Key Enhancements to the WFRC/MAG Four-Step Travel Demand Model

Reid Ewing, Ph.D.

Sadegh Sabouri

Keunhyun Park, Ph.D.

Torrey Lyons, Ph.D.

Guang Tian, Ph.D.

THE UNIVERSITY OF UTAH

KEY ENHANCEMENTS TO THE WFRC/MAG FOUR-STEP TRAVEL DEMAND MODEL

Final Report

NITC-RR-1086

by

Reid Ewing
Sadegh Sabouri
Keunhyun Park
Torrey Lyons
Guang Tian

University of Utah

for

National Institute for Transportation and Communities (NITC)
P.O. Box 751
Portland, OR 97207

December 2019

Technical Report Documentation Page

1. Report No. NITC-RR-1086	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Key Enhancements to the WFRC/MAG Four-Step Travel Demand Model		5. Report Date December 2019
		6. Performing Organization Code
7. Author(s) Reid Ewing, Sadegh Sabouri, Keunhyun Park, Torrey Lyons, Guang Tian		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Utah Department of City and Metropolitan Planning 375 South 1530 East, RM 235 AAC Salt Lake City, UT 84112		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 1086
12. Sponsoring Agency Name and Address National Institute for Transportation and Communities (NITC) P.O. Box 751 Portland, Oregon 97207		13. Type of Report and Period Covered Draft Final June 2017 – December 2019
		14. Sponsoring Agency Code

15. Supplementary Notes

Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration

16. Abstract

In a National Transit Institute course on "Coordinating Land Use and Transportation," co-taught by Robert Cervero, Uri Avin, and the PI on this project, the analytic tools session began with a hypothetical: assume that all households, jobs, and other trip generators are concentrated in a walkable village rather than segregated by use and spread across a traffic analysis zone in the standard suburban fashion. The instructor then asks: How would the outputs of conventional four-step travel demand models differ between these two future land use scenarios. The answer, to most participants' surprise, was "Not at all." Conventional four-step travel demand models are used by nearly all metropolitan planning organizations (MPOs), state departments of transportation, and local planning agencies, as the basis for long-range transportation planning in the United States. In the simplest terms, the four-step model proceeds from trip generation, to trip distribution, to mode choice, and finally to route assignment. Trip generation tells us the number of trips generated (produced or attracted) in each traffic analysis zone (TAZ), usually based on some prediction of vehicle ownership. Trip distribution tells us where the trips go, matching trip productions to trip attractions by considering the spatial distribution of productions and attractions as well as the impedance (time or cost) of connections. Particularly tricky are predictions of trips that remain within the same zone. Mode choice tells us which mode of travel is used for these trips, factoring trip tables to reflect the relative shares of different modes. Route assignment tells us what routes are taken, assigning trips to networks that are specific to each mode. A flaw of the four-step model is its relative insensitivity to the so-called D variables. The D variables are characteristics of the built environment that are known to affect travel behavior. The Ds are development density, land use diversity, street network design, destination accessibility, and distance to transit. This report develops a vehicle ownership model (car shedding model), an intrazonal travel model (internal capture model), and mode choice model that consider all of the D variables based on household travel surveys and built environmental data for 32, 31, and 29 regions, respectively, validates the models, and demonstrates that the models have far better predictive accuracy than Wasatch Front Regional Council (WFRC)/Mountainland Association of Governments' (MAG) current models.

17. Key Words travel model enhancements, vehicle ownership model, intrazonal travel model, mode choice model, four-step travel demand model		18. Distribution Statement No restrictions. Copies available from NITC: www.nitc-utc.net	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 68	22. Price

ACKNOWLEDGEMENTS

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government. This report does not constitute a standard, specification, or regulation.

RECOMMENDED CITATION

Ewing, Reid., Sabouri, Sadegh., Park, Keunhyun., Lyons, Torrey., & Tian, Guang. *Key Enhancements to the WFRC/MAG Four-Step Travel Demand Model*. NITC-RR-1086. Portland, OR: Transportation Research and Education Center (TREC), 2019.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Conventional four-step travel demand models are used by nearly all metropolitan planning organizations (MPOs), state departments of transportation, and local planning agencies, as the basis for long-range transportation planning in the United States. In the simplest terms, the four-step model proceeds from trip generation, to trip distribution, to mode choice, and finally to route assignment. Trip generation tells us the number of trips generated (produced or attracted) in each traffic analysis zone (TAZ), usually based on some prediction of vehicle ownership. Trip distribution tells us where the trips go, matching trip productions to trip attractions by considering the spatial distribution of productions and attractions as well as the impedance (time or cost) of connections. Particularly tricky are predictions of trips that remain within the same zone. Mode choice tells us which mode of travel is used for these trips, factoring trip tables to reflect the relative shares of different modes. Route assignment tells us what routes are taken, assigning trips to networks that are specific to each mode.

A flaw of the four-step model is its relative insensitivity to the so-called D variables. The D variables are characteristics of the built environment that are known to affect travel behavior. The Ds are development density, land use diversity, street network design, destination accessibility, and distance to transit. This report develops a vehicle ownership model (car shedding model), an intrazonal travel model (internal capture model), and a mode choice model that consider all of the D variables based on household travel surveys and built environmental data for 32, 31, and 29 regions, respectively, validates the models, and demonstrates that the models have far better predictive accuracy than Wasatch Front Regional Council (WFRC)/Mountainland Association of Governments' (MAG) current models.

Vehicle ownership – the number of private vehicles a household owns – is one of the key inputs to trip generation and mode choice in most four-step models. The problems with existing vehicle ownership models include the use of data from a single region, the consideration of only some D variables, and the use of different metrics to represent the Ds. These issues restrict our understanding of car shedding behavior, that is, the decision to own fewer vehicles as the Ds increase (except distance to transit, which works in reverse). In this report, we pool regional household travel survey data from 32 diverse regions of United States and generate consistent measures for all regions. Next, we use Poisson regression to model vehicle ownership instead of the commonly used multinomial logit (MNL) model. We also use multilevel modeling to account for the dependence of households from a given metropolitan region on characteristics of that region. We compare the results of our model and the Wasatch Front Regional Council's current model against the actual number of vehicles owned by households from the 2012 Utah Travel Study for prediction accuracy. Our model outperforms the current model.

Trip distribution – whether the trip is intrazonal (internal) or interzonal (external) – is one of the essential steps in travel demand forecasting. However, the current intrazonal forecasts based on a gravity model involve questionable assumptions, primarily due to differences in D variables across zones. In this study, we first survey 25 MPOs about how they model intrazonal travel and

find the state of the practice to be dominated by the gravity model. Using travel data from 31 diverse regions in the U.S., we develop an approach to enhance the conventional model by including more built environment D variables and by using multilevel logistic regression. The models' predictive capability is confirmed using k-fold cross-validation. The study results have practical implications for state and local planning and transportation agencies to achieve better accuracy and generalizability in their travel demand modeling.

Mode choice model is the third step of the traditional four-step travel demand model that predicts the mode of travel and may include private automobile, public transportation, walking, bicycling, or other means. However, only a few of the existing mode choice models include nonmotorized modes of travel (i.e., walk and bike). By surveying 25 MPOs about how they model modes of travel, we found that more than half of them do not include nonmotorized modes. WFRC/MAG are two MPOs that model nonmotorized trips, but only with consideration of trip distance, not the built environment characteristics of the TAZs in which travel occurs. We develop advanced mode choice models (i.e., multi-level nested logit regression models) by trip purpose using regional household travel data and built environmental variables from 29. The results confirm the vital role that the built environment plays in shaping people's mode choice behavior. Same as the previous models, we compare the results of our mode choice models and the Wasatch Front Regional Council's current model against the actual shares of motorized and nonmotorized modes within each TAZ from the 2012 Utah Travel Study for prediction accuracy. Once again, our models outperform the current WFRC/MAG models.

1.0   INTRODUCTION

Metropolitan planning organizations (MPOs) coordinate transportation investments from federal, state, and local sources to ensure that regional transportation plans meet performance criteria such as air quality and congestion management. One of the essential ways MPOs determine how to allocate funds is the forecasting of future travel demands. Forecasts are ordinarily made using what is known as the four-step travel demand model.

Conventional four-step models, used by nearly all MPOs, state departments of transportation and local transportation planning agencies to forecast future travel patterns and develop long-range transportation plans, are the basis for long-range transportation planning in the United States. Their importance for project selection cannot be overstated.

In the simplest terms, the four-step model proceeds from trip generation, to trip distribution, to mode choice, and finally to route assignment. Trip generation tells us the number of trips generated (produced or attracted) in each traffic analysis zone (TAZ). Trip distribution tells us where the trips go, matching trip productions to trip attractions by considering the spatial distribution of productions and attractions as well as the impedance (time or cost) of connections. Mode choice tells us which mode of travel is used for these trips, factoring trip tables to reflect the relative shares of different modes. Route assignment tells us what routes are taken, assigning trips to networks that are specific to each mode. The model's behaviors are estimated based on travel patterns distilled from surveyed household trips. The model is calibrated and validated by comparing the predicted trips in the base year to actual travel survey data. The four-step modeling process is visualized below in Figure 1.1.
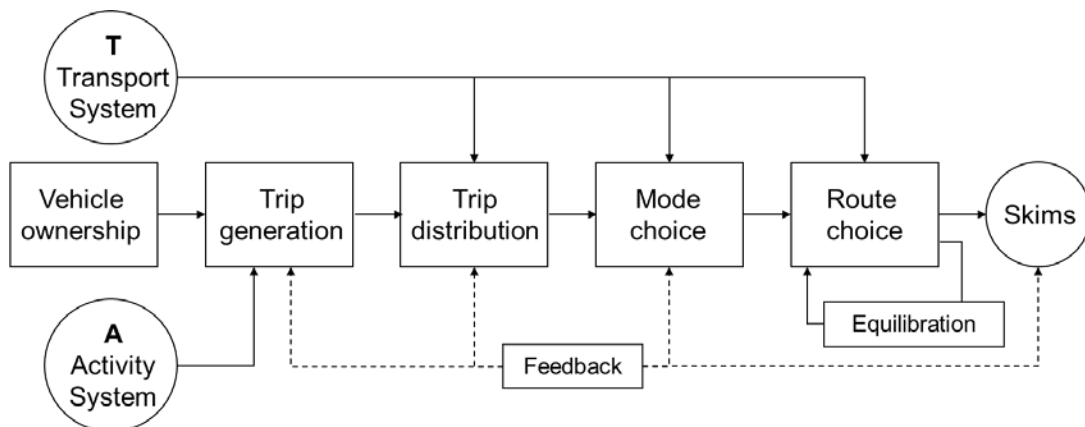
Figure 1.1: Four Step Travel Demand Model (Adapted from McNally, 2007)

1.1   LIMITATIONS OF THE FOUR-STEP PROCESS

The most important limitation of the conventional four-step travel demand modeling and forecasting process is the failure to account for the full effects of the built environment on travel outputs at each step. The built environment affects household travel decisions in multiple ways, many of which are not captured in the conventional process (Cervero, 2006; Davidson et al.,

2007; Ferdous et al., 2012; McNally, 2000; Pinjari and Bhat, 2011; Pont et al., 2013; Rouwendal and Nijkamp, 2004; Van Acker and Witlox, 2011; Walters et al., 2000).

These models currently are underspecified, which is to say that important variables are omitted. In particular, conventional models fail to fully account for local land use patterns, street network designs, and urban design features—indeed, the entire built environment at the scale of a neighborhood or activity center. In many four-step models, vehicle ownership is treated as a function of sociodemographic variables only (or largely), and the phenomenon of car shedding as the built environment becomes more compact is not accounted for. In many models, only trips by vehicle are modeled, and trip rates are related only to sociodemographic characteristics of people, not characteristics of place. Bicycling, in particular, is seldom treated as a separate transportation mode. In nearly all four-step models, households, jobs, and other trip generators are assumed to be located at a single point, the zone centroid, rather than spread across the traffic analysis zone, and the entire local street network is reduced to one or more centroid connectors to the regional street network. This limits the modeling of intrazonal travel in terms of the local built environment.

With this study, we seek to develop and implement car shedding, intrazonal travel, and walk and bike mode choice models that can be used in conjunction with a conventional four-step model to capture neglected effects of the built environment on travel behavior. These models are calibrated with data from our 32-region household travel database, the largest household travel database of its sort ever assembled. This database has been linked to built environmental data for buffers around geocoded trip ends. These models will pre-process inputs to the four-step process and/or post-process outputs. They will be incorporated into the Wasatch Front Regional Council and Mountainland Association of Governments' (our MPOs) four-step model and, based on this case study, will be offered to other MPOs for incorporation into their models. We have WFRC, MAG, UTA, and UDOT's support to do this work, along with support from the National Institute for Transportation and Communities (NITC). Our work best aligns with the NITC theme of Integrating Multimodal Transportation and Land Use.

Some MPOs are beginning to abandon the traditional four-step travel model in favor of activity/tour-based travel modeling (ABT). As of 2015, in the U.S. ABT modeling was still in its formative stages and not standard practice (Travel Forecasting Resource, 2015). Atlanta Regional Commission, San Diego Association of Government, and New York Metropolitan Transportation Council are some of the pioneering MPOs using this approach. Notwithstanding nearly 30 years of promotion of activity-based modeling (ABM) in the travel modeling literature, we believe that enhancements of the conventional four-step model are still relevant and desirable. As presented in this paper, our survey of MPOs, both big and small, shows that all still use the conventional four-step model, and 20 of 25 MPOs surveyed still use the gravity model for trip distribution. The conventional model and gravity model are still near-universal among small and medium-sized MPOs. As a representative of our local MPO said, when it comes to modeling, MPOs need to be "met where they are at." Meeting MPOs at the current state of the practice and providing an incremental advancement to that practice is the goal of our suggested approach. Our method is meant to be simple and used in connection with the four-step model.

1.2   RESEARCH QUESTIONS

What are the specific research questions addressed in this project?

- How does vehicle ownership vary with the D variables from the travel behavior literature (density, diversity, design, destination accessibility, distance to transit, and demographics)? We would expect car shedding to occur in dense, mixed-use, pedestrian-friendly and transit-served developments, holding sociodemographics constant.

- How does intrazonal travel vary with the D variables? We would expect internal capture of significant numbers of trips to occur in dense, mixed-use, pedestrian-friendly and transit-served developments, holding sociodemographics and employment constant.

- How do walk and bike mode choices vary with the D variables? We would expect high walk and bike mode shares to occur in dense, mixed-use, pedestrian-friendly and transit-served development, holding sociodemographics constant.

The specific outcomes of the project will be equations that predict each of the outcomes listed above (vehicle ownership, intrazonal trip choice by trip purpose, etc.) in terms of D variables of the TAZs themselves and their surrounding environments. The equations will be along the same lines as those already published by the lead investigator. However, the neighborhood variables will be for TAZs rather than the MXDs or buffers. Earlier published work by this team includes:

- J. Gulden, J.P. Goates, and R. Ewing, Mixed-Use Development Trip Generation Model, Transportation Research Record, Vol. 2344, 2013, pp. 98–106;
- R. Ewing, M. Greenwald, M. Zhang, et al., Traffic Generated by Mixed-Use Developments – A Six-Region Study Using Consistent Built Environmental Measures, Journal of the Urban Planning and Development, Vol. 137, Issue 3, 2011, pp. 248-261;
- R. Ewing, M. Bogaerts, M. Zhang, M. Greenwald, and W. Greene, Predicting Transportation Outcomes for LEED-ND Pilot Projects, Journal of Planning Education and Research, Vol. 33, Issue 3, 2013, pp. 265-279;
- R. Ewing, G. Tian, J.P. Goates, M. Zhang, M.J. Greenwald, A. Joyce, J. Kircher, and W. Greene (2014), Varying influences of the built environment on household travel in 15 diverse regions of the United States, Urban Studies, 52(13), 2330–2348; and
- G. Tian, R. Ewing, A. White, J. Walters, J.P. Goates and A. Joyce (2015), Traffic Generated by Mixed-Use Developments—13-Region Study Using Consistent Built Environment Measures, Transportation Research Record, (2500), 116–124.

1.3   METHODOLOGY

In this study, we proposed to estimate vehicle ownership, intrazonal trip choice, and walk and bike mode choice models using data from a 32-region household travel database integrated with built environmental data. The 32 regions are diverse and provide a great measure of external validity to our work. They are: Albany, NY; Atlanta, GA; Boston, MA; Burlington, NC; Charleston, SC; Dallas, TX; Denver, CO; Detroit, MI; Eugene, OR; Greensboro, NC; Hampton

Roads-Norfolk, VA; Houston, TX; Indianapolis, IN; Kansas City, MO; Madison, WI; Miami-Dade, FL; Minneapolis-St. Paul; Orlando, FL; Phoenix, AZ; Portland, OR; Provo-Orem, UT; Richmond, VA; Rochester, NY; Salem, OR; Salt Lake City, UT; San Antonio, TX; Seattle, WA; Springfield, MA; Syracuse, NY; Tampa, FL; West Palm Beach, FL; and Winston-Salem, NC. Note that for intrazonal trip choice and walk/bike mode choice models, we used 31 and 29 of the regions, respectively.

Our first step was to acquire household travel and built environmental data. It proved difficult to obtain travel data with XY coordinates due to concerns over confidentiality. Each dataset has required about three or four months for acquisition and processing. This 32-region database has been collected and processed over seven years.

Our second step was to conduct thorough reviews of the literature on vehicle ownership/car shedding, intrazonal travel/internal capture, and walk and bike mode choice. Only the first of these topics had a relatively recent, comprehensive review by our research team. The literature search was conducted using Transport Research International Documentation (TRID) (which already has been searched), SCOPUS, and Google Scholar.

The third step was to estimate/calibrate three sets of models. The vehicle ownership model is a Poisson model, though two other models were also estimated. The current WFRC/MAG model is a multinomial logit model, which Bill Greene, one of the world's leading econometricians and consultant on this project, says is not preferred for a count variable like vehicle ownership. The intrazonal travel/internal capture model is a binomial logistic regression model, as staying or leaving a zone is a dichotomous choice. The current WFRC/MAG model uses the gravity model and a nearest neighborhood approximation to estimate intrazonal travel, ignoring many of the D variables. The walk and bike models are multinomial logistic regression models, possibly nested. Bill Greene has assisted with the estimation of these models. The current WFRC/MAG model crudely estimates walk and bike trips, and lumps them together.

Given the nested nature of the datasets (with households nested within TAZs and TAZs nested within regions), the modeling will necessarily be multilevel. This is the approach we took in the five articles referenced above. The nesting structure creates a dependence among trips to the same place, and households living in the same place, which violates the independence assumption of ordinary least squares (OLS) regression and leads to inefficient and biased regression coefficients and standard error estimates. That is to say, households in Boston are likely to have very different travel and vehicle ownership patterns than households in Houston, irrespective of their socioeconomic and neighborhood characteristics. Such a nested data structure requires multilevel modeling (MLM) to account for shared characteristics.

2.0 A VEHICLE OWNERSHIP (CAR SHEDDING) MODEL AS A PRE-STEP OF TRAVEL DEMAND MODELING

2.1 INTRODUCTION

Travel demand models are used to predict future traffic volumes for the auto-highway and transit systems based on projections of future land use patterns and future network capacities. The conventional four-step model has become the workhorse of long-range transportation planning. Its steps include trip generation, trip distribution, mode choice, and route choice (traffic assignment) (Beimborn et al., 1996; McNally, 2008; Zhou et al., 2009).

While not always treated as such, vehicle ownership forecasting is a step in the conventional travel demand forecasting process and activity based travel demand models (Castiglione et al., 2015). In conventional travel demand forecasting, it logically follows land use forecasting, before trip generation, which is commonly treated as step one. Vehicle ownership and household size are the most common inputs to household trip generation in the conventional process, and the effects carry through all the remaining steps (Cervero, 2006; Kitamura, 2009; Mwakalonge and Badoe, 2014). In the trip generation step, input files that classify households by household size, vehicle ownership, and one or two other variables, are multiplied by trip generation rates to obtain trip productions by traffic analysis zone and trip purpose. These generated trips are then distributed in the second step, divided among modes in the third step, and assigned to the highway and transit networks in the fourth step. Errors in vehicle ownership, and hence trip generation, propagate through the remaining steps.

In many metropolitan regions, vehicle ownership is not even a modeled input but instead is held constant or extrapolated from existing vehicle ownership patterns (Broadstock et al., 2010; Kim and Susilo, 2013). If it is modeled, vehicle ownership often is related mainly to sociodemographic variables, not so much to built environmental variables (Cao et al., 2007; Cirillo and Liu, 2013; Kitamura et al., 2001; Pinjari et al., 2011). However, in activity-based models, we can see a conspicuous improvement to the vehicle ownership prediction since these models provide "better sensitivity to the influence of urban form, accessibility, and demographics on auto ownership choices" (Castiglione et al., 2015).

In this report, we present vehicle ownership models that contribute to our understanding of vehicle ownership and improve the accuracy of travel demand forecasts in two distinct ways. First, we pool regional household travel survey data from 32 diverse regions of United States and generate consistent measures for all regions. Next, we use Poisson regression to model vehicle ownership instead of the commonly used multinomial logit (MNL) model. We also use multilevel modeling to account for the dependence of households from a given metropolitan region on characteristics of that region. We compare the results of our model and the Wasatch Front Regional Council's current model against the actual number of vehicles owned by households from the 2012 Utah Travel Study for prediction accuracy.

The remainder of this chapter is organized as follows. Section 2 contains a review of studies on vehicle ownership and the phenomenon of car shedding. Section 3 introduces state of the practice in predicting vehicle ownership, and problems associated with these models. Section 4

describes the data and statistical methods used to estimate new multi-regional models. Section 5 presents the results and evaluates the new models relative to the current WFRC/MAG model. Finally, section 6 discusses the results and presents the conclusions.

2.2   LITERATURE REVIEW

Vehicle ownership is of interest from the standpoints of energy, environment, and transportation. Over half of the world's oil and about 30% of total commercial world energy are consumed by the transport sector. In 2013, about 31% of total U.S. CO_2 emissions and 26% of total U.S. greenhouse gas emissions were generated by transportation (EPA, 2015). Vehicle ownership models are used by policy makers to identify factors that affect vehicle miles traveled (VMT), and therefore address problems related to energy consumption, air pollution, and traffic congestion (Dargay and Gately, 2007; Schipper, 2011).

Vehicle ownership is generally treated as a function of households' sociodemographic characteristics. Some studies use income or income per capita to forecast national or global vehicle ownership (Dargay and Gately, 1997; Dargay et al., 2007). Some other sociodemographic characteristics have been reported as good predictors of vehicle ownership, like household size, number of children and workers, and even immigration status (Bhat et al., 2013).

However, there are many studies that have found additional relationships between vehicle ownership and built environmental variables (Ewing & Tilbury, 2002; Schimek, 1996; Van et al., 2010; Zegras; 2010). Households that live in dense, mixed-use, and transit served areas tend to own fewer automobiles, a phenomenon called car shedding; at the same time, they make more walk, bike, and transit trips (Ewing & Tilbury, 2002).

The phenomenon of car shedding is well documented in the literature (Chang, 2006; Cirillo and Xu, 2011; de Jong and Kitamura, 2009). Studies have found that the built environment, characterized by the so-called D variables, affects vehicle ownership after controlling for the sociodemographic characteristics of households. The original 'three Ds', coined by Cervero and Kockelman (1997), are density, diversity, and design, followed later by destination accessibility and distance to transit (Ewing and Cervero, 2001). While not part of the environment, demographics are the sixth D, controlled as confounding influences in travel studies.

Car shedding occurs as the Ds increase (or inversely, as distance to transit decreases). All of the Ds are important, not just density which is the D variable most likely to be included in vehicle ownership models. That is, all of the Ds have been found to be related to vehicle ownership in one study or another, like population and employment density (Bento et al., 2005; Chatman, 2013; Guo, 2013; Hess and Ong, 2002; Pinjari et al., 2011; Ryan and Han, 1999; Zegras, 2010), street network design (Bento et al., 2005; Bhat and Guo, 2007; Guo, 2013; Pinjari et al., 2011), land use diversity (Bento et al., 2005; Cao et al., 2007; Chu, 2002; Hess and Ong, 2002; Zegras, 2010), destination accessibility (Pinjari et al., 2011; Shay and Khattak, 2005), and distance to transit (Bento et al., 2005; Bhat and Guo, 2007; Cao et al., 2007; Chatman, 2013; Guo, 2013; Kim and Kim, 2004; Pinjari et al., 2011; Zegras, 2010).

Additionally, some other variables have also been reported to be related to vehicle ownership, like parking availability (Chatman, 2013; Guo, 2013; Kitamura et al., 2001), housing or neighborhood type (Bhat and Guo, 2007; Bhat and Pulugurta, 1998; Chatman, 2013; Pinjari et al., 2011; Potoglou and Susilo, 2008; Shay and Khattak, 2005; Shay and Khattak 2007; Zegras, 2010), travel attitudes (Cao et al., 2007), and urban area size (Cirillo and Liu, 2013).

The economic and behavioral explanations of car shedding is that the first five Ds affect the accessibility of trip productions to trip attractions, and hence the generalized cost of travel by different modes to and from different locations. This, via consumer choice theory of travel demand (Ben-Akiva and Lerman, 1985; Domencich and McFadden, 1975), affects the utility of different travel choices and hence vehicle ownership. For example, destinations that are closer as a result of higher development density or greater land use diversity may be easier to walk or bike to than drive to. Also, origins that are closer to high quality transit, and hence to destinations regionally via transit, render transit a viable alternative to the automobile. People living in such environments will tend to own fewer vehicles. Also, a household's vehicle fleet can be utilized more efficiently when destinations are close by, as trip chaining and carpooling become more practical. Again, a household can meet its travel activity demands with fewer vehicles.

Vehicle ownership is a household-level variable. To capture car shedding behavior, it is important to define a spatial unit that can best capture a household's built environment. It may be a quarter mile network distance around the household, or much greater. However, due to data availability and confidentiality concerns, aggregated D variables at the TAZ, zip code, or census boundary level are more commonly used (Bhat et al., 2013; Cirillo and Liu, 2013; Guo, 2013; Zegras, 2010). The problems with the existing literature include the use of data from a single region, the use of only some of D variables, and the use of different metrics to represent the Ds. These issues restrict our understanding of car shedding phenomenon.

2.3   CURRENT MODELS AND NEW MODEL

2.3.1 State of the Practice in Vehicle Ownership Modeling

To understand the gap between academic research and practical implementation, we conducted a survey of current vehicle ownership-modeling practices at 25 randomly selected (taking a stratified random sample) Metropolitan Planning Organizations (MPO). We contacted the transportation analysts and modelers in each MPO, asked for and reviewed travel model documentation, and asked for the details of travel models if we could not find the answers in the documentation. Summary findings from our survey are presented in Table 2.1. Although we surveyed MPOs with different population sizes, we focused most heavily on large regions since generally, their MPOs are leaders in using new travel modeling techniques.

The results of our survey show that first of all, the four-step process is still being widely used for regional travel demand modeling. As it was mentioned in the previous section, modeling vehicle ownership is not a mandatory step in the traditional four-step modeling and according to Table 2.1, 14 MPOs do not model vehicle ownership (it remains constant across the forecast years). However, all types of tour-based or activity-based models actually model vehicle ownership. It is worth mentioning that in more complex types of activity-based models, even transit pass and parking pass ownership are modeled as well (see Castiglione et al., 2015 for more details).

The results indicate that only two of the MPOs with populations less than 1 million model vehicle ownership which are Chattanooga-Hamilton County/North Georgia Transportation Planning Organization (CHCNGTPO) and Fresno Council of Government (FresCOG). On the other hand, nine out of 13 MPOs with populations greater than 1 million model vehicle ownership and surprisingly, all of them use logit regression for their estimation. Among these MPOs, eight of them use multinomial logit models: CHCNGTPO, FresnoCOG, Wasatch Front Regional Council (WFRC), East-West Gateway Council of Government (EWGCOG), Southeast Michigan Council of Government (SEMCOG), Boston Region MPO, National Capital Region Transportation Planning Board (NCRTPB) and Chicago Metropolitan Agency for Planning (CMAP). One MPO uses a series of binomial logit models, i.e. Mid-America Regional Council (MARC), one uses nested logit, i.e. Ohio-Kentucky-Indiana Regional Council of Government (OKI), and one uses an ordered logit, i.e. Houston-Galveston Area Council (H-GAC).

Seven of these 25 MPOs are working on developing activity based models. SEMCOG and H-GAC now have both four-step travel demand models and activity based models. But, they have not switched to ABM yet and none of them predicts vehicle ownership in their four-step travel demand models.

CHCNGTPO and OKI are the only MPOs in our survey that have already switched to ABM. CHCNGTPO uses multinomial logit and OKI uses nested logit model to predict vehicle ownership. The OKI model has five choices as shown below in Figure 2.1. The alternatives can be nested in several ways to account for a differential similarity across adjacent and non-adjacent alternatives. Based on the variables and the model that OKI has used, it should have one of the most accurate vehicle ownership estimation among all of the 25 MPOs.
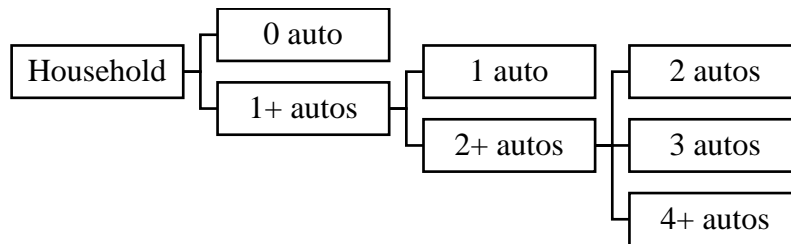
Figure 2.1: Auto Ownership Model Structure of OKI

As it is shown in Table 2.1, vehicle ownership is related mainly to socioeconomic variables and not so much to built environmental variables. To sum up, the results indicate that: 1- The majority of MPOs do not model vehicle ownership, 2- Logit models are the dominant way of predicting vehicle ownership (the problem with these models are discussed in the next subsection) and 3- not much attention has been paid to built environment variables (only one or two of these variables are used, i.e., destination accessibility and density).

Table 2.1: The summary of MPOs models and variables for estimating vehicle ownership

MPO Name	Major City	Population (2010)	Is VO Modeled?	Method and variables used for calculating vehicle ownership
Brunswick MPO	Brunswick	79,626	No	-

RVAMPO	Roanoke	227,507	No	-
Lincoln MPO	Lincoln (Nebraska)	285,407	No	-
North Front Range MPO	Fort Collins	433,178	No	-
CHCNGTPO	Chattanooga	436,669	No	Multinomial Logit Model. Vehicle ownership is sensitive both to various demographic variables such as number of workers, income, number of drivers and accessibility by transit.
ARTS	Augusta	440,134	No	-
Des Moines Area MPO	Urbandale	475,855	No	-
Stanislaus COG	Modesto	514,453	No	-
COMPASS	Meridian	550,359	No	-
AMBAG	Marina	732,667	No	-
CDTC	Albany	823,239	No	-
FresnoCOG	Fresno	930,885	Yes	Multinomial logit model. Variables: household size, housing type, accessibility, household income.
Memphis Urban Area MPO	Memphis	1,077,697	No	-
WFRC	Salt Lake City	1,561,348	Yes	Multinomial logit model. Variables: household size, household income, density of the nearest eight zones, the amount of employment within 30-minutes of transit
METROPLAN Orlando	Orlando	1,837,385	No	-
MARC	Kansas City	1,895,535	Yes	Series of binary logit models. Variables: household income, household size, population density of the TAZ, and highway and transit accessibility from the zone to activity centers.
OKI	Cincinnati	1,981,230	Yes	Nested Logit Model. Variables: Explained in text.
EWGCOG	St. Louis	2,571,253	Yes	Multinomial logit model. Variables: income, household size, worker numbers, as well as highway and transit accessibility.
Boston Region MPO	Boston	3,159,512	Yes	Multinomial logit model. Variables: income (four logit models for four income categories), household size, workers per household, household

				density, employment density, household location, and transit walk-access factors.
SEMCOG	Detroit	4,703,593	No	No in the current model, but yes in the ABM
NCRTPB	Washington	5,068,540	Yes	Multinomial logit model. Variables: household size, household income, area type, and transit accessibility defined as the number of jobs accessible in 45 minutes using the "best" AM transit service. The best transit service is defined as the minimum AM walk-/drive-access transit time among the Metrorail- related transit, i.e. Metrorail only or bus/Metrorail (NCRTB report, 2012)
H-GAC	Houston	5,892,002	No	No in the current model, but yes in the ABM
NCTCOG	Arlington	6,417,630	No	-
NJTPA	Newark	6,579,801	No	-
CMAP	Chicago	8,444,660	Yes	Multinomial logit model. Separate models were estimated and calibrated for three different sized households defined by the total adults (workers plus nonworking adults) in the household. Variables: socioeconomic variables and the location of the household (inner Chicago, rest of Chicago and inner suburbs, mid-suburbs, and far suburbs and fringe).

Abbreviations:

COG: Council of Government
RVAMPO: Roanoke Valley MPO
ARTS: Augusta Regional Transportation Study
CDTC: Capital District Transportation Committee
NCTCOG: North Central Texas COG

COMPASS: Community Planning Association of Southwest Idaho
NJTPA: North Jersey Transportation Planning Authority
AMBAG: Association of Monterey Bay Area Governments

2.3.2 WFRC and MAG's Current Vehicle Ownership Model

As it was explained before, WFRC uses a multinomial logit model (MNL) to forecast vehicle ownership levels based on characteristics of the traveling household and the home location (WFRC/MAG Demand Model Calibration & Validation Report, 2017). It uses household characteristics from the socioeconomic and household income files and land use variables from the employment-within-30-minutes-of-transit and zonal urbanization files to generate auto ownership. This same model is used by Mountainland Association of Governments (MAG).

The autos-by-household size table includes five household categories (1, 2, 3, 4 and 5+ persons per household) and four vehicle categories (0, 1, 2, or 3+ vehicles per household). This

information, along with some summary information, is estimated for every TAZ and is inputted into the trip generation step of the model.

The current model is based on the 2012 household travel survey. The variables determined to be significant in replicating the behavioral characteristics of a household's decision to own or not to own vehicles are the key parameters used in the logit model's utility equations. The constants were calibrated to reflect auto ownership patterns by socioeconomic class from the 2000 Census. All parameters in the utility equations are significant at the 0.05 level, except the parameter for population density for the 2-vehicle choice, which is significant at the 0.10 level.

One problem with this model is its MNL structure. The MNL model treats the number of vehicles owned by a household as a discrete choice, like the choice among discrete modes— driving, taking transit, or walking/biking. That is, it treats vehicle ownership as a nominal variable when, in fact, the number of vehicles owned by a household is a count variable, which can only assume the values of zero, one, two, or some larger positive integer. Although vehicle ownership has been widely modeled as a discrete choice in the literature (Anowar et al., 2014), this may not be the best approach. Since vehicle ownership is a count variable, it seems that a count regression may better fit the data.

Previous studies have done comparisons of model structures, such as MNL, ordered logit (ORL), or ordered probit (ORP), and all have treated vehicle ownership as a discrete choice (Bhat and Pulugurta, 1998; Potoglou and Susilo, 2008). These comparisons have not tested count models – either Poisson or negative binomial – as alternative model structures.

Another problem with the current model is its failure to account for the interdependence of households from the same TAZ. Households are "nested" within TAZs. Households within a given TAZ share the characteristics of that TAZ. This dependence violates the independence assumption of ordinary least squares (OLS) and other types of regression that ignore the nesting structure.

2.3.3 Developing a New Model

This study addresses the issues of existing models in literature and practice in a different manner, by pooling household travel and built environment data from 32 diverse U.S. regions and using a large number of consistently defined and measured built environmental variables to model vehicle ownership. A study using data from, say, Portland, OR, or Houston, TX, can be challenged for relevance to other regions of the country, particularly when different independent variables and models are used in each study. Yet, there are obvious advantages to pooling data in terms of sample size and external validity. A region whose urban form is changing may come to resemble larger and more compact regions over the 20 to 30 years of a travel demand forecast. In this study, improvements to the standard vehicle ownership model include:

- Accounting for the impacts of all D variables on vehicle ownership while controlling for sociodemographic characteristics;
- Using a count regression model (i.e. Poisson regression) along with logit models (i.e. MNL and ORL) and compare the results;

- Using multilevel modeling (MLM) to account for dependence of households in the same TAZ or region on shared TAZ or regional characteristics.

Hence, in this report, we will estimate multilevel MNL, ORL, and Poisson (count regression) models, using all of the D variables to find the best-fit model. Once we find the best-fit model, we will re-estimate the model, using only D variables that can be computed in WFRC/MAG model. The final step will be presenting the results and evaluating the new model relative to the current WFRC/MAG model.

2.4   DATA AND METHODS

2.4.1 Regional Household Travel Survey

The main criterion for inclusion of regions in this study was data availability. Regions had to offer regional household travel surveys with XY coordinates, so we could geocode the precise locations of residences and capture the built environment for households more accurately. It is not easy to assemble databases that meet this criterion, as confidentiality concerns mean that MPOS are often unwilling to share XY travel data.

At present, we have consistent datasets for 32 regions. The resulting pooled dataset consists of 883,695 trips by 91,979 households (see Table 2.2). The average number of household vehicles is 1.92, comparable to 1.74, the national average in 2016 1-year ACS data. The regions are as diverse as Boston and Portland at one end of the urban form continuum and Houston and Atlanta at the other. To our knowledge, this is the largest sample of household travel records ever assembled for such a study outside the National Household Travel Surveys of 2009 and 2017 (NHTS). And relative to NHTS, our database provides much larger samples for individual regions and permits the calculation of a wide array of built environmental variables based on the precise location of households. NHTS provides geocodes (identifies households) only at the census tract level.

Table 2.2: Combined Household Travel Survey Dataset from 32 regions of the U.S.

Regions				
Albany, NY	2009	1,453	12,618	2.02
Atlanta, GA	2011	9,575	93,681	2.11
Boston, MA	2011	7,826	86,915	1.64
Burlington, NC	2009	606	5,111	2.24
Charleston, SC	2009	243	2,098	2.04
Dallas, TX	2009	2,869	27,066	2.05
Denver, CO	2010	5,551	55,056	1.94
Detroit, MI	2005	939	14,690	1.49
Eugene, OR	2011	1,777	16,563	1.82
Greensboro, NC	2009	2,022	17,561	2.09
Hampton Roads-Norfolk, VA	2009	1,957	16,495	2.16
Houston, TX	2008	5,330	59,552	2.27
Indianapolis, IN	2009	3,926	37,473	1.89
Kansas City, MO	2004	3,048	31,779	1.84

Madison, WI	2009	138	1,316	2.12
Miami-Dade, FL	2009	1,428	11,580	1.76
Minneapolis-St. Paul	2010	8,931	79,236	1.81
Orlando, FL	2009	866	7,315	2.00
West Palm Beach, FL	2009	944	7,166	1.70
Phoenix, AZ	2008	4,638	37,811	1.92
Portland, OR	2011	4,513	47,551	1.86
Provo-Orem, UT	2012	1,556	19,255	2.08
Richmond, VA	2009	623	5,123	2.13
Rochester, NY	2011	3,439	23,145	1.81
Salem, OR	2010	1,795	16,231	1.82
Salt Lake City, UT	2012	4,236	44,565	2.04
San Antonio, TX	2007	1,563	14,952	1.90
Seattle, WA	2006	4,965	47,877	1.49
Springfield, MA	2011	850	8,456	1.70
Syracuse, NY	2009	654	5,752	1.94
Tampa, FL	2009	2259	17,538	1.79
Winston-Salem, NC	2009	1,459	12,168	2.15
Total	—	91,979	883,695	1.92

2.4.2 Built Environmental Data

As modal options increase, the need for a second or third household vehicle decreases. Also, as destinations become more accessible to home, vehicles can be used more efficiently, with a carpooling or sequential use of the same vehicle by different household members. Thus, car shedding can occur. All the Ds are represented in our model based on these data:

- Parcel level land use data with detailed land use classifications; from these we can compute detailed measures of land use mix.
- A GIS layer for street networks and intersections; from these we can compute intersection density and percentage of 4-way intersection.
- A GIS layer for transit stops; from these data we can compute transit stop densities.
- Population and employment at the block or block group-level; from these we can compute activity density.
- A GIS layer for TAZs with socioeconomic information (population and employment).
- Travel times for auto and transit travel from TAZ to TAZ (so-called travel time skims); from these, and TAZ employment data, we can compute regional employment accessibility measures for auto and transit.

2.4.3 Variables

The dependent and independent variables used in this study are defined in Table 2.3. Sample sizes and descriptive statistics are also provided. The variables in this study cover most of the Ds, from density to demographics and a total of 11 independent variables is available to explain household vehicle ownership. All variables are consistently defined from region to region.

Table 2.3: Variables Used to Estimate a Vehicle Ownership Model

Variable	Description	N	Mean	S.D.
Dependent variables				
veh	actual number of vehicles owned by household	91,979	1.906	1.045
Independent variables – sociodemographic characteristics				
hhsize_cat	household size of 1,2,3,4 and 5+	91,979	2.403	1.223
employed_cat	number of employed persons in household: 0,1,2, and 3+	91,979	1.184	0.858
dum_income	dummy of income: 1 if lowest income quartile (<35k), 0 otherwise	86,710	0.761	0.427
Independent variables – built environment within TAZs				
actden	activity density within TAZ (pop + emp per square mile in 1000s)	25,735	7.013	21.113
jobpopa	job-population balance within TAZ	25,634	0.545	0.281
intden	intersection density within TAZ	25,729	98.006	80.482
pct4way	percentage of 4-way intersections within TAZ	25,688	25.758	20.106
pctemp10a	percentage of regional employment within 10 minutes by auto	25,686	6.973	11.001
pctemp20a	percentage of regional employment within 20 minutes by auto	25,730	27.449	25.209
pctemp30a	percentage of regional employment within 30 minutes by auto	25,732	49.275	30.175
pctemp30t	percentage of regional employment within 30 minutes by transit	25,732	16.877	21.244

a job-population balance $= 1 - [ABS(employment - 0.2 * population)/(employment + 0.2 * population)]$; ABS = absolute value of expression in parentheses. The value 0.2, representing a balance of employment and population, was found through trial and error to maximize the explanatory power of the variable.

2.4.4 Statistical Analysis

As it was discussed before, to improve the accuracy of WFRC/MAG model and to increase statistical power and external validity, we pooled household data from 32 diverse regions. Our data and model structure are hierarchical, with households "nested" within TAZs and TAZs "nested" within regions. The best statistical approach for nested data is multilevel modeling (MLM), also called hierarchical modeling (HLM). MLM accounts for spatial dependence among observations. OLS and other single-level statistical methods produce biased standard errors and inefficient regression coefficients. MLM overcomes these limitations, accounting for the dependence among observations and producing more accurate coefficient and standard error estimates (Raudenbush and Bryk, 2002).

Households living in a region such as Boston are likely to have very different vehicle ownership characteristics compared to a region such as Houston, regardless of household and neighborhood characteristics. The essence of MLM is to isolate the variance associated with each data level. MLM partitions variance between the household level (Level 1), TAZ level (Level 2) and the regional level (Level 3) and then seeks to explain the variance at each level in terms of D

variables at that level. We can expect to explain a good portion of the variance at Level 1 and Level 2 given the sociodemographic variables and D variables available at these levels. Since we have such a small sample of regions (32 at level 3), we are using fixed effect model to extract all of the variations at this level. In other words, TAZ variance is captured in the random effect term of the Level 2 equation. However, regional variance is captured in the fixed effect term of the Level 3 equation.

The dependent variable we model is a household's vehicle count. We use two discrete choice models, i.e., ORL and MNL since they are being used more frequently by travel demand modelers. Besides, we use count regression model, i.e. Poisson regression model, as well. In principle, two basic regression methods are used to model count variables – Poisson and negative binomial regression. They differ in their assumptions about the distribution of the dependent variable. Poisson regression is appropriate if the dependent variable is equi-dispersed, meaning that the variance of counts is equal to the mean count. Negative binomial regression is appropriate if the dependent variable is over-dispersed, meaning that the variance of counts is greater than the mean count. Popular indicators of over-dispersion are the Pearson and $\chi 2$ statistics divided by the degrees of freedom, so-called dispersion statistics. If these statistics are substantially greater than 1.0, a model is said to be over-dispersed (Hilbe, 2011, pp. 88, 142). By these measures, we have under-dispersion of vehicle counts in our dataset, and the Poisson model is more appropriate than the negative binomial model (see Figure 2). Besides, in the model estimations, only the intercept was allowed to vary randomly across Level 2 units. That is, all of the regression coefficients at Levels 2 were treated as fixed. This is referred to as a random intercept model (Raudenbush and Bryk, 2002).
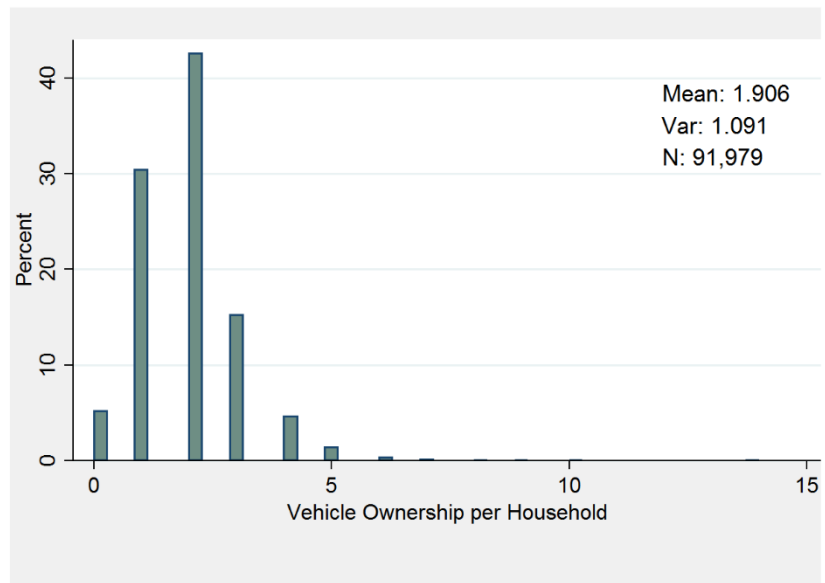
Figure 2.2: The percentage frequency distribution of household vehicle counts

2.5 RESULTS AND EVALUATION

2.5.1 Identifying the Best-Fit Model

The first step is finding the best-fit model. This study adds to the existing literature by comparing two categorical vehicle ownership models, ORL and MNL, and a count data vehicle ownership model, Poisson. Multilevel ORL and MNL models were estimated considering four categories of vehicle ownership: zero, one, two, and 'three or more'. In these three models, we controlled for all of the D variables, even the ones that are not included in Table 2.3, i.e., entropy (measure of land use mix) and transit stop density. By controlling for the socio-demographic variables and all of the D variables, we could better identify the best-fit model. An overall summary of the results for the three models is presented in Table 2.4. Note that all three are fit with fixed region effects and random TAZ effects.

Table 2.4: Summary of the Results for the Three Multilevel Models

	Multinomial Logit	Ordered Logit	Poisson
Log Likelihood (LL(β))	-66107	-68393	-107289
AIC/N	1.443	1.743	2.733
McFadden R2	0.3065	0.2826	0.1540
Correlation(Mean, Veh)	0.6536	0.6527	0.6536
Correlation(IntMean, Veh)	0.6065	0.6039	0.6008
RMSE	0.8964	0.9083	0.8347

Based on this table, MNL has the largest log likelihood or LL(β) and smallest Akaike Information Criterion (AIC). AIC is computed as $-2*LL(\beta)+2K$ where K is the number of parameters in the model (excluding any constant). However, it's not appropriate to compare the logit models with Poisson model on this basis since they are not nested models. Hence, we will compare the correlation measures, which are surprisingly close together.

The computation of the expected number of vehicles in Poisson model is quite straightforward and it is just based on the constant term and the coefficients. However, for the logit models, it is a little tricky. With each estimated model, we computed the expected number of vehicles by computing E[vehicles] = 0*Prob(0) + 1*Prob(1) + 2*Prob(2) + 3*Prob(3+). We then also computed the nearest integer for the expected number of vehicles. To compare the models, we computed the expected value for each model, then the integer nearest to the expected value, and computed the correlation with the actual vehicle count. It is quite surprising how close the three results are. The MNL is very slightly better than the others.

Lastly, we have Root Mean Square Error (RMSE) which measures the standard deviation of the residuals, known as prediction errors. For a given dataset, a lower RMSE shows the better predictive power of the model. Based on the Table 2.4, the Poisson model performs better than the logit models. One of the main reasons that the Poisson model got the edge here is its ability to predict closer to some large values in the sample.

The ORL and MNL models can't use the number of vehicles (veh) as the explained variable - they are inherently categorical, so vehicle category (veh_cat: 0, 1, 2, and 3+) has to be the explained variable. Poisson (or the variants) is a regression model for counts. so veh is the

appropriate left hand side (LHS) variable. The difference is in the top cell. The ORL and MNL models should slightly under predict, simply because they censor the top cell– 3+ is treated as 3. To reduce this prediction error, instead of using 3 for the top cell, we have used the mean of 3 and more vehicles cell which is 3.40. That will mitigate the undercount.

To recapitulate, there are two uses for whatever model got built: (1) Understanding the ownership decision. This means learning responses such as how would vehicle ownership likely change if household size increases, or density decreases. (2) Predicting vehicle ownership. For (1), the behavioral implications of the ordered logit model or the Poisson model are more persuasive. For (2), the three models were extremely similar in how they fit the data. By looking at the correlation, with a very small margin, MNL is the best model. On the other hand, the RMSE of the Poisson model is lower than both MNL and OL. Based on these two uses, we believe that Poisson is the best fit model and for this study, we will use 3-level Poisson model with fixed region effects and random TAZ effects.

2.5.2 Model Results

The best-fit multilevel Poisson regression model for vehicle ownership is shown in Table 2.5. All of the variables are significant at the 0.05 probability level (except employment accessibility within 10 minutes by auto which is significant at 0.06) and also, have the expected signs. The number of vehicles owned by a household increases with household size, number of working members, and household income (1 means low income households). This relationship suggests that bigger households with more workers and higher incomes tend to own more vehicles. We see evidence of car shedding as well. Controlling for socioeconomic variables, vehicle ownership declines with activity density, intersection density, percentage of 4-way intersections, and employment accessibility by auto and transit (percentage of regional employment within 10 and 30-minutes travel time by auto and 30 minutes by transit). These relationships suggest that areas with high population and employment density, good street connections, great transit service, and high accessibility allow direct substitution of transit, walk, and bike travel for automobile travel.

These are variables that we can be confident have a real relationship to vehicle ownership rather than a chance relationship since we have, conservatively, limited our vehicle ownership model to variables significant at the 0.05 level, except the employment accessibility. The McFadden R squared of the model is 0.13. We have shown the pseudo-R2 largely because urban planners are used to dealing with R2s and may want this information. Note that Pseudo-R2s in multilevel Poisson regressions are not equivalent to R2s in ordinary least squares regression, and should not be interpreted the same way. The pseudo-R2 bears some resemblance to the statistic used to test the hypothesis that all coefficients in the model are zero, but there is no construction under which it is a measure of how well the model predicts the outcome variable in the way that R2 does in conventional regression analysis. The goodness of fit and validation of the model are shown in the following section.

Table 2.5: The Results of Three-level Poisson Regression

	coef.	std. err.	t-ratio	p-value
(Intercept)	0.31380	0.02011	15.6	< 2e-16

hhsize_cat1*	0.56480	0.01116	50.619	< 2e-16
hhsize_cat2	0.46790	0.00830	56.363	< 2e-16
hhsize_cat3	0.52560	0.00977	53.795	< 2e-16
hhsize_cat4	0.52060	0.01009	51.619	< 2e-16
employed_cat0**	0.48850	0.01171	41.735	< 2e-16
employed_cat1	0.08804	0.00770	11.44	< 2e-16
employed_cat2	0.19350	0.00825	23.448	< 2e-16
dum_income	-0.27520	0.00737	-37.356	< 2e-16
actden	-0.00597	0.00040	-15.04	< 2e-16
intden	-0.00064	0.00005	-12.361	< 2e-16
pct4way	-0.00083	0.00017	-4.919	8.70E-07
pctemp10a	-0.00065	0.00035	-1.827	0.06764
pctemp30a	-0.00094	0.00017	-5.646	1.65E-08
pctemp30t	-0.00108	0.00018	-6.132	8.69E-10
Salt Lake Region	0.04905	0.01864	2.631	0.00851
Provo-Orem Region	0.01316	0.02474	0.532	0.59481

Sample size: **level 1 – 86489**
        level 2 – 25205
        level 3 – 32

Log likelihood (Full): **-119390.7**

Log likelihood (Null): **-138206.7**

AIC: **238972.5**

BIC: **239131.7**

McFadden R2: **0.1361**

* Household size of 5 is the reference category.

** Three or more employers in a household is the reference category.

An elasticity is a percentage change in one variable with respect to 1% change in another variable. For a count model, the elasticity is just equal to the regression coefficient times the mean value of the independent variable. Thus, for the built environment variables in the best-fitting Poisson model, we compute elasticities of:

Elasticity of vehicle ownership w.r.t. activity density = -0.0059 * 7.013= -0.0413

Elasticity of vehicle ownership w.r.t. intersection density = -0.00064 * 98.006= -0.0627

Elasticity of vehicle ownership w.r.t. percentage of 4-way intersections = -0.00083 * 25.758 = -0.0213

Elasticity of vehicle ownership w.r.t. employment accessibility by auto (within 10 minutes) = -0.00065 * 6.973 = -0.0045

Elasticity of vehicle ownership w.r.t. employment accessibility by auto (within 30 minutes) = -0.00094 * 49.275 = -0.0463

Elasticity of vehicle ownership w.r.t. employment accessibility by transit = -0.00108 * 16.877 = -0.0182

The elasticities of built environmental variables are relatively small in the model, but still significant. Viewed another way, for example, the percentage of regional employment accessible within 30 minutes by transit for our sample ranges from 0 to 99.21. The difference between the household that has 0% access to regional employment within 30 minutes by transit and the one who has 99.21% access is +0.11. So, for a household that is average in all other respects, vehicle ownership will drop from 1.92 to 1.81 as accessibility by transit climbs from the lowest value 0, to the highest value 99.21.

2.6   MODEL VALIDATION

Our approach is theoretically more solid in the sense that it incorporates influential built environment characteristics of TAZ and uses disaggregate data at the individual household level from various U.S. regions. To be used in practical modeling, however, we need to validate our model in comparison with the multinomial logit model used by WFRC. In other words, does our model outperform the current model of WFRC?

Since WFRC ultimately models the average number of vehicles for each of the TAZs, our unit of analysis is the TAZ. The modeled values are compared against the actual average of vehicle ownership by TAZ for the Wasatch Front from the 2012 Utah Travel Survey.

The problem with this approach is that many TAZs have no or only a few households. This raises sampling error issues, meaning that the small number of households in the survey cannot represent all households residing in that TAZ. For instance, only one household in a TAZ that has four cars cannot be a good representative of all households living in that TAZ. Or if a household does not have a vehicle, it doesn't mean that all households have no cars in that specific TAZ. Hence, in order to minimize this sampling error issue, we tried different values for the minimum number of households in a TAZ and determined 10 as a final threshold value for model validation purposes. As it is shown in Table 2.6, even if we don't define this threshold, still our model outperforms the WFRC model.

The correlation between the predicted value versus the actual number of vehicles, along with the root mean square error (RMSE) which were explained in the previous chapters are appropriate measure of model prediction quality between two continuous variables (in this case, the average number of vehicles in TAZs from the survey vs. the model). RMSE is a frequently used measure of the differences between values predicted by a model and the values actually observed. RMSE is a measure of accuracy, to compare forecasting errors of different models for a particular dataset. The smaller the RMSE, the more accurate the model (and the better the predictive power). The RMSE of our model is 0.2293 while this number for the WFRC model is 0.9243. On the other hand, the correlation between the predicted values and the actual average number of vehicles in TAZs in the best-fit model is 0.8506, while this value is only 0.08 in the WFRC model. Based on these results, we can conclude that our model performs way better than the WFRC model.

Table 2.6: Summary of the results

	Best-Fit Model	WFRC Model
RMSE for All TAZs	0.5274	1.1431
Correlation (Predicted vs. Actual) for All TAZs	0.6557	0.0276

RMSE for TAZs With 10 or More Households	0.2293	0.9243
Correlation (Predicted vs. Actual) for TAZs With 10 or More Households	0.8506	0.0882

2.7   CONCLUSIONS

This study estimates a vehicle ownership model using regional household travel data and built environmental variables from 32 diverse regions across the Unites States. The household ownership model is estimated with multilevel Poisson regression. The results show that household vehicle ownership has positive relationships with household size, number of household workers, and household income. Household vehicle ownership has negative relationships with several built environmental variables. Although the elasticities of built environmental variables are smaller than the elasticities of the socioeconomic variables, all are highly significant. Vehicle ownership decreases with activity density, intersection density, percentage of 4-way intersections, and destination accessibility after controlling sociodemographic variables. These findings are consistent with the literature on car shedding.

Such a large dataset also gives the models external validity missing from earlier studies. The model developed in this study can be directly used for travel demand modeling and forecasting not only by WFRC but also by MPOs in other regions of the U.S., especially those medium and small MPOs that have limited resources to collect household travel survey data and estimate a vehicle ownership model of their own.

Based on the results of this study, we would recommend using a count model (Poisson) model over a categorical model (multinomial logit). By comparing the MNL and ORL models, where vehicle ownership is treated as a categorical variable, with the Poisson model, where vehicle ownership is treated as a count variable, this study shows that the Poisson model has slightly better predictive accuracy than the MNL model.

For the urban planning and design practices, this study suggests that car shedding occurs as built environments become more dense, mixed, connected, and transit served. This finding has important implications in the policy and planning practice, where decision makers seek solutions to deal with VMT, emissions, obesity, and other health and environmental concerns.

In terms of limitations, although it covers the standard D variables, this study still omits certain variables that have presumptive effects on household vehicle ownership. Parking supplies and prices, travel attitudes, and residential self-selection may strongly affect household vehicle ownership. A study in New York City shows that free residential street parking increases private car ownership by as much as 9% (Guo, 2013). Individuals who would like to own fewer vehicles and want to use alternative modes may choose to live in neighborhoods that support such lifestyle choices. We have no ability to control for these self-selection effects in this multi-region study, as most of the underlying household surveys do not include relevant attitudinal questions. Failure to control for these effects may lead to erroneous estimates of model parameters that may result in overestimating or underestimating the impact of built environment changes on vehicle ownership. We have elsewhere argued that self-selection effects are small compared to built

environmental effects, and that self-selection is as likely to result in enhanced as attenuated built environmental effects (Ewing and Cervero, 2010; Ewing and Cervero, 2017).

3.0   INTRAZONAL OR INTERZONAL? IMPROVING INTRAZONAL TRAVEL FORECASTS IN A FOUR-STEP TRAVEL DEMAND MODEL

3.1   INTRODUCTION

A major weakness of conventional travel demand models is that they tend to predict intrazonal trips with poor accuracy. To quote a reviewer of this chapter, "The limitations of the gravity model are well known and it cannot be expected to deal with trips that travel what is really an unknown average distance. Practitioners have tried to overcome this limitation mostly with heuristic approaches to estimate an average travel distance for intrazonal trips. The main reason for this is that intrazonal trips are not particularly interesting in themselves but their number affects all the other interzonal trips estimated by a trip distribution or destination choice model."

Trips are classified as intrazonal if their origin and destination are contained within the same traffic analysis zone (TAZ). Intrazonal trips are a minor consideration in the four-step travel demand modeling process, despite the fact that they typically amount to 10% or more of all trips in household travel surveys. They are treated like any other zonal interchange in the trip distribution step. Trip productions and attractions are modeled as occurring at a single point in the four-step model, the zone centroid, and the entire local street network on which intrazonal trips occur is reduced to one or more centroid connectors to the external street network. This means that intrazonal trips should be modeled differently than interzonal trips.

This chapter presents a new method for modeling intrazonal trips that addresses the major identified shortcomings of traditional approaches to intrazonal trip modeling in two ways. First, we employ a novel dataset with disaggregated travel survey data coupled with TAZ-specific built environmental measurements. This rich dataset allows us to account for differences in important built environment measures like activity density, street connectivity, and mixed land uses and how they impact intrazonal trip making. The second significant improvement over standard intrazonal modeling efforts is the use of discrete choice modeling. Where traditional methods employ the gravity model which merely measures the attraction potential of a destination less its impedance from an origin on a uniform, aggregated network, discrete choice modeling actually integrates elements of behavior and utility maximization. We use binomial logistic regression, which models the decision of whether to stay within the zone or to leave, as a discrete choice dependent on built environment characteristics within the traffic analysis zone. This method more accurately represents the behavioral aspects inherent in individual travel decision making.

Our chapter proceeds as follows. First, we discuss the most common method in use for trip distribution within and across transportation analysis zones, namely the gravity model, and known limitations of the method. Then we present results from a survey of 25 MPOs of different sizes from across the US, determining their method-in-use for distributing trips. Then we describe our new method, developed as a substitute and improvement upon the commonly used approach. Finally, we present results using our method, validate the models, and conclude with their implementation.

3.2   LIMITATIONS OF THE GRAVITY MODEL

Various methods have been developed for forecasting intrazonal trips as a component of conventional four-step modeling. However, limitations of the methods raise concerns about the ability of conventional travel demand modeling to adequately account for intrazonal trips. This section considers some methods in common use and their limitations.

One of the most glaring issues with travel demand modeling and the gravity model is that it is done at a relatively aggregate level. Hamilton (1989) was one of the first to point out this issue, stating that as data become more aggregated the assumptions on which the models are run become more and more compromised. Varying sizes of TAZs could lead to differing likelihoods that trips will be intrazonal (Hamilton, 1989; Moeckel and Donnely, 2015; Okrah, 2016).

Cervero (2006) provides a critique of the conventional approach to four-step modeling that makes a similar point, while also emphasizing the importance of considering localized information on built environment characteristics. He asserts that in the conventional four-step process, "fine-grained land use mixes, local street connectivity, and pedestrian amenities, do not influence intrazonal trip estimates." This is a general criticism of four-step models, but is particularly apropos to the modeling of intrazonal trips. The failure to consider local land use and street network patterns potentially leads to an underprediction of intrazonal trip rates in densely developed areas.

Research investigating intrazonal travel empirically in relation to characteristics of the local built environment is scant, but some findings are pertinent to this discussion. Modeling intrazonal travel in Gainesville, Florida, Ewing and Tilbury (2002) found that built environment variables (the D variables of development density, land use diversity, street network design, destination accessibility, and distance to transit) rival or sometimes exceed the explanatory power of the gravity formula used to estimate intrazonal trips in a conventional four-step model. This finding has two implications: first, that conventional models are ill-suited to predict intrazonal trips, and second, that sketch planning models that account for these other variables can correct the problem to a degree. One land-use variable, an entropy measure, appeared consistently significant in their models of intrazonal travel for different trip purposes. This variable, derived from Property Appraisers' parcel-level data using GIS, captured the following mix of land uses: pedestrian-oriented retail uses; finance, insurance, and real estate offices; general office buildings; and commercial lodging. Also, highly significant in the authors' models was the presence of a grocery store (for home-based shopping and non-home-based trips) and a public school (for home-based social-recreational and other trips).

Examining intrazonal trip characteristics, Greenwald (2006) found that mode choice for these trips is affected by urban form. The choice of mode, in turn, then affects trip distribution, as non-motorized trips are more likely to stay close to their origin. However, as Greenwald cautions, there is a threshold effect in the ability of the built environment to affect travel behavior; at some point, changes to the economic diversity of a TAZ start showing decreasing impacts on mode choice.

Although research is limited on intrazonal travel measured empirically in relation to D variables, there has been more work on methods for forecasting intrazonal travel as a component of the

four-step model. The trip distribution step in the conventional four-step model relies on measuring trip impedance, essentially a measure of the time it will take to travel from a trip origin to a destination. The most common method for capturing impedance is to employ a gravity model, but the standard gravity model disregards local land use and street network patterns. Facile approaches to intrazonal trip distribution are common, including the use of uniform intrazonal trip rates derived from travel surveys as well as simple runs of a gravity model. In the latter case, impedances must be estimated based on intrazonal travel times. Impedances for intrazonal trips are technically zero in the four-step model, since both origins and destinations are located at the same point in space, the zone centroid (Horner and Murray 2001; Bhatta & Larsen 2011). Therefore, intrazonal travel times must be crudely approximated, usually by factoring the size of a TAZ or travel time to adjacent zones.

The traditional four-step model treats intrazonal trips exactly like all trips within the trip distribution step. The basic approach is to use a gravity model to determine the number and proportion of trips being made from a specific origin zone to a specific destination zone. The gravity model works under the assumption that the trips produced at an origin and attracted to a destination are directly proportional to the number of trip productions at the origin and the number of trip attractions at the destination, and inversely proportional to the travel time impedance between the original and destination. The standard form of the gravity model is depicted below:

$$T_{ij} = P_i \frac{A_j F_{ij} K_{ij}}{\sum_x A_x F_{ix} K_{ix}}$$ **(Equation 3-1)**

where T_{ij} is trips produced at i and attracted at j; P_i is total trip production at i; A_j is a total trip attraction at j; F_{ij} is the travel impedance between i and j; K_{ij} is the socioeconomic adjustment factor for interchange ij (Anas 1985).

A relatively large body of literature has been published on techniques for estimating intrazonal impedances in the gravity model, in other words for estimating the F_{ij} values in the above formula. Early methods were based on assumptions that vastly simplified the problem, such as one advanced by Batty (1976). In this method, Batty assumed a constant population density over an evenly spread circular zone. His equation for estimating intrazonal travel cost was as follows:

$$c_{ii} = \frac{r_i}{\sqrt{2}}$$ **(Equation 3-2)**

where c_{ii} is travel cost and r_i is the radius of the zone.

Venigalla et al. (1999) suggest a relatively simple method in which intrazonal trip impedance is calculated by merely dividing the trip length and time to the nearest zone centroid in half, sometimes referred to as the nearest neighbor approximation. Others have assumed that intrazonal travel time is two-thirds the time to the nearest neighboring zone, or equal to a set fraction of the average travel time to two or more adjacent zones.

These methods have obvious shortcomings, such as the necessity to make assumptions that zones are circular in shape and demonstrate homogeneous population densities. A marginal improvement to this method was made by Dowling et al. (2005), who divided each zone into 13 concentric squares. The authors then determined mean distance by averaging the distances from the zone centroid to the perimeter of each of the squares. Finally, they used a table of speeds by area type and time of day to compute travel time from the intrazonal distances.

In some regions, the method of calculating intrazonal impedance is based on the zone's total area as well as the average travel speed of the zone. This approach is one of the earliest to be developed (Lamb 1970). The average intrazonal trip distance is approximated by one half of the square root of the zone's area, and the conversion to time in minutes is made with the intrazonal speed in miles per hour and the constant 60 to convert hours into minutes (Martin & Mcguckin 1998).

$$Intrazonal\ Time = \frac{0.5 \times \sqrt{(zonal\ Area)} \times 60}{Intrazonal\ Speed\ (Area\ Type)}$$ **(Equation 3-3)**

Whatever approximation is used, the result flies in the face of findings from our empirical research. Using the gravity model, the larger the zone area is, the greater the impedance is and the smaller the proportion of intrazonal trips becomes. In fact, however, we determined empirically that all else being equal, larger zones capture a higher proportion of total trips generated within the zone. We discuss our research findings on this topic in more detail below.

3.3   STATE OF THE PRACTICE IN INTRAZONAL TRAVEL MODELING

To understand the gap between academic research and practical implementation, we conducted a survey of current intrazonal travel-modeling practices at 25 MPOs in the U.S. We selected MPOs with various population sizes: three MPOs with a service area population of less than 300,000, nine MPOs between 300,000 and 1 million, and 13 MPOs with more than 1 million population. We focused mostly on large regions because we assume that their MPOs are leaders in using new travel modeling techniques.

The survey findings are presented in Table 3.1 with their population size, trip distribution model, and intrazonal trip forecast method. The results of our survey show that the four-step travel demand modeling process is still being widely used for regional travel modeling. All surveyed MPOs use the conventional four-step model.

The model that is used most commonly for estimating trip distribution is the gravity model. Out of 25, 20 MPOs use the gravity model for trip distribution – both intrazonal and interzonal. The next most widely used method is the destination choice model, a type of trip distribution or spatial interaction model, which is formulated as a discrete choice model, typically employing a logit model. The destination choice model can be thought of as a generalization of the gravity model. In the gravity model, most MPOs use nearest neighbor approximations for calculating the intrazonal travel time, while the number of adjacent zones included in the equation varies from one (the nearest zone; e.g., COMPASS, StanCOG) to four (e.g., ARTS, CHCNGTPO, Memphis, Brunswick).

Basically, the MPOs treat intrazonal trips just like interzonal trips, and the only zone-specific attributes accounted for are trip productions at the zone centroid, trip attractions at the zone centroid, and a crude estimate of intrazonal travel time to create separation between the two – except for CMAP which is not based on the travel time (see Table 3.1). It is worth mentioning that six of them (FresnoCOG, NCTCOG, SEMCOG, OKI, NJTPA and CMAP) are working on activity-based modeling, which is the state-of-the-art in travel modeling. While some of them are almost done with this process, they have not completely switched to ABM yet.

Table 3.1: The summary of MPOs methods for calculating trip distribution and intrazonal trips (as of March 2018; sorted by population size)

MPO Name	Major City	Population (2010)	Trip Distribution Model	Method for Calculating Intrazonal Trips
CMAP	Chicago, IL	8,444,660	Gravity with Intervening Opportunities	Both inter-zonal and intra-zonal trips are modeled together based on zone size, trip cost, and available destinations, and then separated based on impedance (time, cost, etc.)
NJTPA	Newark, NJ	6,579,801	Gravity	The intrazonal time was calculated using half of the sum of time from two closest "nonzero" zones, and then multiplied it by 0.60
NCTCOG	Arlington, TX	6,417,630	Gravity	Nearest Neighbor Rule (0.5 of three zones)
H-GAC	Houston, TX	5,892,002	Atomistic Model (a gravity-analogy-based model)	…by dividing existing zones into atoms a more realistic interchange of intrazonal trips and short (less than five minutes) trips among adjacent zones is defined
NCRTPB	Washington, DC	5,068,540	Gravity	The intra-zonal times have been set to 85% of the minimum inter-zonal time
SEMCOG	Detroit, MI	4,703,593	Destination Choice Model	Intra-zonal travel time is calculated based on 4 nearest neighbor zones
Boston Region MPO	Boston, MA	3,159,512	Gravity	Nearest neighbor rule (0.5 of 3 zones)
EWGCOG	St. Louis, MO	2,571,253	Gravity/ Destination Choice Model	For Home-Based Work: Gravity model/ For Other types: Destination choice model
OKI	Cincinnati, OH	1,981,230	Gravity	Half of the average travel time to the nearest three zones
MARC	Kansas City, MO	1,895,535	Destination choice model	The nearest neighbor rule was used to estimate the intrazonal travel times
METROPLAN Orlando	Orlando, FL	1,837,385	Gravity	The nearest neighbor rule with terminal time as the constraining variable.
WFRC	Salt Lake City, UT	1,561,348	Gravity	Intrazonal travel time as a function of the area of the zone and the average travel speed
Memphis Urban Area MPO	Memphis, TN	1,077,697	Destination Choice Model	The intrazonal travel times are computed by taking half the average travel time to the four closest neighboring zones
FresnoCOG	Fresno, CA	930,885	Gravity	100% and 33.3% the average time to the nearest adjacent TAZ for urban and rural areas, respectively

CDTC	Albany, NY	823,239	Gravity	A travel time of 6 minutes is assumed for intrazonal trips (trips within the same zone)
AMBAG	Marina, CA	732,667	Gravity	Intra-zonal travel times were computed based on the average time to the nearest 3 zones
COMPASS	Meridian, ID	550,359	Gravity	Travel times: 50% time to the nearest zone
Stanislaus COG	Modesto, CA	514,453	Gravity	Intrazonal travel times are estimated based on 50% of the travel time to the nearest adjacent zone
Des Moines Area MPO	Urbandale, IA	475,855	Gravity	Three neighbor zones for the calculation of average travel time were chosen and a final factor, 0.5, was applied to the end result
ARTS	Augusta, GA	440,134	Gravity	Intrazonal times were created by the travel purpose+ Matrix function using half of the average travel time to the nearest four TAZ's
CHCNGTPO	Chattanooga, TN	436,669	Destination Choice Model	The intrazonal travel time is calculated as half the average travel time to the four closest neighboring zones
North Front Range MPO	Fort Collins, CO	433,178	Gravity	Intrazonal travel time is calculated as a function of the travel time required to reach the closest adjoining zone
Lincoln MPO	Lincoln, NE	285,407	Gravity	Intrazonal travel time has been calculated by multiplying the distance to the single nearest neighbor by 75%
RVAMPO	Roanoke, VA	227,507	Gravity	Two adjacent zones are used to compute the intrazonal travel time during the trip distributions
Brunswick MPO	Brunswick, GA	79,626	Gravity	Intrazonal times were created by the Travel Purpose + Matrix function using half of the average travel time to the nearest four TAZ's

3.4   METHODOLOGY

3.4.1 Data

For 31 regions (Table 3.2), household travel surveys were collected from MPOs. The surveys were conducted between 2006 and 2012. While conducted by individual regional organizations such as metropolitan planning organizations (MPOs) or State Departments of Transportation, the regional household travel surveys have quite similar structure and questions, akin to U.S. DOT's National Household Travel Survey (NHTS). To gather comprehensive data on travel and transportation patterns, the survey data consistently includes, but is not limited to, household demographic information, vehicle ownership information, and data about one-way trips taken during a designated 24-hour period on a weekday, including travel time, mode of transportation, and purpose of trip information. The survey data have exact XY coordinates so we could geocode the precise locations of households and the precise origins and destinations of trips. The regional survey data were acquired from individual MPOs or state DOTs with confidentiality agreements. The pooled dataset consists of 843,287 trips produced by 89,768 households within 25,469 traffic analysis zones (TAZs) in 31 regions.

The 843,287 trips were classified as either intrazonal (produced and attracted within the same TAZ) or interzonal trips (produced in one TAZ and attracted to another). On average, intrazonal

trips account for 10.7% of total trips. This is a significant share of total trips. We computed intrazonal trip shares by trip purpose from the regional household travel surveys. The result is presented in Table 3.2. The shares vary from region to region. For example, intrazonal home-based work trips make up only 2.9% of all home-based work trips on average, ranging from 1.3% in Eugene to 5.9% in Madison. Intrazonal home-based other trips (excluding work and and shopping-related ones) make up 14.4% of all home-based other trips on average, ranging from 7.4% in Eugene to 26.0% in Palm Beach. This large variance may reflect differences in zone size, land use and street network patterns, or even socio-demographics. The need to model intrazonal travel, in terms of these variables, is evident. In this paper, we show results from modeling intrazonal travel in relation to the D variables for the 31 regions, based on the regional household travel surveys.

Table 3.2: Percentage of Intrazonal travel by trip purpose from travel surveys

	HBW	HBShp	HBOth	NHBW	NHBNW
Albany, NY	3.2	8.5	21.9	9.5	15.0
Atlanta, GA	3.4	9.8	17.4	10.6	15.9
Boston, MA	2.9	7.3	15.3	10.6	12.6
Burlington, NC	4.5	4.4	13.1	10.3	11.0
Dallas, TX	2.3	6.4	15.9	7.7	11.6
Denver, CO	2.8	4.6	11.5	8.0	11.6
Detroit, MI	2.0	8.9	9.6	6.2	9.9
Eugene, OR	1.3	3.2	7.4	7.1	8.2
Greensboro, NC	1.9	5.0	15.1	8.7	12.0
Hampton Roads–Norfolk, VA	2.8	7.8	19.4	11.4	14.6
Houston, TX	3.1	8.4	14.7	6.5	11.8
Indianapolis, IN	2.5	3.8	11.0	7.4	12.7
Kansas City, MO	4.8	11.0	16.8	9.8	15.1
Madison, WI	5.9	4.8	13.8	12.6	13.0
Miami, FL	1.7	5.0	13.4	6.7	10.9
Minneapolis–St. Paul, MN-WI	3.0	5.2	9.0	7.8	12.3
Orlando, FL	2.1	6.2	21.8	9.5	12.5
Palm Beach, FL	2.6	8.0	26.0	9.3	11.9
Phoenix, AZ	2.8	10.5	20.2	9.3	13.5

Portland, OR	3.3	7.8	14.9	16.7	17.1
Provo-Orem, UT	3.3	4.6	19.1	6.6	10.5
Richmond, VA	2.2	5.6	17.9	9.9	11.1
Rochester, NY	2.8	5.7	9.3	5.8	12.2
Salem, OR	2.4	0.9	8.7	6.7	9.6
Salt Lake City, UT	2.7	4.2	15.0	6.2	10.6
San Antonio, TX	2.8	5.5	10.9	6.2	10.7
Seattle, WA	1.5	7.0	11.1	10.5	10.0
Springfield, MA	4.0	8.2	15.2	16.3	17.5
Syracuse, NY	1.4	5.9	15.7	7.6	10.7
Tampa, FL	4.2	8.3	21.5	8.3	12.7
Winston-Salem, NC	3.2	4.5	14.0	5.7	11.1
Total	**2.9**	**6.9**	**14.4**	**9.2**	**12.8**

Also, we collected land use data at the parcel level with detailed land use classifications, so we could study land use intensity and mix down to the parcel level for the same year as the household travel survey. We also gathered GIS data layers for streets, population and employment for TAZs, and travel times between zones by different modes, again for the same years as the household travel survey. Built environmental variables were computed for each TAZ and assigned to households within the TAZ.

3.4.2 Variables

In this study, the D variables of the built environment were measured and used to predict the intrazonal travel. The measurement of the D variables and their expected effect on travel behavior are summarized in Table 3.3. Some dimensions capture closely related qualities (e.g., diversity and destination accessibility). Still, it is a useful framework used to organize the empirical literature and provide order-of-magnitude insights (Ewing and Cervero 2010). The dependent and independent variables used in this study are defined in Table 3.4. Sample sizes and descriptive statistics are also provided.

For home-based trip (home-based-work, home-based-shopping, and home-based-other) models, the D variables of the TAZ where the home is located were used to characterize the built environment of the TAZ. For the non-home-based-work trip model, the D variables of the TAZ where the workplace is located were used to characterize the built environment of the TAZ. For the non-home-based-non-work trip model, the D variables of the TAZ where the trip origin is located were used to characterize the built environment of the TAZ.

Table 3.3: The D Variables (Ewing et al. 2015)

D Variable	Measurement
Density	Density is always measured as the variable of interest per unit of area. The area can be gross or net, and the variable of interest can be population, dwelling units, employment, or building floor area. Population and employment are sometimes summed to compute an overall activity density per areal unit.
Diversity	Diversity measures pertain to the number of different land uses in a given area and the degree to which they are balanced in land area, floor area, or employment. Entropy measures of diversity, wherein low values indicate single-use environments and higher values more varied land uses, are widely used in travel studies. Jobs-to-housing or jobs-to-population ratios are less frequently used.
Design	Design measures include average block size, proportion of four-way intersections, and number of intersections per square mile. Design is also occasionally measured as sidewalk coverage (share of block faces with sidewalks); average building setbacks; average street widths; or numbers of pedestrian crossings, street trees, or other physical variables that differentiate pedestrian-oriented environments from auto-oriented ones.
Destination accessibility	Destination accessibility measures ease of access to trip attractions. It may be regional or local (Handy 1993). In some studies, regional accessibility is simply distance to the central business district. In others, it is the number of jobs or other attractions reachable within a given travel time, which tends to be highest at central locations and lowest at peripheral ones. The gravity model of trip attraction measures destination accessibility. Local accessibility is a different animal. Handy (1993) defines local accessibility as distance from home to the closest store.
Distance to transit	Distance to transit is usually measured as an average of the shortest street routes from the residences or workplaces to the nearest rail station or bus stop. Alternatively, it may be measured as transit route density, distance between transit stops, or the number of stations per unit area. In this literature, frequency and quality of transit service are overlooked.

Table 3.4: Descriptive statistics for our variables

Variable	Description	N	Mean	Median	S.D.
Outcome Variable (level 1: trip)					
Intrazonal	trip remaining internal to TAZ (1=intrazonal, 0=interzonal)	843,287	0.11	0.00	0.31
trip purpose	five trip purpose: home-based-work (HBW), home-based-shopping (HBShp), home-based-other (HBOth), non-home-based-work (NHBW), non-home-based-non-work (NHBNW)	-	-	-	-
Explanatory Variables (level 2: TAZ)					
totpop	total population within TAZ	25,396	1,832.76	1406.00	1,664.44

totemp	total employment within TAZ	25,396	611.60	283.00	1,065.82
area	gross land area of TAZ in square miles	25,396	1.82	0.50	10.57
actden	activity density within TAZ (pop + emp per square mile in 1000s)	25,396	7.05	4.29	21.14
jobpop[a]	job-population balance within TAZ	25,396	0.55	0.57	0.28
Intden	intersection density within TAZ	25,396	98.39	85.47	80.52
pct4wy	percentage of 4-way intersections within TAZ	25,396	25.80	20.22	20.10
pctemp10a	percentage of regional employment within 10 minutes by car	25,396	6.93	3.06	11.01
pctemp20a	percentage of regional employment within 20 minutes by car	25,396	27.4	18.9	25.2
pctemp30a	percentage of regional employment within 30 minutes by car	25,396	49.3	50.5	30.2
pctemp30t	percentage of regional employment within 30 minutes by transit	25,396	16.81	7.84	21.26

(a) JOBPOP = 1 − [ABS(employment − 0.2 * population)/(employment + 0.2 * population)], where ABS is absolute value of expression in parentheses (Ewing et al., 2015). The value 0.2, representing a balance of employment and population, was found through trial and error to maximize the explanatory power of the variable.

3.4.3 Analysis Methods

We treated intrazonal/interzonal travel as a binary choice, and hence modeled it with multilevel binomial logistic regression. We modeled intrazonal travel for the 31 regions. A binomial logistic regression predicts the probability that an observation falls into one of two categories of a dichotomous dependent variable (intrazonal or interzonal travel, in this case) based on multiple independent variables (in our case, the TAZ-level D variables and the three regional variables).

A three-level model was required to represent the nested nature of the dataset, with multiple trips nested within TAZs and TAZs nested within regions. Multilevel modeling accounts for dependence among observations. All trips within a given TAZ share TAZ characteristics and all TAZs within a given region share regional characteristics. This dependence violates the independence assumption of standard regression. Standard errors of regression coefficients will consequently be underestimated. Moreover, coefficient estimates will be inefficient. Multilevel models overcome these limitations, producing more accurate coefficient and standard error estimates (Raudenbush and Bryk 2002). The three-level model used in this study partitions variance among the trip level (Level 1), the TAZ level (Level 2), and the regional level (Level 3) and uses level-specific variables to explain the variance at each level.

A multi-level model is implemented the same way as a single-level model; values of the independent variables are substituted for the variables in equations, multiplied by coefficients, and summed to get the log odds. Then, by exponentiating the log-odds, we can compute the odds

of intrazonal trips and the probability of intrazonal trips, which is equal to (odds of intrazonal trips / (1 + odds of intrazonal trips)).

The final models were chosen based on three considerations – 1) whether the sign of a coefficient is expected or not (for example, total employment in a TAZ is expected to have a positive relationship with the share of intrazonal trips. If not, we drop that variable), 2) statistical significance of the explanatory variable, and 3) the overall model fit based on the pseudo-R-squared values.

3.4.4 Model Validation

To test how well the intrazonal models are able to predict intrazonal travel, we evaluated the predictive performance of our five models—one for each trip purpose—by running k-fold cross-validation on our datasets (Fielding and Bell, 1997; Hair et al., 1998). Using the same data to estimate parameters and to test predictive accuracy may overestimate model validity. In k-fold cross-validation, the data are divided into k equal partitions. One partition is withheld, and the model is fitted with the remaining data. As Borra and Ciaccio (2010) suggest, data were randomly divided into ten folds: 90% of the data (training data) used for model fitting and 10% of the data withheld for model validation in each iteration.

The receiver operating characteristic (ROC) curves and the areas under ROC curves (AUC) are appropriate measures to evaluate prediction capability of logistic regression models (Greiner et al., 2000; Hanley and McNeil, 1982; Meng, 2014; Zweig and Campbell, 1993). For the ROC curves, the rate of true-positives is plotted on the vertical axis and the rate of false-positives is plotted on the horizontal axis. Then the ROC statistics, AUC, provides the predictive accuracy of the logistic models, with values from 0.5 (no predictive power) to 1.0 (perfect prediction). In this study, the ROC curves were first used to visualize prediction capability of our models using only the left-out partition that was not used in model fitting. Predictive accuracy is then assessed by calculating the areas under ROC curves (AUC). This procedure is repeated for each of the k partitions, and the AUC values are averaged to obtain the mean AUC value.

In addition to the k-fold validation, we also validate our models against a conventional practice—the gravity model. How much more accurate is our model than the gravity model? Instead of modeling it, there are a few regions using a constant value, a region-wide proportion of intrazonal trips by trip purpose, to estimate intrazonal trip distribution. Is our model better than that simplest approach?

To prove the validity of our model, we compare our model with two other models – a gravity model and a constant model (using a region-wide average proportion of intrazonal trips by trip purpose) using data from two regional MPOs—Wasatch Front Regional Council (WFRC) and Mountainland Association of Governments (MAG). Two regions are selected because we can obtain intrazonal proportions by TAZ from their gravity models. Thus, our unit of analysis is the TAZ. The modeled values are compared against the actual proportion of intrazonal trips by trip purpose by TAZ from the 2012 Utah Household Travel Survey.

The problem with this approach is that many TAZs have no or only a few trips. This raises sampling error issues, meaning that the small number of trips in the survey cannot represent all

trips occurring in that TAZ. For example, if a TAZ has only one trip (which is internal) from the survey, it gets 100% intrazonal trip probability. Thus, we tried different values in the minimum number of trips in a TAZ to minimize the sampling error and determined 20 as a threshold for model validation purposes.

Root mean square error (RMSE) is an appropriate measure of model prediction quality between two continuous variables (in this case, the proportion of intrazonal trips in the survey vs. a model). RMSE is a frequently used measure of the differences between values predicted by a model and the values actually observed. RMSE is a measure of accuracy, to compare forecasting errors of different models for a particular dataset. The smaller the RMSE, the more accurate the model (and the better the predictive power).

3.5   RESULTS

3.5.1 Intrazonal Trip Share Models

Tables 3.5 to 3.9 show the results of multilevel binomial logistic regressions for intrazonal trips by trip purpose. The intercept in the tables is the constant of the models, which is the expected mean value of log-odds of Y (intrazonal trip share) when all independent variables are zero. The coefficients are log-odds of a trip being intrazonal not interzonal for a one-unit change in the specific independent variable. By exponentiating the log-odds, we can compute the odds of intrazonal trip and the probability of intrazonal trip, which is equal to (odds of intrazonal trips / (1 + odds of intrazonal trips)).

Different D variables are shown to be significant predictors of intrazonal trips for different trip purposes. All relationships are as expected. To summarize, total employment (demographic variable) is positively associated with the share of intrazonal trips for all five trip purposes. Total population (demographic variable) is positively associated with the share of intrazonal trips for home-based-shopping, home-based-other, and non-home-based-none-work purposes. Area size has a positive association with the intrazonal trip likelihood for home-based-work, home-based-shopping, home-based-other, and non-home-based-none-work trips. Activity density is only included in non-home-based-work model. Land use diversity variable, job-population balance, is positively related to the share of intrazonal trips for all home-related trip purposes but home-based-work trips.

Destination accessibility – the percentage of jobs available within 10-minute, 20-minute, or 30-minute by car or 30-minute by transit – is negatively associated with the share of intrazonal trips for all five trip purposes. This implies that the more jobs immediately outside of the given TAZ, the more likely a trip crosses the zone boundary for specific trip purposes. A measure of street network design – the percentage of four-way intersections – is positively associated with intrazonal trip likelihood only for home-based-shopping and non-home-based-work trips. Lastly, regional variables are not statistically significant in any models, and so were dropped.

Table 3.5 Home-based-work models

	coef.	std. err.	z-value	p-value	odds ratio

	coef.	std. err.	z-value	p-value	odds ratio
intercept	-4.683	0.112	-41.706	< 0.001	0.007
totemp	0.0003	0.00003	10.430	< 0.001	1.0003
area	0.009	0.003	3.111	0.002	1.010
pctemp20a	-0.007	0.002	-3.290	0.001	0.993

Sample size: level 1 – 121,200; level 2 – 19,656; level 3 – 31
Log likelihood: -13,033; AIC: 26,078; pseudo-R-squared: 0.01

Table 3.6: Home-based-shopping models

	coef.	std. err.	z-value	p-value	odds ratio
intercept	-4.426	0.121	-36.532	< 0.001	0.012
totemp	0.0003	0.00002	14.841	< 0.001	1.0003
totpop	0.0001	0.00001	3.605	< 0.001	1.0001
area	0.004	0.002	1.994	0.046	1.004
jobpop	0.754	0.104	7.276	< 0.001	2.125
intden	0.001	0.000	2.961	0.003	1.001
pct4way	0.007	0.002	4.103	< 0.001	1.007
pctemp20a	-0.005	0.002	-2.920	0.004	0.995

Sample size: level 1 – 134,454; level 2 – 20,301; level 3 – 31
Log likelihood: -27,701; AIC: 55,422; pseudo-R-squared: 0.02

Table 3.7: Home-based-other models

	coef.	std. err.	z-value	p-value	odds ratio
intercept	-2.744	0.088	-31.297	< 0.001	0.064
totemp	0.0001	0.00001	7.397	< 0.001	1.0001
totpop	0.0001	0.00001	10.689	< 0.001	1.0001
area	0.005	0.001	3.285	0.001	1.005
Jobpop	0.333	0.059	5.673	< 0.001	1.395
intden	0.0004	0.0002	2.015	0.044	1.0004
pctemp10a	-0.006	0.002	-2.716	0.007	0.994

Sample size: level 1 – 256,004; level 2 – 22,273; level 3 – 31
Log likelihood: -92,914; AIC: 185,845; pseudo-R-squared: 0.01

Table 3.8: Non-home-based-work models

	coef.	std. err.	z-value	p-value	odds ratio
intercept	-2.603	0.084	-31.053	< 0.001	0.074
totemp	0.00005	0.00002	2.672	0.008	1.00005
actden	0.003	0.001	2.564	0.010	1.003
pct4way	0.003	0.001	3.003	0.003	1.003
pctemp30a	-0.003	0.001	-2.717	0.007	0.997

Sample size: level 1 – 86,763; level 2 – 16,200; level 3 – 31
Log likelihood: -25,060; AIC: 50,136; pseudo-R-squared: 0.002

Table 3.9: Non-home-based-non-work models

	coef.	std. err.	z-value	p-value	odds ratio
intercept	-2.096	0.040	-52.431	< 0.001	0.123
totemp	0.00004	0.00001	3.848	< 0.001	1.00004
totpop	0.00001	0.00001	2.299	0.021	1.00001
area	0.004	0.001	4.137	< 0.001	1.004
pctemp10a	-0.004	0.001	-2.457	0.014	0.996
pctemp30t	-0.002	0.001	-3.196	0.001	0.998

Sample size: level 1 – 183,066; level 2 – 20,156; level 3 – 31
Log likelihood: -67,680; AIC: 135,375; pseudo-R-squared: 0.002

3.5.2 Model Validation Result

After fitting the models with the full data, we assessed the predictive power of the five intrazonal models using 10-fold cross-validation. Travel data were randomly split into ten equal-sized groups. The validation dataset, 10% of the data, was used to validate the model which was fitted using the other 90% of the data through multilevel logistic regression.

As a result of the 10-fold cross-validation, we obtained average AUCs by trip purpose. The average AUCs range from 0.671 for the non-home-based-non-work model to 0.887 for the home-based-work model (Figure 3.2). The AUC provides the predictive accuracy of the logistic models, with values from 0.5 (no predictive power) to 1.0 (perfect prediction). Following Swets (1988) and Manel et al. (2001), models with an AUC value ranging between 0.7 and 0.9 as 'useful applications' and those with values greater than 0.9 as being of 'high accuracy.' Thus, most models can be considered useful applications. The non-home-based-non-work is lower than

the threshold of 0.7, implying a need for a different, more advanced modeling approach such as generalized additive model (Hastie and Tibshirani, 1990).
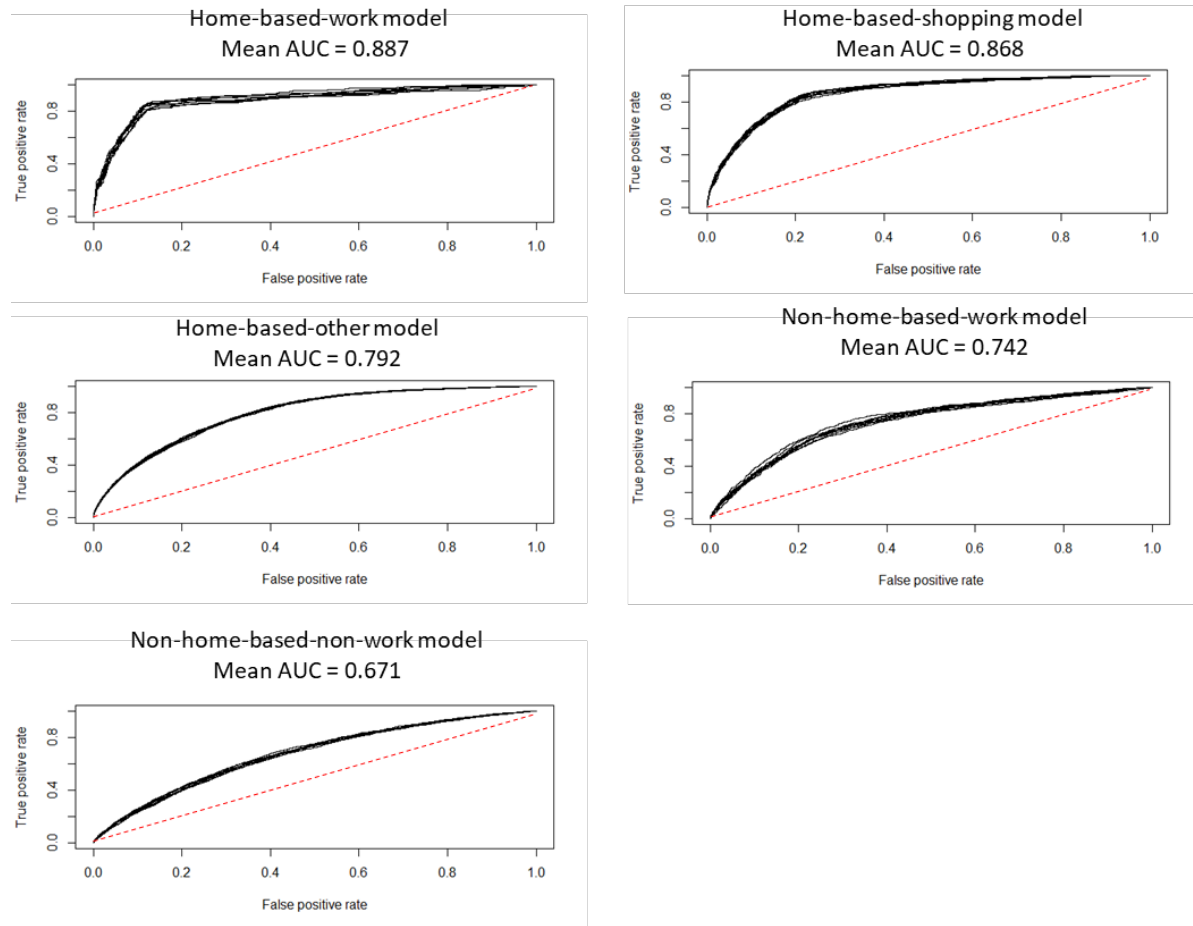
Figure 3.1: Model validation (1): Receiver operating characteristic (ROC) curves and the area under the ROC (AUC) statistics for measuring predictive power of the models

In addition to the k-fold validation, we validated our models against a conventional practice—the gravity model. We compare our model with two other models – a gravity model and a constant model (using a region-wide average proportion of intrazonal trips by trip purpose) using travel survey data from the 2012 Utah Household Travel Survey.

Table 3.10 shows that our model outperforms other models for all five trip purposes. The error rate of gravity model is significantly higher than that of our model (more than ten-fold in most models), and even higher than the constant model using an identical region-wide value of intrazonal proportion for each trip purpose.

Table 3.10: Model validation (2): Root Mean Square Error (RMSE): The smaller the RMSE, the more accurate the model and the better the predictive power.

	HBW	HBShp	HBOth	NHBW	NHBNW
WFRC/MAG Gravity model	0.076	0.101	0.199	0.055	0.112
Constant model	0.047	0.082	0.170	0.064	0.090

Our model	0.007	0.010	0.017	0.020	0.029

3.6   CONCLUSIONS

Conventional four-step models, used by virtually all metropolitan planning organizations (MPOs), state departments of transportation, and local transportation planning agencies to forecast future travel patterns, are the basis for long-range transportation planning in the United States. Trip distribution is one of the critical steps in travel demand forecasting. In the model structure, it consists of two categories – intrazonal trips and interzonal trips. As Bhatta & Larsen (2011) explained, intrazonal trips cannot be ignored, due to the impact they have on important aspects of transportation, such as congestion and pollution. For modeling intrazonal trips, there are two important components: 1) predicting whether a trip will be intrazonal and 2) determining the impedance of intrazonal trips. Little attention has been given to the former component, and in this study, we developed an approach to enhance the conventional gravity model for predicting intrazonal trips by including more built environment D variables and using a more robust modeling method.

In the first step, we surveyed 25 MPOs about how they model intrazonal travel. The finding shows the dominance of the gravity model with nearest neighbor assumptions, while a few regions are currently in the process of shifting to activity-based modeling. However, the current model involves validation errors, probably due to differences in zone size, land use, and street network patterns, none of which should be overlooked. The need to model intrazonal travel in terms of the built environment variables is evident. Thus, by using multilevel binomial logistic regression models and regional household travel survey data from 31 U.S. regions, we proved that different D variables are significant predictors of intrazonal trips for different trip purposes. Model validation results confirm that our models are useful for prediction purposes.

There is broad interest in the planning and policy communities in developing accurate tools to predict the consequences of land use and transportation strategies on travel demands. State, regional and local organizations such as state departments of transportation and metropolitan planning organizations, public health organizations, transit agencies, and city and county planning commissions are also eager to have a reliable means of evaluating growth scenarios and planning alternatives. To this end, the results of this study could be used in travel demand modeling practice, especially for the hundreds of medium- and small-sized MPOs. It is worthwhile to note that two regional MPOs, Wasatch Front Regional Council (WFRC) and Mountainland Association of Governments (MAG), are incorporating our models into their four-step models in the transportation modeling software, Cube, to improve the accuracy of travel forecasts. Because we estimated models based on 31-region database, the models have external validity, and are generalizable for future changes on land use and transport toward more compact, mixed-use, and transit-supportive developments.

The first and most obvious limitation to this study is the fact that we are proposing a novel approach to the less than novel practice of four-step travel demand modeling. As we described in the introduction, the state-of-the-art is activity-based modeling (ABM). Many of the shortcomings of the trip-based approach to travel modeling such as the inability to consider the potential sequencing of trips, are rectified by the application of ABM. However, while ABM is

the state-of-the-art in travel demand modeling, trip-based modeling is still the state of the practice for small to medium-sized MPOs. Besides, our small sample of large MPOs seems to have some bias toward those which are still using the four-step model. Although our survey indicates that some of the largest MPOs with the highest capacities are either using or developing ABMs, the majority of MPOs continue to use the four-step model. We contend that an incremental improvement to the tool that is currently the most ubiquitous among travel modelers is a valuable contribution to the practice.

4.0   MODULE FOR FORECASTING NON-MOTORIZED TRAVEL MODE CHOICES IN A TRAVEL DEMAND MODEL

4.1   INTRODUCTION

The need to account for walking and cycling in travel demand models has long been recognized, and many planning agencies have incorporated non-motorized travel into their models. In the traditional four-step model, mode choice is typically the third step in the process, following trip distribution and preceding network assignment. In the modeling process, the outputs of trip distribution are person trip tables, which are matrices of trips where the rows and columns represent trip productions and trip attractions, respectively, and aggregated typically at the traffic analysis zone (TAZ) level. In the next step, the mode choice process determines the share of trips made by each mode of transportation for each zone pair and trip purpose from the matrices.

Conventional four-step models have traditionally focused on motorized transport (Okrah, 2016). However, as the Federal Highway Administration guidebook (1999) stated, properly forecasting bicycle and pedestrian travel and advancing the state of the practice in this area have a variety of uses such as estimating the benefits of a proposed project (e.g., number of users served, reduction in automobile emissions and energy consumption), prioritizing projects based on the most significant interest to existing users or on the greatest payoff in attracting new bicyclists or walkers, planning bicycle and pedestrian networks, identifying and correcting deficiencies in existing networks, and planning for their safety.

In tour-based models (e.g., activity-based models), mode choice is usually separated into two stages, the tour level and the trip level. As the names imply, trip-level mode choice is estimated for each trip between every two stops on a tour, and it is dependent on tour mode choice. In both four-step and tour-based models, mode choice is determined using probabilities for each mode estimated from the characteristics of the trip, the modes, the traveler, and the environment in which the travel occurs (Travel Forecasting Resource; www.tfresource.org).

While over the past two decades a vast body of literature has investigated the influence of built environment on travel mode choice behavior (Zhang, 2004; Lee et al., 2014; Khan et al., 2014; Munshi, 2016), only a few agencies have considered the built characteristics in which the travel occurs. In other words, for modeling travel mode choices, metropolitan planning organizations (MPOs) often control for trip, mode, and sociodemographic characteristics, but not for built environmental variables. Moreover, not all MPOs consider nonmotorized modes of travel (e.g., walk and bike) in their mode choice modeling process. Wasatch Front Regional Council (WFRC) and Mountainland Association of Governments (MAG) are two MPOs that model nonmotorized trips, but only with consideration of trip distance, not the built environment characteristics of the TAZs in which travel occurs. As we will show in the next sections, one of the critical reasons that some MPOs do not include walk and bike modes in their models is lack of data, especially for bike trips.

The primary purpose of this research is to illustrate that the built environment, characterized by the so-called D variables (i.e., development density, land use diversity, street design, distance to transit, and destination accessibility) can play an essential role for individuals to choose

nonmotorized modes of transportation. Although the magnitude of the effect might be different from one study to another, D variables can encourage more nonmotorized travel and transit use. For instance, one would expect to see more nonmotorized trips in a neighborhood equipped with higher densities, more mixed uses, more grid-like street patterns, and better destination accessibility.

In terms of the modeling approach, mode choice (in both trip-based and tour-based models) is formulated as a discrete choice corresponding to the specific tour or trip modes. Two popular discrete choice models are multinomial logit (MNL) and nested logit (NL) or nested multinomial logit. WFRC/MAG used a nested multinomial logit mode choice model to estimate the split between non-motorized (walk/bike) and motorized (auto and transit) trips.

In this study, we use data from 29 diverse regions across the U.S., and control for the five Ds of the built environment. We present a nested logit mode choice model for three different trip purposes. Since the trip data have a nested structure (i.e., trips are nested within TAZs within regions) and, as a result, the impact of the built environment might greatly vary from one TAZ to another (or from one region to another), we use a two-level nested logit model with trips and their TAZs at level one and regions at level two. We then compare our model results with the WFRC/MAG model to examine the accuracy of our model.

The remainder of this section is organized as follows. Section 4.2 contains a review of studies on travel mode choice with the emphasis on nonmotorized modes of travel. Section 4.3 introduces the state of the practice in predicting walk and bike modes of travel. Section 4.4 describes the data and statistical methods used to estimate a new multiregional model by trip purpose. Section 4.5 presents the results of the three models, and then Section 4.6 evaluates the new models relative to the WFRC/MAG models. Finally, Section 4.7 discusses the results and presents the conclusions.

4.2 LITERATURE REVIEW

Travel mode choice is the part of the traditional four-step travel demand model that predicts the mode of travel and may include private automobile, public transportation, walking, bicycling, or other means. The central concept of travel mode choice models is to identify the relationships between travelers' mode choice and the contributing factors (Ding & Zhang, 2016; Ewing et al., 2004; Kim et al., 2007). Although MPOs have historically neglected non-motorized travel in the regional travel forecasting models due to the lack of consistent built environment data, limited records of non-motorized trips, and frankly, lack of interest (until relatively recently) in nonmotorized modes (Lie et al., 2012; Singleton & Clifton, 2013; Zhang, 2015), recent studies linking the built environment with travel behaviors have explored mode choice modelling frameworks and variables that affect whether people walk or bicycle. In this section, we will conduct a thorough literature review on walk and bike mode choices to identify the optimal modeling process and built environment attributes that have been proven to have a significant impact on such non-motorized travel mode choices.

4.2.1 Built Environment Factors Affecting Walk and Bike Travel Mode Choices

In travel studies, as explained in the previous section, the influence of the built environment has often been identified along five principal dimensions—density, diversity, design, destination accessibility, and distance to transit (Cervero & Kockelman, 1997; Ewing & Cervero, 2001; Ewing et al., 2009). While it may have limitations to cover all of the factors examined by previous studies, using the five dimensions as measures of the built environment will provide a valuable framework to navigate and encapsulate complex built environment influences on walk and bike mode shares.

A review of 16 published studies on non-motorized travel mode choices is summarized in Table 4.1. To forecast non-motorized travel mode choices, researchers use three different outcome variables—non-motorized mode in which walk and bike travel modes are integrated, walk travel mode and bike travel mode. While early studies tended to estimate the likelihood of taking non-motorized travel modes (Bento et al., 2005; Zhang, 2004; Kockelman, 1997; Cervero & Kockelman, 1997), relatively recent studies acknowledged the differences between walk mode choice and bike mode choice and thus separate them in the modeling process (Frank et al., 2008; Reilly & Landis, 2002; Hamre & Buehler, 2014; Ferrell et al., 2015; Rajamani et al., 2003; Mitra, 2011; Ozbil & Peponis, 2012; Ewing et al., 2009; Ewing et al., 2004; Kim et al., 2007; Aziz et al., 2017; Khan et al., 2014).

Studies have often included density, diversity, and design variables in walk and bike mode choice models (see Table 1). Regarding density and diversity variables, although trip purposes and effect sizes vary across studies, walk mode choice tends to have positive relationships with higher population and job densities, greater retail floor area ratios, more diverse land use characteristics, and shorter distances to closest commercial uses (Hamre & Buehler, 2014; Reilly & Landis, 2002; Frank et al., 2008; Ferrell et al., 2015; Rajamani et al., 2003; Ozbil & Peponis, 2012). Similarly, bike mode choice is positively related to higher population densities and greater mix of land uses (Ferrell et al., 2015; Hamre & Buehler, 2014), but higher job and population densities have occasionally been found to result in less biking (Khan et al., 2014).

While sharing density and diversity variables, walk and bike mode choice models are differentiated by specific design factors. Mainly, walk mode choice models were apt to encompass detailed pedestrian environment factors, in which better street connectivity, greater block density, smaller fraction of cul-de-sac streets, more sidewalk coverage, and wider sidewalks increase the likelihood of walking (Ozbil & Peponis, 2012; Ewing et al., 2004; Ewing et al., 2009; Aziz et al., 2017). Bike models contain variables about bicycling facilities and bike lanes, showing more of those bike-friendly environments are motivating factors (Hamre & Buehler, 2014; Aziz et al., 2017).

A few studies include variables related to destination accessibility and distance to transit. For example, better job accessibility and shorter walking time (or walking distance) are associated with a greater chance of walking (Kockelman, 1997; Ewing et al., 2004; Ewing et al., 2009; Kim et al., 2007), while shorter time to destination (e.g., school) is related to a greater chance of biking (Ewing et al., 2004).

Table 4.1: Comparison of built environment factors for non-motorized, walk, and bike mode choices

Authors	Methods	Significant factors				
		Density	**Diversity**	**Design**	**Destination access**	**Distance to transit**
Non-motorized mode choice						
Bento et al. (2005)	MNL	Population density (-)	Job-housing balance (-)	-	-	Supply of rail transit (+)
Zhang (2004)	MNL, NL	Population density (+), Job density (+)	Entropy of land use balance (+)	Street connectivity (+)	-	-
Kockelman (1997)	LGR	-	Land use mix (+)	-	Job accessibility by walking (+)	-
Cervero & Kockelman (1997)	LGR	-	-	Sidewalk width (+), Proportion front and side parking (+)	-	-
Walk mode choice						
Hamre & Buehler (2014)	MNL	Population Density (+)	-	-	-	-
Reilly & Landis (2002)	MNL	Population Density (+)	Distance to closest commercial use (-)	-	-	-
Frank et al. (2008)	NL	Retail floor area ratio (+)	Land use mix (+)	Intersection Density (+)	-	-
Ferrell et al. (2015)	MNL	Population density (+)	-	4-way intersection density (+)	-	-
Rajamani et al. (2003)	MNL	-	Land use mix (+)	% Cul-de-sac street (-)	-	-
Mitra (2011)	BNL	-	Jobs-to-population ratio (-)	block density (+)	-	-
Ozbil & Peponis (2012)	LNR	-	Mixed-use entropy (+)	Street connectivity (+)	-	-
Ewing et al. (2004)	MNL	-	-	Average sidewalk coverage (+)	Walk time to school (-)	-
Ewing et al. (2009)	MNL	-	-	Intersection density (+), Sidewalk Coverage (+)	Jobs within one mile (+)	-
Aziz et al. (2017)	MNL	-	-	Sidewalk width (+)	-	-
Khan et al. (2014)	MNL	-	-	3-way intersection density (+), 4-way	-	-

Authors	Methods	Significant factors				
		Density	**Diversity**	**Design**	**Destination access**	**Distance to transit**
				intersection density (+)		
Kim et al. (2007)	MNL	-	-	Park and ride lot at the station (-)	-	Distance between home and station (-)

Bike mode choice

Authors	Methods	Density	Diversity	Design	Destination access	Distance to transit
Ferrell et al. (2015)	MNL	Population density (+)	Mixed use (+)	4-way intersection density (+)	-	-
Hamre & Buehler (2014)	MNL	Population density (+)	Urban core (+)	Bikeway supply (+)	-	-
Khan et al. (2014)	MNL	Population + Job density (-)	-	4-way intersection density (+)		
Aziz et al. (2017)	MXL	-	-	Bike land length (+), Fraction open space (+)	-	-
Ewing et al. (2004)	MNL	-	-	-	Bike time to school (-)	-

Notes:
We use the following abbreviations:
Method: MNL = Multinomial logit regression
                  NL = Nested logit regression
                  BNL = Binomial regression
                  LGR = Logistic regression
                  LNR = Linear regression
Factors: (+) = positive relationship
                  (-) = negative relationship

4.3   STATE OF THE PRACTICE IN WALK AND BIKE MODELING

It has been less than three decades that non-motorized travel mode choice has been included in the regional travel demand models and, so far, different modeling frameworks and a wide range of factors have been examined to improve the models (Singleton & Clifton, 2013). Like the two previous chapters, we conducted a survey of current walk/bike mode choice modeling practices at 25 randomly selected MPOs in mid-2018 to understand the gap between academic research and practical implementation. Summary findings from our survey are presented in Table 4.2. Although we surveyed MPOs with different population sizes in their regions, we focused most heavily on large regions since, generally, their MPOs are leaders in using new travel modeling techniques. In this section, we will also discuss the results of Singleton and Clifton's study in 2013, in which they did a comprehensive review of the non-motorized travel mode choice modeling of the 48 largest MPOs across the U.S.

The results of our survey show that:

- One MPO (Des Moines Area) does not conduct mode choice modeling.

- The StanCOG model uses an adjustment procedure rather than a full mode choice analysis step. Since the percentage of transit trips is small in Stanislaus County, currently the StanCOG travel model does not include a separate mode choice analysis step.

- In addition to these two MPOs, 13 MPOs have mode choice models to predict the use of motorized modes (auto and transit) but do not model non-motorized mode choices. So, in total, 15 MPOs do not predict the share of walk/bike mode trips. It should be noted that North Jersey Transportation Planning Authority (NJTPA) and Chicago Metropolitan Agency for Planning (CMAP) (among these 15 MPOs) split non-motorized and motorized trips after trip generation and before trip distribution. Yet, they do not model non-motorized trips in the mode choice step.

- Multinomial logit and nested logit are the two dominant model forms in the mode choice step. Although multinomial logit is the most popular one in the U.S. (Fresno report, 2014), our results show that among MPOs surveyed that predict non-motorized trips, five of them utilize nested logit and four of them use a multinomial logit model. The Lincoln MPO is the only MPO that simply uses a distance-based algorithm to determine non-motorized mode share.

- Out of these nine MPOs, seven of them only use travel distance or travel time to predict the probability of walk/bike mode choice. One (Association of Monterey Bay Area Governments) uses trip time and total employment density as predictor variables, and one (Memphis Urban Area MPO) uses household income and population density as predictor variables

Table 4.2: Non-Motorized Mode Choice in Travel Demand Modeling of MPOs

MPO Name	Major City	Population (2010)	Walk and Bike Mode Choices
Brunswick MPO	Brunswick	79,626	Does not model non-motorized travel.
Roanoke Valley MPO	Roanoke	227,507	Does not model non-motorized travel.
Lincoln MPO	Lincoln (Nebraska)	285,407	The Lincoln MPO Travel Model uses a distance-based algorithm to determine the non-motorized mode share. Local information is only available for commute trips, which are defined similar to but not exactly the same as the home-based work trips (HBW) in the Lincoln MPO Travel Model. For the remaining trip purposes, data was borrowed from another region. After reviewing available data sources, including NHTS data, San Luis Obispo, CA was selected as the source model for non-motorized trip shares.
North Front Range MPO	Fort Collins	433,178	The NFR Model uses a mode choice structure that nests multiple multinomial choices. Non-motorized trips are divided to walk and bike and the probabilities of these trips are calculated based on the walk and bike time.
Chattanooga-Hamilton County/North Georgia Transportation Planning Organization	Chattanooga	436,669	**ABM**[*]: The tour main mode sub-model is structured as a multinomial logit with the following eight mode options: Drive-to-Transit, Walk-to-Transit, School Bus, Shared Ride (3 or more persons), Shared Ride (2 persons), Drive Alone, Bicycle, Walk. Roundtrip road distance is the only variable used to determine walk or bike trips.
Augusta Regional Transportation Study	Augusta	440,134	Does not model non-motorized travel. In the ARTS model, the mode choice component includes "motorized person trips" and splits these into auto and transit trips.

Des Moines Area MPO	Urbandale	475,855	The Des Moines Area MPO does not conduct mode choice modeling.
Stanislaus COG	Modesto	514,453	Does not model non-motorized travel. The StanCOG model uses an adjustment procedure rather than a full mode choice analysis step.
Community Planning Association of Southwest Idaho	Meridian	550,359	The COMPASS mode choice model uses a nested logit structure with five alternatives. The non-motorized nest includes walk and bicycle modes and their probabilities are estimated based on trip distance.
Association of Monterey Bay Area Governments	Marina	732,667	The updated mode choice model for the AMBAG RTDM utilizes a nested logit based model structure. The estimated models are a series of logit models (multinomial or nested) that vary by trip purpose and by peak/off-peak periods. For most purposes, the following travel modes are estimated: Auto drive alone, Auto shared ride (carpool), Transit, Walk, and Bike. The probabilities of walk and bike trips are predicted based on trip time and total employment density.
Capital District Transportation Committee	Albany	823,239	Does not model non-motorized travel. For other modes, multinomial logit.
Fresno Council of Governments	Fresno	930,885	The Fresno County mode choice models use a multinomial logit formulation. The Fresno COG Model includes a mode choice step which divides trips into drive alone, shared ride 2 people, shared ride 3+ people, local bus, regional bus, BRT, walk and bike.
Memphis Urban Area MPO	Memphis	1,077,697	Nested Logit model. For some trip purposes, there are no bike trips. So, this mode is excluded. The variables used for predicting the probability of non-motorized trips are households income and population density.
Wasatch Front Regional Council + MAG	Salt Lake City	1,561,348	A nested multinomial logit mode choice model is used to estimate the split among non-motorized (walk/bike) and motorized (auto and transit) trips. Trip distance is the only predictor of the non-motorized share.
METROPLAN Orlando	Orlando	1,837,385	Does not model non-motorized travel. For the rest, nested logit form.
Mid-America Regional Council	Kansas City	1,895,535	Does not model non-motorized travel. For the rest, nested logit model.
Ohio-Kentucky-Indiana Regional Council of Governments	Cincinnati	1,981,230	**ABM.** The mode choice model does include non-motorized choices and is a multinomial logit model.
East-West Gateway Council of Government	St. Louis	2,571,253	Does not model non-motorized travel. For the rest, nested logit model.
Boston Region MPO	Boston	3,159,512	multinomial logit form. No bike mode. Walk time is the only predictor for walk probability.
Southeast Michigan COG - First Gravity, Then Destination Choice Model	Detroit	4,703,593	Based on current version, they do not model non-motorized travel. For other modes, nested logit. But for their enhanced model (**ABM.** will be finished in this year), they are going to use non-motorized, divided by walk and bike as well.
National Capital Region Transportation Planning Board	Washington	5,068,540	Does not model non-motorized travel.

Houston-Galveston Area Council	Houston	5,892,002	Does not model non-motorized travel. For the rest, nested logit model.
North Central Texas COG	Arlington	6,417,630	Does not model non-motorized travel. For the rest, Nested logit models are used for HBW and HNW trips, and a multinomial logit model is used for NHB trips.
North Jersey Transportation Planning Authority	Newark	6,579,801	Using binomial logit model to split non-motorized and motorized trips after trip generation and before trip distribution. But, does not model non-motorized travel at mode choice. For other modes, nested logit.
Chicago Metropolitan Agency for Planning	Chicago	8,444,660	Splitting non-motorized and motorized trips after trip generation and before trip distribution. But, for mode choice model, they did not model non-motorized travel. For other modes, multinomial logit model.

* ABM: Activity-Based Modeling.

Our findings are almost consistent with Singleton and Clifton's (2013) study. They found that among the 48 largest MPOs in the U.S., 18 of them (38%) do not model non-motorized travel. For the rest, they divided the modeling framework into six categories: Two MPOs (4%) use a cross-classification model to perform separate non-motorized and motorized trip generation processes; five MPOs (10%) use a percentage, linear regression, or binary logit model to split non-motorized and motorized trips after trip generation and before trip distribution (same as NJTPA and CMAP); five MPOs (10%) use a binary logit model to split non-motorized and motorized trips after trip distribution and before mode choice; four MPOs (8%) use a multinomial or nested logit mode choice model with only non-motorized modes combined; four MPOs (8%) use a multinomial logit mode choice model with walk and bicycle modes but not within a non-motorized nest; and, finally, 10 MPOs (20%) use a nested logit mode choice model that considers walk and bicycle modes within a non-motorized nest. It is worth mentioning that only four MPOs were using activity-based modeling (tour-based modeling).

4.4   DATA AND METHODS

4.4.1 Regional Household Travel Survey

In this study, we have consistent datasets for 29 regions. The resulting pooled dataset consists of 810,030 trips by 86,400 households. Table 4.3 shows the share of each mode by trip purpose by region. The three trip purposes in this study are home-based work (HBW), home-based other (HBO), and non-home-based (NHB). Like the WFRC model, the home-based shop (HBShp) trips are merged with HBO trips, and NHB consists of both non-home-based work and non-home-based non-work trips.

According to Table 4.3, the mode shares vary from region to region. For instance, in a region like Boston, MA, 30% of trips are generated by non-motorized modes for the NHB trip purpose, while this value for the Wasatch Front Regional Council is only 6%. Note that, on average, the share of non-motorized modes for HBO is higher than other trip purposes, ranging from 5% in San Antonio, TX, to 25% in Seattle, WA.

Table 4.3: Travel Mode Shares (%) by Trip Purpose from Travel Surveys

Region	HBW		HBO		NHB	
	non-motorized	motorized	non-motorized	motorized	non-motorized	motorized
Atlanta, GA	204 (1.47%)	13603 (98.52%)	2764 (6.33%)	40853 (93.66%)	2306 (7.51%)	28393 (92.48%)
Boston, MA	862 (8.39%)	9407 (91.6%)	9055 (22.63%)	30951 (77.36%)	9770 (30.01%)	22775 (69.98%)
Burlington, NC	19 (3.44%)	532 (96.55%)	284 (10.2%)	2500 (89.79%)	104 (6.29%)	1549 (93.7%)
Dallas, TX	38 (1%)	3729 (98.99%)	1471 (10.11%)	13068 (89.88%)	352 (4.26%)	7894 (95.73%)
Denver, CO	504 (6.35%)	7427 (93.64%)	3374 (12.53%)	23541 (87.46%)	3411 (17.99%)	15541 (82%)
Detroit, MI	100 (4.4%)	2169 (95.59%)	913 (13.15%)	6028 (86.84%)	268 (5.96%)	4224 (94.03%)
Eugene, OR	346 (14.63%)	2018 (85.36%)	1301 (16.87%)	6408 (83.12%)	1058 (17.29%)	5058 (82.7%)
Greensboro, NC	23 (1.34%)	1685 (98.65%)	882 (9.35%)	8546 (90.64%)	289 (5.18%)	5284 (94.81%)
Hampton Roads–Norfolk, VA	21 (1.2%)	1725 (98.79%)	1171 (12.09%)	8512 (87.9%)	285 (5.78%)	4642 (94.21%)
Houston, TX	51 (0.58%)	8726 (99.41%)	1633 (5.25%)	29415 (94.74%)	433 (2.63%)	16011 (97.36%)
Indianapolis, IN	162 (2.84%)	5526 (97.15%)	1482 (8.37%)	16209 (91.62%)	1081 (8.41%)	11763 (91.58%)
Kansas City, MO	59 (1.22%)	4747 (98.77%)	919 (5.7%)	15179 (94.29%)	257 (2.7%)	9260 (97.29%)
Madison, WI	12 (7.94%)	139 (92.05%)	114 (15.76%)	609 (84.23%)	51 (12.2%)	367 (87.79%)
Miami, FL	23 (1.94%)	1161 (98.05%)	1073 (15.58%)	5811 (84.41%)	270 (8.5%)	2906 (91.49%)
Minneapolis–St. Paul, MN-WI	510 (4.94%)	9800 (95.05%)	3153 (9.25%)	30921 (90.74%)	1776 (8.49%)	19129 (91.5%)
Orlando, FL	5 (0.58%)	853 (99.41%)	625 (14.65%)	3639 (85.34%)	158 (7.52%)	1941 (92.47%)
Palm Beach, FL	14 (2.29%)	595 (97.7%)	705 (15.45%)	3856 (84.54%)	114 (5.94%)	1805 (94.05%)
Phoenix, AZ	57 (1.64%)	3417 (98.35%)	2768 (13.58%)	17606 (86.41%)	528 (5.76%)	8632 (94.23%)
Portland, OR	677 (11.06%)	5444 (88.93%)	4209 (19.82%)	17018 (80.17%)	4676 (25.94%)	13348 (74.05%)
Provo, UT	121 (4.65%)	2479 (95.34%)	1569 (13.85%)	9757 (86.14%)	293 (6.04%)	4554 (93.95%)
Rochester, NY	171 (3.43%)	4804 (96.56%)	754 (6.88%)	10195 (93.11%)	291 (4.23%)	6582 (95.76%)
Salem, OR	169 (6.68%)	2360 (93.31%)	990 (12.3%)	7057 (87.69%)	685 (12.97%)	4595 (87.02%)
Salt Lake City, UT	299 (4.38%)	6513 (95.61%)	2769 (11.2%)	21948 (88.79%)	732 (6.17%)	11122 (93.82%)
San Antonio, TX	28 (1.27%)	2169 (98.72%)	417 (5.01%)	7891 (94.98%)	65 (2.34%)	2712 (97.65%)

Seattle, WA	950 (14.82%)	5456 (85.17%)	5085 (25.18%)	15103 (74.81%)	2835 (23.82%)	9066 (76.17%)
Springfield, MA	78 (6.01%)	1218 (93.98%)	709 (16.44%)	3603 (83.55%)	444 (16.02%)	2327 (83.97%)
Syracuse, NY	15 (2.56%)	570 (97.43%)	381 (11.81%)	2843 (88.18%)	85 (6.49%)	1224 (93.5%)
Tampa, FL	33 (2.16%)	1490 (97.83%)	1507 (14.07%)	9202 (85.92%)	250 (4.96%)	4787 (95.03%)
Winston-Salem, NC	23 (1.91%)	1177 (98.08%)	545 (7.92%)	6330 (92.07%)	208 (5.56%)	3527 (94.43%)
Average	**4.32%**	**95.68%**	**12.47%**	**87.53%**	**9.55%**	**90.45%**

4.4.2 Built Environment Data

Like the previous two chapters, we control for the 5D variables at the TAZ level. The variables are based on these data:

- Parcel-level land use data with detailed land use classifications; from these, we compute detailed measures of land use mix.
- A GIS layer for street networks and intersections; from these, we compute intersection density and percentage of 4-way intersections.
- A GIS layer for transit stops; from these data, we compute transit stop densities.
- Population and employment at the block or block group-level; from these, we compute activity density.
- A GIS layer for TAZs with socioeconomic information (population and employment). Travel times for auto and transit travel from TAZ to TAZ (so-called travel time skims); from these, and TAZ employment data, we compute regional employment accessibility measures for auto and transit.

4.4.3 Variables

The dependent and independent variables used in this study are defined in Table 4.4. Sample sizes and descriptive statistics are also provided. The variables in this study cover most of the Ds, from density to demographics. WFRC and MAG have asked us not to include variables they do not predict or cannot predict for future years. Hence, we have not included land use entropy and transit stop density variables. However, the street network variables can be assumed constant in built up areas, and probably can be predicted by looking at neighboring zones or scenario plans. Additionally, activity density, job-population balance, and accessibility measures can be estimated from future population and employment data. Note that we have also dropped the household income variable since we used it in the previous chapter to predict vehicle ownership. All in all, a total of 16 independent variables are available to predict mode choices and all of them are consistently defined from region to region.

Table 4.4: Variables used to estimate mode choice model

Variable	Description	N	Mean	S.D.
Outcome Variable				

mode	Mode choice (1= walk, 2= bike, 3= transit, 4= car)	810,030	-	-
trip purpose	Trip purpose: home-based work (HBW), home-based other (HBO), non-home-based (NHB)	-	-	-
Choice-Specific Variables				
time_w	travel time by walk	799,216	99.56	228.96
time_b	travel time by bike	799,216	27.88	64.11
time_t	travel time by transit	799,216	24.89	57.24
time_c	travel time by car	799,216	11.62	26.71
Socio-Demographic and Built Environment Variables				
hhsize	household size	810,030	3.06	1.51
employed	number of employed persons in household	810,030	1.44	0.91
veh	number of vehicles owned by households	810,030	2.12	1.05
lnactden	natural log of activity density within TAZ (pop + emp per square mile in 1000s)	810,030	1.25	1.26
jobpop[a]	job-population balance within TAZ	810,030	0.57	0.27
intden	intersection density within TAZ	810,030	103.35	76.57
pct4way	percentage of 4-way intersections within TAZ	809,555	24.55	19.25
pctemp10a	percentage of regional employment within 10 minutes by auto	809,253	7.50	11.57
pctemp20a	percentage of regional employment within 20 minutes by auto	809,867	29.44	26.28
pctemp30a	percentage of regional employment within 30 minutes by auto	809,886	51.31	30.12
pctemp30t	percentage of regional employment within 30 minutes by transit	809,886	19.25	23.60
Regional Variable				
region	a set of regional dummy variables	29	-	-

[a] job-population balance = $1 - [ABS(employment - 0.2 * population)/(employment + 0.2 * population)]$; ABS = absolute value of expression in parentheses. The value 0.2, representing a balance of employment and population, was found through trial and error to maximize the explanatory power of the variable.

4.4.4 Analysis Method

Our data and model structure are hierarchical, with trips "nested" within TAZs within regions. The best statistical approach for nested data is multilevel modeling (MLM), also called hierarchical modeling (HLM). MLM accounts for spatial dependence among observations. Regression and other single-level statistical methods may produce biased coefficients and biased standard errors and misleading predictions. In some settings, MLM overcomes these limitations, accounting for the dependence among observations and producing more accurate coefficient and standard error estimators (Raudenbush and Bryk, 2002).

The essence of MLM is to partition the variation of an outcome variable associated with each data level. However, a choice model is not based on partitioning variance; it is about predicting probabilities of discrete outcomes. Still, as we showed in Table 4.3, travel modes (especially the share of non-motorized trips) are very different from one region to another. As such, the best approach would seem to be some sort of 3- or 4-level model (since trips made by individuals are nested within households). The model of choice is a nested logit model. NLOGIT is the only

available statistical package that can estimate a nested logit model (which will be explained later in this subsection) with a 3-level structure. However, the massive size of our dataset exceeded the internal limit for 3-level models in NLOGIT. So, in this study, we are considering a 2-level model. We are using fixed-effect models to control for the impacts on choice at the region level.

On the other hand, our outcome variable is categorical, and we can partition the choice set. One of the best statistical approaches for this kind of outcome variable is the nested logit (NL) model, which was first proposed by McFadden in 1978. It is a generalization of the multinomial logit model that is based on the idea that groups of similar alternatives may be joined in several groups, partitions or nests. Essentially the NL structure allows the choice set to be partitioned in such a way that the constant variance assumption holds among alternatives in the same partition while allowing differential variance between partitions and correlation among alternatives within a partition (Hensher et al., 2015). For example, the four alternatives in our case study are partitioned in such a way that walk and bike are separated from transit and personal vehicle (auto) to reflect the presence of unobserved attributes that have more in common with respect to their utility influence within nonmotorized modes and within motorized modes than between nonmotorized and motorized modes.

Hence, it can be said that while the multinomial logit model treats all alternatives equally, the NL model includes intermediate branches grouping a subset of alternatives, like a tree structure. Figure 1 shows the nesting structure of our outcome variable. Unlike MNL, in the NL model the probability of a mode of travel like walking is equal to the joint probability of choosing the nonmotorized mode (i.e., Prob(*nonmotorized)) and choosing the walk option (Prob(walk/ nonmotorized)):*
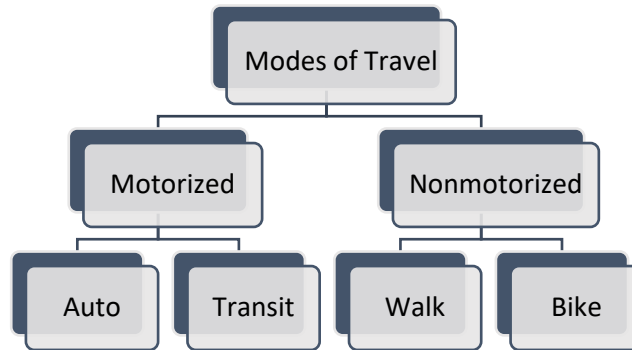Prob(walk,nonmotorized) = Prob(nonmotorized) × Prob(walk |nonmotorized)

Figure 4.1: Nesting structure of the dependent variable

If we abbreviate "nonmotorized" and "motorized" to *nmt* and *mt*, respectively, the conditional choice probability for each alternative among the nested alternatives (conditioned on the choice of the nest at the higher level) will be:

$$P_{walk/nonmotorized} = \frac{\exp\left(\frac{V_{walk}}{\theta_{nmt}}\right)}{\exp\left(\frac{V_{walk}}{\theta_{nmt}}\right) + \exp\left(\frac{V_{bike}}{\theta_{nmt}}\right)}$$ **(Equation 4-1)**

$$P_{bike/nonmotorized} = \frac{\exp\left(\frac{V_{bike}}{\theta_{nmt}}\right)}{\exp\left(\frac{V_{walk}}{\theta_{nmt}}\right) + \exp\left(\frac{V_{bike}}{\theta_{nmt}}\right)}$$

(Equation 4-2)

$$P_{transit/motorized} = \frac{\exp\left(\frac{V_{transit}}{\theta_{mt}}\right)}{\exp\left(\frac{V_{transit}}{\theta_{mt}}\right) + \exp\left(\frac{V_{auto}}{\theta_{mt}}\right)}$$

(Equation 4-3)

$$P_{auto/motorized} = \frac{\exp\left(\frac{V_{auto}}{\theta_{mt}}\right)}{\exp\left(\frac{V_{auto}}{\theta_{mt}}\right) + \exp\left(\frac{V_{transit}}{\theta_{mt}}\right)}$$

(Equation 4-4)

Where *P* is the choice probability, *V* corresponds to the specific characteristics of the alternative, and θ_{branch} is the log sum parameter (also called the dissimilarity parameter, inclusive parameter, or the nesting coefficient) for the specific branch (motorized or nonmotorized). Basically, it is a function of the underlying correlation between the unobserved components for pairs of alternatives in that nest, and it characterizes the degree of substitutability between those alternatives (Koppelman & Bhat, 2006). The functions V_{mode} are the deterministic parts of the random utility models:

$$V_{mode} = \alpha_{region} + \sum_{k=variables} \beta_k x_k$$

(Equation 4-5)

The marginal choice probabilities for each of the nests would be:

$$P_{nmt} = \frac{\exp(V_{nmt} + \theta_{nmt}\Gamma_{nmt})}{\exp(V_{mt} + \theta_{mt}\Gamma_{mt}) + \exp(V_{nmt} + \theta_{nmt}\Gamma_{nmt})}$$

(Equation 4-6)

$$P_{mt} = \frac{\exp(V_{mt} + \theta_{mt}\Gamma_{mt})}{\exp(V_{mt} + \theta_{mt}\Gamma_{mt}) + \exp(V_{nmt} + \theta_{nmt}\Gamma_{nmt})}$$

(Equation 4-7)

Where Γ measures the expected maximum utility among the nested alternatives and is given by the log sum of the exponents of the nested utilities:

$$\Gamma_{mt} = \ln[\exp(V_{walk}/\theta_{mt}) + \exp(V_{bike}/\theta_{mt})]$$

(Equation 4-8)

$$\Gamma_{nmt} = \ln[\exp(V_{auto}/\theta_{nmt}) + \exp(V_{transit}/\theta_{nmt})]$$

(Equation 4-9)

The terms V_{branch} accommodate additional effects that would be expected to impact the branch choice but not the mode choice given the branch. We have no such effects in our model, so $V_{motorized} = V_{nonmotorized} = 0$. The inclusive value parameters, θ_{nmt} and θ_{mt} reflect the dissimilarity of the scaling of the random components of the two branches. The model reverts to the simpler multinomial logit model if $\theta_{nmt} = \theta_{mt} = 1$.

4.5 RESULTS

The maximum likelihood estimates of the nested logit models for different trip purposes are shown in Tables 4.5-7 (auto is the reference category). The final models were chosen based on two considerations: 1- whether the sign of a coefficient is expected or not (for example, it is expected to see higher vehicle trips (compared to non-motorized modes and transit) as the

number of vehicles owned by the household increases. If not, we drop that variable), 2- statistical significance of the explanatory variable (we dropped the variables that were not significant for at least two trip purposes). It should be noted that some of the regions did not have all of the modes of travel. For instance, in the San Antonio (TX) travel survey there is no record of bike trips for NHB trip purposes, and in Burlington (NC) there is no record of transit trips for HBW trip purposes. Since we are using fixed region effects, we needed to drop these regions from our analysis. All in all, our model for HBW trips consists of 20 regions, while the two other models have information for 28 regions.

According to Tables 4.5-7, our choice-specific variable (i.e., travel time) is statistically significant for all of the trip purposes and has a negative sign. That is, the higher the travel time of a mode, the lower the related utility. In other words, the probability of the corresponding choice would decrease. For instance, if the travel time for *walk* increases, the probability of walk will fall and the probabilities of the other three modes will increase.

Number of vehicles owned by households (*veh*) is the other important significant variable across all three models. The results show that if the vehicle ownership of a household increases, the probability of using the personal vehicle over other modes will increase. *Veh* has the lowest coefficient for transit mode, meaning that the increase in vehicle ownership lowers the probability of choosing transit over other modes.

According to Table 4.5, if the household size increases, travelers tend to use bike and transit over their car (walk mode is not statistically significant). We have already controlled for the number of vehicles. Since households have a limited number of cars and their workplaces might be at different locations, some of the members need to choose other modes of travel, namely bike and transit for HBW trip purpose and walk, bike, and transit for NHB trip purpose. For HBO trip purpose the probability of walking decreases, while the probability of transit increases as households get larger (bike mode is not significant). The negative sign for walking makes sense since for these trip purposes, carpooling and traveling together is more common. As a reminder, HBO trip purpose consists of both home-based other and home-based shopping trips.

Table 4.5: Results of the fixed effect nested logit model for HBW trips

Variable	Estimate	Std. Error	Z Value
walk:(intercept)	-0.71305	0.1378	-5.1743***[1]
bike:(intercept)	-4.12209	0.3561	-11.5741***
transit:(intercept)	-4.96735	0.2736	-18.1533***
Time	-0.02084	0.0008	-25.1233***
walk:hhsize	0.01614	0.0157	1.0306
bike:hhsize	0.14998	0.0157	9.5724***
transit:hhsize	0.24468	0.0193	12.7036***
walk:veh	-0.33655	0.0237	-14.2263***
bike:veh	-0.21299	0.0270	-7.8911***
transit:veh	-1.26329	0.0314	-40.2503***
walk:lnactden	0.29165	0.0225	12.9439***

bike:lnactden	-0.14553	0.0285	-5.1085***
transit:lnactden	0.17849	0.0338	5.2853***
walk:pct4way	0.00164	0.0008	2.0074*
bike:pct4way	0.00853	0.0009	9.1691***
transit:pct4way	0.00710	0.0011	6.5861***
walk:pctemp30a	-0.00346	0.0014	-2.4979*
bike:pctemp30a	0.00755	0.0022	3.4628***
transit:pctemp30a	0.01696	0.0023	7.5079***
walk:pctemp30t	0.00260	0.0018	1.4628
bike:pctemp30t	0.00980	0.0022	4.4542***
transit:pctemp30t	0.00696	0.0025	2.7927**
walk:SLC Region	0.06839	0.1094	0.6251
bike:SLC Region	2.08125	0.2930	7.1029***
transit:SLC Region	2.54220	0.2134	11.9136***
walk:Provo-Orem Region	-0.10655	0.1438	-0.7409
bike:Provo-Orem Region	2.05539	0.3283	6.2609***
transit:Provo-Orem Region	2.28747	0.2673	8.5578***
iv:motor	0.47541	0.1204	3.9481***
iv:nonmotor	2.22330	0.0981	22.6641***

Number of regions: 20
Log-Likelihood: -15989
McFadden R^2: 0.33183

[1] Significance codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1

In terms of D variables, we see that a couple of them are statistically significant and have the expected signs. In the HBW model, as activity density, percentage of 4-way intersections, and percentage of jobs accessible within 30 minutes by auto and transit increase, the probabilities of walking, biking, and using transit increase. The two exceptions are bike mode for activity density and walk mode for the percentage of jobs reachable within 30 minutes by auto, which have negative signs. For the first exception, this might be due to the fact that TAZs with higher activity density are usually more crowded and probably not safe and efficient (in terms of travel duration) for bike users. For the second exception, the possible explanation for the lower probability of walk trips compared to personal vehicle trips is that individuals tend to walk (and have access to jobs) mostly for longer-distance trips.

Table 4.6: Results of the fixed effect nested logit model for HBO trips

Variable	Estimate	Std. Error	Z Value
walk:(intercept)	0.47034	0.0359	13.0992***[1]
bike:(intercept)	-2.86572	0.1067	-26.8683***
transit:(intercept)	-1.94207	0.1152	-16.8555***
time	-0.09814	0.0003	-314.883***
walk:hhsize	-0.04072	0.0038	-10.703***
bike:hhsize	-0.00680	0.0096	-0.7076
transit:hhsize	0.04588	0.0124	3.7032***

walk:veh	-0.31391	0.0053	-59.2256***
bike:veh	-0.16005	0.0132	-12.1132***
transit:veh	-0.96448	0.0157	-61.2389***
walk:pct4way	0.00462	0.0003	17.8001***
bike:pct4way	0.00627	0.0006	9.9308***
transit:pct4way	0.00420	0.0007	5.7418***
walk:pctemp30t	0.00630	0.0003	19.1271***
bike:pctemp30t	0.00702	0.0009	8.006***
transit:pctemp30t	0.00688	0.0013	5.2614***
walk:SLC Region	0.46231	0.0351	13.1893***
bike:SLC Region	0.91995	0.1012	9.0919***
transit:SLC Region	0.29379	0.1194	2.4599*
walk:Provo-Orem Region	0.42209	0.0387	10.9028***
bike:Provo-Orem Region	0.50353	0.1191	4.2276***
transit:Provo-Orem Region	-1.23882	0.2192	-5.6528***
iv:motor	2.72154	0.0445	61.1692***
iv:nonmotor	1.58639	0.0120	132.2214***

Number of regions: 28
Log-Likelihood: -121240
McFadden R^2: 0.34605

1 Significance codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1

Table 4.7: Results of the fixed effect nested logit model for NHB trips

Variable	Estimate	Std. Error	Z Value
walk:(intercept)	-2.87930	0.1081	-26.6485***[1]
bike:(intercept)	-3.24170	0.1063	-30.5106***
transit:(intercept)	-0.24649	0.0148	-16.6038***
time	-0.01123	0.0002	-63.6916***
walk:hhsize	0.02022	0.0064	3.1527**
bike:hhsize	0.11703	0.0075	15.5387***
transit:hhsize	0.00213	0.0010	2.1724*
walk:veh	-0.06758	0.0097	-6.9433***
bike:veh	-1.08760	0.0104	-104.2272***
transit:veh	-0.02334	0.0016	-14.6607***
walk:lnactden	0.09354	0.0118	7.9266***
bike:lnactden	0.27945	0.0122	22.9042***
transit:lnactden	0.00807	0.0015	5.4579***
walk:pct4way	0.00159	0.0004	3.7203***
bike:pct4way	0.00068	0.0006	1.2053
transit:pct4way	0.00028	0.0001	3.8149***
walk:pctemp10a	0.01691	0.0016	10.7216***
bike:pctemp10a	0.00304	0.0016	1.9547.
transit:pctemp10a	-0.00129	0.0002	-5.8653***

walk:pctemp30t	-0.00401	0.0006	-6.3851***
bike:pctemp30t	0.01752	0.0008	23.3014***
transit:pctemp30t	0.00170	0.0001	17.4064***
walk:SLC Region	1.02140	0.1066	9.5853***
bike:SLC Region	0.41049	0.1225	3.3505***
transit:SLC Region	-0.05457	0.0187	-2.9198**
walk:Provo-Orem Region	1.23870	0.1188	10.4232***
bike:Provo-Orem Region	-0.76462	0.2214	-3.4538***
transit:Provo-Orem Region	-0.12951	0.0279	-4.6469***
iv:motor	-0.35659	0.0387	-9.2133***
iv:nonmotor	9.02280	0.1368	65.9438***

Number of regions: 28
Log-Likelihood: -104630
McFadden R^2: 0.3757

1 Significance codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1

In both the HBO and NHB models, percentage of 4-way intersections (*pct4way*) and percentage of jobs accessible within 30 minutes by transit (*pctemp30t*) are significant and have the expected signs. An increase in any of these variables will result in a higher probability for walk, bike, and transit modes. The only exception is walk mode for the *pctemp30t* variable in the NHB model where the negative coefficient seems counterintuitive and requires further investigation.

In the NHB model, activity density is also significant and negatively impact the use of auto mode over other modes. The fourth important D variable, which is positive and significant for walk and bike (with respect to personal vehicle), is job accessibility by auto (within 10 minutes). These are short-distance trips wherein walk and bike modes are more competitive and transit mode is less competitive relative to the auto mode.

All in all, compared to the WFRC walk/bike mode choice model which only controls for the trip distance, in this study we were able to control for most of the critical sociodemographic and built environment variables. Our results confirm that in all models, some D variables will reduce the share of vehicle trips and will encourage travelers to use non-motorized modes of travel, as well as transit.

We also ran the likelihood ratio (LR) test to compare our model results with fixed effect multinomial logit models where alternatives (here, travel modes) are not grouped. As it was shown in the previous sections, MNL is one of the popular approaches among MPOs in modeling travel mode choices. The LR test is 2*[Log-Likelihood(NL) – Log-Likelihood(MNL)] which is a chi-squared with 2 degrees of freedom with 5.99 as the critical value (Hensher et al., 2015). Below is the result of the LR tests.

HBW: 2 * [15989 (NL) - 16331 (MNL)] = 684
HBO: 2 * [121240 (NL) - 123340 (MNL)] = 4200
NHB: 2 * [104630 (NL) - 113090 (MNL)] = 16920

The NL models greatly outperform the MNL models in terms of this important metric. All of these values are many times the critical value. In addition, we tested the overall significance of

the regional effects which again would be a likelihood ratio test. The test statistic is 2*[logL(regional) - logL(no regional)] where logL(no regional) represents a log-likelihood of a model without regional dummies. It is chi-squared with "3 × number of regions" degrees of freedom (DF). The number of regions should be multiplied by 3 since we have three sets of regional dummies in the models. As it is shown in Table 4.8, all of the values are way above the critical points, suggesting that multilevel modeling approach with region effects performs much better.

Table 4.8: Test result of the overall significance of the regional effects

Model	logL(regional)	logL(no regional)	DF	Test Result	Critical Value (p=0.05)
HBW	-15989	- 17283	60	2588	43.18
HBO	-121240	-129010	84	15540	63.87
NHB	-104630	-114500	84	19740	63.87

4.6   MODEL VALIDATION

Our approach is theoretically more robust in the sense that it incorporates influential built environment characteristics of TAZ and uses disaggregate data at the individual trip level from various U.S. regions. To be used in practical modeling, however, we need to validate our three models in comparison with the WFRC models. In other words, how much more accurate are our models than the WFRC models? Since WFRC models only report total number of motorized and non-motorized trips at the TAZ level, we first compute the average share of each four modes of travel by TAZ based on our models' predictions, and then summed the bike and walk shares (to obtain total nonmotorized share) and transit and automobile shares (to obtain total motorized share).

The modeled values are compared against the actual proportion of motorized and nonmotorized shares by trip purpose by TAZ from the 2012 Utah Household Travel Survey. Like the intrazonal modeling report, the problem with this approach is that many TAZs have no or only a few trips from the Household Travel Survey. This raises sampling error issues, meaning that the small number of trips in the survey cannot represent all trips occurring in that TAZ. For example, if a TAZ has only one trip (which is walk) from the survey, it gets 100% nonmotorized trip probability. If it has only one trip (which is auto), it gets 0% nonmotorized trip probability. Thus, we tried different values in the minimum number of trips in a TAZ to minimize the sampling error and set 20 as a final threshold for model validation purposes.

Root mean square error (RMSE) and the correlation between the predicted values and actual values are appropriate measures of model prediction quality between two continuous variables (in this case, the proportion of nonmotorized and motorized trips in the survey vs. the models). RMSE is a frequently used measure of the differences between values predicted by a model and the values actually observed. The smaller the RMSE, the more accurate the model (and the better the predictive power).

Tables 4.9 and 4.10 show that our model outperforms other models for all three trip purposes. In terms of RMSE, our models have lower values compared to the WFRC models. Based on the threshold that we have chosen, in the WFRC models, surprisingly, the predicted values for some

trip purposes (both for motorized and nonmotorized) have a negative correlation with the actual observed values from the survey which suggest weak estimation ability. Note that the error rates of WFRC models for motorized mode share are way higher than our models (at least four times as high), suggesting their models to be much less accurate.

Table 4.9: Model performance comparison: RMSE and Correlation of nonmotorized mode share

	HBW-nonmotorized		HBO-nonmotorized		NHB-nonmotorized	
	our model	WFRC	our model	WFRC	our model	WFRC
RMSE	0.0693	0.0985	0.1113	0.1478	0.0752	0.0976
Correlation	0.4353	-0.1082	0.5283	-0.0031	0.6156	0.0123

Table 4.10: Model performance comparison: RMSE and Correlation of motorized mode share

	HBW-motorized		HBO-motorized		NHB-motorized	
	our model	WFRC	our model	WFRC	our model	WFRC
RMSE	0.0693	0.3215	0.1113	0.4016	0.0752	0.3428
Correlation	0.4353	0.0566	0.5283	0.0414	0.6156	-0.0059

4.7   CONCLUSION

This study estimated mode choice models by trip purpose using regional household travel data and built environmental variables from 29 diverse regions across the United States. The models are estimated with nested logit specification. The results show that walk, bike, and transit probabilities decrease (with respect to the personal automobile as the reference category) as the vehicle ownership of a household increases. Household size is another sociodemographic variable that we controlled for and, based on the trip purpose, non-motorized modes can have both higher and lower probabilities relative to the auto mode. Travel time as our choice-specific variable has a negative sign in all three models, which implies that the higher the travel time of a mode, the lower the relative utility and the lower the probability that a given traveler will choose that particular mode.

Our study results confirm the vital role that the built environment plays in shaping people's mode choice behavior. Overall, almost all of the Ds that we could control for in this report are found to be significant and negatively correlated with car use for one trip purpose or another, after controlling for sociodemographic variables. From urban planning and transportation public policy perspectives, this study suggests that sustainable modes of travel become more favorable as the built environment becomes dense (in the HBW and NHB models); mixed (in the NHB model); well-connected (in all three models); transit-served (in the HBO and NHB models); and job accessible (in all three models). These findings have important implications for policy and planning practice, where decision-makers seek solutions to promote sustainable transportation and deal with vehicle miles traveled, congestion, greenhouse gas emissions, obesity, and other health and environmental concerns.

Such a large dataset also gives the models external validity missing from earlier studies. The model developed in this study can be directly used for travel demand modeling and forecasting, not only by WFRC but also by MPOs in other regions of the U.S. Based on the results of this study, we would recommend using multilevel nested logit models over multinomial logit models. NL is a choice method specifically designed to recognize the possibility of different variances across the modes of travel and some correlation among sub-sets of modes, and in this study we showed that NL models perform better than MNL models.

This study has some limitations as well. Although the standard D variables are covered, this study still omits certain variables that have presumptive effects on people's mode choice behavior. Parking supplies and prices, travel attitudes, and residential self-selection may strongly affect the choice of travel. The second limitation of this study is about the modeling approach. We were unable to control for a possible random effect by TAZs in each region mainly due to the complexity of the model. Also, the recent literature suggests that analyzing the relationships between travel mode choice and the built environment can be improved through the use of vehicle ownership as a mediating variable rather than an exogenous variable (e.g., see Van Acker and Witlox, 2010; Ding et al., 2017). Hence more advanced methodologies such as multilevel integrated structural equation models and discrete choice models may be required in the future.

5.0   REFERENCES

1. Anas, A.: The combined equilibrium of travel networks and residential location markets. Regional Science and Urban Economics, 15(1), 1-21 (1985)
2. Anowar, S., Eluru, N., & Miranda-Moreno, L. F. (2014). Alternative modeling approaches used for examining automobile ownership: a comprehensive review. Transport Reviews, 34(4), 441-473.
3. Aziz, H.M., Nagle, N., Morton, A., Hilliard, M., White, D., Stewart, R. (2017). Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: A random parameter model using New York City commuter data, Transportation, pp. 1-23.
4. Batty, M.: Urban Modeling: Algorithms, Calibrations, Predictions. Cambridge University Press, London (1976)
5. Beimborn, E, Kennedy, R, & Schaefer, W. (1996). Inside the blackbox: making transportation models work for livable communities. Retrieved from https://www4.uwm.edu/cuts/blackbox/blackbox.pdf
6. Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand. Cambridge, MA: MIT Press.
7. Bento, A. M., Cropper, M. L., Mobarak, A. M., & Vinha, K. (2005). The effects of urban spatial structure on travel demand in the United States. Review of Economics and Statistics, 87(3), 466-478.
8. Bhat, C. R., & Guo, J. Y. (2007). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. Transportation Research Part B: Methodological, 41(5), 506-526.
9. Bhat, C. R., & Pulugurta, V. (1998). A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. Transportation Research Part B: Methodological, 32(1), 61-75.
10. Bhat, C., Paleti, R., Pendyala, R., Lorenzini, K., & Konduri, K. (2013). Accommodating immigration status and self-selection effects in a joint model of household auto ownership and residential location choice. Transportation Research Record: Journal of the Transportation Research Board, (2382), 142-150.
11. Bhatta, B. P., & Larsen, O. I.: Are intrazonal trips ignorable? Transport policy, 18(1), 13-22 (2011)
12. Borra, S., & Di Ciaccio, A. (2010). Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. Computational statistics & data analysis, 54(12), 2976-2989.
13. Broadstock, D. C., Collins, A., & Hunt, L. C. (2010). Modelling car trip generations for UK residential developments using data from TRICS. Transportation planning and technology, 33(8), 671-678.
14. Cao, X., Mokhtarian, P. L., & Handy, S. L. (2007). Cross-sectional and quasi-panel explorations of the connection between the built environment and auto ownership. Environment and Planning A, 39(4), 830-847.

15. Capital District Transportation Committee: The CDTC Step Model, Appendix E: Documentation of CDTC Step Model and Planning Assumptions (2010)

16. Castiglione, J., Bradley, M. and Gliebe, J. (2015), Activity-Based Travel Demand Models: A Primer, SHRP 2 Report S2-C46-RR-1. Transportation Research Board.

17. Cervero, R. (2006). Alternative approaches to modeling the travel-demand impacts of smart growth. Journal of the American Planning Association, 72(3), 285-295.

18. Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity, and design. Transportation Research Part D: Transport and Environment, 2(3), 199-219.

19. Cervero, R.: Alternative approaches to modeling the travel-demand impacts of smart growth. Journal of the American Planning Association, 72(3), 285-295 (2006)

20. Chang, J. S. (2006). Models of the relationship between transport and land-use: A review. Transport Reviews, 26(3), 325-350.

21. Chatman, D. G. (2013). Does TOD need the T? On the importance of factors other than rail access. Journal of the American Planning Association, 79(1), 17-31.

22. Chu, Y. L. (2002). Automobile ownership analysis using ordered probit models. Transportation Research Record: Journal of the Transportation Research Board, (1805), 60-67.

23. Cirillo, C., & Liu, Y. (2013). Vehicle Ownership Modeling Framework for the State of Maryland: Analysis and Trends from 2001 and 2009 NHTS Data.Journal of Urban Planning and Development, 139(1), 1-11.

24. Cirillo, C., & Xu, R. (2011). Dynamic discrete choice models for transportation. Transport Reviews, 31(4), 473-494.

25. CMAP & CATS.: Go to 2040: Travel Model Documentation for Chicago Metropolitan Agency for Planning (2014)

26. Community Planning Association of Southwest Idaho: Regional Travel Demand Forecast Model Calibration and Validation Report for Ada and Canyon County in Idaho, Report Number 06-2017.

27. Dargay, J., Gately, D. (1997), Vehicle ownership to 2015: Implications for energy use and emissions, Energy Policy, 25 (14-15), 1121-1127.

28. Dargay, J., Gately, D., & Sommer, M. (2007). Vehicle ownership and income growth, worldwide: 1960-2030. The Energy Journal, 143-170.

29. Davidson, W., Donnelly, R., Vovsha, P., Freedman, J., Ruegg, S., Hicks, J., ... & Picado, R. (2007). Synthesis of first practices and operational research approaches in activity-based travel demand modeling. Transportation Research Part A: Policy and Practice, 41(5), 464-488.

30. De Jong, G. C., & Kitamura, R. (2009). A review of household dynamic vehicle ownership models: holdings models versus transactions models. Transportation, 36(6), 733-743.

31. Ding, C., Wang, D., Liu, C., Zhang, Y, & Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance, Transportation Research Part A: Policy and Practice, Volume 100, Pages 65-80, https://doi.org/10.1016/j.tra.2017.04.008.

32. Ding, L., Zhang, N. (2016). A Travel Mode Choice Model Using Individual Grouping Based on Cluster Analysis, Procedia Engineering, Volume 137, Pages 786-795, https://doi.org/10.1016/j.proeng.2016.01.317.

33. Domencich, T. A., & McFadden, D. (1975). Urban Travel Demand-A Behavioral Analysis Amsterdam: North Holland Publishing Company.

34. Dowling, R. G.: Predicting air quality effects of traffic-flow improvements: final report and user's guide (Vol. 535). Transportation Research Board (2005)

35. Ewing, R., & Cervero, R. (2001). Travel and the built environment: a synthesis. Transportation Research Record: Journal of the Transportation Research Board, (1780), 87-114.

36. Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. Journal of the American planning association, 76(3), 265-294.

37. Ewing, R., & Cervero, R. (2017). Does compact development make people drive less? The answer is "yes." Journal of the American Planning Association, 83(1), 19-25.

38. Ewing, R., & Tilbury, K. (2002). Sketch models for integrated transportation and land use planning. Gainesville. FL: Metropolitan Transportation Planning Organization.

39. Ewing, R., Greenwald, M. J., Zhang, M., Walters, J., Feldman, M.,Cervero, R., … Thomas, J. (2009). Measuring the impact of urban form and transit access on mixed use site trip generation rates—Portland pilot study. Washington, DC: U.S. Environmental Protection Agency.

40. Ewing, R., Schroeer, W., & Greene, W. (2004). School Location and Student Travel Analysis of Factors Affecting Mode Choice. Transportation Research Record, 1895(1), 55–63. https://doi.org/10.3141/1895-08

41. Ewing, R., Tian, G., Goates, J. P., Zhang, M., Greenwald, M. J., Joyce, A., ... Greene, W. (2015). Varying influences of the built environment on household travel in 15 diverse regions of the United States. Urban Studies, 52(13), 2330–2348.

42. Ewing, R., Tilbury, K.: Sketch models for integrated transportation and land use planning. Gainesville, FL.: Metropolitan Transportation Planning Organization (2002)

43. Ferdous, N., Vana, L., Bowman, J., Pendyala, R., Giaimo, G., Bhat, C., ... & Anderson, R. (2012). Comparison of Four-Step Versus Tour-Based Models for Prediction of Travel Behavior Before and After Transportation System Changes. Transportation Research Record: Journal of the Transportation Research Board, (2303), 46-60.

44. Ferrell, C. E., Mathur, S., & Appleyard, B. S. (2015). Neighborhood Crime and Transit Station Access Mode Choice-Phase III of Neighborhood Crime and Travel Behavior.

45. Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation, 24(1), 38-49.

46. Frank L, Bradley M et al. (2008) Urban form, travel time, and cost relationships with tour complexity and mode choice. Transportation 35(1):37–54

47. Fresno Council of Governments. Fehr & Peers Transportation Consultant: Travel Demand Model: Model Description & Validation Report for Fresno Council of Governments (2014).

48. Giuffrè, O., Grana, A., Giuffrè, T., & Marino, R. (2013), Accounting for Dispersion and Correlation in Estimating Safety Performance Functions. An Overview Starting from a Case Study, Modern Applied Science, 7(2); 11-23.

49. Greenwald, M. J.: The relationship between land use and intrazonal trip making behaviors: Evidence and implications. Transportation Research Part D: Transport and Environment, 11(6), 432-446 (2006)

50. Greiner, M., Pfeiffer, D., & Smith, R. D. (2000). Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Preventive veterinary medicine, 45(1-2), 23-41.

51. Guo, Z. (2013). Residential street parking and car ownership: a study of households with off-street parking in the New York City region. Journal of the American Planning Association, 79(1), 32-48.

52. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). Multivariate data analysis (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: Prentice hall.

53. Hamilton, B. W.: Wasteful commuting again. Journal of political economy, 97(6), 1497-1504 (1989)

54. Hamre, Andrea & Buehler, Ralph. 2014. Commuter Mode Choice and Free Car Parking, Public Transportation Benefits, Showers/Lockers, and Bike Parking at Work: Evidence from the Washington, DC Region. Journal of Public Transportation, 17 (2): 67-91. DOI: http://doi.org/10.5038/2375-0901.17.2.4

55. Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), 29-36.

56. Hastie, T. J. & Tibshirani, R. J. (1990). Generalized Additive Models. Chapman & Hall/CRC.

57. Hensher, D., Rose, J. and Greene, W., (2015). Applied Choice Analysis, 2nd ed., Cambridge University Press, Cambridge.

58. Hess, B. D., & Ong, P. (2002). Traditional neighborhoods and automobile ownership. Transportation Research Record: Journal of the Transportation Research Board, (1805), 35-44.

59. H-GAC.: 2012 Model Validation and Documentation Report, Houston-Galveston Area Council Regional Travel Models, Appendix 4 (2014)

60. Hilbe, J. M. (2011). Negative binomial regression. Cambridge University Press.

61. Horner, M. W., & Murray, A. T.: Excess commuting and the modifiable areal unit problem. Urban Studies, 39(1), 131-139 (2002)

62. Hüls, A., Frömke, C., Ickstadt, K., Hille, K., Hering, J., Von Münchhausen, C., Hartmann, M., & Kreienbrock, L. (2017), Antibiotic Resistances in Livestock: A Comparative Approach to Identify an Appropriate Regression Model for Count Data, Frontiers in Veterinary Science, Vol. 4, Article 71.

63. Khan, M., Kockelman, K.M., Xiong, X., (2014). Models for anticipating non-motorized travel choices, and the role of the built environment. Transport 35, 117–126. http://dx.doi.org/10.1016/j.tranpol.2014.05.008.

64. Kim, B.Y., Fleming, G.G., Lee, J.J., Waitz, I.A., Clarke, J.-.P., Balasubramanian, S., Malwitz, A., Klima, K., Locke, M., Holsclaw, C.A., Maurice, L.Q., Gupta, M.L., (2007). System for assessing Aviation's Global Emissions (SAGE). Part 1: Model description and inventory results. Transportation Research – D 12, 325–346.

65. Kim, H. S., & Kim, E. (2004). Effects of public transit on automobile ownership and use in households of the USA. Review of Urban & Regional Development Studies, 16(3), 245-262.

66. Kim, N. S., & Susilo, Y. O. (2013). Comparison of pedestrian trip generation models. Journal of Advanced Transportation, 47(4), 399-412.

67. Kimley-Horn and Associates, Inc., Cambridge Systematics, Inc. and HNTB: Direction 2040: Long Range Transportation Plan: Travel Demand Model Documentation for Memphis Urban Area, Appendix G (2014)

68. Kimley-Horn and Associates, Inc.: Chattanooga Travel Demand Model Development Highway and Transit Assignment, Technical Memorandum #6, Final Version 3.0 (2013)

69. Kitamura, R. (2009). A dynamic model system of household car ownership, trip generation, and modal split: model development and simulation experiment. Transportation, 36(6), 711-732.

70. Kitamura, R., Akiyama, T., Yamamoto, T., & Golob, T. (2001). Accessibility in a metropolis: Toward a better understanding of land use and travel. Transportation Research Record: Journal of the Transportation Research Board, (1780), 64-75.

71. Kockelman, K. M. (1997). Travel behavior as a function of accessibility, land use mixing and land use balance: Evidence from the San Francisco Bay Area. Transportation Research Record, 1607, 116–125.

72. Koppelman, F S, Bhat, C. (2006). Self-instructing course in mode choice modeling: multinomial and nested logit models, Federal Transit Administration, US Department of Transportation, http://www.civil.northwestern.edu/people/koppelman/PDFs/LM_Draft_060131Final-060630.pdf

73. Kordi, M., Kaiser, C. & Fotheringham, A. S.: A possible solution for the centroid-to-centroid and intra-zonal trip length problems, In International Conference on Geographic Information Science, Avignon (2012)

74. Lamb, G. M.: Introduction to transportation planning: 4-trip distribution and model split. Traffic Engineering and Control April, 628-630 (1970)

75. Lee, J.S., Nam, J., Lee, S.S., (2014). Built environment impacts on individual mode choice: an empirical study of the Houston-Galveston metropolitan area. Int. J. Sustain. Transport. 8 (6), 447–470. http://dx.doi.org/10.1080/15568318.2012.716142.

76. Liu, F., Evans, J. E., & Rossi, T. (2012). Recent Practices in Regional Modeling of Nonmotorized Travel. Transportation Research Record: Journal of the Transportation Research Board, (2303), 1–8. https://doi.org/10.3141/2303-01

77. LSA Associates, Inc.: North Front Range Regional Travel Model: Model Process, Parameters, and Assumptions (2007)

78. LSA Associates, Inc.: Travel Demand Model: Model Development and Validation Report for Lincoln Metropolitan Planning Organization (2011)

79. Manel, S., Ceri Williams, H., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38, 921-931.

80. Martin, W. A., & McGuckin, N. A.: Travel estimation techniques for urban planning (Vol. 365). Washington, DC: National Academy Press (1998)

81. McNally, M. (2000). The activity-based approach (chap. 4), In Handbook of transport modeling, edited by D. Hensher and K. Button. Amsterdam: Pergamon.

82. McNally, M. (2008). The four step model, Center for Activity Systems Analysis. Retrieved from http://escholarship.org/uc/item/0r75311t

83. McNally, M. G. (2007). The four-step model. In Handbook of Transport Modelling: 2nd Edition (pp. 35-53). Emerald Group Publishing Limited.

84. Meng, Q. (2014). Modeling and prediction of natural gas fracking pad landscapes in the Marcellus Shale region, USA. Landscape and Urban Planning, 121, 109-116.

85. METROPLAN Orlando: 2040 Long Range Transportation Plan, Technical Report 8: Model Development and Application Guidelines, Final Adopted Plan (2016)

86. Mitra, R. (2012). School Travel Mode Choice Behaviour in Toronto, Canada (Doctoral dissertation).

87. Moeckel, R., & Donnelly, R. (2015). Gradual rasterization: redefining spatial resolution in transport modelling. Environment and Planning B: Planning and Design, 42(5), 888-903.

88. Munshi, T., 2016. Built environment and mode choice relationship for commute travel in the city of Rajkot, India. Transp. Res. Part D 44, 239–253. http://dx.doi.org/10.1016/j.trd.2015.12.005.

89. Mwakalonge, J. L., & Badoe, D. A. (2014). Trip generation modeling using data collected in single and repeated cross-sectional surveys. Journal of Advanced Transportation, 48(4), 318-331.

90. National Capital Region Transportation Planning Board, Metropolitan Washington Council of Governments: Calibration Report for the TPB Travel Forecasting Model, Version 2.3, on the 3,722-Zone Area System (2012)

91. Okrah M.B. (2016) Handling Non-Motorized Trips in Travel Demand Models. In: Wulfhorst G., Klug S. (eds) Sustainable Mobility in Metropolitan Regions. Studien zur Mobilitäts- und Verkehrsforschung. Springer VS, Wiesbaden

92. Okrah, M. B. (2016). Handling Non-Motorized Trips in Travel Demand Models. In Sustainable Mobility in Metropolitan Regions (pp. 155-171). Springer VS, Wiesbaden.

93. Ozbil, A. and Peponis, J. (2012). The Effects of Urban Form on Walking to Transit, in Proceedings of Eighth International Space Syntax Symposium, Greene, M. and Reyes, J. and Castro. A. (Eds), Santiago de Chile: PUC, 2012.

94. PBS&J.: ARTS 2035 Long Range Transportation Plan, Appendix B: The Travel Demand Model for the ARTS MPO (2010)

95. PBS&J.: The Travel Demand Model for the Brunswick MPO (2010)

96. Pinjari, A. R., & Bhat, C. R. (2011). Activity-based travel demand analysis. A Handbook of Transport Economics, 10, 213-248.

97. Pinjari, A. R., Pendyala, R. M., Bhat, C. R., & Waddell, P. A. (2011). Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. Transportation, 38(6), 933-958.

98. Plaza, C. V., & da Silva, A. N. R.: Intrazonal Trip Distances: An Estimation Approach and Application to a Case Study. In Transportation Research Board 94th Annual Meeting, No. 15-0383 (2015)

99. Pont, K., Wadley, D., Ziviani, J., & Khan, A. (2013). The influence of urban form and family decision making on children's travel to school. Journal of Urban Design, 18(3), 363-382.

100. Potoglou, D., & Susilo, Y. (2008). Comparison of vehicle-ownership models.Transportation Research Record: Journal of the Transportation Research Board, (2076), 97-105.

101. Rajamani, J., Bhat, C. R., Handy, S., Knaap, G. & Song, Y. (2003) Assessing the Impact of Urban Form Measures in Nonwork Trip Mode Choice after Controlling for Demographic and Level-of-Service Effects. Paper presented at the Transportation Research Board Annual Meeting, Washington, DC, January 2003.

102. Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods, 2nd Ed. Sage Publications: Thousand Oaks, CA.

103. Raudenbush, S. W., Bryk, A. S., & Cheong, Y. F., et al. (2010). HLM 7: Hierarchical Linear and Nonlinear Modeling. Scientific Software International, Chicago, IL.

104. Reilly, M., Landis, J., (2002). The Influence of Built-Form and Land Use on Mode Choice: Evidence From the 1996 Bay Area Travel Survey. University of California Transportation Center.

105. Ribeiro Jr, E.E., Zeviani, W.M., Bonat, W.H, Demetrio, C.G.B. & Hinde, J. (2018), Reparametrization of COM-Poisson Regression Models with Application in the Analysis of Experimental Data, arXiv:1801.09795

106. Rouwendal, J., & Nijkamp, P. (2004). Living in Two Worlds: A Review of Home-to-Work Decisions. Growth and change, 35(3), 287-303.

107. RSG, Inc. (2017), WFRC/MAG Demand Model Calibration & Validation Report, Version 8.1.

108. RVAMPO: Constrained Long-Range Transportation Plan 2035 (2011), Available at: http://rvarc.org/chapters-4-and-5-technical-process-clrtp-2035/

109. Ryan, J., & Han, G. (1999). Vehicle-ownership model using family structure and accessibility application to Honolulu, Hawaii. Transportation Research Record: Journal of the Transportation Research Board, (1676), 1-10.

110. Schimek, P. (1996). Household motor vehicle ownership and use: How much does residential density matter?. Transportation Research Record: Journal of the Transportation Research Board, (1552), 120-125.

111. Schipper, L. (2011). Automobile use, fuel economy and CO 2 emissions in industrialized countries: encouraging trends through 2008?. Transport Policy, 18(2), 358-372.

112. Schwartz, W. L., C. D. Porter, G. C. Payne, J. H. Suhrbier, P. C. Moe, and W. L. Wilkinson (1999). Guidebook on Methods to Estimate Non-Motorized Travel: Overview of Methods. Publication No. FHWA-RD-98-165. FHWA, U.S. Department of Transportation, 1999, www.fhwa.dot.gov/tfhrc/safety/pubs/vol1/title.htm, Accessed August 29, 2019.

113. Shay, E., & Khattak, A. (2005). Automobile ownership and use in neotraditional and conventional neighborhoods. Transportation Research Record: Journal of the Transportation Research Board, (1902), 18-25.

114. Shay, E., & Khattak, A. (2007). Automobiles, trips, and neighborhood type: comparing environmental measures. Transportation Research Record: Journal of the Transportation Research Board.

115. Singleton, P.A., Clifton, K.J. (2013). Pedestrians in Regional Travel Demand Forecasting Models: State-of-the-Practice (Paper 13-4857). Presented at the 92nd Annual Meeting of the Transportation Research Board, Washington, DC, January 13–17.

116. StanCOG: The StanCOG Transportation Model Program: Documentation and Validation Report, Version 1 (2010). Avilable at: http://www.stancog.org/trans-model.shtm

117. Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285-1293.

118. The Association of Monterey Bay Area Governments: Association of Monterey Bay Area Governments Regional Travel Demand Model Technical Report (2015)

119. The Des Moines Area MPO: Travel Demand Model: Documentation & User Guide (2006)

120. United States Environmental Protection Agency (EPA), (2015). Overview of Greenhouse Gases. Retrieved from http://www3.epa.gov/climatechange/ghgemissions/gases/co2.html

121. Van Acker, V., & Witlox, F. (2011). Commuting trips within tours: how is commuting related to land use? Transportation, 38(3), 465-486.

122. Van Acker, V., Witlox, F., 2010. Car ownership as mediating variable in car travel behavior research using a structural equation modelling approach to identify its dual relationship. J. Transp. Geogr. 18 (1), 65–74. http://dx.doi.org/10.1016/j.jtrangeo.2009.05.006.

123. Venigalla, M. M., Chatterjee, A., & Bronzini, M. S.: A specialized equilibrium assignment algorithm for air quality modeling. Transportation Research Part D: Transport and Environment, 4(1), 29-44 (1999)

124. Walters, G., Ewing, R., & Schroeer, W. (2000). Adjusting Computer Modeling Tools to Capture Effects of Smart Growth: Or" Poking at the Project Like a Lab Rat". Transportation Research Record: Journal of the Transportation Research Board, (1722), 17-26.

125. Zegras, C. (2010). The built environment and motor vehicle ownership and use: Evidence from Santiago de Chile. Urban Studies, 47(8), 1793-1817.

126. Zhang, M. (2004). The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong, Journal of the American Planning Association, 70:3, 344-360, DOI: 10.1080/01944360408976383

127. Zhang, Y. (2015). Microsimulating Active Transportation Mode Choice using Smartphone-based Travel Survey and Transportation Tomorrow Survey Data. Retrieved from https://tspace.library.utoronto.ca/handle/1807/71456

128. Zhou, Z., Chen, A., & Wong, S. C. (2009). Alternative formulations of a combined trip generation, trip distribution, modal split, and trip assignment model. European Journal of Operational Research, 198(1), 129-138.

129. Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry, 39(4), 561-577.