Multimodal Connections with Transitways: Ridership, Access Mode, and Route Choice Implications

Andrew Guthrie, Yingling Fan
Humphrey School of Public Affairs

Alireza Khani, Jacqueline Nowak
Department of Civil, Environmental, and Geo-Engineering
University of Minnesota

March 2019
Multimodal Connections with Transitways: Ridership, Access Mode, and Route Choice Implications

Andrew Guthrie, Yingling Fan, Alireza Khani, and Jacqueline Nowak

Humphrey School of Public Affairs
University of Minnesota
301 19th Ave S
Minneapolis, MN 55455

Center for Transportation Studies
University of Minnesota
University Office Plaza, Suite 440
2221 University Avenue SE
Minneapolis, MN 55414

Transitways—premium transit corridors employing technologies such as Light Rail Transit or Bus Rapid Transit—often depend on a variety of access and egress modes to connect users with their trip origins and destinations. This study seeks to create better understanding of how users access transitway stations by applying mode choice models, route choice models, and direct ridership models. Choice models were applied to revealed-preference transit passenger data from the Twin Cities show key components to user decisions regarding how to reach high-quality transit. To explore users’ choice of routes through the transit systems, schedule-based shortest path and multi-criterion shortest path algorithms were combined to investigate whether transit riders choose to take the shortest path between their origin and destination, a subjectively shortest path, or neither. In terms of ridership models, Poisson regression model were used to estimate average weekday boardings at transitway stations in 10 regions around the United States as a function of pedestrian, bicycle, and bus connections.
MULTIMODAL CONNECTIONS WITH TRANSITWAYS: RIDERSHIP, ACCESS MODE, AND ROUTE CHOICE IMPLICATIONS

FINAL REPORT

Prepared by:
Andrew Guthrie
Yingling Fan
Humphrey School of Public Affairs
University of Minnesota
Alireza Khani
Jacqueline Nowak
Department of Civil, Environmental, and Geo- Engineering
University of Minnesota

MARCH 2019

Published by:
Center for Transportation Studies
University of Minnesota
University Office Plaza, Suite 440
2221 University Avenue SE
Minneapolis, MN 55414

This report represents the results of research conducted by the authors and does not necessarily represent the views or policies of the Center for Transportation Studies, the University of Minnesota, or the sponsoring organizations of the Transitway Impacts Research Program.

The authors, the Center for Transportation Studies, the University of Minnesota, and the sponsoring organizations of the Transitway Impacts Research Program do not endorse products or manufacturers. Any trade or manufacturers’ names that may appear herein do so solely because they are considered essential to this report.

TIRP funding partners and program supporters: Anoka County, Center for Transportation Studies, Center for Urban and Regional Affairs, City of Bloomington, City of Minneapolis, City of Saint Paul, Dakota County, Federal Transit Administration, Hennepin County, Hennepin–University Partnership, Humphrey School of Public Affairs State and Local Policy Program, Metropolitan Council, Metro Transit, Minnesota Department of Transportation, Ramsey County, University Metropolitan Consortium, Washington County.
# TABLE OF CONTENTS

**CHAPTER 1: Introduction**

1.1 Access Mode Choice Modeling ................................................................. 1
1.2 Route Choice Modeling ........................................................................... 1
1.3 Station-Level Ridership Modeling ........................................................... 2
1.4 Report Organization ................................................................................ 3

**CHAPTER 2: Literature Review**

2.1 Access Mode Choice .............................................................................. 4
2.2 Route Choice ........................................................................................... 5
2.3 Station-Level Ridership .......................................................................... 7
   2.3.1 Bus Connections to Transitway Ridership ........................................... 7
   2.3.2 Non-Motorized Access to Transitways ................................................. 8
   2.3.3 Summary .......................................................................................... 9

**CHAPTER 3: Methods**

3.1 Access Mode Choice Modeling Approach ............................................. 10
   3.1.1 Model Specification ......................................................................... 13
   3.1.2 Model Estimation ............................................................................ 14
3.2 Route Choice Modeling ......................................................................... 14
   3.2.1 Multicriterion Shortest Path ............................................................... 14
   3.2.2 Schedule-Based Shortest Path ............................................................ 16
   3.2.3 Algorithm ....................................................................................... 17
   3.2.4 Illustrative Example ......................................................................... 18
   3.2.5 Data ............................................................................................... 21
   3.2.6 Procedure ....................................................................................... 21
3.3 Station Level Ridership Modeling ........................................................... 22
3.3.1 Measuring Multimodal Connections to Stations ................................................................. 23
3.3.2 Variables and Model Specification .................................................................................. 27

CHAPTER 4: Mode Choice Modeling Results .............................................................................. 31
4.1 Model Validation ................................................................................................................. 34

CHAPTER 5: Route Choice Modeling Results .............................................................................. 37
5.1 Example of Route Overlap Computation: .......................................................................... 39

CHAPTER 6: Station-Level Ridership Modeling Results ............................................................. 43
6.1 Initial Models ...................................................................................................................... 43
6.2 Simplified Models ............................................................................................................. 45
6.2.1 Model Predictions ....................................................................................................... 46
6.3 Final Model Refinements ................................................................................................ 47
6.3.1 Final Model Predictions ............................................................................................ 49

CHAPTER 7: Conclusions ....................................................................................................... 52
REFERENCES ......................................................................................................................... 55
LIST OF FIGURES

Figure 3-1: Transitway access mode shares from the 2016 On Board transit survey of the Twin Cities Metropolitan Area .................................................................................................................................................. 11

Figure 3-2: Histogram of transitway access times separated by mode. Modes with less than 1% share are not pictured ................................................................................................................................................................. 12

Figure 3-3: Example of a network for bicriterion analysis. Paths include travel time and travel cost components ............................................................................................................................................................. 15

Figure 3-4: Paths plotted by cost and time for bicriterion shortest path (left) and paths plotted by in-vehicle time, number of transfers, and walking time (right) .................................................................................................................. 16

Figure 3-5: Sample time space diagram for three alternative paths between an origin and destination pair ........................................................................................................................................................................ 17

Figure 3-6: Example of the shortest path algorithm implementation. Both left and right begin with the top initialization. The left is a baseline scenario with no time weighting. The right side weights walking time as three times worse than in-vehicle or waiting time. .................................................................................. 20

Figure 3-7: Example of Areas Reachable by Bus in 15 Minutes ................................................................................. 24

Figure 3-8: Example of Areas Reachable by Bicycle in 15 Minutes ............................................................................... 25

Figure 3-9: Example of Areas Reachable on Foot in 15 Minutes ...................................................................................... 26

Figure 3-10: Comparison of Boardings and Areas Reachable by Bus in 15 Minutes ......................................................... 27

Figure 3-11: Comparison of Boardings and Areas Reachable by Bicycle in 15 Minutes ................................................ 28

Figure 3-12: Comparison of Boardings and Areas Reachable on Foot in 15 Minutes ...................................................... 29

Figure 4-1: Comparison of model predictions of access mode to mode reported in survey data ................................. 36

Figure 5-1: Unlinked trip comparison between AFC data and schedule-based shortest paths for different weighting scenarios ........................................................................................................................................ 38

Figure 5-2: Route overlap (top) and link overlap (bottom) between paths generated by each of the five weighting schemes ........................................................................................................................................ 40

Figure 5-3: Share of passenger paths from AFC data that have the greatest route overlap when compared to each of the five weighting scenarios ........................................................................................................... 41

Figure 5-4: Histogram of passengers with certain percentages of route overlap between AFC data and the five weighting scenarios ........................................................................................................................................ 42
LIST OF TABLES

Table 3-1: List of variables used for model estimation ................................................................. 12

Table 3-2: Example link travel times corresponding with Figure 4. Values for In-Vehicle and Waiting times are constant across scenarios. Walking values are used on the left side of Figure 4 and Weighted Walking values are used on the right side ............................................................................................................. 19

Table 3-3: Weighting scheme and description of each test scenario ............................................... 22

Table 3-4: Regions and Systems Included ...................................................................................... 22

Table 4-1: Model estimation results for base model. Parameter values normalized to the time coefficient, B_TIME .......................................................................................................................... 31

Table 4-2: Model estimation results for interactive term model. Parameter values normalized to the time coefficient, B_TIME .......................................................................................................................... 32

Table 4-3: Model estimation results for relaxed model. Parameter values normalized to transit time coefficient, B_TIME_TRANSIT .................................................................................................................. 33

Table 4-4: Results of Pearson chi-squared goodness of fit test for all models ............................... 35

Table 5-1: Number of passengers assigned to paths for each weighting scenario out of 1,093 .......... 37

Table 6-1: Descriptive Statistics .................................................................................................... 43

Table 6-2: Initial Model Results ................................................................................................... 44

Table 6-3: Simplified Models ........................................................................................................ 45

Table 6-4: Descriptive Statistics, New Variables .......................................................................... 48

Table 6-5: Final Poisson Regression Results .................................................................................. 48
EXECUTIVE SUMMARY

Transitways—premium transit corridors employing technologies such as Light Rail Transit (LRT) or Bus Rapid Transit (BRT)—offer an exceptional level of service for high-demand corridors in a transit system. In this sense, transitways can be understood as the arterials, so to speak, of a transit system, with local bus routes, bicycling and walking playing the roles of collectors and local streets. Much as an efficient road system depends on good local connections between sources of travel demand, major thoroughfares, and vice-versa, we hypothesize that the utility (and consequently ridership) of a transitway system depends on the quality of connections offered between homes, destinations and transitway stations.

The growing share of intermodal transport requires analysis of how passengers make choices when planning their routes. Choice models applied to revealed-preference transit passenger data from the Twin Cities show key components to user decisions regarding how to reach high-quality transit. Passenger paths reported in the Metropolitan Council’s 2016 Onboard Transit Survey combined with paths generated using Google Directions Application Programming Interface (API) form the choice set for analysis of transitway access mode. Multinomial logit structures are applied to the choice set. Distance and time factors are consistently significant, along with gender for biking, driving, and drop off and cost when paying transitway fare with a reloadable payment card. Implications include better understanding of user interaction with the transitway system of the Twin Cities and areas where better integration is needed.

To explore users’ choice of routes through the transit systems, schedule-based shortest path and multi-criterion shortest path algorithms are combined to investigate whether transit riders choose to take the shortest path between their origin and destination, a subjectively shortest path, or neither. A label setting algorithm is applied to transit schedule data along with five distinct time weighting schemes chosen to represent extreme attitudes toward trip attributes: high in-vehicle time, high waiting time, high walking time, and high transfer penalty. Results show a significant amount of overlap between paths generated for different weighting schemes, despite their extremity. This could indicate limited route options due to low redundancy and therefore low network resilience. Paths extracted from automated fare collection data align most closely with a baseline scenario with no extreme time weighting (i.e., the shortest travel time path). Future work will include investigation of intermediate weighting schemes and clustering analysis of passengers.

Existing research acknowledges the importance of multimodal connections to transitways, but mostly approaches the measurement of the quality of those connections from an overly simplistic perspective. This is particularly true of the measurement of the quality of local transit connections; common approaches to measurement include a simple count of connecting bus routes or numbers of trips per day.

To address these shortcomings of existing research, we develop an innovative, improved method for measuring the quality of bus, bicycle and pedestrian connections to transitways employing contemporary geospatial analysis techniques. We measure the quality of multimodal connections to
transitway stations as the area reachable by each mode in 15 minutes of travel on the actual street and/or transit network. Transit travel times include walking and waiting time. Bicycle and pedestrian travel times only consider local streets as defined by Census TIGER/Line data.

Based on these data and control variables, we estimate a set of ordinary least squares regression models, followed by a refined Poisson regression model to explain variation in average weekday boardings at transitway stations in 10 regions around the United States as a function of the areas reachable by bus, bicycle and walking in 15 minutes. Based on our results, for each additional square mile of area reachable by bus from transitway stations, we expect an increase of 20% in average weekday boardings at that station. Bicycle and pedestrian connections fail to produce the expected positive relationships with boardings, quite possible due to a lack of nationally consistent street network data incorporating presence of sidewalks and bicycle facilities, or other link-level bicycle and pedestrian friendliness factors. Figure E-1 shows model predictions of boardings at the station level for values spanning the range of areas reachable by bus in 15 minutes from studied stations.

Figure E-1: Model Predictions

The model results clearly show a strong relationship between the quality of bus connections provided at transitway stations and ridership at those stations. Further, the results also show the value of an approach to measuring the quality of those bus connections for ridership modeling that is more sophisticated than a simple count of routes or trips per day.

A valuable first step in attempting to improve transitway ridership via improved bus connections would be to calculate travel time-based bus travel sheds around stations and assess whether their relative sizes and the destinations they offer access to fit with the intended roles of the stations they serve in the transit system. This process would allow planners to systematically evaluate the degree of access a
transfer station actually offers to destinations it is intended to provide transfers to. Should any service improvements be deemed necessary to enhance access to transitway stations, service development planners could then evaluate the effects of multiple service improvement scenarios to both maximize areas reachable and ensure that specific desired destinations fall within them.

Data availability for transit analysis has made great strides in recent years—the General Transit Feed Specification (GTFS) data format in and of itself has been transformative—to the point that the in-vehicle leg(s) of a transit trip is well provided for. This research shows, however, that understanding the in-vehicle leg alone is insufficient for predicting transit use. This issue has implications even for bus access to transitways, as each bus trip to a transitway station has its own access mode, which current data do not allow sophisticated consideration of. Contemporary geospatial analysis techniques show great promise for improving the practice of transit planning by allowing planners to consider the aggregate impacts of a myriad of potential, complex multimodal trips. These analysis techniques, however, are only as good as the data they are based on. As such, systematically collecting and publishing comprehensive data on the pedestrian and bicycle systems must be considered a high priority for further improving our understanding of the ridership impacts of multimodal connections to transitways.
CHAPTER 1: INTRODUCTION

Transitways—premium transit corridors employing technologies such as Light Rail Transit (LRT) or Bus Rapid Transit (BRT)—offer an exceptional level of service for high-demand corridors in a transit system. “Exceptional” here refers both to the quality of service provided and the fact that such transit corridors are quite literally the exception to the rule; due to high implementation costs, transitways can never serve a majority of residents, workplaces or other destinations directly. Even accounting for the willingness of transitway users to walk distances of half a mile or more (Cao & Jordan, 2009), reasonable walk-sheds of transitway stations are unlikely to cover large portions of most regions in the foreseeable future. As such, while pedestrian access is crucial for their success, transitways must depend on a variety of access and egress modes to connect users with their trip origins and destinations. Understanding these access and egress modes requires us to explore how transit users choose access modes, how they choose their routes through the transit system itself and how the quality of multimodal connections available at stations affect transit use.

1.1 ACCESS MODE CHOICE MODELING

Significant investments into the transitway system of the Twin Cities in recent years motivate studies of how users interact with the system. Light rail, commuter rail, and bus rapid transit lines make up the transitway system in the Twin Cities. Plans currently in place show rapid expansion of the system with extensions to existing light rail lines and new arterial and dedicated guideway bus rapid transit lines. While some new lines will replace existing local bus service, others will lead to new right-of-way acquisition and potential for development around station areas. Land use around station areas has potential to shape the types of modes used to access the system.

This study seeks to create better understanding of how users access transitway stations. Data from a 2016 On-board transit survey shows which modes users take to get to transitway stations but does not explain how these choices are made. This analysis considers walking, regular transit, driving, drop off, biking, and taxi or transportation network companies (TNCs), such as Uber or Lyft, as potential access modes. Survey respondents use each of these modes in varying frequency. With new modes of shared transportation making multimodal trips increasingly available, better understanding of how to integrate the transit system is necessary.

1.2 ROUTE CHOICE MODELING

Varying attitudes and service levels can make a great impact on transit riders’ patterns. Understanding how public perceptions and transit service interact is key to developing an effective and useful transit assignment model. As their name suggests, choice models assume that users have choices and that the factors impacting these choices can be extracted from data. However, if a transit rider has only one path option, can any insight be made into his or her preferences?

As part of an ongoing study in multimodal transportation modeling in the Twin Cities, this paper investigates how Twin Cities transit riders interact with transit service, namely, how they choose paths.
Factors including waiting time, walking time, and transfers will be considered in evaluating how riders make choices about their paths. The goal of the analysis is greater understanding of limitations in the current Twin Cities transit service by determining whether transit riders choose the shortest path, a subjectively shortest path, or neither and whether or not they have options to do so. For example, if riders prefer not to make a transfer, do they have an alternative route that will enable this, or if one route is removed, is there an alternative available? We make the assumption that system resilience and ability of users to make choices regarding their routes are inherently connected. Thus, the questions to be addressed are: Do Twin Cities transit riders have options when choosing paths, and given a set of paths to choose between, how is the choice made?

To answer these questions, automated fare collection (AFC) data is compared to the outcomes of a multi-criterion schedule-based shortest path algorithm. Different weights are applied to different trip time components: walking, waiting, and transfer penalty in relation to in-vehicle time. If the system is resilient, one expects different subjectively “shortest” paths when trip time components are weighted to different extremes, representing passengers with different attitudes. The number of paths that coincide between survey data and shortest path results for each weighting scheme indicates attitudes of transit riders in the Twin Cities.

1.3 STATION-LEVEL RIDERSHIP MODELING

In this sense, transitways can be understood as the arterials, so to speak, of a transit system, with local bus routes, bicycling and walking playing the roles of collectors and local streets. Much as an efficient road system depends on good local connections between sources of travel demand, major thoroughfares, and vice-versa, we hypothesize that the utility (and consequently ridership) of a transitway system depends on the quality of connections offered between homes, destinations and transitway stations.

This is a simple hypothesis to propose, and frankly unsurprising on its face. Multimodal connections to transitways exist in a complex urban environment, however, and any sophisticated exploration of their relationship with transitway ridership must consider that complexity in detail. In addition, the state of multimodal connections at any given station is much more nuanced than a simple binary one of presence versus absence. For example, a transitway station also served by an hourly local bus route and one served by a mix of half a dozen local and limited stop routes running every quarter hour both have connecting bus service, but the level of service provided by their respective bus connections is by no means comparable.

Further, some bus connections with transitways take place in a high_amenity bus transit center integrated with the transit station, while others take place at a curbside stop several blocks away. Even given identical numbers of routes and frequencies of service, these two circumstances do not provide equivalent levels of service, nor do they provide equivalent travel times due to differing walking times between transitway and connection.
This need for nuanced analysis of connections’ quality is greater still for bicycle and pedestrian connections. While in theory both modes are available universally, the mere fact that it is possible to cycle or walk to a transitway station does not mean the trip will be an attractive, pleasant or even safe one. Even as network distance bike- and walk-sheds are becoming a standard practice in planning and evaluating transitway station areas, distances do not tell the whole story. Riding even a few hundred yards on a high-speed arterial may represent a much greater practical barrier to using a bicycle for transitway access than riding several miles on an off-street bicycle facility. Walking a quarter mile in the grass and gravel beside a suburban highway also represents a rather different experience from walking the same distance on quiet neighborhood streets with wide sidewalks.

Understanding the implications of differing mixtures and qualities of multimodal connections requires consideration of the wide variation in the degree to which they are provided for around transitway stations, as well as the surrounding environments in which they function. To that end, we explore the relationship between multimodal connections and transitway ridership through an analysis of transitway stations in 16 regions across the United States designated as peer regions to the Twin Cities by the Metropolitan Council. For each station in these systems, we model ridership as a function of area reachable from each station by bus, bicycle and on foot in 15 minutes, as well as control variables describing the regional accessibility provided by the transitway and the station area’s population, economy and built environment. For bicycling and walking, we adjust the time required to traverse links in the street system based on their functional class and literature on perceptions of time in different street environments. In other words, larger, busier streets are assigned a higher travel “cost” for cyclists and pedestrians than smaller, quieter ones.

1.4 REPORT ORGANIZATION

The following section reviews the planning literature on transitway access mode choice and route choice, as well as the relationship between multimodal connections and transitway ridership. Subsequent chapters introduce our study areas, as well as our measurement and modeling approaches, present the specifications and results of our models and offer policy recommendations for planning effective multimodal connections to transitways.
CHAPTER 2: LITERATURE REVIEW

All transit trips are inherently multimodal, requiring an access mode from trip origin to boarding stop and an egress mode from alighting stop to trip destination. Due to generally long stop spacings and a common role in transit systems focused primarily on rapid, regional mobility, this is especially true of transitway trips. This chapter reviews the planning literature related to the role multimodal connections to transitways in the broader transit system.

2.1 ACCESS MODE CHOICE

Numerous different analyses have been done to investigate how different transportation attributes impact users’ choices (or lack thereof) and thus inform policy. This section will discuss methodologies and their policy implications more broadly, then take a deeper look at the use of discrete choice modeling in transportation specifically.

Literature concludes that built-environment characteristics have a greater impact on travel behavior than social or personal attributes, though distinct population groups portray different transportation choices and behaviors (Lucas, Philips, Mulley, & Ma, 2018; Mejia-Dorantes, 2018; Ding, Cao, & Wang, 2018). In particular, when comparing the impacts of trip attributes between non-car and above-average mileage car users (where and when to go, cost, speed, reliability, environmentally friendly, physically active), attributes like speed and reliability impacted above-average car users more than other attributes. Policy implications of this include recommendations to make non-car modes more user-friendly (improvement to built environment) to encourage use rather than focusing on environmental impacts (personal convictions) (Hoffmann, Abraham, Skippon, & White, 2018).

In addition to impacting travel choices, built environment and transportation system characteristics can impact users’ travel satisfaction. De Vos investigates the effect of mode choice on travel satisfaction. Results show low satisfaction with public transit and high satisfaction with cycling. Dissatisfaction is highest when travelers do not get to take their preferred mode of transportation. Policy implications include making diverse modes available for travelers to increase overall satisfaction with transportation system. Integration of more desirable modes with public transit may increase satisfaction with the transit system as a whole (De Vos, 2018).

A robust amount of literature on application of discrete choice models to transportation questions such as route or mode choice exists. Discrete choice models can provide insight to the choices of users for whom transportation systems exist. Complexity in models increases with application to multi-modal transportation systems, however understanding how users make choices in these complex systems has become increasingly important in recent years. Hoogendoorn-Lanser describes the need to think about “modal merge as opposed to the traditional notion of modal split” in response to declining transit ridership and cost of private vehicle use (Hoogendoorn-Lanser & Bovy, 2005). Indeed choice models can help inform where to invest in the transportation system be it service integration, travel time, or passenger comfort (Guo & Wilson, 2011).
Numerous model structures exist to answer questions regarding choice, from complex to relatively simple. Recent literature describes use of a recursive structure in discrete choice modeling. Such a structure allows for separation of system aspects like transfers from environmental components like stops (Tien, Bastin, & Frejinger, 2017). Choice set estimation also incorporates recursion, allowing paths to not be considered in total as path enumeration adds significant complexity to the modeling process (Nassir, Hickman, & Ma, 2018). Some work takes a different strategy altogether by avoiding use of graph structures for choice modeling at all. A so-called “wayfinding” construct considers user cognition in complex spaces (like transit stations) rather than their potential paths (Ruetschi & Timpf, 2005). No consensus exists as to whether adding complexity to choice models significantly improves model performance, and more complex models often require more assumptions (Guo & Wilson, 2011; Bhat & Pulugurta, 1998).

The phenomenon of path overlap adds motivation for added complexity to transportation related choice models. Redundancy in transportation networks violates independence assumptions required for route choice modeling in particular and sometimes mode choice. Network GEV models account for path overlaps by establishing utility covariance (Papola & Marzano, 2013). Others, like C-logit models and path size correction models, incorporate measures that quantify overlap between paths (Cascetta, Nuzzolo, Russo, & Vitetta, 1996; Ben-Akiva & Bierlaire, 2000; Frappier, Morency, & Trepanier, 2018).

In addition to diversity in modeling structure, discrete choice modeling literature covers a wide variety of components influencing user choices, particularly when incorporating transit route choice. Most incorporate in-vehicle time, walking time, waiting time, and transfer penalties (Frappier, Morency, & Trepanier, 2018; Raveau, Guo, Munoz, & Wilson, 2014; Redman, Friman, Gaerling, & Hartig, 2013). Transfer penalties incur complexity in and of themselves as they represent “a purely psychological aspect ... affected by the transfer environment” (Guo & Wilson, 2011). More novel parameters of transit choice modeling include geometric distortion of route away from the shortest path between a user’s origin and destination (Frappier, Morency, & Trepanier, 2018; Raveau, Guo, Munoz, & Wilson, 2014). User perception parameters like time reliability, prior experience, security, and esthetics can also be of interest (Guirao, Garcia-Pastor, & Lopez-Lambas, 2016). Distinction is made between user characteristics (socio-demographics, habits) and contextual characteristics (trip purpose, weather, time of day) (Grison, Burkhardt, & Gyselinck, 2017; Dieleman, Dijst, & Burghouwt, 2002).

Little research has been done in the Twin Cities to investigate transit access mode choice. This analysis seeks to bridge that gap by adapting an established framework, multinomial logit model estimation, to answer questions about what impacts users’ choices to access high-quality transit options.

### 2.2 ROUTE CHOICE

A healthy amount of literature regarding multimodal route choice modelling exists (Bovy and Hoogendoorn-Lanser 2005, Hoogendoorn-Lanser et al. 2006, Bliemer et al. 2008, Arentze and Molin 2013, Anderson et al. 2014). Route choice modeling in multimodal networks introduces complexities beyond standard auto route choice modeling. Many different model structures, though generally of logit type, are proposed to account for these complexities in route choice. However, the components of each
model tend to be location-specific and dependent on available data. Studies vary in use of stated preference and revealed preference data (Anderson et al. 2014, Bliemer et al. 2008). Each of the papers cited above consider how utility of different trip components vary relative to one another.

An alternative to the traditional choice model, “A model and algorithm for multicriterion route-mode choice” is described by Dial (Dial 1979). The basis for the model in Dial’s example is defining all modes between an origin and destination by their respective cost and time. The set of “attractive” modes forms a convex hull in the plot of cost versus time. Declaring the choices in this way allows for variance in people’s value of time. People with a higher value of time will opt for faster but more expensive modes (like driving) while people with a lower value of time will opt for slower but less expensive modes (like walking) (Dial 1979). A similar concept is used here, however the choice set proposed for each passenger is multi-dimensional as opposed to two-dimensional with walking time, waiting time, transfers, and in-vehicle time as potentially separate dimensions. While Dial’s algorithm produces intermediate points in the cost versus time continuum, this analysis will focus on two extremes for each dimension.

Determining which aspects of a trip to quantify such as differences in preferences for walking, waiting, and transferring can be a challenge in choice modelling. While this paper does not address this idea explicitly by testing different choice parameter schemes, results do shed light onto which parameters tend to be more important. Hoogendoorn-Lanser et al. do take an in-depth look into different trip aspects in their 2006 paper, considering correlations between “parameter estimates for which attribute values are correlated” and “parameter estimates for which attribute values are not related” (Hoogendoorn-Lanser et al. 2006). These correlations might be between number of transfers and transfer time or streetcar and bus indicators, respectively. Differences between access, egress, and the actual transit trip for different modes as well as how to account for transfer penalty are considered (Hoogendoorn-Lanser et al. 2006). The Twin Cities’ transit network is less diverse than the network in the Netherlands used by Hoogendoorn-Lanser et al. (2006) with only two light rail lines, two bus rapid transit lines, and many standard bus lines, thus comparing differences and correlations across modes is not feasible.

In an evaluation of different choice models for auto trips, Zhu and Levinson aim to determine whether Twin Cities drivers choose the shortest travel time path. They find that forty percent of travelers take paths that overlap at least ninety percent with the shortest path between their origin and destination. They surmise that this number is high due to lack of alternatives for paths that are either very long or quite short (Zhu and Levinson 2015).

A lack of alternatives, or redundancy, is an indicator of a lack of network resilience. While there are different components of resilience (“redundancy, diversity, efficiency, autonomous components, strength adaptability, collaboration, mobility, safety, and the ability to recover quickly”) (Murray-Tuite 2006), we focus on the structure of a transit network and how it works with itself to serve passengers. An outcome of this study is measure of overlap between paths given different user choice preferences.
In the process of the literature review, no works combining the ideas of scheduled-based shortest path and multicriterion shortest path were found, and studies investigating whether people choose the shortest path focused on auto use, not transit. This analysis will bridge these gaps to provide an alternative to traditional choice modeling of understanding at transit user preferences.

2.3 STATION-LEVEL RIDERSHIP

If we understand travel as a derived demand, the usefulness of any system of transportation can be described in terms of the destinations it provides access to. Specifically in the case of transit, empirical evidence bears this out: research conducted by the University of Minnesota’s Accessibility Observatory finds that the number of jobs reachable by transit within a half hour’s travel from any given location explains roughly four fifths of transit’s commute mode share from that location (Owen & Levinson, 2015). While convenient pedestrian access and egress are rightly understood as critically important for effective transit, the small numbers of locations transitways in any given region are able to serve directly means much of the accessibility they provide depends on bus connections to their stations (Fan, Guthrie, & Levinson, 2011).

Looking beyond the travel possibilities represented by accessibility, the actual travel behavior of transitway users supports this contention: non-pedestrian access and egress modes play a considerably larger role in transitway trips than conventional, local bus trips. In addition, users appear to tolerate longer walks to and from transitway stations than they generally can be expected to for local bus stops (Cao & Jordan, 2009). This access/egress pattern represents a departure from traditional local transit practice dating back to the streetcar era: providing a differentiated service that provides rapid travel between stations while requiring significant travel to reach stations versus providing uniform, slow service between stops while minimizing distances to origins and destinations (Cervero, 1998; Warner, 1978).

2.3.1 Bus Connections to Transitway Ridership

A significant body of literature describes the strength of relationships between bus connections and transitway ridership (Durning & Townsend, 2015; Kuby, Barranda, & Upchurch, 2004; Upchurch & Kuby, 2014). Much of this literature employs a Direct Ridership Modeling (DRM) approach to estimate ridership at the station level in a one-step multiple regression analysis, as opposed to the four-step models more commonly used at the regional scale. DRMs do not seek to determine overall trip origins, destinations or routes, merely the number of people who pass through transitway stations. As such, they can be sensitized to a wide variety of local factors such as bus connections (Cervero, 2006). In general, however, this existing body of research is primarily focused on building comprehensive understanding of the determinants of transitway ridership—or on identifying the most important determinants—rather than exploring the importance and effects of specific ridership determinants such as bus connections (Durning & Townsend, 2015; Kuby et al., 2004).

Given that the accessibility provided by transit strongly predicts transit ridership (Owen & Levinson, 2015; Renne, Hamidi, & Ewing, 2016) and that bus access/egress legs are crucial to the regional
accessibility transitways provide (Fan et al., 2011), it is reasonable to expect that higher levels of local bus accessibility in station areas will translate to higher ridership at the stations they surround (Chandra, Bari, Devarasetty, & Vadali, 2013). Accessibility, however, includes factors beyond transit agencies’ control, most importantly local land use. While accessibility is a function of both mobility and destination density, recent research indicates local land use is actually more important than travel speed in determining accessibility in the context of actual conditions found in 21st century cities (Levine, Grengs, Shen, & Shen, 2012).

In general urban planning terms, accessibility measures offer significant advantages over mobility measures because they directly measure what one can accomplish by traveling a given amount of time, as opposed to merely how much distance one can cover in that time, as well as how they highlight the importance of integrated transportation and land-use planning (Levinson & Krizek, 2005). In the specific case of guiding transit agencies in providing effective local transit connections to transitways, however, mobility given existing land use may allow for more direct planning and action.

2.3.2 Non-Motorized Access to Transitways

Traditionally, access by non-motorized modes—particularly pedestrian access—is seen as central to successful transit service. The long-recognized importance of density in transit planning stems from the fact that the denser the station area, the more people and destinations there will be within walking distance of it (Cervero, 2002; Pushkarev & Zupan, 1977a). Indeed, in many ways the increasingly prominent transit-oriented development movement can be seen as an attempt to massively increase the quantity of residents and destinations within an easy walk of transit (Cervero & Kockelman, 1997; A. Guthrie & Fan, 2016). It is important to consider what makes a walk (or bicycle ride) easy, however? Answering this question requires consideration of how best to measure the quality of bus, bicycle and pedestrian connections to transitways, as well as acknowledging that “convenience” for bicycle and pedestrian transportation involves factors beyond distance and travel time, such as exposure to traffic and perceived safety (Tilahun, Thakuriah, Li, & Keita, 2016).

The walkability of station area street networks can be measured using pedestrian environment audits (Tucker, Ostwald, Chalup, & Marshall, 2005; Werner, Brown, & Gallimore, 2010). The labor-intensive nature of the pedestrian environment audit process, however, often leads walkability-focused ridership research to employ measures which can be generated from existing, publicly available data.

These measures frequently include factors such as intersection density (Renne et al., 2016), street network connectivity (Asadi-Shekari, Moeinaddini, & Zaly Shah, 2013) and sidewalk completeness (Woldeamanuel & Kent, 2015). The former two consider the area practically reachable on foot in a reasonable distance, while the latter considers the safety and pleasantness of walking that distance. None of these traditional measures, however, directly consider the area one can reach in a reasonable, pleasant walk.
2.3.3 Summary

Convenient access and egress legs have long been understood as important for promoting transit use. However, the complexity of transit trips including an access leg on the street network, one or more transit legs tied to schedules and stop locations, followed by an egress leg on the street network have traditionally led to the use of simplified proxy measures for access/egress convenience. These measures include residential and/or employment density within some pre-determined, Euclidean distance of stations (Pushkarev & Zupan, 1977b) and numbers of bus routes connecting with stations (Kuby et al., 2004). Measures such as these approximate the convenience of transitway access and egress legs, but contain simplifying assumptions which significantly reduce their precision. For example, while station area density is a function of the number of people living and/or working within walking distance of stations, the definition of station areas based on Euclidean distance fails to account for the significant variation in areas reachable within any given actual walking distance with different street networks. Measuring bus connections in terms of numbers of routes ignores variation in both frequency and speed of those routes. Recent advances in Geographic Information Systems (GIS) and computing technology (Abdullahi & Pradhan, 2017) as well as the development and broad adoption of the standardized General Transit Feed Specification (GTFS) format for transit route and schedule data (Andrew Guthrie, Fan, & Das, 2017) enable a more sophisticated approach to measuring the convenience of transitway access and egress. The following chapter describes the development of such a method for this research.
CHAPTER 3: METHODS

This chapter describes the research approach employed for each of the three primary tasks. It is organized by individual task due to the variety of data sets and analysis approaches in this research.

3.1 ACCESS MODE CHOICE MODELING APPROACH

Data for this analysis comes from the 2016 On Board Transit Survey commissioned by the Metropolitan Council, the Metropolitan Planning Organization in the Twin Cities. The original survey contains over 30,000 records of surveyed passengers. Passengers’ trip information includes origin, destination, purpose, hour of the day of the trip, routes taken, and access mode. Figure 3-1 shows transit way access mode shares from the survey. The most detailed trip information is listed for the route on which a passenger was surveyed including a unique stop identifier. Other trip segments do not include stop information (Metropolitan Council, 2017).

To ensure correct information for modeling, records were selected in which the route surveyed was a transitway and also the first transitway a passenger used in their trip, about 6,000 records. Doing so ensured correct calculation of distance, time, and cost of access to the transitway station where the passenger boarded.

Google Directions API was used to calculate time, and distance of different access modes (Google, 2018). The user enters origin, destination, departure time, and mode for a desired trip, and the API returns distance and time. While the API provided trip time information for walking, transit, driving, drop off, and biking modes, taxi or transportation network company travel distance required estimation. Because wait times vary for these services, a random amount of time between five and ten minutes was added to drive time for this mode option.
Transitway Access Mode Share

Figure 3-1: Transitway access mode shares from the 2016 On Board transit survey of the Twin Cities Metropolitan Area

A more refined version of the Directions API was used for transit access, as specific route information is required for cost calculation. For example, express routes have higher cost than regular routes. Although attempts were made to match transit routes to routes users actually reported using prior to boarding the transitways, lack of specific time information in the survey made this difficult. The On Board Survey only provided an hour window in which the passenger departed. Additionally, the Directions API does not allow historical queries, and transit schedule data has most certainly changed since the survey was administered, requiring the assumption that routes and transit travel times have not changed significantly since fall 2016. Figure 3-2 shows the distribution of transitway access times by mode.
Cost calculation required additional assumptions. Driving cost estimation assumed AAA’s average cost per mile, 0.564 cents, which includes operating costs such as gas maintenance, plus ownership costs including depreciation and insurance (AAA, 2017). Transit fare for different times of day and different service type dictated transit costs. Taxi or transportation network company costs are based on the pricing structure of Uber. The cost includes a booking fee ($2.50), per minute fee ($0.20), and per mile fee ($1.50), all of which are variable. If the cost of the trip is under the current minimum fee ($8.50 but also variable), the minimum fee will be used (Uber, 2018).

Table 3-1: List of variables used for model estimation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Duration of time to access a transitway using regular transit</td>
<td>Mean</td>
</tr>
<tr>
<td>Time&lt;sub&gt;Transit&lt;/sub&gt;</td>
<td>9.7</td>
<td>11.8</td>
</tr>
<tr>
<td>Time&lt;sub&gt;Drive&lt;/sub&gt;</td>
<td>Duration of time to access a transitway using a personal vehicle</td>
<td>4.7</td>
</tr>
<tr>
<td>Time&lt;sup&gt;Walk&lt;/sup&gt;</td>
<td>Duration of time to access a transitway by walking</td>
<td>21.8</td>
</tr>
<tr>
<td>Time&lt;sup&gt;Bike&lt;/sup&gt;</td>
<td>Duration of time to access a transitway by biking</td>
<td>7.9</td>
</tr>
<tr>
<td>Time&lt;sup&gt;TNC&lt;/sup&gt;</td>
<td>Duration of time to access a transitway by TNC</td>
<td>12.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cost</th>
<th>[U.S. Dollars]</th>
<th>[U.S. Dollars]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost&lt;sup&gt;Transit&lt;/sup&gt;</td>
<td>Cost of transit fare for given route and time of day in USD</td>
<td>$2.09</td>
</tr>
<tr>
<td>Cost&lt;sup&gt;Drive&lt;/sup&gt;</td>
<td>Cost to drive a personal vehicle to access transitway</td>
<td>$0.77</td>
</tr>
<tr>
<td>Cost&lt;sup&gt;TNC&lt;/sup&gt;</td>
<td>Cost of hiring a TNC operator to provide transit access</td>
<td>$9.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distance</th>
<th>[miles]</th>
<th>[miles]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance&lt;sup&gt;Transit&lt;/sup&gt;</td>
<td>Length of transit access distance in miles</td>
<td>1.3</td>
</tr>
<tr>
<td>Distance&lt;sup&gt;Drive&lt;/sup&gt;</td>
<td>Length of auto access distance in miles</td>
<td>1.4</td>
</tr>
<tr>
<td>Distance&lt;sup&gt;Walk&lt;/sup&gt;</td>
<td>Length of walking access distance in miles</td>
<td>1.1</td>
</tr>
<tr>
<td>Distance&lt;sup&gt;Bike&lt;/sup&gt;</td>
<td>Length of biking access distance in miles</td>
<td>1.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>User Characteristics</th>
<th>Count of 1</th>
<th>Count of 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1 if user is female, 0 otherwise</td>
<td>2746</td>
</tr>
<tr>
<td>Go to Card</td>
<td>1 if user paid transitway fare with a Go to Card, 0 otherwise</td>
<td>2295</td>
</tr>
<tr>
<td>Total Records</td>
<td>6393</td>
<td></td>
</tr>
</tbody>
</table>

3.1.1 Model Specification

The literature review described numerous types of predictors used in transportation choice modeling. This analysis incorporates both trip-dependent parameters and user-specific qualities in a multinomial logit model structure. Because this analysis seeks information about access to transitways, regular transit access is used as the reference category. The model to be fitted is:
where

\[ \logit(\pi^j) = \log \frac{\pi^j}{\pi^{\text{Transit}}} = \frac{\pi^{jT}}{\pi^{\text{Transit}}} \beta^j \text{ for } j = 2 \text{ to } 6 \]

(Dobson & Barnett, 2008)

Where

\( \pi^j \) = probability of mode j (drive, drop off, bike, walk, or taxi)

\( \pi^{\text{Transit}} \) = probability of taking transit

\( \pi^{jT} \) = vector of independent variables in consideration (travel time, distance, cost, etc.)

\( \beta^j \) = vector of model coefficients to be estimated, each corresponding to an independent variable

Three different structures of linear predictors \((\pi^{jT} \beta^j)\), referred to as “utility equations” in subsequent sections, are tested in this analysis: a base model, an interactive term model, and a relaxed base model. The base model specification contains uniform time, cost, and distance coefficients for each mode. Bike and walk modes are assumed to have no cost. The interactive model specification also contains uniform time and distance coefficients for each mode. Numerous parameters were tested in this specification (income, parking availability, gender, and transitway payment type among others), with only significant parameters included in the final specification: costs of passengers using a “GotoCard” (refillable automatic payment card) and gender for driving, drop off, and biking. Because policy implications motivate this work, differences in user perception of time and distance between modes is of interest. Therefore, the “relaxed” model structure allows for differences in coefficients for time, cost, and distance between modes.

3.1.2 Model Estimation

Biogeme software, an open source maximum likelihood estimation software, was used for model estimation. Pythonbiogeme files for discrete choice model estimation were modified to desired utility equation structures (Bielaire, 2016).

3.2 ROUTE CHOICE MODELING

This analysis seeks to apply a schedule-based shortest path algorithm to a multicriterion route choice algorithm to obtain a more holistic idea of how diverse passengers choose paths through a multimodal network. We will answer the questions: how many people take the shortest path or the subjectively shortest path, and do different preferences for route attributes lead to significantly different subjectively shortest paths?

3.2.1 Multicriterion Shortest Path

Using multiple criteria to determine the “subjectively” shortest path acknowledges different individuals’ preferences for route aspects. While Dial gave the example of cost and time (thus selecting different
paths based on different values of time), this analysis goes beyond two dimensions to see how preferences for in-vehicle time, waiting time, walking time, and transfers change transit riders’ paths. Consider the two-dimensional example in Figure 3-3.

![Figure 3-3: Example of a network for bicriterion analysis. Paths include travel time and travel cost components](image)

A traveler has a choice between two paths: an express path with a toll or a regular path with no toll. Using econometric models, we know that if the traveler’s value of time is greater than $10 per hour, the traveler will take the path 1 as the total perceived cost will be less than that of path 2. Dial’s algorithm finds these value of time “break points” where travelers opt for an alternative path with different travel time and cost between an origin and destination due to having a different personal value of time.

Plotting a point for each path with time on the abscissa and monetary cost on the ordinate, a convex hull of the shortest path for each value of time emerges (Figure 3-2a). Different modes are shown for illustrative purposes. A traveler would likely never consider walking and flying for the same trip, but the two options illustrate the idea of natural extreme points.

This study will consider more than two path dimensions as shown in Figure 3-2b. Figure 3-2b shows three of the four dimensions to be incorporated: in-vehicle time, walking time, and number of transfers. Like in the illustration in Figure 3-2a, each point in the Figure 3-2b represents a path between a specific origin and destination. Rather than value of time being the difference between points, personal perception of one of the three aspects relative to the other aspects is what differs. For example, the red point represents the path of a person with a high aversion to walking relative to taking transfers and in-vehicle time. The blue point is a path of a person with high aversion to in-vehicle time while the green point is for high aversion to transfers.
Figure 3-4: Paths plotted by cost and time for bicriterion shortest path (left) and paths plotted by in-vehicle time, number of transfers, and walking time (right)

In this study we will not determine all of the intermediate gray points, but rather will generate the paths represented by the blue, red, and green points, plus another point to represent paths for people with a high aversion to waiting and a point where a person considers all trip aspects equally.

3.2.2 Schedule-Based Shortest Path

This study employs a schedule-based shortest path algorithm to capture the intricacies of transit trips such as precise waiting and walking times for transit vehicles and distance to and from stops, respectively. Consider the example time versus space diagram of three potential paths between an origin and a destination (Figure 3-5). A passenger could choose to leave immediately, but would have to take one walking transfer and one waiting transfer. The next option is to wait a few minutes, have shorter transit time, and no transfers, but a later arrival time. The final option is to wait longer and transfer once. Incorporating the exact transit schedule as opposed to a frequency based algorithm ensures exact paths are generated, thus allowing for algorithm output to be compared to automated fare collection data in a precise way.
The algorithm we use allows the user to input different weights for time spent in-vehicle, walking, and waiting as well as a separate transfer penalty. Changing the weights of each factor is critical for the multicriterion shortest path concept, and by extension, so is having precise timing information for each path.

In order to apply a shortest path algorithm, transit schedule data (General Transit Feed Specification, GTFS, data) is transformed into a set of time dependent links and nodes. Individual transit trips, different vehicles that run the same route but depart at different times, make up a portion of the links with each link having specific departure and arrival times from a node. Transfers, access, and egress make up the remaining links in the network. These links are based on scheduled transit trips as well, but their time “cost” is computed based on straight-line distance between points: origin location and transit stop, transit stop and transit stop, or transit stop and destination location. Transfers can be either waiting transfers or walking transfers. The nodes in the network are transit stops with the added dimension of time. A single physical location will have multiple nodes for every time a transit vehicle makes a stop.

3.2.3 Algorithm

A label setting algorithm is developed for this analysis using the time dependent network described in the previous section. The following logic describes how the algorithm is used in the larger scope of the shortest path assignment process.

\[ N \]  
Set of all nodes

\[ n_i \]  
Node i
Step 1: Create N and M. Read in entire transit network (from GTFS) plus walking and waiting links and passenger demand.

Step 2: Create access and egress links for a specific individual, add to M.

Step 3: Initialize $l_i = \infty$, $p_i = \text{NULL}$. Label the $l_{\text{origin}} = 0$ and $p_{\text{origin}} = -1$. Initialize $L = \{N\}$ and $\bar{L} = \{\}$. 

Step 4: Select the node $n_i$ with the minimum $l_i$ as the current node. Remove $l_i$ from $L$ and add it to $\bar{L}$. For all links extending from the $n_i$, if $l_k > w_j t_{ik} + l_i$, $l_k = w_j t_{ik} + l_i$ and $p_k = n_i$.

Step 5: If $L = {}$ or $\bar{L} = \{N\}$, continue to step 6. Otherwise, repeat step 4.

Step 6: Beginning at the $n_{\text{destination}}$, record $p_i$, then trace back to $n_{\text{origin}}$ based on each node’s $p_i$. This set of nodes is the shortest path between the origin and destination. Store this path.

Step 7: Remove the individual passenger’s access and egress links from M.

Step 8: Return to step 2 and repeat process for all passengers.

When considering a transfer penalty ($T$) as opposed to the parametric weighting factors ($w_j$), the algorithm gains an extra complexity in step 4. Rather than simply considering the time it takes to traverse the links between in-vehicle links, the transfer penalty is added to that time. $T$

### 3.2.4 Illustrative Example

Consider the example shown in Figure 3-6 and Table 3-2. A passenger has three parallel paths to choose from. Table 3-2 shows the time required to traverse each link, sorted by type. In the left column of Figure 3-6, all link weights are equal at 1.0, whereas in the right column of Figure 3-6, walking links are weighted to show a strong aversion to walking relative to waiting and in-vehicle time (column 4 in Table 3-2).
Table 3-2: Example link travel times corresponding with Figure 4. Values for In-Vehicle and Waiting times are constant across scenarios. Walking values are used on the left side of Figure 4 and Weighted Walking values are used on the right side.

<table>
<thead>
<tr>
<th>In-Vehicle</th>
<th>Walking</th>
<th>Waiting</th>
<th>Weighted Walking, ( w_{walk} = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link Name</td>
<td>Time, ( t_{ik} ) (minutes)</td>
<td>Link Name</td>
<td>Time, ( t_{ik} ) (minutes)</td>
</tr>
<tr>
<td>V1</td>
<td>20</td>
<td>W1</td>
<td>5</td>
</tr>
<tr>
<td>V2</td>
<td>20</td>
<td>W2</td>
<td>10</td>
</tr>
<tr>
<td>V3</td>
<td>50</td>
<td>W3</td>
<td>5</td>
</tr>
<tr>
<td>V4</td>
<td>20</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>V5</td>
<td>20</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 3-6: Example of the shortest path algorithm implementation. Both left and right begin with the top initialization. The left is a baseline scenario with no time weighting. The right side weights walking time as three times worse than in-vehicle or waiting time.
The first row of Figure 3-6 shows the initialization stage of the shortest path algorithm: all nodes have initial labels and predecessors (pred). In the second row, labeling is partially done and differences between weighting schemes appear on node 5. With original weighting, node five has a label of 5 minutes while with the new weighting scheme node 5 has a label of 15 minutes. The final labels are shown in the third row of Figure 3-6. Of greatest importance are the differences between the labels and predecessors of node 9. Initially, l9 = 55 minutes while p9 = 5. After a different weighting scheme is introduced, l9 = 65 minutes and its p9 = 4. This simple example shows how different passenger preferences can impact perceived travel time to alter the subjectively shortest path.

3.2.5 Data

Data availability has been a significant influencer of methodology, with best available data being AFC data from a special fare card available to University of Minnesota Twin Cities students, known as the U-Pass. U-Pass data has been taken from its raw form of unlinked transit rides and transformed via a trip chaining algorithm to create full passenger trips. Trip chaining work has been conducted independently from this analysis (Kumar et al. 2018).

An origin destination matrix of U-Pass riders has been created from the trip chained U-Pass data and is used as an input for this study. Due to privacy concerns, the home addresses of U-Pass holders have not been provided, thus individuals’ actual origins and destinations have not been inferred. Therefore, this study cannot infer rider preferences regarding access and egress times.

To complement the AFC data, general transit feed specification (GTFS) data has been acquired for the same time period, March 2016, from Metro Transit. GTFS data provides the schedule to create the network described in the Algorithm section.

3.2.6 Procedure

The schedule-based shortest path algorithm was run for five scenarios with different trip aspect weighting schemes. Each scenario is described in Table 3-3.
Table 3-3: Weighting scheme and description of each test scenario.

<table>
<thead>
<tr>
<th>Base Scenario</th>
<th>In-Vehicle Scenario</th>
<th>Walking Scenario</th>
<th>Waiting Scenario</th>
<th>Transfer Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Weights (In-Vehicle, Walking, Waiting, Transfer Penalty)**

| 1.0, 1.0, 1.0, 0 | 50.0, 1.0, 1.0, 0 | 1.0, 50.0, 1.0, 0 | 1.0, 1.0, 50.0, 0 | 1.0, 1.0, 1.0, 300 |

**Descriptions**

- Hypothetical passengers consider all aspects equally
- Hypothetical passengers prefer to be in transit vehicles as little as possible
- Hypothetical passengers prefer to walk as little as possible
- Hypothetical passengers prefer to wait as little as possible
- Hypothetical passengers avoid making transfers as much as possible

The output of the shortest path algorithm is the shortest path for each individual passenger with a preferred departure time from the origin.

3.3 STATION LEVEL RIDERSHIP MODELING

This research focuses on transitways as the highest level of hierarchical 21st century transit systems. As such, we focus on new start systems that provide service clearly differentiated both in terms of service type and role within the broader transit system from surrounding bus service. Specifically, we include Heavy Rail Transit (HRT), Light Rail Transit (LRT) and dedicated guideway Bus Rapid Transit (BRT). Rapid bus stations are not included as observations due to a similar role in transit systems to local bus service, though rapid bus routes are included to calculate areas reachable by bus from stations.

To produce results valuable to planners in the Twin Cities metropolitan area, we focus on ten HRT, LRT and BRT systems operating in regions identified by the Metropolitan Council as peer regions for transit system performance or regional transit investment level comparisons (Metropolitan Council, 2010) for which ridership data were available either publicly or by correspondence. Table 3-4 shows the systems and regions included, as well as the number of stations in each.

Table 3-4: Regions and Systems Included

<table>
<thead>
<tr>
<th>Region</th>
<th>System</th>
<th>Mode</th>
<th>Stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Francisco Bay</td>
<td>BART</td>
<td>HRT</td>
<td>24</td>
</tr>
<tr>
<td>Dallas-Fort Worth</td>
<td>DART</td>
<td>LRT</td>
<td>48</td>
</tr>
<tr>
<td>City</td>
<td>Agency</td>
<td>Mode</td>
<td>Area</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Denver</td>
<td>RTD</td>
<td>LRT</td>
<td>32</td>
</tr>
<tr>
<td>Houston</td>
<td>Metro</td>
<td>LRT</td>
<td>24</td>
</tr>
<tr>
<td>Twin Cities</td>
<td>MVTA</td>
<td>BRT</td>
<td>4</td>
</tr>
<tr>
<td>Twin Cities</td>
<td>Metro Transit</td>
<td>LRT</td>
<td>33</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>PAT</td>
<td>BRT</td>
<td>27</td>
</tr>
<tr>
<td>Portland</td>
<td>TriMet</td>
<td>LRT</td>
<td>72</td>
</tr>
<tr>
<td>Seattle</td>
<td>Sound Transit</td>
<td>LRT</td>
<td>11</td>
</tr>
<tr>
<td>San Jose</td>
<td>VTA</td>
<td>LRT</td>
<td>44</td>
</tr>
<tr>
<td>Phoenix</td>
<td>Valley Metro</td>
<td>LRT</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td>341</td>
</tr>
</tbody>
</table>

### 3.3.1 Measuring Multimodal Connections to Stations

This research adopts a multimodal access/egress mobility approach to measuring the quality of bus, bicycle and pedestrian connections to transitway stations. To directly measure the quality of the connections themselves, while allowing us to hold the influence of local land use constant, we compute the precise geographic area reachable by each mode in 15 minutes travel time from each station, while separately considering local density and regional accessibility. We use 15 minutes as our cutoff time for access/egress travel as a compromise between allowing for bus travel waiting time and restricting non-motorized access to a plausible distance, equating to no more than three-quarters of a mile network distance for walking (Cao & Jordan, 2009), assuming an average speed of three miles per hour, and two and one-half miles for cycling (Hochmair, 2015), assuming an average speed of ten miles per hour.

Areas reachable by bus vary considerably as a function of service levels provided. We base these areas on General Transit Feed Specification (GTFS) data published by transit agencies, in all cases using the most recent data available. We remove transitways from each agency’s GTFS feed, allowing us to calculate the area reachable from each station by transit without using the transitway. The corresponds to the area for which a local transit access/egress leg of a transitway trip is feasible.

Areas reachable by bicycling and walking vary as well based on the street network. In addition, as mentioned in the previous chapter, bicycle and walking access is affected by a number of factors beyond distance. Measuring pedestrian and bicycle friendliness of individual links in the street systems at a national scale is difficult, however, due to a lack of data on sidewalks and bicycle facilities in the nationally-available (and comparable) geospatial street data provided by the Census Bureau.

The OpenStreetMap open-source mapping project offers both sidewalk and bicycle facility data. These data, however, are not available in all regions and have questionable data quality in regions where they are available (Hochmair, Zielstra, & Neis, 2013). (A spot-check of the Twin Cities region, well-known to
the authors, confirms this incompleteness.) Due to the unfeasibility of manually checking every street and path in every station area considered, we base bicycle and pedestrian access on areas reachable without using either primary or secondary roads in the Census Bureau’s TIGER/Line road data. While imperfect, this approach at least screens out roads we can be reasonably sure are unpleasant and/or unsafe for walking or bicycling.

Figure 3-7: Example of Areas Reachable by Bus in 15 Minutes
Figure 3-8: Example of Areas Reachable by Bicycle in 15 Minutes
Figure 3-9: Example of Areas Reachable on Foot in 15 Minutes
As an example, Figures 3-7, 3-8 and 3-9 show the areas reachable within 15 minutes by bus, bicycle and walking from part of the Twin Cities transitway system. The variability of the size of areas reachable by bus is immediately apparent, underscoring differences in levels of bus service provided at different stations. Much less variation appears in areas reachable by bicycle or pedestrian travel, the former in particular. Areas reachable by bicycle are consistently the largest—larger even than bus, likely due to a lack of waiting time and ability to travel at a similar average speed to a local bus in any direction allowed by the street network. The lack of variation observable here in the bicycle and pedestrian areas may arise in part from the relatively regular street network present in the urban area shown, and in part from a relative lack of nuance in the street data employed.

3.3.2 Variables and Model Specification

The most critical variables in this research are average weekday boardings and the sizes of areas reachable by bus, bicycle and walking in 15 minutes of travel from stations. Some understanding of the relationships between these variables is important for determining a modeling approach. Figures 3-10, 3-11 and 3-12 present scatter plots with boardings on the Y axis and the area reachable by one of the three modes on the X axis.
Though a small number of high-ridership outlier stations are present, Figure 3-10 does appear to show the expected positive relationship between boardings and the quality of bus connections. The relationship also appears relatively linear, suggesting a linear regression technique for modeling the data.

![Figure 3-11: Comparison of Boardings and Areas Reachable by Bicycle in 15 Minutes](image)

The relationship appears somewhat less clear in Figures 3-11 and 3-12, though at first glance, the basic positive relationship between boardings and areas reachable appears to hold. The apparent relative weakness of the relationship may result from the previously-mentioned lesser degree of variation in the size of areas reachable on streets designated as bikeable or walkable due to lack of detailed national bicycle and pedestrian system data.

Based on these observations, we estimated an initial ordinary least squares regression model of the data including the following variables:

- **Boardings**—The number of people boarding at each transitway station on an average weekday. (Response variable.)
- **Area Reachable by Bus**—The area, in square miles, reachable in 15 minutes’ travel by bus from the station. (Key explanatory variable.)
Figure 3-12: Comparison of Boardings and Areas Reachable on Foot in 15 Minutes

- **Area Reachable by Bicycle**—The area, in square miles, reachable in 15 minutes’ travel by bicycle from the station. (Key explanatory variable.)
- **Area Reachable on Foot**—The area, in square miles, reachable in 15 minutes’ travel on foot from the station. (Key explanatory variable.)
- **Park & Ride**—Dummy variable identifying stations with park-and-ride facilities. (Control variable.)
- **Heavy Rail**—Dummy variable identifying heavy rail rapid transit stations. (Control variable.)
- **Bus Rapid Transit**—Dummy variable identifying dedicated guideway bus rapid transit stations. Light rail is omitted as the reference category. (Control variable.)
- **Job Accessibility**—The count of jobs reachable from the census block containing the station in 30 minutes’ transit travel. (Control variable.)
- **Population Density**—The density (in people per square mile) of population in the census tract containing the station. (Control variable.)
- **Miles from Central Station**—The Euclidean distance, in miles, from the station to the station closest to the heart of the region’s central business district.
- **Median Household Income**—The median household income, in 2016 dollars, of the census tract containing the station.
Early model runs also included variables describing the racial mix, age distribution and household composition of station area populations. None of these variables was ever significant, and model goodness of fit improved on their removal. As a result, in the following chapter, we report the results of the initial OLS model and several simplifications of it. We also report a refined, Poisson regression model developed from the initial model to improve consideration of waiting time for bus connections and station area population.
CHAPTER 4: MODE CHOICE MODELING RESULTS

Table 4-1: Model estimation results for base model. Parameter values normalized to the time coefficient, B_TIME

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Parameters</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>4802</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial Log-likelihood</td>
<td>-8145.749</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Log-likelihood</td>
<td>-4152.942</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rho squared</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rho squared bar</td>
<td>0.489</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
<th>Normalized Value*</th>
<th>Standard Error</th>
<th>t-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike Constant</td>
<td>-2.07</td>
<td>39.429</td>
<td>0.0957</td>
<td>-21.6</td>
<td>0</td>
</tr>
<tr>
<td>Drive Constant</td>
<td>-1.79</td>
<td>34.095</td>
<td>0.0823</td>
<td>-21.72</td>
<td>0</td>
</tr>
<tr>
<td>Drop Off Constant</td>
<td>-2.78</td>
<td>52.952</td>
<td>0.101</td>
<td>-27.46</td>
<td>0</td>
</tr>
<tr>
<td>Taxi/TNC Constant</td>
<td>-4.72</td>
<td>89.905</td>
<td>0.262</td>
<td>-18.06</td>
<td>0</td>
</tr>
<tr>
<td>Walk Constant</td>
<td>1.3</td>
<td>-24.762</td>
<td>0.0392</td>
<td>33.25</td>
<td>0</td>
</tr>
<tr>
<td>Cost</td>
<td>0</td>
<td>0.000</td>
<td>0.00157</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.372</td>
<td>7.086</td>
<td>0.0518</td>
<td>-7.17</td>
<td>0</td>
</tr>
<tr>
<td>Time</td>
<td>-0.0525</td>
<td>1.000</td>
<td>0.00389</td>
<td>-13.5</td>
<td>0</td>
</tr>
</tbody>
</table>

*Values normalized to B_TIME

Table 4-1 shows estimation results of the base model including cost, distance, and time, normalized to the time coefficient. Beyond the cost coefficient, all estimated coefficients significantly differ from zero. Walking has the highest baseline utility followed by transit, driving, biking, drop off, and taxis and transportation network companies. The value of the walk coefficient interprets as walking access to transit removes about twenty-five perceived minutes of time from the trip. Likewise, adding a mile of distance adds a perceived seven minutes to the trip.

The base model has the lowest final log-likelihood value and rho squared values of any model. Log-likelihood measures how well estimated coefficients predict the given data (Ben-Akiva & Lerman, 1985).
Biogeme seeks to maximize log-likelihood in its estimation process. Therefore, the higher the value, the better the model predictability. Rho squared values can be interpreted similarly to an $R^2$ value of normal linear models.

Table 4-2: Model estimation results for interactive term model. Parameter values normalized to the time coefficient, $B_{\text{TIME}}$

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
<th>Normalized Value*</th>
<th>Standard Error</th>
<th>t-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike Constant</td>
<td>-1.85</td>
<td>1.985</td>
<td>0.109</td>
<td>-17.02</td>
<td>0</td>
</tr>
<tr>
<td>Gender/Bike Constant</td>
<td>-0.853</td>
<td>0.915</td>
<td>2.06E-01</td>
<td>-4.14</td>
<td>0</td>
</tr>
<tr>
<td>Drive Constant</td>
<td>-1.95</td>
<td>2.092</td>
<td>0.102</td>
<td>-19.22</td>
<td>0</td>
</tr>
<tr>
<td>Gender/Drive Constant</td>
<td>0.271</td>
<td>-0.291</td>
<td>1.29E-01</td>
<td>2.1</td>
<td>0.04</td>
</tr>
<tr>
<td>Drop Off Constant</td>
<td>-2.98</td>
<td>3.197</td>
<td>0.13</td>
<td>-22.92</td>
<td>0</td>
</tr>
<tr>
<td>Gender/Drop Off Constant</td>
<td>0.34</td>
<td>-0.365</td>
<td>1.76E-01</td>
<td>1.94</td>
<td>0.05</td>
</tr>
<tr>
<td>Taxi/TNC Constant</td>
<td>-4.52</td>
<td>4.850</td>
<td>0.264</td>
<td>-17.12</td>
<td>0</td>
</tr>
<tr>
<td>Walk Constant</td>
<td>1.26</td>
<td>-1.352</td>
<td>0.0412</td>
<td>30.47</td>
<td>0</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.386</td>
<td>0.414</td>
<td>5.28E-02</td>
<td>-7.3</td>
<td>0</td>
</tr>
<tr>
<td>GoToCard/Cost</td>
<td>-0.0854</td>
<td>0.092</td>
<td>2.43E-02</td>
<td>-3.51</td>
<td>0</td>
</tr>
<tr>
<td>Time</td>
<td>-0.932</td>
<td>1.000</td>
<td>6.55E-02</td>
<td>-14.21</td>
<td>0</td>
</tr>
</tbody>
</table>

*Values normalized to $B_{\text{TIME}}$

Table 4-2 shows estimation results of the interactive term model. Although tests used many different combinations of interactive terms, few added significantly to the model. Most results match those of the base model with a few exceptions. The baseline utility of bike access to transit increased to surpass that
of driving. However, both of these modes have significant gender parameters. The bike gender parameter translates to an additional perceived minute of travel time for females using bikes while the driving gender parameter indicates a reduction of about a third of a minute of travel time for females who drive. Females who get dropped off also see improvement in utility.

The Go to Card parameter has some nuance to interpretation as it represents the interaction between perceived access cost and transit payment type. The Go to Card parameter is non-zero for users who reported paying for transitway service with a Go to Card. Users paying with Go to Cards must have previously purchased a card and loaded money onto it. Amount of money

Table 4-3: Model estimation results for relaxed model. Parameter values normalized to transit time coefficient, B_TIME_TRANSIT

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
<th>Normalized Value*</th>
<th>Standard Error</th>
<th>t-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike Constant</td>
<td>-3.73</td>
<td>53.1339</td>
<td>0.27</td>
<td>-13.85</td>
<td>0</td>
</tr>
<tr>
<td>Drive Constant</td>
<td>-3.37</td>
<td>48.0057</td>
<td>0.275</td>
<td>-12.28</td>
<td>0</td>
</tr>
<tr>
<td>Drop Off Constant</td>
<td>-4.39</td>
<td>62.5356</td>
<td>0.293</td>
<td>-15</td>
<td>0</td>
</tr>
<tr>
<td>Taxi/TNC Constant</td>
<td>-7.55</td>
<td>107.55</td>
<td>2.2</td>
<td>-3.43</td>
<td>0</td>
</tr>
<tr>
<td>Walk Constant</td>
<td>-0.153</td>
<td>2.17949</td>
<td>0.221</td>
<td>-0.69</td>
<td>0.49</td>
</tr>
<tr>
<td>Driving Cost</td>
<td>-69.5</td>
<td>990.028</td>
<td>23.7</td>
<td>-2.93</td>
<td>0</td>
</tr>
<tr>
<td>Drop Off Cost</td>
<td>-90.4</td>
<td>1287.75</td>
<td>31.4</td>
<td>-2.88</td>
<td>0</td>
</tr>
<tr>
<td>TNC Off Cost</td>
<td>0.262</td>
<td>-3.7322</td>
<td>0.211</td>
<td>1.24</td>
<td>0.22</td>
</tr>
<tr>
<td>Transit Cost</td>
<td>-0.557</td>
<td>7.93447</td>
<td>0.107</td>
<td>-5.2</td>
<td>0</td>
</tr>
<tr>
<td>Bike Distance</td>
<td>1.84</td>
<td>-26.211</td>
<td>0.508</td>
<td>3.63</td>
<td>0</td>
</tr>
<tr>
<td>Metric</td>
<td>Coefficient</td>
<td>Standard Error</td>
<td>t-Statistic</td>
<td>P-Value</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------------</td>
<td>-------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Drive Distance</td>
<td>39</td>
<td>-555.56</td>
<td>13.4</td>
<td>2.92</td>
<td>0</td>
</tr>
<tr>
<td>Drop Off Distance</td>
<td>50.7</td>
<td>-722.22</td>
<td>17.7</td>
<td>2.86</td>
<td>0</td>
</tr>
<tr>
<td>TNC Distance</td>
<td>-0.665</td>
<td>9.47293</td>
<td>0.399</td>
<td>-1.67</td>
<td>0.1</td>
</tr>
<tr>
<td>Transit Distance</td>
<td>-0.215</td>
<td>3.06268</td>
<td>0.0539</td>
<td>-4</td>
<td>0</td>
</tr>
<tr>
<td>Walk Distance</td>
<td>-3.06</td>
<td>43.5897</td>
<td>1.17</td>
<td>-2.61</td>
<td>0.01</td>
</tr>
<tr>
<td>Bike Time</td>
<td>-0.28</td>
<td>3.9886</td>
<td>0.0818</td>
<td>-3.42</td>
<td>0</td>
</tr>
<tr>
<td>Drive Time</td>
<td>-0.0379</td>
<td>0.53989</td>
<td>0.0338</td>
<td>-1.12</td>
<td>0.26</td>
</tr>
<tr>
<td>Drop Off Time</td>
<td>0.0102</td>
<td>-0.1453</td>
<td>0.0398</td>
<td>0.26</td>
<td>0.8</td>
</tr>
<tr>
<td>Taxi/TNC Time</td>
<td>-0.0608</td>
<td>0.8661</td>
<td>0.0993</td>
<td>-0.61</td>
<td>0.54</td>
</tr>
<tr>
<td>Transit Time</td>
<td>-0.0702</td>
<td>1</td>
<td>0.00803</td>
<td>-8.74</td>
<td>0</td>
</tr>
<tr>
<td>Walk Time</td>
<td>0.101</td>
<td>-1.4387</td>
<td>0.0571</td>
<td>1.77</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*Values normalized to B_TIME_TRANSIT

available on the card reduces each time the card owner makes a trip and pays with the card. Users see money deduction immediately when they use their card. The value for the Go to Card parameter indicates that users with Go to Cards perceive the cost of their access to transitways negatively, equivalent to adding a fraction of a minute of perceived travel time. Although not all users have paid the cost of access with a Go to Card (only those using regular transit may have), the Go to Card parameter is still significant.

Table 4-3 shows estimation results of the relaxed base model. This model allows for differences in user perception of access cost, distance, and time. All parameters are normalized to the transit time coefficient. This model shows transit to have the highest baseline utility of any access mode, though the walking coefficient does not differ significantly from zero and can be considered equal to the baseline utility of transit access. Neither of the on demand, motorized access modes, driving and drop off have time as a significant predictor, likely attributable to the correlations between cost, distance, and time. Distance travelled by these modes contributes positively to their utilities, with a mile of driving taking the equivalent of 555 minutes of perceived transit travel time. Conversely, cost has a significantly negative impact on driving and drop off utilities, exceeding the impact of distance. Transit results remain consistent with other models. Transit time, distance, and cost contribute negatively to its utility.

### 4.1 MODEL VALIDATION

To better quantify how well each model fits the original survey data, Table 4-4 shows results of Pearson chi-squared goodness of fit tests for each model. The following formula calculates the chi-squared statistic, \( \chi^2 \):
\[ X^2 = \sum_{i=1}^{6} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i} \]

Where \( i \) represents different modes. To determine the observed mode generated by each model, Biogeme simulated results using data excluded from model estimation and calculated probabilities for each mode for each passenger. The highest probability mode for each passenger is considered the “observed” mode. The access mode reported in survey data is the “expected” mode. For each access mode \( i \), the number of passengers expected and the number of passengers observed are counted. These values are used in the equation above to compute the test statistic.

The null hypothesis of the Pearson chi-squared test is that the model fits the data well with the alternative being that the model does not fit well. The test statistic must be compared to the chi-squared distribution with given degrees of freedom to either reject or fail to reject the null hypothesis. Each model has degrees of freedom equal to number of observations minus parameters estimated, shown in Table 4-4. Table 4-4 also contains computed \( X^2 \) test statistics and resulting probabilities, or p-values, that the null hypothesis is true. The p-values calculated are extremely small, so the null hypothesis is rejected with the conclusion that the models do not fit the data well.

Table 4-4: Results of Pearson chi-squared goodness of fit test for all models

<table>
<thead>
<tr>
<th>Model</th>
<th>Base</th>
<th>Interactive</th>
<th>Relaxed</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X^2 )</td>
<td>611.0</td>
<td>5419.8</td>
<td>670.7</td>
</tr>
<tr>
<td>DOF</td>
<td>2049</td>
<td>2046</td>
<td>2036</td>
</tr>
<tr>
<td>p-value</td>
<td>1.66E-228</td>
<td>3.68E-303</td>
<td>6.62E-197</td>
</tr>
</tbody>
</table>

Figure 4-1 provides a graphical representation of each models’ predictive ability. A deeper look into the data shows that different modes have different fit. For the base and relaxed models, walking access has a high success rate of prediction and driving has a moderate success rate while other modes have a poor success rate. The interactive model switches the success rates of walking and driving and shows small improvement in prediction of biking access.
Figure 4-1: Comparison of model predictions of access mode to mode reported in survey data
CHAPTER 5: ROUTE CHOICE MODELING RESULTS

The network analyzed contains 490,272 nodes and 3,034,213 links which can be broken down as follows:

- 1,599 zones
- 13,701 stops
- 9,068 trips
- 490,272 nodes
- 481,204 transit links
- 470,139 waiting transfer links
- 2,082,870 walking transfer links

The computational time ranges from 1 hour, 45 minutes to 3 hours, 20 minutes for 1,093 passengers on a local server with Xeon E5-2650 2.20GHz CPU, 64 GB of RAM, and 64 bit operating system.

In running the algorithm for the Twin Cities transit network, paths were not found for all passengers. Table 5-1 shows the number of passengers assigned to paths for each scenario. Reasons for lack of found paths are believed to stem from selection of various parameters and some possible data errors. Access and egress walking distances were limited to 0.1 miles because passengers’ start and end locations are taken from their first and last AFC tag. Walking transfers are also limited to 0.1 miles, and waiting transfers are limited to 20 minutes. Distances are all Euclidean, not network based.

Table 5-1: Number of passengers assigned to paths for each weighting scenario out of 1,093

<table>
<thead>
<tr>
<th>Base Scenario</th>
<th>In-Vehicle Scenario</th>
<th>Walking Scenario</th>
<th>Waiting Scenario</th>
<th>Transfer Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>373</td>
<td>372</td>
<td>375</td>
<td>376</td>
</tr>
<tr>
<td>Passengers with paths generated</td>
<td>373</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To understand how results from the scheduled-based shortest path algorithm compare to reality, each unlinked trip made by each passenger for each scenario was compared to AFC data. Figure 5-1 shows the result. Note that to be considered “a match,” the route number, transit trip identifier, or stop identifier must be the same as the AFC data for the same sequence number. For example, if in the AFC data, Passenger A takes routes 2, 6, then 10, but in the schedule-based shortest path output takes routes 2, 10, then 6, only one of Passenger A’s unlinked trips match. Such a specific distinction of unlinked trip number is not made for other results, yet is included here for completeness. Because of
the highly specific nature, the low matching rates are not wholly surprising, particularly for matching transit trip identifiers, as they are both spatially and temporally specific, unlike routes and stops.

Though no clear winner emerges among scenarios for best matching rates, the high walking time scenario appears in the top two best in each case. The caveat to be noted with the performance of the high walking time scenario is the inaccuracy of origin and destination locations. Due to the inability to infer actual origins and destinations, walking times for access and egress are artificially short. Thus, the tendency of the high walking time scenario to match well may be more a result of a passenger’s origin being directly adjacent to the stop chosen in reality, increasing the likelihood of also selecting the real route and trip. Weighting the walking time of the trip indirectly coerces the passenger to go directly to the closest transit stop which happens to be the stop chosen in reality.

Despite four of the scenarios being similar in terms of matching reality of AFC data, one scenario, the high waiting time scenario, emerges as a clear loser. This could be indicative of relative ambivalence to waiting for transit.

![Figure 5-1: Unlinked trip comparison between AFC data and schedule-based shortest paths for different weighting scenarios](image)

Figure 5-1 indicates that different paths were generated for each weighting scenario, but does not really say the extent to which paths differ, or anything about the resilience of the network. Figure 5-2 lists overlap between paths generated for different scenarios for two different cases in an effort to shed light on the network’s resilience. The first is route overlap. When considering the paths taken by the same passenger for two scenarios, the route overlap is the percentage of routes taken that are the same. Route overlap is shown because it gives a good sense of the options available to the transit user. Though
a passenger may opt to take a different trip to obtain a better transfer, that passenger’s options may still be limited to a single route, which would not be reflected in the second overlap measure.

5.1 EXAMPLE OF ROUTE OVERLAP COMPUTATION:

Passenger A, High Transfer Penalty take routes 902, and 2

Passenger A, High In-Vehicle takes routes 16, 2, 61

Overlap = 1 route overlapping / 4 total routes = 25%

The second overlap measure compares all “links” shared across scenarios, excluding access and egress links. A link connects two nodes in the network be it two transit stops via transit vehicle or walking, or the same transit stop at different time points via waiting. Link overlap more accurately reflects how similar two paths are than route overlap.

As expected, the values for route overlap are greater than those for link overlap with the highest overlap for both cases being between the base scenario and the high transfer scenario. This shows that despite a strong preference for avoiding transfers, passengers likely still are forced to take them. Indeed, half of the route overlap values are over 50%, indicating that passengers are limited in their transit route choices, and that the transit network lacks route redundancies.

High in-vehicle time and high waiting time share the least in terms of routes and links. It seems logical that in order to avoid time in-vehicle a passenger would wait longer for a faster one, or vice versa.

In addition to network resilience a goal of this analysis is to look into transit user preferences. To do so, we assess which scenario best explains passengers’ paths found in the AFC data. Route overlap is used as a metric for comparison (Figure 5-3). For each passenger, the share of route overlap between AFC and algorithm output for each weighting scenario was computed. Across the five scenario overlaps, the highest overlap was selected for each passenger and awarded one “point.” If two scenarios were tied, each was given half a point. The total points for each weighting scenario were added and divided by the total overall.
Figure 5-2: Route overlap (top) and link overlap (bottom) between paths generated by each of the five weighting schemes.
Figure 5-3 shows that in general, Twin Cities transit users do not have extreme opinions regarding in-vehicle time, waiting time, walking time, or transfers as the base scenario is most like 95% of transit trips. Implementation of an extreme transfer penalty accounts for the second greatest share at 3.14%. Thus, we conclude that transit riders’ paths can generally be best explained as taking the shortest path, not a subjectively shortest path. However, just 1.5% of passengers’ paths can be fully explained by any of the extreme scenarios presented. The remaining passenger paths would require intermediate points using a multicriterion shortest path algorithm.

An alternative way of looking at how well different scenarios explain real passenger behavior is shown in Figure 5-4. Figure 5-4 is a histogram of percent route overlap for each weighting scenario. The greatest number of algorithm-generated paths have between 0 and 25% route overlap with AFC data for all scenarios. The transfer penalty scenario has the most passengers with overlap between 75 and 100%, but not significantly more than other scenarios.
Figure 5-4: Histogram of passengers with certain percentages of route overlap between AFC data and the five weighting scenarios.
Table 6-1 presents descriptive statistics for the variables included in the analysis. It supports the earlier conclusion of greater variation being present in the areas reachable by bus than by other modes, given the large range and standard deviation of the variable relative to its mean. While the range of areas reachable by bicycle is large as well, its standard deviation is much smaller relative to its mean. Not surprisingly, given the similar calculation process, areas reachable on foot follow a similar pattern.

Table 6-1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Weekday Boardings</td>
<td>2,399.82</td>
<td>4,699.70</td>
<td>3.00</td>
<td>43,956.00</td>
</tr>
<tr>
<td>Area Reachable by Bus</td>
<td>1.74</td>
<td>0.91</td>
<td>0.22</td>
<td>6.02</td>
</tr>
<tr>
<td>Area Reachable by Bicycle</td>
<td>12.21</td>
<td>1.79</td>
<td>2.51</td>
<td>15.18</td>
</tr>
<tr>
<td>Area Reachable on Foot</td>
<td>1.01</td>
<td>0.20</td>
<td>0.24</td>
<td>1.32</td>
</tr>
<tr>
<td>Park &amp; Ride</td>
<td>0.33</td>
<td>0.47</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Heavy Rail</td>
<td>0.07</td>
<td>0.26</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Bus Rapid Transit</td>
<td>0.09</td>
<td>0.29</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Job Accessibility (x 1,000)</td>
<td>146.70</td>
<td>132.39</td>
<td>3.45</td>
<td>694.96</td>
</tr>
<tr>
<td>Population Density</td>
<td>6,657.39</td>
<td>5,616.23</td>
<td>34.90</td>
<td>40,811.12</td>
</tr>
<tr>
<td>Miles from Central Station</td>
<td>9.81</td>
<td>13.28</td>
<td>0.00</td>
<td>50.75</td>
</tr>
<tr>
<td>Median Household Income</td>
<td>58,530.42</td>
<td>31,058.21</td>
<td>11,410.00</td>
<td>162,690.00</td>
</tr>
</tbody>
</table>

6.1 INITIAL MODELS

Table 6-2 presents the results of the initial linear regression models estimated to explain variation in average weekday boardings at the station level as a function of multimodal connections and control variables. All models include 341 observations. The initial model includes all service areas, as well as the park-and-ride status of stations. The model produces an adjusted $R^2$ value of 0.59, indicating that it explains 59% of the observed variation in transitway boardings. The area reachable by bus is highly significant, with the expected positive relationship to boardings. Though also significant, area reachable by bicycle, however, has an unanticipated negative sign. Area reachable on foot—intended to correspond to the ease of pedestrian access—is unexpectedly insignificant. The definitions of these last two variables are both based on the local street network, unavoidably introducing some degree of correlation between them. Indeed, they produce a bivariate correlation coefficient of 0.69, which,
though not high enough to represent a definite problem, indicates the possibility of multicollinearity in models including both variables.

Table 6-2: Initial Model Results

<table>
<thead>
<tr>
<th></th>
<th>Bus, Bike, Ped, P&amp;R</th>
<th>Bus, Bike</th>
<th>Bus, Ped</th>
<th>Bus, P&amp;R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adjusted R²</strong></td>
<td>0.59</td>
<td>0.59</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Area Reachable by Bus</strong></td>
<td>643.87 ***</td>
<td>620.62 ***</td>
<td>714.60 ***</td>
<td>546.88 **</td>
</tr>
<tr>
<td><strong>Area Reachable by Bicycle</strong></td>
<td>-326.09 **</td>
<td>-351.92 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Area Reachable on Foot</strong></td>
<td>-429.58</td>
<td>-2511.77 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Park &amp; Ride</strong></td>
<td>-100.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heavy Rail</strong></td>
<td>7508.58 ***</td>
<td>7432.93 ***</td>
<td>7880.68 ***</td>
<td>8218.57 ***</td>
</tr>
<tr>
<td><strong>Bus Rapid Transit</strong></td>
<td>-736.64</td>
<td>-711.53</td>
<td>-708.54</td>
<td>-547.83</td>
</tr>
<tr>
<td><strong>Job Accessibility (x 1,000)</strong></td>
<td>10.31 ***</td>
<td>10.35 ***</td>
<td>9.87 ***</td>
<td>8.71 ***</td>
</tr>
<tr>
<td><strong>Population Density</strong></td>
<td>0.10 ***</td>
<td>0.10 ***</td>
<td>0.10 **</td>
<td>0.09 **</td>
</tr>
<tr>
<td><strong>Miles from Central Station</strong></td>
<td>-32.33 *</td>
<td>-33.13 **</td>
<td>-30.91 *</td>
<td>-35.25 **</td>
</tr>
<tr>
<td><strong>Median Household Income</strong></td>
<td>0.02 ***</td>
<td>0.02 ***</td>
<td>0.02 ***</td>
<td>0.03 ***</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>1977.62</td>
<td>1845.32</td>
<td>-5.01</td>
<td>-2190.80 ***</td>
</tr>
</tbody>
</table>

Legend: *p<0.1, **p<0.05, ***p<0.01

To evaluate what effect, if any, is produced by including all variables in the same model, we also estimated models with area reachable by bus and one other station access variable, shown in the right three columns of Table 6-2. Omitting the two insignificant variables from the original model yields very little change in the coefficients, signs or p-values of the remaining variables. Goodness of fit is unchanged, with an R² value of 0.59, just as in the initial model.

Retaining the variable describing area reachable on foot while eliminating area reachable by bicycle and park-and-ride yields a stronger positive, significant relationship between area reachable by bus and boardings. Area reachable on foot is also now significant, but with a counter-intuitive negative sign. Other variables and goodness of model fit once again show relatively little change.

Finally, retaining only area reachable by bus and park-and-ride shows a slightly weaker positive relationship between bus access and boardings. Park-and-ride is still insignificant. Again, other variables’ results are remarkably similar, indicating a stable model apart from the unexpected behavior of the bicycle and pedestrian access variables.
6.2 SIMPLIFIED MODELS

The initial models show an empirical relationship between bicycle and pedestrian access and transitway boardings that flies in the face of well-established theory regarding transit demand—that ease of non-motorized access ought to yield higher ridership. The empirical relationship exists, for this specific set of stations, under the definition of “bicycle and pedestrian access” used in this research. As related in the previous chapter, however, that definition is not ideal due to data limitations. It is possible that area reachable in 15 minutes’ travel without use of primary or secondary roads is simply not an adequate measure of the convenience, attractiveness or safety of non-motorized travel. In all honesty, it appears more likely this is the case than that our basic understanding of the direction of the relationship between bicycle and pedestrian friendliness and transit use is wrong.

In addition, the consistent insignificance of the park-and-ride variable casts doubt on its usefulness as a predictor of transitway ridership in the specific circumstances of this research. Again, we do not believe this is because park-and-ride facilities do not contribute ridership to transitways, or that otherwise similar stations without commuter parking would not have lower ridership. It is more likely that the presence of park-and-ride facilities is associated with other factors which reduce ridership—such as commonly suburban locations, for instance. Even so, the variable contributes little to our understanding of the importance of multimodal connections to transitway ridership.

For these reasons, to remove the risk of multicollinearity and simplify the interpretation of the results, we estimate a further model employing area reachable by bus as the only access variable, shown in Table 6-3. Once again, the remaining variables and adjusted $R^2$ change remarkably little, indicating a stable, robust model of the data. In the first simplified model (shown in the column labeled “Bus”), all variables except the dummy variable identifying bus rapid transit stations are now significant at at least the 5% level. Since this variable has never achieved statistical significance in any model variation (indicating, all else equal, no difference between station-level light rail and dedicated guideway bus rapid transit boardings), we proceeded to estimate a final model excluding the BRT dummy variable. This model is shown in the right column of Table 6-3. Further interpretation is based on this model.

Table 6-3: Simplified Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bus</th>
<th>No BRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted $R^2$</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>Area Reachable by Bus</td>
<td>546.88 **</td>
<td>579.45 ***</td>
</tr>
<tr>
<td>Heavy Rail</td>
<td>8216.37 ***</td>
<td>8236.09 ***</td>
</tr>
<tr>
<td>Bus Rapid Transit</td>
<td>-547.55</td>
<td></td>
</tr>
<tr>
<td>Job Accessibility</td>
<td>8.72 ***</td>
<td>8.80 ***</td>
</tr>
<tr>
<td>Population Density</td>
<td>0.09 **</td>
<td>0.09 **</td>
</tr>
</tbody>
</table>
The most striking result is the effect size and high significance of the area reachable by bus variable. The coefficient it produces indicates that, all else equal, for each additional square mile of area reachable within 15 minutes bus travel from stations we expect an additional 579 average weekday boardings. Given that the range of the area variable extends from 0.2 square miles to 6 square miles, variations in bus access to transitway stations can plausibly account for significant differences in ridership, especially given that these results refer to individual stations, not entire lines.

The importance of job accessibility in explaining station-level boardings is notable as well—predicting roughly 9 additional boardings for each additional thousand accessible jobs. Given that the blocks containing studied stations offer accessibility ranging from roughly 3,500 jobs to nearly 700,000, the practical significance of job accessibility in explaining transitway boardings is also considerable.

Other variables behave much as expected, with heavy rail transit and station area population density associated with more transitway boardings and distance from the central station of the system associated with fewer. The exception is that median household income has a positive relationship with boardings, despite generally higher rates of transit use among low-income individuals. This effect may result from higher employment and/or labor force participation in higher-income areas, rail lines better serving higher-wage commutes, some combination of the two and/or other factors.

6.2.1 Model Predictions

To demonstrate the predictions produced by the final model more comprehensively than is possible by simply interpreting the coefficients of individual variables, Figure 6-1 shows average weekday boardings as predicted by the models at various values of area reachable by bus. All other continuous variables are held at their mean values; dummy variables are held at mode.
The graph shows model predictions at values of area reachable by bus roughly spanning the range of the variable. At an area reachable by bus of zero, the model predicts 972 boardings on an average weekday. This increases by 579 boardings for every additional square mile of area reachable by bus, translating to 4,449 with six square miles reachable by bus.

### 6.3 FINAL MODEL REFINEMENTS

The linear models detailed above describe the relationship between area reachable by bus and transitway boardings while controlling for population density near the station. In order to offer a more intuitive interpretation of the three-way relationship between areas reachable by bus, boardings and the density of those areas, we estimated a final model replacing population density in the census tract containing the station with the aggregated population of all census block groups which intersect the area reachable by bus from the station within 15 minutes. In addition, to account more directly for connecting bus service frequency as well as travel speed, we recalculated our bus connection area variable as the mean area reachable at five minute intervals in the 8:00 am to 9:00 am hour. (Population with bus access is thus also the mean population with access in that hour.)

Though the original linear model became unstable with the introduction of the new population variable, the use of a Poisson regression model solved this issue. In addition, the Poisson model including mean area reachable and mean population with access proved able to generate significant results for modal and demographic control variables excluded from earlier models. Table 6-4 presents descriptive statistics for the new variables included in the final model. Though mean area reachable by bus is broadly similar to the simple 8:00 am area reachable variable in the initial models, it has a somewhat smaller range and standard deviation, suggesting using the mean area reachable during the 8:00 am to 9:00 am hour reduces the influence of waiting times based on an arbitrary departure time. Population with bus access to stations varies significantly between observations, underscoring the range of
connection area sizes and the range of their densities. The percentage of residents ages 18 to 39 in the tract containing the station are included as a measure of the prevalence of age groups most likely to use transit. Percentage of tract residents of color is included both as a predictor (people of color are generally more likely to use transit than non-Hispanic whites) and a measure of transitway impacts (as desirable transit service—indicated by high ridership—may correlate with neighborhood social change and displacement).

Table 6-4: Descriptive Statistics, New Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Area Reachable by Bus</td>
<td>1.789</td>
<td>0.786</td>
<td>0.235</td>
<td>4.742</td>
</tr>
<tr>
<td>Mean Population with Bus Access to Station</td>
<td>33.182</td>
<td>24.739</td>
<td>1.506</td>
<td>196.158</td>
</tr>
<tr>
<td>% Residents ages 18-39</td>
<td>0.421</td>
<td>0.138</td>
<td>0.130</td>
<td>0.931</td>
</tr>
<tr>
<td>% People of Color</td>
<td>0.507</td>
<td>0.241</td>
<td>0</td>
<td>0.991</td>
</tr>
</tbody>
</table>

Table 6-5 presents the results of the final regression model. As a non-linear model, Poisson regression produces coefficients which cannot be interpreted as directly as linear regression. To aid interpretation, therefore we also report Incidence Rate Ratios (IRRs) for all explanatory variables. An IRR sets a coefficient of 0 (i.e. no statistical relationship) equal to 1, and represents the percentage change expected in the response variable for one unit of change in the explanatory variable. For example, if an explanatory variable with a unit of miles returns an IRR of 1.1, we expect a 10% increase in the response variable for 1 mile increase in that explanatory variable. If that same variable were to return an IRR of 0.9, however, we would expect a 10% decrease in the response variable for 1 mile increase in that explanatory variable.

Table 6-5: Final Poisson Regression Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>IRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Area Reachable by Bus</td>
<td>0.182196 ***</td>
<td>1.199850</td>
</tr>
<tr>
<td>HRT</td>
<td>1.418263 ***</td>
<td>4.129942</td>
</tr>
<tr>
<td>BRT</td>
<td>-1.178974 ***</td>
<td>0.307594</td>
</tr>
</tbody>
</table>
All variables are highly significant. The IRR of 1.19985 for Mean Area Reachable by Bus indicates that for every additional square mile of area reachable on average at weekday peak service, we expect roughly a 20% increase in transitway boardings. Heavy rail stations have roughly four times as many boardings as light rail stations, while BRT stations have nearly one third as many. Stations with park-and-ride facilities have roughly 20% more boardings than otherwise similar stations without. Other variables all show the expected relationship with boards, with the exceptions of median income, as in the original models, and % People of Color.

### 6.3.1 Final Model Predictions

Figure 6-2 shows predicted boardings for various values of Mean Area Reachable by Bus. We again hold all explanatory variables except Area Reachable by Bus at their mean or modal values, meaning this graph represents a light rail station without a park-and-ride, with mean values for all other explanatory variables. Rising from just over 1,000 for a station with no bus connections to nearly 2,700 for a station with bus connections reaching an average area of five square miles within 15 minutes, the final model predicts less variation in boardings as a function of bus connections that the initial models. The relationship still has genuine practical significance, however, and it must be remembered that the area reachable has been effectively disaggregated into its physical area (representing service quality) and the people within that area.
To show how station level boarding change based on the population in the area reachable by bus, Figure 6-3 shows final model predictions at values of Mean Population With Bus Access ranging from 0 to 100,000. (This does not represent the full range of the variable; high-end outliers are excluded to show detail in the range including most stations.) Again, all other variables are held at the mean or modal values. Beginning at roughly 1,200 boardings for a station with no population with bus access, boardings rise to around 2,100 for a station with 100,000 people able to gain access within a 15 minute bus ride.
Figure 6-3: Final Model Predictions, Population With Bus Access
CHAPTER 7: CONCLUSIONS

This report estimated transitway access mode choice models using data from a transit onboard survey distributed in the Twin Cities. Mode choices considered include walking, regular bus, driving, drop off, biking, and taxi or transportation network company. A multinomial logit structure models passenger choice with time, distance, cost, and gender considered as predictors. Results show that distance and time of access mode significantly influence mode choice, as expected. A relaxed model allowing different modes to have different coefficients for time, distance, and cost showed different modes to have significantly different parameters estimated, indicating different user perceptions. A model with interacting terms showed gender to have a significant impact on certain modes, namely biking and driving. The model estimated female users to have negative perceptions to biking and positive perceptions to driving, possibly indicating barriers to bike usage when compared to other modes. Future policies should be considered to help eliminate this gender bias.

Walking access represents the overwhelming majority of transitway users. Whether by inherent choice or due to built-environment characteristics cannot be said. Nevertheless, priority ought to be taken in facilitating walkability of transitway stations. Relatively low utility of biking access indicates room for improvement of bike access to transitway stations. Improved bike infrastructure (lanes and storage space) and access to bikes are necessary to increase utility and use of bikes as a transitway access mode. Bike access to transitways should be reevaluated with the implementation of dockless bikeshare in the Twin Cities.

The option to drive presents a limitation of this work. Only a small percentage of transitway stations have official park and rides, but certainly lack of an official park and ride need not preclude users from driving to other locations then transferring to the transitway system. The idea of “hide and ride” as the behavior is sometimes referenced proved challenging to model given variable costs of parking and unknown distances users would need to walk between parking and transitway stations. Considering hide-and-ride behavior in future station area plans might still be prudent, though future analysis on the behavior should be done first.

Future plans for this work include estimation of egress mode choice models to see if egress behaviors are different than access behaviors and time of day evaluation. People doing shift work or who have non-traditional work hours probably interact differently with transit than those travelling in regular AM and PM peak hours. More complicated nested models might also be considered to further understand differences between the impacts of personal characteristics and built environment attributes.

The goals of the route choice analysis were to determine whether Twin Cities transit users take the shortest path, a subjectively shortest path, or neither, and if they have different options. This idea was tested by using Twin Cities transit AFC data in a schedule-based shortest path algorithm with five different weighting schemes: a base scenario, high transfer penalty, and high waiting, walking, and in-vehicle times. Results show that while the shortest path between two points does not exactly match a path chosen by the user (Figure 5-1), users do take paths that are most like the shortest path versus a
subjectively shortest path generated using an artificial, extreme weighting scheme (Figure 5-3). There are, however, shares of passengers that align most closely with paths generated using an extreme transfer penalty, a high walking penalty, and a high waiting penalty (Figure 5-4). Unsurprisingly, we infer that there are transit riders who avoid taking transfers and waiting and walking as much as possible.

A secondary goal of this analysis is to establish whether transit riders have choices within the Twin Cities network to accommodate their preferences (Figure 5-2). While paths generated for different weighting schemes are not identical, they do share a considerable amount of overlap (minimum 19% for link overlap, 39% for route overlap). This shows that passengers are limited in their options and can’t necessarily find paths to match their unique preferences. In a world of many transportation choices, this is a serious downfall of transit. Lack of choices also stands to be an issue when viewed through the lens of resilience. If one route is critical for all a passenger’s paths, minor delays or outages of that route become highly problematic.

These results establish a foundation for future work investigating preferences of Twin Cities transit riders. Implementation of Dial’s algorithm for route choice to find intermediate points of trip component preference would be illuminating, as would clustering analysis of passengers to highlight distinct transit use patterns. Eventually a route choice model will be generated for the Twin Cities.

The ridership model results clearly show a strong relationship between the quality of bus connections provided at transitway stations and ridership at those stations. Further, the results also show the value of an approach to measuring the quality of those bus connections for ridership modeling that is more sophisticated than a simple count of routes or trips per day. By directly measuring the area it is possible to reach by a brief bus trip from transitway stations while controlling for surrounding density, regional job accessibility at the station and local social conditions, our models are able to directly evaluate the relationship between local bus mobility and ridership.

This ability has important implications for the practice of transit planning. Using the GTFS data format and industry-standard GIS software, it is a relatively simple matter to calculate the actual mobility provided by bus routes serving or passing near a transitway station. In addition, it is also possible to edit GTFS data for purposes of assessing the mobility impacts of hypothetical service changes at stations where greater local bus mobility is found to be desirable (Andrew Guthrie et al., 2017).

A valuable first step in attempting to improve transitway ridership via improved bus connections would be to calculate travel time-based bus travel sheds around stations and assess whether their relative sizes and the destinations they offer access to fit with the intended roles of the stations they serve in the transit system. This process would allow planners to systematically evaluate the degree of access a transfer station actually offers to destinations it is intended to provide transfers to. Should any service improvements be deemed necessary to enhance access to transitway stations, service development planners could then evaluate the effects of multiple service improvement scenarios to both maximize areas reachable and ensure specific desired destinations fall within them.

It is important to note, however, that our results do not necessarily mean large increases in bus service at all transitway stations would lead to corresponding increases in transitway ridership. Bus service is
not randomly distributed; on the contrary, it is generally tailored as closely as possible to demand and existing ridership. As a result, our results show it is important to locate new transitway stations so that they intersect corridors with high levels of local transit demand able to support high levels of connecting bus service, as well as to connect heavily-used routes to existing transitway stations as conveniently as possible.

The lack of similarly useful results from the pedestrian and bicycle variables have implications for practice as well, though of a different kind. They show the importance of developing detailed, accurate, standardized data on the pedestrian and bicycle systems for improving the practice of transitway planning. Specifically, the lack of any reliable standard data source for sidewalk presence, pedestrian crossings, bicycle facilities, etc. severely hampers analysis of the very transit access and egress modes planners arguably expend the most time and energy to promote. It is quite difficult to rigorously evaluate the ridership benefits of high-quality pedestrian and bicycle facilities without systematic information on whether such facilities exist.

Data availability for transit analysis has made great strides in recent years—the GTFS data format in and of itself has been transformative—to the point that the in-vehicle leg(s) of a transit trip is well provided for. This research shows, however, that understanding the in-vehicle leg alone is insufficient for predicting transit use. This issue has implications even for bus access to transitways, as each bus trip to a transitway station has its own access mode which current data do not allow sophisticated consideration of. Contemporary geospatial analysis techniques show great promise for improving the practice of transit planning by allowing planners to consider the aggregate impacts of a myriad of potential, complex multimodal trips. These analysis techniques, however, are only as good as the data they are based on. As such, systematically collecting and publishing comprehensive data on the pedestrian and bicycle systems must be considered a high priority for further improving our understanding of the ridership impacts of multimodal connections to transitways.
REFERENCES


Mejia-Dorantes, L. (2018). An example of working women in Mexico City: How can their vision reshape transport policy? doi:https://doi.org/10.1016/j.tra.2018.05.022


