RESEARCH

UDOT Keeping Utah Moving

Utah Department of Transportation - Research Division
4501 South 2700 West - P.O. Box 148410 - SLC,UT 84114-8410

Report No. UT-19.14

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

Prepared For:

Utah Department of Transportation
Research & Innovation Division

Submitted By:

Department of City and Metropolitan Planning
University of Utah

Authored By:

Reid Ewing
Keunhyun Park
Sadegh Sabouri
Torrey Lyons
Guang Tian

DISCLAIMER

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the U.S. Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore.

ACKNOWLEDGMENTS

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals on the Technical Advisory Committee for helping to guide the research:

- Khaisy Vonarath, UDOT Planning
- Walt Steinvorth, UDOT Planning
- Kevin Nichol, UDOT Research and Innovation
- Jon Larsen, Salt Lake City
- Tim Hereth, Mountainland Association of Governments
- Chad Worthen, WFRC
- Ted Knowlton, WFRC
- Shawn Seager, MAG

TECHNICAL REPORT ABSTRACT

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UT- 19.14	N/A	N/A

4. Title and Subtitle Key Enhancements to the WFRC/MAG Conventional Four-Step Travel Demand Model	5. Report Date June 2019 6. Performing Organization Code
7. Author(s) Reid Ewing. Keunhyun Park, Sadegh Sabouri, Torrey Lyons, Guang Tian	8. Performing Organization Report No.
9. Performing Organization Name and Address Department of City & Metropolitan Planning College of Architecture + Planning, 220 AAC University of Utah 375 S 1530 E Salt Lake City, Utah 84112	10. Work Unit No. 5H07921H 11. Contract or Grant No. 16-8752
12. Sponsoring Agency Name and Address Utah Department of Transportation 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410	13. Type of Report & Period Covered Final Feb 2017-May 2019 14. Sponsoring Agency Code PIC No. UT15.601

15. Supplementary Notes

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

16. Abstract

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) and an intrazonal travel model (internal capture model) that consider all of the D variables based on household travel surveys and built environmental data for 32 and 31 regions, respectively, validates the models, and demonstrates that the models have far better predictive accuracy than WFRC/MAG's current models.

17. Key Words	18. Distribution Statement	23. Registrant's Seal
travel model enhancements vehicle ownership mode intrazonal travel model	Not restricted. Available through: UDOT Research Division 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410 www.udot.utah.gov/go/research	N/A

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	59	N/A

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

UNIT CONVERSION FACTORS

Preferably, present all measurements in the report in inch-pound or U.S. Customary system units. For non-conforming units, give data conversion units in parentheses throughout the report, or include applicable unit conversions here.

(Example) Units used in this report and not conforming to the UDOT standard unit of measurement (U.S. Customary system) are given below with their U.S. Customary equivalents:

- 1 meter (m) = 3.28 feet (ft)
- 1 kilometer (km) = 0.62 mile (mi.)
- Etc.

(Alternatively, the following conversion factors table may be included. Enlarge to fit.)

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. (Adapted from FHWA report template, Revised March 2003)

LIST OF ACRONYMS

FHWA	Federal Highway Administration
UDOT	Utah Department of Transportation

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) and an intrazonal travel model (internal capture model) that consider all of the D variables based on household travel surveys and built environmental data for 32 and 31 regions, respectively, validates the models, and demonstrates that the models have far better predictive accuracy than WFRC/MAG's 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.

1.0 INTRODUCTION

1.1 Problem Statement

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 metropolitan planning organizations (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.

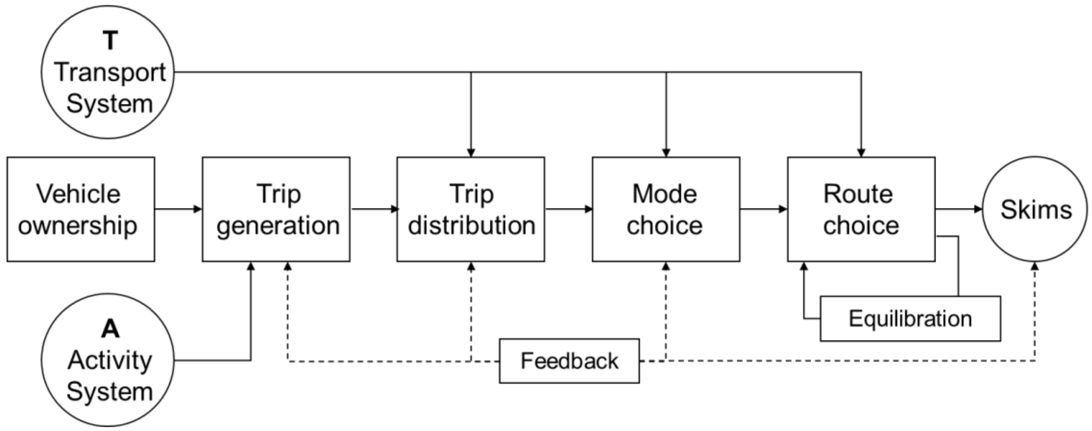

Figure 1.1. Four Step Travel Demand Model (Adapted from McNally, 2007)

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 asked: 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."

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. 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 precludes the modeling of intrazonal travel in terms of the local built environment.

1.2 Objectives

With this study, we seek to develop and implement car shedding and intrazonal travel 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 for the vehicle ownership model and our 31-region household travel database for the intrazonal travel model. For vehicle ownership, we have precise XY geocodes for all 86,710 households in the database. For intrazonal travel, we have precise XY geocodes for trip ends for all 843,287 trips in the database. We lost one region from the latter database for lack of XY geocodes for all trip ends. This is the largest household travel database of its sort ever assembled. This database has been linked to built environmental data for TAZs around geocoded households and trip ends. These models will pre-process inputs to the four-step process. 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.

1.3 Scope

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.

The specific outcomes of the project are equations that predict each of the outcomes listed above (vehicle ownership, intrazonal trip choice by trip purpose) in terms of D variables of the TAZs themselves and their surrounding environments. The equations are 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, & 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 & A. Joyce (2015). Traffic Generated by Mixed-Use Developments—13-Region Study Using Consistent Built Environment Measures. Transportation Research Record. (2500), 116–124.

1.4 Outline of Report

Chapter 2 is on vehicle ownership modeling. Chapter 3 is on intrazonal travel modeling. The body of each chapter covers the following subjects:

- Introduction
- Research Methods
- Data Collection

- Data Evaluation (or Analysis)
- Conclusions

1.5 Research Methods

In this study, 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 data set has required about three or four months for acquisition and processing. This 32/31 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 and intrazonal travel/internal capture. 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 two 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 crude estimate of intrazonal travel time based on TAZ area, ignoring many of the D variables.

Given the nested nature of the data sets (with households nested within TAZs and TAZs nested within regions, the modeling will necessarily be multi-level. 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 multi-level modeling (MLM) to account for shared characteristics.

2.0 CAR SHEDDING MODEL

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.

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 report 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 are 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 consideration of only some 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 at all (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 predictions among all of the 25 MPOs.

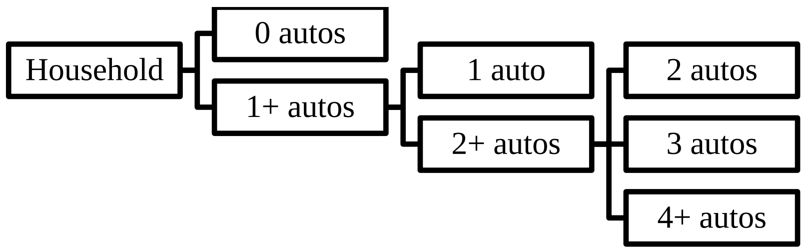

Figure 2.1. Auto Ownership Model Structure of OKI

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 is 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 MPO 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.

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.

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 or categorical 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. As such, a count regression model better fits the data.

Previous studies have compared vehicle ownership 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 a MNL 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 multi-level 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 predicted in the 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 surveys

At present, we have consistent household data sets for 32 regions. The resulting pooled data set 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	Survey Date	Surveyed Households	Surveyed Trips	Mean of Household Vehicles
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 four-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 (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 are 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
jobpop[a]	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 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 a fixed effect model to extract all of the variation 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 ownership. We estimate two discrete choice models, i.e., ORL and MNL since they are used more frequently by travel demand modelers in other regions. Besides, we estimate a count regression model, i.e., Poisson regression model. 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 data set, and the Poisson model is more appropriate than the negative binomial model (see Figure 2.2).

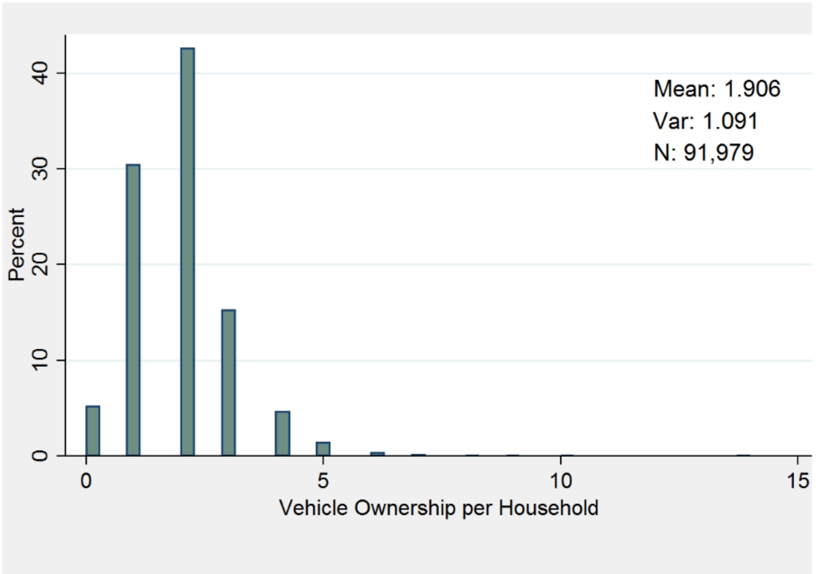

Figure 2.2. The percentage frequency distribution of household vehicle counts

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 regional 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 R^2	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

The computation of the expected number of vehicles in Poisson model is quite straightforward and is 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 slightly better than the others.

Lastly, we have the 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. Thus, 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+ cell which is 3.40. That will mitigate the undercount.

In sum, 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 and we will use 3-level Poisson model with fixed regional 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 0.14. We have shown the pseudo-R^2 largely because urban planners are used to dealing with R^2s and may want this information. Note that Pseudo-R^2s in multilevel Poisson regressions are not equivalent to R^2s in ordinary least squares regression, and should not be interpreted the same way. The pseudo-R^2 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 R^2 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 R^2: **0.1361**

* Household size of 5 is the reference category.
** Three or more workers in a household is the reference category.

An elasticity is a percentage change in one variable with respect to one percent 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 percent and averages 16.88 percent. The difference in vehicle ownership between the household that has average access to regional employment by transit and the one that has the maximum is about -11 percent. So, for a household that is average in all other respects, vehicle ownership will drop from 1.92 to 1.71 vehicles per household as transit accessibility climbs from average to the maximum.

2.6 Model Validation

Our approach is theoretically more solid in the sense that it incorporates influential built environment characteristics of TAZs 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 Wasatch Front Travel Model. In other words, does the Poisson model outperform the current model?

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 in the survey. 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 representation 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 the Poisson model outperforms the Wasatch Front Travel Model.

The correlation between the predicted value versus the actual number of vehicles, along with the root mean square error (RMSE) which was explained in the previous section, are appropriate measures 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 the Poisson model is 0.2293 while this number for the Wasatch Front Travel 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 Wasatch Front Travel Model. Based on these results, we can conclude that our model performs much better than the Wasatch Front Travel Model.

Table 2.6. Summary of the results

	Best-Fit Model	Wasatch Front Travel 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 in the WFRC Travel Survey	0.8506	0.0882

2.7 References

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.

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

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand. Cambridge, MA: MIT Press.

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.

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.

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.

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.

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.

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.

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.

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

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.

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

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.

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

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.

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

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

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

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.

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.

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

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

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

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.

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

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.

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.

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.

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

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

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.

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.

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

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.

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.

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

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

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.

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

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.

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.

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

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

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

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

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

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

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.

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

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.

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.

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

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

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

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.

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

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.

3.0 INTRAZONAL TRAVEL MODEL

3.1 Introduction

In conventional travel demand modeling, 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 percent or more of all trips in household travel surveys. They are treated like other zonal interchanges in the trip distribution step, even though they are inherently different. In the four-step model, trip productions and attractions are modeled as occurring at a single point, 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. There is no travel time between productions and attractions, and therefore one has to be synthesized, ordinarily without regard to micro-land use patterns and street network characteristics.

This report presents a new method for modeling intrazonal trips that addresses the shortcomings of traditional approaches to intrazonal trip modeling in two ways. First, we employ a dataset with disaggregated travel survey data coupled with TAZ-specific built environmental measurements. This 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. Second is the use of discrete choice modeling. Where traditional methods employ the gravity model which 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.

This report 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.

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 a misrepresentation 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:

$$T_{ij} = \frac{A_j F_{ij} K_{ij}}{\sum_{allzones} A_x F_{ij} K_{ix}} x P_i$$

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:

$$c_{ii} = \frac{r_i}{\sqrt{2}}$$

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). This is the approach taken by WFRC and MAG.

$$\textit{Intrazonal Time} = \frac{0.5\times\sqrt{(\text{zonal Area})}\times 60}{\text{Intrazonal Speed (Area Type)}}$$

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 later.

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	Intra-zonal travel time is calculated based on 4

			Choice Model	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	For Home-Based Work: Gravity model For Other types: Destination choice model	The nearest neighbor rule (half of the average distance to 3 nearest zones)
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/ Destination Choice Model	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 percent and 33.3 percent 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
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 percent 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 Our Methodology

3.4.1 Data

For 31 regions (Table 3.2), household travel surveys were collected from MPOs. The surveys were conducted between after 2006. 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 data set 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 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 and trips within the TAZ.

3.4.2 Variables

In this study, the D variables of the built environment were measured and used to predict 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	S.D.
Intrazonal	trip remaining internal to TAZ (1=intrazonal, 0=interzonal)	843,287	0.11	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)	-	-	-
totpop	total population within TAZ	25,396	1,832.76	1,664.44
totemp	total employment within TAZ	25,396	611.60	1,065.82
area	gross land area of TAZ in square miles	25,396	1.82	10.57
actden	activity density within TAZ (pop + emp per square mile in 1000s)	25,396	7.05	21.14
jobpop[(a)]	job-population balance within TAZ	25,396	0.55	0.28
Intden	intersection density within TAZ	25,396	98.39	80.52
pct4wy	percentage of 4-way intersections within TAZ	25,396	25.80	20.10
pctemp10a	percentage of regional employment within 10 minutes by car	25,396	6.93	11.01
pctemp20a	percentage of regional employment within 20 minutes by car	25,396	27.4	25.2
pctemp30a	percentage of regional employment within 30 minutes by car	25,396	49.3	30.2
pctemp30t	percentage of regional employment within 30 minutes by transit	25,396	16.81	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 household 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.

3.5 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 and 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. If a TAZ has only one trip (which is external) from the survey, it gets 0% 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).

3.6 Results

3.6.1 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 the change in 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**
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					

3.7 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 data set, 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).

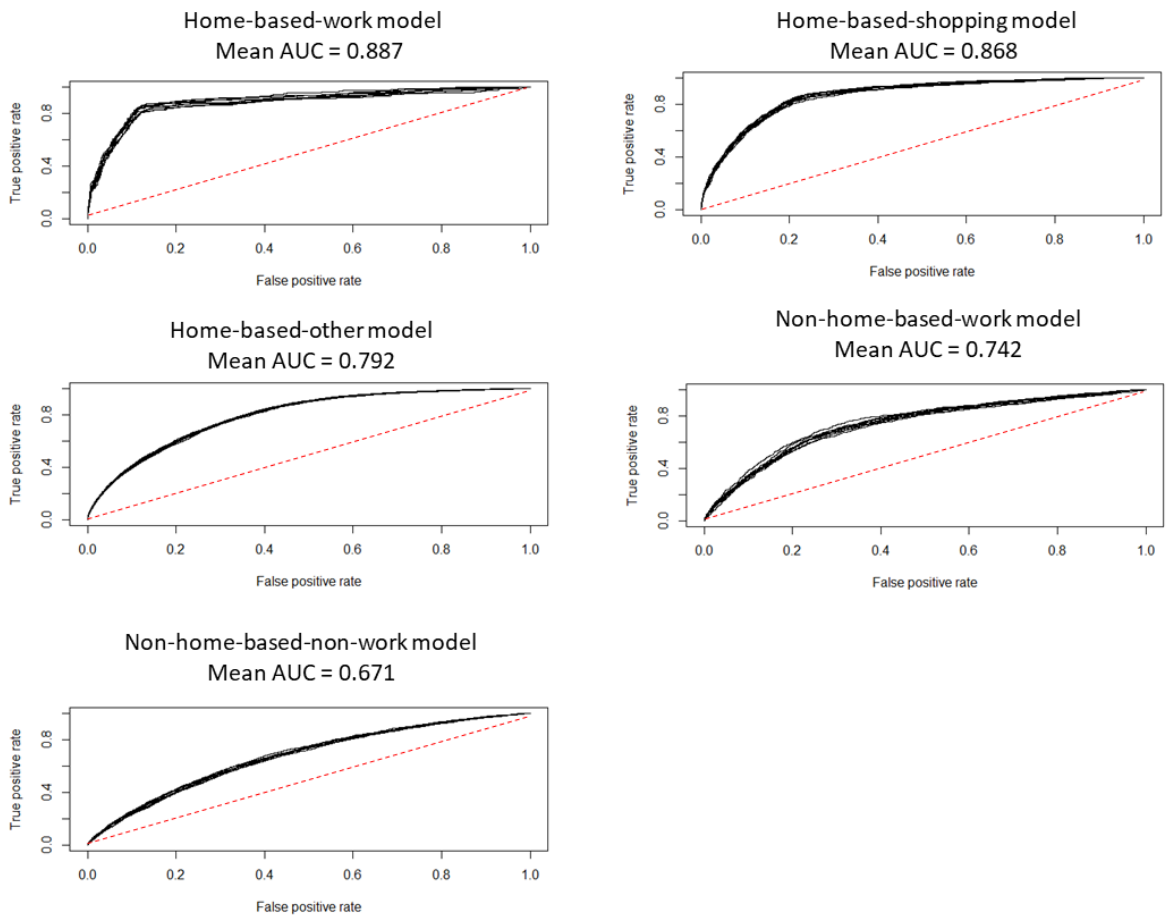

Figure 3.2. Model validation (1): Receiver operating characteristic (ROC) curves and the area under the ROC (AUC) statistics for measuring predictive power of the models

In addition to the k-fold validation, we 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**

3.8 References

Anas, A.: The combined equilibrium of travel networks and residential location markets. Regional Science and Urban Economics, 15(1), 1-21 (1985)

Batty, M.: Urban Modeling: Algorithms, Calibrations, Predictions. Cambridge University Press, London (1976)

Bhatta, B. P., & Larsen, O. I.: Are intrazonal trips ignorable? Transport policy, 18(1), 13-22 (2011)

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.

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

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

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

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.

Dowling, R. G.: Predicting air quality effects of traffic-flow improvements: final report and user's guide (Vol. 535). Transportation Research Board (2005)
Ewing, R., Tilbury, K.: Sketch models for integrated transportation and land use planning. Gainesville, FL.: Metropolitan Transportation Planning Organization (2002)
Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. Journal of the American Planning Association, 76(3), 265–294.
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.
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.
Fresno Council of Governments. Fehr & Peers Transportation Consultant: Travel Demand Model: Model Description & Validation Report for Fresno Council of Governments (2014).
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)
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.
H-GAC.: 2012 Model Validation and Documentation Report, Houston-Galveston Area Council Regional Travel Models, Appendix 4 (2014)
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.
Hamilton, B. W.: Wasteful commuting again. Journal of political economy, 97(6), 1497-1504 (1989)
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.
Hastie, T. J. & Tibshirani, R. J. (1990). Generalized Additive Models. Chapman & Hall/CRC.
Horner, M. W., & Murray, A. T.: Excess commuting and the modifiable areal unit problem. Urban Studies, 39(1), 131-139 (2002)

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

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)

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)

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

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

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

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

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.

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

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.

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

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

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)

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

PBS&J.: ARTS 2035 Long Range Transportation Plan, Appendix B: The Travel Demand Model for the ARTS MPO (2010)
PBS&J.: The Travel Demand Model for the Brunswick MPO (2010)
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)
Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousdand Oaks, CA: Sage Publications.
RVAMPO: Constrained Long-Range Transportation Plan 2035 (2011), Available at: http://rvarc.org/chapters-4-and-5-technical-process-clrtp-2035/
StanCOG: The StanCOG Transportation Model Program: Documentation and Validation Report, Version 1 (2010). Avilable at: http://www.stancog.org/trans-model.shtm
Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285-1293.
The Association of Monterey Bay Area Governments: Association of Monterey Bay Area Governments Regional Travel Demand Model Technical Report (2015)
The Des Moines Area MPO: Travel Demand Model: Documentation & User Guide (2006)
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)
Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry, 39(4), 561-577.

4.0 CONCLUSIONS AND RECOMMENDATIONS

4.1 Car Shedding Model

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. A pre-step to the four-step process is estimating vehicle ownership by TAZ for some future (target) year. 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 by WFRC and by MPOs in other regions of the U.S., especially 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) 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 better predictive accuracy than the MNL model.

For urban planning and design, this study suggests that car shedding occurs as built environments become more dense, mixed, connected, and served by transit. This finding has important implications for 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).

4.2 Intrazonal Travel Model

Trip distribution is one of the critical steps in travel demand forecasting. It is the second step in the conventional four-step model, and for nearly all MPOs is accomplished with the gravity model. Intrazonal trip distribution is treated like trip distribution to any other zone, except that the fact that all trip productions and attractions are treated like they occur at a single point in space, the zone centroid, and the entire local street network is reduced to one or more connectors to the external street network. The larger the TAZ, the smaller the intrazonal trip share, just the reverse of what we observe in practice.

In our model structure, intrazonal travel is treated as a discrete choice between 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 findings show 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. 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 showed 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. 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, and many large ones. 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.