Modeling the Commute Mode Share of Transit Using Continuous Accessibility to Jobs

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Andrew Owen

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

David M. Levinson, Adviser

September, 2013

Abstract

This research develops an accessibility-based model of aggregate commute mode share, focusing on the share of transit relative to auto. It demonstrates the use of continuous accessibility — calculated continuously in time, rather than at a single or a few departure times — for the evaluation of transit systems. These accessibility calculations are accomplished using only publicly-available data sources. Multiple time thresholds for a cumulative opportunities measure of accessibility are evaluated for their usefulness in modeling transit mode share. A binomial logit model is estimated which predicts the likelihood that a commuter will choose transit rather than auto for a commute trip based on aggregate characteristics of the surrounding area. Variables in this model include demographic factors as well as detailed accessibility calculations for both transit and auto. The model achieves a ρ^2 value of 0.597, and analysis of the results suggests that continuous accessibility of transit systems may be a valuable tool for use in modeling and forecasting. It may be possible to apply these techniques to existing models of transit ridership and mode share to improve their performance and cost-effectiveness.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

It is increasingly common for urban transportation planning agencies to establish goals of increasing transit ridership. For example, in 2010 the Metropolitan Council, the metropolitan planning organization for the Minneapolis–Saint Paul, MN area, adopted the goal of doubling transit ridership in its region by 2030 [28]. In seeking that and similar goals, the Council and other agencies will be guided in part by the answer to the qustion: what makes a traveler choose to make a trip by transit rather than some other mode?

In almost all cases, that "other mode" is driving: according to estimates from the American Community Survey for the years 2007–2011, 86.5% of commute trips in the Twin Cities metropolitan area were made by car. Driving is a very different experience than using transit: driving is on-demand while transit is schedule-based; a train passenger can read a book while a driver ought not; a bus passenger can talk to other passengers while a solo driver is relatively isolated; a driver may pay to store her vehicle at the end of a trip while a streetcar passenger pays to board at the beginning of his.

When researchers and planners describe transit and driving as separate *modes* they are focusing on these differences. However, transit systems and road networks share a fundamental purpose: as transportation *systems*, they both are created with the intent that people will use them to reach destinations by paying some cost (in time and/or money). When viewed from this standpoint, it becomes apparent that travel by transit and travel by auto, regardless of their myriad differences, can be compared along two fundamental dimensions: the set of destinations to which they provide access, and the cost of reaching those destinations. Destinations and their cost of access are integrated by measures of accessibility; this study investigates the value of considering the accessibility provided by transit and auto in predicting mode share.

This analysis targets two main research goals:

1. To investigate the feasibility of modeling aggregate commute mode share at origins using detailed accessibility measurement for auto and transit, and

2. To test the hypothesis that transit accessibility calculation methods which reflect the ways that transit accessibility varies over time provide more accurate models of mode share than methods which use transit accessibility at a single point in time.

The motivation is not to obviate or dismiss the importance of mode share models that rely on detailed demographic data collection, but rather to facilitate transit mode share modeling and forecasting (*e.g.* sketch planning) in the absence of such data. Public and open data sources such as the General Transit Feed Specification (GTFS), OpenStreetMap (OSM), and the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) data allow low-cost calculation of the accessibility provided by transit systems. If an appropriate accessibility metric is able to describe a meaningful proportion of the variation in transit mode share, it can be used to increase the accuracy of current models or decrease the cost of future modeling efforts.

Chapter 2

Background

2.1 Accessibility

Accessibility marries the simpler concept of mobility with an understanding that travel is driven by a desire to reach destinations. It is important to distinguish between *individual accessibility* and *locational accessibility*: the former seeks to characterize the ease with which travelers might reach their destinations, subject to constraints of ability, budget, and other barriers; the latter examines accessibility as a spatial phenomenon by considering the costs and benefits of the potential trips offered by transportation systems between origins and destinations of interest. Horner [13] explored this distinction in the literature and notes that individual accessibility measures are generally poor at "producing ...generalized assessments of intraurban structure," while locational accessibility measures are more useful for "understanding relationships between transportation and land use."

Geurs and Van Wee [10] provide a taxonomy of accessibility measures and draw a similar distinction between locational and utility-based accessibility measures, and additionally identifies *infrastructure-based* measures which focus chiefly on the conditions of a transportation system and only secondarily (if at all) on the origins and destinations served by it. Metrics which indicate congestion or speeds on highway systems fall into this category.

Locational accessibility can be a particularly useful tool for transportation planners because it provides a way to evaluate the properties of transportation systems at a level that is aggregate enough to avoid the vagaries of individual users' preferences and constraints, but still detailed enough to provide guidance for planning at the city and regional level. It can be especially useful for multi-modal transportation planning because it is able to provide a level playing field for evaluating modes relative to one another [1]. This is achieved by setting aside the many particular differences between transportation *modes* and considering their relative merits as transportation *systems*.

Many different implementations of locational accessibility measurement are possible. El-Geneidy and Levinson [8] and provide a practical overview of historical and contemporary approaches. Most contemporary implementations can be traced at least back to Hansen [12], who proposes a measure where potential destinations are weighted by a gravity-based function of their access cost and then summed:

$$A_i = \sum_j O_j f\left(C_{ij}\right) \tag{2.1}$$

$$\begin{aligned} A_i &= \text{accessibility for zone } i \\ O_j &= \text{number of opportunities in zone } j \\ C_{ij} &= \text{time cost of travel from } i \text{ to } j \\ f\left(C_{ij}\right) &= \text{weighting function} \end{aligned}$$

The specific weighting function $f\left(C_{ij}\right)$ used has a tremendous impact on the resulting accessibility measurements, and the best-performing functions and parameters are generally estimated independently in each study or study area [14]. This makes comparisons between modes, times, and study areas challenging. Levine et al. [22] discuss these challenges in depth during an inter-metropolitan comparison of accessibility; they find it necessary to estimate weighting parameters separately for each metropolitan area and then implement a second model to estimate a single shared parameter from the populations of each. Geurs and Van Wee [10] also note the increased complexity introduced by the cost weighting parameter.

Perhaps the simplest approach to evaluating locational accessibility is discussed by Ingram [14] as well as Morris et al. [29]. *Cumulative opportunity* measures of accessibility employ a binary weighting function:

$$f\left(C_{ij}\right)=\begin{cases}1 & \text{if } C_{ij}\leq t\\ 0 & \text{if } C_{ij}>t\end{cases} \tag{2.2}$$

$t =$ travel time threshold

Accessibility is calculated for specific time thresholds and the result is a simple count of destinations that are reachable within each threshold. This approach involves both advantages and disadvantages. Both calculation and interpretation of the accessibility measure are dramatically simplified, but accessibility must be reported separately for each time threshold of interest, and the model cannot be finely calibrated to account for varying user preferences, values of time, etc.

2.1.1 Accessibility of Transit Systems

Lei and Church [21] provide a review of approaches to evaluating the accessibility provided by transit systems. Developments fall into two categories: changes in the techniques used to calculate travel times by transit, and changes in the ways those travel times are employed to calculate accessibility. The chief technical challenge in evaluations of transit accessibility has been calculating travel times. Prior to the mid-2000s, evaluations of accessibility in transit systems generally operated on simplified representations of transit networks. For example, a bus route might be assigned an average speed, a trip frequency, and hours of service. From these, travel times by transit are estimated rather than measured. Polzin et al. [33], Beimborn et al. [3], Wu and Hine [38], and Shen [34] follow this general approach. More aggregate evaluations of accessibility, such as those by Kawabata [15, 16] and Kawabata and Shen [17], make use of average travel times reported by transit commuters.

The introduction of the general transit feed specification in 2005 [11] made detailed transit schedules more widely available, while increases in generally-available computing power made their use more feasible. Krizek et al. [19], Lei and Church [21], Benenson et al. [4], Mavoa et al. [23], Owen and Levinson [32], and Dill et al. [7] demonstrate various calculations of transit accessibility using detailed transit schedules.

Despite their technical differences, these studies of transit accessibility are remarkably consistent in the selection and use of travel times to calculate accessibility. In almost every case, the accessibility provided by transit is derived from a single travel time between each origin/destination pair.

Some work has addressed this limitation. Polzin et al. [33] proposes a "time-of-day-based" evaluation of transit accessibility, and discusses the fact that transit service levels vary throughout the day. However, the ultimate focus is on variation in *demand*: after calculating accessibility on a simple hypothetical two-route transit network, the results are scaled based on the distribution of passenger trips throughout the day. Mavoa et al. [23] address the issue of accessibility variation by reporting a transit frequency measure alongside the accessibility value for each analysis zone. However, the accessibility values themselves are based on travel times calculated at a single departure time. Similarly, Dill et al. [7] include a single-departure-time accessibility variable when modeling transit ridership in addition to nine other variables describing local service levels.

Lei and Church [21] propose a method for evaluating transit accessibility that is sensitive to travel time variations throughout the day. This approach calculates accessibility by using detailed schedule information to find the minimum travel time in an arbitrary trip departure window. Owen and Levinson [32] follow a similar approach, guided by the earlier work of Krizek et al. [19]. While this makes the selection of a departure time less arbitrary, it still makes the assumption that transit users are willing and able to adjust their departure time, within an arbitrary window, in order to achieve this optimal travel time.

Fan et al. [9] provide the clearest example of how transit accessibility can be evaluated across multiple departure times. Accessibility values are calculated using travel times based on departures at each hour of the day; these are averaged to produce a single accessibility metric which incorporates travel times at multiple departure times.

2.1.2 Continuous Transit Accessibility

Anderson et al. [2] propose a method for implementing a measurement of transit accessibility that captures the way that accessibilty fluctuates continuously over time as trips approach and depart. For each trip departure n, the departure time and a vector C representing the travel times provided by that trip to all reachable destinations m are retained. A vector O is also established which provides the number of opportunities at each reachable destination:

$$T_n = \langle c_1, c_2, ..., c_m \rangle$$
$$O = \langle o_1, o_2, ..., o_m \rangle$$

Using the selected accessibility function, the accessibility provided at the departure time for each trip is calculated. A time sampling interval is then selected and the calculation moves backwards through time from each trip departure time, applying the same accessibility function to the next trip's travel time vector with the current time offset subtracted from each element. When the departure time d for the previous trip n is reached, the process is restarted using the travel time vector for that trip if the accessibility provided by the previous trip is greater than the accessibility provided by waiting for the next trip. Thus is it possible to calculate accessibilty A at each time point t:

$$A_t = \begin{cases} A_{T_1+\Delta t_1,O} & \text{if } t \leq d_1 \\ A_{T_2+\Delta t_2,O} & \text{if } d_1 < t \leq d_2 \\ ... \\ A_{T_n+\Delta t_n,O} & \text{if } d_{n-1} < t \leq d_n \end{cases} \tag{2.3}$$

Once accessibility is calculated at every time point of interest, it can be treated as a continuous variable over time.

2.2 Models of Commute Mode Choice, Mode Share, and Transit Ridership

McFadden's work of the 1970s and 1980s explored the theories of discrete choice and their application to the field of transportation [25, 24]. Most importantly, "The Measurement of Urban Travel Demand" [25] demonstrates the application of logit models to transportation mode choice. In McFadden's study, mode choice is estimated at the individual level using single estimated travel times via auto and transit for each individual's home–work trip [26]. Today, it is possible to calculate in far greater detail the benefits provided by entire transportation systems, and to evaluate all possible trips from a given origin, rather than only the trips actually taken by a set of survey respondents.

Mode share within an analysis zone summarizes the result of local individuals' mode choices. Research into mode share and mode choice has explored a wide range of factors which potentially contribute to individuals' mode choice process. Taylor et al. [36] provide a review of prior research into transit mode share which identifies two major avenues of investigation: *descriptive* analyses which focus on "traveller attitudes and perceptions," and *causal* analyses which examine "environmental, system, and behavioral characteristics." Within each, individual factors are identified as *external* if they are generally outside the direct control of transit planners and managers (*e.g.* population, income), or *internal* when they are endogenous to a specific transit system implementation (*e.g.* fares, vehicle design).

Using this classification, accessibility-based investigations of transit mode share can be described as causal, because they specifically avoid reliance on demographics or traveller preferences. They include both internal and external factors: travel times are the direct result of routing and scheduling decisions made by the operating agency, while the spatial distributions of opportunities are not.

It appears that in practice it is more common to model transit use by predicting ridership as a quantity, rather than a share. Dill et al. [7] provide an excellent example of contemporary transit ridership modeling as well as a comparison of three transit ridership models implemented by transit agencies in the Portland, OR region. The best-performing model achieves an adjusted R^2 of 0.69 using 29 independent variables categorized as "socio-demographic variables," "transit service variables," "transportation infrastructure variables." and "land use variables."

Lei and Church [21] discuss the use of accessibility measurement to make comparisons between different transportation modes, pointing out that service level indicators alone cannot facilitate direct comparisons between modes. Church and Marston [6] specifically argue for the use of relative accessibility as a tool to understand decision-making in transportation. Benenson et al. [4] proposes specific methods for comparing transit and auto accessibility, with the goal of identifying "accessibility gaps" where the difference between the accessibility levels of the two systems is greater than expected or desired.

Chapter 3

Data

Any analysis involving accessibility requires data describing the locations of origins and destinations, the cost of traveling between them, and the opportunities available at each.

3.1 Auto

Automobiles travel across the network of public roads and highways. Calculating travel times through this network requires two types of information: data describing the structure of the network, and data describing the cost of travel along individual links in the network. The Metropolitan Council, the metropolitan planning organization for the Twin Cities region, maintains a model of the regional road network. It provides a network topography for freeway, arterial, and collector roadways in the region, designed to model travel using transportation analysis zones (TAZs) as origins and destinations. The most recent version of this network was updated in 2009, and it provides an adequately accurate representation of the state of the regional road network in 2010, with model links conflated to match actual geometry. It does not model local roads, but instead provides "dummy links" which connect the centroid of each TAZ to adjacent arterial links. These are coded so that they may only be used for direct access to or from a TAZ centroid; they may not be used for travel between zones.

By itself, this model only describes the structure of the road network; per-link speed information is needed in order to generate travel times. The Twin Cities' regional freeway network is very well-instrumented, and data recorded by loop detectors throughout the system are archived by the Minnesota Traffic Observatory, operated and hosted by the University of Minnesota. Archived loop detector data for every weekday in 2010 provided the basis for average freeway link speeds.

Freeway speeds are derived from direct traffic observations made by embedded loop detectors which record the observed traffic volume and detector occupancy at 30-second intervals. MnDOT provides and estimated average effective vehicle length for each detector, which allows the calculation of speed from volume and occupancy. These 30-second speed measurements are aggregated by averaging to 5-minute time slots. Finally, the 5-minute speed measurements taken between 7 and 9 AM on weekdays during the year 2010 are averaged for each detector to produce a representation of average AM peak period speeds. The resulting detector speeds are assigned to links in the model network based on location.

In contrast to freeways, local arterials and collectors are only sparsely instrumented. Average arterial and collector link speeds are estimated from speed measurements made during a regional GPS-based travel survey in 2008 [39]. This data represents a very accurate measurement of traffic speeds at specific locations and specific times. Speeds for unobserved links are estimated from the samples collected on similar links.

3.2 Transit

Transit users interact with a different type of network than automobile drivers. Instead of navigating physical infrastructure, transit users move through a more abstract network of bus and rail routes provided by the transit operator. Metro Transit, the primary transit operator in the Twin Cities region, provides a publicly-available general transit feed specification (GTFS) dataset.

The schedule used in this analysis was published on November 9, 2009 and describes transit services in operation through March of 2010. It includes bus and rail services operated by all fixed-route transit providers in the Twin Cities area, including Metro Transit, MVTA, SouthWest Transit, the University of Minnesota, and others. Travel time calculations were performed using the schedule for Wednesday, December 2, 2009; transit services schedules for this day were typical for the entire 2009–2010 period.

Transit travel time calculations include off-vehicle time costs: waiting at stations as well as time spent accessing an initial station, accessing any required transfers, and accessing the destination after disembarking. This requires a detailed representation of pedestrian facilities in order to calculation walk times between origin and destination census blocks and transit stations. OpenStreetMap [30] provides an open-source dataset with sufficient detail for this purpose. Specifically, the pedestrian network is comprised of OpenStreetMap features with the “footway,” “pedestrian,” and “residential” tags.

3.3 Labor and Employment

Data describing the distribution of labor and employment in the region are drawn from the U.S. Census Bureau’s Longitudinal Employer-Household Dynamics program (LEHD). The workplace area characteristic dataset for 2010 provides Census block-level estimates of employee home and work locations.

In general, LEHD is a very useful data source for accessibility evaluation because it is updated yearly and is drawn from actual payroll records collected at the state level — in this case, by the Minnesota Department of Employment and Economic Development. However, it is important to recognize the fact that LEHD data is *synthetic*: while it is based on actual payroll records, the published results are created by an algorithm designed to produce data which are statistically similar to the underlying data, and which converge to the same distribution when aggregated. An analysis by Spear [35] of LEHD data in transportation analysis found LEHD to be a useful source of both home and work location data, but identified shortcomings related to job locations of federal workers. Tilahun and Levinson [37] demonstrate the use of LEHD data in contemporary transportation research.

3.4 Mode Share

The American Community Survey (ACS) collects data describing commute mode choice during its annual national survey of households and individuals. This analysis uses the 2007–2011 5-year estimate ACS dataset. For each of the 2,085 block groups in the study area, the ACS estimated the total number of workers, which includes the workers who report commuting by each mode as well as workers who report working from home.

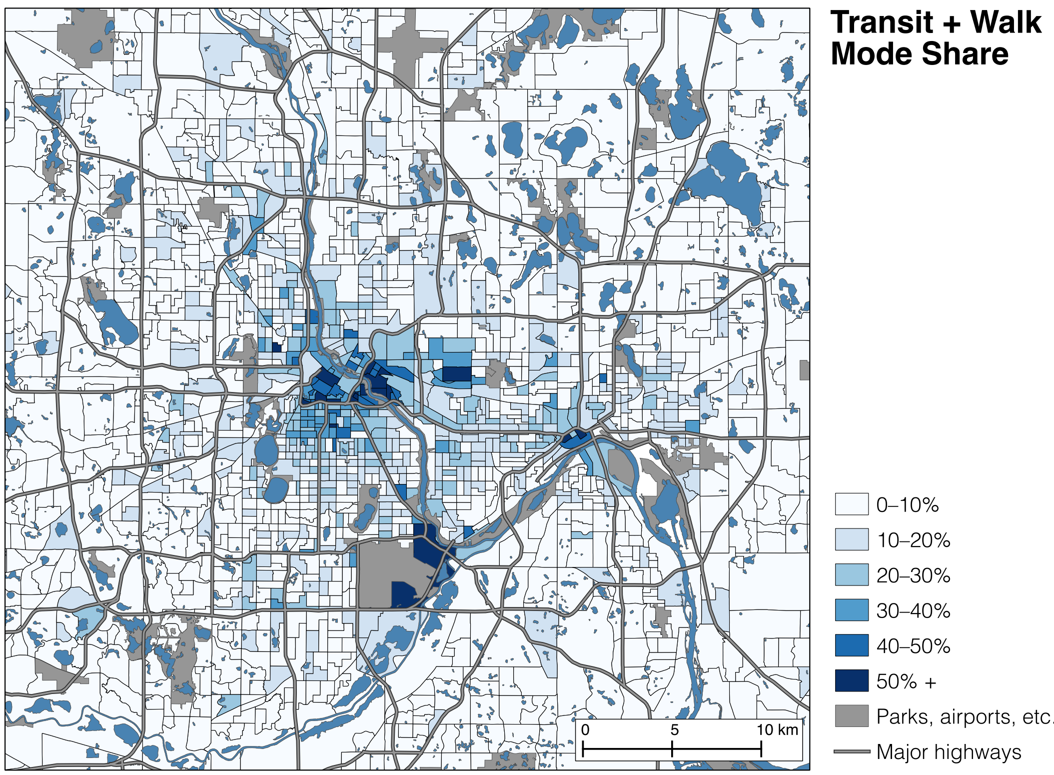

Figure 3.1: Commute mode share of transit + walking (ACS 2007–2011 estimates).

The ACS tracks commute mode at a level of detail not necessary for this analysis, so the mode categories are regrouped according to Table 3.1. Because of the specific methodology used to calculate transit accessibility (discussed in section 4.1), transit and walking are considered a single mode. All modes that involve motorized vehicles on public roads are grouped as "Auto." In order to model a binary mode choice, workers who commute via bicycle, who work at home, or who commute via some other means are excluded. Figure 3.1 illustrates transit mode share in the study area.

Table 3.1: Grouping of ACS "means of transportation to work" responses

Transit	**Auto**	**Excluded**
Bus or trolley bus	Drove alone	Bicycle
Streetcar or trolley car	Carpooled	Taxicab
Subway or elevated		Worked at home
Railroad		Other means
Ferryboat		
Walked		

Chapter 4

Methodology

4.1 Calculating Travel Times by Transit

Travel times by transit are calculating using OpenTripPlanner [31], an open-source software package sponsored by Portland's TriMet. OpenTripPlanner is a graph-based transit routing system which operates on a unified graph including links representing road, pedestrian, and transit facilities and services.

The time cost of travel by transit is comprised of several components. *Initial access time* refers to the time cost of traveling from the origin to a transit stop or station. *Initial wait time* refers to the time spent after reaching the transit station but before the trip departs. *On-vehicle time* refers to time spent on board a transit vehicle. When transfers are involved, *transfer access time* and *transfer wait time* refer to time spent accessing a secondary transit station and waiting there for the connecting trip. Finally, *destination access time* refers to time spend traveling from the final transit station to the destination. All of these components are included in the calculation of transit travel times.

This analysis makes the assumption that all access portions of the trip — initial, transfer(s), and destination — take place by walking at a speed of 1.38 meters/second along designated pedestrian facilities such as sidewalks, trails, etc. On-vehicle travel time is derived directly from published transit timetables, under an assumption of perfect schedule adherence.

An unlimited number of transfers are allowed. This is somewhat unusual among evaluations of transit accessibility. In many cases travel times are limited to trips involving no more than one or two transfers; this is justified by the observation that in most cities a very large majority (often over 90%) of observed transit trips involve no more than two transfers. However, the shortest-path algorithms typically employed in these evaluations are single-constraint algorithms: they are guaranteed to find the shortest path only when given a single constraint (typically, travel time). When the path search tree is pruned based on an additional constraint such as number of transfers (or, in some cases, transfer wait time), these algorithms provide no insurance against a shorter trip, requiring additional transfers, remaining undiscovered in the pruned space [18, 20, 31].

Given the realities of transit networks, it likely that cases where (for example) a three-transfer itinerary provides a faster trip than a two-transfer itinerary are relatively rare. However, given the goal of evaluating the full accessibility provided by a transit system rather than simply the accessibility that is likely to be utilized, this analysis prefers the algorithmically correct approach of using travel time as the single routing constraint and leaving the number of transfers unconstrained.

Just as there is no upper limit on the number of vehicle boardings, there is no lower limit either. Transit and walking are considered to effectively be a single mode. The practical implication of this is that the shortest path by “transit” is not required to include a transit vehicle. This may seem odd at first, but it allows the most consistent application and interpretation of the travel time calculation methodology. For example, the shortest walking path from an origin to a transit station in some cases passes through potential destinations where job opportunities exist. In other cases, the shortest walking path from an origin to a destination might pass through a transit access point which provides no trips which would reduce the origin–destination travel time. In these situations, enforcing a minimum number of transit boardings would artificially inflate the shortest-path travel times. To avoid this unrealistic requirement, the transit travel times used in this analysis are allowed to include times achieved only by walking.

4.2 Calculating Travel Times by Auto

Travel times by auto are calculated using a shortest-path search on a graph defined by the Metropolitan Council's 2009 regional road network model and the speed data sources discussed in section 3.1.

It should be noted that auto travel times do no include an estimation of access times at the origin (*e.g.*, walking from home to a car) or at the destination (*e.g.*, walking from a parking space to work). While the former is likely negligible given the typical residential urban forms in the study area, the latter has the potential to be significant. Unfortunately, the spatial resolution of the road and highway network model available for this analysis is not detailed enough to allow accurate estimation of destination access times.

4.3 Calculating Accessibility to Jobs

Using the travel time matrices described above, cumulative opportunities accessibility to jobs, by both auto and transit, is calculated for each origin according to Equation 2.2 and using values of 10, 20, 30, 40, 50, and 60 minutes for the time threshold (t). For each origin, this process identifies all destinations reachable within t minutes and sums the number of jobs at those destinations. For auto accessibility, this takes place using TAZs as origins and destinations; for transit, Census blocks are used as origins and destinations. Later, the geographic resolution of all accessibility data is standardized as described in section 4.5.

4.4 Calculating Time-Continuous Transit Accessibility

Transit accessibility to jobs is evaluated as described in section 4.3 using every minute in the 7–9 AM peak period as a potential departure time. Figure 4.3 illustrates how accessibility varies minute by minute at a single census block. Accessibilty increases as departure times at nearby stops approch, and then drops after trips depart. Deep troughs in the accessibility profile are associated with times with few or no upcoming trip departures at nearby stops, while sustained periods of high accessibility are associated with periods providing frequent departures. Because of these fluctuations, the average

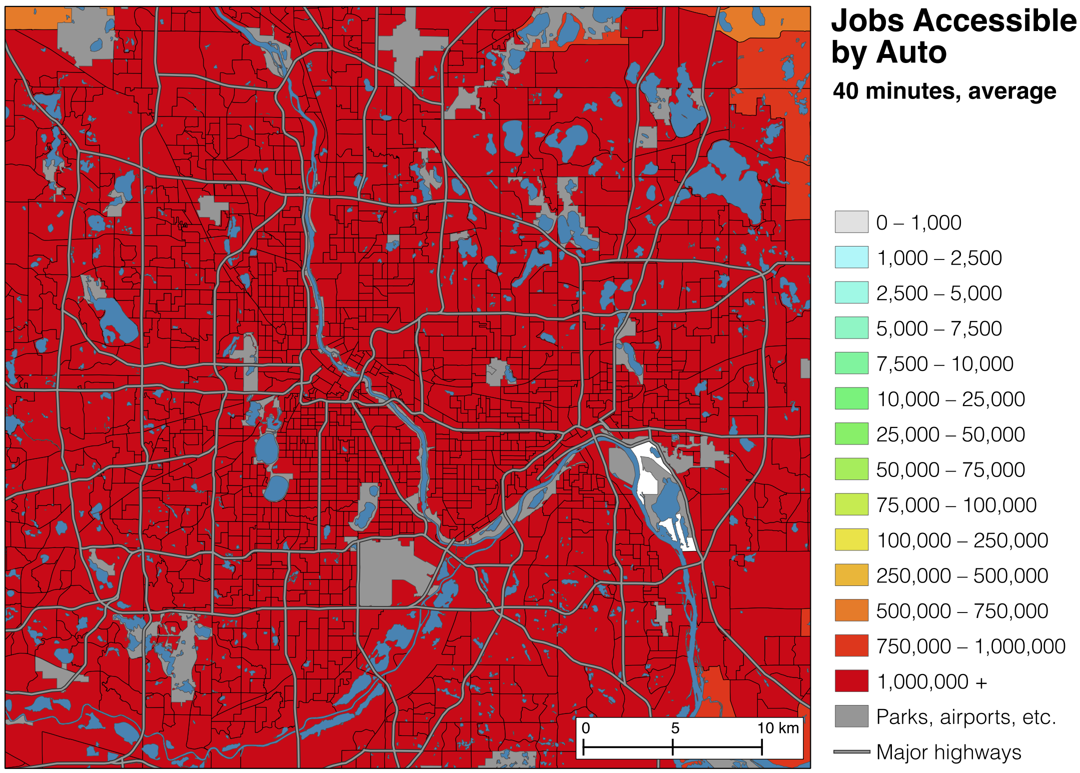

Figure 4.1: Jobs accessibile within 40 minutes by auto

accessibility over the 7–9 AM peak period is significantly lower than the maximum accessibility value over the same period.

Figure 4.4 provides a map of average accessibility over the 7–9 AM peak period for each block group. Average accessibility values are across the board lower than the maximum accessibility values (compare with Figure 4.2), but the magnitude of the difference between the maximum and the average accessibility values varies throughout the region.

In addition to average accessibility, calculation of time-continuous transit accessibility allows computation of various measures of variance. For each block group, the standard deviation, variance, and coefficient of variation are calculated over the 7–9 AM peak period. As shown in Figure 4.5, transit accessibility varies more widely in some parts of the metro area than others. In central Minneapolis, downtown Saint Paul, and

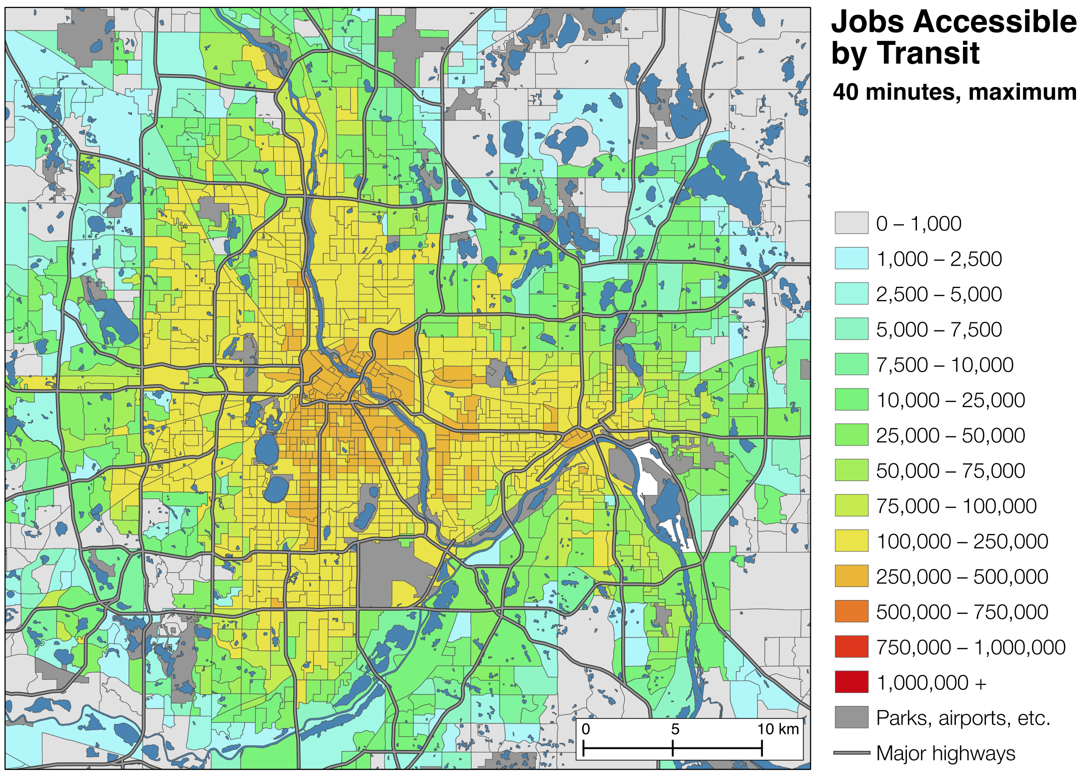

Figure 4.2: Jobs accessible within 40 minutes by transit 7–9 AM (maximum)

a few places in between, very low accessibility coefficients of variation suggest that high-frequency transit service is preventing gaps between trip departures where accessibility would otherwise drop.

In contrast, the inner-ring suburbs, especially to the west and northwest of Minneapolis as well as to the south of Saint Paul, the accessibility coefficients of variation tend to be high — in many places, the standard deviation of accessibility is several times the local average value. This is not necessarily a negative phenomenon; it merely indicates a different approach to transit service. Transit routes in these areas tend to be fairly infrequent, running every half hour or hour. Between these trips accessibility is very low. Additionally, these areas often have express commuter serivce during the AM and PM peak periods, which provides very high accessibility per trip but remain low-freqency. Transit accessibility in these areas varies sharply over time, and the local coefficients of variation reflect that.

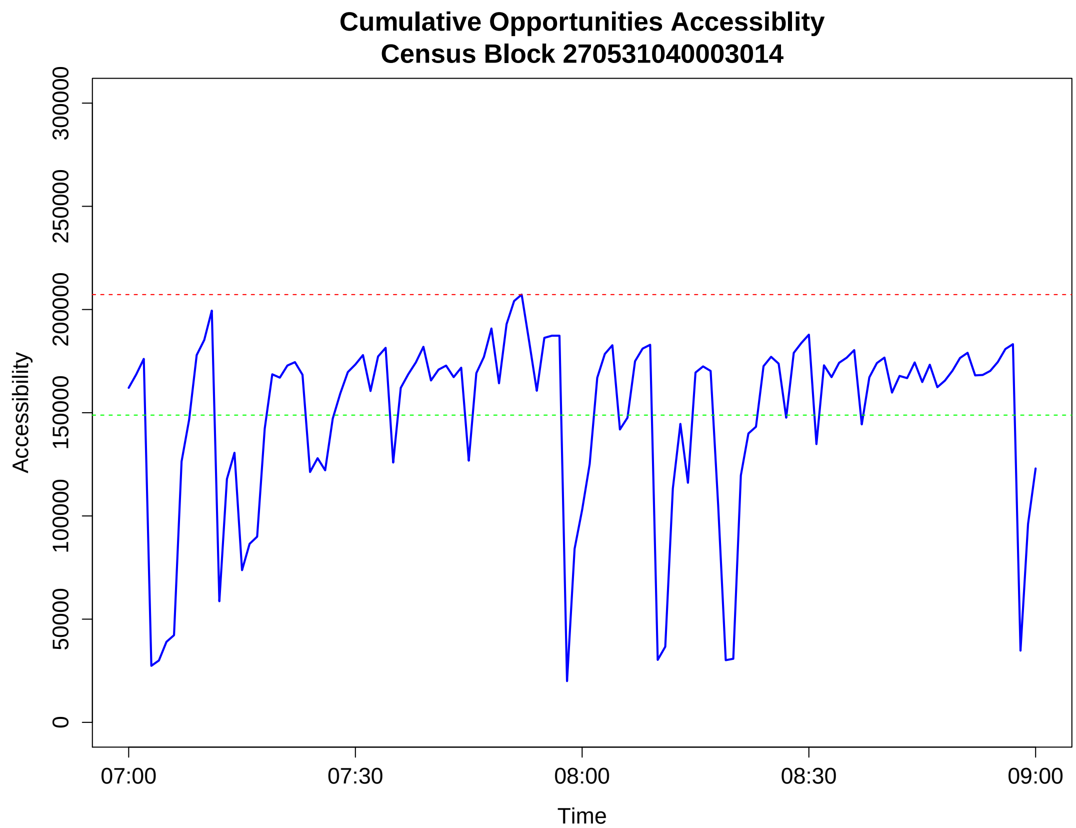

Figure 4.3: Time-continuous transit accessibility for a single Census block

Beyond the inner-ring suburbs, the accessibility coefficients of varation are very low, or even zero. These areas are effectively outside the fixed-route transit service area: walking access to the nearest transit stops takes long enough (equal to or greater than the cumulative opportunity threshold) that transit service provides no increase in accessibility relative to walking alone.

4.5 Geographic Standardization

The three principal inputs to this analysis are available at three different levels of geographic resolution. Auto accessibility is calculated at the TAZ level based on the level of detail provided by the Metropolitan Council's current road network model;

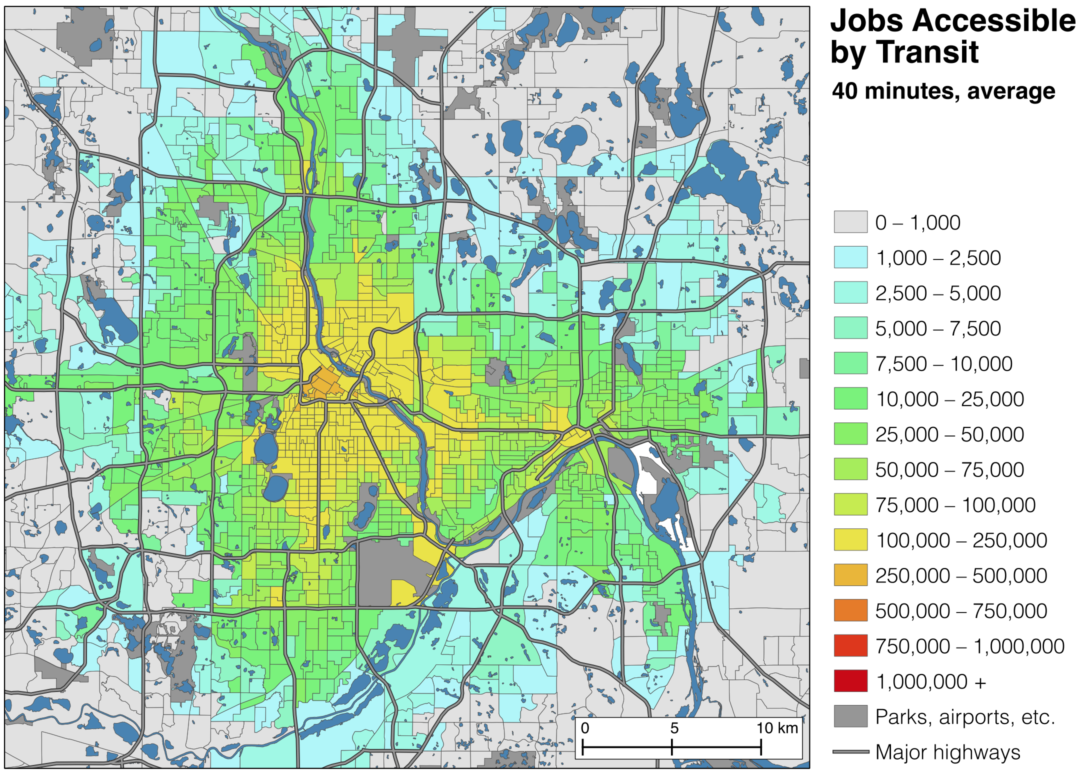

Figure 4.4: Jobs accessible within 40 minutes by transit, 7–9 AM (average)

transit accessibility is calculated at the block level using detailed GTFS schedule and OpenStreetMap network data; mode share data collected and estimated by the ACS is available at the block group level. If these data sources are to be used together they must be made comparable in a reasonable way.

Auto accessibility values are assigned to block groups based on centroid inclusion. Each block group is assigned the accessibility value of the TAZ which includes its centroid. Because there are roughly twice as many block groups in the Twin Cities metropolitan area as there are TAZs, neighboring block groups sometimes receive the same auto accessiblity value.

Transit accessibility is aggregated to the block group using worker-weighted averaging. Within each block group, the accessibility values for each contained block are assigned weights proportional to the local number of resident workers, and then averaged using these weights. Thus the final accessibilty value for each block group represents

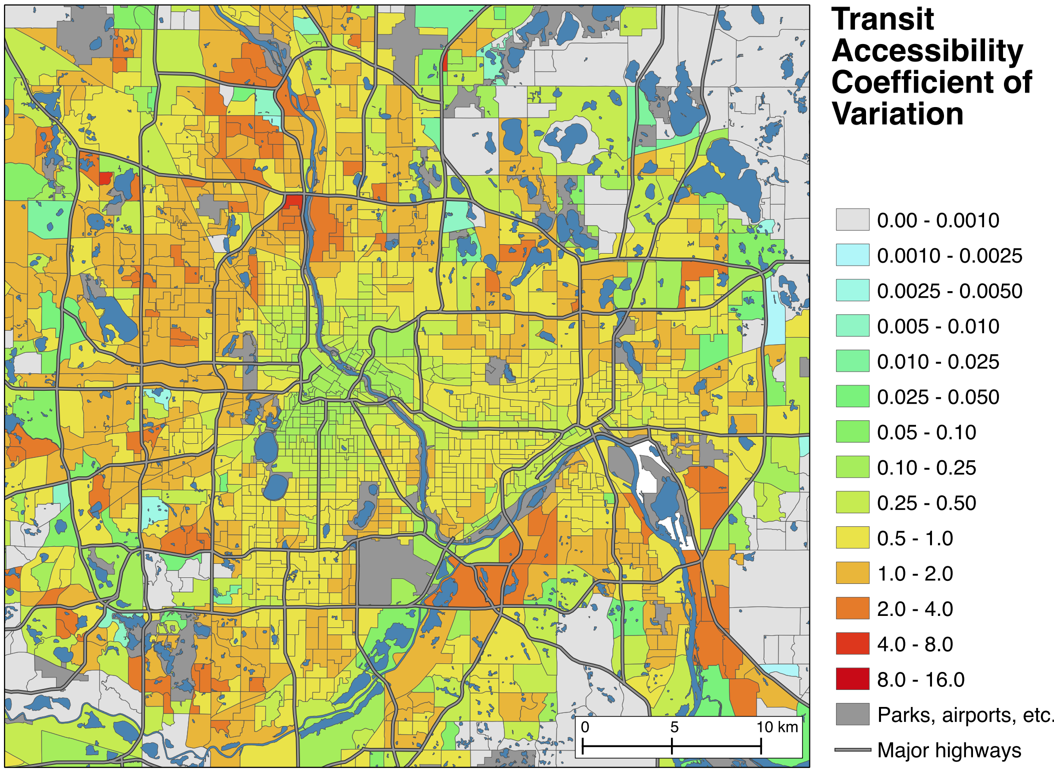

Figure 4.5: Transit accessibility coefficient of variance

the accessibility experienced by the average worker living within it.

4.6 Model Formulation

Transit mode share is considered to be the aggregate outcome of individual commuters' mode choices. The choice modeled here is the choice between 1) using transit or walking, and 2) using an automobile or an auto-like mode. The grouping of modes in discussed in section 3.4; these will subsequently be referred to simply as "transit" and "auto." As guided by McFadden [25], McFadden et al. [26], and numerous subsequent mode choice models, the probability of choosing a mode is modeled as a function of the utility provided by that mode and other modes, where the utility is a linear function of a set of parameters for each mode:

$$P_1 = \frac{1}{1 + e^{-U}} \tag{4.1}$$
$$U = \beta_o + \beta_1 x_1 + ... + \beta_n x_n$$

This is the form used for all models following. Transit is always regarded as "choice 1," so a positive coefficient will always indicate a positive effect on the probability of choosing transit. The following sections describe the various model formulations tested in this analysis.

4.6.1 Single-Variable Binomial Logit

It is useful to begin with an extremely simple model. Models using a single independent variable have great ease of interpretation and visualization, both of which are beneficial to lay the groundwork for more complicated models.

The accessibility ratio models propose that the mode share of transit depends on the ratio of transit accessibility to auto accessibility in each block group:

$$U = \beta_0 + \beta_1 \left(\frac{A_t}{A_a} \right) \tag{4.2}$$

This model will be tested with accessibility values calculated using each of the six time cost thresholds, with transit and auto values for the same thresholds paired together.

While this is a simplification of the classic mode choice model formulation presented above, it retains a reasonable interpretation. Because auto accessibility (A_a) is always significantly higher than transit accessibility (A_t) in the study area, it is very likely that the jobs reachable by transit within a given time threshold are a subset of the jobs reachable by auto. As a greater proportion of jobs are reachable by both transit and auto with a given time threshold, it is reasonable to expect that commuting by transit becomes more attractive (relative to commuting by auto) to more workers. Lei and Church [21] identifies earlier literature supporting a direct relationship between the relative travel times of transit and auto and the mode share of transit.

Using this basic model form, single-variable models will be estimated using, in turn, maximum transit accessibility (model S1) and average transit accessibility (model S2) as the numerator in the utility function (Equation 4.2).

4.6.2 Multivariate Binomial Logit

Multivariate binomial logit models are developed next. The first set of models use independent variables derived only from evaluations of the transit and auto transportation systems. These are labeled M1 through M4. They explore the relationships between transit mode share and various combinations of average transit accessibility, maximum transit accessibility, and variance of transit accessibility. Various combinations of these variables are tested to discover which provides the best-performing representation of the utility provided by the transit system. Auto accessibility is included in all models as a representation of the utility provided by driving. Models are estimated using groups of parameters for each 10-minute accessibility threshold increment; the same threshold is used for all parameters in each model. The best-performing model is extended in model M5 to include variables derived from socioeconomic and demographic data. Table 4.1 describes each variable used in the models.

Table 4.1: Variable Descriptions and Statistics

Variable	**Description**	**Mean†**	**S.D.†**
A_{Tm}	Maximum transit accessibility (100,000 jobs)	0.874‡	0.995‡
A_{Ta}	Average transit accessibility (100,000 jobs)	0.358‡	0.557‡
A_{Tv}	Transit accessibility variance (100,000 jobs)	8.449‡	12.06‡
A_A	Auto accessibility (100,000 jobs)	13.48‡	2.95‡
I	Mean household income ($1,000)	69.89	30.64
S	Mean household size	2.51	0.53
V	Mean vehicles per household	1.77	0.45
W	Percent white, non-hispanic population	81.30	22.89
B	Percent of population 25+ with B.A./B.S. or higher	25.70	11.92

† *Over 2,082 block groups*

‡ *40-minute threshold*

4.7 Model Evaluation

In order to evaluate the several models that will be presented, it will be necessary to measure two things. First, it must be possible to evaluate how well each model fits the data. Second, it must be possible to evaluate how each model performs relative to the others.

Model fit is evaluated using the pseudo-R^2 for logistic regression labeled as ρ^2 by McFadden [24, 25], and calculated from the likelihood measures of the specified model and the null model. In an evaluation of pseudo-R^2 measures in the context of logistic regression, Menard [27] identifies several desirable qualities of ρ^2. Most importantly, it was found to have an "intuitively reasonable interpretation as a proportional reduction in error measure."

Relative model performance is evaluated using the Akaike information criterion (AIC). AIC estimates the amount of information lost by using the model instead of the data, and penalizes models for including more parameters. The latter is an important property given that this analysis depends on comparisons between models with different numbers of parameters. [5]

Chapter 5

Results and Discussion

5.1 Performance of Single-Variable Models

Table 5.1 provides a comparison of the performance of the single-variable S1 and S2 models. Model S1, which uses maximum transit accessibility, peaks in fit at the 30-minute cumulative opportunity (where ρ^2=0.498) threshold before tapering off as the threshold increases. In contrast, model S2, which uses average transit accessibility, improves rapidly in fit as the threshold increases to 30 minutes, but then continues to improve slightly through the 60-minute threshold. In both models, thresholds below 30 minutes provide very poor fit. Figure 5.1 and Figure 5.2 illustrate the fit of these models using the best-performing thresholds.

Table 5.1: Performance of single-variable models

	S1		**S2**	
Threshold	AIC	ρ^2	AIC	ρ^2
10	126300	0.168	130600	0.139
20	105720	0.302	127314	0.159
30	**75951**	**0.498**	81515	0.462
40	76050	0.497	73066	0.518
50	77281	0.490	71622	0.527
60	78962	0.479	**71413**	**0.528**

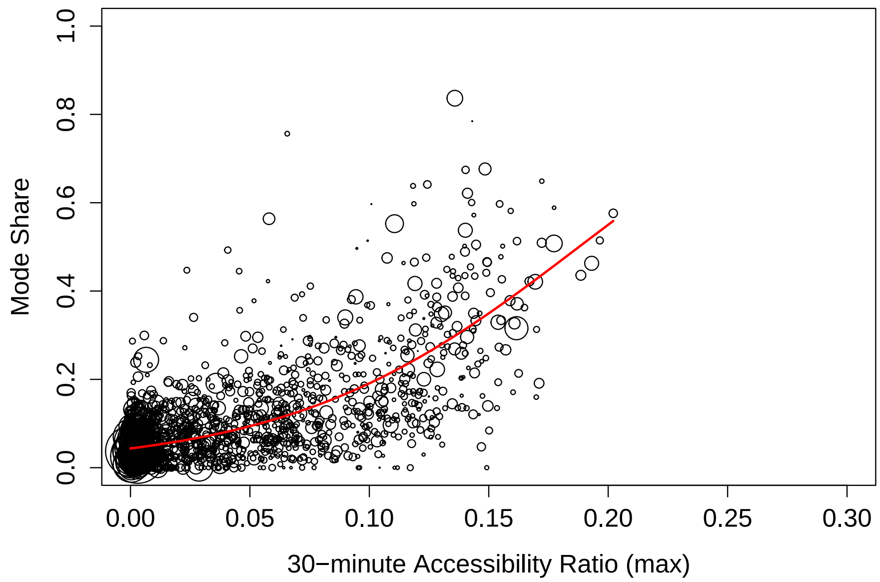

Figure 5.1: RM model fit. ρ^2=0.498

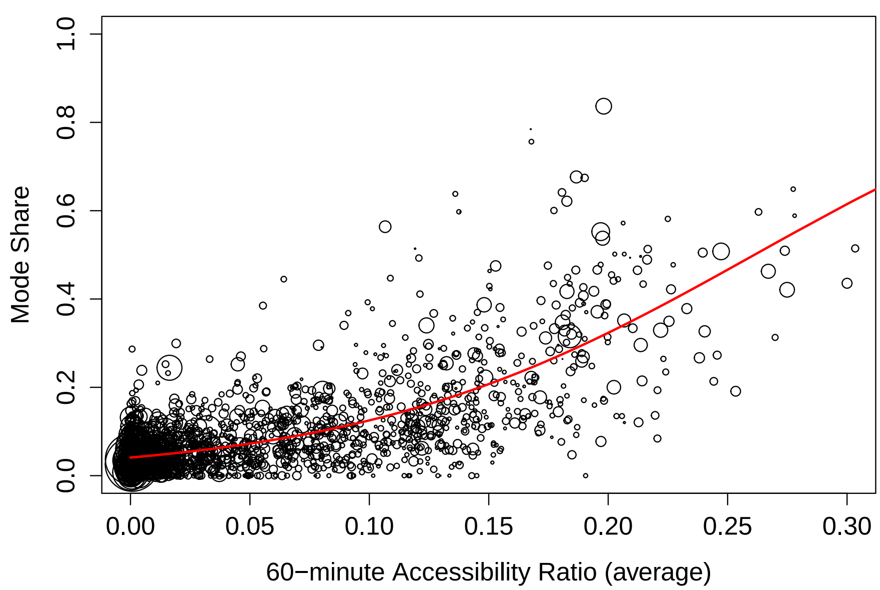

Figure 5.2: RA model fit. ρ^2=0.528

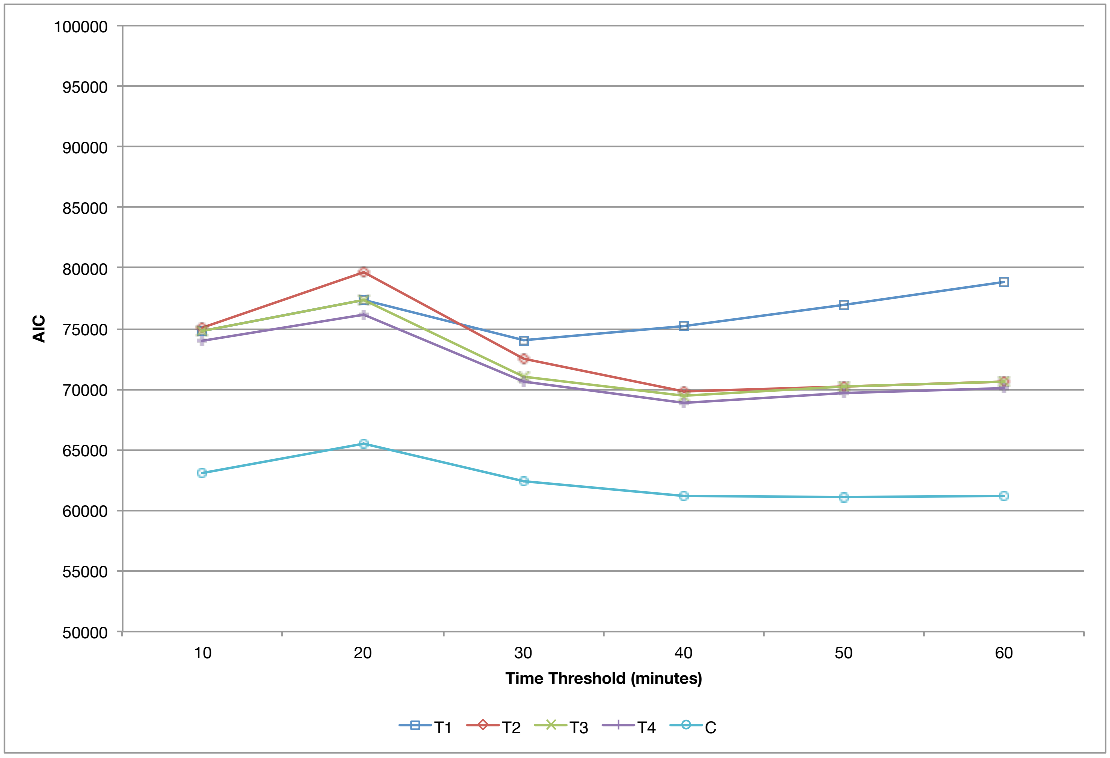

Figure 5.3: Comparison of multivariate models (lower values indicate better performance).

5.2 Performance of Multivariate Models

Figure 5.3 and Table 5.2 summarize the relative performance of the various multivariate models at each accessibility time threshold, as measured by the AIC of each model. Lower AIC values indicate more information preserved in the model, and thus greater relative model performance. Higher ρ^2 values indicate a better fit between the model and the data.

As with the single-variable models, a comparison of the M1 and M2 models indicates that at time thresholds below 30 minutes, maximum transit accessibility provides a better-performing model than does average transit accessibility, but the opposite is true at time thresholds of 30 minutes and higher. However, at time thresholds below 30 minutes, model performance is relatively low regardless of what parameters are included. The most significant improvements in model performance appear when moving from a

Table 5.2: Model Performance Using Various Time Thresholds

	Model M1		Model M2		Model M3		Model M4		Model M5	
Threshold	AIC	ρ^2	AIC	ρ^2	AIC	ρ^2	AIC	ρ^2	AIC	ρ^2
10	74786	0.506	75027	0.504	74786	0.506	73952	0.512	63091	0.583
20	77309	0.489	79606	0.474	77297	0.490	76198	0.497	65436	0.568
30	**74043**	**0.511**	72547	0.521	71012	0.531	70657	0.533	62422	0.588
40	75272	0.502	**69783**	**0.539**	**69460**	**0.541**	**68858**	**0.545**	61219	0.596
50	76990	0.491	70257	0.536	70179	0.537	69682	0.540	**61097**	**0.597**
60	78786	0.480	70676	0.533	70672	0.533	70080	0.537	61151	0.596

20-minute to a 30-minute accessibility threshold.

An important difference between models M1 and M2 is that while the accuracy of model M peaks at the 30-minute accessibility threshold and then declines as the threshold increases, the accuracy of model A is highest at the 40-minute threshold but then remains relatively steady through the 50- and 60-minute thresholds.

When combined in model M3, maximum and average transit accessibility together provide a consistently better fit for the data than either does individually, with model fit peaking at the 40-minute threshold. Across all time thresholds, the addition of variance of accessibility to form the MAV model provides a small but statistically significant improvement in model fit. The M4 model at a 40-minute threshold provides both the best fit of the transportation-based models, as indicated by a ρ^2 of 0.545, and the best preservation of information relative to the data, as indicated by an AIC of 68858.

The introduction of socioeconomic and demographic variables in model M5 has the effect of improving model fit at all thresholds. The difference in performance between model M5 and models M1-M4 is much greater than than between any two T models.

All models improve in performance when moving from a 10-minute threshold to a 20-minute threshold; this is an unexpected result. One possible explanation lies with the fact that the transit accessibility calculation method allows pure walking trips as well. If, as seems possible, only a small share of transit commute trips are shorter than 30 minutes while only a small share of walking commute trips are longer than 20 minutes, it may be that the 20-minute transit/walk accessibility does not align well with actual traveler behavior.

Table 5.3 lists the coefficients, and their significance, as estimated in the various models. In general, the accessibility coefficients follow expected and intuitive patterns. Increases in both maximum and average accessibility are associated with increases in transit mode share, while increased variance of transit accessibility is associated with lower transit mode share.

However, the coefficients for the auto accessibility parameter are unexpectedly positive – indicating that higher auto accessibility is associated with higher transit mode share. First, as illustrated in Figure 3.1 and Figure 4.4, auto accessibility is, throughout the study area, much higher than transit accessibility — often by multiple orders of magnitude. Also, the higher speeds associated with the road network mean that auto accessibility varies far more gradually over space than does transit accessibility. Combined, these suggest that meaningful comparisons of accessibility between transit and auto may require more careful investigation of the scale and scope of their variation. Second, , auto accessibility and transit mode share are both highest in downtown Minneapolis and decline with distance. This is true of many other

In model M5, the signs and magnitudes of coefficients of the sociodemographic variables generally follow the patterns suggested by previous literature. Transit mode share is negatively associated with household income, a relationship that is found in nearly all investigations of transit ridership in North America. Some of the demographic variable results are also intuitive: transit mode share has a negative association with increasing share of white, non-hispanic population, and a negative association with increased ownership of private vehicles. However, the coefficients for household size and education (% with bachelor's degrees or higher) are unexpected. Dill et al. [7] found negative associations between these variables and transit ridership.

Also, it is important to note that in this model the effect of maximum transit accessibility is not statistically significant. This presents a challenge for an accessibility-based model, through both average transit accessibility and variance of transit accessibility maintain their significance. This unexpected finding regarding maximum transit accessibility may be related to factors such as residential location selection or work arrival time flexibility, both of which have logical associations with the demographic variables included in the model.

Table 5.3: Commute Model Choice Model Results

	Model 1		Model 2		Model 3		Model 4		Model 5	
Variable	Value	z	Value	z	Value	z	Value	z	Value	z
A_{Tm}	0.963	181.9^{***}			0.152	18.1^{***}	0.423	30.7^{***}	0.008	0.7
A_{Ta}			1.1	235.1^{***}	0.902	74.7^{***}	0.669	44.0^{***}	0.652	49.6^{***}
A_{Tv}							-0.013	-24.4^{***}	-0.002	-5.4^{***}
A_A	0.055	46.5^{***}	0.089	53.3^{***}	0.075	71.9^{***}	0.068	37.5^{***}	0.052	16.1^{***}
I									-0.004	-19.6^{***}
S									0.250	29.7^{***}
V									-0.835	-50.2^{***}
W									-0.117	-5.1^{***}
B									0.596	15.5^{***}
ASC	-3.54	-298.7^{***}	-4.14	-182.8^{***}	-4.01	-173.4^{***}	-3.95	-171.6^{***}	-2.63	-45.6^{***}
ρ^2	0.511		0.539		0.541		0.545		0.597	
AIC	74043		69783		69460		68858		61097	
Threshold	30		40		40		40		50	
n	2082									

****Significant at 0.001 level or better*

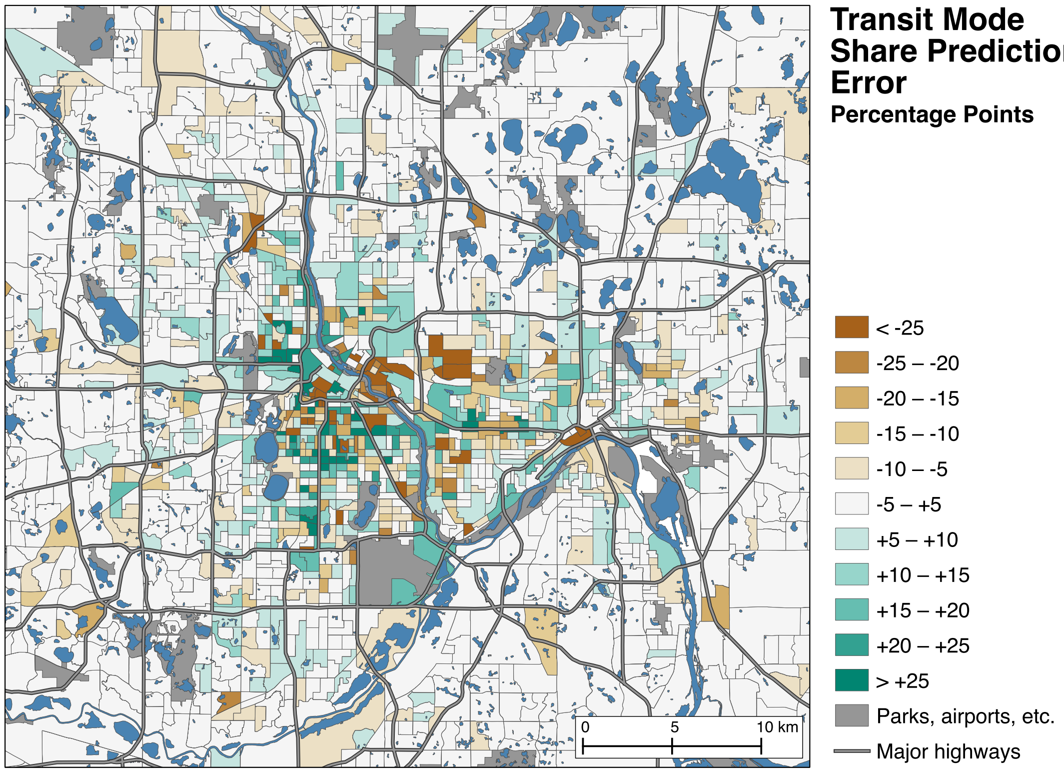

Figure 5.4: Spatial distribution of model error

5.3 Spatial Distribution of Model Error

Figure 5.4 illustrates the spatial distribution of error in the best-performing model (model M5). It is important to recognize that the model error is not distributed evenly or randomly across space, but that several distinct spatial patterns are apparent. These have implications for model interpretation and directions for future research.

Areas where the model error is strongly negative — that is, where the model predicts a lower transit mode share than is estimated by the ACS — are largely associated with “special case” land uses or urban forms that are unusual in the study area. The campuses of the University of Minnesota are apparent as two clusters of strongly negative model error in the central metro area. The block groups which make up these clusters contain housing used by large numbers of graduate and undergraduate students. It is likely that jobs held by these residents are on or near the University campus; these populations are also typically associated with low auto ownership. Neither of these factors are accounted for in the model.

Similarly, model error is moderately to strongly positive in much of southwest Minneapolis, particularly along Hennepin Avenue and south of Lake Harriet. The block groups which make up these areas typically have household incomes higher than the regional average; increased income is typically associated with lower use of transit.

Chapter 6

Conclusion

Thanks to improvements in data sources and processing capabilities, detailed accessibility calculations are practical today which would have been previously unachievable. This analysis has built on earlier work which established methods for calculating continuous transit accessibility in two primary ways.

First, it has demonstrated the feasibility of accessibility-based mode share modeling. Though questions remain as to the most appropriate model specification, accessibility-based modeling can provide useful results even in the absence of socioeconomic and demographic data. When such data is added to an accessibility-based model, accessibility parameters retain a statistical and practical significance. Second, it has demonstrated the value of accessibility evaluation methods for transit systems which reflect the variation of accessibility over small time scales. Though most accessibility-based investigations of transit use evaluate accessibility only at a single departure time, this research shows that model fit may be improved by incorporating time-averaged accessibility and/or measures of accessibility variation over time.

These results suggest that while the specific model formulations presented here can be improved upon, continuous evaluations of the accessibility provided by transit systems are a promising metric for use in ridership and mode share modeling. The techniques demonstrated allow detailed calculations of transit accessibility to be implemented at low cost and with data that are generally publicly available. It may be possible to apply these techniques to more robust existing models of transit ridership and mode share to improve their performance and cost-effectiveness.

References

[1] Anderson, P., Levinson, D. M., and Parthasarathi, P. (2012). Using Twin Cities destinations and their accessibility as a multimodal planning tool. Technical Report 2012-05, Minnesota Department of Transportation.

[2] Anderson, P., Owen, A., and Levinson, D. (2013). The time between: Continuously-defined accessibility functions for scheduled transportation systems. In *92nd Annual Meeting of the Transportation Research Board*.

[3] Beimborn, E. A., Greenwald, M. J., and Jin, X. (2003). Accessibility, connectivity, and captivity: impacts on transit choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1835(1):1–9.

[4] Benenson, I., Martens, K., and Rofé, Y. (2010). Measuring the gap between car and transit accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, 2144(1):28–35.

[5] Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, 2nd edition.

[6] Church, R. L. and Marston, J. R. (2003). Measuring accessibility for people with a disability. *Geographical Analysis*, 35(1):83–96.

[7] Dill, J., Schlossberg, M., Ma, L., and Meyer, C. (2013). Predicting transit ridership at the stop level: The role of service and urban form. In *92nd Annual Meeting of the Transportation Research Board*.

[8] El-Geneidy, A. M. and Levinson, D. M. (2006). Access to destinations: Development

of accessibility measures. Technical Report MN/RC-2006-16, Minnesota Department of Transportation.

[9] Fan, Y., Guthrie, A., and Levinson, D. (2013). Impact of light-rail implementation on labor market accessibility: A transportation equity perspective. *Journal of Transport and Land Use*, 5(3):28–39.

[10] Geurs, K. T. and Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12(2):127–140.

[11] Google, Inc. (2013). https://developers.google.com/transit/gtfs/.

[12] Hansen, W. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2):73–76.

[13] Horner, M. W. (2004). Spatial dimensions of urban commuting: A review of major issues and their implications for future geographic research. *The Professional Geographer*, 56(2):160–173.

[14] Ingram, D. R. (1971). The concept of accessibility: A search for an operational form. *Regional Studies*, 5(2):101–107.

[15] Kawabata, M. (2003). Job access and employment among low-skilled autoless workers in US metropolitan areas. *Environment and Planning A*, 35(9):1651–1668.

[16] Kawabata, M. (2009). Spatiotemporal dimensions of modal accessibility disparity in Boston and San Francisco. *Environment and planning. A*, 41(1):183.

[17] Kawabata, M. and Shen, Q. (2007). Commuting inequality between cars and public transit: The case of the San Francisco Bay area, 1990-2000. *Urban Studies*, 44(9):1759.

[18] Korkmaz, T. and Krunz, M. (2001). Multi-constrained optimal path selection. In *Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 834–843. IEEE.

[19] Krizek, K. J., Iacono, M., El-Geneidy, A., Liao, C.-F., and Johns, R. (2009). Access to destinations: Application of accessibility measures for non-auto travel modes. Technical Report 2009-24, Minnesota Department of Transportation.

[20] Kuipers, F., Van Mieghem, P., Korkmaz, T., and Krunz, M. (2002). An overview of constraint-based path selection algorithms for qos routing. *Communications Magazine, IEEE*, 40(12):50–55.

[21] Lei, T. and Church, R. (2010). Mapping transit-based access: integrating gis, routes and schedules. *International Journal of Geographical Information Science*, 24(2):283–304.

[22] Levine, J., Grengs, J., Shen, Q., and Shen, Q. (2012). Does accessibility require density or speed? A comparison of fast versus close in getting where you want to go in us metropolitan regions. *Journal of the American Planning Association*, 78(2):157–172.

[23] Mavoa, S., Witten, K., McCreanor, T., and O'Sullivan, D. (2012). GIS based destination accessibility via public transit and walking in Auckland, New Zealand. *Journal of Transport Geography*, 20(1):15–22.

[24] McFadden, D. (1974a). *Conditional logit analysis of qualitative choice behavior*, pages 105–142. Academic Press.

[25] McFadden, D. (1974b). The measurement of urban travel demand. *Journal of Public Economics*, 3(4):303–328.

[26] McFadden, D., Johnson, M., and Reid, F. (1973). *Travel Demand Forecasting Study*. Institute of Urban and Regional Development, University of California.

[27] Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, 54(1):17–24.

[28] Metropolitan Council (2010). 2030 transportation policy plan.

[29] Morris, J. M., Dumble, P., and Wigan, M. R. (1979). Accessibility indicators for transport planning. *Transportation Research Part A: General*, 13(2):91–109.

[30] OpenStreetMap Foundation (2013). http://openstreetmap.org.

[31] OpenTripPlanner (2013). http://opentripplanner.org.

[32] Owen, A. and Levinson, D. (2012). Annual accessibility measure for the Twin Cities metropolitan area. Technical Report 2012-34, Minnesota Department of Transportation.

[33] Polzin, S. E., Pendyala, R. M., and Navari, S. (2002). Development of time-of-day-based transit accessibility analysis tool. *Transportation Research Record: Journal of the Transportation Research Board*, 1799(1):35–41.

[34] Shen, Q. (2006). Job accessibility as an indicator of auto-oriented urban structure: a comparison of Boston and Los Angeles with Tokyo. *Environment and Planning B: Planning and Design*, 33:115–130.

[35] Spear, B. D. (2011). Improving employment data for transportation planning. Technical report, American Association of State Highway and Transportation Officials.

[36] Taylor, B. D., Miller, D., Iseki, H., and Fink, C. (2009). Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A: Policy and Practice*, 43(1):60 – 77.

[37] Tilahun, N. and Levinson, D. (2011). Work and home location: Possible role of social networks. *Transportation Research Part A: Policy and Practice*, 45(4):323–331.

[38] Wu, B. M. and Hine, J. P. (2003). A PTAL approach to measuring changes in bus service accessibility. *Transport Policy*, 10(4):307–320.

[39] Zhu, S. (2010). *The Roads Taken: Theory and Evidence on Route Choice in the wake of the I-35W Mississippi River Bridge Collapse and Reconstruction*. PhD thesis, University of Minnesota.