Center for Advanced Multimodal Mobility Solutions and Education

Project ID: 2017 Project 03

Forecasting Ridership for Commuter Rail in Austin

Final Report

by

Scott Kilgore, M.S.
The University of Texas at Austin

and

Randy Machemehl, Ph.D., P.E. (ORCID ID: https://orcid.org/0000-0002-4045-2023)
Professor, Department of Civil and Environmental Engineering
The University of Texas at Austin
301 E. Dean Keeton Street, Stop C1761, Austin, TX 78712
Phone: 1-512-471-4541; Email: rbm@mail.utexas.edu

for

Center for Advanced Multimodal Mobility Solutions and Education
(CAMMSE @ UNC Charlotte)
The University of North Carolina at Charlotte
9201 University City Blvd
Charlotte, NC 28223

September 2018

ACKNOWLEDGMENTS

This project was funded by the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE @ UNC Charlotte), one of the Tier I University Transportation Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT), under the FAST Act. The authors are also very grateful for all of the time and effort spent by DOT and industry professionals to provide project information that was critical for the successful completion of this study.

DISCLAIMER

Table of Contents

List of Figures

List of Tables

EXECUTIVE SUMMARY

Growing cities like Austin, Texas continue to see the need to improve commuter rail options to make people's daily travels in an increasingly congested network easier. Therefore, understanding the way people go about accessing (walking, biking, driving, etc.) boarding stations is fundamentally important to characterizing commuter rail travel. To deal with the growing transportation needs, Capital Metro is proposing the addition of commuter rail services in several corridors where publicly-owned rail right of way is available. Forecasting ridership for such services is problematic due to a lack of experience with rail access modal choices and the potential operational state of the transport system due to rapid growth.

The goal of this study was to examine the influence of access modes from a commuter's decision-making process while understanding the characterization at each boarding station. An onboard survey was deployed on Capital Metro's MetroRail Red Line, revealing access mode patterns and trip purposes for each train station. Then, a binomial logit model was used to determine whether a rider may choose to access the Red Line by walking or driving to the station. This study illustrates a case involving a 32-mile stretch of rail and nine stations where we model the commuters' decision-making process and future trips relating preferences in travel. Whether train passengers decided to walk, bike, ride a bus, or drive with the convenience of locating a park-and-ride facility, data collected based on distances and choice of access mode lead to generalizations of an individual's preference for their trips.

With a geographic information system (GIS) perspective of the city, evaluating demographic and socioeconomic data gathered from each commuter helped to depict the area influenced by urban sprawl. After which, boarding locations were identified in accordance with how the rail passengers were willing to access each station. For instance, the Central Business District (CBD) within downtown Austin describes a commuter who prefers walking rather than any of the other identified modes since the individual is in close proximity to entertainment and social activities alike.
The research carried out suggests that denser areas see a higher number of people willing to walk to the boarding station. A preference for walking was observed at the Downtown Station and Plaza Saltillo Station for entertainment and social trips. On the other hand, people boarding at stations further from CBD often take advantage of parking available at the stations thus their preferred access mode was typically driving. Travelers boarding at park-and-ride stations and for school trips were also found to prefer driving to the station. The model can be used to understand Red Line riders' decision-making, and may be used to predict access modes for a given trip to inform long-term metropolitan planning models.

Finally, this report offers an initial glimpse into the preferences of commuter rail riders in the Austin, TX area, and how such preferences influence the access modes riders use to get to the station. The model specified in this research could be expanded to include other access modes, such as biking or riding the bus, in addition to walking and driving to the station. More data would need to be collected to have enough information to estimate additional access modes. If more data were collected, individual models could be estimated for access mode decision-making at each station, rather than having one model for all stations with dummy variables for each. The same modeling approach used for access mode in this research could be applied to riders' egress trip mode choice as well.

Chapter 1. Introduction

1.1 Problem Statement

Austin, Texas has seen significant sustained population growth for nearly two decades. Rapid growth has led to increasing strain on Austin's transportation network, with considerable congestion during peak periods. Congestion reduces the reliability of travel times and lowers quality of life. As growth continues, the introduction of transit as a solution to reduce congestion has garnered much attention.

Public transportation in Austin is provided by the Capital Metropolitan Transportation Authority (Capital Metro) primarily via bus. As buses are susceptible to roadway congestion, and suffer poor public perception, rail has emerged as a potential solution to shift more travelers to transit. The Red Line MetroRail commuter rail, Austin's first rail transit line, opened in 2010. The Red Line connects Austin's northern suburbs with the downtown central business district (CBD), and is primarily meant to serve commuters who live in the suburbs and work in the CBD.

The Red Line consists of nine stations, which span 32 miles of track connecting downtown Austin to the northern suburb of Leander (Figure 1). Of the nine stations, the northernmost three are suburban park-and-ride stations (Leander Station, Lakeline Station, and Howard Station), with the other six offering no sanctioned parking. Three of the stations (Plaza Saltillo Station, Highland Station, and Crestview Station) currently are or are planned to be surrounded by dense, mixed-use transit-oriented development. Kramer Station is located in an industrial area, with a feeder bus providing service to the nearby Domain shopping center. The nearest stop to the University of Texas at Austin, one of the largest regional employers, is Martin Luther King (MLK) Jr. Station. A feeder bus provides access to the university from MLK Jr. Station, which is located in a residential neighborhood. Downtown Station serves as the gateway to downtown Austin, a major regional employment center. With Red Line stations existing in such varying contexts, the surroundings of each station and its amenities must be accounted for.

Figure 1: MetroRail Red Line Map

As Capital Metro considers additional commuter rail lines, it is important to understand the performance of the existing line, and the behavior of Austin's rail commuters. One key aspect of riding rail is the access trip to the station. Mode choice models often omit the varying choices for the access trip to the station. The factors which influence mode choice for the access trip to rail are poorly understood. To better comprehend how the Red Line functions today, and predict how future commuter rail lines in Austin may perform in terms of ridership, this research aims to understand what influences access mode choice for rail commuters in Austin.

1.2 Objectives

For this research, the focus is only access mode choice modeling, since commuter rail in Austin is limited to the nine stations on the MetroRail Red Line.

1.3 Expected Contributions

Austin Texas is one of the most rapidly growing cities in the United States. Current estimates indicate over 150 people per day are moving to Austin. To deal with the growing transportation needs, Capital Metro is proposing the addition of commuter rail services in several corridors where publicly-owned rail right of way is available. Forecasting ridership for such services is problematic due to a lack of experience with rail access modal choices and the potential operational state of the transport system due to rapid growth. The research team has developed dynamic traffic assignment algorithm for such problems, however, it currently estimates the time and cost of access in very rudimentary ways. This work will develop robust predictive tools for assessing modes used for accessing the proposed commuter rail systems, which will improve future efforts to forecast commuter rail ridership.

1.4 Report Overview

The remainder of this report is organized as follows: Chapter 2 provides an overview of previous work on creating access mode choice models, Chapter 3 describes the data collection process and solution methodology, and lastly, Chapter 4 reviews conclusions gleaned from this work, and directions for future research.

Chapter 2. Literature Review

2.1 Introduction

This chapter provides a review and synthesis of the state-of-the-art and state-of-the-practice literature on the commuter rail access mode problem.

2.2 Commuter Rail Access Mode Choice Models

Revealed preference surveys are powerful tools to understand what drives people's choices. Onboard surveys have long been employed by transit agencies to gather feedback from public transit riders, and are regularly conducted by many agencies. Researchers have used transit agencies' onboard surveys as a source of revealed preference data to model traveler behavior. Fan et. al (1993) sourced revealed preference travel data from a 1987 sample of periodical onboard surveys administered by GO Transit in Toronto, Canada. From this information, Fan et. al (1993) built logit models to predict station choice and access modes for commuter rail users. Vijayakumar et. al (2011) created a statistical model to describe driving distances to commuter rail stations in Montreal, Quebec, Canada from 2003 origin-destination (O-D) survey data collected by the regional transit provider, the Agence Métropolitaine de Transport (AMT). Chakour & Eluru (2013) used 2010 AMT-collected onboard survey data to develop a latent segmentation nested logit model to predict station choice and access mode for commuter rail passengers in Montreal. Bergman et. al (2011) used onboard survey data to develop multinomial and nested logit models describing access mode to suburban commuter rail stations in Portland, Oregon. The survey data used in Bergman et. al (2011) were collected by Tri-Met, the transit agency serving the Portland, Oregon region. Park et. al (2013) administered a mail-back survey of travelers at a single commuter rail station in Mountain View, California. However, Park et. al (2013) modeled walking access to a single station, rather than system-wide, and therefore used a different survey format than other research, despite collecting similar information. Necessary information to model access mode choice includes the trip origin, chosen access mode, rail transit boarding station, and select demographic information. Additional data to enhance models may consist of station characteristics, transit service attributes, and location socioeconomic data, which can be collected from various sources.

Several model estimation tools can be used when modeling access mode to rail transit. The most basic of these is ordinary least squares (OLS) regression. This model type was used by Vijayakumar et. al (2011) to estimate driving distances to commuter rail stations based on trip distances, station parking provision, transit service quality, and demographic qualities such as age and race. A second regression model was developed to model the peak hour boarding per station based on parking supply, street connectivity, and service population. Vijayakumar et. al (2011) were successful in implementing OLS regression for these purposes because the dependent variables are numerical. When modeling choice behavior, more advanced models can be helpful.

One of the most widely used forms of choice models is the logit model. The logit model makes selections from modeler-defined choices using various data inputs. Logit models select from the available choices, which must represent all the choices available to a decision-maker. Park et. al (2013) used two binomial logit (BL) models to describe access mode choices with options to walk or drive. Individual-level and neighborhood-level variables were used to model access mode choices for the subset of survey respondents who walked or drove to the station. Park et. al (2013) used detailed characteristics of paths between origin and destinations to determine whether rail riders would access a station by walking or driving. Similarly, Kim et. al (2006) used multinomial logit (MNL) to model access mode choices to the MetroLink light rail system in the St. Louis metropolitan area. Data from an onboard survey administered by the local metropolitan planning organization (MPO) were used to build a model predicting traveler behavior when selecting between four different access modes. Logit models assume that all alternatives of the choice set are independent of one another.

Nested logit (NL) models relax the independence of alternatives assumption inherent to MNL models. Choice sets are nested within one another to group similar choices. When modeling mode choice to access the Westside Express (WES) in Portland, Oregon, Bergman et. al (2011) initially employed an MNL model to select between different modes. However, it was observed that certain modes were more likely to be substituted for one another in the decision-making process (e.g. active modes vs. motorized modes). As such, Bergman et. al (2011) employed an NL model to accordingly group similar alternatives, and found the NL model performed better than the MNL model. Fan et. al (1993) used an NL model structure when modeling rider boarding station selection and access mode to the commuter rail system in Toronto, Canada. In their model, Fan et. al (1993) specified that access mode is chosen in the top nest and boarding station be chosen in the bottom nest, since the set of accessible boarding stations in the system is dependent on access distance. Debrezion et. al (2008) used an NL structure to model access mode and station choice for Dutch rail riders. Model specifications with access mode in the top nest and station choice in the top nest were tested. Debrezion et. al (2008) concluded that selecting access mode in the top nest yields more accurate decision-making than choosing boarding station first. Chakour & Eluru (2013) used a latent segmentation NL model, which optimized the nesting of access mode and boarding station choices for each decision-maker. The model consisted of a BL model, which selected one of two NL model specifications to make the choices of access mode and boarding station. Chakour & Eluru (2013) applied this model to commuter rail travelers in Montreal, Quebec, Canada and found that roughly two-thirds selected station then mode, with the remaining one-third selecting mode then station. For this research, the focus is only access mode decision-making, since commuter rail in Austin is limited to the nine stations on the MetroRail Red Line. However, it is informative to see how access mode and boarding station have been shown to influence one another in traveler choices.

2.3 Summary

A comprehensive review and synthesis of the current and historical research and development of access mode choice models has been presented in the previous section.

Chapter 3. Solution Methodology

3.1 Introduction

As mentioned in the literature review, a revealed preference survey is a typical method of obtaining the necessary data to build a mode choice model. This section reviews the survey data collection effort and the model building process.

3.2 Data Collection

For this research, an onboard survey was conducted on the MetroRail Red Line over the course of three consecutive days (August 22-24th, 2017). The survey was administered in a paper form, with a total of fifteen questions requiring roughly three minutes to complete. Surveys were distributed to every passenger who would accept a form upon boarding the train. Respondents were encouraged to keep the pen distributed with their survey as an incentive to complete it. An online version of the survey, using Qualtrics software, was also made available to respondents. Student volunteers from The University of Texas at Austin administered the survey by handing out forms, checking in with respondents, answering questions, and collecting finished forms. Each day of the survey, surveyors rode different trains across the Red Line schedule, in order to catch as many different riders as possible. With a limited number of student volunteers, only one train could be surveyed at a time, so the three days of surveying allowed data to be collected from nearly every scheduled train in a day on the Red Line. Data collection was conducted on a Tuesday, Wednesday, and Thursday to capture standard weekday ridership. Due to constraints on the availability of student volunteers, the survey was conducted when the University of Texas was not in session, though the Austin Independent School District (ISD) was in session the week of the survey. While not generally best practice, this concession was accepted since very few University of Texas students live in the northern suburbs of Austin, and faculty and staff who commute on the Red Line would still be traveling to campus despite classes not being in session, as classes began the following week.

The information collected in the survey focused on traveler one-directional trips. The survey asked respondents to identify their access mode, trip origin, trip purpose, vehicle availability and ownership, household size and income, gender, age, education level, alighting station, egress mode, trip destination, and parking availability at their destination. Current time was recorded, and forms were color-coded based on boarding station, so respondents would not have to indicate their boarding station. To preserve privacy, the nearest intersection to trip origins and destinations were requested instead of specific addresses. Additionally, ranges of incomes and ages were provided for respondents to select from, so they would not have to provide such sensitive information exactly. With this information, the characteristics of each traveler's access trip can be determined, along with their general demographics.

A total of 1,203 survey responses were collected across three consecutive days. As with any real-world data collection effort, there were some complications. Many respondents refused to provide certain information, especially locations. Anonymity was allowed by asking respondents for the nearest intersection to their origins and destinations. However, 441 respondents did not provide origin locations, indicating that they may have been uncomfortable providing location information, or uncertainty in how to respond for some respondents.

Additionally, there was confusion about the wording of the question asking for access mode. The question asked "[h]ow did you get to the rail today", which caused confusion for some respondents taking the survey in the afternoon. Despite instructions from the student volunteers, some respondents were confused by this question and appeared to provide the access mode for their morning trip to the station, rather than the access mode for their current trip. As such, the raw Downtown Station survey data showed a much larger share than expected of riders driving to the station, even though there is no parking available, and the nearby street parking is expensive and time-limited. The responses where travelers indicated that they drove to the Downtown Station were checked against respondents' indicated egress mode, and were corrected by using the assumption that travelers did not own two vehicles, one at each end of their trip. There were 13 such responses which were omitted, since the actual access mode used is unknown. Another question which caused confusion among respondents was the question asking for parking availability at their destination. A much larger share of respondents than expected indicated that there is parking available at their destination, and showed little variation among destination stations. The large variability in land uses around the stations on the Red Line casts doubt on this conclusion, especially near the Downtown Station, where permit parking costs hundreds of dollars a month. Asking riders about the "availability" of parking, and only allowing categorical "yes" or "no" responses appears to have caused respondents to consider the existence of parking near their destination rather than their ability to use it.

Of the 1,203 responses collected, 706 had complete access trip information. In addition to the sampling issues discussed, responses without viable locations were omitted for modeling purposes, due to incomplete information. For example, some respondents indicated parallel streets or only one street when asked about the intersection nearest their starting point. An "other" access mode category was provided on the survey, but the 5 responses which selected it were omitted for modeling. Survey responses were checked for clarity, and some illogical responses were removed or corrected if possible. For example, several respondents switched their trip starting point and ending point. Outliers were identified and removed as necessary. This research focuses on access trips to the station, so the 706 complete and logical access trip responses were eligible for modeling purposes, regardless of the quality of the corresponding egress trip data. The spatial distribution of origin locations in relation to Red Line stations is shown in Figure 2.

Figure 3 shows a breakdown of the complete survey responses by time of day. As expected, most surveys were collected during peak times when ridership of the Red Line is very high. The morning peak (6 am to 9 am) accounted for 43.9% of survey responses, with an additional 35.1% of responses collected during the afternoon peak (4 pm-7 pm). The temporal distribution of survey responses is logical since the bulk of ridership on the Red Line is comprised of commuters.

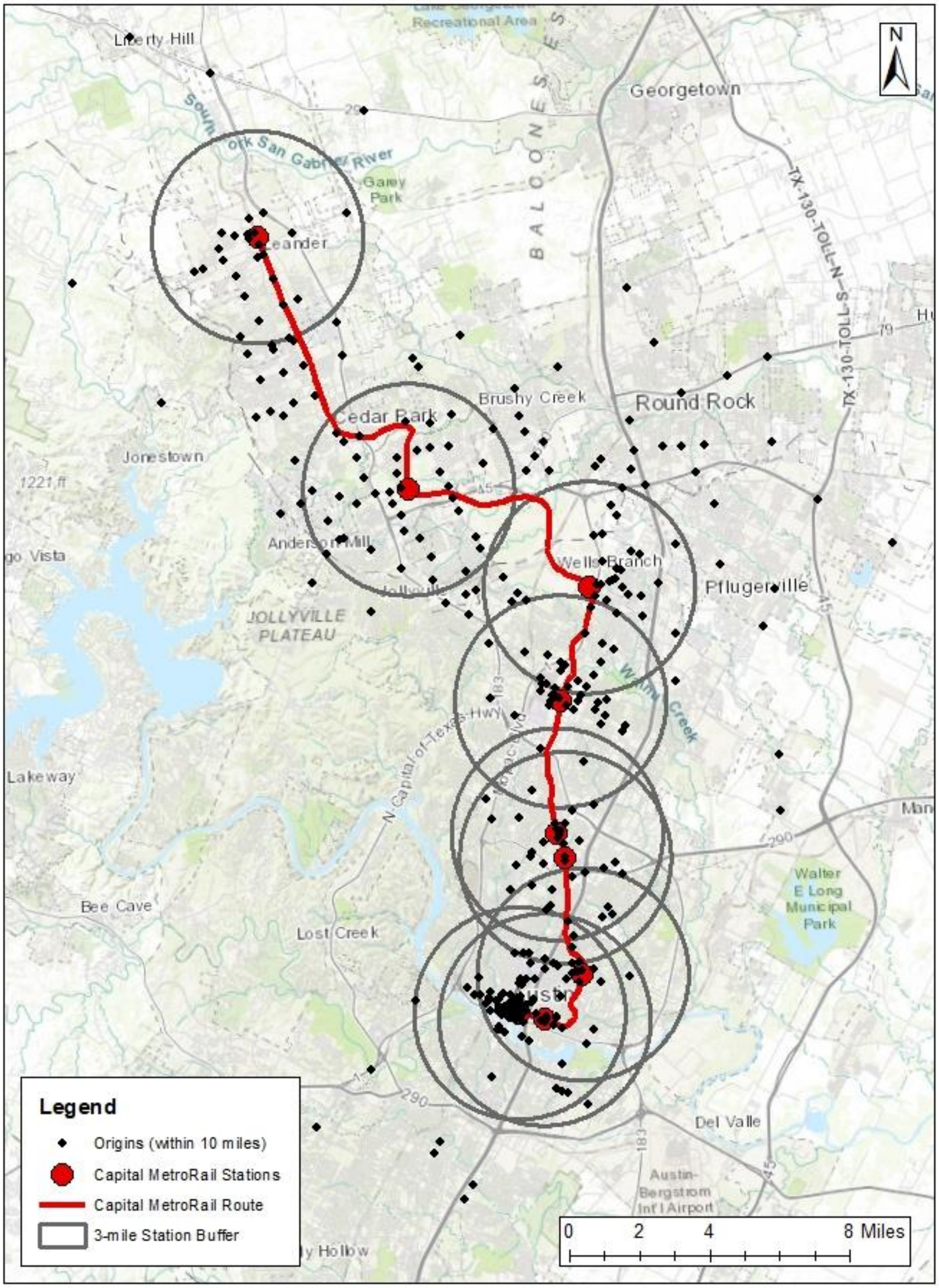

Figure 2: Spatial Distribution of Survey Trip Origins

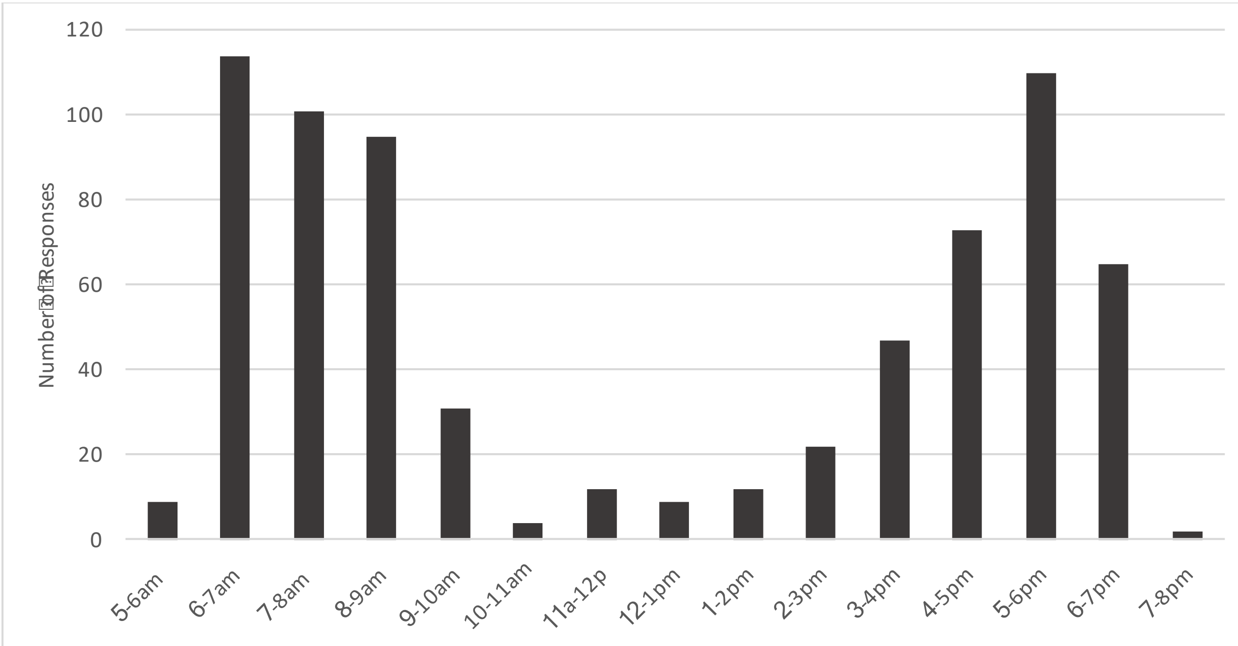

Figure 3: Survey Responses by Time of Day Across All Days of Sampling

Figure 4 shows the aggregate distribution of survey responses by boarding station. The highest numbers of surveys were distributed to riders boarding at Downtown Station, Kramer Station, and the three park and ride stations: Leander Station, Lakeline Station, and Howard Station. Generally, most trips on the Red Line are taken as round-trips, and the data reflects this generalization. Some respondents refused to take a survey on both their morning and afternoon trip, so the observed trend of round trips is slightly diluted. However, there are certainly some one-way trips captured as well.

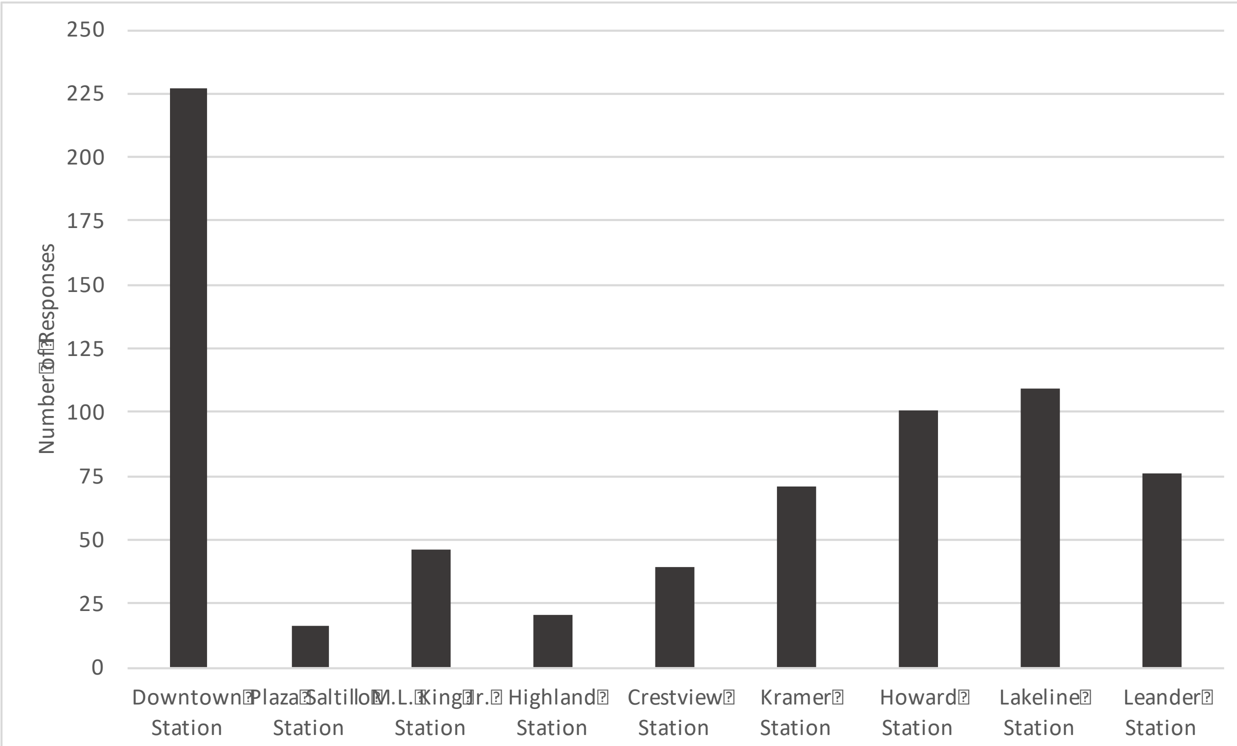

Figure 4: Survey Responses by Boarding Station Across All Days of Sampling

Access modes aggregated across all responses are shown in Figure 5. Access trips are dominated by the walk and drive modes, despite the lack of parking available at most of the stations. When broken out by stations, a clear division emerges between the suburban park-and-ride stations and the urban stations without parking. Table 1 shows the access mode share by stations, aggregated across all times of day. The park-and-ride stations show a pronounced majority of riders accessing the station by vehicle, with the majority driving and a significant portion of riders being dropped off. Stations without parking are much more varied. Downtown Station shows a prevalence of the walk mode for access trips. MLK Jr. Station, connected to The University of Texas at Austin campus by feeder bus, has the highest share of riders accessing the station by bus.

Table 1: Access Mode Share by Stations, Access Distance, and Time of Day

Boarding Station	Access Mode	Number of Responses	Percent of Responses	Average Access Distance (miles)	Standard Deviation of Access Distance	Number during AM Peak (6-9am)	Number during PM Peak (4-7pm)
Downtown Station	Driven by someone else	11	5%	1.10	1.30	2	8
	Ridesharing	3	1%	3.37	2.28	2	1
	Walked all the way	184	81%	0.36	0.32	4	136
	Walked to bus then rode bus	13	6%	1.80	2.04	3	6
	Biked all the way	14	6%	0.96	0.96	4	6
	Biked to bus then rode bus	2	1%	5.60	1.38	0	0
	Total	227	100%	0.60	0.98	15	157
Plaza Saltillo Station	Driven by someone else	2	13%	3.66	2.22	1	1
	Ridesharing	1	6%	0.33	—	1	0
	Walked all the way	9	56%	0.29	0.21	2	3
	Walked to bus then rode bus	1	6%	0.09	—	1	0
	Biked all the way	3	19%	1.48	1.17	3	0
	Total	16	100%	0.93	1.38	8	4
M.L. King Jr. Station	Drive and Park	2	4%	0.91	1.15	2	0
	Driven by someone else	3	7%	0.49	0.29	2	0
	Walked all the way	11	24%	0.46	0.61	6	2
	Walked to bus then rode bus	20	43%	1.50	0.62	1	11
	Biked all the way	10	22%	0.78	0.58	8	0
	Total	46	100%	1.00	0.75	19	13
Highland Station	Drive and Park	4	19%	4.77	5.19	4	0
	Driven by someone else	1	5%	6.67	—	1	0
	Walked all the way	8	38%	0.45	0.61	0	4
	Walked to bus then rode bus	5	24%	2.93	3.17	2	0
	Biked all the way	2	10%	0.86	0.15	1	0
	Biked to bus then rode bus	1	5%	0.10	—	0	1
	Total	21	100%	2.18	3.20	8	5
Crestview Station	Drive and Park	2	5%	1.80	2.45	1	1
	Driven by someone else	7	18%	2.85	3.22	3	2
	Ridesharing	4	10%	1.42	0.96	2	2
	Walked all the way	16	41%	0.55	0.99	4	5
	Walked to bus then rode bus	4	10%	1.56	1.35	1	3
	Biked all the way	6	15%	1.60	0.92	4	2
	Total	39	100%	1.38	1.79	15	15
Kramer Station	Drive and Park	12	17%	1.32	1.19	10	2
	Driven by someone else	9	13%	2.18	2.91	5	1
	Ridesharing	3	4%	1.04	0.83	1	2
	Walked all the way	19	27%	0.50	0.34	4	14
	Walked to bus then rode bus	9	13%	1.71	1.46	5	1
	Biked all the way	17	24%	0.73	0.56	3	11
	Biked to bus then rode bus	2	3%	4.33	3.86	1	0
	Total	71	100%	1.19	1.55	29	31
Howard Station	Drive and Park	66	65%	4.45	4.15	56	1
	Driven by someone else	17	17%	2.18	1.16	7	8
	Walked all the way	7	7%	0.37	0.21	3	2
	Walked to bus then rode bus	8	8%	3.09	2.25	3	1
	Biked all the way	2	2%	0.95	0.50	1	1
	Biked to bus then rode bus	1	1%	7.15	—	1	0
	Total	101	100%	3.63	3.69	71	13
Lakeline Station	Drive and Park	76	70%	3.18	2.18	61	4
	Driven by someone else	11	10%	2.83	1.76	8	1
	Ridesharing	3	3%	4.23	2.69	0	0
	Walked all the way	9	8%	0.74	0.29	3	3
	Walked to bus then rode bus	2	2%	1.52	0.70	0	1
	Biked all the way	8	7%	2.11	1.34	6	1
	Total	109	100%	2.87	2.10	78	10
Leander Station	Drive and Park	57	75%	2.95	4.62	54	0
	Driven by someone else	11	14%	2.06	1.71	9	0
	Walked all the way	2	3%	0.91	0.52	0	1
	Biked all the way	6	8%	1.76	1.53	5	0
	Total	76	100%	2.67	4.10	68	1
All Stations	Drive and Park	219	31%	3.40	3.67	188	8
	Driven by someone else	72	10%	2.19	2.01	38	21
	Ridesharing	14	2%	2.28	2.06	6	5
	Walked all the way	265	38%	0.40	0.42	26	170
	Walked to bus then rode bus	62	9%	1.90	1.72	16	23
	Biked all the way	68	10%	1.16	1.00	35	21
	Biked to bus then rode bus	6	1%	4.52	3.02	2	1
	Total	706	100%	1.79	2.61	311	249

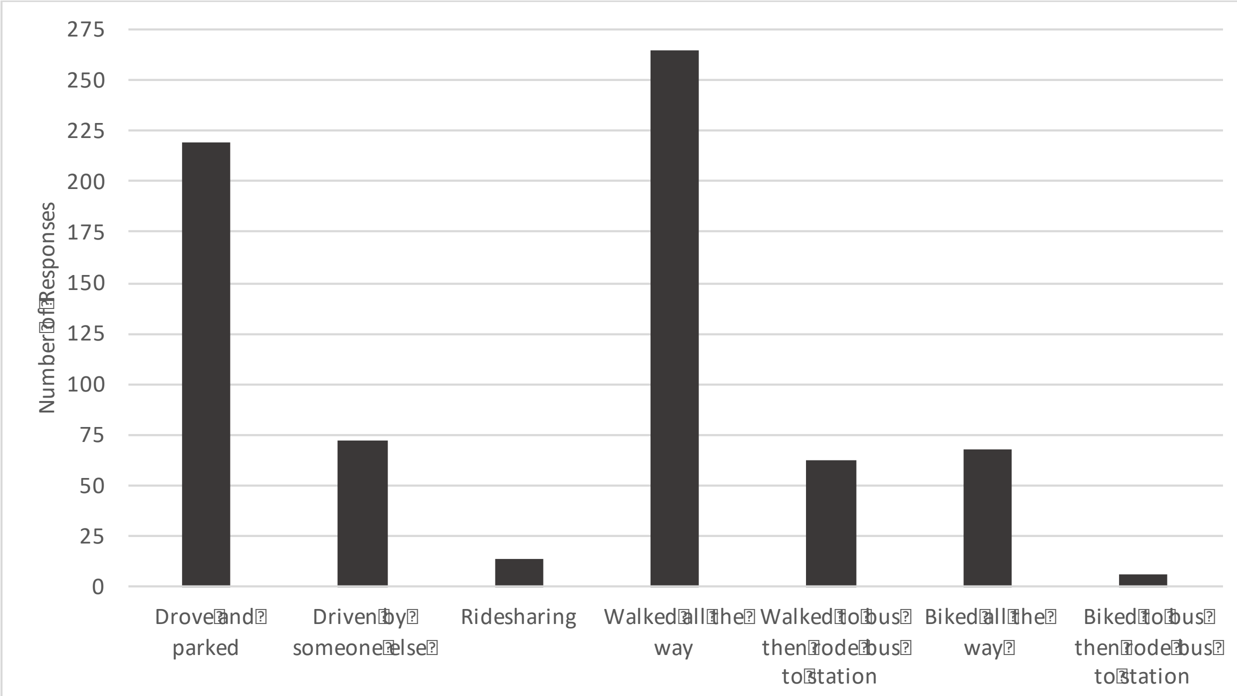

Figure 5: Survey Responses by Access Mode Across All Days of Sampling

Table 2 shows the share of trip purposes for all responses. The vast majority of trips surveyed were work trips, with a 92.2% share of responses. The next most common trip type was entertainment and social trips, which accounted for 5.1% of responses. School travel only accounted for 1.6% of trips, while personal care and shopping trips accounted for 0.4% and 0.7% of trips, respectively.

Table 2: Survey Responses by Trip Purpose

Trip Purpose	Number of Responses	Percent of Responses
Work	651	92.2%
School	11	1.6%
Entertainment/Social	36	5.1%
Shopping/Errand	5	0.7%
Personal Care/Doctor	3	0.4%
Total	706	100.0%

The age distribution of responses is shown in Figure 6. The majority (52.7%) of respondents were between the ages of 25 and 39, with another 25.5% of respondents between 40 and 54. People between ages 18 to 24 and 55 to 69 made up 10.5% and 10.9% of responses, respectively. No one under the age of 18 was surveyed, and none of the responses eligible for modeling purposes were from people 70 or older. The age distribution in the data is in line with expectations for users of the Red Line. Most of the people traveling on the Red Line are working-age commuters, who live in the suburbs and commute downtown for work. As people become more established in their career, they are more likely to have reserved parking or are more willing to purchase a parking pass for an expensive downtown lot. Therefore, a bias toward younger working-age riders is expected for the Red Line, and the collected data supports this hypothesis.

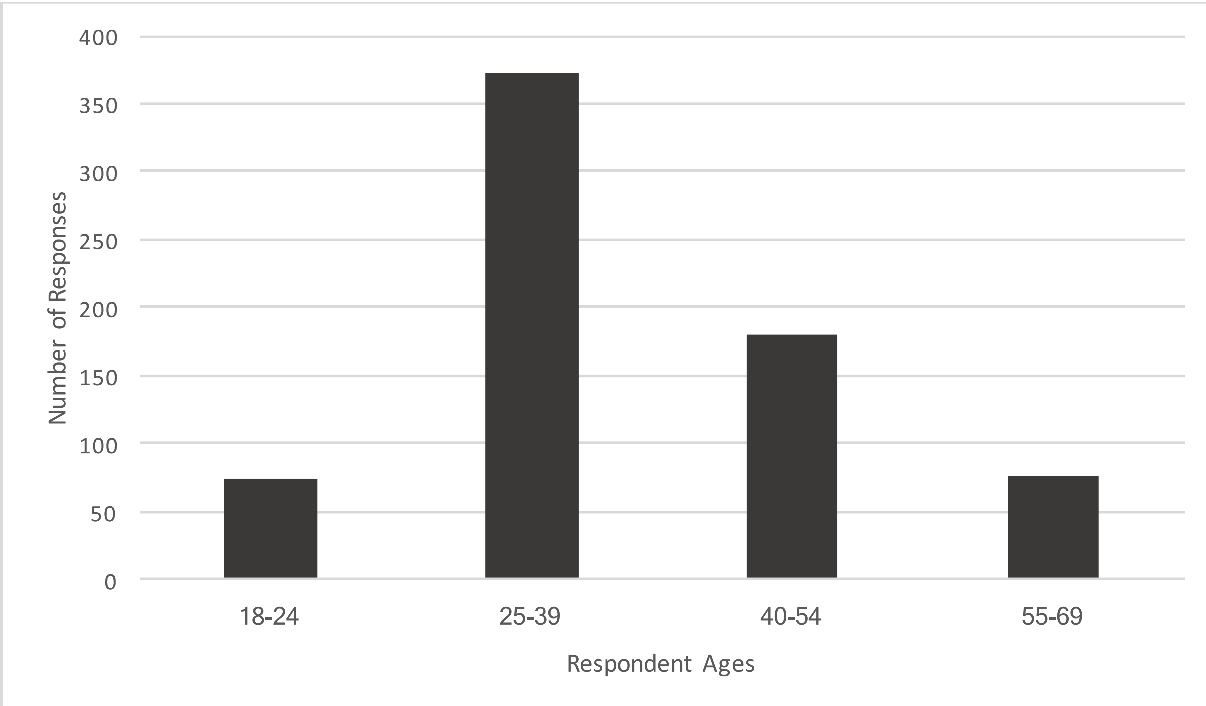

Figure 6: Survey Responses by Age

The survey form allowed respondents to select three options to specify their gender: Male, Female and Other. Table 3 shows the breakdown of responses by gender. A surprisingly large portion of respondents was male, with 66.1% of respondents identifying as male. An additional 32.6% of respondents identified as female, with the other category making up 0.8% of responses. It was expected that the percentage of male and female riders on the Red Line would be similar to the gender bias of the Austin workforce, with men and women making up nearly equal portions. It is hypothesized that the higher proportion of males riding the Red Line could be due to the nature of employment in the CBD being more biased toward male workers.

Table 3: Survey Responses by Gender

Gender	Number of Responses	Percent of Responses
Male	467	66.1%
Female	230	32.6%
Other	6	0.8%
No Response	3	0.4%
Total	706	100.0%

Riders on the Red Line are generally well-educated and wealthy. Table 4 and Table 5 show responses by education and income levels, respectively. The education categories for some college, bachelor's degree, and advanced degree together account for 93.8% of people surveyed. Only 8.5% of respondents live in households with income under $35,000 annually.

Table 4: Survey Responses by Education Level

Education Level	Number of Responses	Percent of Responses
Less than High School	1	0.1%
High School/GED	24	3.4%
Some College	140	19.8%
Technical School	18	2.5%
Bachelor's Degree	354	50.1%
Advanced Dregree	168	23.8%
No Response	1	0.1%
Total	706	100.0%

Table 5: Survey Responses by Income Level

HH Income	Number of Responses	Percent of Responses
$34,999 or less	60	8.5%
$35,000 to $74,999	189	26.8%
$75,000 to $99,999	135	19.1%
$100,000 +	283	40.1%
No Response	39	5.5%
Total	706	100.0%

When considering travel choices, vehicle ownership can greatly limit options available to travelers. Commuters on the Red Line are mostly from car-owning households: only 7.2% of respondents lived in households without a vehicle. Vehicle ownership information is shown in Table 6.

Table 6: Survey Responses by Household Vehicle Ownership

HH Vehicle Ownership	Number of Responses	Percent of Responses
0	51	7.2%
1	222	31.4%
2	330	46.7%
3+	101	14.3%
No Response	2	0.3%
Total	706	100.0%

3.3 Modeling

The first step in the modeling process is to see how the dependent variable changes with each independent variable. With interval or ratio scale data, scatter plots can be used to visualize the relationship between two variables. In this case, however, the dependent variable being modeled is the access mode chosen, which is a nominal scale measurement. Additionally, all of the independent variables are nominal or ordinal scale measurements with the exception of the access distance. With these types of data, it is not possible to use scatter plots to determine the relationship between each independent variable and the dependent variable. In lieu of scatter plots, from which a correlation coefficient can be found, the contingency coefficient was calculated to determine the relationship between each independent variable and the dependent variable.

The contingency coefficient is a statistic which determines the ability of one categorical variable to predict another categorical variable. To calculate the contingency coefficient, the chi-squared value for two related sets of categorical variables must be found. Once the chi-squared value and contingency coefficient are found for each set of the variables, the contingency coefficient is compared to the chi-squared distribution table to determine if it is significantly different from zero. The number of degrees of freedom is calculated by the number of categories for the independent variable minus one multiplied by the number of categories for the dependent variable minus one. The equations for chi-squared and the contingency coefficient (C) are shown in Figure 7.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$C = \sqrt{\frac{\chi^2}{N+\chi^2}}$$

Figure 7: Equations for Chi-Squared and the Contingency Coefficient

An additional constraint on the process dictates that the expected values when calculating chi-squared should all be above one. If any expected values are below one, then the corresponding categories should be removed from consideration. When calculating chi-squared for access mode across all independent variables, this constraint was invoked for several of the access mode categories. Since the majority of responses used the walk or drive access modes, only the walk and drive modes were selected for modeling. Using only these two modes ensured that enough data was available to build a quality model. Certain categories of potential predictor variables were removed in this process as well. For example, the “other” gender option was removed for having expected values below one, as were the shopping and personal care trip purposes.

Using a significance level of 0.05, a total of four predictor variables were found to be statistically significant predictors of access mode. Access trip distance, age, vehicle ownership, trip purpose (excluding the shopping and personal care categories), and boarding station were all found to be statistically significant predictors between the walk and drive access modes. Education level, gender, income and household size were also tested, but were not significant predictors of access mode.

Narrowing down the scope of the model to predict between walk and drive modes resulted in the exclusion of responses using other modes when estimating the model. Additionally, the walk and drive mode responses used were narrowed down to only include responses for which these two modes could reasonably compete. This resulted in responses with access distances greater than three miles to be excluded from the modeling responses, as shown by the station buffers in Figure 2. Respondents from households without a vehicle were also removed for modeling purposes because they do not really have a choice between the walk and drive modes. All told, 371 responses were used to estimate the final model. Table 7 shows a description of the final modeling responses by access trip distance and time of day. Demographic data for the final model responses are shown in Table 8 and Table 9.

Table 7: Model Responses by Access Mode, Distance, and Time of Day

Boarding Station	Access Mode	Number of Responses	Percent of Responses	Average Access Distance (miles)	Standard Deviation of Access Distance	Number during AM Peak (6-9am)	Number during PM Peak (4-7pm)
Downtown Station	Drive and Park	0	0%	—	—	0	0
	Walked all the way	174	100%	0.36	0.33	2	129
	Total	174	100%	0.36	0.33	2	129
Plaza Saltillo Station	Drive and Park	0	0%	—	—	0	0
	Walked all the way	8	100%	0.26	0.20	2	3
	Total	8	100%	0.26	0.20	2	3
M.L. King Jr. Station	Drive and Park	2	15%	0.91	1.15	2	0
	Walked all the way	11	85%	0.46	0.61	6	2
	Total	13	100%	0.53	0.67	8	2
Highland Station	Drive and Park	2	22%	0.67	0.06	2	0
	Walked all the way	7	78%	0.51	0.63	0	3
	Total	9	100%	0.55	0.55	2	3
Crestview Station	Drive and Park	1	6%	0.07	—	1	0
	Walked all the way	15	94%	0.33	0.49	4	4
	Total	16	100%	0.32	0.48	5	4
Kramer Station	Drive and Park	10	42%	0.90	0.61	8	2
	Walked all the way	14	58%	0.43	0.30	1	12
	Total	24	100%	0.62	0.51	9	14
Howard Station	Drive and Park	31	86%	1.49	0.80	25	1
	Walked all the way	5	14%	0.43	0.23	3	1
	Total	36	100%	1.34	0.83	28	2
Lakeline Station	Drive and Park	42	88%	1.79	0.84	38	1
	Walked all the way	6	13%	0.76	0.31	2	2
	Total	48	100%	1.66	0.86	40	3
Leander Station	Drive and Park	42	98%	1.06	0.85	41	0
	Walked all the way	1	2%	1.28	—	0	0
	Total	43	100%	1.06	0.84	41	0
All Stations	Drive and Park	130	35%	1.37	0.88	117	4
	Walked all the way	241	65%	0.38	0.37	20	156
	Total	371	100%	0.73	0.76	137	160

Table 8: Model Responses by Trip Purpose

Trip Purpose	Number of Responses	Percent of Responses
Work	349	94.1%
School	3	0.8%
Entertainment/Social	19	5.1%
Total	371	100.0%

Table 9: Model Responses by Household Vehicle Ownership

HH Vehicle Ownership	Number of Responses	Percent of Responses
1	114	30.7%
2	204	55.0%
3+	53	14.3%
Total	371	100.0%

After narrowing down to the final model responses, the age parameter was no longer significant, and was not included in the model. With significant model parameters determined, a binomial logit model was created to predict between walk and drive access modes given a traveler's household vehicle ownership, trip purpose, boarding station, and access distance to their boarding station. Microsoft Excel software was used to generate the model. The final model specification is shown in Figure 8, where *d* is the access trip distance in miles, and all other variables are dummy variables which take a value of 1 if the corresponding boarding station, vehicle ownership, or trip purpose applies to a traveler, and a value of 0 otherwise. The probability of selecting walk or drive modes is given by raising *e* to the power of *L* divided by one plus *e* raised to the power of *L*. The walk mode is represented by zero, and the drive mode by 1, so responses with probabilities less than 0.5 are more likely to walk, and those with probabilities greater than 0.5 are more likely to drive.

$$L = -8.99 - 1.50\,d + 1.58\,d^2 - 1.97\,(\text{PlazaSaltillo}) + 6.70\,(\text{MLK}) + 7.61\,(\text{Highland}) + 6.19\,(\text{Crestview}) + 8.74\,(\text{Kramer}) + 10.24\,(\text{Howard}) + 10.14\,(\text{Lakeline}) + 12.62\,(\text{Leander}) + 3.37\,(\text{School}) - 0.22\,(\text{Entertainment}) - 0.37\,(\text{TwoVehicles}) + 0.43\,(\text{ThreeVehicles})$$

$$P(\text{Walk or Drive}) = \frac{e^L}{1+e^L}$$

Where:

d is access trip distance to the station, in miles

"PlazaSaltillo," "MLK," "Highland," "Crestview," "Kramer," "Howard," "Lakeline," and "Leander" are dummy variables which take a value of 1 for responses boarding at that station, and 0 for all other stations

"School" and "Entertainment" are dummy variables which take a value of 1 for responses with that trip purpose, and a value of 0 otherwise

"TwoVehicles" and "ThreeVehicles" are dummy variables which take a value of 1 for responses with corresponding household vehicle ownership, and a value of 0 otherwise

e is Euler's constant

Figure 8: Final Binomial Logit Model Specification

To arrive at the final model specification, dummy variables were created for each category of the categorical variables. From these, a reference category for each categorical variable was selected. The reference category for boarding station was selected as Downtown Station, the work purpose was chosen as the reference for trip purpose, and households with one vehicle were selected as the reference category for vehicle ownership. As such, variables for these categories are not included in the final model, and the coefficients on other categories for the corresponding variables indicate how those categories affect decision-making with respect to the reference category. For example, the coefficient on the parameter for Leander Station indicates the access mode choice impact of boarding at Leander Station as compared to Downtown Station.

Access trip distance was tested as a first-, second-, and third-order parameter to determine the best fit. The best fit is determined by the greatest log-likelihood value for the model. The maximum log-likelihood was found for the model with second-order access trip distance included. This allows the impact of the access trip distance to vary parabolically rather than linearly, as it would if only the first-order term were included. Adding in the third-order term decreased the log-likelihood as compared to the second-order model, indicating that the third-order term does not improve the predictive ability of the model.

Looking at the final model specification in Figure 8, the relative influence of different aspects of people's access trips can be determined. The model specifies the walk mode as "0" and driving as "1" meaning that positive coefficients indicate aspects of access trips that encourage driving, and negative coefficients indicate aspects which encourage walking. The model constant is negative, indicating an initial bias toward the walk mode. This is likely due to the selection of Downtown Station as the reference boarding station, since there is a strong preference to access Downtown Station by walking. Interestingly, the first-order access distance term is negative, while the second-order term is positive. The parabola defined by these two terms is concave up, and is negative until access distance is over one mile, and then becomes positive and increases rapidly. In terms of impacting decision-making, this suggests that there is a slight preference for walking to the station for trips under one mile, but a strong preference to drive for trips over a mile, all else held constant. A graph of the parabolic effect of the distance term on the utility function is shown in Figure 9.

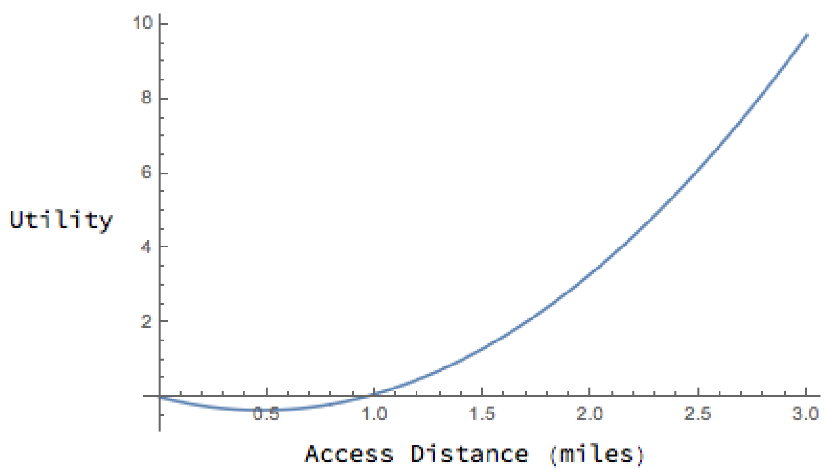

Figure 9: Parabolic Effect of the Access Distance Term on the Utility Function

Looking at the effect of boarding stations, Plaza Saltillo Station was the only station with a negative coefficient, thus encouraging riders to access the station by walking. The other stations had positive coefficients, with the three park and ride stations having the largest coefficients. The park and ride station coefficients were large enough to counteract the negative constant of the model, indicating that there is an overall preference to drive to the park and ride stations. This result is logical since riders boarding the rail at park and ride stations are more likely to drive to the station, as was observed in the survey data. Crestview Station, Highland Station, and MLK Jr. Station also had positive coefficients, but with smaller magnitude than the constant, so that there is a slight preference to walk to those stations. Kramer Station has a positive coefficient nearly equal to the constant, indicating that there is no preexisting preference to walk or drive when accessing Kramer Station.

The two trip purpose categories have smaller coefficients than the boarding station categories, but show interesting results. Unexpectedly, the coefficient for school trips was positive, indicating a greater preference to drive to the station than work trips. This is unexpected as students generally have a lower value of time than professionals and would be expected to use less expensive modes than driving. Part of this result could be due to the low number of students responding to the survey. Entertainment and social trips had a much smaller coefficient than school trips, and had a negative sign. This shows a slight preference for walking to the station, but the magnitude of the coefficient for entertainment trips is comparatively very small.

Household vehicle ownership appears to have little impact on rider's preference between walking and driving to the station. As compared to riders from one vehicle households, riders from households that own two vehicles are slightly less likely to drive. The two vehicle household coefficient is negative, but its magnitude is very small, indicating that there is not much difference in decision-making for individuals from two vehicle households. Similarly, the coefficient for households with three or more vehicles is positive with a small magnitude, slightly increasing the likelihood an individual would choose to drive to the station.

Out of the 371 responses used to generate the model, only 26 are incorrectly predicted. Therefore, the model accurately predicts access mode for 93.0% of the responses used to estimate it. Summary

3.4 Summary

The objective of this chapter is to present the data collection process and the access choice model that was built using revealed preference survey data.

Chapter 4. Summary and Conclusions

4.1 Introduction

This chapter summarizes the findings of this work and provides a succinct overview of potential directions for future research. As mentioned in the first chapter, this work highlights the preferences of commuter rail riders in the Austin, TX area, and how such preferences influence the access modes riders use to get to the station. Understanding these preferences enables the improvement of future ridership forecasting efforts.

4.2 Summary and Conclusions

In this research, a binomial logit model was developed to predict commuter rail riders' access mode to their boarding station in Austin, TX. Revealed preference survey data were collected from commuter rail riders and were used to build a logit model predicting walk versus drive access modes. Access distance, boarding station, trip purpose, and household vehicle ownership were used to predict access mode. The model developed shows how certain characteristics influence Austin rail riders' access mode decision-making. It can also be used to estimate how a given traveler may choose to access the Red Line given information about their demographics and access trip.

Riders are more likely to walk to the station when boarding the Red Line at Downtown Station or Plaza Saltillo Station and are more likely to drive when boarding at park-and-ride stations. Riders traveling for school were more likely to drive than those traveling for work, and people traveling for entertainment were found to be slightly more likely to walk than those traveling for work. Riders from households with two cars were found to be slightly more likely to walk than those from households with one vehicle, and riders from households with three or more vehicles were found to be slightly more likely to drive than those from one vehicle households.

4.3 Directions for Future Research

This research offers an initial glimpse into the preferences of commuter rail riders in the Austin, TX area, and how such preferences influence the access modes riders use to get to the station. The model specified in this research could be expanded to include other access modes, such as biking or riding the bus, in addition to walking and driving to the station. More data would need to be collected to have enough information to estimate additional access modes. If more data were collected, individual models could be estimated for access mode decision-making at each station, rather than having one model for all stations with dummy variables for each. The same modeling approach used for access mode in this research could be applied to riders' egress trip mode choice as well.

References

1. Altes, K. (2014). [MetroRail Red Line map]. Retrieved November 10, 2017, from http://ecohomesaustin.homestead.com/SearchHomesMetroRailStations.html
2. Bergman, A., J. Gliebe and J. Strathman. 2011. Modeling Access Mode Choice for Inter-Suburban Commuter Rail. Journal of Public Transportation, Vol. 14, No. 4, 2011.
3. Chakour, V. & N. Eluru. 2013. Analyzing Commuter Train User Behavior: A Decision Framework for Access Mode and Station Choice. Transportation (2014) 41:211–228
4. Debrezion et al. 2008. Modelling the joint access mode and railway station choice. Transportation Research Part E 45 (2009) 270-283.
5. Fan, KS, E Miller & D Badoe. 1993. Modeling Rail Access Mode and Station Choice. Transportation Research Record 1413, 1993. http://onlinepubs.trb.org/Onlinepubs/trr/1993/1413/1413-006.pdf
6. Kim et al. 2006. Analysis of Light Rail Rider Travel Behavior: Impacts of Individual, Built Environment, and Crime Characteristics on Transit Access. Transportation Research Part A 41 (2007) 511-522.
7. Park et al. 2013. To Walk or Not to Walk: Testing the Effect of Path Walkability on Transit Users' Access Mode Choices to the Station. International Journal of Sustainable Transportation 9 (2015): 529-541.
8. Vijayakumar, N., A. El-Geneidy and Z. Patterson. 2011. Driving to Suburban Rail Stations: Understanding Variables That Affect Driving Distance and Station Demand. Transportation Research Record 2219 (p. 97-103), 2011.