.

Mandel, Benedikt, Marc Gaudry, and Werner Rothengatter. “A Disaggregate Box-Cox Logit Mode Choice Model of Intercity Passenger Travel in Germany and its Implications for High Speed Rail Demand Forecasts.” *The Annals of Regional Science* 31 (1997): 99–120.

McLeod, A. I., and C. Xu. “bestglm: Best Subset GLM.” http://cran.r-project.org/web/packages/bestglm/vignettes/bestglm.pdf (accessed July 26, 2013).

Morgul, Ender Faruk, and Kaan Ozbay. “Simulation-Based Evaluation of a Feedback-Based Dynamic Congestion Pricing Strategy on Alternate Facilities.” No. 11-3535. Paper presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C., January 23–27, 2011.

Mousavi, Amir, Jonathan Bunker, and Brian Lee. “A New Approach for Trip Generation Estimation for Traffic Impact Assessments.” Paper presented at the 25th ARRB Conference, Shaping the future: linking policy research and outcomes, Perth, Australia, September 23–26, 2012.

O'Neill, Wende A., and Eugene Brown. "Long-Distance Trip Generation Modeling Using ATS." *Transportation Research E-circular Number E-C026*, Washington, D.C.: Transportation Research Board, 2001.

Pels, Eric, Peter Nijkamp, and Piet Rietveld. "Access to and Competition Between Airports: A Case Study for the San Francisco Bay Area." *Transportation Research Part A* 37 (2003): 71–83.

Psaraki, Voula, and Costas Abacoumkin. "Access Mode Choice for Relocated Airports: The New Athens International Airport." *Journal of Air Transport Management* 8 (2002): 89–98.

Racca, David P., and Edward C. Ratledge. "Project Report for 'Factors That Affect and/or Can Alter Mode Choice.'" Project Report, Newark, DE: Center for Applied Demography & Survey Research, University of Delaware, 2004.

Salanova, Josep Maria, Miquel Estrada, Georgia Aifadopoulou, and Evangelos Mitsakis. "A Review of the Modeling of Taxi Services." *Procedia—Social and Behavioral Sciences* 20 (2011): 150–161.

Schaller, Bruce. "A Regression Model of the Number of Taxicabs in U.S. Cities." *Journal of Public Transportation* 8 (2005): 63–78.

Schaller, Bruce. "Entry Controls in Taxi Regulation: Implications of US and Canadian Experience for Taxi Regulation And Deregulation." *Transport Policy* 14 (2007): 490–506.

Schaller Consulting. "The New York City Taxicab Fact Book." March 2006. www.schallerconsult.com/taxi/taxifb.pdf (accessed November 1, 2013).

Schmöcker, Jan-Dirk, Mohammed A. Quddus, Robert B. Noland, and Michael G. H. Bell. "Estimating Trip Generation of Elderly and Disabled People." *Transportation Research Record* 1924 (2005): 9–18.

Schreiber, Chanoch. "The Economic Reasons for Price and Entry Regulation of Taxicabs." *Journal of Transport Economics and Policy* 9 (1975): 268–279.

Schwanen, Tim, and Patricia L. Mokhtarian. "What Affects Commute Mode Choice: Neighborhood Physical Structure or Preferences toward Neighborhoods?" *Journal of Transport Geography* 13 (2005): 83–99.

"TAG UNIT 3.5.6: Values of Time and Vehicle Operating Costs." London: Department for Transport, October 2012.

Tam, Mei Ling, William H. K. Lam, and Hing Po Lo. "The Impact of Travel Time Reliability and Perceived Service Quality on Airport Ground Access Mode Choice." *Journal of Choice Modelling* 4 (2011): 49–69.

Toner, Jeremy P. “The Welfare Effects of Taxicab Regulation in English Towns.” *Economic Analysis and Policy* 40 (2010): 299–312.

Train, Kenneth E. *Discrete Choice Methods with Simulation*, 2nd Ed., Cambridge: Cambridge University Press, 2009.

Trip Generation Manual, 9th Ed. (Washington, D.C.: Institute of Transportation Engineers, 2012).

Turvey, Ralph. “Some Economic Features of the London Cab Trade.” *The Economic Journal* 71 (1961): 79–92.

Yan, Xin, and Xiao Gang Su. *Linear Regression Analysis: Theory and Computing*. Singapore: World Scientific Publishing Co., Pte., Ltd., 2009.

Yazici, M. Anil, Camille Kamga, and Kyriacos C. Mouskos. “Analysis of Travel Time Variability in New York City Based on Day-of-Week and Time-of-Day Periods.” *Transportation Research Record* 2308 (2012): 83–95.

Zhang, Guohui, Yinhai Wang, Heng Wei, and Ping Yi. “A Feedback-Based Dynamic Tolling Algorithm for High-Occupancy Toll Lane Operations.” *Transportation Research Record* 2065 (2008): 54–63.

Zito, R., G. D’Este, and M. A. P. Taylor. “Global Positioning Systems in the Time Domain: How Useful a Tool for Intelligent Vehicle-Highway Systems?” *Transportation Research Part C* 3 (1995): 193–209.

ABOUT THE AUTHORS

ERIC J. GONZALES, PH.D.

Dr. Eric Gonzales is currently an Assistant Professor in the Department of Civil and Environmental Engineering at the University of Massachusetts Amherst, and previously he was an Assistant Professor at Rutgers University. His research interests are in the operation, management, and design of multimodal transportation systems. He has experience in macroscopic modeling of urban street networks and developing models and theory for how to allocate scarce street space to multiple transportation modes. He also has expertise in the planning and management of public transportation systems, including demand responsive systems. Dr. Gonzales received a Ph.D. in Civil and Environmental Engineering at UC Berkeley in 2011. He received recognition as the University of California Transportation Center's Outstanding Student of the Year for 2010-2011 and as an Eno Transportation Foundation Fellow in 2010.

CI (JESSIE) YANG, M.S.

Ci (Jessie) Yang works as a Graduate Assistant in Civil and Environmental Engineering at Rutgers University, specializing in transportation. She received her Master of Science in Environmental Engineering from Texas A & M University, Kingsville, and Bachelor of Science in Environmental Science from People's University of China. Ms. Yang is currently working with GPS data from taxis in New York City in order to study travel demand and mode choice. Ms. Yang plans to become a transportation researcher to help design an intelligent, efficient, and sustainable transportation system worldwide.

ENDER FARUK MORGUL, M.S.

Ender F. Morgul works as a Graduate Research Assistant at the Civil and Urban Engineering Department in Polytechnic Institute of New York University. He received his B.S. from Bogazici University and M.Sc. in Civil Engineering from Rutgers University. His research interests include modeling and prediction of driver behavior, transportation economics, urban freight operations, big data and GPS-based transportation data analysis. His 2010 M.Sc. thesis investigated dynamic congestion pricing using large-scale traffic simulations.

KAAN OZBAY, PH.D.

Kaan M.A. Ozbay is Professor at the Department of Civil and Urban Engineering at NYU-Poly and Center for Urban Science and Progress (CUSP). Dr. Ozbay's research interest in transportation covers advanced technology and sensor applications; incident and emergency management; development of real-time control techniques for traffic, traffic safety, application of artificial intelligence, and operations research techniques in network optimization; development of simulation models for transit and automated highway systems; and transportation economics. He is co-editor of numerous books, including, most recently, Dynamic Traffic Control & Guidance published by Springer Verlag's "Complex Social, Economic and Engineered Networks" series published in 2013. He has published more than 300 refereed papers in scholarly journals and conference proceedings. Prior to NYU, he was a tenured full professor at the Rutgers University Department of Civil and Environmental Engineering. Since 1994, Dr. Ozbay, has been the Principal Investigator and Co-Principal Investigator of 77 projects funded at a level of more than $11,00,000 by National Science Foundation, NJDOT, NYMTC, NY State DOT, New Jersey Highway Authority, USDOT, FHWA, VDOT, CUNY University Transportation Research Center (UTRC), Department of Homeland Security, and USDOT ITS Research Center of Excellence.

PEER REVIEW

San José State University, of the California State University system, and the MTI Board of Trustees have agreed upon a peer review process required for all research published by MNTRC. The purpose of the review process is to ensure that the results presented are based upon a professionally acceptable research protocol.

Research projects begin with the approval of a scope of work by the sponsoring entities, with in-process reviews by the MTI Research Director and the Research Associated Policy Oversight Committee (RAPOC). Review of the draft research product is conducted by the Research Committee of the Board of Trustees and may include invited critiques from other professionals in the subject field. The review is based on the professional propriety of the research methodology.

MTI FOUNDER

Hon. Norman Y. Mineta

Directors

Karen Philbrick, Ph.D.
Executive Director

Hon. Rod Diridon, Sr.
Emeritus Executive Director

Peter Haas, Ph.D.
Education Director

Donna Maurillo
Communications Director

Brian Michael Jenkins
National Transportation Safety and Security Center

Asha Weinstein Agrawal, Ph.D.
National Transportation Finance Center

MNTRC
MINETA NATIONAL TRANSIT RESEARCH CONSORTIUM

SAN JOSÉ STATE UNIVERSITY

Funded by U.S. Department of Transportation