UNIVERSITY OF CALIFORNIA, IRVINE

Unraveling the Effects of Land Use Planning and Energy Policy on Travel Behavior

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Planning, Policy, and Design

by

Harya S Dillon

Dissertation Committee:
Professor Douglas Houston, Chair
Professor John Hipp
Professor Jae Hong Kim

2017

DEDICATION

To

Humanity

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude to my dissertation committee chair, Prof. Douglas Houston for his role in guiding me developing and finishing my dissertation. I am grateful to my committee members, Prof. John Hipp and Prof. Jae Hong Kim, for their valuable comments and feedback. I am indebted to Prof. Martha Feldman, Director of the Doctoral program, for her support during the toughest time in my doctoral research career. In addition, a thank you to Prof. Marlon Boarnet, Prof. Rodolfo Torres, Prof. Ken Chew, Prof. Victoria Basolo, Prof. Victoria Beard, Prof. Dan Stokols, and Prof. Nick Marantz, for their exemplary mentorship and support at various stages of my doctoral studies. I look forward to the day I can do the same for someone else. I thank Prof. Jean-Daniel Saphores for his role in the early stages of my research.

This dissertation research cannot escape the benefit of support from the School of Social Ecology, Institute of Transportation Studies-UCCONNECT dissertation fellowship, and the US DOT Pacific Southwest Region University Transportation Center fellowship. I am also grateful to the Department of Urban Planning & Public Policy for the generous financial support.

Finally, I would like to extend my heartfelt gratitude to my family and friends for their loving support.

Curriculum Vitae

Education

Ph.D. in Urban Planning & Public Policy June 2017

University of California, Irvine

Dissertation: *Unraveling the Effects of Land Use Planning and Energy Policy on Travel Behavior.* Chair: Prof. Douglas Houston

Master of Urban and Regional Planning July 2011

University of California, Irvine

Capstone project: *Land Use, Gasoline Prices and Travel Behavior: The case of Southern California*

Bachelor of Science in Engineering March 2005

Department of City and Regional Planning, School of Civil Engineering and Planning

Insititut Teknologi Bandung, Indonesia

Senior Thesis: *Traffic Impact Assessment of Urban Megaprojects*

Academic Publications (refereed journal articles)

Dillon, H. S., Saphores, J. D., & Boarnet, M. G. (2015). The impact of urban form and gasoline prices on vehicle usage: Evidence from the 2009 National Household Travel Survey. *Research in Transportation Economics*, *52*, 23-33.

Dillon, H. S. (2015). Social Optimality of Cordon Area Congestion Pricing in a Monocentric City. *Journal of Regional and City Planning*, *26*, 43-53.

Kim, J. H., Hipp, J. R., Basolo, V., & **Dillon, H. S.** (2017). Land Use Change Dynamics in Southern California: Does Geographic Elasticity Matter? *Journal of Planning Education and Research*.

Marantz, N. J., **Dillon, H. S.,** (2017 – Forthcoming). Do State Affordable Housing Appeals Systems Backfire? A Natural Experiment. *Housing Policy Debate*

Honors and Awards

USDOT Pacific Southwest Region University Transportation Center fellowship	2017
Social Ecology Dean's Dissertation Fellowship	2016
ITS UCCONNECT Dissertation Fellowship	2016
Social Ecology Alumni Award (UC Irvine)	2015
Fulbright Scholarship	2009-2011
School of Social Ecology Merit-based Financial Award (UC Irvine)	2009-2011

Abstract of the Dissertation

Unraveling the Effects of Land Use Planning and Energy Policy on Travel Behavior

By

Harya S Dillon

Doctor of Philosophy in Planning, Policy, and Design

University of California, Irvine, 2017

Professor Douglas Houston, Chair

This three-essay dissertation focuses on understanding linkages between urban form, travel behavior, ownership of alternative fuel vehicles, active commuting, congestion, fuel consumption, and air pollution (including greenhouse gas emissions). These essays estimated different specifications of Generalized Structural Equation Models (GSEM) to explicitly account for residential self-selection and vehicle choice endogeneities.

The first essay analyzes the influence of land use policies and gasoline prices on driving patterns. I estimated a Generalized Structural Equation Model (GSEM) with a Tobit-link specification on a Southern California subsample of the 2009 National Household Travel Survey (NHTS). These data haves a quasi-experimental nature thanks to large exogenous variation in gasoline price during the survey period. I analyzed separately home-based work trips and non-work trips under the hypothesis that households have more flexibility to adjust their non-work trips when gasoline prices change, whereas most of the literature does not take trip purpose into account. To measure urban form, which is treated as a latent construct, I used fine-grained geospatial information including population density, land use mix, employment density, distance to employment centers and transit availability. I found that, in the short run, households drive 0.171% less for non-work trips when gasoline prices increase by 1%, while work trips are not

responsive to gasoline price changes. This suggests that, in the short term, higher fuel prices reduce discretionary driving such as shopping and recreational trips, but they do not affect non-discretionary driving such as commuting trips. My results also suggest that policies that seek to increase transit service and housing opportunities near employment centers will reduce driving.

The second essay investigates the impact of government incentives such as access exemption to High Occupancy Vehicle (HOV) lanes and parking privileges on household ownership of Alternative Fuel Vehicles (AFVs) using Generalized Structural Equation Models (GSEM), and accounts for residential self-selection, household demographics and ambient political-environmentalism. I analyzed geocoded travel diary data from the 2012 California Household Travel Survey (CHTS), linked with fueling station data from the US Department of Energy Alternative Fuels Data Center and precinct level election data from the UC Berkeley Statewide Database. My findings suggest that, on average, households with alternative fuel vehicles drive approximately 10 miles more on weekdays and about 0.5 miles more on non-discretionary trips than otherwise similar households. In addition, households who live closer to a freeway with HOV lanes, work closer to an AFV charging facility (that provides free parking), and are likely supportive of pro-environmental measures are more likely to own alternative fuel vehicles.

The third essay examines the influence of urban form on transit use and non-motorized travel (NMT, including biking and walking) for households (with at least one employed adult) in Los Angeles and Orange Counties in California based on 2009 National Household Travel Survey (NHTS) data. The objectives of the research are (1) to assess several methods for measuring urban form features in the near-residence and near-workplace environments and (2) to assess the importance of these urban form features on transit use and NMT after accounting for

the influence of these features on household vehicle ownership and residential selection. Results provide insights into the relative influence of several specifications of population density, transit access and walkability measures on transit use and NMT for commute and non-work trips. Reduced form models suggest that the dominant determinant of discretionary travel is household socio-demographic status. In terms of residential selection, lower income, younger, and smaller households are more likely to choose a dense, pedestrian friendly, and transit rich neighborhood. In terms of vehicle ownership, households living in high density, pedestrian friendly, and transit rich neighborhoods are less likely to own vehicles. After accounting for the influence of urban form on vehicle ownership and residential selection, workplace transit accessibility has greater influence on transit commuting than transit access near a household's residence. Results vary by how urban form is specified and by the source of travel data. Finally, there is some evidence that population density affects active travel for discretionary purposes.

1 Introduction

Understanding linkages between urban form, travel behavior, vehicle choice, congestion, and air pollution (including greenhouse gas emissions) is one of the main challenges facing urban transportation systems in the United States and California in particular. According to the Texas A&M Transportation Institute (TTI) 2015 Urban Mobility Scorecard (UMS), in 2014 congestion cost the economy $160 Billion in delays and wasted fuel (TTI & INRIX, 2015). About $ 23.9 Billion or 15% of that cost was borne by Californians. To put this figure into perspective, California accounts for 13.2% of the US economy (Bureau of Economic Analysis, 2014) and about 12% of US population (US Census, 2017). The Los Angeles and San Francisco regions have been consistently ranked among the most congested cities in the U.S. Therefore, while the average Californian is more productive, his/her travel behavior is disproportionately detrimental to the environment.

The impacts of transportation on local air quality and greenhouse gas (GHG) emissions have been of concern to policymakers for some time. Since the 1980s, proponents of the New Urbanism movement have sought to promote socially and environmentally desirable travel behavior, namely less driving and more walking and transit use, by changing the urban environment. The popularity of these ideas can be attributed to the increasing social and environmental costs of driving (Boarnet & Crane, 2001; Brownstone & Golob, 2009). Today, many MPOs and transportation planners in the United States have turned to land use planning and urban design measures to rein in automobile use (SCAG, 2016).

As in many other parts of the country, transportation accounts for over half of ozone precursors and particulate matter emissions, and for nearly 40 percent of greenhouse gas emissions in California (CA Governor Executive Order No. B-16, 2012; CARB, 2013). One of the responses from legislators is California's Assembly Bill 32 (AB 32), the Global Warming Solutions Act 2006 and State Bill 375 (SB 375) also known as The Sustainable Communities and Climate Protection Act of 2008. SB 375 requires 18 Metropolitan Planning Organizations (MPOs) to develop Sustainable Community Strategies (SCS) that demonstrate how they will meet GHG reduction targets set by the California Air Resources Board, through integrated land use, affordable housing and transportation planning strategies (CARB, 2013). Once adopted by an MPO, its SCS is incorporated into a Regional Transportation Plan (RTP), a planning document that is required for federal infrastructure funding. In particular, the Southern California Association of Governments (SCAG) developed plans that are explicitly aimed at reducing the use of private cars and increasing the use of transit, walking and bicycling by coordinating transit corridor investments with housing supply trajectories (SCAG, 2016). In November 2016, Los Angeles County voters generously passed Measure M which is expected to generate $120 billion in public transit investment from a half percent increase in sales tax for the next 40 years.

Furthermore, to reduce the local, regional and global air pollution from transportation, California has decided to reduce its transportation petroleum use by 50% by 2030. One key strategy to achieve this ambitious goal is to increase the share of alternative fuel vehicles (AFVs) on the road (CA Governor Executive Order No. B-16, 2012). Since travel behavior is influenced by policies that shape urban form and driving costs (such as gasoline taxes and tolls), informed policy choices play a pivotal role in improving the region's overall economic competitiveness.

This three-essay dissertation seeks to contribute to our understanding of linkages between urban form, travel behavior, ownership of alternative fuel vehicles, active commuting, congestion, fuel consumption, and air pollution (including greenhouse gas emissions). These three essays share some common research design and methodological approaches: (1) using disaggregate data at the household level, and (2) addressing endogeneity of residential selection and/or vehicle ownership.

In this dissertation, the household is the assumed to be the behavioral agent making travel decisions. Household members typically share resources and common goals – maximizing common utility under common budget constraints. A household is typically defined as co-residents sharing the same housing unit and a fixed co-location which affects home-based travel decisions. It is also common for household members to share vehicles and make collective destination decisions. Furthermore, travel surveys used in this research, the 2009 National Household Travel Survey (NHTS) and the 2012 California Household Travel Survey (CHTS), used geographically stratified sampling techniques to target sampled households. Thus, household members are not identically and independently distributed within the household. That said, intra-household dynamics such as car sharing, budgeting, and priority among members, are outside the scope of this research.

In a geographically stratified household travel survey, residential location influences the probability of being observed in the sample. Sampled household had previously made its residential decision and therefore residential self-selection can be understood as a sampling problem. Although households were recruited using random-digit dialing, the probability of being sampled/observed in a particular location/neighborhood is not random. The 2009 NHTS and 2012 CHTS are cross-sectional surveys; they do not repeat measurements for households and thus I cannot control for time-invariant factors that might affect residential choices. Likewise, households had previously made their vehicle ownership decisions prior the survey date. Vehicle ownership or access and travel behavior outcomes such as trip rate, travel distance, and mode choice, may present a causal feedback. For these reasons, addressing endogeneities is an important contribution of this dissertation to travel behavior research. As such, these essays estimated different specifications of Generalized Structural Equation Models (GSEM) to explicitly account for residential self-selection and vehicle choice endogeneities (Brownstone, 2008; Brownstone & Golob, 2009).

The remaining chapters are organized as follows. Chapter 2 analyzes the influence of land use policies and gasoline prices on driving patterns. I estimated a Generalized Structural Equation Model (GSEM) with a Tobit-link specification on a Southern California subsample of the 2009 National Household Travel Survey (NHTS). I found that, in the short run, households drive 0.171% less for non-work trips when gasoline prices increase by 1%, while work trips are not responsive to gasoline price changes. This suggests that, in the short term, higher fuel prices reduce discretionary driving such as shopping and recreational trips, but they do not affect non-discretionary driving such as commuting trips. My results also suggest that policies that seek to increase transit service and housing opportunities near employment centers will reduce driving.

Chapter 3 investigates the impact of government incentives such as access exemption to High Occupancy Vehicle (HOV) lanes and parking privileges on household ownership of Alternative Fuel Vehicles (AFVs) using Structural Equation Modeling techniques, and accounts for residential self-selection, household demographics and ambient political-environmentalism. I estimated a Generalized Structural Equation Models (GSEM) with a logit-link specification on the 2012 California Household Travel Survey (CHTS). My findings suggest that, on average, households with alternative fuel vehicles drive approximately 10 miles more on weekdays and about 0.5 miles more on non-discretionary trips than otherwise similar households. In addition, households who live closer to a freeway with HOV lanes, work closer to an AFV charging facility (that provides free parking), and are likely supportive of pro-environmental measures are more likely to own alternative fuel vehicles.

Chapter 4 examines the influence of urban form on transit use and non-motorized travel (NMT, including biking and walking) for households (with at least one employed adult) in Los Angeles and Orange Counties in California based on 2009 National Household Travel Survey (NHTS) data. Results provide insights into the relative influence of several specifications of population density, transit access and walkability measures on transit use and NMT for commute and non-work trips. Reduced form models suggest that discretionary travel is dominated by household's socio-demographic status. Results vary by how urban form is specified and by source of travel data. Finally, there is some evidence that population density affects active travel for discretionary purposes.

Finally, Chapter 5 provides concluding remarks and suggestions for future travel behavior research.

References

Boarnet, M. G., & Crane, R. (2001). *Travel by design: The influence of urban form on travel*. Oxford University Press.

Brownstone, D. (2008). Key relationships between the built environment and VMT. Transportation Research Board, 87(August), 466–478. Retrieved from http://onlinepubs.trb.org/Onlinepubs/sr/sr298brownstone.pdf

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, *65*(1), 91–98. doi:10.1016/j.jue.2008.09.002

California Governor Exec. Order B-16-2012, (2012) (https://www.gov.ca.gov/news.php?id=17472)

Southern California Association of Governments (2016). *The 2016-2040 Regional Transportation Plan/Sustainable Communities Strategy (2016 RTP/SCS or Plan)*. Los Angeles, CA: Author.

The Texas A&M Transportation Institute and INRIX (2015). *Urban Mobility Scorecard*. College Station, TX: Author. Retrieved from: https://mobility.tamu.edu/ums/

US Department of Commerce, Bureau of Economic Analysis (BEA) (2014). Widespread But Slower Growth in 2013 [Press release]. Retrieved from http://www.bea.gov/ newsreleases/regional/ gdp_state/ 2014/pdf/gsp0614.pdf

US Department of Commerce, Census Bureau (2017). California quick facts. Retrieved from https://www.census.gov/quickfacts/CA

California Air Resources Board (2014, August 13). Senate Bill 375 - Regional Agencies and Data in Target Setting. Retrieved from https://www.arb.ca.gov/cc/sb375/mpo/info.htm

2 The Impact of Urban Form and Gasoline Prices on Vehicle Usage: Evidence from the 2009 National Household Travel Survey[1]

This paper relies on generalized structural equation modeling (SEM) to tease out the relationship between land use, gasoline prices and travel behavior. We analyze data from the Southern California subsample of the 2009 National Household Travel Survey (NHTS), which has a quasi-experimental nature thanks to large exogenous variations in gasoline prices during the administration of the NHTS (March 2008-April 2009). Our joint models of residential urban form, vehicle efficiency choice, and vehicle use account for residential self-selection and endogeneity of vehicle preferences in order to explain vehicle miles traveled (VMT) for both work and non-work trips. Residential urban form is treated as a latent construct that reflects observed variables such as population density, land use diversity and distance to employment centers. Our results suggest that in the short run, households drive 0.15% less for all trips and 0.18% less for non-work trips when gas prices increase by 1%, while work trips are not responsive to gasoline price changes. Moreover, the direct effect of residential urban form on driving is statistically significant for total and non-work VMT, but it has no impact on work trips. We also find that owners of more fuel efficient vehicles tend to be more educated, Asian and younger (under 30). Moreover, households in low density neighborhoods are more likely to have a higher income, to be older than 65 and either White or Asian; these households tend to own more vehicles per driver. Finally, our results show that accounting for the nature of trips is important for understanding the short term price elasticity of travel. Keywords: land use; travel behavior; structural equation modeling; gasoline prices.

2.1 Introduction

How does urban form affect household travel behavior? How much do households adjust their travel behavior and vehicle usage in the short-run when gasoline price changes? These questions, which are at the forefront of transportation planning and policy debates, have far reaching impacts on communities that are highly dependent on automobiles such as Southern California, but few studies have analyzed urban form and gas prices together. The relationship between land use and travel behavior has been the subject of many studies in recent years (e.g., Bagley & Mokhtarian, 2002; Boarnet, 2011; Cao, Mokhtarian, & Handy, 2009a, 2009b, 2010; Chao & Qing, 2011; De Abreu e Silva *et al.*, 2007; Houston *et al.*, 2014; Lovejoy *et al.*, 2013; Mokhtarian & Cao, 2008). As new mandates such as California's 2008 Senate Bill 375 have directed policymakers to reduce vehicle miles traveled by changing urban form, there have been lingering debates about the relative efficacy of land use strategies and pricing strategies in reducing travel, but few studies have been able to examine both questions, in large part due to the cross-sectional nature of most transportation data sets.

In this paper, we take advantage of the fact that the 2009 National Household Travel Survey (NHTS)[2] was administered during a time of large gas price changes to examine how gas prices and land use influence household vehicle miles traveled (VMT). We analyze a southern California sub-sample of the 2009 NHTS. As in previous studies (Bagley & Mokhtarian, 2002; Boarnet & Crane, 2001; Boarnet, 2011; Brownstone & Golob, 2009; Cervero & Murakami, 2010; Chao & Qing, 2011; De Abreu e Silva et al., 2007; Gillingham, 2013; J. Hong, Shen, & Zhang, 2013; Kim & Brownstone, 2013), we rely on common land use metrics but instead of treating them as observed variables that capture urban form, we conceptualize urban form as an

[2] U.S. Department of Transportation, Federal Highway Administration, 2009 National Household Travel Survey. URL: http://nhts.ornl.gov

unobserved (latent) construct characterized by various land use variables in a measurement model, which we estimate using confirmatory factor analysis (CFA). This approach acknowledges the complexity of measuring urban form and relaxes the assumption that observable metrics measure urban form without errors.

In addition, we analyze separately home-based work trips and non-work trips under the starting hypothesis that households have more flexibility to adjust vehicle miles traveled (VMT) for their non-work trips when gasoline prices change, whereas most of the literature does not take trip purpose into account when analyzing the impacts of gasoline price changes and urban form on VMT. Work trips are trips to and from a workplace whereas non-work trips capture all other trips including trips for shopping, social and recreational purposes. We investigated travel behavior with a finer classification of trip purposes but trip chaining behavior substantially complicated the classification of non-work trips so we focus on this simpler classification here.

This paper is organized as follows. The next section presents a limited review of some recent papers to motivate our work. In Section 3, we present our data before outlining our methodology in Section 4. In Section 5, we discuss our results starting with model fit and continuing with result interpretation. Finally, in Section 6 we summarize our findings, outline limitations of our analyses, and present suggestions for future work.

2.2 Literature Review

To motivate our modeling choices and contextualize our results, we first present a brief review of selected papers on urban form and travel behavior, highlight some methodological treatments of endogeneity in travel behavior research, including structural equation modeling, and finally review a few salient papers dealing with travel behavior and gas price elasticity.

2.2.1 Urban Form and Travel Behavior

As indicated in recent reviews (e.g., see Boarnet *et al.*, 2011; Boarnet, 2011; Brownstone, 2008; Ewing & Cervero, 2010) most papers concerned with the relationship between urban form and travel behavior estimate reduced form models where measures of travel behavior (VMT, transit use, walking, trip frequency) are regressed on individual and household socio-demographic variables (i.e., age, ethnicity, education level, household income and size) and measures of urban form (density, diversity, design, and accessibility) around each household's residence. A wealth of empirical studies report that some features of the built-environment, particularly high density and measures of transportation access to employment, are associated with lower vehicle use (e.g., see Bento *et al.*, 2005; Brownstone, 2008; Bhat and Guo, 2007; Chen *et al.*, 2008; and Fang, 2008). Based on this literature, the 2009 National Academy of Sciences/National Research Council (NAS/NRC) report concluded that the elasticity of VMT with respect to population density ranges between -0.05 and -0.12.

However, as shown by Bento *et al.* (2005), simultaneous changes on multiple facets of urban form can have much larger effects on driving. They analyzed travel diary data from the 1990 Nationwide Personal Transportation Survey (NPTS) and a variety of urban form measures from 114 U.S. metropolitan areas, including population, housing, and jobs distributions, transit supply and road density. They found that the magnitude of the elasticity of VMT with respect to most measures of urban form is small (≤0.07 in absolute value) but that the elasticity of VMT with respect to comprehensive measures of the built environment could be as large as -0.4 when different facets of urban form change jointly.

Greenhouse gas (GHG) reduction laws such as California Senate Bill 375 have elevated urban form and travel behavior research, and particularly questions about the magnitude of the elasticity of driving with respect to urban form, to the center of policy discussions. Regulatory mandates assume that urban form has a causal impact on driving behavior, and especially on vehicle miles traveled (VMT), but the link between the two remains controversial. Skeptics argue that urban form (and the spatial distribution of population in an urban area) is shaped, in part, by highway development and that highway expansions are designed to accommodate driving, thus blurring the direction of causality, so efforts to understand this relationship ultimately need to rely on quasi-experimental research for an unambiguous answer (Baum-Snow, 2007; Funderberg *et al.*, 2010; Duranton & Turner, 2012).

Using the 1947 plan for the U.S. National Highway System to instrument for the built Interstate Highway System, Baum-Snow (2007) concluded that instead of an 18% decline from 1950 to 1990, central city population would have increased by 8% if the Interstate Highway System had not been built. Duranton and Turner (2012) reached similar conclusions with job growth: a 10% increase in highway stock caused a 1.5% growth in employment over 20 years (1983-2003). Since new highways created new employment centers outside of traditional city centers and given that the increase in highway miles was much larger outside city centers, highway expansion did have a causal effect in decentralizing employment away from city centers. Funderburg *et al.* (2010) provide similar results based on a case study of highway expansions in Orange County, California. They find that new highways are associated with increases in employment in nearby census tracts on the order of a few thousand new jobs. Duranton and Turner (2012) go further, showing that additional highway capacity induces more driving. Their results confirm that urban form policies (highway expansion here) have a causal impact on driving.

However, merely demonstrating the presence of this causal effect is not sufficient to justify reshaping urban form to limit VMT as a climate change policy instrument. Instead, it is critically important to correctly measure the elasticity of VMT with respect to urban form (Boarnet *et al.*, 2011; Boarnet, 2011; Brownstone, 2008) because underestimating it would neglect a potentially important avenue to reduce greenhouse gases, while overestimating its value would waste public investments on a large scale.

Possible sources of bias that plague the travel behavior literature are endogeneity associated with residential and vehicle choice self-selection (Bhat & Guo, 2007; Cao *et al.*, 2009a-b; Mokhtarian & Cao, 2008). The former implies that travel behavior influences residential location, and the latter implies that households select their vehicles (and their characteristics, such as fuel efficiency) partly based on their transportation preferences.

2.2.2 Methodological Response to the Endogeneity Problem in Travel Behavior Research

Controlling for residential self-selection is important for obtaining unbiased estimates of the impact of urban form on travel behavior (Bagley & Mokhtarian, 2002; Boarnet & Crane, 2001; Boarnet, 2011; Brownstone & Golob, 2009; Cao et al., 2009a; Kim & Brownstone, 2013; Mokhtarian & Cao, 2008). Typical travel diary data, such as collected by the NHTS, surveyed a cross-section of households to capture their travel behavior after they chose their residence and their vehicles; urban form and vehicle fuel efficiency are not randomly assigned by a clever research design, but confounded in travel behavior. Addressing this potential self-selection problem is therefore the biggest hurdle in travel behavior research.

Another self-selection issue that affects driving is vehicle fuel economy (Bhat & Guo, 2007; Brownstone & Golob, 2009; Fang, 2008; Kim & Brownstone, 2013). Households with fuel efficient vehicles are less sensitive to fluctuations in gas prices. Conversely, households who prefer driving may drive vehicles that are more fuel efficient to reduce spending, and may be more willing to live in a location that necessitates commuting. Unobserved preferences for residential location may thus jointly affect vehicle fuel economy and driving behavior, leading Brownstone and Golob (2009) and Kim and Brownstone (2013) to treat residential density and vehicle fuel efficiency as jointly-endogenous in their structural models. In addition, households may choose their vehicles based on future expectations about gasoline prices. Gillingham (2013) argues that households with foresight about fuel prices would buy more efficient vehicles, thus creating an additional source of endogeneity in driving behavior models.

Recent studies suggest that a rich set of socio-demographic controls (which are available in recent travel diary surveys such as the 2009 NHTS) can reduce the bias associated with residential self-selection (Boarnet, 2011; Brownstone, 2008; Cao *et al.*, 2009a, 2010). However, including a rich set of socio-economic control variables in disaggregated, single equation models does not guarantee that households in some neighborhoods do not differ by some unobserved trait that affects their travel behavior (Brownstone & Golob, 2009; Kim & Brownstone, 2013). The gold standard for addressing this problem with cross-sectional data is to construct a joint-model of residential urban form and travel behavior (Boarnet & Crane, 2001; Brownstone & Golob, 2009; Kim & Brownstone, 2013).

2.2.3 Structural Equation Modeling and Travel Behavior Research

In this study, we correct for self-selection by specifying a system of simultaneous equations where urban form, vehicle fuel efficiency and household VMT are jointly determined so it is useful to review some applications of SEM to travel behavior research. For brevity, we restrict ourselves to a few directly relevant papers. Golob (2003) provides a nice review of structural equation modeling (SEM) and its applicability to travel behavior research.

SEM techniques have often been applied to models of vehicle use and ownership for cross-sectional data to simultaneously capture the mutual causal effects between vehicle ownership and use (Golob, 2003, p. 12). While they endogenized vehicle ownership, many of these studies treated land use characteristics/urban form as exogenous and ignored residential self-selection issues. Studies that address residential self-selection in travel behavior – urban form interactions remain scarce (Van Acker, Mokhtarian, & Witlox, 2014, p. 89), even though strong evidence of self-selection bias exists (Cao *et al.*, 2009a; Mokhtarian & Cao, 2008).

In a SEM framework, residential self-selection can be accounted for by specifying residential urban form and measures of travel behavior as jointly-endogenous (Brownstone & Golob, 2009; Kim & Brownstone, 2013). However, SEM techniques can only confirm causality claims from theoretically informed models. Causality claim should not only satisfy correlation and non-spuriousness, but also temporality (i.e., cause must precede effect). While spuriousness and correlation can be confirmed by testing estimated coefficients, models relying on cross-sectional data may not satisfy temporality. To address that problem, a non-recursive model with a feedback loop between two endogenous variables can be used to satisfy temporality, although identification may be more difficult in such a setting.

2.2.4 Driving and Gas Price Elasticity

Since late 1970s US policymakers have been interested in how households adjust their behavior in response to changes in real gasoline prices in order to design more effective energy and environmental policies (Congressional Budget Office, 2008; Puller & Greening, 1999).

From aggregate trip frequency and highway traffic data collected in California, the Congressional Budget Office (CBO) (2008) concluded that the short-run effect of gasoline price changes is small because households do not have time to change their vehicles or to move. They found that in areas with no rail transit, the number of freeway trips declined by 0.7% for every $0.50 increase in gasoline prices on weekdays, which is equivalent to an (arc) elasticity of -0.044.

To tease out short-run and long-run impacts of gasoline prices on driving and fuel consumption, several papers have relied on time series analyses (e.g., see Puller & Greening, 1999, or Small & van Dender, 2007). As fuel economy standards increase, people use less fuel ceteris paribus. As cars become more fuel efficient, people buy more efficient vehicles and may then drive more on a gallon of fuel, leading to the so-called "rebound effect."

After summarizing studies focusing on how VMT changes with gasoline prices, Puller & Greening (1999) reported elasticities ranging from -0.26 in the short-run to -0.86 in the long-run. Results from their two-equation model applied to 9 years of household-level Consumer Expenditure Survey (CES) panel data gave a long-run elasticity of non-business gasoline demand of -0.35 after accounting for changes in household vehicle stocks.

Small and van Dender (2007) estimated the elasticity of vehicle miles traveled (VMT) with respect to gas price. They found that total driving is not highly responsive to gasoline prices: a 1% increase in gasoline prices is expected to reduce driving by only 0.02% to 0.03% in the short-run. They partly justified this lack of responsiveness by noting that increases in real household income over time have reduced the share of disposable income used for driving.

Bento, Hughes, & Kane (2013) used a different approach based on weekly traffic flow data in High Occupancy Vehicle (HOV) lanes reserved for vehicles with two or more passengers in Los Angeles County. Unlike other large U.S. cities, most HOV lanes in southern California have permanent restrictions. This allows associating weekly changes in average gasoline prices with changes in traffic flow in these lanes. Their results show that a 1% increase in gasoline price implies 1 additional carpool per hour in HOV lanes. More carpools imply less VMT for the same level of passenger-miles traveled.

Also in California, Gillingham (2013) estimated the medium-run elasticity of driving with respect to gasoline price based on a 2001-2003 panel of vehicle smog-check data from registered vehicles. His quantile regression results reveal the presence of heterogeneity in elasticity values, which range from -0.33 to -0.17, with a higher responsiveness for households with larger income, perhaps due to within-household vehicle-switching.

2.3 Data

2.3.1 NHTS Travel Diary Data

This paper analyzes geocoded data from a subset of Southern California (Los Angeles, Orange, Ventura, San Bernardino and Riverside counties) residents who participated in the 2009 National Household Travel Survey (NHTS). The geocoded raw data include the latitude and longitude of each household location and travel destination.

We focus on households that own at least one vehicle, which represents approximately 96% of households in our NHTS sample and corresponds to 3,752 households; however, because of missing income information, our sample size is 3,511. Since this paper investigates how households respond to changes in gasoline prices in the short-run, we assume that gas price changes do not influence vehicle ownership decisions.

The 2009 National Household Travel Survey (NHTS), which was conducted from March, 2008 to April, 2009, includes a one-day diary of all household travel, and collected detailed household socio-demographic information. Participating household members were asked to record all of their trips on an assigned day, and to estimate the length of each trip in blocks; each block was then converted to 1/8 of a mile to estimate vehicle miles traveled (VMT). We construct our dependent variable from these data by summing distance travelled for work and non-work trips for each household, taking care not to double count distances when more than one household member traveled in the same vehicle.

2.3.2 Gas Prices

Fortuitously, the 2009 NHTS covered a period of remarkable gas price variations. Gas prices during the 2009 NTHS survey period varied from a low of $1.75 per gallon on the week of December 12, 2008 to a high of $4.46 per gallon on the week of June 20, 2008 (see **Figure 2-1**). The survey period covered the dramatic price increases in late spring and early summer of 2008, the price declines of the summer of 2008 (which accelerated rapidly during the onset of the Great Recession), and then a period of price rebound and relative price stability from January to April, 2009. Importantly, these price changes were almost certainly supply curve shocks: price increases likely reflected strong and growing world oil demand, and possibly the effect of unfavorable exchange rates versus the U.S. dollar in late spring 2008, while the price declines reflected a sudden drop in world demand. These price changes were large and exogenous to households, so the NHTS provides travel data for a natural experiment during a period of large and rapid gas price changes.

NHTS gas data are average prices for the week a household was surveyed, which for our southern California sub-sample, are the U.S. Energy Information Administration (EIA) weekly average for the Western Region of the U.S.

2.3.3 Measures of Urban Form

To measure urban form, we used fine-grained geospatial information; our land use variables include population density, land use mix, employment density, distance to employment centers and transit availability. Using Geographical Information System (GIS) software, we assigned land use characteristics to each household based on residential location. We relied on 2010 census data to measure population density at the block group level. As a proxy for employment accessibility, we extracted from Census Transportation Planning Products (CTPP) data the number of jobs in circles with a 10 miles radius centered on each household's residence.

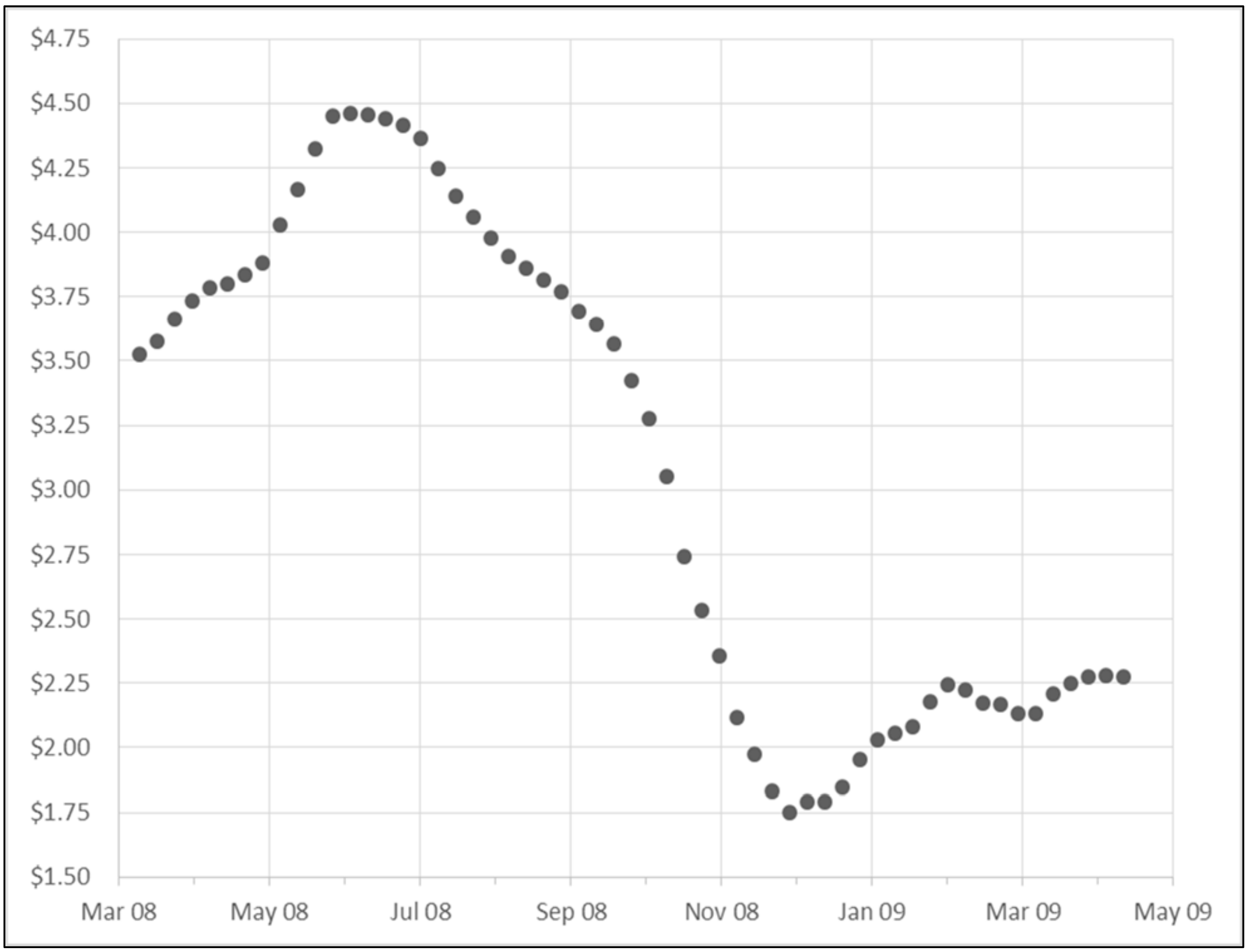

Figure 2-1 Gas Prices (in US $) versus Survey Date

Land use mix was measured from SCAG's 2005 parcel level land use data. This information is available in GIS, which allowed calculating an index of land use diversity at the block group level using the entropy formula (e.g., see De Abreu e Silva, Golob, & Goulias, 2007) as follows:

$$H_i = \frac{-1}{\ln(k)} \sum_{j=1}^{k} p_{ij} \ln(p_{ij}), \qquad (1)$$

where H_i is the land use entropy of block group i, and p_{ij} is the area's proportion of land use of type j in block group i. SCAG's land use database stores parcel level land use based on 150 categories, which we condensed into k=15 major types of land uses.[3]

Households with access to transit service have the opportunity to substitute driving with transit. While transit is not a perfect substitute for driving, certain conditions (e.g., high gas prices or parking constraints) may make transit preferable to driving, depending on trip purpose. Households with higher opportunities to use transit are expected to use transit as a substitute when gasoline becomes prohibitively expensive.

Transit service density data were obtained from SCAG's transit network and from the Los Angeles Metropolitan Transit Authority GIS database. The shapefile we received contains data about each transit stop, including information on service frequency and connectivity. To measure level of service, we aggregated the level of service for each transit stop. By using proximity analysis with ArcGIS, we calculated transit service level by aggregating the number of

[3] These land uses category are: single family residential, multi-family residential, other residential types (e.g. mobile home and trailer parks) commercial and services, industrial, transportation, communication and utilities, public facilities such as government offices, schools and libraries, military land uses, mixed development, open space and recreational, vacant urban land, agricultural, vacant non-urban land (e.g. abandoned mines), body of water and facilities, and other land uses.

transit lines within ¼ mile from each household's residence. We created buffers for household residence and measured the aggregated level of service by transit stops located within these buffers to capture the extent to which a household has the opportunity to use transit as an alternative mode of transportation. Summary statistics for our variables are presented in **Table 2-1**.

Table 2-1 Summary Statistics (N=3,511 Households)

Variable	Mean	Std. dev.	Min	Max
Dependent variables				
Log of daily household VMT	2.98	1.85	-2.20	6.14
Log of daily household work VMT	-0.21	2.58	-2.20	5.48
Log of daily household non-work VMT	2.45	2.10	-2.20	5.87
Log of vehicle fuel efficiency (miles per gallon)	3.33	0.28	2.60	4.72
Gasoline price				
Price of gasoline on week of travel (in $ / gallon)	5.68	0.33	5.17	6.10
Household characteristics				
Midpoint of annual household income (in $1,000)	62.40	32.22	2.50	100.00
Household Annual Income is more than $100,000	0.28	0.45	0	1
Household size	2.85	1.46	1	13
Number of workers	1.17	0.95	0	5
Number of children under 16	0.37	0.75	0	5
Number of persons between 16 and 24	0.26	0.57	0	4
Vehicles per licensed driver	1.19	0.50	0.25	7.00
Respondent characteristics				
Age between 16 and 29	0.05	0.22	0	1
Age between 30 and 44	0.21	0.41	0	1
Age between 45 and 64	0.47	0.50	0	1
Age 65 and up	0.27	0.44	0	1
Education: less than High School	0.07	0.26	0	1
Education: High School degree	0.18	0.39	0	1
Education: some college	0.32	0.47	0	1
Education: Bachelor's degree	0.24	0.43	0	1
Graduate or professional degree	0.18	0.38	0	1
White	0.71	0.45	0	1
Hispanic	0.22	0.41	0	1
Black	0.05	0.22	0	1
Asian	0.07	0.26	0	1
Other ethnicity	0.02	0.14	0	1
Residential Urban Form Characteristics[a]				
Population density (10,000 person/square mile)[c]	1.27	2.14	0.00	45.34
Housing units per square mile (thousands)	3.34	3.43	0.05	30.00
Workers per square mile (census tract; thousands)	1.91	1.69	0.03	5.00
Percentage of renter occupied housing	0.32	0.27	0.00	0.95
Land use entropy (diversity index)	0.35	0.16	0.00	0.82
Transit lines within 0.25 mile[b] (/100)	0.03	0.14	0.00	5.13
Inverse distance to nearest sub-center (in 1/miles)[b]	0.27	0.64	0.01	33.46
Day of Travel				

Monday	0.14	0.35	0	1
Tuesday – Thursday	0.45	0.50	0	1
Friday	0.14	0.35	0	1
Weekend (Saturday and Sunday)	0.27	0.44	0	1

Notes:
[a] Measured at the block group level unless stated otherwise.
[b] Results from spatial proximity analysis.
[c] Extremely low density neighborhoods are found in the Southern California deserts.

2.4 Methodology

Like Brownstone & Golob (2009) and Kim & Brownstone (2013), we account for residential self-selection effects by specifying a system of recursive simultaneous equations where all causal effects are directed at vehicle usage. However, our model specification differs from theirs in several ways. First, their endogenous variables are different from ours: they consider total annual miles driven by all household vehicles, total annual household fuel usage measured in gallons of gasoline equivalents per year, and housing units per square mile in Census block group; in contrast, our endogenous variables are the log of VMT by the household on the survey day, the log of the fuel efficiency of the most efficient household vehicle, and urban from. Second, we treat urban form as a latent variable, which is obtained using a measurement modeling approach. Third, we estimate separate models for work trips and for non-work trips to explore the price sensitivity of households for different trip purposes, which Brownstone & Golob (2009) and Kim & Brownstone (2013) did not do. Moreover, our model allows directly estimating the short-run elasticity of driving with respect to gasoline price via one of our structural equations.

Our generalized SEM model is therefore comprised of a structural component and a measurement component. The former models the relationship between our endogenous variables using simultaneous equations where the direction of causality is explicitly specified (Acock, 2013; Bollen, 1989; Kline, 2005). The measurement component accounts for how our latent construct (urban form) is measured by observed variables; it is estimated via confirmatory Factor

Analysis (CFA) jointly with the structural component using quasi-maximum likelihood.

For each trip purpose, our model can be written as follows (vectors and matrices are in bold; n is the sample size; and k is the number of household explanatory variables):

$$\begin{cases} \mathbf{v}^* = \alpha\mathbf{g} + \beta_{12}\mathbf{m} + \beta_{13}\mathbf{f} + \mathbf{\Gamma}_1\mathbf{X} + \boldsymbol{\varepsilon}_1, \\ \mathbf{m} = \beta_{23}\mathbf{f} + \mathbf{\Gamma}_2\mathbf{X} + \boldsymbol{\varepsilon}_2, \\ \mathbf{f} = \mathbf{\Gamma}_3\mathbf{X} + \boldsymbol{\varepsilon}_3, \\ \mathbf{u} = \mathbf{\Lambda}_\mathbf{u}\mathbf{f} + \boldsymbol{\varphi}, \end{cases} \tag{2}$$

with

$$\mathbf{v} = \begin{cases} \mathbf{v}^* & \text{if } \mathbf{v}^* > 0, \\ \mathbf{0} & \text{if } \mathbf{v}^* \leq 0, \end{cases} \tag{3}$$

where:

- **v** is an $n\times1$ vector of the logarithm of household VMT and **v*** is the associated latent tobit variable;
- **m** is an $n\times1$ vector of the logarithm of vehicle fuel efficiency (in mpg); for simplicity, vehicle fuel efficiency is represented by the most fuel efficient vehicle in the household;
- **f** is an $n\times1$ vector of urban form latent variables;
- **g** is an $n\times1$ vector of the logarithm of gasoline prices;
- **X** is an $n\times k$ matrix of k household explanatory variables;
- $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, and $\mathbf{\Gamma}_3$ are $n\times nk$ matrices. $\mathbf{\Gamma}_\mathrm{j} = \left(\gamma_{j1}\mathbf{I_n} \;...\; \gamma_{jk}\mathbf{I_n}\right)$ for $\mathrm{j}\in\{1,2,3\}$, with $\gamma_{ji} \in\Re$ the γ_{ji}s are unknown coefficients to estimate, and $\mathbf{I}_n$ is the $n\times n$ identity matrix; we imposed additional restrictions on the $\mathbf{\Gamma}_j$s to over-identify the model: like Brownstone and Golob

(2009), we removed non-significant household explanatory variables from the VMT and urban form equations. As a result, education variables are present only in the vehicle selection equation, where they serve as weak instruments for vehicle selection.

- **u** is a (7n)×1 vector of stacked variables for residential urban form, which includes population density, housing unit density, employment density, % of renter occupied housing, land use entropy, inverse distance to the nearest sub-center[4], and transit access;
- Λ_u is an (7n)×n matrix given by $\boldsymbol{\Lambda}_{\mathrm{u}} = \begin{pmatrix}\lambda_1 \mathbf{I_n} & \dots & \lambda_7 \mathbf{I_n}\end{pmatrix}$, with $\lambda_i \in \Re$ ($i \in \{1,\dots,7\}$);
- The ε_is for $i \in \{1,2,3\}$ and $\boldsymbol{\varphi}$ are respectively n×1 and 7n×1 error vectors; and
- α, β_{12}, , β_{13}, β_{23}, γ_{ji} ($j \in \{1,2,3\}$, $i \in \{1,\dots,k\}$, in $\boldsymbol{\Gamma}_j$), and λ_i ($i \in \{1,\dots,7\}$ in Λ_u) are unknown model parameters to estimate; like SEM, generalized SEM minimizes the difference between the sample covariance and the covariance predicted by the model (Bollen, 1989).

2.4.1 Structural Component

The first three equations in (2) and equation (3) represent the structural component of our model, whose structure reflects causal paths (see Figure 2-2). In this component, **v** (and **v***), **m**, and **f** are vectors of endogenous variables, **X** is a matrix of exogenous household characteristics, and α, β_{12}, , β_{13}, β_{23}, γ_{ji} ($j \in \{1,2,3\}$, $i \in \{1,\dots,k\}$, in $\boldsymbol{\Gamma}_j$) are structural parameters.

In the third equation of (2), residential urban form (**f**), which is an endogenous latent construct, is assumed to be selected by households based on their characteristics prior making travel decisions. In the second equation, the log of vehicle fuel efficiency (**m**) is explained by

[4] The distance to sub-centers accounts for the polycentric nature of Southern California's urban structure, which has multiple employment centers. For detailed explanations of how sub-centers are determined, see Lee (2007) who analyzed sub-centers in Los Angeles and other major metropolitan areas.

residential urban form and by household characteristics. In the first equation, the log of vehicle miles traveled (**v**) is explained by the log of gas prices, the log of vehicle efficiency, residential urban form, and household characteristics.

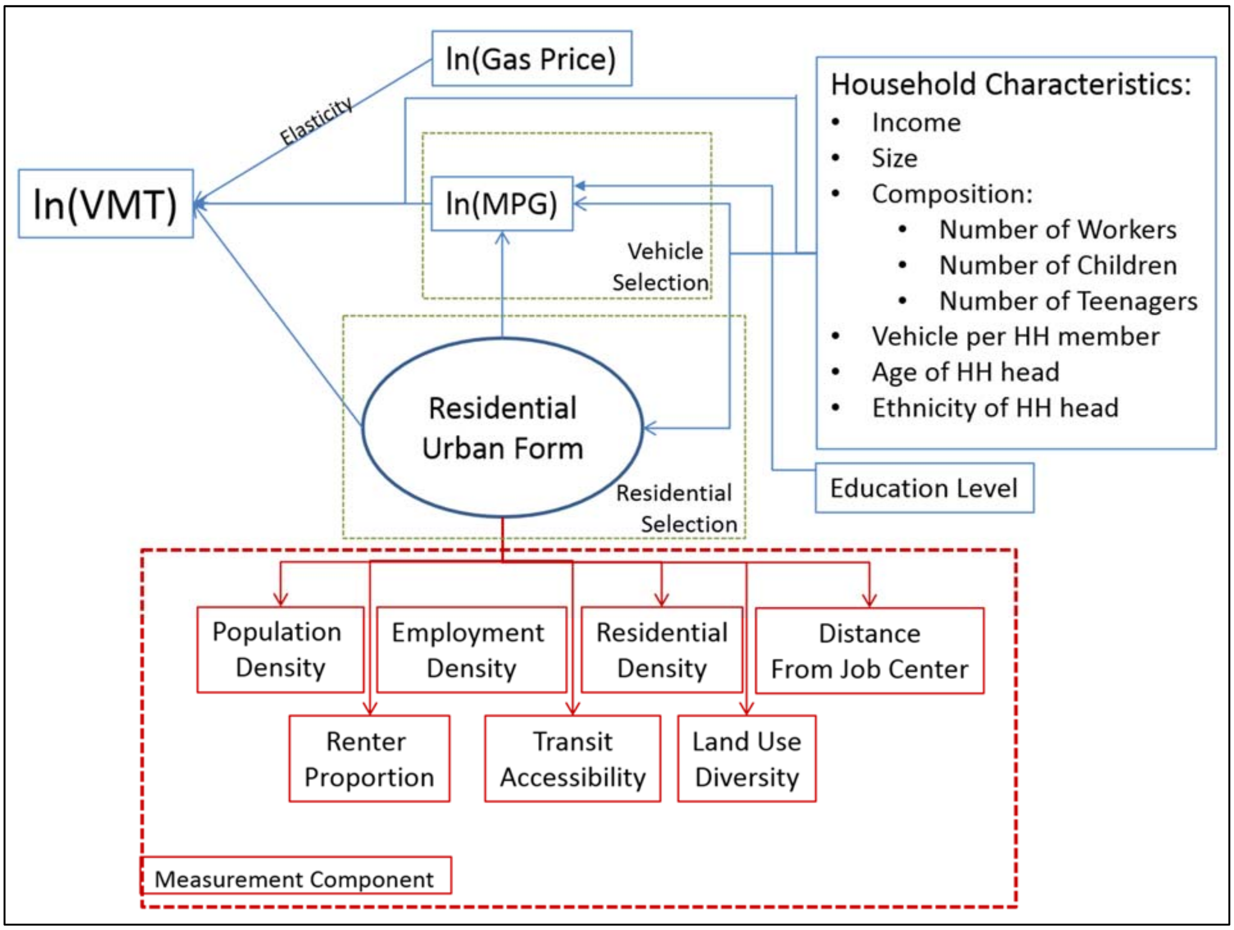

Figure 2-2 Model Structure

We therefore have a recursive system with all causality paths directed at vehicle usage; since we also assume the ε_is to be uncorrelated, our model is guaranteed to be identifiable (Bollen, 1989, pp. 95–98; Kline, 2005, pp. 105–107). In SEM, identification requires at least as many observations as free model parameters (df $\geq$ 0) and every latent variable/construct must be assigned a scale (Bollen, 1989, pp. 103–104; Kline, 2005, p. 105).

2.4.2 Measurement Component for Urban Form

The fourth equation is the measurement component of our model, which defines residential urban form as a latent construct that depends on population density, housing unit density, employment density, % of renter occupied housing, inverse distance to the nearest sub-center, land use entropy and transit access. These variables are stored in the vector **u**. Moreover, Λ_u is a matrix of measurement coefficients obtained by confirmatory factor analysis (CFA).

The treatment of residential urban form as a latent construct merits explanation. Since home based trips originate from residences, the surrounding land use patterns capture the configuration of potential activities that could influence travel behavior in general and driving in particular. Although this relationship is abstract, has no natural scale, and is partly unobservable, it can be characterized by variables such as land use entropy, density, transit access and distance from employment centers so it can be modeled as a latent variable or construct (Bollen, 1989, p. 180). As noted by Bollen (1989) abstract constructs have long been a part of many disciplines including quantum physics, biology, medicine, psychology, sociology, and thermodynamics (e.g., temperature is a latent variable measured by the height of a column of mercury). Conceptualizing urban form as a construct has been under-explored in the travel behavior literature (Golob, 2003; Van Acker *et al.*, 2007; van Wee, 2009) where urban form is often proxied by density.

Based on findings from our literature review, we started with a simple CFA measurement model with correlations of measurement errors (φ) restricted to zero. Following the SEM literature (Bollen, 1989, pp. 233–235, 296–305; Kline, 2005, pp. 176–191), we then modified our starting model after reviewing its fit to allow for theoretically justifiable correlated measurement errors for population density, residential density, land use entropy (diversity), employment density, proportion of renter-occupied housing units, transit access and distance from employment sub-centers, to reflect that places with high population density tend to have a more diverse land use mix, employment density and transit access. For this model to be identified, we scaled λ_1 (the coefficient of population density) to 1, which scales urban form to population density (in deviation scores), after correcting for correlations of population density with other variables reflecting urban form.

2.4.3 Model Interpretation

To discuss our results, we rely on elasticities whenever appropriate. For the first equation in (2), we rely on the expression of the elasticity for a tobit model (Wooldridge, 2002, p. 522):

$$\frac{\partial(\mathbf{v} \mid \mathbf{X}, \mathbf{v} > 0)}{\partial x_j} = \beta_j \left\{1 - \boldsymbol{\lambda}(\mathbf{w})\left[\mathbf{w} + \boldsymbol{\lambda}(\mathbf{w})\right]\right\}, \tag{4}$$

where $\mathbf{v}$ is the vector of log(household VMT); x_j is a log-transformed right hand-side variable; β_j is the model parameter of x_j; $\mathbf{w} = (\alpha\mathbf{g} + \beta_{12}\mathbf{m} + \beta_{13}\mathbf{f} + \boldsymbol{\Gamma}_1\mathbf{X}) / \sigma$, where σ is the standard deviation of $\boldsymbol{\varepsilon}_1$; and $\boldsymbol{\lambda}(\mathbf{w}) = \phi(\mathbf{w}) / \Phi(\mathbf{w})$ is a vector of inverse Mills ratios. This expression is readily adapted when x_j is not a log transformed variable.

2.5 Results

We estimated the model described above for total household trips, work trips, and non-work trips using quasi-maximum likelihood, the approach Stata 13.1 resorts to when the variance covariance matrix of the estimators is calculated using the Huber-White-Sandwich estimator. We chose that option because it relaxes the assumption that errors are identically and normally distributed, and requires only that errors be independently distributed. Results are presented in Tables 2-3 (A)-(B); they are discussed after addressing model fit (**Table 2-2**).

Table 2-2 Fit Statistics for the Urban Form Measurement Model

	χ^2 [df] (p-value)[a]	χ^2/df[a]	CFI[b]	TLI[c]	RMSEA[d]
Critical values	p> 0.05	< 5	> 0.95	> 0.95	< 0.05
Urban Form Measurement Model					
Model without correlated errors	862.74 [14] (<0.001)	61.62	0.85	0.77	0.13
Model with correlated errors	14.41 [6] (0.03)	2.40	1.00	0.99	0.02

Notes:

a. The chi-square (χ^2) statistic or chi-square goodness of fit, measures whether the observed covariance matrix is similar to the covariance matrix predicted by the model: if it is not significant, the model is regarded as acceptable. With small samples, the χ^2 statistic lacks power so some researchers have proposed the relative chi-square (χ^2 divided by the number of degrees of freedom, denoted here by χ^2/df).

b. The Comparative Fit Index (CFI) is an incremental fit index, which has been shown to have good power and robustness (Iacobucci, 2010, p. 97). It ranges from zero (worst fit) to one (best fit), where values above 0.97 indicate good fit and values between 0.95 and 0.97 denote acceptable fit (Schermelleh-Engel et al., 2003, p. 42).

c. The Tucker-Lewis Index (TLI) compares χ2/df of the proposed model to χ2/df for a null model; it is interpreted like the CFI (Schermelleh-Engel et al., 2003, p. 41).

d. The root mean square error of approximation (RMSEA) estimates the amount of error of approximation per model degree of freedom while taking sample size into account. A RMSEA value under 0.03 suggests excellent fit, and a value between 0.05 and 0.03 is considered good (Kline, 2005, p. 139).

2.5.1 Model fit

Model fit in SEM refers to the ability of a model to reproduce the observed variance-covariance matrix. A number of fit statistics have been developed for SEM but they typically do not carry over to Generalized SEM (GSEM) because common fit statistics depend on the assumptions that observed endogenous, observed exogenous, and latent endogenous variables are jointly normally distributed, which clearly does not hold for binary or censored variables. We therefore report only common fit statistics for our residential urban form measurement model (see **Table 2-2**).

Table 2-2 show five common fit statistics. The chi-square (χ^2) statistic measures whether the observed covariance matrix matches the model covariance matrix: if the χ^2 statistic is not significant, fit is deemed acceptable. Unfortunately, with small samples, the χ^2 statistic lacks power, and in large samples, it tends to over reject (Iacobucci, 2010, pp. 91, 95; Schermelleh-Engel *et al.*, 2003, pp. 31–33). To remedy these problems, some researchers have proposed the relative χ^2, i.e., χ^2 divided by the number of degrees of freedom (denoted by χ^2/df in Table 2).

A notable disadvantage of the χ^2 statistic is that it decreases with additional model parameters, which tends to favor models with many parameters (Schermelleh-Engel *et al.*, 2003, p. 33). To avoid this shortcoming, incremental fit indices, which are not sensitive to sample size and penalize for over-parameterization, have been developed. We report here the Comparative Fit Index (CFI), which has been shown to have good power and robustness (Iacobucci, 2010, p. 97). The CFI ranges from 0 (worst fit) to 1 (best fit), where values ≥0.97 indicate good fit and values between 0.95 and 0.97 denote acceptable fit (Schermelleh-Engel *et al.*, 2003, p. 42). We also report the popular Tucker-Lewis Index (TLI), which compares χ^2/df of the proposed model with χ^2/df for a null model. The TLI is interpreted like the CFI (Schermelleh-Engel *et al.*, 2003, p. 41).

Finally, we report the root mean square error of approximation (RMSEA), which estimates the amount of error of approximation per model degree of freedom while taking sample size into account. A RMSEA value of zero implies exact fit, a value under 0.03 suggests excellent fit, and a value between 0.05 and 0.03 is considered good (Kline, 2005, p. 139).

Table 2-2 shows that our residential urban form measurement model with correlated errors performs well, compared to our starting residential urban form measurement model with correlated errors, because it satisfies the criteria of all five shown fit indices.

2.5.2 Results Interpretation

SEM decomposes the mediating effect of residential selection on travel behavior by estimating direct, indirect, and total effects of endogenous and exogenous variables. Here, direct effects refer to how household characteristics directly influence VMT, and indirect effects, which can be interpreted as sorting (self-selection) effects, capture how they influence VMT through residential choice (Golob, 2003). Total effects are the sum of direct and indirect effects.

Before starting our overview of direct effects, we need to mention that all seven urban form variables considered for the latent urban form variable have highly significant coefficients (<0.001) and are positively correlated with the urban form latent variable (except for entropy), so higher densities (population, employment and residential) with more renters and better transit accessibility all contribute to a higher value of the urban form latent variable (coefficients are omitted for brevity). To better link our discussion with **Tables 2-3 (A)-(B)**, the value of the estimated coefficient of a variable discussed is provided in parentheses with its significance level.

Table 2-3 (A) GSEM Structural Model Coefficients / Direct Effects

Direct Effects	Household VMT			Vehicle Selection	Residential Selection
	All Trips	**Non-Work trips**	**Work trips**		
Exogenous ↓ | Endogenous →	Log(total VMT)	Log(non-work VMT)	Log(work VMT)	Log-MPG	Res. Urban Form
Residential Urban Form	-0.140***	-0.142***	-0.051*	0.009**	- -
Log(Vehicle fuel efficiency)	0.105	0.179	0.007	- -	- -
Log(Price of gasoline)	-0.186**	-0.265***	0.006	- -	- -
Household characteristics					
Midpoint of annual HH income	0.012***	0.011***	0.000	0.000	-0.011***
Annual Income is > $100,000	-0.274***	-0.220**	0.011	0.011	0.060
Household size	-0.003	0.093**	0.004	0.004	-0.183***
Number of workers	0.511***	0.242***	1.203***	- -	- -
Number of children under 16	0.073	0.070	0.066	- -	-0.045
Number of persons between 16 and 24	0.191***	0.356***	-0.036	- -	- -
Vehicles per licensed driver	0.115	0.114	-0.055	- -	-0.438***
Respondent characteristics					
Age: 16 to 29	-0.085	-0.027	-0.132	0.044**	0.797***
Age: 30 to 44	-0.032	-0.214**	0.021	-0.003	0.415***
Age: 65 and up	0.007	0.274***	-0.395***	0.013	-0.499***
Hispanic	-0.116	-0.334***	0.259**	0.005	0.526***
African American	0.018	-0.149	0.112	-0.022	0.590***
Asian	0.136	-0.050	0.327**	0.067***	0.045
Other ethnicity	0.151	0.341	-0.262	0.006	0.134
Education: High School degree	- -	- -	- -	0.018	- -
Education: some college	- -	- -	- -	0.013	- -
Education: Bachelor's degree	- -	- -	- -	0.055***	- -
Graduate or professional degree	- -	- -	- -	0.109***	- -
Day of Travel:					
Monday	-0.039	-0.102	-0.166	- -	- -
Friday	0.015	0.099	-0.205*	- -	- -
Weekend	-0.504***	0.023	-1.885***	- -	- -

Table 2-3 (B) Indirect and Total Effects

	Log(total VMT)		Log(non-work VMT)		Log(work VMT)	
	Indirect Effects	Total Effects	Indirect Effects	Total Effects	Indirect Effects	Total Effects
Residential Urban Form	0.001	-0.139***	0.001	-0.140***	0.000	-0.051**
Log(Vehicle fuel efficiency)	- -	0.105	- -	0.179	- -	0.007
Log(Price of gasoline)	- -	-0.186*	- -	-0.265***	- -	0.006
Household characteristics						
Midpoint of annual household income	0.002	0.013***	0.001	0.012***	0.000	0.000
Annual Income > $100,000	-0.007	-0.281***	0.006	-0.214**	0.002	0.013
Household size	0.026	0.023	0.022	0.118***	0.009	0.013
Number of workers	0.000	0.511***	0.000	0.242***	0.000	1.203***
Number of children under 16	0.006	0.079	0.002	0.073	0.001	0.067
Number of persons between 16 and 24	0.000	0.191**	0.000	0.356***	0.000	-0.036
Vehicles per licensed driver	0.061	0.176***	0.051	0.172***	0.021	-0.034
Respondent characteristics						
Age: 16 to 29	-0.107	-0.192	-0.068	-0.107	-0.032	-0.164
Age: 30 to 44	-0.058	-0.090	-0.050	-0.270***	-0.020	0.001
Age: 65 and up	0.071	0.078	0.047	0.325***	0.018	-0.377***
Hispanic	-0.073	-0.189**	-0.061	-0.403***	-0.025	0.234***
African American	-0.085	-0.067	-0.102	-0.263**	-0.040	0.072
Asian	0.001	0.137	-0.005	-0.059	-0.007	0.320***
Other ethnicity	-0.018	0.133	-0.015	0.324*	-0.006	-0.268
Education: High School degree	0.002	0.002	0.002	0.002	0.000	0.000
Education: some college	0.001	0.001	0.001	0.001	0.000	0.000
Education: Bachelor's degree	0.006	0.006	0.010	0.008	0.000	0.000
Graduate or professional degree	0.011	0.011	0.022	0.018	0.001	0.001
Day of Travel						
Monday	- -	-0.039		-0.124	- -	-0.166
Friday	- -	0.015	- -	0.090	- -	-0.205**
Weekend	- -	-0.504***	- -	-0.005	- -	-1.885***

Note: * p-value≤0.1, ** p≤0.05, *** p≤0.01

2.5.2.1 Direct effects

Let us first discuss coefficient estimates for the first equation in (2), which explains VMT. One of our main results from Table 3A is that the elasticity of VMT with respect to gas prices (based on Equation (4)) equals -0.15 (p-value=0.01) for all trips, -0.18 for non-work trips (p-value=0.01), but it is not significantly different from 0 for work trips. Overall, the value of the elasticity of total VMT with respect to gas price is consistent with previous studies (e.g., see Small & Dender, 2007, p. 22).

Vehicle fuel efficiency does not seem to affect VMT for work trips in the short run. However, urban form does; its effect on VMT is relatively small but it depends on trip-purpose as non-work trips (-0.142***) are slightly more sensitive to urban form than total trips (-0.140***) and clearly more than work trips (-0.051*). Because population density is negatively correlated with distance to job centers, a higher population density implies a closer proximity to job centers and thus work trips require less driving.

Annual household income appears to increase total VMT (0.012***) and non-work VMT (0.011*) only mildly, but it does not seem to affect work trips VMT. We note, however, that households in the highest income category tend to drive less, which is somewhat surprising.

Household composition plays different roles for different trip purposes. Households with more workers and young adults drive more overall (0.511*** and 0.191***) and the former have substantially more non-work trip miles (1.203***).

Age groups are also predictive of travel behavior. Compared to households in the 45-65 age group, households older than 65 tend to drive more for non- work trips (0.274***) and less for work trips (-.0.395***) likely because they have reached retirement age. Conversely, households in the 30-44 age group tends to drive less for non-work trips (-0.213**), but their work VMT is not significantly different from the work VMT of the 45-65 baseline age group.

Ethnicity also appears to play a role in explaining VMT. Compared to a White household, a Hispanic household with similar characteristics tends to drive less for non-work trips (-0.334***), but more to reach work (0.259**). Moreover, Asian households tend to drive more to reach work (0.327**), but their non-work VMT are statistically indistinguishable from the non-work VMT of comparable White households. There are also no statistically significant differences in VMT between African American and White households.

Relatively few variables are statistically significant in the vehicle fuel efficiency equation (second from last column in Table 2-3 (A)). First, urban forms with a larger index (i.e., denser) foster the selection of more fuel efficient vehicles, but only mildly. Second, we see that younger adults (0.044**) and Asian households (0.067***) are more likely to choose more fuel efficient vehicles, while African American, Hispanic, and White households have statistically similar vehicle fuel efficiency preferences. Importantly, results also show that education status is important for understanding preferences for more fuel efficient vehicles as households with more education (i.e., BS/BA or higher) are more likely to own vehicles that are more fuel efficient.

A number of variables impact how households select the urban form around their residence (last column of Table 3A). First, we see that households with higher incomes tend to choose neighborhood with a lower urban form index (-0.011***), which is not surprising because higher income neighborhoods tend to have large lot houses. Likewise, larger households prefer neighborhoods with a lower urban form index (-0.183***), and so do households with a vehicle to licensed drivers ratio (-0.438***). Age seems especially important as younger adults 0.797*** for the 16 to 29 group and 0.415***) prefer urban forms with a higher index that the 45 to 64 baseline group, whereas the reverse holds for older households (-0.499***). Ethnicity also matters as Hispanic (0.526***) and African American (0.590***) households tend to live in areas with a larger urban form index, while Asian and White households seem to have similar residential urban form preferences.

2.5.2.2 Indirect and Total Effects

Table 2-3 (B) reports indirect and total effects of socio-economic and demographic variables on VMT, vehicle fuel efficiency, and urban form selection. Indirect effects refer to how socio-economic variables affect driving through residential self-selection and vehicle fuel efficiency selection.

The most important message from this analysis is that the indirect effects of residential urban form on VMT through vehicle efficiency choices are not statistically significant. A comparison of the magnitude of direct and indirect effects shows that while self-selection may exist, its size has at best a small impact on total effects.

2.6 Conclusions

This paper presents a generalized SEM model that jointly explain residential urban form, vehicle efficiency choice, and vehicle use while accounting for residential self-selection and endogeneity of vehicle preferences in order to explain vehicle miles traveled (VMT) for both work and non-work trips. Residential urban form is treated as a latent construct that reflects observed variables such as population density, land use diversity and distance to employment centers.

First, our results show that trip purpose matters: in the short run, households drive 0.15% less for all trips and 0.18% less for non-work trips when gas prices increase by 1%, while work trips are not responsive to gasoline price changes. Second, the direct effect of residential urban form on driving is statistically significant for total and non-work VMT, but it has no impact on work trips. Third, the indirect effects of residential urban form on VMT through vehicle efficiency choices are not statistically significant.

Our research is not without limitations. In the 2009 NHTS data we analyzed, trip miles were calculated based on self-reported origin and destinations. Ideally, GPS devices should be deployed to accurately measure actual VMT.

Future work could consider intra-household vehicle substitution behavior where household with multiple vehicles may choose to drive different vehicles in respond to price changes. A natural follow up would be to explore more in depth how household adjust their travel behavior in response to increase in gasoline prices (do they organize their trips differently, forego travel altogether, or do they use other modes?). Another area of improvement would be to refine the residential-selection component by exploring how households choose neighborhoods based on amenities such as local public school quality as well as physical features of the neighborhood.

Finally, the next iteration of this paper will explore the modeling value of representing urban form by a latent variable by contrasting current models with simplified models where urban form is replaced by population density.

2.7 References

Acock, A. C. (2013). *Discovering Structural Equation Modeling Using Stata*. Stata Press.

Bagley, M. N., & Mokhtarian, P. L. (2002). The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *Annals of Regional Science*, *36*(2), 279–297.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, (May). Retrieved from http://www.jstor.org/stable/25098858

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. doi:10.1016/j.trb.2005.12.005

Boarnet, M. (2011). A Broader Context for Land Use and Travel Behavior, and a Research Agenda. *Journal of the American Planning Association*, *77*(3), 197–213. doi:10.1080/01944363.2011.593483

Boarnet, M., & Crane, R. (2001). The influence of land use on travel behavior : specification and estimation strategies. *Transportation Research Part A: Policy and Practice*, *35*(9), 823–845.

Boarnet, M., Houston, D., Ferguson, G., & Spears, S. (2011). Land use and vehicle miles of travel in the climate change debate: Getting smarter than your average bear. In Y.-H. Hong & G. K. Ingram (Eds.), *Climate Change and Land Policies* (pp. 151–187). Puritan Press Inc. Retrieved from https://www.lincolninst.edu/pubs/2038_Land-Use-and-Vehicle-Miles-of-Travel-in-the-Climate-Change-Debate

Bollen, K. A. (1989). *Structural Equations With Latent Variables*. Wiley.

Brownstone, D. (2008). Key relationships between the built environment and VMT. *Transportation Research Board*, *87*(August), 466–478. Retrieved from http://onlinepubs.trb.org/Onlinepubs/sr/sr298brownstone.pdf

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, *65*(1), 91–98. doi:10.1016/j.jue.2008.09.002

Cao, X. (Jason), Mokhtarian, P. L., & Handy, S. L. (2009a). *Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings*. *Transport Reviews* (Vol. 29, pp. 359–395). doi:10.1080/01441640802539195

Cao, X. (Jason), Mokhtarian, P. L., & Handy, S. L. (2009b). The relationship between the built environment and nonwork travel: A case study of Northern California. *Transportation Research Part A: Policy and Practice*, *43*(5), 548–559. doi:10.1016/j.tra.2009.02.001

Cao, X. (Jason), Mokhtarian, P. L., & Handy, S. L. (2010). Neighborhood Design and the Accessibility of the Elderly: An Empirical Analysis in Northern California. *International Journal of Sustainable Transportation*, *4*(6), 347–371. doi:10.1080/15568310903145212

Cervero, R., & Murakami, J. (2010). Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environment and Planning A*, *42*(2), 400–418. doi:10.1068/a4236

Chao, L., & Qing, S. (2011). An empirical analysis of the influence of urban form on household travel and energy consumption. *Computers, Environment and Urban Systems*, *35*(5), 347–357. doi:10.1016/j.compenvurbsys.2011.05.006

Congressional Budget Office. (2008). *Effects of gasoline prices on driving behavior and vehicle markets*. The Congress of The United States. Retrieved from http://www.cbo.gov/sites/default/files/01-14-gasolineprices.pdf

Crane, R. (2000). The Influence of Urban Form on Travel: An Interpretive Review. *Journal of Planning Literature*, *15*(1), 3–23. doi:10.1177/08854120022092890

De Abreu e Silva, J., Golob, T. F., & Goulias, K. G. (2007). Effects of Land Use Characteristics on Residence and Employment Location and Travel Behavior of Urban Adult Workers. *Transportation Research Record: Journal of the Transportation Research Board*, *1977*(-1), 121–131.

Duranton, G., & Turner, M. a. M. (2012). Urban Growth and Transportation. *The Review of Economic Studies*, *79*(4), 1407–1440. doi:10.1093/restud/rds010

Ewing, R., & Cervero, R. (2001). Travel and the Built Environment: A Synthesis, *1780*(01), 87–114. Retrieved from http://trb.metapress.com/content/A1W1712RW1225372

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, *76*(3), 265–294. doi:10.1080/01944361003766766

Fang, H. A. (2008). A discrete–continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B: Methodological*, *42*(9), 736–758. doi:10.1016/j.trb.2008.01.004

Gillingham, K. (2013). Identifying the elasticity of driving: Evidence from a gasoline price shock in California. *Regional Science and Urban Economics*, *47*, 13–24. doi:10.1016/j.regsciurbeco.2013.08.004

Golob, T. F. (2003). Structural equation modeling for travel behavior research, *37*, 1–25.

Hong, J., Shen, Q., & Zhang, L. (2013). How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. *Transportation*, *41*(3), 419–440. doi:10.1007/s11116-013-9462-9

Houston, D., Boarnet, M., Ferguson, G., & Spears, S. (2014). Can Compact Rail Transit Corridors Transform the Automobile City? Planning for More Sustainable Travel in Los Angeles. *Urban Studies*, 1–34.

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*(1), 90–98. doi:10.1016/j.jcps.2009.09.003

Kim, J., & Brownstone, D. (2013). The impact of residential density on vehicle usage and fuel consumption: Evidence from national samples. *Energy Economics*, *40*, 196–206. doi:10.1016/j.eneco.2013.06.012

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). Guilford.

Lee, B. (2007). "Edge" or "Edgeless" Cities? Urban Spatial Structure in U.S. Metropolitan Areas, 1980 to 2000. *Journal of Regional Science*, *47*(3), 479–515. doi:10.1111/j.1467-9787.2007.00517.x

Lovejoy, K., Sciara, G.-C., Salon, D., Handy, S. L., & Mokhtarian, P. L. (2013). Measuring the impacts of local land-use policies on vehicle miles of travel: The case of the first big-box store in Davis, California. *Journal of Transport and Land Use*, *6*(1), 25–39. doi:10.5198/jtlu.v6i1.336

Mokhtarian, P. L., & Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological*, *42*(3), 204–228. doi:10.1016/j.trb.2007.07.006

Puller, S. L., & Greening, L. a. (1999). Household adjustment to gasoline price change: an analysis using 9 years of US survey data. *Energy Economics*, *21*(1), 37–52. doi:10.1016/S0140-9883(98)00006-1

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models : Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Small, K., & Dender, K. Van. (2007). Fuel efficiency and motor vehicle travel: the declining rebound effect. *The Energy Journal*, *28*(1). doi:10.5547/ISSN0195-6574-EJ-Vol28-No1-2

Van Acker, V., Mokhtarian, P. L., & Witlox, F. (2014). Car availability explained by the structural relationships between lifestyles, residential location, and underlying residential and travel attitudes. *Transport Policy*, *35*, 88–99. doi:10.1016/j.tranpol.2014.05.006

Van Acker, V., Witlox, F., & Van Wee, B. (2007). The Effects of the Land Use System on Travel Behavior: A Structural Equation Modeling Approach. *Transportation Planning Technology*, *30*(4), 331–353. doi:10.1080/03081060701461675

Van Wee, B. (2009). Self-Selection: A Key to a Better Understanding of Location Choices, Travel Behaviour and Transport Externalities? *Transport Reviews*, *29*(3), 279–292.

3 Incentives to Promote Household Ownership of Alternative Fuel Vehicles: Effectiveness and Unintended Effects

California, where transportation accounts for over half of ozone precursors and particulate matter emissions, as well as nearly 40 percent of greenhouse gas emissions, has adopted the ambitious goal of reducing petroleum use in transportation by 50 percent by 2030. One of the proposed strategies to achieve this goal is to increase the share of alternative fuel vehicles (AFVs) on the road. In California, incentives to foster the addition of AFVs include the removal of occupancy requirements to access HOV lanes, parking privileges and workplace charging facilities for Hybrid Electric and Electric vehicles. Although popular, the effectiveness of these incentives is not well known. In addition, they may induce additional driving due to AFVs lower variable driving costs. To start answering these questions, this paper analyzes the 2012 California Household Travel Survey using a generalized structural equation model that accounts for residential self-selection, household demographic characteristics, and environmentalism. My findings suggest that, households who live closer to a freeway with HOV lanes, work closer to parking lots with AFV privileges, and are likely to support pro-environmental measures are more likely to own AFVs. However, these households are also likely to drive more than if they had conventional vehicles. Comparison between direct and total effects suggests that while expansion of HOV lanes (OR=1.004**) and parking incentives (OR=1.017***) will increase the adoption of alternative fuel vehicles, its utilization will also increase although the effect is relatively small. On average, a household moving 1 mile closer to an HOV lane would induce driving by an additional 0.12 miles per month. Working 10 miles closer to a parking lot with AFV privileges would induce a household to drive an additional 0.51 miles per month. These results imply that HOV and parking incentives comes with unintended consequences. Increasing awareness about environmental problems might yield higher policy pay-off. Keywords**:** Alternative fuel vehicles; Generalized Structural Equation Modeling; Incentives; HOV access.

3.1 Introduction

As in many other parts of the country, transportation accounts for over half of ozone precursors and particulate matter emissions, and for nearly 40 percent of greenhouse gas emissions in California (CA Governor Executive Order No. B-16, 2012). To reduce the local, regional and global air pollution from transportation, California has decided to reduce its transportation petroleum use by 50% by 2030. One key strategy to achieve this ambitious goal is to increase the share of alternative fuel vehicles (AFVs) on the road, an approach that has also been adopted in Europe, for example (Tscharaktschiew, 2015). In this study, only vehicles eligible for Clean Air Vehicle decals (which waive the occupancy requirement for accessing high occupancy vehicle (HOV) lanes on freeways) are classified as AFVs.

However, expanding the market share of AFVs is no trivial task. First, because they rely on new technologies and are typically produced is smaller numbers, AFVs tend to cost more than conventional vehicles (i.e., vehicles with internal combustion engines). Second, the refueling and maintenance infrastructure for some AFVs (i.e., electric or hydrogen vehicles) is currently lacking, which is a major impediment to their adoption. And third, potential buyers may have questions about the reliability, the durability, and the maintenance costs of some AFVs.

To overcome these obstacles, a few incentives have been put in place by federal, state, and local governments (Gallagher & Muehlegger, 2011; Diamond, 2008, 2009; Sallee, 2011; Sierzchula, Bakker, Maat, & van Wee, 2014; Beresteanu, & Li, 2011; Greene, Patterson, Singh, & Li, 2005; Shewmake & Jarvis, 2014; Shewmake, 2012; Kwon & Varaiya, 2008). In addition to various subsidies, the adoption of AFVs has been encouraged by operational (non-monetary) incentives. In California, the Clean Air Access program was designed to promote the adoption of AFVs by offering eligible vehicles access to High Occupancy Vehicle (HOV) lanes without occupancy requirement (Shewmake & Jarvis, 2014; Shewmake, 2012). When it was introduced in 2004, the rationale was that offering access to underutilized HOV lanes would be a zero-cost incentive that could boost AFV sales. However, there is no consensus about the overall effectiveness of California's HOV access program, which was originally introduced to encourage carpooling (Shewmake & Jarvis, 2014; Shewmake, 2012; Kwon & Varaiya, 2008). Parking incentives have also been offered to AFV drivers, either in the form of cheaper parking in preferred locations, or as free or discounted fuel (for electric and hydrogen vehicles).

One drawback of these incentives, however, is that they might induce additional driving since AFVs have lower variable costs, as shown for fuel efficient vehicles (e.g., see Puller, & Greening, 1999; Small & Van Dender, 2007). This would offset some of the environmental benefits of AFVs even if AFVs run mostly on renewable energy sources.

In this context, the goal of this paper is to analyze vehicle choice decisions at the household level to understand how they responded to incentives (HOV access and parking privileges). More specifically, using data from the 2012 California Household Travel Survey, I estimated a generalized structural equation model that explains the number of AFVs owned by households based on their socio-economic characteristics, their environmental views (proxied by voting data on environmental proposals), and incentives (HOV access and parking privileges) while endogenizing residential location (proxied by residential density) and vehicle utilization.

3.2 Literature Review

To select my model variables, motivate modeling framework and contextualize the results, I will review selected papers on vehicle ownership, political environmentalism, and incentives for AFVs.

3.2.1 Vehicle Ownership Modeling

Studies concerned with AFVs began to emerge after AFV incentive policies and programs were introduced in the early 1990s. At the Federal level, the 1990 Clean Air Act Amendments first allowed waiving the occupancy requirement of high-occupancy vehicle (HOV) lanes for low-emission and energy-efficient vehicles. The Transportation Equity Act for the 21st Century (TEA-21) expanded these measures by providing incentives for the purchase of low-emission and energy-efficient vehicles before the Safe, Accountable, Flexible, and Efficient Transportation Equity Act (SAFETEA-LU) broadened the set of HOV eligible vehicles.

Several papers relevant to my study are part of the rich empirical literature on vehicle choice. Recent papers (Bhat & Guo, 2007; Fang, 2008; and Salon, 2009) have proposed extensions of classical discrete choice models, by endogenizing household residential neighborhood characteristics as well as vehicle utilization.

In the first paper to control for residential self-selection in a transportation cross-sectional study, Bhat & Guo (2007) estimated a mixed-logit model on San Francisco Bay Area data to understand the impact of the built environment on travel behavior. They assumed that Traffic Analyzes Zone (TAZ) capture the characteristics of residential neighborhoods and that households select their vehicles based on make, model and fuel-efficiency. However, vehicle utilization was assumed to be exogenous, which desensitized vehicle use from policies. They found that density and other built environment attributes affect residential location as well as vehicle ownership. In addition, both household and built environment characteristics influence household vehicle ownership decisions. Income is the most important factor in residential sorting as low income households may choose (or are constrained to) high density neighborhoods that make low-cost commuting possible, impacting the number of cars they own. Finally, they reported that a well-specified model with extensive socio-demographic and neighborhood characteristics can adequately account for residential and vehicle choice endogeneities.

Fang (2008) was interested in household preference for fuel-efficient vehicles. She estimated her model on the California sub-sample of the 2001 National Household Travel Survey (NHTS) without additional land use or location data, which prevented her from considering residential location. She captured residential characteristic through residential density and regional indicators such as rail availability, location in a regional MSA (Los Angeles MSA, San Francisco MSA, etc.), and a rural/urban indicator. To account for residential self-selection and vehicle choice endogeneity, she developed a Bayesian Multivariate Ordered and Tobit model to jointly estimate the influence of residential characteristics on vehicle fuel efficiency and use. Her results suggest that residential density affects vehicle ownership choices; households who live in high density neighborhoods are less likely to own trucks. Increasing density is also associated with the likelihood of owning no vehicle. Moreover, higher density implies less driving, both for cars and trucks.

Salon (2009) estimated a multinomial logit model to jointly explain residential selection, vehicle ownership (total holding), and commuting mode on New York City data, with a focus on household modal-choice. Due to computational limitations, she assumed that work location is exogenous. Like Fang (2008) and Bhat & Guo (2007), she found that population density has a substantial impact on vehicle ownership; in her study, living farther from midtown Manhattan increases the utility of car ownership while higher density has an opposite effect.

These studies, however, did not consider the effectiveness of specific energy and environmental policies. Furthermore, as noted by Choo and Mokhtarian (2004), vehicle choice studies should consider attitudes, environmental beliefs, lifestyle, and/or personality characteristics if this information is available.

3.2.2 The Role of Environmentalism

Several studies have examined how political environmentalism may explain AFV ownership. In this paper, environmentalism broadly refers to the belief that the government has an important role to play in improving environmental quality. See Kahn (2007) for an overview of environmentalism and market behavior or Guber (2003) for a comprehensive discussion. Political environmentalism was either proxied by membership in an environmental organization or reflected in survey responses about environmental issues and the role of government.

Kahn (2007) estimated a negative binomial count regression model for various types of vehicles on a 2005 proprietary vehicle registration dataset for Los Angeles County. He found that, controlling for income, population size, population density and racial mix, a higher share of registered Green Party voters - a proxy for environmentalism - is positively associated with a higher demand for hybrids electric vehicles (HEVs) at the census tract level.

Studies of sales and market share analyses at the State-level (Sallee, 2011) and in Virginian counties (Diamond, 2008) found results similar to Kahn (2007). While primarily focused on the role of incentives, Gallagher & Muehlegger (2011) examined whether the adoption of HEVs correlates with environmental and energy security preferences. They proposed using state-level Sierra Club membership per capita as a proxy for environmentalism and relied on this approach to explain sales of HEVs per capita, without accounting for incentives. Their result suggests that an increase in Sierra Club membership increases the sales of HEVs.

Diamond (2008) primarily analyzed the impact of HOV incentives on HEV ownership in Virginia. His regression model controls for the share of Green Party votes (a proxy for environmentalism) in explaining county-level HEV market share. His results show that the number of Green Party votes is positively associated with the market share of HEVs.

Aggregate level studies, however, do not consider household heterogeneity. Although Kahn (2007) explicitly assumed that households tend to Tiebout-sort into "like minded" communities, it is not clear that census tracts spatially delineate communities. Moreover, using statistical inferences based on aggregate data to examine the effectiveness of AFV policies and incentives assumes that household's AFV ownership decision can be deduced from spatially aggregated unit where households reside (also known as an ecological fallacy). Furthermore, aggregate data typically do not allow controlling for residential location, which has been shown to affect travel behavior and vehicle ownership decisions (Fang, 2008; Salon, 2009; Dillon, Saphores, & Boarnet, 2015; Brownstone, 2008; Brownstone & Golob, 2009; Kim & Brownstone, 2013).

3.2.3 The Role of Incentives

Among incentives that can influence a household to purchase a HEV, I can distinguish between financial incentives that offset the purchase price of a vehicle, and operational incentives such as occupancy exemptions for accessing HOV lanes or parking privileges in urban areas. AFV incentives were designed to stimulate consumer demand and create a viable market for AFV manufacturers. This second-best government intervention can be justified on environmental and energy independence grounds. While skeptics may argue that manufacturers may capture an excessive share of these incentives, voters may favor these incentives for environmental reasons. Furthermore, the literature shows that consumers can benefit from incentives that promotes higher fuel efficiency. Consumers captured a significant share of tax benefits and of the cash-for-clunkers program (Diamond, 2009; Beresteanu, & Li, 2011).

Diamond (2009) analyzed a state-level panel that covers substantial variations in gasoline prices. Using a fixed-effect model, he found that gasoline prices crowd out incentives and that HOV access has a stronger effect than financial incentives. This finding is consistent with Beresteanu & Li (2011) who analyzed 2001 NHTS data and proprietary vehicle sales data.

Gallagher & Muehlegger (2011) used quarterly national vehicle sales to evaluate the relative efficacy of state sales tax waivers, income tax credits, and non-tax incentives (HOV access) on HEV market penetration. They reported that even though state sales tax waivers tend to be less generous than state income tax credits, the mean sales tax waiver (valued at $ 1,077 in 2011) is associated with three times the increase in sales of the mean income tax credit. Note, however, that only four states offered sales taxes waivers for AFVs. Gallagher & Muehlegger (2011) also found that the HOV access program is positively correlated with HEV sales in Virginia but little evidence that opening HOV lanes to HEVs has a positive impact on HEV sales in other states.

While not specifically focused on HEVs, Greene *et. al* (2005) analyzed 2,000 sales data to quantify the impact of 'feebates' (fees and rebates) on vehicle fuel economy. Feebates are a market-based alternative in which vehicles with fuel consumption rates above a "pivot point" are charged fees while vehicles below receive rebates. They concluded that the clear majority of fuel economy increase is due to the adoption of fuel economy technologies rather than shifts in sales.

3.3 Data

To understand how vehicle choice decisions at the household level respond to incentives (such as HOV lane access and parking privileges), I combined geocoded data from the 2012 California Household Travel Survey (CHTS) with incentive data (access to HOV lanes, parking privileges, and proximity to AFV refueling stations) and a measure of political environmentalism based on voting data. Summary statistics for my variables are presented in **Table 3-1**.

3.3.1 CHTS Data

The 2012 CHTS collected detailed socio-economic information, travel information on a pre-determined survey day, and vehicle information from 42,431 randomly selected households from all 58 counties in California. The 2012 CHTS was administered over a full year starting in January 2012. The geocoded raw data include the latitude and longitude of each household residential location and travel destinations. Out of 42,431 households randomly selected for the survey, only 37,744 households (~86.6%) completed the travel diary, and only 17,807 households (~ 42%) traveled to work or school on the survey date. After removing households without any vehicle, the sample size reduced to 14,477 households (~34%). Discarding records with missing information, my final sample has 12,030 households (~28% of recruited households.) From this rich dataset, I used information about household, their vehicles, and how much they drive.

Based on my literature review (Fang, 2008; Salon, 2009; Dillon, Saphores, & Boarnet, 2015; Brownstone, 2008; Brownstone & Golob, 2009; Kim & Brownstone, 2013), I used a wide range of socio-demographic variables characterizing households (income, household structure; size, number of children under 16, number of young adults 16 to 24, and number of workers) or household heads (age, ethnicity, and educational attainment).

Table 3-1 Summary statistics (N = 12,030 households)

Variable	Min	P25	Mean	P75	Max	SD
Endogenous Variables						
Count of Household AFVs	0	0	0.072	0	2	0.269
Work and School VMT	0	4.255	14.625	21.164	179.352	13.935
Residential density (10K person /sq. mile)	.00004[a]	0.203	0.697	0.909	11.363	0.727
Incentives & Environmentalism						
Political Environmentalism	0	0.403	0.511	0.612	1	0.151
Distance to HOV lane (miles)	0	1.206	33.231	38.616	327.542	54.816
Distance to AFV parking (miles)	0	0.097	5.045	1.466	172.24	17.317
Household Characteristics						
Midpoint of annual HH Income (in $1,000)	0	62.5	98.915	125	250	62.079
HH annual income is more than $250,0000	1	0	0.044	0	1	0.206
Household size	0	2	2.977	4	8	1.4
Number of children under 16	0	0	0.598	1	7	0.979
Number of persons between 16 and 24	1	0	0.343	1	5	0.659
Number of workers	0	1	1.725	2	6	0.739
Education: High School degree	0	0	0.137	0	1	0.344
Education: some college	0	0	0.296	1	1	0.456
Education: Bachelor's degree	0	0	0.284	1	1	0.451
Graduate or professional degree	0	0	0.223	0	1	0.416
Age: 18 to 31	0	0	0.070	0	1	0.255
Age: 32 to 47	0	0	0.284	1	1	0.451
Age: 67 and up	0	0	0.061	0	1	0.24
Hispanic	0	0	0.192	0	1	0.394
Black	0	0	0.029	0	1	0.167
Asian	0	0	0.065	0	1	0.246
Native American	0	0	0.042	0	1	0.201
Other ethnicity	0	0	0.003	0	1	0.059

[a] Extremely low density neighborhoods are found in the Californian High Deserts

Household income enters my model as two variables: I used the midpoint of each CHTS income interval and a binary variable for households with an annual income over $250,000 (the top income bracket is open ended), which make up 4.2% of my sample. Household size, the number of children under 16, the number of young adults between 16 and 24, and the number of workers are count variables. Age enters my model as categorical variables to capture generational effects on residential choice, travel behavior, and AFV ownership (the 48-67 age group is the baseline.) I followed a similar strategy for educational attainment with people with less than a high school education as the baseline. For ethnicity, White is the baseline category, with binary variables for the other ethnic groups (see Table 1 for details.)

Second, I used detailed vehicle information on make, model-year, and fuel type to determine eligibility for Clean Air Vehicle (CAV) decals that allow driving a vehicle with a single occupant in HOV lanes (see http://dmv.ca.gov/portal/dmv/detail/vr/decal for details). In this study, only vehicles that are eligible for HOV access decals are classified as AFVs. I then used this information to construct my dependent variable; which is the AFV ownership status for each household. Out of 12,030 sampled households, 6.95% (836 HHs) own at least one AFV. Overall, households in my sample own 868 AFVs: 86.87% are HEVs, 3.34% are PHEVs, 7.83% are EVs and 1.96% runs on Crude Natural Gas (CNGs).

Third, I accounted for vehicle utilization because previous studies have reported that vehicle use impacts vehicle choice decisions (Fang, 2008; Salon, 2009; Dillon, Saphores, & Boarnet, 2015; Brownstone, 2008; Brownstone & Golob, 2009; Kim & Brownstone, 2013). I focused on non-discretionary travel (such as travel to work or school) for two reasons. First, it allows us to ignore spatial and temporal variations in gas prices (remember that the 2012 CHTS was administered over one year and that gas prices in California can vary substantially at a given point in time; e.g., see https://www.gasbuddy.com/GasPrices/California) which would impact the number of discretionary miles driven but not non-discretionary miles since they are not responsive to short-term variations in gas prices (Dillon, Saphores, & Boarnet, 2015). Second, households who drive more for non-discretionary purposes are more likely to consider AFVs because these vehicles have lower operating costs. To calculate non-discretionary miles driven from the 2012 CHTS one-day travel diaries, I summed the distance of work and school trips on weekdays for each household, taking care not to double count when more than one household member traveled in the same vehicle. Visited places were then geocoded and trip distances were computed using the PostGIS 2.0's ST_Length function.

3.3.2 Incentives: HOV network and Alternative Fuel Stations

My model can only help us understand the impact of incentives that were not uniformly available to all CHTS respondents, so I did not include federal incentives in my models. The incentives I considered are HOV access with no minimum vehicle occupancy and parking privileges.

Households who could drive on freeways with HOV lanes to reach desired locations would benefit from an HOV exemption. I proxied this benefit by calculating the proximity of each household residence in my sample to the nearest freeway with HOV lanes using Geographical Information System (GIS) software. The location of HOV lanes (as of 2010, to ensure that the incentive precedes the effect) was obtained from the California Department of Transportations' GIS database (see http://www.dot.ca.gov/hq/tsip/gis/datalibrary/).

Another incentive is proximity from home or work to a parking facility that gives discounts to AFV drivers or that allocates space for recharging electric vehicles or PHEVs and CNG refueling stations. The U.S. Department of Energy Alternative Fuels Data Center (AFDC) (http://www.afdc.energy.gov/locator/stations/) provides detailed geospatial information on alternative refueling stations. I used the location information provided by the CHTS to calculate the shortest distance between these facilities and work location of households in my sample. A map showing the location of alternative refueling stations is shown in **Figure 3-1**. Likewise, I used 2010 data to test whether AFV ownership (the effect) is caused by the incentive (policy).

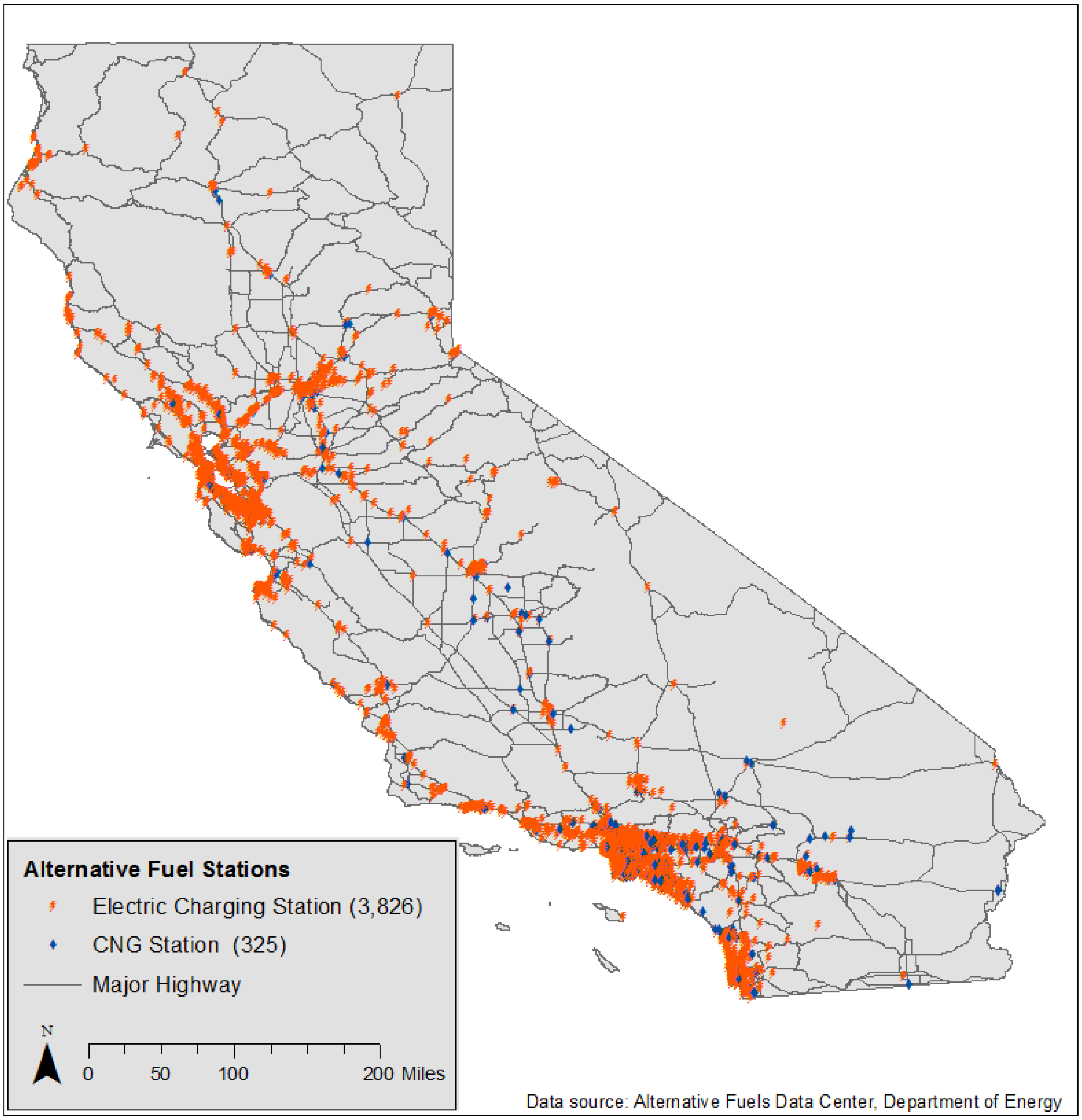

Figure 3-1 Alternative Vehicle Refueling Stations

3.3.3 Political Environmentalism

Following Guber (2003) and Kahn (2007), I hypothesized that AFV ownership may also depend on households' willingness to voluntarily reduce the external costs of their mobility, and more generally on their environmental views. Unfortunately, household-level information about environmental beliefs was not collected by the CHTS so I followed Kahn's (2007) strategy. I assumed that households tend to self-select into "like minded" communities as argued by Tiebout (Kahn & Matsusaka, 1997; Kahn, 2007) and used precinct-level voting data as a proxy for household's environmental views. More specifically, I went through the California Ballot Propositions on environmental issues available from the Berkeley Law School's Statewide Database (http://statewidedatabase.org), which stores precinct level voting data, and selected the most relevant propositions on the ballot slightly before or after the 2012 CHTS. Like Kahn & Matsusaka (1997), I used the share of votes in favor of positions that were endorsed by prominent pro-environment organizations, such as the Sierra Club, the Audubon Society, the Nature Conservancy, the Natural Resources Defense Council, the Environmental Defense Fund and the National Wildlife Federation.

Table 3-2 California Environmental Ballot Propositions, 2010 – 2014

Proposition (Year-Number) and Brief Description [a]	**Average (Standard deviation)**	**Normalized factor loadings** [b]
2010 – 21. Vehicle License Fee for Parks. Would have increased vehicle license fee by $18 a year to increase funding for State Parks. Endorsed position: Support	44.3% (13.82)	0.89
2010 – 23. Suspension of AB 32 (2010), the "Global Warming Act of 2006". Would have suspended implementation of comprehensive greenhouse gas reduction programs, including renewable energy and cleaner fuel requirements. Endorsed position: Opposition	62.4% (13.55)	0.92
2014 – 01. Water Bond (AB 1471) Authorizes obligation bonds for state water projects, including ecosystem protection and restoration. Endorsed position: Support	65.06% (11.1)	0.57

Notes: [a.] Endorsed position describes the position of prominent pro-environment organizations, such as the Sierra Club, the Audubon Society, the Nature Conservancy, the Natural Resources Defense Council, the Environmental Defense Fund and the National Wildlife Federation.
[b.] Normalized factor loadings are results from the Principal Component Analysis. The factor score for political environmentalism is the sum of the percentage of endorsed votes, weighted by the normalized factor loadings. This weighted sum is then normalized to be between 0 and 1. A higher value of political environmentalism indicates higher willingness to spend on environmental goods and support for greenhouse gas reduction programs. Cronbach's Alpha is 0.84, with a KMO statistics of 0.617 and a highly significant Bartlett's test ($p<0.0001$).

Table 3-2 provides brief a summary of the three propositions I selected and how they fared at the ballot box. The second column reports the average and standard deviation (in parentheses) of the proportion of endorsed votes. The last column reports normalized factor loadings (see next section.)

3.4 Modeling framework

To explain the household AFV ownership, I specified a recursive Generalized Structural Equation Model (GSEM) with a Logit link function that endogenizes residential self-selection and non-discretionary driving (for work and school purposes). Structural Equation Modeling (SEM) has been applied many times to models of vehicle use and ownership to capture the endogenous causal effects between vehicle ownership and use (Dillon, Saphores, & Boarnet, 2015; Brownstone, 2008; Kim & Brownstone, 2013; Brownstone & Golob, 2009). However, SEM can only handle continuous dependent variables so I had to resort to GSEM to handle my binomial response dependent variable.

My conceptual model is shown in **Figure 3-2**. Causal flows between two variables are represented by an arrow. A variable is endogenous when an arrow is directed towards it. Variables from which arrows only depart are exogenous. All causal paths are directed towards the household ownership of AFV which reflects the recursivity of my model.

I assumed that households first choose their residential neighborhood (characterized by its density) based on their exogenous socio-demographic characteristics, prior to traveling and selecting their vehicles. Then, the amount of non-discretionary driving for work and school is determined by incentives (parking and access to HOV lanes), household socio-economic characteristics, and residential density. Higher density neighborhoods are more likely to be found near employment centers and better transit access – land use features that may reduce the need for driving (Dillon, Saphores, & Boarnet, 2015). A recursive path model is appropriate here because non-discretionary driving is not sensitive to vehicle fuel economy, which is a proxy for the variable cost of driving– a quantity that households may want to minimize when considering the purchase of an AFV. Lastly, I assumed that political environmentalism, proxied by environmental voting outcomes, directly influences household vehicle choices (households who live in neighborhoods with higher concerns for the environment are more likely to own AFVs than conventional vehicles) along with residential density and non-discretionary driving for work and school.

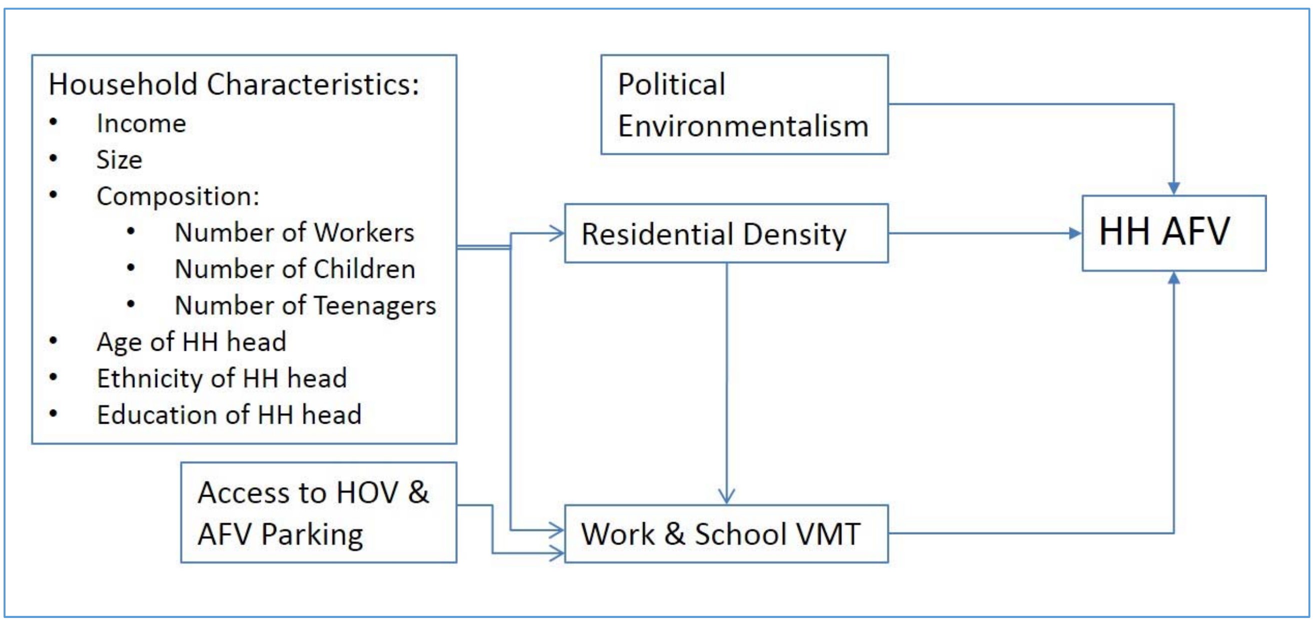

Figure 3-2 Conceptual Model Structure

Since my model does not include any latent variable measurement component, it only has a structural component, i.e., a simultaneous equation system. I used principal component analysis (PCA) to measure political environmentalism because I only have three indicators of political environmentalism, instead of a Confirmatory Factor Analysis (CFA) which would require more than three indicators to guarantee model over-identification (Kline, 2005, p. 172; Bollen, 1989, pp. 238-241).

3.4.1 Principal component analysis for Political Environmentalism

I hypothesize the existence of Political Environmentalism, a latent construct that captures political attitudes and beliefs about the environment. I used Principal component analysis (PCA) to estimate this latent construct that explains the variation in precinct-level voting outcomes for the three environmental propositions shown in **Table 3-2**. This led to a single factor (see below). To simplify its interpretation, I normalized this factor to be between 0 and 1, where 1 corresponds to high pro-environmental beliefs. Normalization is done by subtracting the factor score to the minimum and dividing by the range of factor score (maximum less the minimum).

I assessed the adequacy of my factor using a standard approach. First, I used Bartlett's test of sphericity to check for the appropriate level of inter-correlation between the voting outcomes analyzed. Inter-correlations must be sufficiently high to limit the number of factors, but not too high to avoid multicollinearity, which I detected using the Kaiser-Meyer-Olkin (KMO) statistic – a measure of sampling adequacy. I also used Cronbach's alpha to measure the reliability of my factor.

For a PCA model to work well, Bartlett's test should reject the null hypothesis that the correlation matrix is an identity matrix. The KMO statistic (which ranges between 0 and 1, with small values suggesting that the variables do not have enough in common) should be larger than 0.6 to be satisfactory. Finally, Cronbach's alpha (which has a maximum value of 1) has been suggested to be at least 0.7 (Kline, 2005, p. 59).

3.4.2 GSEM with a Logit link

Before specifying a structural model, I estimated 'reduced form' logit model to assess goodness of fit. I also used variance inflation factors to detect multicollinearity among my explanatory variables (none was found).

Overall, I have a recursive model with causality paths directed at AFV ownership; since I also assume that the error terms are uncorrelated, my model is guaranteed to be identified (Kline, 2005; Bollen, 1989). In estimating unknown model parameters, GSEM minimizes the difference between sample covariance and the covariance predicted by the model.

3.5 Results

I estimated my model using quasi-maximum likelihood, which is the approach used by Stata 14.2 when the covariance structure is obtained using the robust-Huber-White-Sandwich estimator. This option relaxes the assumption that errors are identically and normally distributed, and requires only the errors to be independently distributed.

In GSEM, as in SEM, model fit refers to the ability of a model to reproduce the observed variance-covariance matrix (Kline, 2005; Bollen, 1989). While several fit statistics have been developed for SEM, they are not valid for GSEM because those fit statistics assume that endogenous variables are jointly normally distributed, which is clearly not the case for a count variable. I therefore report only the common fit statistics for my principal component analysis.

3.5.1 Principal component analysis for Political Environmentalism

Results of the principal component analysis are presented in the last column of **Table 3-2**. I obtained a single factor for political environmentalism using the sum of the percentage of endorsed votes (item score) at the precinct level, weighted by the corresponding factor loading shown in the last column. This weighted sum was normalized to be between 0 and 1 to simplify interpretation. A higher value of political environmentalism indicates both a higher willingness to spend on environmental goods and support for greenhouse gas reduction programs.

My factor passed standard fitness tests. Bartlett's test of sphericity yielded a Chi-square statistic of 3,007 (df = 3) which is overwhelming evidence against the null-hypothesis ($p<0.0001$) that the voting outcomes analyzed are not correlated. My KMO measure of sampling adequacy gave a value of 0.619, which suggests that the voting outcomes considered have enough in common to warrant a PCA. Finally, Cronbach's alpha is 0.84, which suggests that my political environmentalism factor has good internal consistency (Kline, 2005, p. 59).

3.5.2 GSEM Results

GSEM decomposes the mediating effect of residential selection on AFV ownership by estimating direct, indirect and total effects of endogenous and exogenous variables. Direct effects refer to how household characteristics directly influence the level of AFV ownership, and indirect effects capture how they influence AFV ownership through other variables. Total effects are the sum of direct and indirect effects. **Table 3-3** presents my results, which include structural model coefficients (direct effects) as well as indirect and total effects.

3.5.2.1 Direct effects

Let us first discuss coefficient estimates which reflect the direct effects of other endogenous and exogenous variables on AFV ownership (Column I of **Table 3-3**).

One of my main results is that households who drives a lot for non-discretionary purposes, such as commuting to work and school, are more likely to own AFVs (OR = 1.009***). In contrast, households who reside in higher density neighborhoods are less likely to own AFVs (OR = -0.85***), which is not surprising since suburban households may favor AFVs compared to urban families because of their longer commute and the lower cost per mile driven of AFVs. As expected, households in neighborhoods with a higher level of political environmentalism are more likely to be AFV owners (OR=4.73***).

Exogenous household characteristics also play a role in AFV ownership decision. Notably, as their income increases households are more likely to own AFVs (OR=1.008***), although this does not hold for higher income households (households making more than $250,000 annually). These households are less likely to own AFVs (OR=0.72***), perhaps because they are less sensitive to costs or tend to decide on vehicles for attributes other than its fuel efficiency. Interestingly, household structure does not directly impact AFV ownership. However, educational attainment of the head of household is an important predictor of AFV ownership. Compared to households with less than a high school degree, households with a bachelor degree (OR=1.67**) and a graduate degree (OR=2.25***) are more likely to be AFV owners. Lastly, ethnicity also impacts AFV ownership: Hispanic (OR=0.65***) and Black (OR=0.51***) households are less likely to be AFV owners compared to otherwise similar White households. Several variables impact households non-discretionary (from home to work and school) commuting (Column II of **Table 3-3**).

As expected, households who reside in denser areas drive fewer miles for work and school (-0.265**). Households with better access to highways with HOV lanes (-0.004**) and who work in places with access to AFV parking privileges (-0.017***) tend to drive more non-discretionary work and school miles, although this effect is relatively small. These results should not be surprising as HOV lanes and parking privileges makes driving more attractive compared to alternative modes and thus incentivize vehicle utilization in general.

Table 3-3 Generalized SEM structural model coefficients

	Direct Effects			Indirect Effects	Total Effects
Column number	I	II	III	IV	V
Exogenous ↓ | Endogenous →	**HH AFV**	**Work & School VMT**	**Residential Density**	**HH AFV**	**HH AFV**
Work and School VMT	0.009***	--	--	--	0.009***
Residential Density	-0.165***	-0.265**	--	-0.002***	-0.167***
Political environmentalism	1.554***	--	--	--	1.554***
Distance to HOV lane	--	-0.004**	--	0.000	-0.004**
Distance to AFV parking	--	-0.017***	--	0.000	-0.017***
Household (HH) characteristics					
Midpoint of annual HH income	0.008***	0.006***	-0.001***	0.000***	0.008***
HH annual income is more than $250K	-0.325***	-0.182	0.047	-0.009***	-0.334***
Household size	-0.062	0.171	-0.036***	0.008	-0.055
Number of children under 16	0.001	-0.709***	-0.011	-0.005	-0.004
Number of persons 16 to 24	0.049	-0.353**	0.017	-0.006	0.043
Number of workers	-0.044	1.267***	-0.005	0.012	-0.032
Education: High School degree	-0.247	0.086	-0.186***	0.032	-0.215
Education: some college	0.178	0.404	-0.183***	0.034	0.212
Education: Bachelor's degree	0.515**	0.034	-0.090***	0.015**	0.531**
Graduate or professional degree	0.812***	-0.319	-0.060**	0.007***	0.820***
Age: 18 to 31	-0.396**	0.629**	0.227***	-0.032**	-0.429**
Age: 32 to 47	-0.034	0.391**	0.136***	-0.019	-0.053
Age: 68 and up	-0.147	-0.921***	-0.030	-0.003	-0.151
Hispanic	-0.430***	0.521**	0.256***	-0.038***	-0.468***
Black	-0.675***	0.384	0.479***	-0.077***	-0.751***
Asian	0.210*	0.415	0.365***	-0.057*	0.153*
Native American	-0.192	-0.232	0.021	-0.006	-0.197
Other Ethnicity	0.408	-1.186	0.112	-0.029	0.378

Notes: *, **, & *** denote p-values ≤0.1, p≤0.05, and p≤0.01 respectively. N = 12,030 HH, Log-likelihood, AIC, & BIC scores are -74,816.6, 149,765.1, and 150,268.9 respectively.

Among socio-economic variables, I first see that household structure plays an important role for non-discretionary travel. Households with more children under 16 (-0.709***) and young adults aged 16 to 24 (-0.353**) have less discretionary driving since children under 16 typically go to schools closer to home, but those with more workers drive more (1.267***) for non-discretionary purposes since more workers imply more driving to the workplace. Compared to the 48-67 baseline age group, younger households are more likely to drive more (0.629** for the 18 to 31 age group and 0.391** for the 32 to 47 age group), while older households (aged 68 and up) drive less for non-discretionary travel (-0.921**) as they are more likely to be retired.

It is also interesting to discuss how socio-economic variables significantly influence household select their residential location (Column III of **Table 3-3**). First, I see that households with higher incomes tend to choose lower density neighborhood (-0.001***), which is not surprising because higher income neighborhoods tend to have large lot houses. Likewise, larger households prefer neighborhoods with lower density (-0.036***) because they typically need housing with larger lots. Age of the head of household seems especially important as younger adults (0.227*** for the 18 to 31 group and 0.136*** for the 32 to 47 group) prefer neighborhoods with higher density than the 48 to 67 baseline group. Ethnicity also matters as Hispanics (0.256***), Black (0.479***), and Asians (0.365***) tend to live in higher density compared to their White socio-economic counterparts.

3.5.2.2 Indirect and total effects

The last two columns of **Table 3-3** reports indirect and total effects of endogenous and socio-demographic exogenous variables on AFV ownership, non-discretionary commuting, and residential density selection. Indirect effects (product-sum of relevant direct effects) refer to how these variables affect AFV ownership decisions through residential self-selection and work and school commuting.

I found several variables that exhibits interesting indirect effect on AFV ownership. Both the estimated direct effect of residential density on household AFV ownership and estimated indirect effect through work and school commuting are positive (-0.165*** & -0.002*** respectively), resulting in a stronger negative total effect (-0.167***). Households living in high density urban neighborhood does not commute as much as their suburban counterparts, and those who live in low density suburban neighborhoods might prefer AFVs because of their lower operating cost – and indeed households who commute more for work and school are more likely to prefer AFVs. The negative indirect effect points to the existence of a mediating effect of work and school commutes on the residential density effects on AFV preferences. This suggests that while households who live in lower density neighborhoods are more likely to be AFV owners, they do so even more because their longer commute.

The indirect effect of HOV and parking incentives on AFV ownership through work and school commuting is relatively mild compared to its direct effect. This suggests that while HOV exemptions (OR=1.004**) and parking privileges (OR=1.017***) for AFV owners may promote the adoption of AFV, they do so mostly by encouraging longer commutes (as cost per mile drops) – which might not be socially desirable.

Lastly, mediating effects also exist for households with higher educational attainment. Households holding graduate degrees are more likely to prefer AFVs, but at the same time, they are also more likely to choose lower residential density and thus would have a longer commute. This suggests that living in a high density urban neighborhood suppresses the effect of educational attainment on AFV preferences. However, on average, the indirect effect through suburban living preferences and longer commute eclipsed the effect of higher-density urban living for households with higher educational attainment (0.820*** for graduate degrees).

3.6 Conclusions

This paper presents a Generalized SEM model with a Logit link that jointly explains household ownership of AFVs, vehicle use for work and school purposes, and residential density while accounting for residential self-selection, political environmentalism, and the impact of incentives in the form of HOV access and parking privileges. To build my dataset, I combined household level data from the 2012 California Household Travel Survey with geospatial information about residential density, voting outcomes on selected California environmental ballot propositions, an information about the HOV network and the location of refueling stations for alternative vehicles.

My methodology offers several advantages. First, I used household-level data with a rich set of socio-demographic variables to account for household heterogeneity. Second, I used proximity to freeways with HOV lanes and AFV parking facility to measure operational (non-monetary) incentives for AFVs while most previous studies (Gallagher & Muehlegger, 2011; Diamond, 2008; 2009) relied on binary variables and used administrative units to measure the impacts of this type of incentives. Third, my GSEM framework endogenizes residential selection and non-discretionary driving in explaining AFV ownership.

My results show that political environmentalism matters. Households who live in neighborhoods favorable to pro-environment agendas are more likely to own AFVs, which is consistent with previous studies (Gallagher & Muehlegger, 2011; Diamond, 2008; Kahn, 2007). Second, while AFV incentives such as HOV exemption and parking privileges may nudge households to be AFV owners, they do so by encouraging longer commutes. Households who live farther from a freeway with HOV lanes or work farther from parking lots that favor AFVs (-0.004** and -0.017*** respectively) are less likely to be AFV owners. Thus, expanding access to HOV lanes with no occupancy requirement and granting AFV drivers parking privileges will increase AFV ownership. However, this will also increase non-discretionary driving, although the effect is relatively small. These results suggest that parking privileges provide a greater incentive. On average, a household moving 1 mile closer to an HOV lane would induce driving by an additional 0.12 miles per month. Working 1 mile closer to a parking lot with AFV privileges would induce a household to drive an additional 0.51 miles per month. This is further evidence that HOV programs come with unintended consequences (Shewmake & Jarvis, 2014; Shewmake, 2012; Kwon & Varaiya, 2008).

Lastly, a longer commute has a mediating effect on how residential density affects AFV ownership (indirect effect = -0.002***). Low density suburban households are more likely to own AFVs partly because of higher driving needs. This suggest that, while policies promoting high density neighborhood will reduce driving needs (direct effect = -0.265***), its effect on reducing likelihood of AFV ownership is considerably smaller (direct effect = -0.165***).

My research is not without limitations, which are mostly due to data availability. One limitation is my use of votes on ballot propositions to proxy for environmental beliefs. A second limitation is that the 2012 CHTS, like most travel diary surveys, provides only a cross-sectional snapshot of household travel behavior so I can only test a unidirectional relationship between vehicle utilization and AFV ownership via a recursive GSEM model. It would be of interest to explore if there is a bi-directional link between vehicle miles driven and AFV ownership. Testing this relationship would require estimating a non-recursive model on panel survey data. This is left for future work.

3.7 References

Beresteanu, A., & Li, S. (2011). Gasoline prices, government support, and the demand for hybrid vehicles in the United States. *International Economic Review*, *52*(1), 161-182.

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. doi:10.1016/j.trb.2005.12.005

Bollen, K. A. (1989). *Structural Equations With Latent Variables*. Wiley.

Brownstone, D. (2008). Key relationships between the built environment and VMT. *Transportation Research Board*, *87*(August), 466–478. Retrieved from http://onlinepubs.trb.org/Onlinepubs/sr/sr298brownstone.pdf

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, *65*(1), 91–98.

California Governor Executive Order B–16-2012, (2012) (https://www.gov.ca.gov/news.php?id=17472)

Choo, S., & Mokhtarian, P. L. (2004). What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice.*Transportation Research Part A: Policy and Practice*, *38*(3), 201-222.

Diamond, D. (2008). Impact of High Occupancy Vehicle (HOV) Lane Incentives for Hybrids in Virginia. Journal of Public Transportation, 11 (4): 39-58.

Diamond, D. (2009). The impact of government incentives for hybrid-electric vehicles: Evidence from US states. *Energy Policy*, *37*(3), 972-983.

Dillon, H. S., Saphores, J. D., & Boarnet, M. G. (2015). The impact of urban form and gasoline prices on vehicle usage: Evidence from the 2009 National Household Travel Survey. *Research in Transportation Economics*, 52, 23-33.

Fang, H. A. (2008). A discrete–continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B: Methodological*, *42*(9), 736–758. doi:10.1016/j.trb.2008.01.004

Gallagher, K. S., & Muehlegger, E. (2011). Giving green to get green? Incentives and consumer adoption of hybrid vehicle technology. *Journal of Environmental Economics and management*, *61*(1), 1-15.

Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological*, *37*(1), 1-25.

Greene, D. L., Patterson, P. D., Singh, M., & Li, J. (2005). Feebates, rebates and gas-guzzler taxes: a study of incentives for increased fuel economy. *Energy policy*, 33(6), 757-775.

Guber, D. L. (2003). *The grassroots of a green revolution: Polling America on the environment*. MIT Press.

Kahn, M. E., & Matsusaka, J. G. (1997). Demand for environmental goods: Evidence from voting patterns on California initiatives. *JL & Econ.*, *40*, 137.

Kahn, M. E. (2007). Do greens drive Hummers or hybrids? Environmental ideology as a determinant of consumer choice. *Journal of Environmental Economics and Management*, *54*(2), 129-145.

Kim, J., & Brownstone, D. (2013). The impact of residential density on vehicle usage and fuel consumption: Evidence from national samples. *Energy Economics*, *40*, 196–206.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). Guilford.

Kwon, J., & Varaiya, P. (2008). Effectiveness of California's high occupancy vehicle (HOV) system. *Transportation Research Part C: Emerging Technologies*, *16*(1), 98-115.

Puller, S. L., & Greening, L. a. (1999). Household adjustment to gasoline price change: an analysis using 9 years of US survey data. *Energy Economics*, 21(1), 37–52.

Sallee, J. M. (2011). The surprising incidence of tax credits for the Toyota Prius. *American Economic Journal: Economic Policy*, *3*(2), 189-219.

Salon, D. (2009). Neighborhoods, cars, and commuting in New York City: A discrete choice approach. Transportation Research Part A: Policy and Practice, 43(2), 180-196.

Shewmake, S., & Jarvis, L. (2014). Hybrid cars and HOV lanes.*Transportation Research Part A: Policy and Practice*, *67*, 304-319.

Shewmake, S. (2012). Can carpooling clear the road and clean the air? Evidence from the literature on the impact of HOV lanes on VMT and air pollution. *Journal of Planning Literature*.

Sierzchula, W., Bakker, S., Maat, K., & van Wee, B. (2014). The influence of financial incentives and other socio-economic factors on electric vehicle adoption. Energy Policy, 68, 183-194.

Small, K., & Dender, K. Van. (2007). Fuel efficiency and motor vehicle travel: the declining rebound effect. *The Energy Journal*, *28*(1). doi:10.5547/ISSN0195-6574-EJ-Vol28-No1-2

Tscharaktschiew, S. (2015). How much should gasoline be taxed when electric vehicles conquer the market? An analysis of the mismatch between efficient and existing gasoline taxes under emerging electric mobility. *Transportation Research Part D: Transport and Environment*, 39, 89-113.

4 Measuring urban form as a latent construct in travel behavior research: A case study of transit use and non-motorized travel in Southern California

This essay examines the influence of urban form on transit use and non-motorized travel (NMT, including biking and walking) for households (with at least one employed adult) in Los Angeles and Orange Counties in California based on 2009 National Household Travel Survey (NHTS) data (N=2,182 HH). The objectives of the research are (1) to assess several methods for measuring urban form features in the near-residence and near-workplace neighborhoods and (2) to assess the importance of these urban form features on transit use and NMT after accounting for the influence of these features on household vehicle ownership and residential selection. Results provide insights into the relative influence of several specifications of population density, transit access and walkability measures on transit use and NMT for commute and non-work trips. Reduced form models suggest that discretionary travels are dominated by household's socio-demographic status. In terms of residential selection, lower income, younger, and smaller households are more likely to choose a dense, pedestrian friendly, and transit rich neighborhood. In terms of vehicle ownership, households living in high density, pedestrian friendly, and transit rich neighborhoods are less likely to own vehicles. After accounting for the influence of urban form on vehicle ownership and residential selection, workplace transit accessibility has greater influence on transit commuting than transit access in household's residence. Results vary by how urban form is specified and by source of travel data. Finally, there are some evidence that population density affects non-motorized travel for discretionary purposes.

4.1 Introduction

Transit and non-motorized travel (NMT, including biking and walking) represent sustainable modes of transportation. Compared to automobiles, they require less resources and are less injurious to the environment for each miles-traveled (Kuzmyak & Dill, 2012). Contemporary planning has been increasingly focused on integrated land use, transit and active transportation plan in response to rising fuel prices, environmental awareness, and public health concerns. In car-dependent regions, such as Southern California, transportation planning organizations have also invested in these modes to meet social justice mandates (Spears, Boarnet, & Houston, 2016; Houston, Boarnet, Ferguson, & Spears, 2014; Southern California Association of Governments, 2016).

Pursuant to California's Sustainable Communities and Climate Protection Act of 2008 (SB 375), Metropolitan Planning Organizations (MPOs) are required to develop integrated transit investment and housing allocation plans to meet greenhouse gas (GHG) emission standards. In particular, the Southern California Association of Governments (SCAG) developed plans that are explicitly aimed at reducing the use of private cars and increasing the use of transit, walking and bicycling by coordinating transit corridor investments with housing supply trajectories (SCAG, 2016). In November 2016, Los Angeles county voters generously passed Measure M which is expected to generate $120 billion in public transit investment from a half percent increase in sales tax for the next 40 years.

These policy-goals largely rely on a large body of evidence about how transit improvements and land use changes affects travel behavior. The literature suggests that household travel behavior is influenced by a tangle of personal and household characteristics, socioeconomic status, and urban form[5]. Urban form is typically operationalized using a composite set of spatial metrics representing features near residences or travel destinations related to design (street connectivity and walkability), diversity (land use mix), and density (land use intensity). This behavioral framework begs the question: how well do these metrics capture how urban form influences travel behavior? More specifically, how useful are empirical models relying on these metrics in informing policy aimed at improving transit and active transportation.[6]

To answer these questions, the objectives of the research are (1) to assess several methods for measuring urban form features in the near-residence and near-workplace environments and (2) to assess the importance of these urban form features on transit use and NMT after accounting for the influence of these features on household vehicle ownership and residential selection. In particular, this research compares travel behavior models where urban form is measured simply by population density, to a more complex approach where urban form is conceptualized as an unobserved (latent) construct, or factors, that underlie multiple observed land use variables. These variables include measures of walkability and transit access, which may better explain walking and transit use respectively, both at the trip's origin (home) and the destination (workplace).

---

[5] In this dissertation, urban form refers to the physical patterns, layouts, features, and structures of urban space. I used it as a 'catch-all' that is interchangeable with land use and built environment. Semantics aside, I am approaching these concepts from a model-selection perspective.

[6] To partly answer these questions, this essay considers household's transit usage and NMT because they are likely to be complements; transit users have to augment their travel by walking or biking to and from their final destination. Park and ride is another thing to consider and it is beyond the scope of this paper.

4.2 Literature review

To motivate my modeling strategies and contextualize my results, I first review of the literature on urban form and travel behavior. I focus on empirical papers between 1997 to 2016 that examined transit usage and NMT and various methods to define, specify, operationalize, and capture urban form. Finally, I discuss methodological treatments of residential and vehicle ownership self-selection in this field.

4.2.1 Urban form and travel behavior

Since the 1980s, proponents of the New Urbanism movement have sought to promote socially and environmentally desirable travel behavior, namely less driving and more walking and transit use, by changing the urban environment. The popularity of these ideas can be attributed to the increasing social and environmental costs of driving. Today, many MPOs and transportation planners in the United States have turned to land use planning and urban design measures to rein in automobile use (e.g., Boarnet & Crane, 2001; Ewing & Cervero, 2010) and promote non-motorized travel (NMT) and transit use (e.g., Khan, Kockelman & Xiong, 2014).

The influence of urban form (including land use, urban design, and other physical features of the urban landscape) on travel behavior is the most heavily researched subject in urban transportation planning (Boarnet & Crane, 2001; Ewing & Cervero, 2001, 2010). As reported in recent reviews (e.g., see Boarnet *et al.*, 2011; Boarnet, 2011; Brownstone, 2008; Ewing & Cervero, 2010), previous studies concerned with the relationship between urban form and travel behavior estimate models where measures of travel behavior (VMT, transit use, walking, trip frequency) are regressed on individual and household socio-demographic variables (i.e., age, ethnicity, education level, household income and size) and measures of urban form around each household's residence.

Boarnet & Crane (2001) proposed that the near-residence configuration of urban form influences travel behavior because it shapes the time cost of travel. This proposal is rooted in a microeconomic theory / behavioral framework which assumes travel behavior (trip length, trip frequency, mode choice) represents a decision shaped by the cost of travel and budget constraints, both in monetary terms and time. For example, street grids, mixed land uses, transit access, and pedestrian-oriented neighborhoods change the time cost of travel or the relative cost across modes. Aesthetic design elements could make walking trips more pleasing and thus shape the 'psychological' or subjective cost of travel. This subjective cost can be measured by perceived neighborhood walkability (Cao, Handy & Moktharian, 2006; Voorhees *et al.*, 2010). On the other hand gasoline price, congestion pricing, tolls, price of parking, and price of transit ridership affect the monetary cost of travel.

In the same vein, de Dios Ortuzar & Willumsen (2011, pp. 493-7) provide an exhaustive review on how land use planning decisions—assumed to be exogenous—may affect accessibility and spatial-activity patterns. Small & Verhoef (2007, pp. 12-14) thoroughly discussed the need to account for endogeneity of land use in travel behavior research and available econometric techniques to handle this modeling problem.

Urban form alone, however, does not explain travel behavior in full. Spears, Houston, and Boarnet (2013) studied the role of socio-psychological factors on transit ridership. They used Confirmatory Factor Analysis (CFA) to capture respondents' attitude towards transit riders and safety perception, and fitted logistic and Tobit regressions to explain transit patronage on a sample of 279 Los Angeles, CA households. They found that these socio-psychological factors matter, even after controlling for demographic and urban form variables. For example, attitude towards transit and safety concerns affect transit ridership decisions.

4.2.2 Early conceptualization of urban form in travel behavior research

Some researchers have sorted elements of urban form into sometime overlapping categories with words beginning with D: density, diversity, design, destination accessibility, and distance to transit service, and demand management (Cervero & Kockelman, 1997; Ewing & Cervero, 2001, 2010; Cervero & Duncan, 2003; Ewing *et al*., 2009). Empirical results, however, show that this categorization of urban form is not axiomatic (Cervero & Kockelman, 1997; Cervero & Duncan, 2003). Density is measured as the variable of interest per unit of area, gross or net, where the variable can be population, residential units, and employment. Diversity of land use mix in an area is usually measured with an entropy index (see following section for the formula) where low values indicate single-use environments and higher values more varied land uses. An entropy index of one, the maximum, suggests that all types of land uses are equally represented within the area. Design typically captures street-network characteristics, including connectivity, sidewalk coverage, pedestrian crossings and other metrics that set pedestrian-oriented environment apart from automobile-oriented ones.

Cervero & Kockelman (1997) relied on factor analysis to extract a small number of underlying factors of urban form dimensions that can be used to represent relationships among sets of interrelated urban form variables. They used data from the Association of Bay Area Governments (ABAG) land-use inventory in 1990 to calculate 27 urban form variables, and extract three underlying factors—density, diversity, and design— using factor analysis (p. 210).

However, only 12 out of 27 candidate variables were used in the final extraction as factor loadings with a magnitude under 0.30 were discarded. The first factor, which captures density, accounts for 47.6% of the total variation. The second factor, accounting for 18% of the total variation captures the design dimension, and more specifically walking quality (pedestrian-friendliness or walkability). Their analysis suggests that walking quality is positively associated with the provision of sidewalks and street lights in flatter areas with planted strips, and negatively associated with block length and lighting spacing. While their factor-analytic results cannot attest to the three-dimensionality of urban form, Cervero and Kockelman (1997) found that the concept of urban form may be defined along distinct elements, and that these elements are associated with travel behavior, although in ways that are complex and difficult to fully capture. Multiple regression results from Cervero and Kockelman (1997) showed overall that neighborhoods with better walking quality are related to less driving and that density, land use diversity, and pedestrian-oriented neighborhoods encourage non-auto travel.

Urban form and active travel behavior is also a critical issue in public health (Cervero & Duncan, 2003; Frank *et al.*, 2005). Obesity, for example, is more prevalent in areas are not considered walkable. In the same vein of Cervero & Kockelman (1997), Cervero & Duncan (2003) relied on principal component analysis to capture the dimensionality of urban form, both at trip origin and destination, before using the factor-extracts as covariate to explain active travel behavior. Like Cervero & Kockelman (1997), they found two, instead of three, dimensions of urban form: density (intensity of land uses) and design (walkability, pedestrian friendliness, and street-connectivity). Their multinomial-logit regression suggests that land-use diversity is the strongest predictor of walking (mostly for social and shopping purposes).

Frank *et al.* (2005) measured urban form using residential density, street connectivity (intersection density), and land use mix. Physical activity was measured using portable accelerometers worn at the hip of the respondent. Normalized measures of urban form (z-scores) were combined into a near-residence walkability index. This index is calculated as a weighted sum of the normalized value, and the weights were determined in an ad hoc manner to maximize explanatory power. Their linear regression models indicate that their walkability index explains a substantial fraction of the variance in physical activity after controlling for gender, age, education, and ethnicity.

4.2.3 Strategies in measuring urban form

More recent studies benefited from GIS to better capture environmental attributes, which allowed researchers to observe heterogeneity of urban form in greater detail (Leslie *et al.*, 2007; Voorhees *et al.*, 2010; Manaugh & El-Geneidy, 2011; Giles-Corti *et al.*, 2011; Khan, Kockelman & Xiong, 2014). Furthermore, if spatial data are available in sufficient granularity, researchers no longer need to rely on predetermined neighborhood boundaries, such as census tracts or transportation analysis zones, which was a constraint for Cervero & Kockelman (1997) and Cao, Handy & Mokhtarian (2006). Leslie *et al.* (2007) demonstrates how almost any type of spatial data can be used to develop and derive measures of urban form to more directly account for near-residence environments and its influence on travel behavior. The following papers have used GIS capabilities to generate measures of urban form in active travel behavior research.

Voorhees *et al.* (2010) studied the effect of self-reported neighborhood characteristics and perception on walking to school by 8th grade girls who lived within 1.5 miles from their school. Using GIS, they created a ½ mile street network buffer around each girl's home ("synthetic neighborhoods") where they captured various features of urban form including street connectivity (alpha, beta, and gamma indices; see next section for complete definition), block size, population density, land use mix, and availability of active destinations (e.g., parks, swimming pools, basketball courts, etc.). After geocoding home addresses, they related each buffer to its socioeconomic characteristics from the 2000 Census. Their survey also captured perceived neighborhood characteristics using a ten-item questionnaire about safety, aesthetics, and access to facilities. However, they did not control for residential self-selection. They reported that girls who perceived that their neighborhood is safe for walking are twice as likely to walk to or from school compared to those who do not. Moreover, girls with more total destinations available in their neighborhood were 38% more likely to walk to or from school whereas girls living in neighborhoods with larger block sizes were 20% less likely to walk.

Manaugh & El-Geneidy (2011) analyzed household travel data from the Montreal Origin-Destination Survey (2003) to explore how well 4 measures of walkability—walkability index as suggested by Frank *et al.* (2005), walk opportunities index, pedestrian shed, and Walk Score® – explain walking behavior. Like Frank *et al.* (2005), their walkability index is the sum of z-scores of urban form variables. Additionally, they created a measure of walk opportunity index, which reflects accessibility, composed of the number of retail businesses, combined with weights to represent network connectivity and the importance of each retail type (see also Kuzmyak *et al.*, 2005). Pedestrian shed is measure of accessibility that is conceptualized as a 'walkable catchment' area from a given point (see also Porta & Renne, 2005). It is calculated as the proportion of area that is truly accessible by sidewalk availability and connectivity, an attribute of the network, to the total area that would have been accessible on a 10 minute walk (a circle).

Finally, Walk Score®, a proprietary index of walkability, is used at the zip-code level. Walk Score® uses a patented method that assigns points based on distance to amenities, combined with a decay function (see walkability.com/methodology). The decision to walk on a home-based trip to school and shops, was separately modeled as a binary variable in a logistic model. Manaugh & El-Geneidy (2011) found that the Walk Score® index explains more of the variation in walking trips to shopping than other indices of walkability. For walking to school, pedestrian shed was found to be the best walkability predictor.

Duncan, Alstadt, Whalen, and Melly (2013) investigated the validity of Walk Scores and Transit Scores using Spearman correlations of these scores with GIS generated metrics, correcting for spatial autocorrelation. They collected a school-based sample of 1,292 residential addresses from public high school students in Boston, MA, and used this geocoded information to draw Walk Score and Transit Score. They use buffer analysis to generate GIS metrics from multiple spatial databases such as MassGIS, ESRI Business Analyst InfoUSA, Census demographic and housing data, etc. Their results offer evidence that, at larger spatial scales, Walk Score is a good, convenient tool to measure certain aspects of walkability and transit availability. Duncan *et al.* (2013), however, did not use travel or physical activity data to inspect the validity of Walk Score and Transit Score as a measure of walkability and transit availability respectively. If Walk Score is indeed a measure of walkability and transit availability, then neighborhoods that report higher scores should predict higher levels of walking and transit use (i.e. in an experiment, as thermometer reading, a measure for temperature, increases, water should eventually come to a boil).

Houston (2014) investigates the influence of Modifiable Areal Unit Problem (MAUP) to analyze the effect of urban form on moderate and vigorous physical activities. He uses a high quality GPS accelerometer data from 143 participating households which generated 5,066 locations – a fine-grained data for physical activities. Urban form metrics, which included exposure to green spaces, came from multiple sources. Green spaces data was comprised of tree canopy, irrigated and non-irrigated grass cover come from high-resolution land cover map. Intersection density as a measure of street connectivity wass used a proxy for walkability. Land use and employment data come from 2011 InfoUSA and SCAG employment database in GIS format. To evaluate scale effects, Houston (2014) fitted separate logistic regression for each size of the three buffer- and grid-based scales. He found that green space cover was significantly associated with physical activity at the 100 meter grid, but not at larger grid sizes. Green space cover was significantly positively associated with physical activity at all buffer sizes (50 m, 250m, and 500m), but with a decreasing magnitude of influence as buffer size increase, suggesting the need for distance-weighting in specifying urban form in the context of travel behavior or physical activity.

Giles-Corti *et al.* (2011) examine the impact of neighborhood walkability within 2 kilometers of public schools on children's walking to school behavior. They used GIS to develop two measures of walkability: pedestrian shed and traffic exposure, and created an index using the sum of decile-scores (1= least walkable, 10=most walkable) of each measure. They found that the likelihood of walking to school was increased by greater walkability but is reduced by traffic exposure. Connected street networks provide direct routes to school, and thus make for a walkable neighborhood centered on a school. But when designed for heavy traffic (i.e., lack of traffic calming measures), pedestrian safety might compromise the potential for children to walk to school.

Khan, Kockelman & Xiong (2014) fit several generalized linear model specifications on detailed travel data from the Puget Sound Regional Council (PSRC) 2006 household travel survey to evaluate the effects of urban form on active travel modes. Their survey data includes 10,510 individuals across 4,741 households residing in Seattle metropolitan area. In this travel survey, each trip was connected to an origin and destination parcel, thanks to very fine parcel-level information inventoried by PSRC. Their models did not control for self-selection and thus the marginal-effect may be overestimated (p. 125). The results, however, stress the importance of street connectivity, transit (bus-stop) availability, and land use diversity in policies aimed to promote active transportation.

4.2.4 Residential and vehicle ownership self-selection

While seminal in operationalizing urban form in the urban planning and public health literatures, Cervero and Kockelman (1997) and Frank *et al.* (2005) did not address the problem of residential and vehicle self-selection which may bias estimates of the effect of urban form on travel behavior (Cao, Handy & Mokhtarian, 2006; Brownstone, 2008). Urban form may not be a causal factor on travel behavior. To a certain extent, it could mask the effect of socio-economic factor (i.e. income, employment status, household composition) that do influence travel.

Cao, Handy & Mokhtarian (2006) was the first to examine residential self-selection on the effect of urban form on pedestrian behavior. To control for residential self-selection they explicitly asked their survey respondents whether proximity to stores was an important factor in choosing their current residence. Cao, Handy & Mokhtarian (2006) surveyed 1,368 households in 6 middle-income Austin, Texas neighborhoods. They collected data on household walking behavior and their perception of the neighborhood's walkability, combining it with measures of street-connectivity measures at the neighborhood level. They fitted a Negative Binomial regression model to explain walking frequency, using perceived (self-reported) and actual neighborhood urban form characteristics as observed independent variables (they did not attempt to extract any latent factors). Their results suggest that, after controlling for self-selection, neighborhood characteristics impacted strolling (leisure walking) frequency. They also reported that perceived characteristics had more explanatory power than actual (objective) measures of street connectivity.

Establishing causal links in travel behavior research, especially those focusing on active travel, has been difficult mostly because of the cross-sectional nature of travel survey data. Wasfi *et al.* (2016) overcame this problem by using the Canadian National Population Health Survey (NPHS), a longitudinal database, to study the effect of neighborhood walkability and walking behavior. Respondents who changed residential locations during the study period were exposed to various levels of changes in urban form, and thus provided a quasi-experiment of changes in walking behavior associated with changes in walkability, measured by Walk Score®. While the decision to change residential location may very well be a self-selection process, the household does not have the choice over the level of change in neighborhood walkability. After taking care of attrition, their fixed-effect model suggest that exposure to more walkable neighborhoods and moving from less walkable to more walkable neighborhoods was associated with increases in utilitarian walking, even for individuals who were otherwise inactive during their leisure time.

Previous studies have noted the importance of addressing vehicle ownership endogeneity in travel behavior research (e.g. Dill, 2004; Kuzmyak & Dill, 2012; Bento *et al*., 2005; Mannering, 1986; Zegras, 2010; Houston, Boarnet, Ferguson, & Spears, 2015; Dillon, Saphores, Boarnet, 2015). Dill (2004) and Kuzmyak & Dill (2012) report that vehicle ownership and availability is perhaps the most telling explanatory factor of transit and NMT use. On average, in households with fewer vehicles than licensed drivers (vehicle ratio less than 1) walk 12.3 percent and bike 1.6 percent of their daily trips, while households with more vehicles than licensed drivers (vehicle ratio more than 1) only walk 7 percent and bike 0.8 percent for their daily trips.

Houston, Boarnet, Ferguson, and Spears (2015) used a two-stage-regression approach to control for vehicle ownership endogeneity in estimating the impact of transit infrastructure on walking and transit usage. They found that households in high density neighborhoods and higher transit service were associated with lower vehicle ownership and utilization. Households within 0.5-2.0 miles of a rail transit station tended to walk and use transit more and that this relationship was stronger for households within 0.5 miles. Spears, Boarnet, and Houston (2015) used an experimental design to estimate the effects of light rail transit investment in Los Angeles and confirmed that proximity to rail transit stations reduced household driving and increased rail transit use.

Table 4-1 Selected Review of Transit & NMT Papers

Author(s) (Year)	Question of interest	Data	Methods	Key findings
Cervero & Kockelman (1997)	Influence of the built environment (density, diversity and design) on travel demand	Bay Area Travel Survey (BATS) 1990, N=896 HHs ABAG land-use inventory (1990), field surveys of N=50 neighborhoods	Factor Analysis Multiple regression	Density, land-use diversity, and pedestrian-oriented design reduced trip-rates, and encouraged non-auto travel
Cervero & Duncan (2003)	Influence of the built environment (density, diversity, and design) on active travel	Bay Area Travel Survey (BATS) 2000	Factor Analysis Multinomial logit model for mode choice	Land-use diversity was the strongest predictor of walking (mostly for social and shopping purposes)
Frank *et al*. (2005)	Influence of urban form on physical activity	Metropolitan Atlanta's Regional Transportation and Air Quality (N=523)	Linear regression	Measures of land-use mix, residential density, and intersection density were positively related with number of minutes of moderate physical activity per day
Cao, Handy & Mokhtarian (2006)	Influence of the built environment and residential self-selection on pedestrian behavior	Primary travel survey (1995). N= 1,368 adults in 6 neighborhoods of Austin, TX	negative binomial regression	After accounting for self-selection, neighborhood characteristics impact strolling frequency
Leslie *et al*. (2007)	Measuring built environment features that influence physical activity (walking and biking)	Physical Activity in Localities and Community Environments (PLACE), Australia N=32 communities	Described GIS spatial data methodology to measure built environment attributes	Demonstrates GIS capability to capture built environment influences on physical activity Computes decile scores and indices of built environment attributes

Author(s) (Year)	Question of interest	Data	Methods	Key findings
Voorhees *et al*. (2010)	Influence of actual and perceived neighborhood features on walking to or from school	Trial of activity in Adolescent Girls Survey of 8th grade girls. N= 890 girls, 36 schools	Nested multivariate (mixed effects) logistic regression	Safety, distance to school, and street design influences walking to and from school
Manaugh & El-Geneidy (2011)	Correlation of walkability and travel behavior	Montréal origin–destination survey (2003) N=17,394 households	Logistic regression	Walkability does not have the same correlation with travel behavior for all individuals or households
Giles-Corti *et al*. (2011)	Impact of walkability on walking to school	Survey of school children in Perth, Australia (2007) N=1,314 children, 238 schools	Logistic regression	Connected street network and low traffic volume increases the likelihood of walking to school
Spears, Houston & Boarnet (2013)	The role of socio-psychological factors on transit ridership	N=279 HH in south Los Angeles, CA	Confirmatory Factor Analysis, logistic regression, Tobit model	Attitudes towards transit and safety perception affect transit ridership
Duncan *et al*. (2013)	Validity of Walk Scores and Transit Scores	School-based sample of students in Boston, MA (N=1,292)	Spearman correlations of Walk Scores and Transit Scores with GIS-generated metrics	At larger spatial scales, Walk Score is good, convenient tool to measure certain aspect of walkability and transit availability
Khan, Kockelman & Xiong (2014)	The effects of built environment variables on the use of non-motorized travel modes	Puget Sound Regional Council (PSRC) 2006 household travel survey N = 4,741 HHs	Generalized linear models (not control for self-selection): 1) Household vehicle ownership (Poisson); 2) Non-motorized trip generation (Zero-Inflated Negative Binomial); 3) Intrazonal versus interzonal trip type binary logit; 4) Interzonal: Destination choice models, by trip purpose (MNL); 5) Intrazonal	Higher rates of non-motorized trips and lower vehicle ownership levels are associated with street connectivity, transit availability and land use diversity

Author(s) (Year)	Question of interest	Data	Methods	Key findings
			& interzonal mode choice by trip purpose (MNL); 6) VMT and non-motorized MT (Tobit)	
Houston (2014)	The influence of Modifiable Areal Unit Problem in the urban form – physical activity	GPS and accelerometer data from 143 participating households (N=5066 pings)	Difference-in-means test between urban form measures captured using circular buffers and grid areas	Scale and zone configurations affect the influence of green space and land use on walking
Wasfi *et al*. (2016)	Relationship between utilitarian walking levels and neighborhood walkability through longitudinal analyses of a population cohort	Canada's National Population Health Survey (N=2976; biannual assessments 1994–2006)	Mixed effects ordered logistic regression Fixed effects logistic regression	Exposure to more walkable neighborhoods and moving from less walkable to more walkable neighborhoods are associated with increases in utilitarian walking, even for otherwise inactive individuals
Houston *et al*. (2015)	The role of rail transit on driving reduction	Los Angeles sub sample of CHTS 2012 (N=8219) and supplemental National Travel and Activity Survey sample (N=383)	Two-stage regression with instrumented vehicle ownership; negative-binomial, multinomial logit, and Tobit models	Features of the rail service and urban form 'maturity' affect travel outcomes
Spears, Boarnet & Houston (2016)	The impact of light rail transit investment	Seven-day travel log, before and after, N = 285 HH; control = 114 HH, experimental = 171 HH	Difference-in-difference, t-test, between-group before-after test, using (quasi) experimental data	Proximity to new light rail stations reduces driving

4.3 Data description

4.3.1 NHTS travel diary and typical commute data

The travel data used in this paper come from the National Household Travel Survey (NHTS) 2009, which obtained data from 4,659 households in the Los Angeles and Orange Counties of California. The geocoded raw data include the latitude and longitude of each household location and travel destination. Since I focus on active commuting (travel to work), I selected households with have at least one worker, which represents approximately 66% of households in the NHTS sample and corresponds to 3,078 households. However, only 70% of these households provided accurate information about their workplace location, resulting in a final sample size of 2,182 households.

The 2009 NHTS was conducted from March, 2008 to April, 2009, includes a one-day travel diary of all household trips and detailed household socio-economic and demographic information. Participating household members were also asked about their employment status and their typical transportation mode to work. There are two sources of information for household's commuting behavior: travel mode reported in the travel diary and the reported typical commute mode. The travel diary data also contain information for purpose for each trip which was used to identify work commute trips and non-work trips. This paper examines factors associated with typical commute mode, actual commute mode (based on the travel diary), and non-work discretionary trips (based on the travel diary). Active of non-motorized travel (NMT) is defined as walking or biking.

4.3.2 Socio-economic and demographic characteristics

Following my literature review, I include a rich set of household socio-economic and demographic variables (income, number of workers, number of licensed drivers, number of vehicles, and household composition) and the characteristic of head of the household (age, ethnicity, and highest educational attainment).

For household income, the midpoint of each NHTS 2009 income bracket is used. Since income bracket is top-coded at $100,000, I created a binary variable for households in that bracket. Household compositional variables, which were found to be predictive of travel behavior, consist of the number of children (under 16 years old), number of young adults (aged 16 to 24), and household size. For race, age, and educational attainment, being White, aged 45 to 64 (Baby Boomers), and earned less than a high school diploma are baselines.

Table 4-2 compares the racial and ethnic composition between the sample and the study area's population based on American Community Survey (ACS) 2008-2012. Table 4-2 suggests that in Los Angeles and Orange Counties, NHTS 2009 severely oversampled Whites and under-sampled Hispanics. It also moderately oversampled Asian and undersampled Blacks.

Table 4-2 Racial and ethnic composition

	NHTS 2009 Sample (N=2,182 HH)	ACS 2008-2012 (12.7 million)
White	68.4%	32.4%
Hispanic	20.8%	43.8%
Black	5.1%	6.8%
Asian	9.7%	14.7%
Other ethnicity	2.9%	2.2%

4.3.3 Land use variables: density, diversity, and transit access

To measure urban form, I used fine-grained geospatial information; my variables include population density, land use diversity, employment density, and distance to employment centers (as a proxy to job-housing balance), transit availability, and street-network connectivity. Using Geographical Information System (GIS) software, I assigned land use characteristics to each household based on both residential and workplace locations. I relied on 2010 census data to measure population density at the block group level. As a proxy for employment accessibility, I extracted from Census Transportation Planning Products (CTPP) data the number of jobs in buffer areas with a 10 miles radius centered on each household residence.

Land use mix was measured from SCAG's 2005 parcel level land use database. This information is available in GIS, which allowed calculating an index of land use diversity at the block group level using the entropy formula (e.g., see Dillon, Saphores & Boarnet, 2015) as follows:

$$H_i = \frac{-1}{\ln(C)} \sum_{j=1}^{C} p_{ij} \ln(p_{ij}), \qquad (1)$$

where H_i is the land use entropy of block group i, and p_{ij} is the area's proportion of land use of type j in block group i. SCAG's land use database stores parcel level land use based on 150 categories, which we condensed into C=15 major types of land uses.[7]

[7] These land uses category are: single family residential, multi-family residential, other residential types (e.g. mobile home and trailer parks) commercial and services, industrial, transportation, communication and utilities, public facilities such as government offices, schools and libraries, military land uses, mixed development, open space and recreational, vacant urban land, agricultural, vacant non-urban land (e.g. abandoned mines), body of water and facilities, and other land uses.

Transit service data were obtained from SCAG's transit network and from the 2008 Los Angeles Metropolitan Transit Authority (LA Metro) GIS database. The shapefile contains data about each transit stop, including information on service frequency and connectivity. To measure level of service, I aggregated the level of service for each transit stop. By using proximity analysis with ArcGIS, I calculated transit availability by aggregating the number of transit lines within ¼ mile from each household residence and work location. Aggregate level of service was measured by transit stops located within these buffers to capture the extent to which a household has the opportunity to use transit to commute.

4.3.4 Street connectivity

Advocates of New Urbanist and neo-traditional planning concepts include street connectivity as a key component for good neighborhood design. Street networks that are more grid-like are preferred over networks that include many cul-de-sacs and long blocks, thus increasing walking distances between destinations. The increased distances are thought to discourage walking and bicycling and, thus, active commuting. There is also debate over how to measure connectivity and what levels of connectivity are appropriate. The current debate is particularly unclear, and hard to test empirically, because street connectivity is proposed to meet multiple, sometime conflicting objectives. In addition, most efforts to date have focused on the street network, which may differ from the pedestrian and bicycle network (Dill, 2004; Kuzmyak & Dill, 2012).

In this study, I used GIS to calculate measures of street network and bikeways connectivity, both for residential and workplace locations. Street network data come Caltrans GIS Database (Caltrans, 2010; source: http://www.dot.ca.gov/hq/tsip/gis/datalibrary/). Bikeways data for each county come from the Los Angeles County Metropolitan Transportation Authority

(LA Metro, 2010; source: http://developer.metro.net/introduction/bikeways-data/download-bikeways-data/) and the Orange County Transportation Authority (OCTA, 2012; http://www.octa.net/Plans-and-Programs/GIS-Data/GIS-Data-Download/).

Intersection density is measured as the number of intersections per square mile, aggregated at the block group level. A higher number indicates more intersections and, presumably, higher connectivity. Street density is measured as the number of linear miles of streets per square mile of land, which was measured for both street and bikeways. A higher number indicates more streets and, presumably, higher connectivity. Street density and intersection density are likely to be strongly and positively correlated.

Transportation geographers have developed other measures of network connectivity based on graph theory which relies on links and nodes (Dill, 2004, pp. 5-7). Link-Node Ratio (Beta index) is a simple measure of connectivity and equals the number of links divided by the number of nodes within in a study area within a block group. Links are defined as roadway or pathway segments between two nodes. The Alpha index uses the concept of a circuit – a finite, closed path starting and ending at a single node. It evaluates connectivity using the ratio of the actual and the maximum possible number of independent circuits. Thus, this index is independent to the number of nodes. Trees roadway configurations and simple networks will have a value of 0. A value of 1 indicates a completely connected network. This index is also called "meshedness" coefficient in the literature on planar networks. The maximum number of circuits is expressed as (2*#nodes – 5) and the number of actual circuit as (#links - # nodes +1).

The Gamma index is a ratio of the number of links in the network to the maximum possible number of links between nodes. The maximum possible number of links is expressed as

3*(# nodes – 2) because the network is abstracted as a planar graph. Values for the gamma index range from 0 to 1 and can be expressed as a percentage of connectivity (e.g., a gamma index of 0.48 means that the network if 48 percent connected).

4.3.5 Walk Score® and Transit Score®

Walk Score and Transit Score are, respectively, measures of walkability (or pedestrian friendliness) and transit availability that were originally developed by Front Seat in 2007, before it was bought by Redfin, a technologized real-estate agency, in 2014. These scores were generated by a proprietary web-based algorithm, using publicly available geographical data. For any input location, the algorithm computes a weighted-sum based on distance to various amenities. Weighting is based on amenity type and a distance decay function up to 1.5 miles (~30 minute walk) (source: https://www.walkscore.com/methodology.shtml). These sums are normalized to a score between 0 (poor) to 100 (excellent).

Walk Score is computed using population density data, network connectivity metrics, and distance to amenities. Transit Score uses data released from transit agencies to generate scores based on distance to nearest stop and level of service (frequency and transit mode; rail or bus). While Walk Score is increasingly used in transportation planning, housing, and public health research (https://www.walkscore.com/methodology.shtml), the literature on Transit Score is sparse. For this study, I used an Application Programming Interface (API) to extract scores for each household and work location. Summary statistics for my variables are presented in Table 3.

Table 4-3 Summary Statistics (N=2,182)

Variable	Min	P25	Mean	P75	Max	Std. dev.
Endogenous Variables / Travel Behavior outcomes						
Typical use Transit to work (Binary)	0.000	0.000	0.058	0.000	1.000	0.234
Typical use of NMT Travel to work (Binary)	0.000	0.000	0.038	0.000	1.000	0.192
Transit Commuting / Home-Based Work (Binary)	0.000	0.000	0.029	0.000	1.000	0.169
NMT Commuting / Home-Based Work (Binary)	0.000	0.000	0.025	0.000	1.000	0.155
Transit Discretionary trips (Binary)	0.000	0.000	0.045	0.000	1.000	0.208
NMT Discretionary trips (Binary)	0.000	0.000	0.407	1.000	1.000	0.491
Vehicles per licensed driver	0.000	1.000	1.119	1.000	7.000	0.492
Household Characteristics						
Midpoint of annual HH income (in $1,000)	2.500	47.500	71.263	100.000	100.000	31.200
=1 if HH income [100K, ~)	0.000	0.000	0.397	1.000	1.000	0.489
Household size	1.000	2.000	2.885	4.000	10.000	1.360
Number of kids <16 in HH	0.000	0.000	0.383	1.000	4.000	0.734
Number of worker(s)	1.000	1.000	1.511	2.000	5.000	0.666
Head of HH Characteristics						
Age between 16 and 29	0.000	0.000	0.053	0.000	1.000	0.224
Age between 30 and 44	0.000	0.000	0.261	1.000	1.000	0.439
Age between 45 and 64	0.000	0.000	0.540	1.000	1.000	0.498
Age 65 and up	0.000	0.000	0.145	0.000	1.000	0.352
Education: less than High School	0.000	0.000	0.065	0.000	1.000	0.246
Education: High School degree	0.000	0.000	0.140	0.000	1.000	0.347
Education: some college	0.000	0.000	0.288	1.000	1.000	0.453
Education: Bachelor's degree	0.000	0.000	0.275	1.000	1.000	0.447
Graduate or professional degree	0.000	0.000	0.233	0.000	1.000	0.423

Variable	Min	P25	Mean	P75	Max	Std. dev.
White	0.000	0.000	0.684	0.000	1.000	0.465
Hispanic	0.000	0.000	0.208	0.000	1.000	0.406
Black	0.000	0.000	0.051	0.000	1.000	0.220
Asian	0.000	0.000	0.097	0.000	1.000	0.296
Other ethnicity	0.000	0.000	0.029	0.000	1.000	0.167
Residential Urban Form[a]						
Population density (10,000 person/square mile)[c]	0.012	0.530	1.742	1.773	59.448	3.157
Housing units per square mile (thousands)	0.050	1.500	4.274	7.000	30.000	4.140
Workers per square mile (census tract; thousands)	0.025	0.750	2.333	3.000	5.000	1.686
Percentage of renter occupied housing	0.000	0.050	0.344	0.600	0.950	0.291
Inverse distance to nearest sub-center (in 1/miles)	0.051	0.158	0.294	0.332	6.262	0.306
Transit service within 0.25 miles (/100)	0.000	0.000	0.044	0.060	5.130	0.135
Land use entropy (diversity index)	0.000	0.224	0.332	0.446	0.727	0.155
Bikeway density	0.009	2.632	6.371	9.466	22.064	4.586
Alpha index	0.000	0.123	0.219	0.291	1.000	0.130
Gamma index	0.340	0.418	0.482	0.531	1.000	0.087
Link/Node Ratio (Beta index)	0.800	1.233	1.410	1.553	2.333	0.237
Intersection density	0.002	0.042	0.063	0.081	0.287	0.031
Street Density	0.006	0.095	0.132	0.166	0.395	0.054
Workplace Urban Form[a]						
Intersection density	0.000	0.003	0.006	0.008	0.025	0.003
Street density	0.006	0.084	0.126	0.160	0.432	0.058
Alpha index	0.000	0.150	0.257	0.333	1.000	0.141
Gamma index	0.340	0.437	0.508	0.559	1.000	0.094
Link/Node Ratio (Beta index)	0.667	1.285	1.482	1.642	2.734	0.259
Bikeway density	0.003	2.436	6.558	9.695	22.892	4.814

Variable	Min	P25	Mean	P75	Max	Std. dev.
Transit service within 0.25 miles (/100)	0.000	0.000	0.206	0.130	6.100	0.656
Walk Score® and Transit Score®						
Residential Walk Score	0.000	36.000	52.895	71.000	99.000	23.652
Residential Transit Score	0.000	26.000	35.158	44.000	100.000	17.281
Workplace Walk Score	0.000	47.000	62.538	81.000	99.000	22.457
Workplace Transit Score	0.000	32.000	43.470	53.000	100.000	19.808

Note: P25 and P75 denote the lower and upper quartiles respectively. HH = household.
[a] Measured at the block group level unless stated otherwise.

4.4 Methodology

The objectives of the research are (1) to assess several methods for measuring urban form features in the near-residence and near-workplace environments and (2) to assess the importance of these urban form features on transit use and NMT after accounting for the influence of these features on household vehicle ownership and residential selection. To do so, this essay compares five modeling approaches that share a common recursive conceptual path model that accounts for vehicle and residential endogeneities, but varies by how residential urban form was measured and specified. A path model posits a hypothesized structural relationship between endogenous and exogenous variables that can be written as a system of equations.

This section will proceed as follows. I start with reduced form logit models to explore the influence of urban form and household's socio-demographic characteristics on the probability of transit usage and NMT, without controlling for endogenous self-selection effects. Then, I explain my conceptual path model to address these effects by explicitly parameterize relationships between endogenous variables. Next, I expound upon urban form modelling approaches and elucidate theoretical implication of each approach. Lastly, I specify a recursive Generalized Structural Equation Model (GSEM) explaining transit and active commuting, based on the aforementioned conceptual model.

4.4.1 Reduced form model

A reduced form model was estimated as a starting point in assessing the influence of urban form features on transit usage and NMT, and how these relationships vary across different urban form specifications. In the reduced form regression, the travel behavior outcome is the only endogenous variable and all other variables are assumed to be exogenous – assuming away endogenous self-selection effects (which will be accounted for using GSEM in subsequent analysis).

Reduced form regression in this study is a logistic regression in which transit usage and NMT is specified as a function of urban form and household's socio-demographic characteristics. The purpose of this analysis is to assess the relationships between transit and NMT and the aforementioned independent variables, forgoing self-selection effects that may exist. For each travel behavior outcomes (transit use and NMT for "typical" commute, transit use and NMT for commuting based on travel survey, and transit use and NMT for discretionary, non-work trips based on travel survey), I fit six models to explore various specifications of urban form. The first uses residential population density as the only urban form metric. The second adds walkability and transit availability, represented by intersection density and by number of transit stops within a 0.25 mile buffer respectively. The third adds urban form variables; walkability and transit availability at the workplace. The fourth adds remaining GIS-generated urban form metrics such as land-use diversity (entropy index), employment density, housing unit and rental unit density, and distance to the nearest employment sub-center (taken together, these variables represent job-housing balance). The fifth model uses Transit Score and Walk Score to represent transit availability and walkability. The sixth uses extracts from Exploratory Factor Analysis, which will be explained below, to represent urban form. Table 3 presents a summary on how these reduced form models are specified.

Table 4-4 Specifications of reduced form models

		Walkability	Transit Access
(1)	PopDens Population Density only	n/a	n/a
(2)	SimpleResUFs Simple UF GIS variables	Intersection Density	Transit Stops
(3)	WorkUFs (2) + Workplace UF GIS vars	Workplace Intersection Density	Workplace Transit Stops
(4)	All_UFs (3) + All other UF GIS variables	All UF GIS variables: All the 3-Ds	Transit Stops
(5)	transWalkScore Transit Score® and Walk Score®	Residential and Workplace Walk Score (from API)	Residential and Workplace Transit Score (from API)
(6)	EFA_extracts Density & Street Connectivity factors	Street Connectivity factor (EFA)	Transit Stops

4.4.2 Conceptual path model

Following Dillon, Saphores & Boarnet (2015), the conceptual path model (Figure 4-1) accounts for residential self-selection and vehicle choice endogeneity by specifying a system of recursive simultaneous equations where all causal effects are directed at the travel behavior outcome of interest: transit use and NMT. To circumvent biases stemming from endogeneities, structural equation modeling (SEM) techniques has the advantage of explicitly parametrizing endogenous relationships (Golob, 2009). Residential urban form and vehicle-to-licensed driver ratio (measures vehicle availability at the household level) are specified as endogenous variables; they are assumed to be influenced by household socio-economic characteristics.

In addition, vehicle choice is also assumed to be influenced by transit availability and street connectivity at the workplace. It is assumed that, everything else equal, if the workplace can be conveniently accessed without driving, households might forego car ownership. Consequently, this conceptual model implies that households make residential and vehicle ownership choices before making commuting-mode decisions, based on their socio-economic characteristics and transit access and walkability at the workplace.

While SEM can circumvent endogeneity problems, linear specification of SEM is not suitable to explain the binary dependent variable Y_i (which equals 1 if household "i" commutes using transit and/or NMT, or 0 otherwise). To handle non-continuous dependent variables, Generalized Structural Equation Modeling (GSEM) is selected to estimate these models.

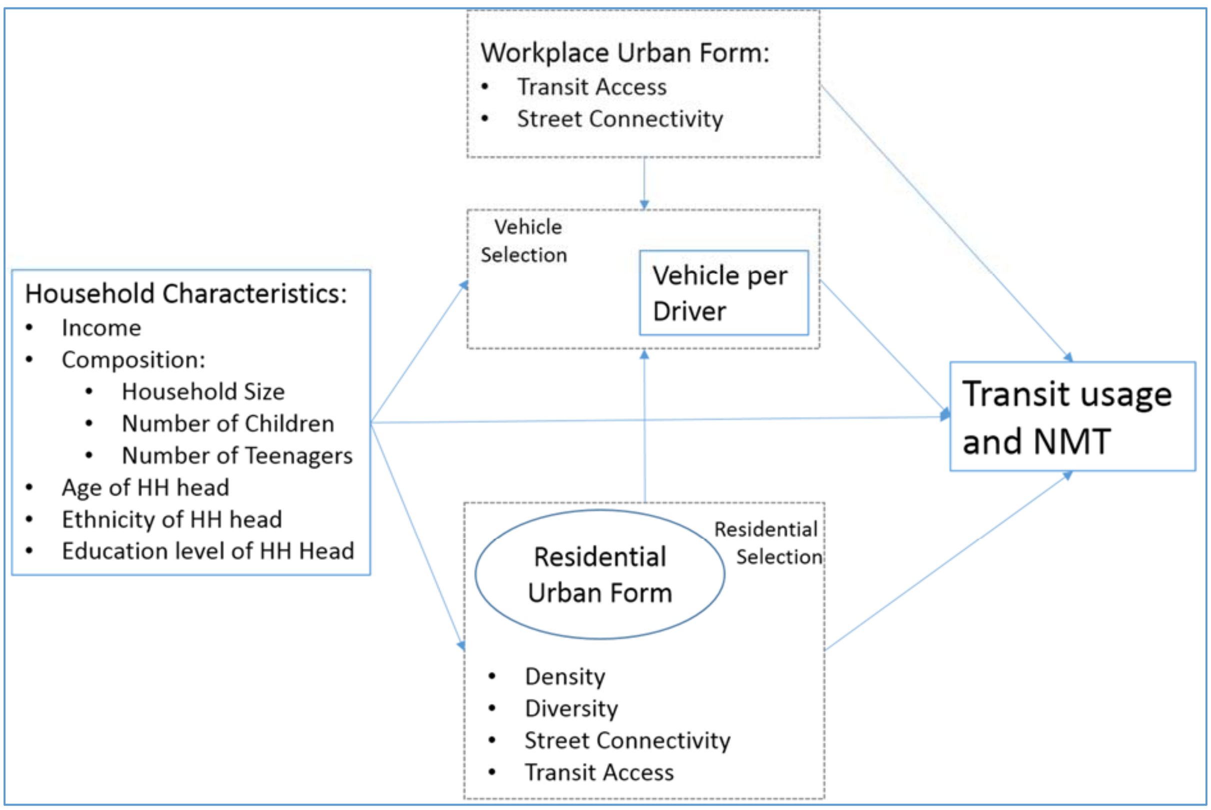

Figure 4-1 Conceptual Path Model

4.4.3 Urban form modeling approaches

Previous studies (Boarnet & Crane, 2001; Cervero & Kockelman, 1997; Ewing & Cervero, 2001, 2010; Ewing *et al*., 2009; Khan, Kockelman & Xiong, 2014) indicated that the configuration of urban form influences travel behavior (trip length, trip frequency, mode choice). By no means axiomatic, three aspects of urban form that could influence travel behavior include at least three categories: density (i.e. population density, residential unit density), diversity (i.e. land use mix), and design (i.e. street-network characteristics, connectivity).

In this paper, I examine five distinct approaches to modeling the effect of urban form on transit usage and NMT, accounting for self-selection effects. The first approach specifies residential urban form with standardized population density only. This approach is similar to Kim & Brownstone (2013) and Brownstone & Golob (2009). The second approach specifies urban form with all measures of land use, including measures of density (population, employment, housing unit, and renter-occupied homes), land use diversity, transit access, and street connectivity (walkability). This approach is similar to Frank *et al.* (2005); Cao, Handy & Moktharian (2006); Voorhees *et al.* (2010); Manaugh & El-Geneidy (2011); Giles-Corti *et al.* (2011); and Khan, Kockelman & Xiong (2014).

The third used Exploratory Factor Analysis (EFA) to extract latent factors (or urban form dimensions) from the aforementioned measures of land use. Measures of land use that do not factor with other variables are entered into the model as observed variables. This approach is similar to Cervero & Kockelman (1997) and Cervero & Duncan (2003). The fourth model used Walk Score® and Transit Score® data to represent urban form. Walk Score and Transit Score use a patented algorithm to calculate pedestrian friendliness and how well a location is served by transit respectively. These measures are normalized to a score between 0 (poor) to 100 (excellent). Lastly, the fifth model specifies urban form as a single latent factor which includes population density, employment density, transit access, land use diversity, and measures of walkability. The latent factor is jointly-estimated using Confirmatory Factor Analysis. To my knowledge, this approach has not been used in the context of transit usage and NMT.

4.4.3.1 Urban form as a latent construct

The treatment of residential urban form as an underlying latent construct (in the third, fourth, and fifth approaches) merits explanation. Commute typically does not generate utility by itself, but from the need of the commuter to arrive at work locations where employment takes place. Since home based trips originate from residences, the surrounding land use patterns capture the configuration of potential activities that could influence travel behavior.

The extent to which surrounding land use has the potential to influence travel behavior is abstract, it has no natural scale and, in principle, is largely unobservable. How residential urban form may influence travel is not directly measurable, but it can be characterized by variables such as land use entropy, density, transit access, street connectivity (walkability) and distance from employment centers. These features fit the criteria for latent factors or constructs (Bollen, 1989, p. 180). Incorporating latent factors or constructs in SEM (and, by extension, GSEM) has been critiqued but Bollen (1989, p. 78) have addressed them elegantly. Abstract constructs have been a part of scientific theories in many disciplines including quantum physics, biology, medicine, psychology, sociology, and thermodynamics.

In essence, Walk Score and Transit Score, used in the fourth approach, operate within the latent construct framework. These scores are (normalized) indices that has no natural scale that seek to explain the extent to which a point location is walkable and served by transit. In theory, Walk Score should be associated with the extent to which a location facilitates walking; a location with higher Walk Score should observe higher levels of walking. The same should also apply to Transit Score – higher scores promise higher transit use. However, since these scores are computed by an external proprietary algorithm, the researcher/user does not have control over the data quality (Duncan *et al.*, 2013).

4.4.3.2 Overview of factor-analytic methods

Factor-analytic approaches have been used to investigate the relationship between sets of observed variables and the urban form latent construct (Cervero & Kockelman, 1997; Cervero & Duncan, 2003). This approach examines the covariation among a set of observed urban form variables to identify the existence of underlying latent constructs (or 'dimensions' thereof). There are two major types of factor analyses in the statistical toolbox of social science: Exploratory Factor Analysis (EFA), used in the third approach, and Confirmatory Factor Analysis (CFA), in the fifth approach. This section will give a brief discussion of each and explain why CFA may be preferred.

EFA is most helpful when the researcher does not have any prior information about the link between the observed and the latent variables (Byrne, 2001, pp. 5-6). EFA helps researcher to explore how and to what extent the observed variables are linked to their underlying latent constructs. A recurring central problem in the application of EFA is deciding the number of unobserved factors/latent constructs that can be used to explain the co-variability in a set of observed variables (Byrne, 2001; Preacher *et al.*, 2013). In practice, the factor model is used to represent observed/measured variables as a linear function of model parameters and unobserved factors/latent constructs.

In numerous cases, the number-of-factors problem, which is inherently a model selection issue, should be considered (Preacher *et al.*, 2013, p. 29). Model selection in general is the practice of evaluating a set of competing theoretical explanations, given the observed data. Models in the social sciences are parametrical representation of competing theories, which can be tested against the observed data using statistical analyses. There are many perspectives on what constitutes a desirable model, but typically it is one that "balances" parsimony and goodness of fit, the ability of a model to capture the data-generating process.

Preacher *et al.* (2013) cautions against the use of EFA to find the "true" number of factors, which is the equivalent of using statistical techniques to find the "one true model". Many methodologists (see Preacher *et al.*, 2013, pp. 30-32) have argued that in most circumstances "there is no true operating model". The hypothetical true model would likely be overly complex to interpret and would not be stable temporally. It would capture the data-generating process only at the time data are observed. Thus all models suffer from misspecification and that the best a researcher can expect from a model is that it provide a useful approximation of the data-generating process.

However, this perspective does not mean that EFA is a pointless vocation. A researcher can still find and retain an 'optimal' number of factors; one that satisfies a given criterion in service of meeting some well-defined scientific goal. In the context of this paper, the goal is to identify theoretically-justifiable model with the highest verisimilitude (proximity to the 'objective truth'), and one that stresses generalizability, the ability to replicate and to cross-validate data that is generated from the same underlying process.

Preacher *et al.* (2013) and Bollen (1989, p. 230-232) argue that eigenvalue-based criteria are too arbitrary to determine the 'optimal' number of factors because it lacks the structure of formal hypothesis testing. To overcome this problem, Preacher *et al.* (2013) suggested RMSEA, the root mean square error approximation, which allows formal hypothesis testing, as a factor retention criterion. In practice, RMSEA is widely used in Confirmatory Factor Analysis (CFA).

Some EFA techniques allow researchers to pre-specify the number of factors and therefore blurs the definition of 'exploratory'.[8] Another limitation of EFA is the inability to explicitly account for correlated measurement errors, which confounds these errors with the latent constructs, potentially leading to ambiguous solutions (Bollen, 1989, p. 232). Lastly, Preacher *et al.* (2013) emphasizes that model selection is not intended to find the true model but rather is intended to identify a parsimonious model that gives reasonable fit. In the context of factor analysis, this strategy involves identifying a theoretically plausible number of factors before data are collected – which leads to CFA.

In contrast to EFA, CFA is appropriate when the researcher has prior knowledge about the underlying structure of the latent constructs (Bollen, 1989; Sorbom, 1989; Bryne, 2001; Kline, 2005), as is the case in this paper. CFA overcomes the shortcomings of EFA in accommodating theoretical and substantive knowledge. Unlike EFA, a researcher can formally test a theoretically-informed model. Once a model is specified, it can be estimated, and goodness of fit can be assessed using theoretically justifiable criterions.

Finally, I turn to algebraic means to explicitly illustrate the methodological differences between EFA and CFA. Given a set of observed variables (x_1, x_2,…, x_p), both EFA and CFA specify a model to represent these variables as functions of model parameters and latent factors as follows (Bollen, 1989, pp. 233-237; Preacher *et al.,* 2013, pp. 30-31):

[8] In reality, researchers begin with data collection in order to test competing theories – parameterized as statistical models. Therefore, EFA can never fully escape prior information.

$$x = \Lambda\xi + \delta \ , \qquad (2)$$

where **x** is a $p \times 1$ vector containing p-observed variables from an individual unit of the data, $\mathbf{\Lambda}$ is a $p \times m$ matrix of factor loadings relating to the p-observed variables to m latent factors, and ξ is an $m \times 1$ vector of latent factors. In this model, $\boldsymbol{\delta}$, a $p \times 1$ vector, is composed of two components: $\boldsymbol{\delta} = \mathbf{s} + \mathbf{e}$, where **s** represents the specific variance associated with each variable and e is the remaining random component in **x**. Since both are errors in **x** with respect to measuring ξ, the latent factor, and assuming that both are uncorrelated with ξ and with each other, $\boldsymbol{\delta}$ is the random errors of measurement. The covariance matrix for **x** can be written:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta_\delta \, , \qquad (3)$$

where $\mathbf{\Sigma}$ is the $p \times p$ population covariance matrix, $\mathbf{\Phi}$ is the covariance matrix of the latent factors ξ, and $\mathbf{\Theta}_{\boldsymbol{\delta}}$ is the covariance matrix for the errors of measurement $\boldsymbol{\delta}$. Parameters in $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\mathbf{\Theta}_{\boldsymbol{\delta}}$ are estimated using the observed data.

The primary interest of EFA is the estimation of factor loadings in $\mathbf{\Lambda}$. However, they are not uniquely identified, and the researcher will usually select a solution for $\mathbf{\Lambda}$ that maximizes some criterion of interpretability. The pattern of high and low factor loadings in $\mathbf{\Lambda}$ identifies groups of variables that depend on the same latent factor(s). In most applications of EFA, the most critical subjective decision is in retaining the 'optimal' number of m-latent factors that account for most of the observed covariation in **x** (i.e., 'estimating' the dimension of $\mathbf{\Lambda}$).

Both EFA and CFA assumes $E(\boldsymbol{\delta}) = 0$ and $E(\boldsymbol{\xi\delta}') = 0$. However, EFA constraints $\mathbf{\Theta}_{\boldsymbol{\delta}}$ to be a diagonal matrix (i.e., $\text{cov}(\delta_i,\delta_j) = 0$ for all $i, j \in \{1, 2, \ldots, p\}$ & $i \neq j$), whereas CFA does not (Sorbom, 1989, p. 380). In fact, CFA leaves open the specification of $\mathbf{\Theta}_{\boldsymbol{\delta}}$ to improve model fit (the ability to reproduce the observed covariance matrix $\mathbf{\Sigma}$), a point I will elaborate on in the next section. Given **x**, the researcher can fit 'competing' theoretically informed CFA models by changing the specification of $\mathbf{\Theta_\delta}$. This is how CFA practice blurs the definition of 'confirmatory'.

In summary, EFA and CFA differs substantially on the structure of $\mathbf{\Theta_\delta}$. In practice, EFA results can inform the researcher about the optimal number of *m*-latent factors that can best explain the data, while CFA can 'explore' specification of $\mathbf{\Theta_\delta}$ to achieve the same modeling goals.

4.4.3.3 Specification issues in Confirmatory Factor Analysis

In a full GSEM, in which CFA is incorporated as a measurement sub-model, residential urban form is specified as a latent construct that captures the extent to which a residential neighborhood may influence transit usage and NMT. The measurement sub-model and conceptualizing urban form as a construct have been under-explored in the travel behavior literature (Golob, 2003; Van Acker *et al.*, 2007; van Wee, 2009). Omitting a measurement sub-model implicitly assumes that observed variables are measured without errors.

In general terms, factor analysis is an approach to detect the existence of an unobserved factor/underlying construct using multiple observable indicators. CFA in the fifth approach is performed to examine the viability a theoretically informed model. I specified a simple CFA measurement model based on the literature and EFA results without specifying any measurement error correlations, and thus restricting it to zero. Model fit in CFA refers to the ability of a model to reproduce the observed variance-covariance matrix. The chi-square (χ^2) statistic measures whether the observed covariance matrix matches the model covariance matrix: if the χ^2 statistic is not significant (i.e., value is closer to zero), fit is deemed acceptable.

Upon review of the model fit, modification of the model can be done by estimating an extra parameter. This can be implemented by allowing measurement errors to be correlated (i.e., re-specifying $\mathbf{\Theta_\delta}$), as long as they are identified, and by entering additional information (urban form variables) into the model. These re-specification steps are permitted to the extent that they are done judiciously (and satisfy statistical criterion) and in a theoretically defensible fashion (Bollen, 1989, pp. 233–235, 296–305; Sorbom, 1989; Kline, 2005, pp. 176–191; Acock, 2013, pp. 26-29). Sorbom (1989) proposed using modification index, a statistical score based on Langrage multiplier tests, for detecting model misspecification and provided guidelines for CFA model re-specification (p. 384).[9] This approach estimates how much χ^2, (measure of 'badness' of model fit), will be reduced when an extra parameter is estimated in the model (adding one degree of freedom). Based on these scores, a CFA analyst should consider adding a parameter that will improve model fit the most, so long as it is theoretically defensible. Sorbom (1989, p. 384) and Acock (2013, p. 27) suggest adding one parameter at a time since modification indices are not additive.

The CFA measurement sub-model reflects the complex nature of the relationship between travel behavior and urban form. Whereas the parsimonious model restricts the covariance between land use indicators to zero, the CFA model allows measurement errors for population density, residential density, land use entropy (diversity), employment density, proportion of renter occupied housing units, transit access, street connectivity, and distance from employment sub-centers to co-vary. The basic idea of correlated measurement errors is that

[9] Modification indices are tests for the statistical significance of the omitted paths in the CFA model. In Stata (version 13 and up), this technique can be implemented by the modification indices postestimation command (estat mindices) (Acock, 2013, pp. 26-27).

places with high population density tend to have a more diverse land use mix, employment density and transit access.

4.4.4 GSEM specification

This essay compares five modeling approaches that share a common conceptual model, but vary by how near-residence and near-workplace environments are measured and specified. The first four approaches do not include a latent factor measurement component. They only have a structural component. The CFA specification is comprised of a structural component (that specifies the influence of urban form, vehicle ratio, and household characteristics on active commuting) and a latent factor measurement component (that links observed urban form variables to a latent factor).

The structural component models the relationship between endogenous variables using a system of simultaneous equations (or conceptual path model) where the direction of causality is explicitly specified (Bollen, 1989; Kline, 2005; Acock, 2013). In a graphical representation (Figure 1), hypothesized causal paths (symbolized by arrows) depart from exogenous variables towards endogenous variables. The measurement component accounts for how the latent factor (residential urban form) is measured by observed variables; it is jointly estimated with the structural component via Confirmatory Factor Analysis (CFA) using quasi-maximum likelihood.

The GSEM model (the fifth and most complex modeling approach) can be written:

$$Pr(Y_i = 1|\tau_i) = \frac{exp(\tau_i \theta)}{1 + exp(\tau_i \theta)} \tag{4}$$

where:

$$
\begin{cases}
\boldsymbol{\tau} = \beta_{11}\mathbf{v} + \beta_{12}\mathbf{f} + \Gamma_1\mathbf{X} + \Delta_1\mathbf{W} + \varepsilon_1, & \text{(5A)} \\
\mathbf{v} = \beta_{22}\mathbf{f} + \Gamma_2\mathbf{X} + \Delta_2\mathbf{W} + \varepsilon_2, & \text{(5B)} \\
\mathbf{f} = \Gamma_3\mathbf{X} + \varepsilon_3, & \text{(5C)} \\
\mathbf{u} = \Lambda_u\mathbf{f} + \varphi, & \text{(5D)}
\end{cases}
$$

In the above:

- $\boldsymbol{\tau}$ is an n×1 vector of latent (unobserved) continuous variables that reflect the utility of active commuting for each household;
- **v** is an n×1 vector of household vehicle ratio;
- **f** is an n×1 vector of urban form latent variables;
- **X** is an n×k matrix of k household explanatory variables;
- **W** is an n×7 matrix of workplace transit availability and street connectivity variables;
- $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, and $\boldsymbol{\Gamma}_3$ are n×nk matrices. $\boldsymbol{\Gamma}_j = \left(\gamma_{j1}\mathbf{I_n} \ \ldots \ \gamma_{jk}\mathbf{I_n}\right)$ for $j \in \{1,2,3\}$, with $\gamma_{ji} \in \Re$ the γ_{ji}s are unknown coefficients to estimate, and $\mathbf{I}_n$ is the n×n identity matrix; I imposed additional restrictions on the $\boldsymbol{\Gamma}_j$s to over-identify the model: like Brownstone and Golob (2009) and Dillon, Saphores and Boarnet (2015), non-significant household explanatory variables were removed from the models;
- $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ are n×n7 matrices. $\boldsymbol{\Delta}_j = \left(\delta_{j1}\mathbf{I_n} \ \ldots \ \delta_{j7}\mathbf{I_n}\right)$ for $j \in \{1,2\}$, with $\delta_{ji} \in \Re$ the δ_{ji} are unknown coefficients to estimate;
- **u** is a (10n)×1 vector of stacked variables for residential urban form, which includes population density, housing unit density, employment density, % of renter occupied

housing, land use entropy, inverse distance to the nearest sub-center[10], transit availability, bikeways density, street density, and intersection density;

- Λ_u is an (10n)×n matrix given by $\mathbf{\Lambda}_u = (\lambda_1 \mathbf{I_n} \quad \ldots \quad \lambda_{10}\mathbf{I_n})$, with $\lambda_i \in \Re$ ($i \in \{1,\ldots,10\}$);
- The ε_is for $i\in\{1,2,3\}$ and $\boldsymbol{\varphi}$ are respectively n×1 and 10n×1 error vectors assumed to be uncorrelated; and
- β_{11}, β_{12}, β_{22}, γ_{ji} ($j\in\{1,2,3\}$, $i\in\{1,\ldots,k\}$, in $\mathbf{\Gamma}$), δ_{ij} ($j\in\{1,2\}$, $i\in\{1,\ldots,7\}$, in $\mathbf{\Delta}$) and λ_i ($i\in\{1,\ldots,10\}$ in $\mathbf{\Lambda}_u$) are unknown model parameters to estimate; like SEM, GSEM minimizes the difference between the sample covariance and the covariance predicted by the model (Bollen, 1989).

The system of equations (5A-C), which is shared by all five modeling approaches, reflects the causal paths shown in Figure 1: in equation (5C), near-residence urban form (**f**) is explained by household socio-economic characteristics—explicitly parameterizing residential selection; in equation (5B), vehicle-to-driver ratio (**v**) is explained by residential urban form, household socio-economic characteristics, and workplace transit access and walkability—explicitly parameterizing vehicle endogeneity; and in equation (5A), transit usage and NMT ($\boldsymbol{\tau}$) is explained by residential urban form, vehicle-to-driver ratio, household socio-economic characteristics, and near-workplace transit access and walkability.

In the first approach, near-residence urban form (**f**) only contains population density. Near-workplace transit availability and walkability are represented by transit stops within ¼ miles and the gamma-connectivity index. In the second approach, all urban form variables enter as observed variables. Near-residence transit availability is represented by transit stops within ¼

[10] The distance to sub-centers accounts for the polycentric nature of Southern California's urban structure, which has multiple employment centers. For detailed explanations of how sub-centers are determined, see Lee (2007) who analyzed sub-centers in Los Angeles and other major metropolitan areas.

miles and near-transit walkability is represented by gamma-connectivity index. In the third, density and walkability are represented by EFA-extracts. In the fourth, transit access and walkability are represented by Transit Score® and Walk Score® respectively.

Equation (5D), the measurement component, defines residential urban form as a latent construct that depends on population density, housing unit density, employment density, % of renter occupied housing, inverse distance to the nearest sub-center, land use entropy, transit availability, bikeways density, street density, and intersection density. These variables are stored in the vector **u**. Moreover, $\mathbf{\Lambda}_u$ is a matrix of measurement coefficients (loadings) obtained by confirmatory factor analysis (CFA).

It follows that $\boldsymbol{\tau}$, **v**, and **f** are endogenous variables while the **X** and **W** matrices are exogenous socio-economic characteristics and the workplace urban form variables respectively (workplace location is assumed to be exogenous). Therefore, I have a recursive system with all causality paths directed at active commuting (Figure 1); since ε_is are also assumed to be uncorrelated, the model is guaranteed to be identified (Bollen, 1989, pp. 95–98; Kline, 2005, pp. 105–107). In GSEM, as in SEM, identification requires at least as many observations as free model parameters (df $\geq$ 0) and every latent factor must be assigned a scale or restrict the variance to one (Bollen, 1989, pp. 103–104; Kline, 2005, p. 105). Unknown model parameters can then be estimated by minimizing the difference between observed covariance and the model-predicted covariance (Bollen, 1989; Kline, 2005). Table 4-5 presents a summary on how these GSEM urban form approaches are specified.

Table 4-5 Urban form modeling approaches using GSEM estimation

Modelling Approach	Density	Transit Access	Walkability
(1)	Population density	n/a	n/a
(2)	Population density	Transit stops (¼ miles)	Gamma Index
(3)	Density factor extract	Transit stops (¼ miles)	Connectivity factor extract
(4)	Population density	Transit Score (from API)	Walk Score (from API)
(5)	Urban form latent construct – jointly estimated as a CFA measurement (sub)component		

4.5 Results

As noted earlier, the objectives of the research are (1) to assess several methods for measuring urban form features in the near-residence and near-workplace environments and (2) to assess the importance of these urban form features on transit use and NMT after accounting for the influence of these features on household vehicle ownership and residential selection.

To achieve the first objective, I used factor analytic approaches, EFA and CFA, to help elucidate underlying dimensions of urban form. Reduced form logit models were fitted to explore the influence of urban form and household's socio-demographic characteristics on the probability of transit usage and NMT, ignoring endogenous self-selection effects as a starting point for a more complex modeling approach. The second objective is met by fitting a recursive Generalized Structural Equations Model (GSEM) under five different methods in representing urban form as explained in the preceding section. Before I compare results from these five GSEM models, I will first report results from reduced EFA, CFA, and reduced form results.

4.5.1 Exploratory Factor Analysis results

As described above, EFA helps elucidate the underlying unmeasurable dimensions of urban form. To assess the adequacy of these dimensions using a standard approach I first used Bartlett's test of sphericity to check for the appropriate level of inter-correlation between the urban form measures. Inter-correlations have to be sufficiently high to limit the number of factors, but not too high to avoid multicollinearity, which I detect using the Kaiser-Meyer-Olkin (KMO) statistic – a measure of sampling adequacy. I also used Cronbach's alpha to measure the reliability of my factors. For the EFA model to work well, Bartlett's test should reject the null hypothesis that the correlation matrix is an identity matrix. The KMO statistic (which ranges between 0 and 1, with small values suggesting that the variables do not have enough in common) should be larger than 0.6 to be desirable. Finally, Cronbach's alpha (which has a maximum value of 1) has been suggested to be at least 0.6 (Sangkapichai & Saphores, 2009, p. 86).[11]

Based on principal factor analysis (the default in Stata), two interpretable factors, walkability and density, were extracted based on the input of 10 urban form variables. To simplify interpretation, Table 4-6 lists EFA results in order of their factor loadings, first on walkability, then on density. From the loadings, it is clear that the first factor, which accounts for 36% of the total variation, represents walkability. The second factor, density, accounts for 25.6% of the total variation – together walkability and density factors explained 61.6% of variation of 10 variables. Three variables that don't factor includes land use entropy (diversity index), street density, and bikeways density.

Table 4-6 Factor loadings for walkability and density

[11] Methodologists agree that there is no gold standard as to how high alpha reliability coefficient should be and that coefficients may not be generalizable to a particular sample. Kline (2005) suggests that 0.7 is adequate and is silent about values lower than that (p. 59).

	Factor Loadings on	
	Walkability Factor	Density Factor
Alpha index	0.9812	
Gamma index	0.9774	
Link/Node Ratio (Beta Index)	0.9813	
Intersection density	0.7077	
Population density (10,000 person/mile^2)		0.7562
Housing units per square mile (thousands)		0.8477
Employment per square mile (thousands)		0.6344
Percentage of renter occupied housing		0.8078
Inverse distance to nearest sub-center (in 1/miles)		0.4390
Transit service within 0.25 miles		0.4635
Variance explained (using oblique rotation)	36.0%	25.6%

Note: KMO statistic is 0.79 and the Bartlett's test was highly significant ($p<0.0001$). Cronbach's Alpha is 0.86 and 0.6 for Walkability and Density respectively.

Factor analysis results provide a multi-variable description of the two hypothesized underlying dimensions of residential urban form: design (walkability or street connectivity) and density. The extracted factors and their link to observed urban form variables are logical and interpretable. Finally, oblique rotation accounted for intercorrelation between the two dimensions of urban form; highly walkable residential locations are also dense. The next section reports findings from Confirmatory Factor Analysis (CFA), using EFA results as a starting point for model specification.

4.5.2 Confirmatory factor analysis results

Before comparing GSEM results, I will discuss results from my CFA measurement model for urban form. First, I will discuss model fit (Table 4-7) before interpreting the loading coefficients. Model fit in CFA refers to the ability of a model to reproduce the observed variance-covariance matrix. The chi-square (χ^2) statistic measures whether the observed covariance matrix matches the model covariance matrix: if the χ^2 statistic is not significant, fit is deemed acceptable. A known disadvantage of the χ^2 statistic is that it decreases with additional model parameters, which favors over-parametrized models (Schermelleh-Engel *et al.*, 2003, p. 33).

To avoid this shortcoming, incremental fit indices, which are not sensitive to sample size and penalize for over-parameterization, have been developed. I report here the Comparative Fit Index (CFI), which has been shown to have good power and robustness (Iacobucci, 2010, p. 97). It ranges from 0 (worst) to 1 (best), where values ≥0.97 indicate good fit and values between 0.95 and 0.97 are acceptable (Schermelleh-Engel *et al.*, 2003, p. 42). I also report the Tucker-Lewis Index (TLI), which can be interpreted like the CFI (Schermelleh-Engel *et al.*, 2003, p. 41).

Finally, I report the root mean square error of approximation (RMSEA), which estimates the amount of error of approximation per model degree of freedom while taking sample size into account. A RMSEA value of zero implies exact fit, a value under 0.03 suggests excellent fit, and a value between 0.05 and 0.03 is considered good (Kline, 2005, p. 139). Table 5 shows that my residential urban form measurement model with correlated errors performs well, compared to my starting model without correlated errors, because it satisfies all five fit criteria considered.

I used EFA results as a starting point for the CFA model specification and fitting a model for each of the underlying urban form dimensions (walkability and density). I followed the steps recommended by SEM literature to improve model fit as explained in the Methodology section (Bollen, 1989, pp. 233–235, 296–305; Sorbom, 1989; Kline, 2005, pp. 176–191; Acock, 2013, pp. 26-29). Unlike EFA, CFA allows measurement errors to be intercorrelated. This feature is useful to 'confirm' one important findings from EFA that the two underlying dimensions of urban form are indeed correlated. Table 4-7 reports five CFA variable specifications (models A-E), both with and without correlated errors.

Comparison of fit indices between models with and without correlated errors offers further evidence that urban form variables are intercorrelated. Moreover, comparison of fit indices across variable specifications suggests that urban form has a strong statistical relationship with density (population, housing unit, percent of rental unit, employment, distance to employment centers, and transit service), diversity (land use entropy), and design/street connectivity (street density, bikeway density, and alpha index) variables, as described in model D with correlated errors.

Table 4-7 Fit Statistics for Residential Urban Form Measurement Model (CFA)

	χ^2 [df] (p-value)[a]	χ^2/df[a]	**CFI[b]**	**TLI[c]**	**RMSEA[d]**
Critical values	p> 0.05	< 3	> 0.95	> 0.95	< 0.05
A. Density variables only					
Model without correlated errors	328.089 [9] (<0.001)	36.4544	0.9096	0.8493	0.1275
Model with correlated errors	10.216 [5] (0.069)	2.0433	0.9985	0.9956	0.0219
B. Density Variables and Diversity					
Model without correlated errors	624.706 [14] (<0.001)	44.6219	0.8401	0.7602	0.1414
Model with correlated errors	10.897 [6] (0.092)	1.8162	0.9987	0.9955	0.0193
C. Density, Diversity, and Bikeway Density					
Model without correlated errors	674.580 [20] (<0.001)	33.7290	0.8312	0.7637	0.1225
Model with correlated errors	17.167 [9] (0.046)	1.9075	0.9979	0.9934	0.0204
D. Density, Diversity, Bikeway Density, and Street Connectivity					
Model without correlated errors	1,430.805 [35] (<0.001)	40.8801	0.7524	0.6817	0.1352
Model with correlated errors	20.482 [14] (0.116)	1.4630	0.9989	0.9963	0.0146
E. Density, Diversity, and Street Connectivity (without Bikeway)					
Model without correlated errors	1,291.950 [27] (<0.001)	47.8500	0.7697	0.6929	0.1465
Model with correlated errors	19.1562 [11] (0.058)	1.7415	0.9985	0.9951	0.0184

Notes:

[a.] The chi-square (χ^2) statistic or chi-square goodness of fit, measures whether the observed covariance matrix is similar to the covariance matrix predicted by the model: if it is not significant, the model is regarded as acceptable. With small samples, the χ^2 statistic lacks power so some researchers have proposed the relative chi-square (χ^2 divided by the number of degrees of freedom, denoted here by χ^2/df). This fit index reflects the 'centeredness' of the chi-square.

[b.] The Comparative Fit Index (CFI) is an incremental fit index, which has been shown to have good power and robustness (Iacobucci, 2010, p. 97). It ranges from zero (worst fit) to one (best fit), where values above 0.97 indicate good fit and values between 0.95 and 0.97 denote acceptable fit (Schermelleh-Engel *et al*., 2003, p. 42).

[c.] The Tucker-Lewis Index (TLI) compares χ^2/df of the proposed model to χ^2/df for a null model; it is interpreted like the CFI (Schermelleh-Engel *et al*., 2003, p. 41).

[d.] The root mean square error of approximation (RMSEA) estimates the amount of error of approximation per model degree of freedom while taking sample size into account. A RMSEA value under 0.03 suggests excellent fit, and a value between 0.05 and 0.03 is considered good (Kline, 2005, p. 139).

Standardized loadings (regression coefficients when variance of the latent variable is constrained to be 1) from model D are presented in Table 4-8. All 10 urban form variables have highly significant loading coefficients (*p*<0.001) and, except for land use entropy and bikeway density, are positively correlated with urban form. Higher densities (population, residential, and employment) with more renters, higher transit availability, and street connectivity all contribute to a higher value of the urban form latent variable. A higher entropy value corresponds to higher land use diversity, which might preclude very large densities and explain the negative loading coefficient. Likewise, neighborhoods with higher densities may have less space for bikeways. The negative loading coefficient can be interpreted as a 'correction' to the contribution of densities to urban form.

Table 4-8 Results of measurement model for urban form

Urban form observed variables	Standardized Loadings	R^2
Population density	0.5873***	0.3449
Residential density	0.7581***	0.5747
Employment density	0.4808***	0.2312
Percentage renter occupied housing	0.6518***	0.4248
Inverse distance to employment sub-center	0.3327***	0.1107
Transit service within 0.25 mile	0.3490***	0.1218
Street density	0.3932***	0.1546
Alpha index	0.6518***	0.4249
Land Use Entropy	-0.0633***	0.0040
Bikeway density	-0.1615***	0.0261

Note: all standard loadings have p-values <0.01. R^2 displays the Bentler-Raykov squared multiple-correlation coefficient (Bentler & Raykov, 2000).

4.5.3 Reduced form logistic Results: comparison of urban form specifications

Reduced form logistic regression models were specified to examine the influence of urban form and household's socio-demographic characteristics on the probability of transit usage and NMT. The purpose of this analysis is to assess relationships between transit usage and NMT and the aforementioned independent variables, forgoing structural relationships that may exist. Results are useful as a starting point for a more rigorous and complex analysis.

I developed a transit and NMT models for each of the following measures of travel behavior outcomes (results are provided in the appendix):

- Probability of Indicating a Typical Commute Mode (Table A.1.1 reports transit models, Table A.1.2 reports NMT models)
- Probability of Using a Mode for Commute Trip, (Based on Trip Diary) (Table A.2.1 reports transit models, Table A.2.2 reports NMT models)
- Probability of Using a Mode for Non-Work Trips, (Based on Trip Diary) (Table A.3.1 reports transit models, Table A.3.2 reports NMT models)

For each of these behavioral outcomes by mode, I generated the following six models to assess the influence of different urban form specifications on the probability of transit usage or NMT (Table 4-4): (1) Residential population density as the only urban form metric, (2) #1 plus residential walkability (intersection density) and transit access (transit stops), (3) #2 plus workplace walkability (intersection density) and transit access (transit stops), (4) #3 plus all other UF variables representing the 3-Ds, (5), #1 plus residential and workplace Walk Score and Transit Score, and (6) EFA-based residential Density and Street Connectivity Factors plus residential transit access (transit stops) and workplace walkability (intersection density) and transit access (transit stops). Before describing and comparing coefficient estimates, I address model performance.

4.5.3.1 Model fit

I rely on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to mechanistically assess model fit. Both try to balance good fit with parsimony using penalized-likelihood criteria, where BIC penalizes complexity more heavily, depending on sample size. AIC's penalty for complexity, on the other hand, does not depend on sample size. AIC and BIC, where smaller values are better, usually agree on which model performs better in explaining the data. However, when it div has a chance of choosing too big a model, regardless of sample size. Since power of the test is increasing in sample size, AIC, then, should be considered when false positive (Type I error) is more acceptable than a false negative.

In explaining transit use for typical commute mode (Table A.1.1), model (6) which specifies urban form with EFA extracts, reports the lowest AIC value (BIC ranks 3rd). Model (1) using density as the only urban form metric does not perform well despite its simplicity (BIC ranks 4th, AIC ranks 5th). Model (3) which adds workplace urban form in explaining transit commute seems to fair rather well in AIC (ranks 2nd) and BIC (lowest).

In explaining NMT commuting (walking or biking), under strictly mechanistic model selection, AIC favors model (5) which uses Walk Score and Transit Score at the residential neighborhood and at the workplace. BIC, on the other hand, favors the simplest model (1). In explaining transit commuting based on the travel survey data, BIC favors the simplest model (1), while AIC favors model (3) where workplace urban form enter as explanatory variables. In explaining active commuting based on travel survey data, BIC favors the simplest model (1), while AIC favors model (5) with Walk Score and Transit Score. However, under strict mechanistic interpretation, both AIC and BIC favor the simplest model when explaining discretionary trips by transit and active travel. This may suggest that discretionary travel are dominated by household's socio-demographic status.

4.5.3.2 Socio-demographic variables

A consistent predictor of transit usage and NMT for commuting and for discretionary trips is vehicle availability. Across all modes reported in the appendix, a higher number of vehicles per adults was associated with a lower likelihood of using these modes. This is largely consistent with the literature (Kuzmyak & Dill, 2012).

The coefficient of income is negative and statistically significant in explaining typical commute by transit (Table A.1.1), but not according to the models estimated on travel diary data (Table A.2.1). While household size is consistently positive, the number of children is consistently negative. These findings suggest that, first, consistent with Kahn & Matsusaka (1997), transit is an inferior commuting mode as higher income households are associated with lower likelihood of transit commuting. Second, based on typical commute data (Table A.1.1), larger households are more likely to use transit for commuting but the number of children (younger than 16) reduces the likelihood transit commuting. However, the travel diary data (Table A.2.1) does not seem to agree. Interestingly, Hispanic households stand out among its socio-demographic peers to be more likely to use transit, both for commuting and discretionary trips, but not for typical commute by NMT (Table A.1.2) and discretionary NMT trips (Table A.3.2) travel.

4.5.3.3 Urban form variables

Comparison of coefficient estimates on urban form variables and how they vary from model to model offer interesting insights. In terms of transit usage (Tables 1.1.1, 1.2.1 and 1.3.1), workplace, but not residential, transit accessibility (transit stops) is positively associated with transit usage for the commute, but not non-work discretionary trips. Transit Score, however, is significant and positively associated with the probability of reporting that transit was the "typical" commute mode (Table 1.1.1) but it is not significantly associated with the probability of actual transit usage based on the travel diary for the work commute (Table 1.2.1) or for transit usage for discretionary trips (Table 1.3.1). Although Houston *et al.* (2015) and Spears, Boarnet, and Houston (2015), suggest a positive association between residential transit availability and transit ridership, these previous studies did not assess the influence of near-workplace transit access. It is surprising that, after controlling for near-workplace transit access, near-resident transit access is no longer significant. This may suggests that near-workplace transit access dominates the effect on household's commuting choices.

Surprisingly, proximity to employment centers are significantly negative in explaining typical commute mode by transit (Table A.1.1), and it is not significant in explaining other travels. When transit availability is specified using Transit Score, population density is no longer statistically significant. EFA extract for density (under Model 6) is statistically significant and positive in explaining typical commute mode (Tables A.1.1. and A.1.2) and discretionary transit and NMT trips (Tables A.3.1 & A.3.2). This may be a clue that the algorithm responsible to generate Transit Score have considered population density such that it dominates the explanatory variable of population density as a stand-alone variable.

In terms of NMT, urban form variables do not explain typical NMT commuting by much. Only population density is positively significant on discretionary NMT trips (Table A.3.2). Only intersection density in the workplace is estimated to be marginally significant on NMT commuting, both for typical commute and travel diary data (Tables A.1.2 & A.2.2). Results from travel diary data are different: Workplace Walk Score is significantly positive but workplace Transit Score is marginally significant and negative. Perhaps the negative sign suggest a correction to Walk Score (Table A.3.2). Surprisingly, EFA extracts for near-residence walkability/street-connectivity does not seem to have any explanatory power in any of the models. Near-residence Walk Score is positive and significant only for discretionary NMT (Table A.3.2), perhaps because the amenities considered by Walk Score are weighted more heavily towards residential needs. This could also suggest that commuting trips choices are dominated by near-workplace urban form, but discretionary NMT is influenced by near-residence walkability.

All considered, these results suggest that typical commute and travel diary are generated by different processes. Furthermore, consistent with Dillon, Saphores, Boarnet (2015), trip purpose matters in transit usage and NMT.

4.5.4 GSEM results: Comparison of path models

GSEM models are used in this section to further assess the importance of these urban form features on transit use and NMT after accounting for the influence of these features on household vehicle ownership and residential selection. Consistent with the approach for reporting the reduced form models, this section reports results in this order:

- Probability of Indicating a Typical Commute Mode (Table 4-9 reports transit models, Table 4-10 reports NMT models)
- Probability of Using a Mode for Commute Trips, (Based on Trip Diary) (Table 4-11 reports transit models, Table 4-12 reports NMT models)

- Probability of Using a Mode for Non-Work Trips, (Based on Trip Diary) (Table 4-13 reports transit models, Table 4-14 reports NMT models)

See Appendix B for the full regression tables. For each of these behavioral outcomes by mode, I examined five recursive GSEM models to assess the effect of urban form specifications in explaining transit usage or NMT (Table 4-5): (1) near-residence population density as the only urban form metric, (2) #1 plus residential walkability (gamma-connectivity index) and transit access (transit stops), (3) uses EFA extract for density and walkability, (4) uses Walk Score and Transit Score, and (5) urban form as a latent construct, jointly estimated in a CFA measurement sub-component. Before describing and comparing coefficient estimates, I address model performance.

4.5.4.1 Model fit

In the GSEM context, AIC and BIC are the appropriate information criteria to compare models. Unlike likelihood-ratio, Wald, and similar testing procedures, the models need not be nested to compare the information criteria; they just need to be estimated on the same set of dependent variables. Because they are based on the log-likelihood function, information criteria are available only after commands that report the log likelihood. In general, smaller is better: given two models, the one with the smaller AIC fits the data better than the one with the larger AIC. As with the AIC, a smaller BIC indicates a better-fitting model. Degrees of freedom corresponds to the number of parameters estimated by the model, an indicator of complexity.

Each table compares five models that share the same conceptual path and thus posit the same behavioral structure. In all models for transit and NMT, both for commuting and discretionary travel, AIC and BIC favor the most simplistic model (1), where density is the only metric for urban form. EFA extracts and the CFA approach offer richer policy insights at a cost of complexity. AIC and BIC values for the two approaches do not differ by much. Models using Walk Score and Transit Score to represent urban form do not offer a better fit compared to competing models, even after reducing estimation complexity. This might indicate presence of multicollinearity that these scores induce on population density.

4.5.4.2 Coefficient estimates

GSEM coefficient are structural parameters that, in unison, represent a theoretically-informed behavioral process. In this research, I specified a system of equations to explicitly parameterize endogenous relationships between residential selection and vehicle ownership decisions on transit ridership and active travel, controlling for self-selection bias. Estimates of direct effects in the residential selection equation suggest that higher income households gravitate toward low density and car dependent neighborhoods that are not well served by transit.

Coefficient estimates from the vehicle ownership equation suggest that higher income households tend to own more cars per licensed drivers. There are some evidence that households living in neighborhoods with high population density, greater walkability and greater transit access, are less likely to own cars.

Coefficient estimates on the travel behavior estimation vary by mode and trip purpose. Households with a higher vehicle ratio are less likely to commute by transit or NMT. This effect is weaker on discretionary travel using NMT modes. Near-workplace transit availability seems to encourage transit commuting after accounting for its influence on vehicle ownership and residential selection. The CFA model reports a negative direct effect of near-workplace walkability (Table 4-9, column 5), which may indicate a correction factor to other urban form metrics. However, the EFA model does not present this ambiguity (Table 4-9, column 3). In the EFA model, near-workplace walkability is not statistically significant.

Comparing transit usage as participant-reported typical commute mode (Table 9.1.a) and transit commuting based on travel diary data (Table 4-11), I found that near-workplace transit retains significance in direct-effects, except when it is specified as Transit Score (column 4). Near-workplace walkability has a statistically significant direct effect on the probability of NMT commute (Table 4-12, row 6), but not in reporting NMT as typical commute mode (Table 4-11). Surprisingly, near-workplace walkability seems to play a role in home-based discretionary NMT trips (Table 4-14, row 6). This may indicate cases where households who use NMT tend to live near the workplace. Across five modeling approaches, there are no clear pattern of how urban form influences commuting and discretionary trips. However, the model with CFA (column 6) seems to suggest that, if the researcher is willing to assume that urban form can be represented as a single latent construct, measured with errors, the effect of various facets of land uses seems to be important in explaining transit use, but not so much for NMT. After controlling for residential and vehicle ownership endogeneities, urban form has an ambiguous direct effect on active commuting. Results vary by how urban form is specified and by source of travel data. Finally, there are some evidence that population density affect active travel for discretionary purposes (Table 4-14, first row).

Table 4-9 GSEM Results: Probability of Indicating Transit as the Typical Commute Mode

Modelling Strategy/ UF Specification	Standardized Density (1)	Std. UF variables (2)	EFA Factor extracts (3)	Walk/Transit Scores (4)	GSEM with CFA (5)
Travel behavior equation (direct effects)					
Density / Urban Form	0.0655*	0.1054	0.0552***	0.1137	0.6821***
Residential Transit	.n/a	0.0607	0.5117	-0.0055	0.0157***
Residential Walkability	.n/a	-1.6940	0.0798	-0.0014	0.0281***
Vehicle ratio	-2.2297***	-2.1707***	-2.1898***	-1.9468***	-2.0692***
Workplace Transit	0.8952***	0.9078***	0.9098***	0.0363***	0.8837***
Workplace Walkability	-11.4149**	3.7045	0.9472	0.0090	-11.8069**
Household Income	-0.0119**	-0.0104**	-0.0121**	-0.0131**	-0.0084
Vehicle ownership equation (Vehicle ratio)					
Density / Urban Form	-0.0161***	-0.0027	-0.0115***	-0.0397***	-0.1246***
Residential Transit	.n/a	-0.0202	-0.1465	-0.0020***	-0.0029***
Residential Walkability	.n/a	-0.3260*	0.0075	-0.0001	-0.0051***
Workplace Transit	0.0044	0.0035	0.0039	-0.0001	0.0077
Workplace Walkability	-1.9992**	-0.7441	-1.1108	-0.0005	-2.0177**
Household Income	0.0029***	0.0027***	0.0028***	0.0027***	0.0023***
Residential selection equation (Density/Urban Form)					
Income → Density	-0.0297***	-0.0094***	-0.0576***	-0.0094***	-0.0250***
Income → Transit	.n/a	-0.0036***	-0.0005***	-0.1692***	-0.0006***
Income → Walkability	.n/a	-0.0066***	-0.0027***	-0.1995***	-0.0010***
Model fit indices					
Log-Likelihood [df]	-7,253.728 [55]	-41,028.760[235]	-14,372.248 [130]	-23,779.179 [75]	-15,266.751 [104]
AIC	14,617.4561	82,527.5184	29,004.496	47,708.358	30,741.502
BIC	14,930.2960	83,864.1977	29,743.935	48,134.958	31,333.054

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n/a means parameter not estimated in the model.

Table 4-10 GSEM Results: Probability of Indicating NMT as the Typical Commute Mode

Modelling Strategy/ UF Specification	Standardized Density (1)	Std. UF variables (2)	EFA Factor extracts (3)	Walk/Transit Scores (4)	GSEM with CFA (5)
Travel behavior equation (direct effects)					
Density / Urban Form	0.0089	-0.0462	0.0149	0.0176	0.3629**
Residential Transit	.n/a	-0.1576	-0.6986	0.0094	0.0084**
Residential Walkability	.n/a	22.1931**	0.1673	0.0122*	0.0151**
Vehicle ratio	-1.4647***	-1.4371***	-1.4615***	-1.4053***	-1.3386***
Workplace Transit	-0.2974	-0.2907	-0.2858	-0.0112	-0.3097
Workplace Walkability	26.2380	22.1931**	22.5301	0.0010	20.5118
Household Income	-0.0106	-0.0092	-0.0105	-0.0093	-0.0081
Vehicle ownership equation (Vehicle ratio)					
Density / Urban Form	-0.0161***	-0.0027	-0.0115***	-0.0397***	-0.1242***
Residential Transit	.n/a	-0.0203	-0.1465	-0.0020***	-0.0029***
Residential Walkability	.n/a	-0.3259*	0.0075	-0.0001	-0.0052***
Workplace Transit	0.0045	0.0036	0.0039	-0.0001	0.0077
Workplace Walkability	-1.9992**	-0.7441	-1.1108	-0.0005	-2.0177**
Household Income	0.0030***	0.0028***	0.0028***	0.0027***	0.0023***
Residential selection equation (Density/Urban Form)					
Income → Density	-0.0297***	-0.0094***	-0.0576***	-0.0094***	-0.0248***
Income → Transit	.n/a	-0.0036***	-0.0005***	-0.1692***	-0.0006***
Income → Walkability	.n/a	-0.0066***	-0.0027***	-0.1995***	-0.0010***
Model fit indices					
Log-Likelihood [df]	-7,220.452 [55]	-40,998.421 [235]	-14,343.850 [130]	-23,730.586 [75]	-15,266.751 [104]
AIC	14,550.905	82,466.842	28,947.700	47,611.172	30,741.502
BIC	14,863.745	83,803.522	29,687.140	48,037.772	31,333.054

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n/a means parameter not estimated in the model.

Table 4-11 GSEM Results: Probability of Using Transit for Commute Trip (Based on Trip Diary)

Modelling Strategy/ UF Specification	Standardized Density (1)	Std. UF variables (2)	EFA Factor extracts (3)	Walk/Transit Scores (4)	GSEM with CFA (5)
Travel behavior equation (direct effects)					
Density / Urban Form	0.0368	-0.1081	0.0395	0.0446	0.8425***
Residential Transit	.n/a	0.0377	0.3137	-0.0054	0.0194***
Residential Walkability	.n/a	-1.7686	0.5144	0.0069	0.0350***
Vehicle ratio	-3.3254***	-3.1965***	-3.2492***	-3.0476***	-3.0941***
Workplace Transit	0.6508***	0.6639***	0.6682***	0.0094	0.6536***
Workplace Walkability	-35.4994	-24.6957	-40.9130	0.0173*	-53.1695
Household Income	-0.0100	-0.0076	-0.0097	-0.0088	-0.0051
Vehicle ownership equation (Vehicle ratio)					
Density / Urban Form	-0.0161***	-0.0027	-0.0115***	-0.0397***	-0.1245***
Residential Transit	.n/a	-0.0203	-0.1465	-0.0020***	-0.0029***
Residential Walkability	.n/a	-0.3259*	0.0075	-0.0001	-0.0052***
Workplace Transit	0.0045	0.0036	0.0039	-0.0001	0.0078
Workplace Walkability	-1.9992**	-0.7441	-1.1108	-0.0005	-2.0194**
Household Income	0.0030***	0.0028***	0.0028***	0.0027***	0.0023***
Residential selection equation (Density/Urban Form)					
Income → Density	-0.0297***	-0.0094***	-0.0576***	-0.0094***	-0.0249***
Income → Transit	.n/a	-0.0036***	-0.0005***	-0.1692***	-0.0006***
Income → Walkability	.n/a	-0.0066***	-0.0027***	-0.1995***	-0.0010***
Model fit indices					
Log-Likelihood [df]	-7,091.057 [55]	-41,028.760[235]	-14,372.248 [130]	-23,607.217 [75]	-15,266.751 [104]
AIC	14,292.113	82,527.5184	29,004.496	47,364.434	30,741.502
BIC	14,604.953	83,864.1977	29,743.935	47,791.034	31,333.054

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n/a means parameter not estimated in the model.

Table 4-12 GSEM Results: Probability of Using NMT for Commute Trip (Based on Trip Diary)

Modelling Strategy/ UF Specification	Standardized Density (1)	Std. UF variables (2)	EFA Factor extracts (3)	Walk/Transit Scores (4)	GSEM with CFA (5)
Travel behavior equation (direct effects)					
Density / Urban Form	0.0239	0.0151	0.0179	0.1349**	0.2910
Residential Transit	.n/a	-0.1025	-0.4942	-0.0124	0.0067
Residential Walkability	.n/a	-3.0940	0.1638	0.0151*	0.0121
Vehicle ratio	-2.0800***	-2.0827***	-2.0933***	-2.1174***	-2.0721***
Workplace Transit	-0.0347	-0.0260	-0.0247	-0.0187**	-0.0198**
Workplace Walkability	64.2736**	62.9698*	61.2229**	0.0275***	0.0288***
Household Income	0.0052	0.0048	0.0052	0.0062	0.0066
Vehicle ownership equation (Vehicle ratio)					
Density / Urban Form	-0.0161***	-0.0027	-0.0115***	-0.0397***	-0.1245***
Residential Transit	.n/a	-0.0203	-0.1465	-0.0020***	-0.0029***
Residential Walkability	.n/a	-0.3259*	0.0075	-0.0001	-0.0052***
Workplace Transit	0.0045	0.0036	0.0039	-0.0001	0.0077
Workplace Walkability	-1.9992**	-0.7441	-1.1108	-0.0005	-2.0219**
Household Income	0.0030***	0.0028***	0.0028***	0.0027***	0.0023***
Residential selection equation (Density/Urban Form)					
Income → Density	-0.0297***	-0.0094***	-0.0576***	-0.0094***	-0.0248***
Income → Transit	.n/a	-0.0036***	-0.0005***	-0.1692***	-0.0006***
Income → Walkability	.n/a	-0.0066***	-0.0027***	-0.1995***	-0.0010***
Model fit indices					
Log-Likelihood [df]	-7,091.057 [55]	-40,896.643 [235]	-14,239.530 [130]	-23,624.947 [75]	-15,131.675 [104]
AIC	14,292.113	82,263.286	28,739.059	47,399.894	30,471.349
BIC	14,604.953	83,599.966	29,478.499	47,826.494	31,062.901

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n/a means parameter not estimated in the model.

Table 4-13 GSEM Results: Probability of Using Transit for Non-Work Discretionary Trips (Based on Trip Diary)

Modelling Strategy/ UF Specification	Standardized Density (1)	Std. UF variables (2)	EFA Factor extracts (3)	Walk/Transit Scores (4)	GSEM with CFA (5)
Travel behavior equation (direct effects)					
Density / Urban Form	0.0488*	0.2899***	0.0409**	0.0877	0.4714**
Residential Transit	.n/a	-0.1403	-0.5312	0.0158	0.1226**
Residential Walkability	.n/a	1.8807	0.1947	-0.0026	0.0468**
Vehicle ratio	-2.5227***	-2.7471***	-2.5598***	-2.4597***	-2.4147***
Workplace Transit	-0.1669	-0.1644	-0.1523	-0.0006	-0.1311
Workplace Walkability	45.4555	66.9505*	43.8533	0.0082	48.8592
Household Income	-0.0116*	-0.0091	-0.0105*	-0.0107*	-0.0092
Vehicle ownership equation (Vehicle ratio)					
Density / Urban Form	-0.0161***	-0.0027	-0.0115***	-0.0397***	-0.1320***
Residential Transit	.n/a	-0.0203	-0.1465	-0.0020***	-0.0324***
Residential Walkability	.n/a	-0.3259*	0.0075	-0.0001	-0.0124***
Workplace Transit	0.0045	0.0036	0.0039	-0.0001	0.0077
Workplace Walkability	-1.9992**	-0.7441	-1.1108	-0.0005	-2.0216**
Household Income	0.0030***	0.0028***	0.0028***	0.0027***	0.0023***
Residential selection equation (Density/Urban Form)					
Income → Density	-0.0297***	-0.0094***	-0.0576***	-0.0094***	-0.0264***
Income → Transit	.n/a	-0.0036***	-0.0005***	-0.1692***	-0.0065***
Income → Walkability	.n/a	-0.0066***	-0.0027***	-0.1995***	-0.0025***
Model fit indices					
Log-Likelihood [df]	-7,188.901 [55]	-40,962.016 [235]	-14,311.027 [130]	-23,698.375 [75]	-15,204.194 [104]
AIC	14,487.802	82,394.032	28,882.054	47,546.750	30,616.389
BIC	14,800.642	83,730.711	29,621.494	47,973.350	31,207.941

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n/a means parameter not estimated in the model.

Table 4-14 GSEM Results: Probability of Using NMT for Non-Work Discretionary Trips (Based on Trip Diary)

Modelling Strategy/ UF Specification	Standardized Density (1)	Std. UF variables (2)	EFA Factor extracts (3)	Walk/Transit Scores (4)	GSEM with CFA (5)
Travel behavior equation (direct effects)					
Density / Urban Form	0.0352*	0.1344*	0.0210*	0.0659	0.3342***
Residential Transit	.n/a	-0.0387	-0.1894	0.0039	0.0874***
Residential Walkability	.n/a	1.1039	0.1214	0.0057**	0.0333***
Vehicle ratio	-0.5414***	-0.5188***	-0.5127***	-0.5327***	-0.4886***
Workplace Transit	-0.0694	-0.0742	-0.0710	-0.0018	-0.0765
Workplace Walkability	33.7564**	29.3096*	29.0479*	0.0026	29.7760*
Household Income	0.0008	0.0013	0.0013	0.0016	0.0028
Vehicle ownership equation (Vehicle ratio)					
Density / Urban Form	-0.0161***	-0.0027	-0.0115***	-0.0397***	-0.1323***
Residential Transit	.n/a	-0.0203	-0.1465	-0.0027***	-0.0326***
Residential Walkability	.n/a	-0.3259*	0.0075	-0.0001	-0.0124***
Workplace Transit	0.0045	0.0036	0.0039	-0.0001	0.0078
Workplace Walkability	-1.9992**	-0.7441	-1.1108	-0.0005	-2.0173**
Household Income	0.0030***	0.0028***	0.0028***	0.0030***	0.0023***
Residential selection equation (Density/Urban Form)					
Income → Density	-0.0297***	-0.0094***	-0.0576***	-0.0094***	-0.0264***
Income → Transit	.n/a	-0.0036***	-0.0005***	-0.1692***	-0.0065***
Income → Walkability	.n/a	-0.0066***	-0.0027***	-0.1995***	-0.0025***
Model fit indices					
Log-Likelihood [df]	-8,319.295 [55]	-42,090.309 [235]	-15,438.143 [130]	-24,825.239 [75]	-16,331.083 [104]
AIC	16,748.589	84,650.617	31,136.287	49,800.478	32,870.166
BIC	17,061.429	85,987.297	31,875.726	50,227.078	33,461.718

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. n/a means parameter not es

4.6 Discussion and conclusions

4.6.1 Place in the literature

Consistent with Cervero & Kockelman (1997) and Cervero & Duncan (2003), my EFA results indicate that urban form can be summarized into two factors; density and design. Similar to Cervero & Kockelman (1997), several of my urban form variables did not factor. In particular, diversity, represented here by land use entropy, does not factor. However, Cervero & Kockelman (1997), who used 7 variables to represent land use diversity at the trip-origin, also found the same result.

My CFA results offer evidence that urban form can be viably represented as a single latent factor. This latent factor can be measured by density, diversity, connectivity (design), and bikeway density variables, when these variables are assumed to be correlated. As noted in the preceding methodology section, CFA allows measurement errors to be correlated while EFA does not. This result offers evidence that urban form influences travel because the configuration of these variables (density, diversity, connectivity) affects the time cost of travel and the relative cost across modes (Boarnet & Crane, 2001, p. 73). However, path model coefficient estimates on the CFA model sometimes offer counterintuitive evidence on the effect of urban form on travel behavior.

Results from the GSEM path models, which account for residential and vehicle-choice endogeneities, are generally consistent with the active transportation literature. Households living in neighborhoods with higher density, more transit access, more connected street network, and are working in a transit-rich workplace are more likely to commute using transit and NMT. Everything else hold equal, these households typically own fewer vehicles per licensed drivers and are poorer compared to households who are less likely to active-commute. Households living in high density neighborhoods with features that support active commute tend to be members of lower income groups. They are also more likely to buy fewer vehicles for each licensed drivers, consistent with Kuzmyak & Dill (2012).

While these models report coefficients that are similar in sign and statistical-significance, with a notable exception on the model where all measures of urban form are entered as observed variables, the size of estimates are different. Comparison of results from different trip purposes and source of travel data suggest that trip purposes matter. Future work might want to consider specification of travel behavior outcome such as count of household's transit trips, etc.

Urban form in the CFA framework can also be interpreted as an index/unit of account for housing goods. Housing is a bundled good. In California, housing supply restrictions exist, distorting housing price. Urban form might be adequate to capture variation in the context of travel behavior.

Finally, GSEM results offer evidence that typical commute and travel diary data were generated by discernably different processes. Each has its own deficiency in portraying travel behavior. For example, aspirational factors might confound how a household reports their typical commute, but happenstance might affect household travel on the survey date and thus not provide an accurate representation of travel behavior. Happenstance is a form of randomness that can be assumed to cancel out on average (centers at zero). At any rate, coefficient estimates from both regressions can be used as bounds to inform planning decisions.

4.6.2 Research limitations

This research presents results on how urban form modeling strategies explain transit usage and NMT using Los Angeles and Orange Counties sub-sample of the 2009 NHTS which oversampled Whites and Asians and undersampled Hispanics and Blacks. In the vast travel behavior literature, urban form has been specified in numerous ways – this research only considers a subset of those. In particular, with more measures for land use diversity, the EFA model might offer different insights. Some urban form measures, such as the connectivity index, was calculated at the block group level, while others, such as transit stops, were calculated within a ¼ mile buffer. It is unclear how to reconcile the two. Furthermore, ¼ mile buffer may not be optimal to capture transit usage. Future research should consistently use buffers with a well-defined distance to construct a household's neighborhood. Another approach would be to conduct the analysis at the trip level such that it can inform trip-chaining behaviors.

Walk Score and Transit Score hold promise to be used jointly in travel behavior studies. They are relatively convenient to use, but at a cost of opacity. Furthermore, Transit Score has not been thoroughly studied like Walk Score. Future research should expound upon Transit Score – taking into account how near-residence transit stop connects to destinations.

While correlated errors dramatically improve CFA model fit (Table 7.D.), this modeling approach treads on the boundary between theoretical justification and overfitting. Specifying a multilevel CFA may relax some of these correlated errors and directly test for the existence of structural ordering in urban form latent factors. Future work should consider multilevel CFA to expound upon the multidimensionality of urban form in travel behavior.

Lastly, future research should do the utmost to overcome historically-disadvantaged community under-sampling in travel datasets. This is important for at least two reasons: (1)

results may not be generalizable to the population, and most importantly (2) transportation and planning organizations rely on these data for policy planning purposes. Bad data led to bad statistics led to bad planning. Planning with bad data only reinforces disenfranchisement.

4.6.3 Implication for policy

Cervero & Kockelman (1997) contends that it is "futile to attempt to isolate unique contribution of each and every observed variable that measures fine aspects of urban form" (p. 210). This research presents results from multiple analytical perspectives. Results offer some evidence of the importance of workplace urban form in promoting transit usage and NMT commuting. This suggests that MPOs should consider improving transit access and walkability in job centers. While results are preliminary, it consistently report the importance of near-workplace urban features in promoting transit and NMT.

The fact that urban form has little direct effect on travel should not deter policy promoting transit investment and improvement. When the policy goal is to reduce car travel, policies that reduces car ownership (and dependency) should count as a policy impact. Although vehicle ownership (or lack thereof) seems to consistently explain transit use and active travel, MPOs should not stop investing in transit improvements. Instead, this as a signal that transit service and housing choices should be improved to assist the travel needs of income-restricted households.

4.7 References

Acock, A. C. (2013). Discovering structural equation modeling using Stata. *Stata Press books*.

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.

Boarnet, M. G., & Crane, R. (2001). *Travel by design: The influence of urban form on travel*. Oxford University Press.

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, *65*(1), 91–98. doi:10.1016/j.jue.2008.09.002

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge.

Cao, X., Handy, S. L., & Mokhtarian, P. L. (2006). The influences of the built environment and residential self-selection on pedestrian behavior: evidence from Austin, TX. *Transportation*, *33*(1), 1-20.

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.

Cervero, R., & Duncan, M. (2003). Walking, bicycling, and urban landscapes: evidence from the San Francisco Bay Area. *American journal of public health*, *93*(9), 1478-1483.

de Dios Ortuzar, J., & Willumsen, L. G. (2011). *Modelling transport*. New Jersey: Wiley.

Dill, J. (2004, January). Measuring network connectivity for bicycling and walking. In *83rd Annual Meeting of the Transportation Research Board, Washington, DC* (pp. 11-15).

Duncan, D. T., Aldstadt, J., Whalen, J., & Melly, S. J. (2013). Validation of Walk Scores and Transit Scores for estimating neighborhood walkability and transit availability: a small-area analysis. *GeoJournal*, *78*(2), 407-416.

Frank, L. D., Schmid, T. L., Sallis, J. F., Chapman, J., & Saelens, B. E. (2005). Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. *American journal of preventive medicine*, 28(2), 117-125.

Glanz, K., Handy, S. L., Henderson, K. E., Slater, S. J., Davis, E. L., & Powell, L. M. (2016). Built environment assessment: multidisciplinary perspectives. *SSM-Population Health*, 2, 24-31.

Giles-Corti, B., Wood, G., Pikora, T., Learnihan, V., Bulsara, M., Van Niel, K. & Villanueva, K. (2011). School site and the potential to walk to school: The impact of street connectivity and traffic exposure in school neighborhoods. *Health & Place*, *17*(2), 545-550.

Houston, D., Basolo, V., & Yang, D. (2013). Walkability, transit access, and traffic exposure for low-income residents with subsidized housing. *American journal of public health*, *103*(4), 673-678.

Kahn, M. E., & Matsusaka, J. G. (1997). Demand for environmental goods: Evidence from voting patterns on California initiatives. JL & Econ., 40, 137.

Khan, M., Kockelman, K. M., & Xiong, X. (2014). Models for anticipating non-motorized travel choices, and the role of the built environment. *Transport Policy*, *35*, 117-126.

Kim, J., & Brownstone, D. (2013). The impact of residential density on vehicle usage and fuel consumption: Evidence from national samples. *Energy Economics*, *40*, 196–206. doi:10.1016/j.eneco.2013.06.012

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. Guilford publications.

Kuzmyak, J., Baber, C., & Savory, D. (2006). Use of walk opportunities index to quantify local accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, (1977), 145-153.

Kuzmyak, J. R., & Dill, J. (2012). Walking and bicycling in the United States: The Who, What, Where, and Why. *TR News*, (280). (URL: https://trid.trb.org/view.aspx?id=1143549)

Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. (2007). Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. *Health & Place*, *13*(1), 111-122.

Manaugh, K., & El-Geneidy, A. (2011). Validating walkability indices: How do different households respond to the walkability of their neighborhood? *Transportation Research Part D: Transport and Environment*, *16*(4), 309-315.

Mannering F. (1986). Selectivity bias in models of discrete and continuous choice: An empirical analysis. Transportation Research Record, 1085, 58–62.

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28-56.

Porta, S., & Renne, J. L. (2005). Linking urban design to sustainability: formal indicators of social urban sustainability field research in Perth, Western Australia. *Urban Design International*, *10*(1), 51-64.

Sangkapichai, M., & Saphores, J. D. (2009). Why are Californians interested in hybrid cars?. *Journal of environmental planning and management*, *52*(1), 79-96.

Small, K. A., & Verhoef, E. T. (2007). *The economics of urban transportation*. Routledge.

Southern California Association of Governments (2016). *The 2016-2040 Regional Transportation Plan/Sustainable Communities Strategy (2016 RTP/SCS or Plan)*. Los Angeles, CA: Author.

Sörbom, D. (1989). Model modification. Psychometrika, 54(3), 371-384.

Voorhees, C. C., Ashwood, J. S., Evenson, K. R., Sirard, J. R., Rung, A. L., Dowda, M., & McKenzie, T. L. (2010). Neighborhood design and perceptions: relationship with active commuting. *Medicine and science in sports and exercise*, *42*(7), 1253.

Wasfi, R. A., Dasgupta, K., Eluru, N., & Ross, N. A. (2016). Exposure to walkable neighborhoods in urban areas increases utilitarian walking: longitudinal study of Canadians. *Journal of Transport & Health*, 3(4), 440-447.

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

5 Conclusions

In this dissertation, I presented three essays that used structural equation modeling techniques to tease out the relationship between urban form, driving patterns, ownership of alternative fuel vehicles, transit use, and active commuting. I focused on urban California where land use regulations have been used to pursue ambitious policy goals. Also, while urban Californians tend to favor progressive environmental policies, they drove a lot more compared to their out-of-state peers. My results have important policy implications for land use, transportation policies, and other incentives at the local and state levels.

First, my results show that trip purpose matters: in the short run, households drive 0.171% less for non-work trips when gas prices increase by 1%, while work trips are not responsive to gasoline price changes. Just considering total VMT in informing policies would mistakenly suggest that driving patterns are not responsive to gasoline price changes. In the short term, higher fuel prices reduce discretionary driving such as shopping and recreational trips, but they do not affect non-discretionary driving such as commuting trips. Second, results suggest that policies that seek to increase transit service and housing opportunities near employment centers will reduce driving. Third, as mentioned in Chapter 3, findings indicate households who live closer to a freeway with HOV lanes, work closer to parking lots with AFV privileges, and are likely to support pro-environmental measures are more likely to own AFVs. However, these households are also likely to drive more than if they had conventional vehicles. While expansion of HOV lanes and parking incentives will increase the adoption of alternative fuel vehicles, its utilization will also increase although the effect is relatively small. These results imply that HOV and parking incentives come with unintended consequences. Increasing awareness about environmental problems might yield higher policy pay-off. Fourth, as mentioned in Chapter 4, a household's socio-demographic status affects travel behavior directly and indirectly through residential location and vehicle ownership choices. Lower income, younger, and smaller households are more likely to choose a dense, pedestrian friendly, and transit rich neighborhood.

Land use interventions have been advocated to partly rein in automobile use, boost transit ridership, and encourage more non-motorized travel in order to improve livability of urban areas. These policies were premised under the distortion in the transportation markets (quantity of miles driven, transit use, vehicle ownership, etc.). However, my dissertation has yet to discuss the distortion in the housing markets (i.e. supply restriction, public goods externalities & unpriced amenities) that may have affected residential choice patterns. Observed distortion in the transportation markets (i.e. congestion; miles-driven is larger than the social optimal) may indeed signal distortion in the housing market. Households may actually prefer to live closer to their workplace or to a transit node in order to minimize commute distance, time or costs, but housing prices might be restrictively expensive or lack nearby amenities that are bundled in housing goods. Unpriced congestion could also distort the housing market. Further research should consider a theoretical framework to jointly address these issues and to investigate whether land use interventions carry unintended consequences.

Future empirical work should consider using a panel design to better control for time-invariant factors affecting residential location and vehicle ownership decisions. A panel design would also allow testing for feedback between travel behavior and vehicle ownership decisions as well as intra-household vehicle substitution behavior in which households with multiple vehicles could choose to drive different vehicles in respond to price changes. A natural follow up would be to explore more in depth how households adjust their travel behavior in response to increases in gasoline prices: Do they organize their trips differently, forego travel altogether, or do they use other modes? Another possibility would be to refine the residential selection component by exploring how households choose neighborhoods based on amenities such as local public school quality and other physical neighborhood features.

Overall, my dissertation research helps improve our understanding of linkages between urban form, travel behavior, congestion, and air pollution. I hope that my work will inform public policy at the local and state levels.