The Components of Transit Time and Their Effects on Trip Mode Choice

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

Ryan Chien

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

Advisors:
Dr. Gerard McCullough

Dr. Greg Lindsey

Dr. Steve Miller

May 2017

Table of Contents

List of Tables

List of Figures

I. Introduction

Public transit, such as taking the bus or the train, is the fastest growing mode of transportation in the Twin Cities. From 2000 to 2010, the transit mode share grew from 2.4 to 3 percent of all trips, while the driving mode decreased from 87.5 to 84 percent of all trips (Metropolitan Council 2015, 26).

The Metropolitan Council forecasts that ridership will roughly double by 2030 (Metropolitan Council 2015, 12.17). An estimated investment of about $31 billion by the year 2040 will be leveraged into maintaining, operating, and moderately expanding the current transit system (Metropolitan Council 2015, 6.68). Up to an additional $18 billion will be spent on larger capital projects for bus and transit system expansion to support a competitive regional economy, dependent on new revenue streams (Metropolitan Council 2015, 6.69-6.70).

In this environmental context, it is crucial to understand why people take the bus, as it is the majority shareholder for transit modes. Bus trips accounted for 70 percent of all transit trips in 2016 (Metro Transit 2017).

The determinants of transit demand and corresponding bus demand are well studied, dating to at least 1974 in, “The Measurement of Urban Travel Demand,” by Daniel McFadden. McFadden identified on-vehicle time, not separating between transit and automobile, and cost per trip as significant determinants of mode share between driving and taking the bus (McFadden 1974, 319). Bus walk time (OVTTwalk), wait time (OVTTwait), and fare interactions were also explored (McFadden 1974, 319).

This early work shows that bus trips are multimodal. They are combinations of walking, waiting, and transit within the vehicle. More recently in 2006, Iseki, Taylor, and Miller conducted a review on transit literature, and found that OVTTwalk, OVTTwait, and in-vehicle transit time (IVTT) represent majority components of the transit choice decision (Iseki, Taylor, and Miller 2006, 8-10). However, the treatment of each of these components is inconsistent across studies. Almost all transit studies are dependent on revealed preference data. Therefore, observations on driving trips are missing OVTTwalk, OVTTwait, and bus in-vehicle transit time (IVTTbus). Likewise, transit trips are missing drive times (IVTTcar). This is the problem of undefined nonselected alternatives (UNA) (Guo and Wilson 2004, 2).

This presents challenges for estimation as each UNA must be collected or proxied. One solution is to leave them out altogether. Another is to apply assumptions, a common one being that travelers value OVTTwalk and OVTTwait equally. The result of either is the same, a loss of understanding between the various components. Therefore, the primary motivation of this paper is to estimate and study the separate components of bus trip time, and compare them to the components of car trip time, regarding their effects on mode share.

II. Literature Review

Marc Gaudry used monthly time series data from 1956 to 1971 in his analysis of urban transit demand in Montreal. OVTTwalk and IVTTbus are averages within months, however they vary across months (Gaudry 1975, 249). He found that average OVTTwalk and average IVTTbus were significant and negatively affected the likelihood of a transit trip (Gaudry 1975, 253-254). IVTTcar was found to be less significant and positively affected the likelihood of a transit trip (Gaudry 1975, 253-254). The elasticity of transit demand with respect to OVTTwait was found to be about double that of IVTTbus, -0.54 and -0.27 respectively (Gaudry 1975, 254). The elasticity of transit demand with respect to IVTTcar was estimated at 0.42 (Gaudry 1975, 254).

Edward Beimborn, Michael Greenwald, and Xia Jin used EMME/2 transit analysis software and the 1994 Portland Metropolitan Services District (Portland Metro) origin-destination (OD) survey data to estimate IVTTcar, IVTTbus, and the sum of walk and wait time called out-of-vehicle transit time (OVTTsum) for each observation (Beimborn, Greenwald, and Jin 2003, 4-5). EMME/2 IVTTcar data was estimated using a shortest-path algorithm (Beimborn 2003, 5), and IVTTbus and OVTTsum were estimated using a minimum-time algorithm (Korhonen, Kangas, and Pursula 1997, 2). Each variable varied across all observations. The authors then estimated a binary logit model between choice of transit versus car. They found that OVTTwalk and OVTTwait were significant and negatively affected the likelihood of choosing transit (Beimborn 2003, 7). The coefficient of OVTTwalk was -0.195, and OVTTwait was -0.118 (Beimborn 2003, 7). Surprisingly IVTTbus and IVTTcar were not found to be significant determinants of transit demand (Beimborn 2003, 7).

Zhan Guo and Nigel Wilson used unspecified geographical information software (GIS) and the 2004 OD survey data from downtown Boston to estimate OVTTwalk, OVTTwait, and in-vehicle transit time for light rail (IVTTrail) for each observation (Guo 2004, 13). It is notable that OVTTwait was calculated as half of the headway time of the subway line (Guo 2004, 13). Therefore, it did not vary within trip observations that used the same subway line. They estimated a binary logit model between subway transit and walking. The authors found that OVTTwalk and OVTTwait were significant and negatively impacted transit demand, with respective coefficients of -1.13 and -0.16 (Guo 2004, 15). IVTTrail was also significant with a coefficient of -0.20 (Guo 2004, 15).

Rongfang Liu, Ram Pendyala, and Steven Polzin used stated preference survey data to estimate a binary logit model comparing light rail transit to driving. IVTT and OVTTsum varied across all observations. It is notable that the authors chose to treat the coefficient of IVTTrail and IVTTcar as equivalent and independent (Liu, Pendyala, and Polzin 1997, 76). They found that IVTT and OVTT were significant and negatively impacted transit demand, with respective coefficients of -0.035 and -0.052 (Liu 1997, 77).

In summary, a multitude of approaches have been tried to accommodate the UNA problem. Assumptions are often used to facilitate estimation. McFadden and Liu assumed that the coefficients of IVTT for the transit mode and IVTTcar were equivalent and independent. Liu further assumed that the coefficients of OVTTwalk and OVTTwait were equivalent. Gaudry assumed homogeneity for OVTTwait and IVTTbus within months, and that the coefficient for OVTTwalk was zero. Wilson assumed that OVTTwait was homogenous within subway lines. The exception is the Beimborn, Greenwald, and Jin study.

III. Research Problem

The primary problem is that UNAs exist for all observations. Previous methods used to estimate UNAs generally impose some assumptions on the time components. Their separate effects are thus obscured. The secondary problems are potentially reduced goodness of fit and attenuation bias of the remaining variables (Cramer 2005).

This study estimates the determinants of transit demand in the Twin Cities, with a focus on OVTTwalk, OVTTwait, IVTTbus, and IVTTcar. A novel use of GIS software is implemented, which imposes fewer assumptions upon the variables than previous studies. The resulting model is then restricted following the assumptions of previous models when possible. The effect on the remaining coefficients and goodness of fit is analyzed.

Finally, an analysis on separate effects of the time components of transit is conducted. This gives transit policymakers additional insight into appropriate actions to support their goals of meeting transit demand.

IV. Model

A. Considerations

Transit demand can be represented by mode share, and its foundation is individual mode choice. This can be modeled using the logit framework, assuming the utility of each mode can be represented by their characteristics, the characteristics of the user, and that the disturbances are random due to incomplete information. The mode with the greater utility is always chosen.

Regarding characteristics, attitude surveys have revealed that, “When asked specifically about the factors they consider when deciding what form of transport to use, people are most likely to mention convenience (67%) and journey time (47%) (Anderson and Stradling 2004, iii).” A control panel of household and individual variables is also needed. Commonly cited reasons for choosing the bus include not being able to drive, and not being able to afford the cost of driving (Anderson and Stradling 2004, 6). Unfortunately, due to data constraints, cost is not included in this study.

Following these considerations, I arrive at the following general specification. The individual difference in the utilities $U(bus)_i - U(car)_i$ is:

$$\begin{aligned} \propto \ &+ \beta_A(Time\ Components\ of\ Transit)_i + \beta_B(Time\ Components\ of\ Driving)_i \\ &+ \beta_C(Transit\ Quality)_i + \beta_D(Driving\ Quality)_i \\ &+ \beta_E(Household\ and\ Individual\ Characteristics)_i + \varepsilon_i \end{aligned}$$

The time components of transit describe the time it takes a traveler to move from an origin to a destination using the bus. Examples are: how much time is spent moving within the bus, walking to the bus, waiting for the bus, waiting within the bus while taking on or dropping off passengers, and time spent transferring from one bus to another. Each should negatively impact the utility of taking the bus, and reduce the probability of choosing the bus.

The time components of driving describe the time it takes a traveler to move from an origin to a destination using a car. Examples are: how much time is spent driving and waiting for traffic. Walking time is not included since it is assumed to be zero. Each should negatively affect the utility of the car, increasing the likelihood the bus is chosen.

Transit quality describe the subjective experience of riding the bus. Examples are: vehicle characteristics such as spaciousness, cleanliness, and comfort. Characteristics of bus stations such as safety, schedule clarity, and protection from weather are also included. Characteristics perceived as good to the bus experience should positively affect the utility of the bus. Bad characteristics such as lack of weather protection should decrease its utility. Good characteristics should increase the probability of choosing the bus.

Driving quality describe the subjective experience of driving. Examples are: the quality of the road surface and parking availability. Good characteristics such as easy parking should increase the car's utility, while bad characteristics such as potholes should decrease it. Good characteristics should decrease the probability of choosing the bus.

Finally, examples of individual and household characteristics include: if the trip taker has a driver's license, age, and household income. A crucial control variable is if the household owns a car. The Metropolitan Council estimates that, "31 percent of transit riders, or about 87,600 travelers, in the region do not own a car (Metropolitan Council 2015, 27)." The effect of these characteristics upon the utility of the car and the bus are expected to be mixed.

B. Variables

Variables were chosen to describe each group based on availability and relevance. The dependent is:

- Tookbus - binary set to 1 if the observation was a bus trip.

The time components of transit are:

- IVTTbus - minutes spent travelling within a bus for a bus trip;
- OVTTwalk - minutes spent walking to, between, and from bus stops during a bus trip;
- OVTTwait - minutes spent waiting at a bus stop during a bus trip;
- nstop - maximum number of stops possible during a bus trip, and;
- boardings - number of boardings during a bus trip.

The time component of driving is:

- IVTTcar - minutes spent driving during a car trip.

No variables were available to describe the components of transit quality or the components of driving quality.

Finally, the control panel of household and individual characteristics are:

- hhincome - scale of household income from 1-18, from least to greatest in $5,000 increments from $5,000-$50,000 corresponding from 1-10, and nonlinear thereafter;
- hhsize - number of people over the age of 5 in a household;
- ncars - number of cars owned by the household;
- ndrivers - number of drivers in a household;
- disabled - binary variable set to 1 if the trip taker is disabled;
- age - scale of the trip taker's age from 1-11, least to greatest;
- educ - scale of education from 1-8, from daycare/pre-school to a postgraduate degree;
- student - binary set to 1 if the trip taker is a current student;
- gender - binary set to 1 if the trip taker is male;
- dwork - binary set to 1 if the trip destination is work;
- owork - binary set to 1 if the trip origin is work, and;
- license - binary set to 1 if the trip taker has a driver's license.

C. Logit Framework for Binary Choice

A binary logit, random utility model is chosen to estimate the specification. The utility of an individual trip mode is modeled as a linear function of the independent variables and a random (due to incomplete information) error term:

$$U(mode)_i = X_i\beta + \mathcal{E}_i$$

The trip taker chooses the bus, $Y_i = 1$, if the utility for taking the bus is greater than the utility of taking the car:

$$Y_i = 1 \;\; iff \;\; U(bus)_i > U(car)_i$$

$$Y_i = 1 \; iff \; U(bus)_i - U(car)_i > 0$$

$$Y_i = 1 \; iff \; \mathcal{E}_i > 0$$

Assuming the error term is takes the form of a logistic distribution, the probability of a trip taker choosing the bus is:

$$P[Y = 1] = \frac{1}{1+e^{-x_i'\beta}}$$

D. Maximum Likelihood Method for Estimating Logit

The maximum likelihood method is chosen to estimate to nonlinear logit function. The likelihood function across N observations is:

$$L(\beta|data) \;=\; \prod_{i=1}^{N}\left(\frac{1}{1+e^{-x_i'\beta}}\right)^{y_i}\left(1-\frac{1}{1+e^{-x_i'\beta}}\right)^{1-y_i}$$

The log likelihood function which is more convenient for maximization is:

$$\ln(L) = \sum_{i=1}^{N} y_i \ln\left(\frac{1}{1+e^{-x_i'\beta}}\right) + (1-y_i)\left(1-\frac{1}{1+e^{-x_i'\beta}}\right)$$

Taking the derivative with respect to β gives the gradient:

$$g = \frac{\partial \ln(L)}{\partial\beta} = \sum_{i=1}^{N}\left(y_i - \frac{1}{1+e^{-x_i'\beta}}\right)x_i$$

Taking the second derivative gives the Hessian:

$$H = \frac{\partial^2 \ln(L)}{\partial\beta\beta\prime} = -\sum_{i=1}^{N}\left(\frac{1}{1+e^{-x_i'\beta}}\right)\left(1-\frac{1}{1+e^{-x_i'\beta}}\right)x_i x_i'$$

Since the Hessian is negative definite, the Log Likelihood is globally concave and a global maximizer exists. We can then use the Newton-Raphson Method to iteratively move our β estimate closer to the true value:

$$\beta_{t+1} = \beta_t + \left[\sum_{i=1}^{N} \left(\frac{1}{1+e^{-x_i'\beta}}\right)\left(1-\frac{1}{1+e^{-x_i'\beta}}\right)x_i x_i'\right]^{-1} \sum_{i=1}^{N}(y_i - \frac{1}{1+e^{-x_i'\beta}})x_i$$

And the marginal effects of each variable at the sample means are:

$$\frac{\partial\beta}{\partial x_{means}} = \left[1-2\left(\frac{1}{1+e^{-X_i\beta}}\right)\right]\left(\frac{1}{1+e^{-X_i\beta}}\right)\left[1-\left(\frac{1}{1+e^{-X_i\beta}}\right)\right]$$

The final specification based on the available variables is:

$$\begin{aligned} U(bus) - U(car) = \\ & \propto + \beta_1 IVTTbus + \beta_2 OVTTwalk + \beta_3 OVTTwait + \beta_4 nstop + \beta_5 boardings \\ & + \beta_6 IVTTcar + \beta_7 hhincome + \beta_8 hhsize + \beta_9 ncars + \beta_{10} ndrivers \\ & + \beta_{11} disabled + \beta_{12} age + \beta_{13} educ + \beta_{14} student + \beta_{15} gender + \beta_{16} dwork \\ & + \beta_{17} owork + \beta_{18} license + \varepsilon \end{aligned}$$

(The derivations above are taken liberally from *Econometric Analysis, Greene 1990,* and *Discrete Choice Analysis, Ben-Akiva and Lerman 1985*.)

E. Software

R version 1.0.136 is the statistical software used (R Core Team, 2016). It is an open-source version of S, which was originally developed by Bell Labratories in the 1990s (Venables and Ripley 2002, v).

Data formatting is handled with the data.table package (Dowle, Srinivasan, Short, and Lianoglou, 2015). Summary statistics are calculated with the pastecs package (Grosjean and Ibanez, 2014). The glm function in the stats package estimates the binary logit model using maximum likelihood (R Core Team, 2016). For thorough documentation on this package, see Venables and Ripley (2002). The Fisher Scoring variant of the Newton-Raphson Method iterates the values of β (Ripley 2002, 186). This method replaces the Hessian with the expected value of the Hessian (Greene 1990, 1100).

Heteroskedasticity robust standard errors are calculated using the vcovHC function in the sandwich package with the HC3 method (Zeileis, 2004). Marginal effects at the sample means are calculated using the predict function in the stats package (R Core Team, 2016). Standard errors for marginal effects are calculated using a bootstrap method with 1000 simulations. Details for this procedure can be found in Fernihough (2011).

The Firth method for bias reduction in cases of quasi-perfect separation is implemented through the logistf package (Heinze and Ploner, 2016). For more details, see Firth (1993).

V. Data

A. 2010 Travel Behavior Inventory

The 2010 Travel Behavior Inventory (TBI) is the cornerstone dataset for this analysis. The survey was conducted from November 2011 to early 2013 (Cambridge Systematics 2014, 1-1). Survey participants described each trip they took for a designated 24-hour period (Cambridge Systematics 2014, 1-3). Participants also provided a socioeconomic profile. An example is provided in Appendix A1.

In total 10,362 households and 21,298 individuals provided usable information. 79,236 trip observations were collected. This was composed of 69,057 car trips and 1,228 bus trips, with the remaining trips split between walking, cycling, school bus, light rail, taxi, motorcycle, ambulance, dial-a-ride, private bus, dial-a-ride, and ‘other’. Only car and bus trips are included in the study dataset, and therefore the logit model estimated is based upon the conditional probability of choosing between the car or bus mode.

The primary purpose of this dataset is threefold. First, it provides OD information for each car and bus trip. Second, it provides the date and time at which each trip began. OD and time data are used by the GIS software to generate estimates of IVTTbus, IVTTcar, OVTTwalk, and OVTTwait (also referred to as the ‘time components,’ or the ‘UNA variables’). Third, it provides a panel of socioeconomic controls.

B. General Transit Feed Specifications (GTFS)

General Transit Feed Specifications are text datafiles that represent public transit routes, stops, schedules, and associated geographic information (Google 2016). Two sources of GTFS data were used for this analysis. The first is the Metro Transit Schedule Data GTFS. This set describes all public transit within the metropolitan

Minneapolis and St. Paul area (Metropolitan Council 2016). The second is the Minnesota Valley Transit Authority GTFS. This set describes Apple Valley, Burnsville, Eagan, Rosemont in Dakota County, Savage, Prior Lake, and Shakopee in Scott County (Minnesota Valley Transit Authority 2015). An example is provided in Appendix A2.

The primary purpose of the GTFS datasets are to provide accurate bus routes, schedules, and stops. These are used by the GIS software to generate estimates of the UNA variables.

C. 2012 Minnesota Roads Shapefile

The 2012 Minnesota Roads shapefile represents all road centerlines within the state of Minnesota (Minnesota Department of Transportation 2012). This shapefile is used by the GIS software to provide linkages between bus stops and destinations, ultimately to estimate the UNA variables. An example is provided in Appendix A3.

D. Estimated UNAs

ArcGIS (ESRI, 2015) and Open Source Routing Machine (OSRM) (OSRM, 2017) software processed data from the TBI, GTFS feeds, and Road Shapefile to estimate UNA variables for all observations. These estimates were then used as data for estimating the logit model. They are proxies for the true UNA variable values. For a detailed discussion on UNA variable estimation, see Appendix B1 and B2.

F. Final Data

Data from the TBI, ArcGIS UNA estimator, and OSRM UNA estimator were merged into the final dataset as depicted by Appendix B5. From ArcGIS came estimates of IVTTbus, OVTTwalk, OVTTwait, number of stops, and boardings. OSRM provided IVTTcar. Twelve variables describing individual and household characteristics came directly from the TBI.

Some final data cleaning steps were undertaken at this point. Observations with OD coordinates beyond the GTFS range of latitude 44.471 to 45.415, and longitude -94.012 to -92.732 were removed, as bus stops are not generated beyond these coordinate lines. Erroneous ArcGIS estimates with negative transit times were removed, for details on why these occur see Appendix B3. 62,442 observations remained in the final dataset. 44,124 were complete while 18,318 were incomplete, as some observations had missing variables due to survey participants’ refusal to answer certain questions, usually regarding income, age, and if they were a licensed driver. Therefore, the final dataset has 44,124 observations.

E. Summary Statistics

Basic summary statistics are presented in Table 1. Of interest are the comparative magnitudes of IVTTbus, OVTTwalk, and IVTTcar. IVTTbus and IVTTcar have nearly the same means at 14.130 and 14.133 respectively. The mean of OVTTwalk is very high however, at 65.176. This is certainly related to the estimation problems experienced using ArcGIS, which are further discussed in Appendix B3.

Table 1: Summary Statistics

	tookbus	IVTTbus	OVTTwalk	OVTTwait	nstop	boardings	IVTTcar
# null	43196	14435	51	19675	19652	19652	0
min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0567
max	1.0000	144.0077	826.5027	222.1600	235.0000	13.0000	109.5933
mean	0.0210	14.1301	65.1759	3.6892	16.8109	1.3631	14.1330
std. dev.	0.1435	20.9298	76.4386	8.1107	26.2590	1.6860	11.3233

	hh_income	hh_size	ncars	ndrivers	disabled	educ	age
# null	0	0	507	270	42389	0	0
min	1.0000	1.0000	0.0000	0.0000	0.0000	1.0000	3.0000
max	18.0000	10.0000	10.0000	7.0000	3.0000	8.0000	11.0000
mean	12.3184	2.4946	2.1406	1.9845	0.0430	6.0941	7.1209
std. dev.	3.3097	1.1917	1.0174	0.7470	0.2234	1.8708	1.6201

	student	gender	dwork	owork	license
# null	40478	24500	37012	37054	1052
min	0.0000	0.0000	0.0000	0.0000	0.0000
max	1.0000	1.0000	1.0000	1.0000	1.0000
mean	0.0826	0.4447	0.1612	0.1602	0.9762
std. dev.	0.2753	0.4969	0.3677	0.3668	0.1526

F. Correlation Coefficient Matrix

Selected correlation coefficients are presented in Table 2 to illustrate the high degree of correlation between the time components of transit and the time components of driving. Multicollinearity inflates the variance of the estimator which can lead to coefficient attenuation toward zero (Cramer 2005).

Table 2: Selected Correlation Coefficients

	IVTTbus	OVTTwalk	OVTTwait	nstop	boardings	IVTTcar
IVTTbus	1.0000	0.1513	0.5921	0.8694	0.8409	0.7054
OVTTwalk	0.1513	1.0000	0.1104	0.1554	0.0501	0.7140
OVTTwait	0.5921	0.1104	1.0000	0.3676	0.5487	0.4760
nstop	0.8694	0.1554	0.3676	1.0000	0.7514	0.5932
boardings	0.8409	0.0501	0.5487	0.7514	1.0000	0.5648
IVTTcar	0.7054	0.7140	0.4760	0.5932	0.5648	1.0000

VI. Results

A. Introduction

Coefficients and marginal effects for the model, henceforth named Model 1, are presented in Table 3. Goodness of fit is discussed in section 6B. New findings are discussed in detail in Section 6C. Findings that confirm robustness that are less noteworthy are discussed in section 6D.

Table 3. Coefficient and Marginal Effects Estimates for Model 1

		Coefficients			Marginal Effects		
		Estimate	Std. Error	Pr(>|z|)	Estimate	Std. Error	Pr(>|z|)
	(Intercept)	2.1416	0.3022	0.0000	0.0340	0.0048	0.0000
Estimated UNAs	IVTTbus	-0.0296	0.0071	0.0000	-0.0005	0.0001	0.0000
	OVTTwalk	-0.0696	0.0070	0.0000	-0.0011	0.0001	0.0000
	OVTTwait	-0.0774	0.0120	0.0000	-0.0012	0.0002	0.0000
	nstop	-0.0260	0.0037	0.0000	-0.0004	0.0001	0.0000
	boardings	0.3144	0.0364	0.0000	0.0050	0.0006	0.0000
	IVTTcar	0.1637	0.0130	0.0000	0.0026	0.0002	0.0000
Direct TBI Variables	hh_income	-0.1365	0.0113	0.0000	-0.0022	0.0002	0.0000
	hh_size	-0.2569	0.0788	0.0006	-0.0041	0.0013	0.0008
	ncars	-1.4583	0.1132	0.0000	-0.0232	0.0018	0.0000
	ndrivers	0.9892	0.1157	0.0000	0.0157	0.0018	0.0000
	disabled	-0.0274	0.1791	0.4392	-0.0004	0.0029	0.4396
	educ	0.0735	0.0221	0.0004	0.0012	0.0004	0.0007
	age	-0.2020	0.0233	0.0000	-0.0032	0.0004	0.0000
	student	0.4075	0.1340	0.0012	0.0065	0.0021	0.0011
	gender	0.2865	0.0770	0.0001	0.0046	0.0012	0.0001
	dwork	1.2726	0.0911	0.0000	0.0202	0.0015	0.0000
	owork	0.9494	0.0921	0.0000	0.0151	0.0015	0.0000
	license	-2.4658	0.1679	0.0000	-0.0392	0.0027	0.0000

Almost all the significant coefficients are in the expected directions. However, the coefficient of boardings is positive, although it is expected to be negative. Transfers are documented to be onerous. An explanation for this inconsistency is the inclusion of walk only trips in the model. Unavailability of transit is one reason these occur (this is further discussed in Appendix B3). When these occur, the boardings variable is set to zero. Therefore, the boardings variable also picks up the effect of the availability of transit. Availability of transit will increase the likelihood of choosing transit, explaining the positive coefficient.

It is also worth noting that the coefficient of IVTTcar is quite large in comparison to the other time components. Although part of this result is certainly structural, it may be overestimated. As noted in Appendix B4, OSRM is not aware of congestion. Almost all the IVTTcar observations are therefore underestimated, and peak-hour observations are likely severely underestimated. This would cause the coefficient to be overestimated in absolute value. Improvement of the OSRM process would likely reduce the magnitude of the coefficient.

B. Goodness of Fit and Model Choice

Two goodness of fit measures are used. The first is the Likelihood Ratio (LR). This compares the goodness of fit of the fully specified model, versus the null model with only a constant. The test statistic is calculated as $-2(L_{model} - L_{null})$, and its critical value is chi-squared distributed with degrees of freedom equal to the number of model coefficients minus one. The null hypothesis is: the fully specified model does not fit significantly better than the null model, and therefore the explanatory variables have no statistically relevant effects on the dependent variable. The null hypothesis is rejected if the test statistic is larger than the critical value.

The second is adjusted rho-squared, which compares the goodness of fit between the fully specified model and the null model. It is calculated as $1 - (L_{model} - K)/L_{null}$. This value is useful for comparison between different model specifications. A larger value indicates a better goodness of fit.

Table 4 presents the LR and adjusted rho-squared for three models. Model 1 is fully specified using all estimated UNA variables separately. Model 2 assumes that the coefficients of IVTTcar and IVTTbus are equivalent and independent. This assumption is found in previous studies such as McFadden (1974), Bhat (1995), and Liu (1997). Model 3 assumes that the coefficients of OVTTwalk and OVTTwait are equivalent, but does not use the Model 2 assumption. This assumption is found in previous studies such as Bhat (1995) and Liu (1997). Details on model specification are found in Appendix B7.

Table 4: Comparison of Goodness of Fit

	Model 1	Model 2	Model 3
Log Lik.	-2790.35	-2252.61	-2790.64
Null Log Lik.	-4501.86	-4501.86	-4501.86
LR Test Val.	3423.02	4498.50	3422.42
K	19	18	18
N	44124	44149	44124
adjrho2	0.3760	0.4956	0.3761
iterations	10	11	10

The log likelihood of Model 1 is -2790.35, and the null log likelihood for all three models is -4501.86. The LR test value is 3423.02 which is larger than the 0.05 percent critical value of 45.97, and the null hypothesis is rejected. The adjusted rho-squared is 0.3760.

The log likelihood of Model 2 is -2252.61. The LR test value is 4498.50, the 0.05 percent critical value is 44.43, and the null hypothesis is rejected. The adjusted rho-squared is 0.4956.

The log likelihood of Model 3 is -2790.64. The LR test value is 3422.42, the 0.05 percent critical value is 44.43, and the null hypothesis is rejected. The adjusted rho-squared is 0.3761.

Each model fits significantly better than the null, therefore none are rejected outright from consideration. Model 2 gives the best fit, followed by Model 1, and then Model 3. However, goodness of fit is not the only criterion for model selection. This study was conducted to analyze the separate effects of IVTTbus, IVTTcar, OVTTwalk, and OVTTwait. This is only possible with Model 1. Therefore, it is selected even though it has the worst adjusted rho-squared. Its fit is good, almost all the explanatory variables are significant, and it presents the best opportunity to understand the separate components.

C. New Findings

This study finds five new results. First, the effects of IVTTcar and IVTTbus are not the same. The Wald test statistic is 104.200 with one degree of freedom, and the 0.01 percent critical value is 15.137. The null hypothesis that the coefficients of IVTTcar and IVTTbus are equivalent is rejected.

This result indicates that Twin Cities travelers likely find driving time more onerous that transit time when choosing between the two modes. Each additional minute of IVTTcar increases the probability of choosing the bus by 0.26 percent at the means. Comparatively, each additional minute of IVTTbus reduces that probability by 0.05 percent.

Model 1 predicts that the bus mode share is 2.10 percent, given the current IVTTcar values. To illustrate the effect of IVTTcar upon mode share, imagine that it is increased by 10 percent across the sample. This hypothetical scenario represents increased traffic congestion due to population growth. All other variables remaining equal, Model 1 predicts that bus mode share would increase to 2.46 percent. Table 5 illustrates the increase in bus mode share given a 10, 20, and 30 percent increase in IVTTcar. See Table 6 for an illustration of the increase in bus mode share given a decrease in IVTTbus; the increase in bus mode share is lower at each level.

Table 5: Bus Mode Share Response to Increases in IVTTcar

Percent Increase	Bus Mode Share Response to: IVTTcar
0% Baseline	2.10%
10%	2.46%
20%	2.89%
30%	3.41%

This finding contradicts previous studies that have assumed there is no perceived difference (McFadden 1974 and Bhat 1995). A literature review in transit psychology reveals this finding has been well documented outside of economics. A study of transit users and drivers using Likert scale stress ratings as the explanatory variables found that driving was the most stressful mode (Legrain, Eluru, and El-Geneidy 2015, 148). However, this is not conclusive as another found transit to be the most stressful mode (Haider, Kerr, Badami 2013, 11), or to be about equal (Horowitz 1981).

These interstudy inconsistencies suggest that the coefficients of IVTTcar and IVTTbus are dependent on unobserved variables. The quality of a bus trip, such as its cleanliness and degree of crowding, should affect the magnitude of IVTTbus. This suggests that separating IVTTcar and IVTTbus is a feasible method to account for the unobserved quality variables in a mode choice model.

Second, each minute of OVTTwalk reduces the log likelihood of choosing the bus by -0.0696. The marginal effect is -0.11 percent. These pass significance testing at the 0.01 percent level. Table 6 illustrates the increase in bus mode share given decreases in OVTTwalk, OVTTwait, and IVTTbus.

Table 6: Bus Mode Share Response to Reductions in OVTTwalk, OVTTwait, and IVTTbus

	Bus Mode Share Response to:		
Percent Reduction	OVTTwalk	OVTTwait	IVTTbus
0% Baseline	2.10%	2.10%	2.10%
10%	2.35%	2.15%	2.19%
20%	2.64%	2.20%	2.28%
30%	3.00%	2.25%	2.38%

Third, the effects of OVTTwalk and OVTTwait are not the same. The null hypothesis that their coefficients are equal can be tested using the Wald method (Greene 1190, 119). The resulting test statistic is 26.8 with one degree of freedom, and the 0.01 percent critical value is 15.137. The null hypothesis is rejected.

These results indicate that OVTTwait is more onerous than OVTTwalk for Twin Cities travelers. Marginal effects presented in Table 3 show that for the average individual, each additional minute of OVTTwait reduces the probability of choosing the bus by 0.12 percent. In comparison, each additional minute of OVTTwalk reduces that probability by 0.11 percent. Illustrated via bus mode share, a 10 percent decrease in OVTTwait will only increase the predicted share from 2.10 to 2.15 percent.

This contradicts previous studies that have assumed there is no perceived difference (Bhat 1995 and Liu 1997). However, previous studies that have included their separate effects show inconsistent results. Some find that OVTTwalk is the more onerous option (Beimborn 2003 and Guo 2004), while others find it to be OVTTwait (McFadden 1974). This suggests that OVTTwalk and OVTTwait are dependent on quality variables not observed in this study. For example, better weather protection at bus stops would likely improve their subjective quality experience. In turn, the OVTTwait coefficient should increase, becoming less negative.

This finding also suggests that the separation of OVTTwalk and OVTTwait allow the capture of their quality characteristics within the time variables. A resulting hypothesis is that OVTTwalk and OVTTwait would be equivalent if all relevant quality characteristics were captured within the model.

Fourth, the effects of OVTTwalk and OVTTwait are more impactful on bus demand than IVTTbus. The coefficients of OVTTwalk and OVTTwait are -0.0696 and -0.0774 respectively. These are more than double the coefficient of IVTTbus, which is -0.0296. See Table 6 for mode share forecasts.

This indicates that travelers within the Twin Cities find walk and wait time more onerous than transit time. This is fully consistent with most revealed preference studies cited by this paper (Gaudry 1975, Bhat 1995, Liu 1997, Beimborn 2003), or partially consistent regarding wait time (McFadden 1974), or walk time (Guo 2004). Previous stated preference studies are also consistent with this finding. A review of 183 British studies found that one minute of OVTTwalk was equivalent to between 1.67 to 2.02 minutes of IVTTbus (Balcombe 2004, 72). Furthermore, one minute of OVTTwait was equivalent to about 1.59 minutes of in vehicle time (Balcome 2004, 77).

Table 7 summarizes the first four findings, presenting previous studies' estimates of the time components of transit compared to this study. There are some differences in method worth noting. McFadden's variables are interactions multiplied by wage, Gaudry's estimates are elasticities, Liu models choice between light rail transit versus car, and Guo models choice between walking and light rail transit. However, the context of previous models is still valuable, and the variables of interest are present.

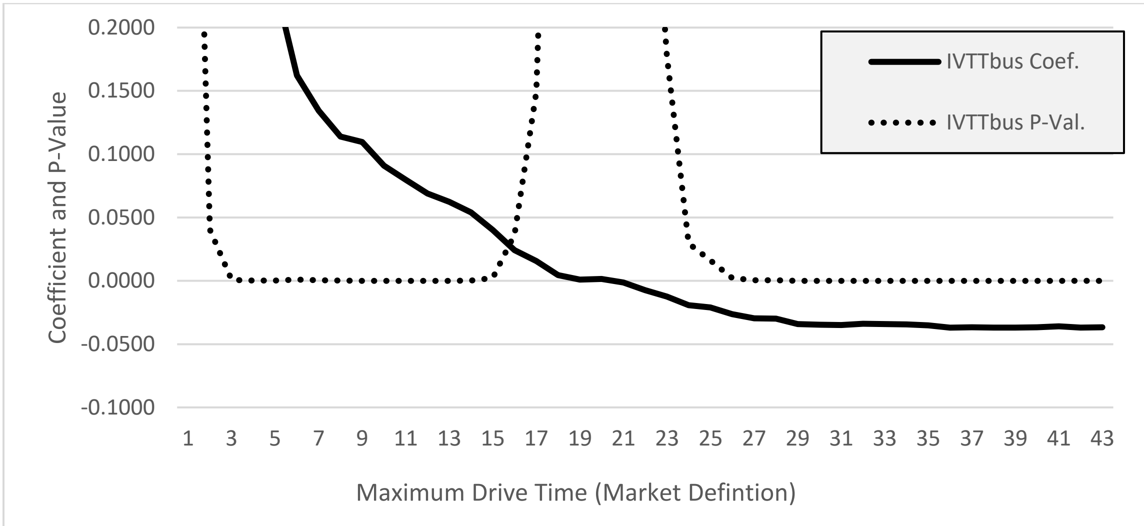
Table 7: Comparison of Time Component Coefficients

Author	McFadden	Gaudry (Elasticities)	Bhat	Liu, Pendyala, and Polzin	Beimborn, Greenwald, and Jin	Guo and Wilson	This study, Model 1
Year	1974	1975	1995	1997	2003	2004	2017
Area	San Francisco Bay Area	Montreal	Intercity, Toronto-Montreal	New York and New Jersey	Portland, OR	Boston	Minneapolis and St. Paul
OVTTwalk	-0.0001				-0.1953	-1.1300	-0.0696
OVTTwait	-0.0171	-0.5400			-0.1180	-0.1600	-0.0774
OVTTsum			-0.0359	-0.0520			
IVTTtransit		-0.2700			0.0011*	-0.2100	-0.0296
IVTTcar		0.4200			-0.0362*		0.1637
IVTTpool	-0.0086		-0.0105	-0.0353			
Model Type	Logit	AR1	HcEV	Logit	Logit	Logit	Logit
Notes	* Indicates insignificant, AR1 is autoregressive order 1, and HcEV is heteroscedasticity extreme value.						

Fifth, the effect and significance of IVTTbus changes across market definitions. In Model 1, the market definition is all trips within the TBI. Short, medium, and long distance trips are included, and the drive time ranges from 0 to 109 minutes.

Figure 1 plots results when the effect of IVTTbus is examined across different market definitions. The X-axis represents the maximum drive time market definition. For example, at the X-axis value of 20, the market is transit with a drive time between 0-20 minutes. The Y-axis represents the coefficient and significance of IVTTbus in a logit model, given the market definition.

Figure 1: IVTTbus Coef. and P-Val. Across Varying Market Definitions, Walk-Only Trips Included

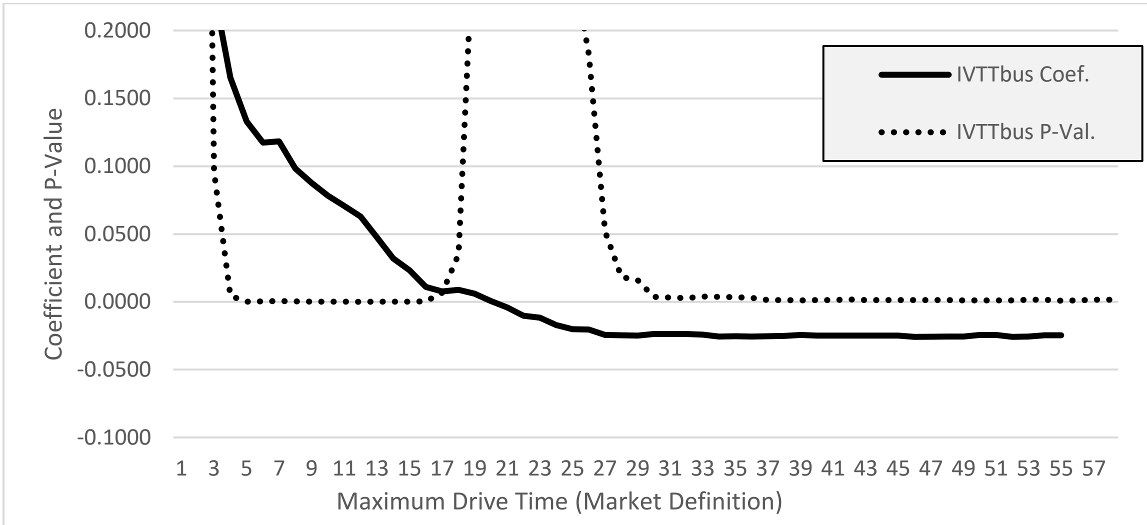


First, Figure 1 indicates that IVTTbus is not a significant determinant at the 0.05 percent level for trips that have a maximum drive time of between 15-26 minutes. For a table of values, see Appendix B6. This is intuitive. It is plausible that for medium-distance trips, the effects of IVTTbus are minimal or insignificant, and the main determinants are walk and wait time. IVTT is also found to be insignificant in Beimborn (2003).

Second, Figure 1 indicates that in a market where drive times are between 3-14 minutes, the IVTTbus coefficient is significant and positive. This suggests the surprising possibility the travelers find utility in bus transit time for short trips. This is suspicious however, and may be caused by walk-only observations where no transit observation is available. Therefore, some non-null IVTTbus values indicate availability of transit, which would make its use more likely.

This suspicion is tested by estimating a secondary regression with walk-only trips removed. The estimate of IVTTbus and its significance are shown in Figure 2. Surprisingly, the same pattern as Figure 1 is found. This indicates that the finding that there is utility in bus transit for short trips may be robust. However, multicollinearity between IVTTbus and boardings is still an issue, and discounts this finding.

Figure 2: IVTTbus Coef. and P-Val. Across Varying Market Definitions, Walk-Only Trips Removed



D. Other Findings

This study finds three results that are in line with previous studies, and confirm robustness. First, each minute of OVTTwait reduces the log likelihood of choosing the bus by -0.0774. The marginal effect is -0.12 percent. These pass significance testing at the 0.01 percent level.

Second, each minute of IVTTcar increases the log likelihood of choosing the bus by 0.1637. The marginal effect is 0.26 percent. These pass significance testing at the 0.01 percent level.

Third, each minute of IVTTbus decreases the log likelihood of choosing the bus by -0.0296. The marginal effect is 0.05 percent. These pass significance testing at the 0.01 percent level.

VII. Conclusion

Unobserved nonselected alternatives are troublesome in transit analysis. This study attempts to estimate them using a novel method, and compare them to previous work. A clearer understanding of the separate effects of each variable can help policymakers better plan future transit networks.

This study finds that for the Twin Cities metropolitan area, each additional minute of OVTTwalk and OVTTwait decrease the average individual's probability of choosing the bus by about 0.1 percent. Each additional minute of IVTTbus reduces that probability by about 0.05 percent. And the effect of IVTTcar is to increase that probability by about 0.26 percent. These weights can be used to model modal change in response to transit policy.

Two alternative models are estimated, and it is found that a pooled IVTT model (Model 2), and a summed OVTT model (Model 3), fit nearly as well or better than a model with separate components (Model 1). This result is surprising and confirms the general robustness of previous studies.

The separate effects of IVTTcar, IVTTbus, OVTTwalk, and OVTTwait are not available in these studies however. Furthermore, they often imply assumptions regarding the components. This study finds that such assumptions are not justified. Each should be treated differently. Another advantage is that estimating their separate effects allows the relative quality of each service to be captured by the time variable. This is useful for policymakers as it is impossible to capture all variables that describe the quality of an individual mode.

One potential application of this study is to improve accessibility metrics. Current metrics are based upon equivalent values of time, such as the number of jobs available within a 30 minute drive or transit trip from a centroid. This study suggests that time is not valued equally across modes. Therefore, weighted metrics based on equivalent time may be better representations of actual accessibility. This study also provides a method of determining the weights.

Regarding the technical method of this study, the UNA estimation process is only partially successful. Although the resulting variables are significant in Model 1, a closer analysis as presented in Appendix B3 and B4 reveals weaknesses in availability and accuracy. There are nearly 15,000 observations where IVTTbus or boardings are equal to zero. These occur when the UNA estimation process finds that transit is not possible for a given route (see details in Appendix B3). Thus, the problem of quasi-perfect separation is encountered during model coefficient estimation.

The Firth bias reduction method is used to check the robustness of Model 1 (Firth, 1993). The coefficients and significance values are similar, and do not discount the results. However, a better UNA estimation process is desirable. One solution is to use historic GTFS feeds instead of the 2016 versions. An archived GTFS from 2014 has been found, before the closing of bus route 50 and the opening of the Green Line, (see Appendix B3 for why this would be beneficial). Rebuilding the ArcGIS network dataset using this information may yield more accurate representations of IVTTbus, OVTTwalk, and OVTTwait.

Improvement of the ArcPy automation process may also yield more accurate representations of the components of bus transit time. The current issues are presented in Appendix B3. One idea is to program code that iterates start time over a range of times to find the closest approximation to reported travel time. Significant computing resources will be necessary for this code. Each individual iteration takes about 15 seconds to run on a high-end processor.

Finally, this study is limited to car and bus modes. It can be extended to include light rail transit bicycling, and walking. The ArcGIS solver can estimate their time components. These variables can be used to estimate a multinomial or nested logit model. Their inclusion would help the model would be more generalizable to overall transit demand forecasting.

VIII. Bibliography

Anderson, Simon and Stephen Stradling. 2004. *Attitudes Towards Car Use and Modal Shift in Scotland.* Scottish Executive Social Research, National Centre for Social Research (NatCen) Scotalnd, and Transport Research Institute, Napier University. http://www.gov.scot/Publications/2004/03/19062/34299.

Balcombe, R., R. Mackett, N. Paulley, J. Preston, J. Shires, H. Titheridge, M. Wardman, P. White. 2004. "The Demand for Public Transport: a Practical Guide." London: Transport Research Laboratory.

Beimborn, Edward, Michael Greenwald, and Xia Jin. 2003. "Accessibility, Connectivity, and Captivity: Impacts on Transit Choice." Transportation Research Record 1835, doi: 10.3141/1835-01.

Ben-Akiva, Moshe and Steven Lerman. 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge: MIT Press.

Bhat, Chandra. 1995. "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice." *Transportation Research Part B: Methodological* 26 no. 6: 471-483, doi: 10.1016/0191-2615(95)00015-6.

Cambridge Systematics, Abt SRBI, and PlanTrans. 2014. "Metropolitan Council Travel Behavior Inventory: Household Travel Survey: Final Report." https://metrocouncil.org/Transportation/Planning-2/Transit-Plans,-Studies-Reports/Other-Studies-Reports/Travel-Behavior-Inventory/TBI-Household-Survey-Report.aspx.

Cramer, JS. 2005. "Omitted Variables and Misspecified Disturbances in the Logit Model." Tinbergen Institute Discussion Papers. https://ideas.repec.org/p/tin/wpaper/20050084.html.

Dowle, M., A. Srinivasan, T. Short, S. Lianoglou with contributions from R. Saporta and E. Antonyan. 2015. data.table: Extension of Data.frame. R package version 1.9.6. https://CRAN.R-project.org/package=data.table.

ESRI. 2015. ArcGIS Desktop: Release 10.4.1. Redlands, CA: Environmental Systems Research Institute.

Firth, David. 1993. “Bias Reduction of Maximum Likelihood Estimates.” *Biometrika* 80, no. 1: 27-38. https://doi.org/10.1093/biomet/80.1.27.

Gaudry, Marc. 1975. “An Aggregate Time-Series Analysis of Urban Transit Demand: The Montreal Case.” *Transportation Research* 9, no. 4: 249-258, doi:10.1016/0041-1647(75)90066-0.

Google. 2016. GTFS Static Overview. https://developers.google.com/transit/gtfs/.

Greene, William. 1990. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Grosjean, Philippe and Frederic Ibanez. 2014. pastecs: Package for Analysis of Space-Time Ecological Series. R package version 1.3-18. https://CRAN.R-project.org/package=pastecs.

Guo, Zhan and Nigel Wilson. 2004. “Assessment of the Transfer Penalty for Transit Trips: Geographic Information System–Based Disaggregate Modeling Approach.” *Transportation Research Record* 1872, doi: 10.3141/1872-02.

Haider, Murtaza, Kenneth Kerr, and Madhav Badami. 2013. “Does Commuting Cause Stress? The Public Health Implications of Traffic Congestion.” *SSRN Electronic Journal,* doi: 10.2139/ssrn.2305010.

Heinze, Georg and Meinhard Ploner. 2016. logistf: Firth's Bias-Reduced Logistic Regression. R package version 1.22. https://CRAN.R-project.org/package=logistf.

Horowitz, Alan. 1981. “Subjective Value of Time in Bus Transit Travel.” *Transportation* 10, no. 149: 149-164, doi: 10.1007/BF00165263.

Iseki, Hiroyuki, Brian Taylor, and Mark Miller. 2006. ”The Effects of Out-of-Vehicle Time on Travel Behavior: Implications for Transit Transfers.” *Institute of Transportation Studies, University of California, Los Angeles, and California PATH Program, University of California, Berkley,* January 18. http://www.its.ucla.edu/wp-content/uploads/sites/6/2014/06/Appendix-A.pdf.

Kennedy, Peter. 2005. “On No! I Got the Wrong Sign! What Should I Do?” *The Journal of Economic Education* 36, no. 1: 77-92, doi: 10.3200/JECE.36.1.77-92.

Korhonen, Annu, Hannu Kangas, and Matti Pursula. 1997. “Back to the basics of EMME/2 transit assignment.” 6th European EMME/2 users’ group conference. https://www.inrosoftware.com/assets/pres-pap/european/eeug97/Europe02_1997.pdf

Legrain, Alexander, Naveen Eluru, and Ahmed El-Geneidy. “Am Stressed, Must Travel: The Relationship Between Mode Choice and Commuting Stress.” *Transportation Research Part F: Traffic Psychology and Behavior* 34: 141-151. http://www.sciencedirect.com/science/article/pii/S1369847815001370.

Liu, Rongfang, Ram Pendyala, and Steven Polzin. 1997. “Assessment of Intermodal Transfer Penalties Using Stated Preference Data.” *Transportation Research Record* 1607: 17-80, doi: 10.3141/1607-11.

McFadden, Daniel. 1974. “The Measurement of Urban Travel Demand.” *Journal of Public Economics* 3, no. 2: 303-328, doi: 10.1016/0047-2727(74)90003-6.

Metro Transit. 2017. *Metro Transit ridership tops 82.6 million in 2016.* https://www.metrotransit.org/metro-transit-ridership-tops-826-million-in-2016.

Metro Transit. 2014. The Last ride on Route 50. https://www.metrotransit.org/the-last-ride-on-route-50.

Metropolitan Council. 2016. Metro Transit Schedule Data - General Transit Feed Specification (GTFS). https://gisdata.mn.gov/dataset/us-mn-state-metc-trans-transit-schedule-google-fd.

Metropolitan Council. 2015. “Thrive MSP: 2040 Transportation Policy Plan.” January 25. https://metrocouncil.org/Transportation/Planning-2/Key-Transportation-Planning-Documents/Transportation-Policy-Plan/The-Adopted-2040-TPP-(1)/Final-2040-Transportation-Policy-Plan/2040-TPP-Complete.aspx.

Metropolitan Council. 2013. “Travel Behavior: The 2010 MSP Travel Behavior Inventory (TBI) Report Home Interview Survey, A Summary of Resident Travel in the Twin Cities Region.” October. https://metrocouncil.org/Transportation/Publications-And-Resources/Travel-Behavior-Inventory.aspx.

Minnesota Department of Transportation (MnDOT). Roads, Minnesota, 2012. https://gisdata.mn.gov/dataset/trans-roads-mndot-tis.

Minnesota Valley Transit Authority. 2015. Minnesota Valley Transit Authority - General Transit Feed Specification (GTFS). https://gisdata.mn.gov/dataset/com-mvta-trans-transit-schedule-google-fd.

Morang, Melinda. 2017. "Add GTFS to a Network Dataset User's Guide." *Yay, transit!* Accessed February 12, 2017. http://transit.melindamorang.com/UsersGuides/AddGTFStoaNetworkDataset/AddGTFStoND_UsersGuide.html.

OSRM. 2017. *OSRM Open Source Routing Machine*. Accessed February 13, 2017. http://project-osrm.org/.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Stradling, Stephen, Michael Carreno, Tom Rye, and Allyson Noble. 2007. "Passenger Perceptions and the Ideal Urban Bus Journey." *Transportation Policy* 14, no. 4: 283-292. http://dx.doi.org/10.1016/j.tranpol.2007.02.003.

Venables, W.N. and B.D. Ripley. 2002. *Modern Applied Statistics With S*. Verlag New York: Springer.

Zeileis, Achim. 2004. "Econometric Computing with HC and HAC Covariance Matrix Estimators." *Journal of Statistical Software 11*, no. 10, doi: 10.18637/jss.v011.i10.

IX. Appendix A

1. TBI Trip Journal Example

Travel: How did you get to Location 2? — DIARY EXAMPLE

1. What type(s) of transportation did you use to go to Location 2?

1st	→	2nd (if needed)	→	3rd (if needed)
1				

1 Car, van, truck	4 Public Bus	7 Amtrak	10 Taxi/Shuttle
2 Walk	5 Light Rail (Hiawatha)	8 Bicycle	11 Dial-A-Ride
3 School Bus	6 Commuter Rail (Northstar)	9 Motorcycle/Moped	12 Other (specify)____________

2. If you used a bus/train for this trip, did you use a pass? ☐ Yes ☐ No --→ How much did you pay?_____

3. If you used a car/van/truck or motorcycle/moped for this trip . . .

A. Were you the . . .? ☒ Driver ☐ Passenger

B. Was this vehicle from your household? ☒ Yes ☐ No

C. Did you pay a toll? ☐ Yes ☒ No

D. How much, in total, did you personally pay for parking? ☐ Nothing

$ ____ ____ . 2 5 Was the rate...? ☒ Hourly ☐ Daily ☐ Monthly ☐ Other

4. A. Including yourself, how many people were with you on this trip? 1 2 3 4+

B. Including yourself, how many were household members? 1 2 3 4+

C. Which household members were with you?

Michael , ______________ , ______________ , ______________

Location 2 — DIARY EXAMPLE

5. When did you arrive at Location 2? ___ 7 : 4 2 ☒ AM ☐ PM

6. Where is this? Anytown Daycare
Name of Location 2

If address already reported, provide location name and *GO TO QUESTION 7*

123 Main St
Street Address

Daycare
Type of Place or Business

Anytown, MN 55401
City, State, Zip Code

Main St & Elm Rd
Nearest Cross Streets

7. A. What was your primary activity at Location 2? (check only ONE box)

☐1 Home – Paid Work	☐8 Other School Activities	☐15 Recreation–Watch
☐2 Home – Unpaid Work	☐9 Quick Stops	☐16 Eat Out
☐3 Home – Other	☐10 Personal Business	☐17 Religious/Community
☐4 Work	☐11 Major Shopping	☐18 Accompany Another Person
☐5 Attend Childcare	☐12 Everyday Shopping	☐19 Pick-Up Passenger
☐6 Attend School	☐13 Social	☒20 Drop-Off Passenger
☐7 Attend College	☐14 Recreation–Participate	☐21 Turn Around

B. Other activities at Location 2, if any? _____ _____ _____

8. When did you leave Location 2? ___ 7 : 4 5 ☒ AM ☐ PM ☐ Did Not Leave

(Cambridge Systematics 2014, 2-18)

2. GTFS Example

Stop_times.txt

```
trip_id,arrival_time,departure_time,stop_id,stop_sequence,pickup_type,drop_off_type

9540880-AUG16-MVS-BUS-Weekday-01,07:52:00,07:52:00,6990,1,0,0
```

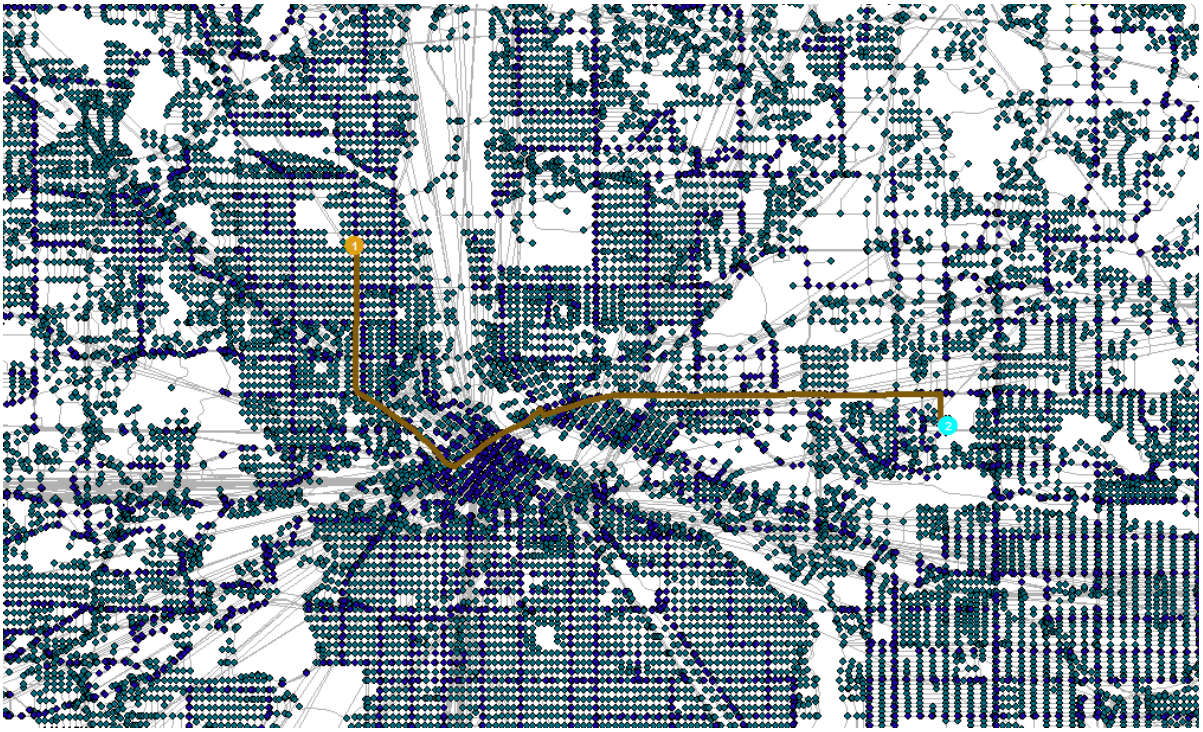
Routes.txt

```
route_id,agency_id,route_short_name,route_long_name,route_desc,route_type,route_url,rou
te_color,route_text_color

2-95,          0,2,"Franklin Av - Riverside Av - U of M - 8th St
SE",3,http://www.metrotransit.org/route/2,000000
```

3. 2012 Minnesota Road Shapefile Example

4. Example of a Successful ArcGIS Solver Run



5. ArcPy Code for Estimation of IVTT, Walk Time, and Wait Time in ArcGIS

```
# Import arcpy module
import arcpy

# Load required toolboxes
arcpy.ImportToolbox("C:/Program Files
(x86)/ArcGIS/Desktop10.4/AddGTFStoaNetworkDataset/AddGTFStoaNetworkDataset_0
_5_2_0/Transit Analysis Tools.tbx")

# set environment variables
arcpy.env.workspace = "C:\\Users\\Ryan
Chien\\Documents\\Minneapolis\\TwinCities3.gdb"
arcpy.env.overwriteOutput = True
arcpy.CheckOutExtension("Network")

# Begin Loop
i = 1
while i < 7000:
    transit_ND = "transit_ND"
    Route = "Route"
    Route__3_ = Route
    Route__2_ = Route
    O_MGRS = "O_MGRS"
    O_MGRS_Select = "C:\\Users\\Ryan
```

```
Chien\\Documents\\Minneapolis\\TwinCities3.gdb\\O_MGRS_Select"
    D_MGRS = "D_MGRS"
    D_MGRS_Select = "C:\\Users\\Ryan
Chien\\Documents\\Minneapolis\\TwinCities3.gdb\\D_MGRS_Select"
    Route__5_ = Route__2_
    Solve_Succeeded = "true"
    output_NALayer = Route__5_
    TwinCities3_gdb = "C:\\Users\\Ryan
Chien\\Documents\\Minneapolis\\TwinCities3.gdb"
    output_junctions = "C:\\Users\\Ryan
Chien\\Documents\\Minneapolis\\TwinCities3.gdb\\Junctions"
    output_turns = "C:\\Users\\Ryan
Chien\\Documents\\Minneapolis\\TwinCities3.gdb\\Turns"
    output_transitedges = "C:\\Users\\Ryan
Chien\\Documents\\Minneapolis\\TwinCities3.gdb\\TransitEdges"
    output_edges = "C:\\Users\\Ryan
Chien\\Documents\\Minneapolis\\TwinCities3.gdb\\Edges"
    EdgesOut_xls = "C:\\Users\\Ryan
Chien\\Documents\\Minneapolis\\RouteResults.4\\EdgesOut" + str(i) + ".xls"

    # Process: Select Origin
    arcpy.Select_analysis(O_MGRS, O_MGRS_Select, "OID = " + str(i))

    # Process: Select Destination
    arcpy.Select_analysis(D_MGRS, D_MGRS_Select, "OID = " + str(i))

    # Get start time
    rows = arcpy.SearchCursor(O_MGRS_Select)
    row = rows.next()
    t_s = row.arcgisTime

    # Process: Make Route Layer
    arcpy.MakeRouteLayer_na(transit_ND, "Route", "TravelTime_withTransit",
"USE_INPUT_ORDER", "PRESERVE_BOTH", "NO_TIMEWINDOWS", "",
"ALLOW_UTURNS", "no_LightRail", "NO_HIERARCHY", "",
"TRUE_LINES_WITH_MEASURES", t_s)

    # Process: Add Locations
    arcpy.AddLocations_na(Route, "Stops", O_MGRS_Select, "", "5000 Meters", "",
"Connectors_Stops2Streets NONE;Streets_UseThisOne SHAPE;TransitLines
NONE;Stops NONE;Stops_Snapped2Streets NONE;transit_ND_Junctions NONE",
"MATCH_TO_CLOSEST", "APPEND", "SNAP", "5 Meters", "INCLUDE",
"Connectors_Stops2Streets #;Streets_UseThisOne #;TransitLines #;Stops
#;Stops_Snapped2Streets #;transit_ND_Junctions #")

    # Process: Add Locations (2)
    arcpy.AddLocations_na(Route, "Stops", D_MGRS_Select, "", "5000 Meters", "",
"Connectors_Stops2Streets NONE;Streets_UseThisOne SHAPE;TransitLines
NONE;Stops NONE;Stops_Snapped2Streets NONE;transit_ND_Junctions NONE",
```

```
"MATCH_TO_CLOSEST", "APPEND", "SNAP", "5 Meters", "INCLUDE",
"Connectors_Stops2Streets #;Streets_UseThisOne #;TransitLines #;Stops
#;Stops_Snapped2Streets #;transit_ND_Junctions #")

    # Process: Solve
    result = arcpy.Solve_na(Route__2_, "SKIP", "CONTINUE", "")

    # If solve is good, copy traversed source features
    if result.getOutput(1) == 'true':

    # Process: Copy Traversed Source Features (with Transit)
        arcpy.CopyTraversedSourceFeaturesTransit_transit(Route__5_, TwinCities3_gdb,
"Edges", "Junctions", "Turns", "TransitEdges")

    # Process: Table To Excel
        arcpy.TableToExcel_conversion(output_edges, EdgesOut_xls, "NAME", "CODE")

    # indexing
    i = i+1
```

X. Appendix B

1. ArcGIS Data Processing

ArcGIS provides estimates of IVTTbus, OVTTwalk, OVTTwait, number of stops, and number of boardings that are specific for individual trip observations. It generates unique estimates based on OD, start time, and date. Six steps were required to build an ArcGIS network dataset for proxy generation.

First, GTFS information was imported into ArcGIS. The experimental ‘Add GTFS to a Network Dataset’ tool, (abbreviated as ‘the tool’), developed by Melinda Morang and Patrick Stevens at ESRI was used (Morang, 2017). The tool interprets GTFS data and converts it into a format understood by ArcGIS. Transit lines, stops, and SQL database containing processed GTFS data were created.

Second, road information from the 2012 Minnesota Rods shapefile was imported into ArcGIS.

Third, links were generated using the tool between transit stops and the nearest road.

Fourth, a network dataset was created. Connectivity rules were defined between transit stops, transit lines, streets, and connectors. Transit stops connect to transit lines and connectors. Transit lines connect to transit stops. Streets connect to connectors. Figure 1 illustrates the rules. This was necessary to prevent pedestrians from walking on transit lines and exiting transit anywhere other than a stop.

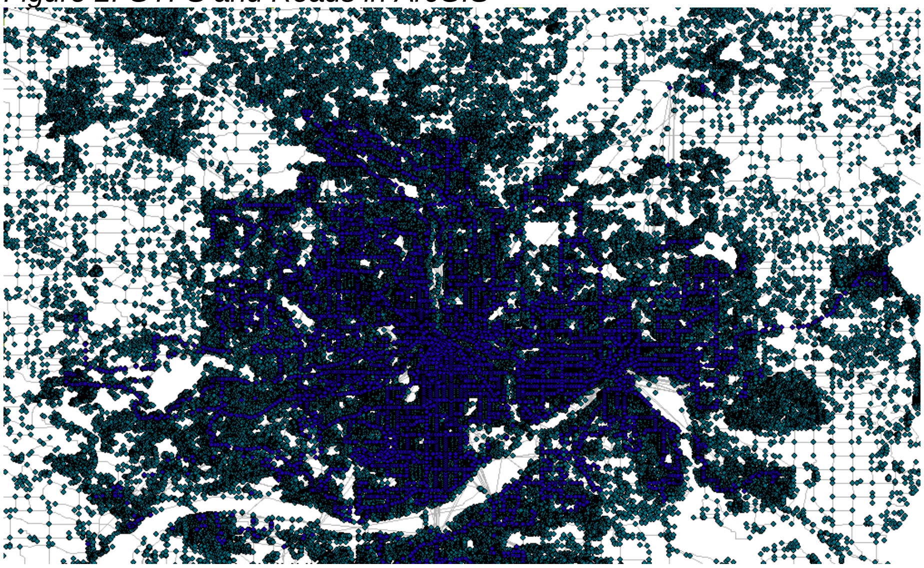
Figure 1. Connectivity Rules for the Network Dataset in ArcGIS

	Transit Stops	Transit Lines	Streets	Connectors
Transit Stops		x		x
Transit Lines	x			
Streets				x
Connectors	x		x	

Fifth, a network transit time evaluator was defined. “An evaluator tells ArcGIS how to calculate the traversal time across elements in the network dataset (Morang 2017).” Walking time was defined by dividing road length by 83.34 meters per minute, a rough estimate for the average persons’ walking speed. IVTTbus and wait time were defined to be dependent on the GTFS database. Boarding time for transit was defined as 0.25 minutes.

Finally, the transit evaluator was linked to the GTFS SQL database. This step interconnected all the preceding elements. Figure 2 shows a zoomed in portion of the completed network dataset.

Figure 2. GTFS and Roads in ArcGIS



The completed network dataset was then ready to be used. The Solve algorithm within the Network Analysis tool, (henceforth referred to as the 'solver'), finds minimum-time paths. It required four pieces of information from the TBI: the origin coordinates, the destination coordinates, the start time, and the start date.

The solver references the GTFS SQL database as part of its minimum-time calculation. The GTFS database used originated in 2016 and had a schedule range from November 2016 to March 2017. However, the TBI start dates ranged from December 2010 to February 2012. The solver cannot process a trip from 2010 since it has no information to reference at that date.

A crucial conversion was undertaken at this point. Each weekday start date from the TBI was converted to November 14, 2016. Each weekend start date was converted to November 19, 2016.

An automated ArcPy python process then ran the solver for each of the 69,788 observations with usable OD coordinates. The solver was required to use combinations of walking and bus transit; light rail was not allowed. The code is shown in Appendix A5. 66,998 observations were successfully estimated. A graphical example of a successful solver run is shown in Appendix A4. About 72 hours of computing time on a 4-core workstation were used running 4 simultaneous solvers. The variables IVTTbus, OVTTwalk, OVTTwait, nstop, and boardings were obtained from this method. A quality test for the generated variables is presented in Appendix B3.

2. Open Source Routing Machine (OSRM) Data Processing

OSRM estimates drive times that are specific for each observation. It finds the shortest path route between OD coordinates (OSRM, 2017). OD coordinates were passed to OSRM using their API, and results were saved. This process was automated using R code. 68,123 observations were successfully estimated. About 24 hours of computing time on a basic laptop were required. IVTTcar was the variable obtained from this process. A quality test for the generated variable is presented in Appendix B4.

3. Testing ArcGIS UNA Estimates for Accuracy

One way to test for the quality of the estimated components was to regress them on reported travel time for the bus observations. The equation estimated was:

$$reported\ travel\ time = \propto\ + \beta_1 IVTT + \beta_2 ovtt\ walk + \beta_3 ovtt\ wait + \beta_4 nstop + \beta_5 boardings + \ \varepsilon$$

Figure 1 presents the results of the linear regression. Only IVTTbus passes significance testing and is in the correct direction. OVTTwalk, OVTTwait, and boardings do not pass significance testing and are in the wrong direction. Nstop is in the correct direction but is not significant. The R-squared is 0.01187, indicating a poor goodness of fit.

Figure 1. Quality Analysis of ArcGIS Proxy

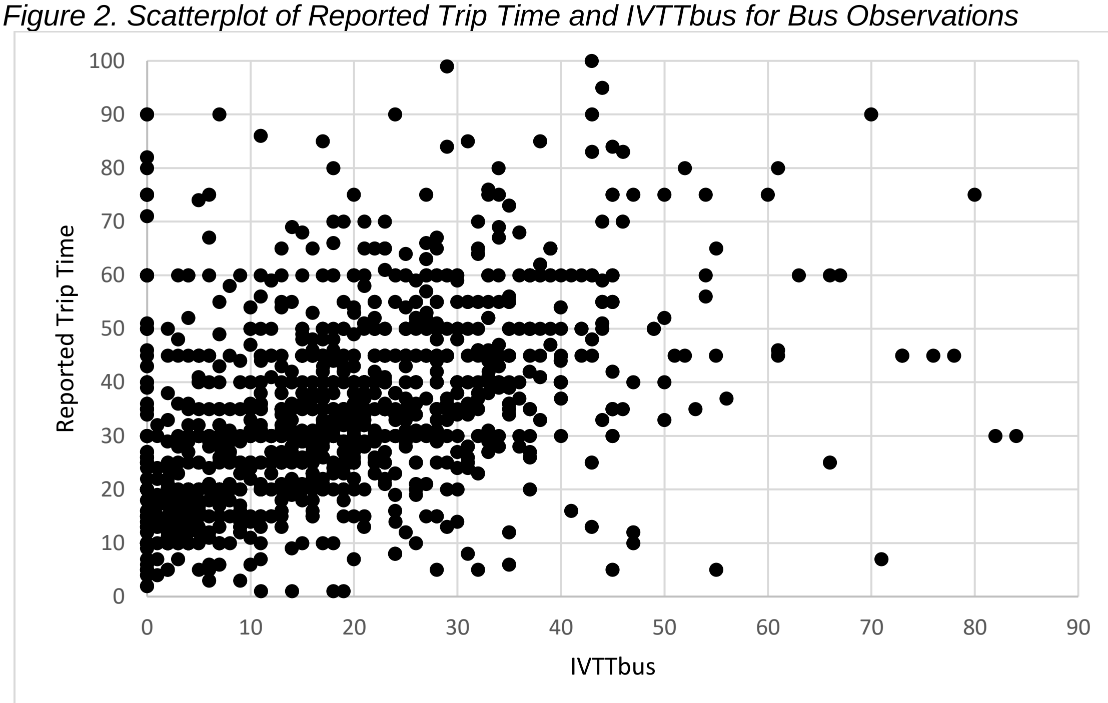
term	estimate	std.error	statistic	p.value	signif. code
(Intercept)	37.0854	2.9004	12.7863	0.0000	***
IVTTbus	0.4170	0.1768	2.3583	0.0185	*
OVTTwalk	-0.0658	0.0586	-1.1214	0.2623	
OVTTwait	-0.1889	0.3376	-0.5593	0.5760	
nstops	0.1364	0.1221	1.1167	0.2644	
boardings	-2.3562	1.2930	-1.8223	0.0687	.

- *Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*
- *Residual standard error: 49.5 on 1162 degrees of freedom*

- *Multiple R-squared: 0.0161,   Adjusted R-squared: 0.01187*

The poor fit can be attributed to three causes: the ArcGIS search algorithm, start time inaccuracies, and transit route changes. First, the solver finds the minimum-time route. A route that requires 10 minutes of riding, 20 minutes of walking, and 5 minutes of waiting is always chosen over a route that requires 21 minutes of riding, 10 minutes of walking, and 5 minutes of waiting. However, travelers may prefer the second trip even though it takes an additional minute over the first. To illustrate this issue, there are approximately 20,000 trips where walking is the only component of the trip. Figure 2 illustrates this issue for the bus observations. Note the large number of estimates where IVTT is equal to zero even though reported triptime is not zero.

Figure 2. Scatterplot of Reported Trip Time and IVTTbus for Bus Observations



Second, the solver is given a specific start time and date converted from the TBI. However, start time misreporting will cause poor solver estimates. Consider the following example: the reported start time is 7:00am. But the actual start time is 6:59am, the traveler walked for one minute, and the bus arrived at 7:00am. No waiting time was incurred. But the solver is passed the 7:00am start time. It calculates that the traveler arrived at the bus stop at 7:01am, and missed the bus. The next bus arrives at 7:15am, so 14 minutes of wait time are incurred.

Third, the Green Line connecting downtown Minneapolis to downtown St. Paul was opened in June 2014, bus route 50 was closed at the same time (Metro Transit 2014). The GTFS data used for this analysis originated in November 2016, and at the time of this study, GTFS archives are unavailable before June 2014. The TBI survey was conducted from 2011 to 2013, and would have included trips taken on bus route 50. Therefore, there is inherent error built into the ArcGIS proxies for trips between downtown Minneapolis and St. Paul, as the preeminent bus route used no longer exists.

4. Testing OSRM UNA Estimates for Accuracy

The accuracy of the IVTTcar proxy was tested by regressing it on reported drive time for the 67,484 drive observations. The estimated equation was:

$$reported\ travel\ time = \propto\ + \beta_1 drivetime + \ \varepsilon$$

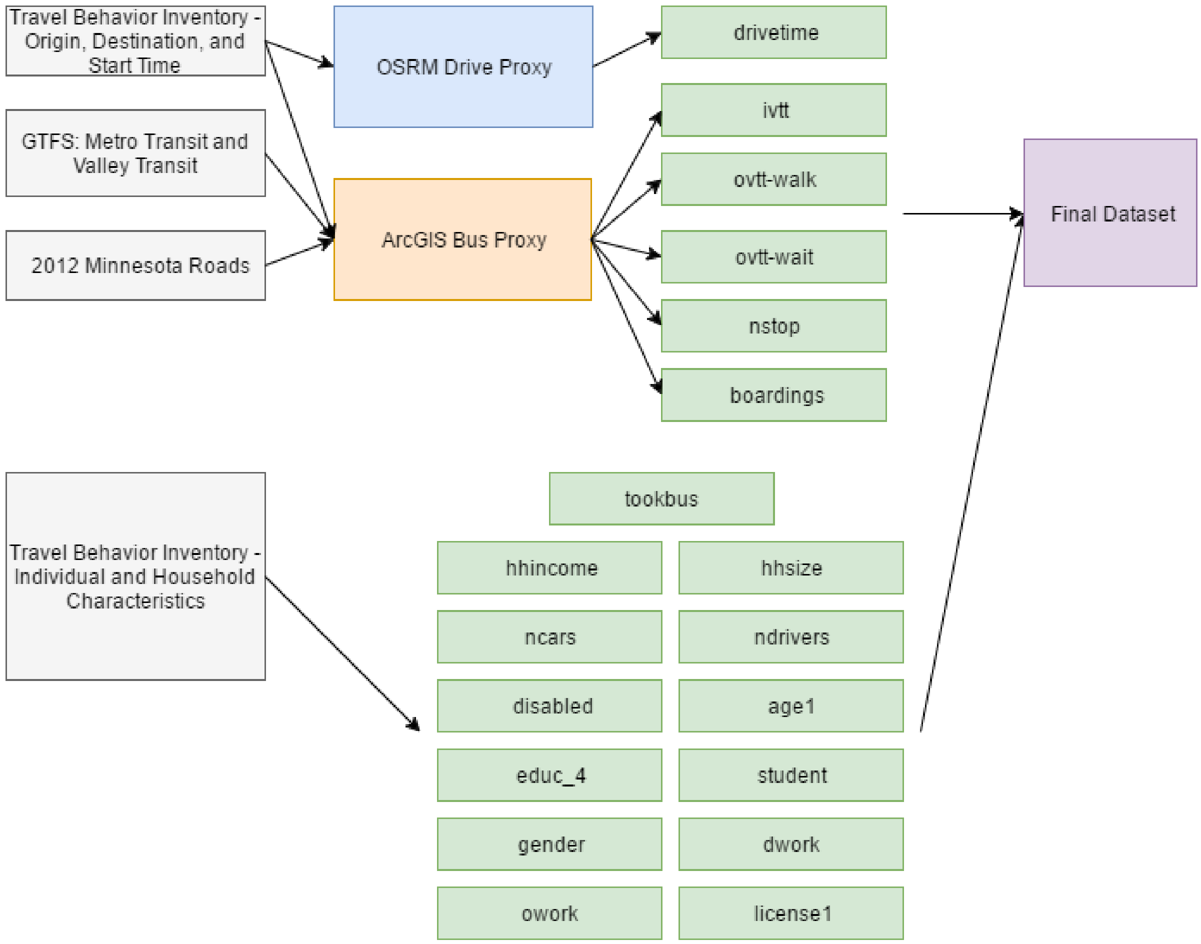
Figure 6 presents the results of the linear regression. The estimated variable's direction is correct, the passes significance testing at better than the 0.1% level, and the goodness of fit is reasonable at 0.4093. The major limitation preventing the OSRM from being more accurate is that it is not traffic aware. As a result, peak-hour drive times are underestimated.

Figure 6: Quality Analysis of OSRM Proxy

term	estimate	std.error	statistic	p.value	signif. code
(Intercept)	7.36068	0.0679656	108.3001	0	***
IVTTcar	0.6798036	0.0031439	216.231	0	***

- *Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*
- *Residual standard error: 12.66 on 67484 degrees of freedom*
- *Multiple R-squared: 0.4093, Adjusted R-squared: 0.4093*
- *F-statistic: 4.676e+04 on 1 and 67484 DF, p-value: < 2.2e-16*

5. Data Conglomeration Flow Chart



6. Table of Coefficients and Significance for Varying Market Definitions Based on Drive Time

dt_max	n	IVTTbus	p	OVTTwalk	p	OVTTwait	p	IVTTcar	p	logLik
2	2588	0.5131	0.6402	-0.1746	0.2276	-0.6385	0.4129	0.7142	0.4682	-44.2
3	5178	0.6333	0.0405	-0.0473	0.4095	-0.2323	0.3561	0.2684	0.5071	-119.3
4	8474	0.4525	0.0008	-0.0275	0.4282	-0.0908	0.5775	0.1582	0.5122	-212.4
5	11792	0.3151	0.0003	-0.0602	0.0133	-0.0617	0.5868	0.1900	0.2290	-324.9
6	15336	0.2336	0.0003	-0.0809	0.0000	-0.0517	0.5361	0.1916	0.0981	-429.5
7	18895	0.1623	0.0009	-0.0782	0.0000	-0.0521	0.4192	0.1384	0.1293	-552.2
8	22261	0.1344	0.0004	-0.0841	0.0000	-0.1052	0.0526	0.2522	0.0005	-722.6
9	25473	0.1140	0.0001	-0.0924	0.0000	-0.1367	0.0011	0.2908	0.0000	-864.8
10	28421	0.1097	0.0000	-0.0942	0.0000	-0.1073	0.0019	0.3112	0.0000	-1031.6
11	31164	0.0908	0.0000	-0.0913	0.0000	-0.1014	0.0017	0.2481	0.0000	-1126.3
12	33624	0.0799	0.0000	-0.0898	0.0000	-0.1010	0.0005	0.2503	0.0000	-1274.5
13	35761	0.0689	0.0000	-0.0873	0.0000	-0.0914	0.0003	0.2274	0.0000	-1378.5
14	37756	0.0623	0.0000	-0.0890	0.0000	-0.0902	0.0001	0.2182	0.0000	-1466.7

15	39496	0.0541	0.0001	-0.0906	0.0000	-0.0928	0.0000	0.2166	0.0000	-1574.9
16	40982	0.0401	0.0022	-0.0948	0.0000	-0.1074	0.0000	0.2270	0.0000	-1681.1
17	42336	0.0241	0.0367	-0.0999	0.0000	-0.1076	0.0000	0.2421	0.0000	-1782.6
18	43439	0.0156	0.1489	-0.0982	0.0000	-0.0982	0.0000	0.2243	0.0000	-1873.5
19	44432	0.0044	0.6653	-0.1000	0.0000	-0.1019	0.0000	0.2364	0.0000	-1955.6
20	45247	0.0010	0.9239	-0.1004	0.0000	-0.0984	0.0000	0.2333	0.0000	-2003.6
21	45935	0.0016	0.8718	-0.1002	0.0000	-0.1016	0.0000	0.2251	0.0000	-2055.3
22	46642	-0.0014	0.8834	-0.0993	0.0000	-0.1049	0.0000	0.2238	0.0000	-2101.7
23	47236	-0.0075	0.4248	-0.0992	0.0000	-0.1009	0.0000	0.2280	0.0000	-2154.7
24	47817	-0.0124	0.1779	-0.0978	0.0000	-0.0955	0.0000	0.2293	0.0000	-2212.4
25	48352	-0.0192	0.0295	-0.0977	0.0000	-0.0892	0.0000	0.2285	0.0000	-2272.3
26	48799	-0.0211	0.0161	-0.0980	0.0000	-0.0887	0.0000	0.2233	0.0000	-2309.1
27	49180	-0.0265	0.0023	-0.0996	0.0000	-0.0916	0.0000	0.2332	0.0000	-2345.4
28	49578	-0.0296	0.0005	-0.0991	0.0000	-0.0905	0.0000	0.2326	0.0000	-2375.0
29	49929	-0.0298	0.0004	-0.0992	0.0000	-0.0896	0.0000	0.2259	0.0000	-2400.6
30	50259	-0.0342	0.0000	-0.1000	0.0000	-0.0885	0.0000	0.2281	0.0000	-2427.7
31	50535	-0.0347	0.0000	-0.0992	0.0000	-0.0900	0.0000	0.2228	0.0000	-2460.9
32	50799	-0.0350	0.0000	-0.0980	0.0000	-0.0876	0.0000	0.2161	0.0000	-2488.4
33	51021	-0.0339	0.0000	-0.0961	0.0000	-0.0902	0.0000	0.2127	0.0000	-2514.1
34	51251	-0.0343	0.0000	-0.0947	0.0000	-0.0895	0.0000	0.2106	0.0000	-2542.2
35	51425	-0.0344	0.0000	-0.0947	0.0000	-0.0913	0.0000	0.2089	0.0000	-2549.0
36	51574	-0.0353	0.0000	-0.0947	0.0000	-0.0931	0.0000	0.2065	0.0000	-2555.3
37	51712	-0.0370	0.0000	-0.0944	0.0000	-0.0873	0.0000	0.2065	0.0000	-2573.7
38	51823	-0.0367	0.0000	-0.0940	0.0000	-0.0855	0.0000	0.2037	0.0000	-2582.8
39	51939	-0.0369	0.0000	-0.0937	0.0000	-0.0855	0.0000	0.2013	0.0000	-2588.0
40	52029	-0.0368	0.0000	-0.0931	0.0000	-0.0844	0.0000	0.1985	0.0000	-2592.1
41	52114	-0.0368	0.0000	-0.0920	0.0000	-0.0833	0.0000	0.1963	0.0000	-2608.9
42	52185	-0.0360	0.0000	-0.0917	0.0000	-0.0829	0.0000	0.1938	0.0000	-2614.9
43	52244	-0.0369	0.0000	-0.0918	0.0000	-0.0824	0.0000	0.1937	0.0000	-2616.7
44	52281	-0.0366	0.0000	-0.0916	0.0000	-0.0807	0.0000	0.1920	0.0000	-2623.6
45	52321	-0.0369	0.0000	-0.0917	0.0000	-0.0807	0.0000	0.1915	0.0000	-2625.2

7. Model Specification and Coefficients for 3 Model Comparison

Model 1	Estimate	Std. Error	Pr(>|z|)	Model 2	Estimate	Std. Error	Pr(>|z|)	Model 3	Estimate	Std. Error	Pr(>|z|)
(Intercept)	2.1416	0.3022	0.0000	(Intercept)	2.2618	0.3175	0.0000	(Intercept)	2.1509	0.3005	0.0000
IVTTbus	-0.0296	0.0071	0.0000	IVTTpool	0.2298	0.0096	0.0000	IVTTbus	-0.0308	0.0066	0.0000
OVTTwalk	-0.0696	0.0070	0.0000	OVTTwalk	-0.0739	0.0074	0.0000	IVTTcar	0.1634	0.0130	0.0000
OVTTwait	-0.0774	0.0120	0.0000	OVTTwait	-0.1501	0.0152	0.0000	OVTTsum	-0.0703	0.0066	0.0000
nstop	-0.0260	0.0037	0.0000	nstop	-0.0637	0.0036	0.0000	nstop	-0.0252	0.0034	0.0000
boardings	0.3144	0.0364	0.0000	boardings	0.1141	0.0401	0.0022	boardings	0.3079	0.0355	0.0000
IVTTcar	0.1637	0.0130	0.0000								
hh_income	-0.1365	0.0113	0.0000	hh_income	-0.1498	0.0125	0.0000	hh_income	0.1365	0.0113	0.0000
hh_size	-0.2569	0.0788	0.0006	hh_size	-0.3240	0.0790	0.0000	hh_size	-0.2560	0.0788	0.0006
ncars	-1.4583	0.1132	0.0000	ncars	-1.5677	0.1243	0.0000	ncars	-1.4577	0.1132	0.0000
ndrivers	0.9892	0.1157	0.0000	ndrivers	1.0349	0.1258	0.0000	ndrivers	0.9890	0.1156	0.0000
disabled	-0.0274	0.1791	0.4392	disabled	-0.0055	0.1850	0.4881	disabled	-0.0274	0.1793	0.4393
educ4	0.0735	0.0221	0.0004	educ4	0.0695	0.0238	0.0018	educ4	0.0735	0.0220	0.0004
age1	-0.2020	0.0233	0.0000	age1	-0.2026	0.0261	0.0000	age1	-0.2019	0.0233	0.0000
student	0.4075	0.1340	0.0012	student	0.5614	0.1453	0.0001	student	0.4065	0.1339	0.0012
gender	0.2865	0.0770	0.0001	gender	0.3111	0.0852	0.0001	gender	0.2848	0.0769	0.0001
dwork	1.2726	0.0911	0.0000	dwork	1.2223	0.1010	0.0000	dwork	1.2724	0.0910	0.0000
owork	0.9494	0.0921	0.0000	owork	0.7352	0.1080	0.0000	owork	0.9486	0.0919	0.0000
license1	-2.4658	0.1679	0.0000	license1	-2.4882	0.1766	0.0000	license1	-2.4654	0.1679	0.0000
Log Lik.	-2790.35				-2252.61				-2790.64		
Null Log Lik.	-4501.86				-4501.86				-4501.86		
K	19				18				18		
N	44124				44124				44124		
rho2	0.3802				0.4996				0.3801		
adjrho2	0.3760				0.4956				0.3761		