Title
Land Use, Land Value, and Transportation: Essays on Accessibility, Carless Households, and Long-distance Travel

Author
Mitra, Suman Kumar

Publication Date
2016

UNIVERSITY OF CALIFORNIA,
IRVINE

Land Use, Land Value, and Transportation: Essays on Accessibility, Carless Households, and Long-distance Travel

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Transportation Science

by

Suman Kumar Mitra

Dissertation Committee:
Professor Jean-Daniel Saphores, Chair
Professor R. Jayakrishnan
Professor Douglas Houston

2016

DEDICATION

In memory of Shiba Prosad Roy (1937-2016)

who was my friend, philosopher, and guide.

&

To

My Parents

in recognition of their patience and support.

Where the mind is without fear and the head is held high;
Where knowledge is free;
Where the world has not been broken up into fragments by narrow domestic walls;
Where words come out from the depth of truth;

–Rabindranath Tagore (1900)

Table of Contents

List of Tables

List of Figures

Acknowledgements

Earning a PhD is something I have dreamt about since I was in high school. My late high school teacher, Mr. Shiba Prosad Roy inspired me to seek a PhD. Conducting PhD research has provided me with an invaluable experience, matured me in so many ways, and marked an important milestone in my career. More importantly, it has brought many beautiful and bright people into my life, and they deserve special mention here.

First and foremost, my deepest gratitude is to my advisor, Professor Jean-Daniel Saphores, who is the best advisor any student could hope for. I feel a deep sense of pride and satisfaction to have worked so closely with him. Jean-Daniel, you are a great professor, an excellent mentor, and a good friend. Without your encouragement, mentorship, and constant guidance, this research might never have been completed. Your insightful comments and constructive criticisms, at different stages of my research, were thought-provoking and very helpful. Besides, your great sense of humor and our conversations about soccer made my PhD life joyful. Thank you for everything.

I would also like to thank my other committee members, Professor R. Jayakrishnan and Professor Douglas Houston, for their guidance and suggestions to improve this research. The Transportation Science program at UC Irvine also gave me a unique opportunity to learn from and interact with many faculty members from across the University, which made these years an insightful and inspiring experience. I would also like to thank Professor Jae Hong Kim, Professor Michael McNally, Professor Scott Bollens, Professor Stephen Ritchie, Professor Wenlong Jin, and Professor Will Recker for their great lectures and inspiring ideas. Additionally, I would like to express my gratitude to the incredible administrative staff at the Institute of Transportation

Studies (ITS) for their kind support: Anne Marie DeFeo, Kathy Riely, Myra Radlow and Ziggy Bates. They are the best!

Next, I would like to express my heartfelt appreciation to my friends Alex and Irene. Living in Irvine without a car is not easy, but Alex and Irene made my life easier. Many thanks are due to my fellow PhD students in ITS, particularly my friends: Anupam, Ashley, Jared, Kate, Kevin, Koko, Kyungsoo and Todd for sharing the good and bad times. They made my graduate life enjoyable. I would also like to acknowledge Rezwana Rafiq, Dr. Yusuf Sarwar Uddin, Dr. Ankoor Bahgat, Dr. Andre Tok, Gavin Ferguson, Dr. Jalal Ahmed, Dr. Paulos Ashebir, Dr. Sarah Hernandez, and Stephanie Mak with whom I have had very good times.

I sincerely thank the Rajshahi Development Authority (RDA) for providing me with the GIS database of Rajshahi City. I would like to acknowledge Mr. Tony Soeller for GIS advice. I also want to thank my colleagues from the Department of Urban and Regional Planning, Bangladesh University of Engineering and Technology for their support and encouragement at different stages of my graduate life. I am grateful for the economic support provided by the ITS Irvine-UCCONNECT fellowship, the UCI Chancellor Club fellowship, the UCI public impact fellowship, and Caltrans.

Lastly, Thanks to my elder brother and sister-in-law for their support, and thanks to my beloved three years old niece for making my days enjoyable during our Skype conversations. Above all, my heartfelt appreciation goes to my Ma and Baba for their constant support, unconditional love, concern, and care. Although I was thousands of miles away from home, my Ma's everyday phone calls made me feel like I lived very close to them. Ma and Baba…I am grateful for everything.

Curriculum Vitae

Suman Kumar Mitra

Institute of Transportation Studies (ITS),
4000 Anteater Instruction and Research Bldg. (AIRB)
University of California, Irvine (UCI),
Irvine, CA 92697-3600

E-mail: skmitra@uci.edu
Phone: +1 (949)-394-7049

EDUCATION

Ph.D. in Transportation Science **June, 2016**
University of California, Irvine — Irvine, California
Advisor: Jean-Daniel Saphores, PhD
Dissertation Title: *Land Use, Land Value, and Transportation: Essays on Accessibility, Carless Households, and Long-distance Travel.*

Master of Urban and Regional Planning (MURP) **July, 2008**
Bangladesh University of Engineering and Technology (BUET), — Dhaka, Bangladesh
Advisor: K.M. Maniruzzaman, PhD
Thesis Title: *Applicability of Artificial Neural Network in Predicting House Rent.*

Bachelor of Urban and Regional Planning (BURP) **June, 2005**
Bangladesh University of Engineering and Technology (BUET) — Dhaka, Bangladesh
Advisor: Ishrat Islam, PhD
Thesis Title: *A Study on Land Price of Dhaka City.*

RESEARCH INTEREST

Land use-transportation interaction; Travel behavior; Sustainable transportation; Spatial econometrics; Geographical Information System (GIS) application; Transportation in developing countries; Transport disadvantage.

PUBLICATION (Selected)

Journal

- **Mitra, S. K.,** and Saphores, J. D. M. (2016). The value of transportation accessibility in a least developed country city–The case of Rajshahi City, Bangladesh. *Transportation Research Part A: Policy and Practice*, *89*, 184-200.
- **Mitra, S. K.** and Saphores, J. D. M. (2016). Travel Pattern of Voluntary and Involuntary Carless Households in California. *Special Issue "Life- Oriented Travel Behavior Analysis" Transportation Research Part A: Policy and Practice,* Under Review (Extended Abstract accepted).

Book Chapters

- Khan, A. Z. and **Mitra, S. K.**, (2011). Transportation Infrastructure of Dhaka City: Status and Challenges", in Hafiz, R., Rabbani, A. K. M. G.(ed), *Urbanization and Urban Development*, Volume-III, Asiatic Society of Bangladesh, Dhaka, Bangladesh.

- Islam, I**, Mitra, S. K.**, Nayeem, A., and Rahman, A. (2007) "Land Price in Dhaka City: Distribution, Characteristics and Trend of Changes." In Jahan, S. Maniruzzaman K.M. (ed.), *Urbanization in Bangladesh Patterns, Issues and Approaches to Planning*, Bangladesh Institute of Planners, Dhaka, 2007.

Conference Presentations and Proceedings

- **Mitra, S. K.** and Saphores, J. (2016). Who are the Carless Households in California?, Presented at UCCONNECT 2016 Student Conference, UCR, CA, February 11-12, 2016.
- **Mitra, S. K.** and Saphores, J. (2016). Exploring the Unnoticed: An Analysis of Voluntary and Involuntary Carless Households in California. Presented at the *95th Transportation Research Board Annual Meeting*, Washington D.C., January 10-14, 2016.
- **Mitra, S. K.** and Saphores, J. (2015). Characterizing Carless Households in California: Evidence from the 2012 California Household Travel Survey, Paper presented in the *56th Annual Transportation Research Forum*, Atlanta, GA, March 12-14, 2015.
- **Mitra, S. K.** and Saphores, J. (2015). The Value of Transportation Accessibility in a Least Developed Country City: The Case of Rajshahi City, Bangladesh, Presented at UCCONNECT 2015 Student Conference, UCSB, CA, February 27-28, 2015.
- **Mitra, S. K.** and Saphores, J. (2015). Transportation Accessibility and Multi-Unit Residential Property Rents: The Case of Rajshahi City, Presented at the *94th Transportation Research Board Annual Meeting*, Washington D.C., January 12-16, 2015.
- **Mitra, S. K.** and Saphores, J. (2014). Impact of Transportation Accessibility on Residential Property Values in Rajshahi City, Bangladesh: A Spatial Hedonic Approach, Paper presented in the *55th Annual Transportation Research Forum*, San Jose, CA, March 13-15, 2014.

AWARDS AND HONORS

- **ITS UCTC Dissertation Fellowship**, University of California, Irvine (UCI), 2016
- **Chancellor's Club for Excellence Fellowship**, UCI, 2015-2016.
- **Public Impact Fellowship,** UCI, 2015-2016.
- **Judges Winner** for best paper presentation**,** AGS Symposium, UCI, 2016
- **Finalist,** UCI Grad Slam, UCI, 2016
- **AGS Travel Grant,** UCI, Fall 2015.
- **Phi Beta Kappa International Scholarship Award**, UCI, 2014.
- **Prime Minister Gold Medal Award**, University Grants Commission, Government of Bangladesh, 2006.

TECHNICAL SKILLS

- Statistical Analysis: STATA, R, SPSS, Matlab.
- Programming Language: C and C++.
- Transportation Planning: Trans CAD.
- Graphics Software: Adobe Photoshop, Corel draw, Auto CAD.
- Mapping Software: Arc Info, Arc View, Arc GIS.

Abstract of the Dissertation

Land Use, Land Value, and Transportation: Essays on Accessibility, Carless Households, and Long-distance Travel

By

Suman Kumar Mitra

Doctor of Philosophy in Transportation Science

University of California, Irvine, 2016

Professor Jean-Daniel Saphores, Chair

During the last two decades, a large body of empirical research has focused on the relationship between land use and travel behavior, and also on the impacts of transportation accessibility on land value. However, significant gaps remain in our understanding of these relationships. In this dissertation, I present three essays on accessibility, carless households, and long-distance travel that will enhance our understandings of relationships among land use, land value, and transportation.

In my first essay, I provide empirical evidence about the magnitude of the value of transportation accessibility as reflected by residential rents in Rajshahi City, Bangladesh. Results of my SARAR (spatial autoregressive model with spatial-autoregressive disturbances) model show a small but statistically significant capitalization of accessibility. Results of this study should be useful for planning transportation infrastructure funding measures in least developed country cities like Rajshahi City.

In my second essay, I assess the joint effects of various socio-economic, life-cycle stage, and land use variables on the likelihood that a household is carless, voluntarily or not, by analyzing data from the 2012 California Household Travel Survey (CHTS). Results of my binary

logit models show the importance of land use diversity and of good transit service to help households voluntarily forgo their vehicles, and downplay the impact of population density and pedestrian-friendly facilities. Results of this study should help planners and policy makers formulate policies to curb automobile dependency and help promote sustainable urban transportation.

My third essay analyzes long-distance data from the 2012 CHTS to understand the influence of different socio-economic, land use, and land value variables on the likelihood that a household commutes long-distance in California. Results of my Generalized Structural Equation Model (GSEM) show that long-distance commuting is negatively associated with mixed density and residential home values (around commuters' residences), but positively related with households' car-ownership. My results also confirm the presence of residential self-selection. The empirical evidence of this study should help formulate land use planning strategies to curb long-distance commuting and thus help reducing vehicle-miles traveled, which is one way of reducing the emission of greenhouse gases from transportation.

Chapter 1. Introduction

Smart growth has gained popularity in the world especially in western countries over the past two decades. A central theme of this movement is creating a built environment that increases accessibility, discourages automobile use, and fosters walking and transit use. This agenda has resonated with many planners and developers around the world and consequently, different policies such as (but not limited to) mixed-use zoning, infill development, brownfield redevelopment, transit-oriented development, as well as the addition of bicycle and pedestrian infrastructure have been adopted by local governments. The connection between transportation and land use lies at the center of these smart growth strategies.

Transportation and land use are inextricably connected. Investments in transportation systems affect land use patterns, urban densities and land values. On the other hand, land use development patterns influence travel behavior. During the last two decades, a large body of empirical research has focused on the relationship between land use and travel behavior and also on the impacts of the transportation infrastructure and accessibility on land value. However, some substantial gaps remain. In this dissertation, I start addressing these gaps with three essays on accessibility, carless households, and long-distance travel, which will contribute to enhancing our understanding of relationships among land use, land value, and transportation.

In my first essay (Chapter 2), I study the nature and magnitude of transportation accessibility impacts on land value in a developing country city, using spatial hedonic models. Poor accessibility due to inadequate roadway network is impairing the economic development of many developing countries. Although opportunities to finance transportation infrastructure through value capture policies have long been recognized, such policies have not been implemented in developing South Asian countries for various reasons. Among them is the absence of results about quantifying the value of transportation enhancements, and in particular how they are reflected in the housing market. This evidence is necessary to formulate appropriate value capture policies. This essay addresses this gap by providing empirical evidence about the value of transportation accessibility as reflected by property rents using a unique data set collected via in-person interviews in 2006 in Rajshahi City, Bangladesh.

My second essay, presented in Chapter 3, contributes to the growing interest in social justice issues in urban transportation planning by characterizing carless households in California based on the 2012 California Household Travel Survey (CHTS). More specifically, in this chapter, I assess the effects of various socio-economic, life-cycle stage, and land use variables on the likelihood that a household is carless, voluntarily or not. I estimate discrete choice models to explain the characteristics of voluntary and involuntary carless households. Results of this study contribute to our understanding of how different land use policies can help encourage low or zero car ownership while avoiding the potentially negative consequences of not owing a car in an automobile-oriented society and thus foster more sustainable transportation.

In my third essay (Chapter 4), I analyze the relationships among land use, land value and long-distance commuting in California using data from the 2012 CHTS. Better understanding and forecasting long-distance commuting is important for multiple reasons: investing in transportation infrastructure, preserving the environment, enhancing social justice, and stimulating the economy. For various reasons, including a dearth of data, long-distance commuting has received limited interest in the US. The goal of this chapter is to fill this gap by assessing the effects of different socio-economic, land use, and land value variables on the likelihood that households commute long-distance by using Generalized Structural Equation Model (GSEM), while taking into account residential self-selection and the endogeneity of car ownership. Results of this study enhance our understanding of how different policies can help curb long-distance commuting and thus help reduce vehicle miles traveled.

Finally, Chapter 5 summarizes my conclusions and proposes suggestions for future work.

Chapter 2. Transportation Accessibility and Land Value: The Case of Rajshahi City, Bangladesh

2.1 INTRODUCTION

Poor accessibility due to inadequate roads or insufficient public transportation has been hampering the economic growth of developing countries (Creightney, 1993; Alam *et al.*, 2004; Sohail *et al*., 2006). Although options to finance transportation accessibility improvements through value capture policies have long been recognized (Smith and Gihring, 2006; Medda, 2012), such policies have not been implemented in the least developed countries. Various reasons explain this situation, including excessive dependence on international financing, poor local governance, and insufficient housing market data that prevent investigating the capitalization of mobility in housing markets. To formulate appropriate policies, however, it is necessary to understand the value of accessibility, which can greatly depend on local conditions, as illustrated by a review of hedonic studies concerned with the value of transportation accessibility (see Table 2.1).

In least developed countries, the difficulty of finding adequate housing market data cannot be understated. Indeed, the quality of available data is typically questionable and collecting primary data is challenging because of low literacy rates (58%) that are not yet compensated by cellular phone penetration (42%) (UN-OHRLLS, 2013). Conducting postal or phone surveys of the general population is therefore typically not an option. This leaves in-person interviews, an approach that is more costly and time-consuming, as the only alternative.

Although the hedonic literature on transportation accessibility is considerable, at least two areas need additional work. First, since most published studies were carried out in developed countries, there is a dearth of research on the value of transportation accessibility in the least developed countries. This is important because urban property markets there differ from those in the West for a variety of reasons, including the size of the rental market (ownership rates are lower than in the West), lower housing standards, scarcity of urban land, and different regulations. Second, even though the importance of accounting for spatial effects has been known for some time (e.g., see Anselin, 1988), few published papers rely on the most recent spatial econometric techniques that address the risk of maximum likelihood estimation in the presence of heteroskedasticity (Drukker *et al.*, 2013).

This paper starts addressing these issues by providing new evidence about how transport accessibility - to major roads, bus stations, and regional train stations - is capitalized in the housing market of a least developed country city based on a unique dataset collected in 2006 via in-person interviews of renters in Rajshahi City, Bangladesh, which we analyze using spatial regression techniques. A better understanding of the value of accessibility has a wide range of practical applications, from assessing the usefulness of innovative land-based tax instruments that hinge on the capitalization of transportation improvements, to informing policy makers about the potential impacts of land development on transportation infrastructure alternatives, particularly in least developed countries such as Bangladesh.

The remainder of this chapter is structured as follows. Section 2.2 reviews selected papers to put our modeling choice in perspective. Section 2.3 presents background information about Rajshahi City and a description of our data. Section 2.4 outlines our methodology and our results are discussed in Section 2.5. After recapping key results, Section 2.6 discusses some policy implications, outlines some limitations, and suggests directions for future work.

2.2 LITERATURE REVIEW

In this section, I start with an overview of hedonic studies dealing with the impact of transportation accessibility on property values before focusing on accessibility in developing countries. I then discuss selected hedonic studies that rely on gravity-based measures and finish with hedonic studies where the dependent variable is apartment rent.

2.2.1 Hedonic Studies of Transportation Accessibility

Land rent theory (Alonso, 1964; Muth, 1969; Mills, 1972) is commonly invoked for justifying that since improved accessibility benefits dwellers of a property, it should translate into higher property values (Dowall and Monkkonen, 2007).

As shown in Table 2.1, many recently published studies dealing with the impacts of transportation infrastructure and accessibility on property values focus on rail transit in North America and in Europe from a variety of perspectives, including different types of trains (e.g., rapid, commuter, or light rail) and different types of properties (single family or multi-family residential, and commercial). Some papers examine the impact of Transit Oriented Development (TOD) while others are concerned with the overall impacts of transportation accessibility. A smaller set of papers focuses on the impact on property values of freeways, limited access roadways, and highway noise.

Possibly because of its diversity, results from this literature appear to be inconsistent. A number of papers find evidence of a positive relationship between accessibly improvements and property values (e.g., Cervero and Kang (2011) for Seoul's Bus Rapid Transit; Debrezion *et al*. (2011b) for road accessibility on office prices in the Netherlands; or Dubé *et al.* (2013) for commuter rail accessibility in Montreal). However, others report that transportation accessibility has no effect or has a negative impact on property values, including Martinez and Viegas (2009) for urban ring roads in Lisbon, Portugal; Chatman *et al.* (2012) for light rail stations in southern New Jersey; and Efthymiou and Antoniou (2013) for various modes in Athens, Greece.

A closer reading of the literature shows, however, that factors such as land use type, transportation facility type, measures of accessibly, the degree of accessibility improvement, the level of development of impacted areas, and location contribute to the diversity in reported findings (Debrezion *et al.*, 2007; Medda, 2012; Mohammad *et al.*, 2013). Methodological choices, and in particular how spatial autocorrelation is modeled, also play a role (Diao, 2015).

2.2.2 Hedonic Studies of Accessibility in Developing Countries

Outside North America and Europe, empirical evidence is more limited. However, recent years have seen hedonic studies published in English that examine the value of accessibility in geographically diverse places such as China (Pan and Zhang, 2008; Zhang and Wang, 2013), Colombia (Rodriguez and Mojica, 2009; Munoz-Raskin, 2010), Hong Kong (Yiu and Wong, 2005; Cervero & Murakami, 2009; Shyr *et al*., 2013), South Korea (Cervero and Kang, 2011), Taiwan (Andersson *et al.*, 2010; Shyr *et al*., 2013), Thailand (Chalermpong, 2007), and Turkey (Celik and Yankaya, 2006). These studies are summarized in Table 2.2.

Among these studies, Chalermpong (2007) stands out as he appears to have pioneered the use of spatial models (spatial lags and spatial error) to study accessibility in developing countries. Only two other studies estimated spatial hedonic models: Rodriguez and Mojica (2009) in Colombia, and Zhang and Wang (2013) in China. With the exception of Cervero and Kang (2011) who relied on multilevel regression models, other papers in Table 2.2 used mostly regression models with possibly transformed (via log or Box-Cox) variables.

Table 2.1 Summary of Selected Studies Dealing with the Impact of Transportation Accessibility on Property Values

Author (Year Published)	Data and Location [Year]	Method	Key Results
Rail Accessibility			
Mohammad *et al.* (2015)	Sale price of 39,308 residential and 3,419 commercial properties, Dubai, UAE [2007-2009 & 2010-2011]	Difference-in-difference estimator and OLS.	The impact of the metro on the value of residential (commercial) properties is largest within 701 to 900 meters of a metro station and is ~13% (76%).
Diao (2015)	Sale price of 10,031 single-family houses, Boston, USA [1998-2007].	Heckman selection model combined with spatial lag and spatial error models.	The willingness to pay (WTP) for subway accessibility is US$10,000 for a property valued at US$325,000. Compared with a Heckman selection model with spatial error, the bias in WTP of a conventional hedonic price model is ~91.0%.
Dubé *et al.* (2014)	Sale price of 27,311 housing, Montreal, Canada [1992-2009].	Spatial difference-in-differences (SDID).	Station proximity increases the growth in house prices as follows: 2.3% (500–1000 m), 3.8% (1000–1500 m), and 5.2% (0–500 m).
Dubé *et al.* (2013)	Sale price of 23,978 single-family houses, Montreal, Canada [1992-2009].	Spatial difference-in-differences, log-linear model	Proximity to a commuter train station translates into a market premium of up to 11%.
Grimes and Young (2013)	Sale price of 5,729 houses, Auckland, New Zealand [1993-2009].	Spatial difference-in-differences with repeat-sales.	Houses located near a station experience a high price increase of 9.9%.
Kim and Lahr (2013)	Sale price of 13,599 repeat-sales residential properties, New York metro area, USA [1991-2009].	OLS and robust regression.	Properties appreciate at an average annual rate of 18.4% higher around light rail stations than in other parts of the study-area.
Lee and Sohn (2013)*	Assessed land prices, Seoul Metropolitan Area, South Korea [no date].	Linear, log-log, semi-log and, Box-Cox transformed models with ML.	Land price of areas along at-grade or elevated railways are much less (US$798 /$m^2$) than those along underground railways, all else being equal.
Chatman *et al.*(2012)	Sale price of 31,470 residential properties, Southern New Jersey, USA [1989-2007].	OLS, Log-linear.	The net impact of the New Jersey River line on the owned housing market is neutral to slightly negative.
Ibeas *et al.* (2012)	Asking price of 1,562 residential properties, Santander, Spain [06/2009].	OLS, Spatial autoregressive model, Spatial Error model, Spatial Durbin model.	Premium of 1.8% for each additional transit line in the study area.

Author (Year Published)	Data and Location [Year]	Method	Key Results
Banister and Thurstain-Goodwin (2011)*	Sale price of residential, commercial and retail properties, London, UK [1997-2003].	Geographically weighted regression.	The JLE is responsible for a 75% increase in the price of residential property values.
Debrezion *et al*. (2011a)	Sale price of 64,095 residential housing units, Netherlands [1996-2001].	OLS, semi-log.	Elasticity of distance to the nearest railway station: –0.01 for Amsterdam, -0.02 for Enschede, and insignificant for Rotterdam.
Pagliara and Papa* (2011)	Residential and non-residential properties within 16 catchment areas and eight control areas, Naples, Italy [2001-2008].	Comparative analysis	Property values in station control areas are lower than in the stations' catchment areas.
Habib and Miller (2008)	Sale price of 250,000 housing units, Greater Toronto Area, Canada [1987-1995].	Two-level spatial and mixed two-level spatio-temporal random effects models, Box-cox transform.	Prices decrease by 0.67% for every additional km from a subway station, and by 0.16% per km from a regional transit station.
Du and Mulley (2006)	Asking price of 2,837 residential properties, Tyne and Wear, UK [05/2004].	Geographically weighted regression model.	The closeness of metro stations raises the prices of properties by more than £20,000.
Gibbons and Machin (2005)	Sale price of 15,943 housing units, London, England [1997-2001]	Spatial difference-in-differences estimator, log-linear.	Price increases by 1.5% for a 1 km reduction in distances to a train station; the capitalized value of the distance reduction was ~£2500 in 2001.
Highways, arterials and overall accessibility			
Xiao *et al*. (2016a)	Sale price of 16, 297 residential properties, Cardiff, Wales [2001-2007].	OLS, log-linear.	Mixed result for different accessibility measures.
Li and Joh (2016)	Sale price of 3,495 condominium and 12,149 single-family houses, Austin, Texas [01/2010-11/2012]	Spatial Cliff-Ord model.	A 1% increase in bike score (transit score) significantly increases condominium property values by 0.30% (0.3946%) and single-family property values by 0.0279% (0.099%).
Xiao *et al*. (2016b)	Asking price for 2,704 residential properties, Nanjing, China [2005-2010].	OLS, First difference, Fixed effect model.	Street network connectivity not only exhibits a positive impact on house price but also a negative one.

Author (Year Published)	Data and Location [Year]	Method	Key Results
Efthymiou and Antoniou (2015)	Sale price of 19,703 and rents of 18,311 residential properties in Athens, Greece [2011 and 2013].	Spatial Error model.	Impact of metro stations (<500 m) declined by 42.5% on purchase prices and by 62.5% on rents; mixed impact depending in transportation system.
Concas (2013)	Sale price of 29,156 single-family detached dwelling, Florida, USA [2000-2011].	Spatial autoregressive difference-in-differences estimator with spatial errors.	Parcels treated with accessibility improvements limited-access roadways exhibit a price premium of 3.4% to 7.3% at project opening.
Efthymiou and Antoniou (2013)	Asking price of 8,066 residences and asking rent price of 8,400 residences, Athens, Greece [09/2011 - 01/2012].	OLS, spatial lag, spatial error, Spatial Durbin, and GWR models.	Find a positive or negative impact depending on the type of transportation system.
Debrezion *et al*. (2011b)	Rent of 11,298 offices, Netherlands [Since 1983].	OLS, Log-linear.	If road accessibility of the workforce doubles, the rental level of offices increases by ~1.5%.
Giuliano *et al*. (2010)	Sale price of 22,552 land parcels, Los Angeles area, USA [2001].	Random coefficient model, log-log.	Adding job and freight accessibility increases land values by 15%.
Martinez and Viegas (2009)	Asking price of 12,488 residential properties, Lisbon, Portugal [02/2007].	OLS and Spatial Lag model, semi log.	Road accessibility coefficients ranges: -11.05% to -7.32% for hierarchy 1; -4.58% to -8.63% for hierarchy 2; & −5.84% to −3.80% for hierarchy 3.
Vadali (2008)	Sale price of 220,000 residences in Dallas County [1979-2000] and 34,643 in Collin County, Texas [1996-2000].	Spatial Lag and Spatial Error model, semi log.	9% premium within 0.25 to 1 mile for detached houses after the opening of a toll way.
Du and Mulley (2007)*	Asking price of residences, Sunderland, UK [08/1999, 04/2002, 03/2003].	Statistical analysis	No positive change in property prices as a result of new transport infrastructure.
Mikelbank (2004)	Sale price of 5,508 single-family detached houses, Columbus, OH, USA [1990].	Spatial lag model, semi log.	A 7% discount applies to houses located within 0.25 miles from a highway
TOD and BRT Accessibility			
McIntosh *et al.* (2015)	Price of 462,476 residential land parcels, Perth, Australia [2001-2011].	OLS, Log-log	A rapidly growing land value increase by up to 40 % due to the introduction of a new rail line.
Kay et al. (2014)	Estimated median market value of 451 block groups, New York, USA [2013].	Spatial Error model.	Block groups one mile from a study transit station are expected to have property values 6.3% lower than block groups one half mile away.

Author (Year Published)	Data and Location [Year]	Method	Key Results
Mathur and Ferrell (2013)	Sale price of 131 single family houses [1991-1995], 421 units [1996-2003], and 227 units [2004-2006], San Jose, California, USA.	OLS, Spatial Lag, Spatial Error and Spatial Durbin model, semi-log.	An average home sale price increases by $21,000 (or 3.2%) for every 50% reduction in the distance between a home and TOD.
Jun (2012)*	Rent data, Seoul, South Korea [No date].	An urban simulation model (SMIUM)	Residential rent increased in the CBD by US $0.60 per m^2 after the introduction of the BRT.
Duncan (2011)	Sale price of 3,374 condominiums, San Diego, California, USA [1997-2002].	OLS-fixed effect model	The estimated station area premium for a good pedestrian neighborhood approaches $20,000 and can exceed 15% of sale value.

Notes: HPM = hedonic price model; OLS = ordinary least squares; GWR = Geographically Weighted Regression; TOD = transit oriented development; BRT = bus rapid transit. * No information available on number of observations and/or sale/asking price.

Table 2.2 Summary of Selected Developing Country Hedonic Studies of Accessibility (2005-2015)

Authors (Year Published)	Data [Period Analyzed] Models	Structural Attributes Transport / Accessibility / Location Attributes Neighborhood and Other Attributes	Key Results
Shyr *et al.* (2013)	• Asking price of 5,291 residences in Hong Kong; sale price of 2,999 residences in Kaohsiung & 4,068 in Taipei, Taiwan [2008]. • Log-linear with fixed-effect and Box-Cox with fixed effects.	• Floor area; house age; floor number. For detailed model: number of floors, lot size (as appropriate), dwelling type. • Network distance to: CBD, secondary centers, nearest MRT/MTR station, airport, high-speed rail station (Taipei and Kaohsiung). • Time of sale; lease expiration. Neighborhood fixed effects or detailed attributes: road width, waterfront within 200 m, commercial zone, residential zone, population density, % of residents with college degree, % of foreign residents, city core.	Transit distance elasticity ranges from -0.016 in Hong Kong to -0.044 in Taipei and -0.072 in Kaohsiung.
Zhang and Wang (2013)	• Asking prices for 592 residences, Beijing, China [1999-2007]. • OLS, spatial lag and spatial error models.	• Housing type; home finishing quality; floor area ratio. • Network distance to nearest transit station; distance to city center; access to: expressway, city rail, health centers, parks, sport facilities. • Green area ratio; dummy variables for Districts 1 and 2; year of sale; home availability.	For every 100 m closer to a transit station, prices increase by 0.35%.
Cervero and Kang (2011)	• Assessed value of 126,426 residential and 61,484 non-residential land parcels, Seoul, Korea [2001-04 & 2005-07]. • Log-log multilevel regression models for residential and non-residential properties: 2001-04 (pre-BRT) and 2005-07 (post-BRT).	• Binary variable for: office, commercial raw land, mixed use, and mixed use raw land; building coverage ratio; floor area ratio. • Binary variables for distance to bus stop (30 m increments from 0 to 300 m; network distance to: nearest freeway ramp, pedestrian entrances; distance to: CBD (City Hall), nearest subway station, nearest urban arterial, Han River. • Density: population, employment; % of residents: with college degree, 40-60 years old, over 60; ratio: park density, developed land, road area, retail area; % of residential permits, % of commercial permits.	Land price premiums: up to 10% for residences within 300 m of BRT stops and >25% for other non-residential uses over a 150 m impact zone.
Andersson *et al.* (2010)	• Sale price of 1,550 dwellings, Tainan Metro Area, Taiwan [2007]. • HPM with log-linear, semi-log and Box-Cox transforms.	• Floor area; lot size; building age; number of floors; shop/dwelling use; street frontage. • Distance to: CBD, HSR station, nearest freeway interchange, Tainan Science-based Industrial Park. • Road width; commercial zone; residential zone; mean household income in district; % population college educated in district.	High-speed rail accessibility has at most a minor impact on house prices.

Authors (Year Published)	**Data [Period Analyzed]** **Models**	**Structural Attributes** **Transport / Accessibility / Location Attributes** **Neighborhood and Other Attributes**	**Key Results**
Munoz-Raskin (2010)	• Asking price of 130,692 housing units, Bogotá, Colombia [2000-2004]. • OLS log-level and level-level.	• Average size of property and number of units in development; property type; year of development. • Within 10 min walk of: BRT, nearest trunk line station, feeder lines, Autopista Norte station; within 5 min (& 5-10 min) of nearest: trunk line station, feeder line. • Categorical variables for socio-economic attributes; interactions between year and some proximity variables.	Properties within 10 min of walking of BRT trunk stations are valued 4.8% higher.
Cervero & Murakami (2009)	• Sale price of 905 flats near 3 rail stations, Hong Kong [2005]. • OLS.	• Building age; building floor; unit size. • Distance to MRT station; presence of main road next to building. • Presence of green space or park near building; R+P project; TOD.	Housing price premiums range from 5% to 30% for R+P projects.
Rodriguez and Mojica (2009)	• Asking price of 1,674 (before) and 2,301 (after) residential properties, Bogotá, Colombia [2001-2006]. • OLS, WLS, and spatial lag models.	• Binary variable for apartment; floor number; categorical variable for building age; number of bedrooms, of bathrooms; floor area; garage spaces. • Within 150 m of BRT right of way; distance to station; binary variables for located within 500 m of major road served by competing public transportation service. • Neighborhood socio-economic stratum; population density; % of area: industrial, commercial, institutional, vacant / empty, park / open space; homicide rate; annual price change; 12 variables that measure whether BRT expansion caused price changes.	Price increases by 13% to 14% after BRT extension.
Pan and Zhang (2008)	• Sale price/m^2 of 503 residential units, Shanghai, China [2007]. • Semi-log OLS model.	• Building age; single use residential; building area. • Distance to: nearest metro station, People Square, sub-centers (Lujiazui and Xujiahui), Nanjing Rd; located in inner ring. • Presence of: neighborhood shopping, elementary school; green area ratio.	Transit proximity premium: about 152 yuan/sq. m for every 100 m closer to a metro station.
Chalermpong (2007)	• Asking price of 226 multifamily units, Bangkok, Thailand [2004-2005]. • OLS, spatial lag, and spatial error linear-linear and log-linear HPM.	• Living area; building age. • Distance to: BTS station, nearest arterial road; number of stations to Siam (the only interchange station of the BTS system). • NA.	A price discount of US$18 for each additional m from an arterial.

Authors (Year Published)	Data [Period Analyzed] Models	Structural Attributes Transport / Accessibility / Location Attributes Neighborhood and Other Attributes	Key Results
Celik and Yankaya (2006)	• Asking price of 360 multifamily residences, Izmir, Turkey [12/2003 - 03/2004]. • Linear and log-linear HPM.	• Floor area; building age; number of floors; corner building; central heating; construction quality. • Walking distance to: nearest rail station, nearest bus stop. • NA.	Within walking distance: proximity to a rail station is valued at $250-300 per m.
Yiu and Wong (2005)	• Sale price of 2,095 flats in Hong-Kong [1991-2001]. • Semi-log HPM.	• Age of property; saleable floor area; floor level. • NA. • Binary variables: 118 for time periods (months), 5 for zones; 2 for before/after; and interactions between time/period and zone variables.	Positive price expectations well before the completion of the tunnel.

Looking at dependent variables, it is noteworthy that out of the 11 papers in Table 2.2, only 5 worked with actual sale prices, and the others analyzed asking prices or assessed land values. Other differences with a number of recent hedonic studies of accessibility performed in North America and Europe (see Table 2.1) include often smaller sample sizes and a smaller set of structural attributes among explanatory variables. These differences reflect the difficulty of collecting large, high quality datasets in many developing countries.

Although understanding the impacts of transportation facilities on property values is of considerable interest around the world, it is therefore not surprising that I could not find published studies of accessibility for least developed country cities in peer-reviewed international journals. Apart from the difficulty of gathering data there, least developed countries often suffer from a dearth of transport investments (especially for transit) compared to North America, Europe, or more advanced economies in Asia.

2.2.3 Measuring Accessibility

Accessibility is typically defined as the ease of reaching desired activities such as employment, retail shopping, and healthcare (Du and Mulley, 2006). A review of selected papers (see Table 2.3) shows that different approaches have been employed for two reasons: there is no uniformly best way to measure accessibility (Handy and Niemeier, 1997; Geurs and Wee, 2004) and consistent data across wide geographical areas are often challenging to source (Halden, 2002).

The two most common approaches for measuring accessibility are minimum distance (in time or space) (e.g., see Ahlfeldt and Wendland, 2011; Andersson *et al.*, 2012; Dubé *et al*., 2013) and gravity potential (e.g., see Ahlfeldt, 2013; or Osland and Thorsen, 2013).

The minimum distance approach has been implemented in various ways. Many studies have relied on straight line distance, which was the initial approach of choice (e.g., see Debrezion *et al.*, 2011a, Mathur and Ferrell, 2013, or Efthymiou and Antoniou, 2015). However, as Geographic Information System (GIS) software has become more powerful, shortest network distance has gained in popularity (e.g., see Baranzini and Schaerer, 2011; Andersson *et al.*, 2012; or Shyr *et al.*, 2013) and so has shortest network travel time (e.g., see Ahlfeldt and Wendland, 2011; Crespo and Grêt-Regame, 2012, 2013; or Schläpfer *et al.*, 2015).

Although minimum distance measures have long been popular, gravity-based accessibility measures have been receiving increasing attention in housing market studies (Ahlfeldt, 2011). These measures are based on the principle that the accessibility of a destination is a decreasing function of the relative distance to other potential destinations with each destination weighted by its size or the number of opportunities available there (Osland and Thorsen, 2008). To implement this approach in the case of employment for example, it is necessary to know the number of jobs available at each destination (employment zone) and the distance or travel time between each origin and each destination. In trip based gravity models, each destination is weighted by the number of trips attracted to that zone (Adair *et al.*, 2000).

While a few studies reported disappointing results with gravity-based accessibility models (Adair *et al.*, 2000; Giuliano *et al.*, 2010), many others have found that they perform much better than the shortest distance approach (e.g., see Ahlfeldt, 2013; Osland and Thorsen, 2013; Wu *et al.*, 2013). Gravity-based models are particularly well-suited for measuring accessibility to employment because job opportunities are likely proportional to the number of residents. In contrast, the accessibility to some services such as education may best be measured with distance to the nearest location (Du and Mulley, 2006). As a result, several studies have relied on mixed approaches with a gravity-based model for job accessibility and distance based measures for access to transport facilities (e.g., see Debrezion *et al.*, 2011b; Diao, 2015).

In summary, the literature suggests that the choice of accessibility measures depends on the purpose and the context of each study but also on data availability.

2.3.4 Hedonic Studies of Apartment Rent

Since the dependent variable in our study is apartment rent, we searched Google Scholar for hedonic studies of apartment rents (not limited to transportation accessibility) published in English between 2005 and 2015 to find relevant determinants of apartment rent that could be derived from available data sources. My findings are summarized in Table 2.4. Only Babalola *et al.* (2013) analyzes data from a developing country; most (11) of the other papers focus on Switzerland, a couple analyze Greek data, and the other 4 work with U.S. data.

I organized explanatory variables in three broad categories: structural attributes, transportation/accessibility/location attributes, and neighborhood and other attributes. The most common structural variables are floor area, number of bedrooms, and building age. Other structural variables are contextual. They include: number of bathrooms; availability of heating, air conditioning, elevators, parking/garage; number of floors and floor level, and apartment type.

Only a handful of studies in Table 2.4 focus primarily on transportation (Löchl and Axhausen, 2010; Efthymiou and Antoniou, 2013, 2015), but most include transportation or accessibility variables in their models. The most common accessibility variables are distance to city center and distance to transportation facilities such as train, tram, or metro stations, bus stops, and airports. Measures of accessibility vary: some studies used straight line distances (Efthymiou and Antoniou, 2013, 2015; Fahrlander *et al.*, 2015), others relied on network distance (e.g., see Banfi *et al.*, 2008; Baranzini *et al.*, 2010; or Baranzini and Schaerer, 2011), and a few used driving time (Löchl and Axhausen, 2010; Crespo and Grêt-Regame, 2012, 2013; Schläpfer *et al.*, 2015).

Table 2.3 Summary of Selected Gravity-Based Hedonic Studies of Accessibility

Author(s) (Year Published)	**Data and Location [Year]**	**Method**	**Key Results**
Ahlfeldt (2013)	Sale price of 60,748 residential properties, Greater London, UK [01/1995-07/2008].	OLS, nonlinear least squares, difference in difference.	The spatial scope of labor market effects is ~60 minutes; doubling accessibility increases the utility of an average household by ~12%.
Osland and Thorsen (2013)	Sale price of 2,788 single-family detached houses, Norway [1997-2001].	OLS and spatial Durbin model (SDM).	The SDM is better than OLS, although spatial externalities are low in this housing market.
Wu *et al.* (2013)*	Housing price, Shenyang, China [2001].	OLS, semi-log model.	Inter-city accessibility has a larger impact on housing prices than inner city accessibility.
Osland and Pryce (2012)	Sale price of 6,269 dwelling units, Glasgow, Scotland [2007].	OLS, spatial error, spatial lag, spatial Durbin model.	The value of accessibility is not monotonic: moving away from an employment node, house prices first rise and then decline.
Ahlfeldt (2011)	Sale price of 33,843 residences, Berlin, Germany [01/2000-12/2008].	Difference-based semi parametric (SP) regressions, OLS, spatial autoregressive model.	The gravity-based approach is better than standard measures to capture accessibility.
Osland and Thorsen (2008)	Sale price of 2,788 single-family detached houses, Norway [1997-2001].	Log-log hedonic models with various accessibility measures estimated via maximum-likelihood or OLS.	Housing prices fall with increasing distance from the CBD even when labor market accessibility is accounted for.
Adair *et al.* (2000)	Sale price of 2648 residences, Belfast, Northern Ireland [1996].	OLS, log-linear model.	The variance in house prices explained by accessibility for Belfast is <2% for most models. It rises to 14% for the West Belfast terraced model.

*Did not mention the number of observation.

Table 2.4 Explanatory Variables in Selected Hedonic Studies of Apartment Rents (2005-2015)

Authors (Year)	Data [Period Analyzed]	Structural Attributes	Transport / Accessibility / Location Attributes	Neighborhood and Other Attributes
Efthymiou and Antoniou (2015)	Rents of 18,311 residences in Athens, Greece [2011 and 2013].	Total area; building age; floor level; parking; fireplace; heat; A/C; storage area; single family or multifamily.	Distance bands to stations (bus, metro, rail, suburban rail, tram), airport, port, marina, ring-road, coastline.	Dummy variables for Northern suburbs; sea view; front orientation.
Füss and Koller (2015)	Rents of 28,728 apartments in the canton of Zurich, Switzerland [2002 to 2014].	Number of rooms; elevator; parking; building age; garage; apartment type.	NA.	NA.
Fahrlände *et al.* (2015)	Rents of 65,301 Zurich properties, Switzerland [2012].	Year of construction; building type; building condition; floor area; number of rooms; floor level.	Straight line distance to local services; binary variables for public transport group, proximity to lake.	Urban center; building zone; landscape quality; exposition; number of services; maximum aircraft noise, nighttime road traffic noise, nighttime rail noise.
Schläpfer *et al.* (2015)	Rents of 162,523 apartments in Switzerland [2001 to 2007].	Room size (mean area); number of rooms; type of apartment; building age.	Travel time to central services by car; inner city location dummy; distance to nearest road / highway; distance to nearest: hill site, major lake, river.	% of area: large buildings, industry, parks, national and fen landscape, forests, species rich grassland; length of: shoreline, hiking & bike trails; meters of: mountain cableways, high voltage lines; number of: mobile antennas, land uses; heritage & UNESCO towns; cultural objects; flood risk; income per capita; year of offer; road & railway noise; tax burden; % foreigners; March solar radiation; view of lake, river.
Efthymiou and Antoniou (2013)	Web-advertised rents of 8,400 residences in Athens, Greece, [09/11 to 01/12].	Total area; building age; floor level; parking; fireplace; heat; A/C; storage; single / multifamily.	Bands for straight line distance to stations (bus, metro, railway, tram), port, airport, ring roads, CBD, coastline, archaeological sites.	Dummy variables for: Northern suburbs, university area, low population density, high education level; sea view; orientation.
Babalola *et al.* (2013)	Rents of 150 houses, Modibbo Adama University	Building age; water supply in house; electricity in house; toilet in house.	Proximity to university.	Mortgage bank credit; tenement rate.

Authors (Year)	Data [Period Analyzed]	Structural Attributes	Transport / Accessibility / Location Attributes	Neighborhood and Other Attributes
	of Technology, Nigeria [no date].			
Crespo and Grêt-Regame (2013, 2012)	Rents of 3500 residences in the Canton of Zürich, Switzerland [12/04 to 10/05].	Floor area; type of property; building age.	Average driving time to the Zurich CBD; regional transit accessibility to employment; distance to nearest rail station; presence of highway nearby.	Population density; % of foreigners; number of jobs in hotels and restaurants nearby; terrain slope; tax level; evening solar exposure; visibility of lake; noise >52 dB.
Baranzini and Schaerer (2011)	Mean annual net rent of 12,932 residential properties in Geneva, Switzerland [2005].	Building year; number of rooms; floor level; building height; attic dwelling.	Network distance to: city center; nearest primary school; nearest public transport stops.	Areas: natural/built environment, water, urban parks, agriculture, industry; diversity indexes for natural/built land uses; tenancy change in past year; views of: natural/built environments, water, parks, agriculture, industry, fountain, cathedral; natural/built view diversity indexes.
Donovan and Butry (2011)	Rents of 1000 houses in Portland, Oregon, USA [10/09 to 01/10].	Finished house area; lot area; house age; number of bedrooms, of bathrooms; heating type; A/C; number of fireplaces; garage.	Straight line distance to city center, to nearest park.	Located in zip code; number of reported crimes within 0.25 miles; area of nearest park; number and crown area of street trees fronting lot, of trees on lot.
Löchl and Axhausen (2010)	Asking rent of 8,592 dwellings in the Canton of Zürich, Switzerland [12/04 to 10/05].	Floor area; elevator; fireplace; number of balconies; garden terrace; construction year.	Travel time to CBD; regional car access to employment; public transport accessibility to employment; Euclidean distance to nearest rail station; proximity to autobahn.	Population density; % of foreigners; number of jobs in hotels and restaurants within 1 km; daily average noise >52db; local income tax; slope; visibility of lake, of landscape; evening solar exposure index.
Baranzini *et al.* (2010)	Rents of 2,840 apartments in Geneva, Switzerland [2003].	Number of building floors, of rooms; floor area; floor level; building age; renovated building; elevator; attic; balcony; separated toilet.	Located in city center; network distance to nearest park, to nearest primary school.	Density of historical buildings; privately/publically owned; owner is insurance company or pension fund; duration of residency; with terrace/garden; noise variables; view of lake, of mountains.
Allen *et al.* (2009)	Rents of 20,131 houses in the Dallas-Fort Worth Metropolitan area,	Building area; age; number of bedrooms, of full and half bathrooms; pool; number of floors; fireplace; brick or	NA.	Pets allowed; security system; no smoking; fenced yard; listing and leased variables.

Authors (Year)	Data [Period Analyzed]	Structural Attributes	Transport / Accessibility / Location Attributes	Neighborhood and Other Attributes
	Texas, USA [2003 and 2004].	wood, siding; central air; central gas heat.		
Banfi *et al.* (2008)	Rents of 6,204 apartments in Zürich, Switzerland [2003].	Floor area, number of rooms; floor level; building age; integrated kitchen; second toilet in dwelling; balcony; elevator; garden; renovated.	Network distance to city center.	Years in residence; PM_{10} concentration; noise annoyance at night; antenna in 200 m radius; for profit or non-profit rental.
Schaerer *et al.* (2008)	Rents of 3,327 Greater Geneva houses and 3,194 Zurich houses, Switzerland [2003].	Construction year; renovated; elevator; floor level; number of rooms; floor area per room; terrace; garden; penthouse; balcony; separated toilet.	*Distance to: lake, nearest forest, nearest park; road traffic.	Old town, northern part of urban area; % area of: water, forest, agricultural, parks; land-use diversity index; private vs. public owner; owner is insurance company or pension fund; duration of residency; lake view; mountain view; daytime noise.
Baranzini *et al.* (2008)	Rents of 42,162 Geneva properties, and 26,489 Zurich properties, Switzerland [2000].	Building construction year; floor level; number of floors; number of rooms; floor area per room and per inhabitant; kitchenette; attic; gas heating.	Network distance to: nearest transportation stop, nearest park, city center.	% of: foreigners, poorly & highly educated foreigners; population density; % of parks in the district; % of trees in the district; head of household characteristics; privately vs. publically owned building; same tenant for ≥5 years; daytime noise.
Wilson and Frew (2007)	Rents of 533 apartments in Portland, Oregon, USA [1992 to 2002].	Number of bedrooms, of bathrooms; fireplace; laundry facility; laundry hookup; exercise area; pool; covered parking; cable hookup; A/C.	Network distance to highway, distance to intersection, distance to city center.	NA.
Valente *et al.* (2005)	Rents of 4,750 apartments in 8 U.S. cities [2002].	Size of average unit; total number of floors in complex; complex age; renovation; number of units in complex.	NA.	NA.

Notes: * Did not mention whether used calculated network distance or straight line distance. A/C designates air conditioning.

Although most studies considered in Table 2.4 include neighborhood characteristics, they share no common neighborhood variables, and several papers even included none in their models. Other variables characterize noise and air quality, building restrictions, landscaping, and views, as well as socio-economic characteristics of the surrounding population or of the building owner, local tax rates, and the availability of credit.

In summary, hedonic studies of apartment rents share only a few core explanatory variables. The inclusion of additional explanatory variables depends on context, the purpose of each study, and (likely) data availability.

2.3 BACKGROUND AND DATA

Our study area is the Rajshahi City Corporation (RCC), which is located in the north-west of Bangladesh (see Figure 2.1) on the banks of the Padma River (also known as the Ganges River). With an area of 48.06 square kilometers (DDC, 2004) and an estimated population of 0.45 million in 2011 (BBS, 2013), it is the fourth largest city of Bangladesh (the metropolitan area has double the area and the population). It is widely known as the 'Silk City' of Bangladesh, as attractive silk products are cheaper and of greater quality there than anywhere else in the country.

Rajshahi is well connected to the rest of Bangladesh by air, road, and rail. Buses serve 20 inter-district and 12 intra-district routes (Bangladesh is organized in 7 divisions and 64 districts), and the national railway operates three stations in Rajshahi City. However, intra-city movements depend entirely on road transportation. Rajshahi's 571 kilometers of roads (BBS, 2013) carry vehicles ranging from push carts to modern cars, although the latter are still very rare: fewer than 1% of households own private cars (Haque, 2014), 8% have motorcycles, and 57% own bicycles. Other vehicles include push carts and rickshaws, which provide both door to door and feeder service to long distance buses and railways as there is no city bus service in Rajshahi City. Official statistics confirms the rarity of private cars in Bangladesh: in January of 2015, there were only 268,246 registered private cars in the whole country (over 166 million people; CIA, 2014), and three quarters of these cars were in Dhaka, the country's capital (BRTA, 2015).

As mentioned above, the dearth of housing market data in developing countries hinders investigating the capitalization of transportation accessibility in housing markets. Bangladesh is no exception: there is no secondary source of housing market data for Bangladeshi cities, and when data are available, there are typically huge differences between official and actual prices (Islam *et al.*, 2007). Moreover, collecting housing data is much harder than in developed countries due to the lower literacy rate (56.1%), which cannot be offset by the penetration of cell phones (63.7%) (BBS, 2012), so the only way to collect reliable data is via in-person interviews.

The housing data analyzed in this study were therefore collected by trained interviewers who randomly selected renters within Rajshahi City during June of 2006. Even though this dataset is older than we would like, it is still highly relevant because local conditions (income, city characteristics, and modal shares) have not changed much over the past few years and they are representative of many cities in South Asia.

Of the 669 renters surveyed, 111 lived in single-unit housing and 558 in multi-unit residential properties. Single-unit properties for rent in Rajshahi City typically consist of single rooms with a bathroom but without a backyard, and their condition is usually not as good as the condition of multi-unit rentals. As a result, their rents are lower than for comparable multifamily units and they are typically inhabited by lower income people. Since single-unit rentals form a distinct sub-market, our analysis focuses on the multi-unit residential properties in our dataset.

Collected socio-economic data of our respondents include gender, literacy rate, and age of each respondent, as well as monthly household income and its main source. These data are summarized in Table 2.5 and contrasted with population data from the Rajshahi City Corporation as data specific to renters in the city are not available.

Table 2.5 Characteristics of Multi-Family Dweller Respondents (N=558) vs. Rajshahi City Population

Variable	Values	Sample (%)	Population (%)
Gender of respondent	*Male*	60.75	51.80
	Female	39.25	48.20
Literacy of Respondent		76.50	74.10
Monthly household income (in BDT*)	*< 5000*	6.99	15.37
	5000 to 6000	10.93	6.90
	6000 to 7999	44.90	12.69
	8000 to 9999	23.67	10.29
	10000 to 12,500	7.66	12.91
	12,500 to 15000	2.72	6.85
	> 15000	0.18	34.99
	Missing	2.95	NA
Main household source of income	*Agriculture and fisheries*	7.50	5.00
	Business	24.10	26.00
	Transport communication	8.10	10.00
	Construction	3.50	4.00
	Salary wage	24.20	34.00
	Artisan / skilled labor	1.76	2.00
	Others	25.70	19.00
	Missing	5.14	NA
Age of respondent (in years)	*20 to 24*	2.56	11.17
	25 to 34	13.74	19.06
	35 to 44	53.38	13.23
	45 to 49	15.60	4.48
	50 to 59	7.51	5.91
	Above 60	1.21	5.73
	Missing	6.00	NA

Sources: in-person interviews performed in 2006, and data from BBS (2011 and 2013)
Note. ♠: 1 USD = 66.89 BDT in June 2006.

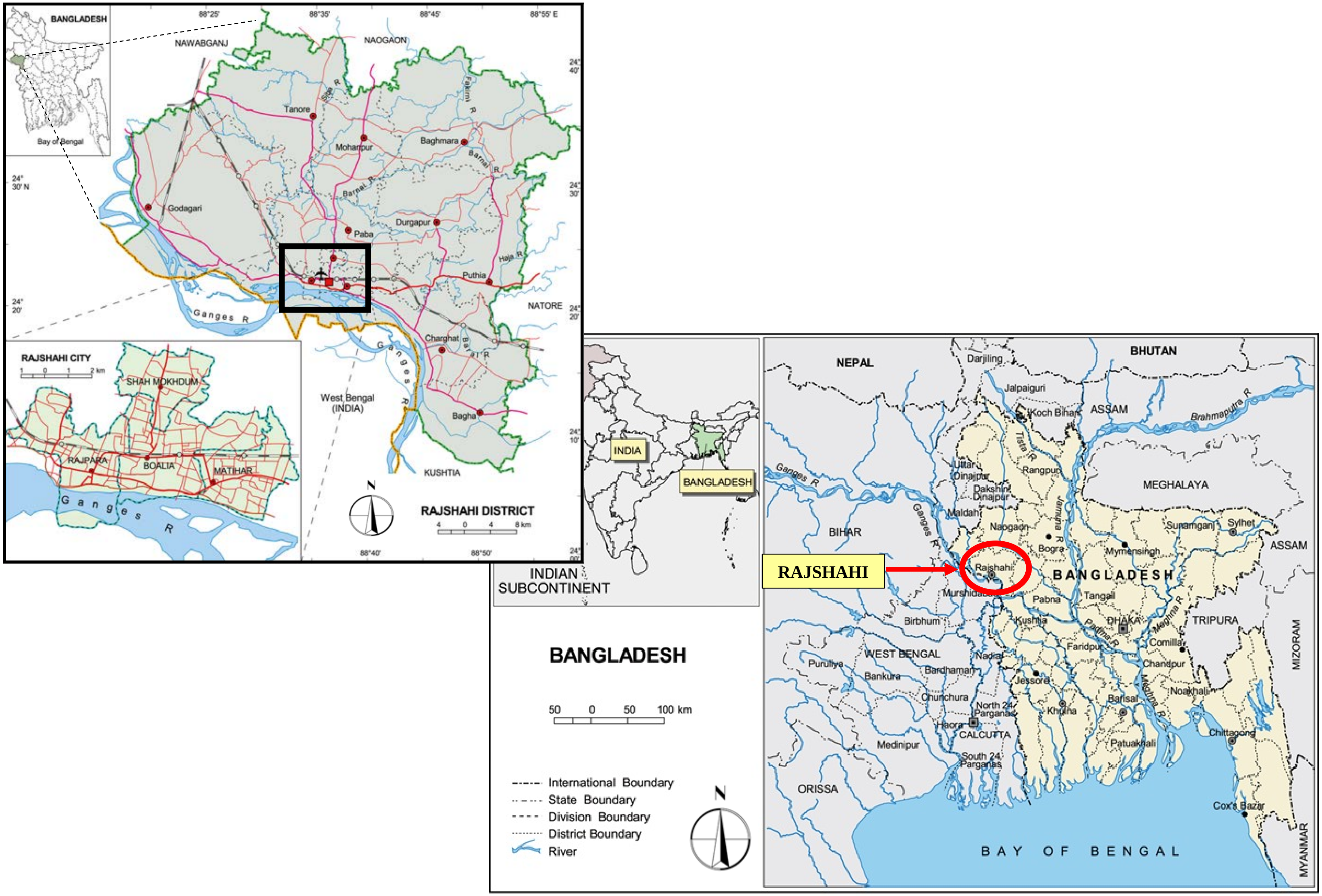


Figure 2.1 Location of Rajshahi City in Bangladesh

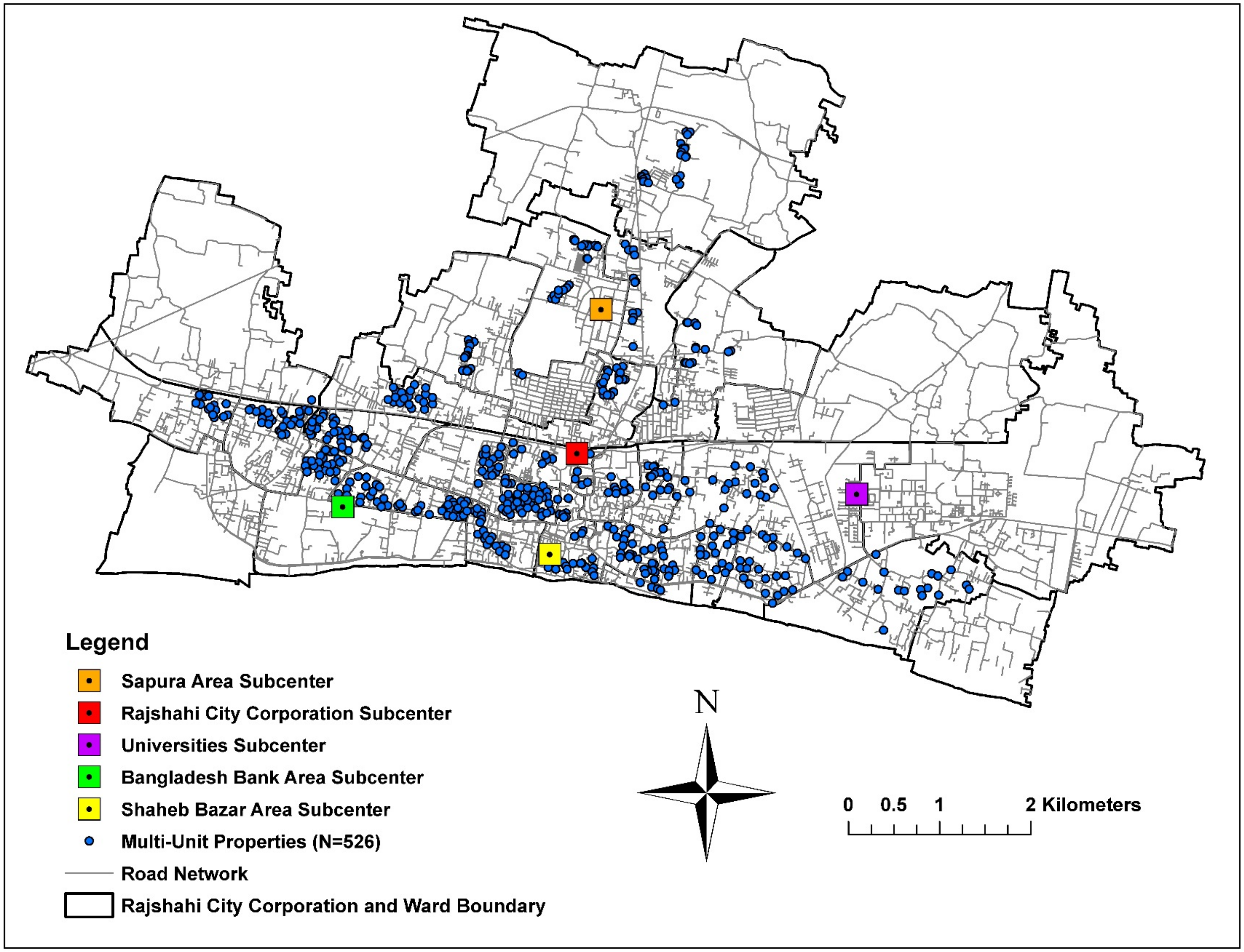


Figure 2.2 Location of Analyzed Multi-unit Properties in Rajshahi City

As expected from a culture where the head of household is typically male and women are uneasy about answering interviews, 60.75 % of respondents are male (vs. 51.80% for the city). The literacy rate of our respondents (76.5%) is comparable to the Rajshahi City average (74.1%). We also note that the very poorest (income < 5,000 BDT: 6.99% vs. 15.37%) and especially the most affluent (income > 15,000 BDT: 0.18% vs. 34.99%) of Rajshahi residents are under-represented in our sample, which is not surprising since our survey targeted renters, although the main sources of income are comparable. Finally, people in the 35 to 59 age groups are over-represented, which again makes sense since it reflects characteristics of heads of households. Overall, there is therefore no glaring reason to doubt that our respondents are representative of renters in Rajshahi City.

Our dependent variable is the monthly rent paid to property owners, which is consistent with standard neoclassical theory (Muth, 1969) where rental prices are assumed to reflect land value. Although most published hedonic studies of the housing market explain sale price, a number of empirical studies have analyzed urban rents (see Table 2.4). Moreover, official sale prices are highly unreliable in Bangladesh as only a fraction of the actual sale price is typically reported to avoid paying taxes (e.g., see Mitra *et al.*, 2005, for Dhaka City). Note that the rents I analyze are free of utility costs (water and electricity), which are paid separately by tenants.

Among the explanatory variables, the following structural characteristics were collected during the survey: usable living area, number of bedrooms, number of bathrooms, and building age. A boxplot of the usable living area variable shows that properties above 2,600 square feet (241.55 m^2) are larger than 1.5 the interquartile range, so I removed these 30 observations from the data used to estimate our hedonic models.

My models do not include lot size because field work indicated that the multi-unit buildings in our sample occupy approximately their entire lot. This was confirmed by examining building data and property boundaries using geographic information system (GIS) software. In addition, since flooding risk is non-negligible in Rajshahi City, I measured the straight line distance from each property in our sample to the nearest drainage network using GIS.

A number of other structural attributes considered in wealthier housing markets (e.g., heating / cooling systems, and the presence of a fireplace, a garage, or a swimming pool) are not relevant for Rajshahi City. Number of building floors, floor level of the apartment considered, and view have also been used in some hedonic studies, but unfortunately this information is not available here (and it is seldom available in the studies summarized in Table 2.1). I do not believe that these variables are important here because most buildings in Rajshahi have less than four floors and only buildings with 6 floors or more are required to have elevators.

To guard against the perils of the bias created by omitting locally constant variables, I included Ward level fixed effects in our models. This caused our sample to lose another two observations that were by themselves in a ward (all other wards have at least 10 observations), so the sample I used for modeling has 526 observations. Other factors such as school quality, crime rates, differentiated municipal tax rates, the quality of local fire services, or residential zoning controls, may also influence residential property values. However, primary school quality and the quality of municipal services are relatively homogeneous in Rajshahi City at the ward level, property tax rates are uniform throughout the city, and no zoning controls were in effect in 2006 when our survey was conducted. To capture the impact of additional unobserved effects, I relied on spatial effects as explained in the next section (also see LeSage and Pace, 2009).

Measures of accessibility make up my third group of explanatory variables. Ideally (see Section 2.3) I would have liked to use a mixed approach like Diao (2015) with a gravity based accessibility measure for employment centers and distance based accessibility measures for other services. Unfortunately, the data necessary to implement gravity-based measures of accessibility were not available which forced me to rely on distance based accessibility measures. Like Shyr *et al.* (2013) and Andersson *et al.* (2012), I measured accessibility in terms of shortest network distance (calculated using ArcView GIS) to major roads, transportation facilities, and activity centers because this approach reflects better than straight line distance the relative travel times of travelers relying on non-motorized transportation.

My first measure of accessibility is the shortest path network distance from a dwelling to the closest of five employment sub-centers (Sapura where several silk factories are located, the Rajshahi City Corporation offices, the University area that includes the Rajshahi University of Engineering and Technology and the University of Rajshahi, a banking center around the Bangladesh Bank, and the Shaheb Bazar area; see Figure 2.2). My second one is the shortest network distance to the nearest intersection with a major arterial road. In addition, a binary variable indicates whether the access road to a dwelling is paved or not.

I also created variables that measure network access distance to the closest of three train stations and to the nearest regional bus stop. My other accessibility variables include binary variables to indicate proximity within 400 m of primary schools (the main neighborhood educational institutions in Rajshahi City), healthcare facilities, and shopping centers. This approach was preferred to shortest network access distance to prevent multicollinearity. In addition, although Rajshahi is not an industrial city, it has a number of factories located in what I call herein small industry employment areas, so I created a variable that reflects accessibility to these employment areas.

Since proximity to transportation facilities may expose residents to additional noise and air pollution, we added binary variables to flag properties within 0.25 km of railway tracks, within 0.3 km of a bus terminal, or within 0.25 km of a highway. Finally, I included a binary variable to indicate proximity (within 0.3 km) to wholesale markets.

Table 2.6 Descriptive Statistics (N=526)

Variable	Mean	Standard Deviation	Min	Max
Dependent Variable				
House rent (in BDT♠, June 2006)	2650	644.49	1300	5300
Structural Variables				
Usable living area (square meters)	142.44	38.12	37.16	241.55
Number of bedrooms	2.78	0.69	1	4
Number of bathrooms	1.58	0.57	1	3
Building age in 2006 (years)	19.29	12.95	2	131
Euclidian distance to nearest drainage network (m)	40.79	78.17	2.3	473.92
Neighborhood Variables				
23 Ward♣ binary variables (see note iv below)				
Accessibility Variables				
Network access distance to nearest of 5 subcenters (Shaheb Bazar, Sapura Area, Rajshahi City Corporation Area, University Area, and Bangladesh Bank Area) (m)	1161.73	535.95	65.72	2688.70
Network access distance to nearest major road (m)	226.94	225.16	0.31	1482.55
Binary Variable: 1 if access road is paved, 0 otherwise	0.66	0.48	0	1
Network access distance to nearest railway station (m)	2270.24	828.72	322.52	4107.65
Network access distance to nearest bus stop (m)	735.3	398.62	42.07	2553.37
Network access distance to nearest small industry (m)	294.97	220.27	0.67	1312.28
Binary: 1 if primary school within 400 m	0.85	0.36	0	1
Binary: 1 if healthcare facility within 400 m	0.53	0.50	0	1
Binary: 1 if shopping center within 400 m	0.07	0.25	0	1
Disamenity Variables				
Binary: 1 if within 0.25 km from railway tracks	0.15	0.36	0	1
Binary: 1 if within 0.30 km of bus terminal	0.04	0.20	0	1
Binary: 1 if within 0.25 km from a highway	0.35	0.48	0	1
Binary: 1 if within 0.30 km of a wholesale market	0.06	0.23	0	1

Notes:

i. ♠: 1 USD = 66.89 BDT in June 2006.

ii. ♣: A ward is the lower-tier administrative unit of Rajshahi City. The number of observations ranges from a low of 10 in ward 20 to a high of 44 in ward 11. Our sample has observations in 25 different wards but the binary variable for Ward 8 was removed to prevent multicollinearity.

iii. “m” designates meters.

iv. Rent data and structural variables were collected during a June of 2006 field survey. Accessibility and disamenity variables were extracted from GIS maps from the Rajshahi Master Plan Project.

Table 2.6 presents summary statistics for the variables considered in this study. An analysis of the multicollinearity of our explanatory variables (either untransformed or log-transformed for continuous explanatory variables) shows that it is not a problem here (all variance inflation factors are < 10) after removing the binary variable for Ward 8. Results with and without this variable show that it has a negligible impact.

2.4 METHODOLOGY

2.4.1 Overview

Following the standard hedonic framework (Rosen, 1974), our starting hedonic price model can be written:

$$\mathbf{P} = f(\mathbf{S}, \mathbf{N}, \mathbf{T}, \mathbf{e}) , \tag{2.1}$$

where **P** is a vector of rental prices; **S**, **N**, and **T** are matrices of structural variables, neighborhood characteristics, and transportation/accessibility/disamenity attributes, respectively; and **e** is a vector of errors. The partial derivative of $f(.)$ with respect to an explanatory variable is an implicit price that represents marginal willingness to pay for the characteristic it represents.

Although Rosen's (1974) framework requires market equilibrium with perfect competition, perfect information, and a continuum of products, Bajari and Benkard (2005) showed that these conditions are not necessary for the hedonic pricing method to be valid. Moreover, MacLennan (1977) argued that equilibrium may be assumed if the housing market does not suffer severe shocks and if the study period is reasonably short, which is the case here.

A difficulty with implementing hedonic models is selecting an appropriate functional form. For ease of interpretation, the dependent variable of many published hedonic analyses is either logarithmically transformed or untransformed (level) (Duncan, 2011), and it is assumed to depend linearly on (possibly transformed) explanatory variables (see Table 2.1). A graphical exploration showed that a log-log functional form (with the logarithm of house rent as dependent variable and log-transformed continuous explanatory variables) is reasonable, but I also estimated level-level and log-level models. I did not consider Box-Cox transformations because they complicate model interpretation and they are not readily available for the spatial models I estimated here (Armstrong and Rodriguez, 2006).

2.4.2 Spatial Dependence and Model

In the housing market, it is natural to expect spatial interactions between nearby properties and it is well known (e.g., see Anselin, 1988) that in the presence of spatial effects, OLS estimates may be biased and inconsistent. These spatial associations could operate via the prices (here the rents) or the structural characteristics of nearby properties, or they could be captured by error terms.

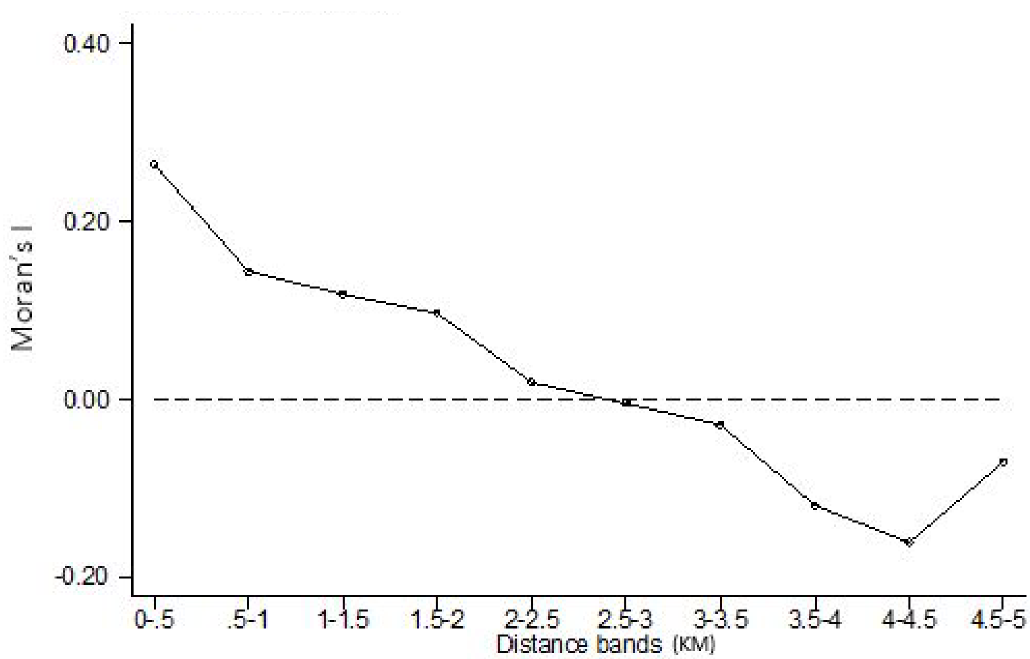


Figure 2.3 Moran's I Spatial Correlogram (log of house rent)

Using Moran's I statistic (Cliff and Ord, 1981), I found that spatial interactions are indeed present among level and log-transformed monthly rents (p-value<0.01) in my dataset and the resulting spatial correlogram (which shows Moran's I versus potential distance bands to assess the extent of spatial dependence) suggested 2.5 km (see Figure 2.3) as an appropriate distance band.

Following Anselin (2005), I then performed Lagrange multiplier (LM) tests for spatial lags and spatial errors. Both tests yielded highly significant statistics ($p<0.001$) with similar magnitudes (20.49 and 17.14 for the LM error and for the LM lag tests respectively), so I estimated a combined spatial-autoregressive model with spatial autoregressive disturbances (SARAR; see Drukker *et al.*, 2013). If N designates sample size and Q is the number of explanatory variables (including a constant), our SARAR model can be written:

$$\begin{cases} \log(\mathbf{P}) = \lambda \mathbf{W} \log(\mathbf{P}) + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \\ \mathbf{u} = \rho \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}, \end{cases} \tag{2.2}$$

where **P** is the N × 1 vector of multi-unit residential property rents; **W** is an N × N spatial weight matrix; λ and ρ are respectively unknown spatial lag and spatial error parameters; **X** is an N × Q matrix of exogenous explanatory variables where continuous variables are log transformed; $\boldsymbol{\beta}$ is a Q × 1 vector of unknown coefficients; **u** is an N × 1 vector of correlated residuals and $\boldsymbol{\varepsilon}$ is an N × 1 vector of independent and identically distributed errors.

In the first equation of (2.2), the term $\lambda \mathbf{W} \log(\mathbf{P})$ reflects the impact of rents of neighboring properties and it accounts for locally constant omitted variables. The second equation of (2.2) captures residual spatial autocorrelation. When ρ=0, Equation (2.2) reduces to a spatial lag model and when λ=0, it becomes a spatial error model; setting both λ and ρ to 0 yields a simple linear regression model.

Estimating SARAR models via maximum likelihood (ML) can lead to biased and inconsistent estimators when errors are heteroskedastic (see Arraiz *et al.*, 2010). To address that problem, Arraiz *et al.* (2010) derived a generalized spatial two-stage least squares (GS2SLS) estimator that relies on instrumental variables and on the generalized-method-of-moments (GMM) to obtain consistent parameter estimates (λ,ρ, and $\boldsymbol{\beta}$ in Equation (2.2)) even with heteroskedastic errors. We therefore used GS2SLS, which is available in Stata, for this study.

2.4.3 Weight Matrix

Since spatial hedonic results may depend on the spatial weights matrix used, we considered several common weight matrices. Our starting weight matrix was obtained by calculating off diagonals terms from $w_{ij} = d_{ij}^{-2}$ if $d_{ij} \leq d$ and 0 otherwise, where d_{ij} is the straight line distance between properties *i* and *j*; and *d* is the bandwidth parameter from Moran's I correlogram (*d*=2.5 km here). Since the weight matrix captures spatial interactions with nearby properties, its diagonal terms are 0, and I normalized its rows to sum to 1 to facilitate the interpretation of results. In my sample, no two distinct observations are at the same location, so $d_{ij}>0$ for $i \neq j$.

I repeated our analysis with two other weight matrices where off-diagonal weights before row standardization are given by $w_{ij} = \exp(d_{ij}^{-2})$ if $d_{ij} \leq d$ and 0 otherwise for one, and $w_{ij} = \exp(-\left(d_{ij} / d\right)^2)$ if $d_{ij} \leq d$ and 0 otherwise for the other: Since all three weight matrices gave very similar results, I report only those for the first weight matrix.

2.4.4 Interpreting Results

While interpreting results for OLS is straightforward, it is more involved for SARAR models because of the spatial lag term $\lambda \mathbf{W} \log(\mathbf{P})$, which creates feedback effects between neighboring properties. Indeed, assuming that $|\lambda| < 1$ and denoting by **I** the N×N identity matrix, we have:

$$\mathbf{V} \equiv (\mathbf{I} - \lambda \mathbf{W})^{-1} = \mathbf{I} + \lambda \mathbf{W} + \lambda^2 \mathbf{W}^2 + \ldots, \tag{2.3}$$

so the first equation of (2.2) becomes (with $\boldsymbol{\omega} \equiv (\mathbf{I} - \lambda \mathbf{W})^{-1} (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}$):

$$\log(\mathbf{P}) = \mathbf{X}\boldsymbol{\beta} + \lambda \mathbf{W}\mathbf{X}\boldsymbol{\beta} + \lambda^2 \mathbf{W}^2 \mathbf{X}\boldsymbol{\beta} + \lambda^3 \mathbf{W}^3 \mathbf{X}\boldsymbol{\beta} + \ldots + \boldsymbol{\omega}, \tag{2.4}$$

which implies that the expected value of the log of the rent of a property depends on a mean value (term $\mathbf{X}\boldsymbol{\beta}$) plus a linear combination of mean values taken by neighboring properties scaled by powers of the spatial lag parameter λ. To better understand these impacts, like Fischer and Wang (2011) I rewrite the first equation of (2.2) to isolate the elements of **X** after moving the dependent variable to the left side and left-multiplying throughout by $\mathbf{V} \equiv (\mathbf{I} - \lambda \mathbf{W})^{-1}$:

$$\begin{pmatrix} \log(P_1) \\ \log(P_2) \\ \vdots \\ \log(P_N) \end{pmatrix} = \mathbf{V} \begin{pmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{pmatrix} + \sum_{q=1}^{Q-1} \beta_q \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \cdots & \mathbf{V}_{1N} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \cdots & \mathbf{V}_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{V}_{N1} & \mathbf{V}_{N2} & \cdots & \mathbf{V}_{NN} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1q} \\ \mathbf{X}_{2q} \\ \vdots \\ \mathbf{X}_{Nq} \end{pmatrix} + \boldsymbol{\omega}, \tag{2.5}$$

where for $(i,j) \in \{1,\ldots,N\}^2$ and for $q \in \{1,\ldots,Q\text{-}1\}$, $\log(P_j)$ is the logarithm of the rent of the j^{th} property; $\mathbf{V} \equiv (\mathbf{I} - \lambda \mathbf{W})^{-1}$ and $\mathbf{V}_{\text{ij}}$ is the i^{th} line and j^{th} column element of **V**; $\mathbf{X}_{\text{jq}}$ is the j^{th} line and q^{th} column element of **X;** and $\boldsymbol{\omega} \equiv (\mathbf{I} - \lambda \mathbf{W})^{-1} (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}$.

If $\mathbf{X}_{jq}$ is a log-transformed continuous variable (i.e., if $\mathbf{X}_{jq}=\log(x_{jq})$), then from Eq. (2.5).

$$\frac{\partial \log(P_i)}{\partial \mathbf{X}_{jq}} \equiv \frac{\partial \log(P_i)}{\partial \log(x_{jq})} = \beta_q \mathbf{V}_{ij}, \tag{2.6}$$

which represents the percentage change in the rent of dwelling i for a 1% change in x_{jq}. It differs from 0 when $\lambda \neq 0$ if observations i and j are neighbors and if $\beta_q \neq 0$, so in that case changing explanatory variable q for observation j affects the rent of observation i. Since a large number of partial derivatives could be non-zero, I follow LeSage and Pace (2009, pp. 36-37), and report for each explanatory variable $q \in \{1,\ldots,Q\text{-}1\}$ the following scalar summary measures:

- Average Direct Impact (ADI_q), obtained by averaging the main diagonal terms of $\beta_q \mathbf{V}$:

$$ADI_q = \beta_q N^{-1} \sum_{i=1}^{N} \mathbf{V}_{ii} \tag{2.7}$$

  It represents the average impact on each observation of changing its own q^{th} explanatory variable, including the feedback passing through neighbors and back to each observation.

- Average Indirect Impact (AII_q), calculated by averaging only off-diagonal terms of $\beta_q \mathbf{V}$:

$$AII_q = \beta_q N^{-1} \sum_{i \neq j} \mathbf{V}_{ij}. \tag{2.8}$$

  It represents spatial spillovers (i.e., impacts on other observations only).

- Average Total Impact (ATI_q), obtained by averaging all row sums of the $\beta_q \mathbf{V}$ matrix; it is the sum of direct and indirect impacts. It is easy to check that since **W** is row-normalized, so are its powers so summing ADI_q and AII_q and simplifying gives:

$$ATI_q = \frac{\beta_q}{1-\lambda}. \tag{2.9}$$

If instead $\mathbf{X}_{jq}$ is a binary or a count variable, changing its value by one unit affects the logarithm of the price of property i as follows:

$$\Delta \log(P_i) = \beta_q \mathbf{V}_{ij}, \tag{2.10}$$

but the expressions of ADI_q, AII_q, and ATI_q are still given by Equations (2.7)-(2.9).

To assess the statistical significance of ADI_q, AII_q, and ATI_q , I followed LeSage and Pace (2009). First, I assumed that $\boldsymbol{\beta}$, λ, ρ and σ^2 are normally distributed with means and covariance matrix obtained from estimating Equation (2.2). Then I performed 10,000 draws, calculated ADI_q, AII_q, and ATI_q for each draw, and then estimated their statistical significance based on their empirical distributions.

Building on the assumed normal distribution of $\boldsymbol{\beta}$, λ, ρ and σ^2, the average expected % change (ADI) in the rent of property i from increasing binary/count variable $\mathbf{X}_{iq}$ by one unit is (based on the expected value of a lognormal distribution; see Casella and Berger, 1990)

$$\left(\frac{\Delta P}{P}\right)_q = N^{-1}\sum_{i=1}^{N}\left[\exp(\beta_q \mathbf{V}_{ii} + 0.5\mathbf{V}_{ii}^2\sigma_q^2) - 1\right], \tag{2.11}$$

where σ_q^2 is the variance of the distribution of $\log(\beta_q)$.

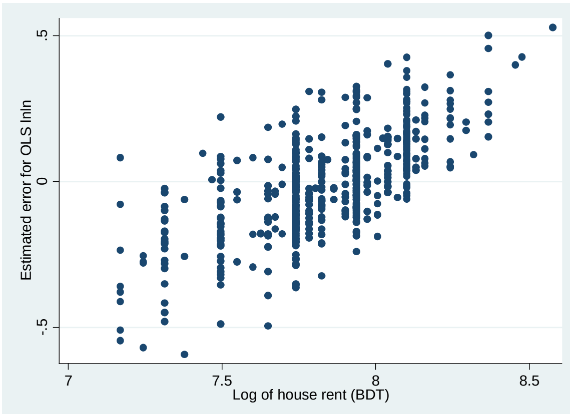
2.5. RESULTS AND DISCUSSION

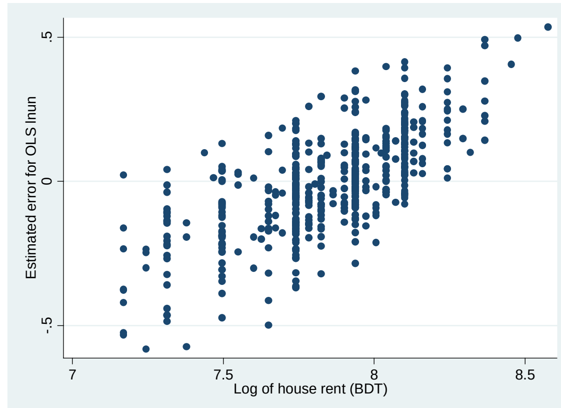
All statistical works were performed with Stata 13; in particular, my SARAR model parameters were estimated using "spreg" with generalized spatial two-stage least squares (GS2SLS) to allow for heteroskedastic errors (Drukker *et al.*, 2013). Results are presented in Table 2.7 and Table 2.8.

2.5.1 Heteroskedasticity and Spatial Dependence

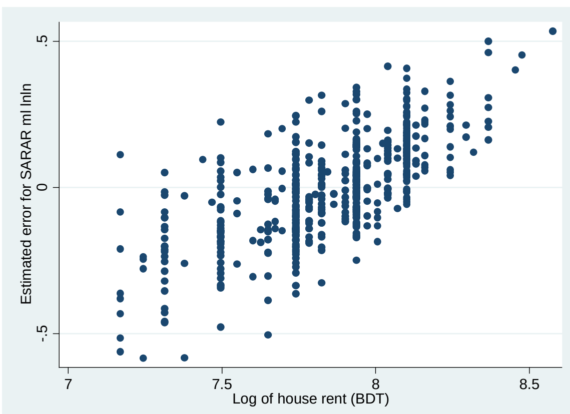
Table 2.7 contrasts maximum likelihood (ML) and generalized spatial two-stage least square (GS2SLS) coefficients of three models (linear-linear, log-linear, log-log) estimated on my N=526 sample. Its main purpose is to illustrate that ML and GS2SLS can yield widely different estimates in the presence of heteroskedasticity, which was detected by graphing estimated residuals versus rent for each model considered (Figure 2.4 and Figure 2.5). This is especially the case here for the spatial lag (λ) and spatial error (ρ) coefficients. With ML, λ is weakly significant for the log-log model and neither λ nor ρ are statistically significant otherwise.

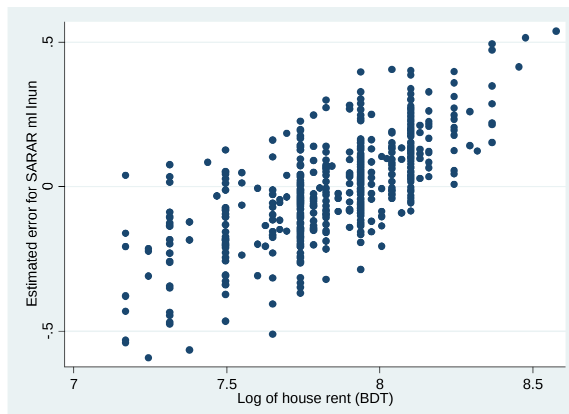
LN-LN **LN-UN**



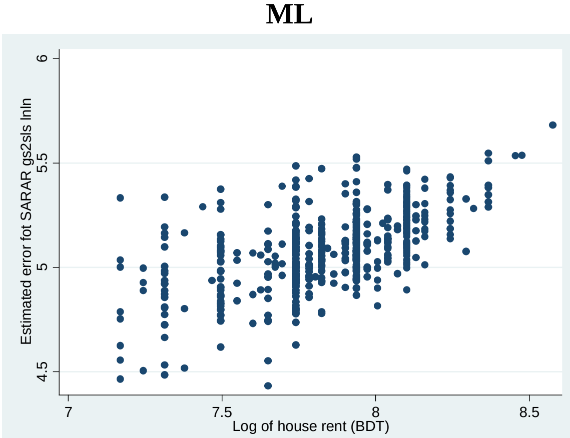


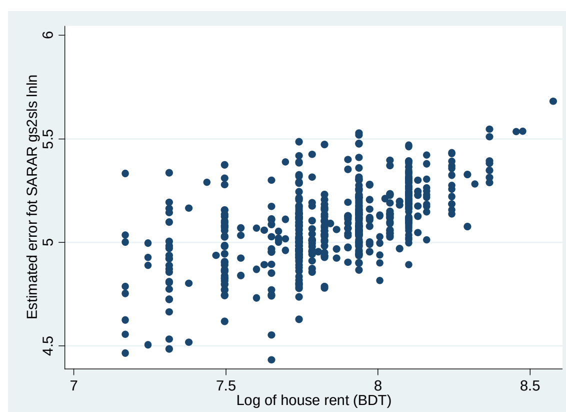
OLS **OLS**





ML **ML**

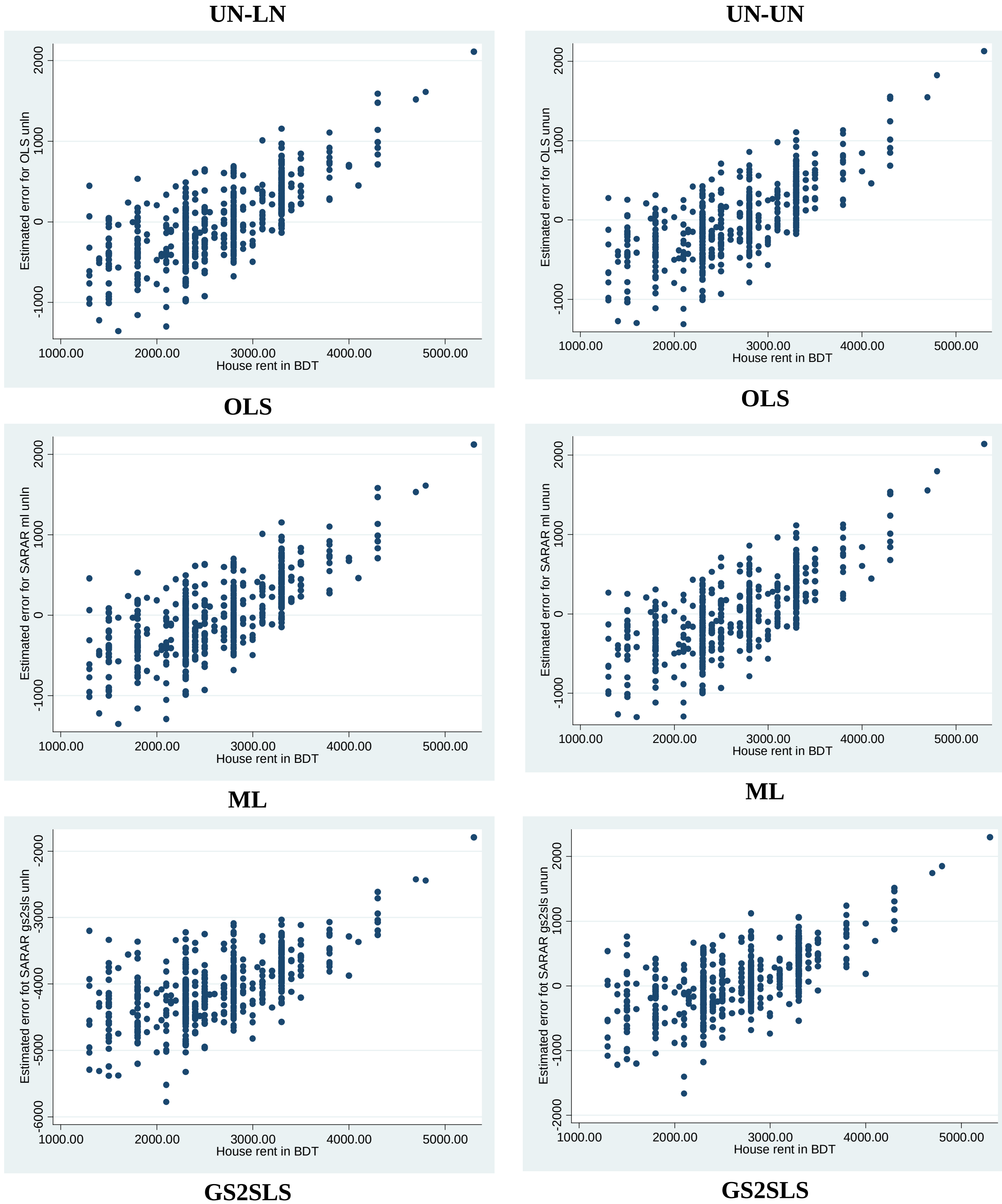




GS2SLS **GS2SLS**

Note: LN-LN: log-log, LN-UN: log-linear. OLS: Ordinary least square, ML: maximum likelihood and GS2SLS: generalized spatial two-stage least square.

Figure 2.4 Errors vs. Rent Graph to Check Heteroskedasticity for Different Model Specifications and Estimation Methods



Note: UN-LN: linear-log, UM-UN: linear-linear. OLS: Ordinary least square, ML: maximum likelihood and GS2SLS: generalized spatial two-stage least square.

Figure 2.5 Errors vs. Rent Graph to Check Heteroskedasticity for Different Model Specifications and Estimation Methods

With GS2SLS, however, both λ and ρ are strongly statistically significant, large, and, as required since we row-normalized our spatial-weight matrix (Kelejian and Prucha, 2010), they are between -1 and 1.

2.5.2 Results Interpretation

In Table 2.8, my results show that structural attributes are strong predictors of rents in Rajshahi City. The statistically significant structural parameters have expected signs. Based on direct impacts, monthly rents increase with usable living area (ADI=0.2565***) and with the number of bedrooms (ADI=0.0804***), but they decrease with building age (ADI=-0.0301*). Since a direct impact can be interpreted as an average elasticity here, a 1% increase in livable area increases monthly rent by 0.2565% while a 1% increase in age decreases monthly rent by 0.0301%. For a monthly rent of 2,650 BDT (US $39.62 based on $1 for 66.89 BDT) (the average in our sample of N=526), this represents 6.80 BDT ($0.102) and 0.80 BDT ($0.012) respectively. Moreover from Equation (2.11), adding one bedroom to a multi-unit dwelling increases its monthly rent by 8.02% or 212.53 BDT ($3.18). However, neither the coefficient of the number of bathrooms nor the coefficient of the log of distance to the nearest drainage network are significant. For the former, it may reflect that bathrooms in Rajshahi City are not as luxurious as they are in the U.S. For the latter, results may indicate that this variable is not a very good predictor of drainage efficacy.

Six of our 23 ward fixed effects binary variables are statistically significant (their values are omitted for brevity), which suggests the presence of locally important ward-level differences. The variables of most interest in this paper relate to transportation accessibility, however. Three of these variables are statistically significant.

First, the coefficient of network access distance to the nearest major road is statistically significant and negative as expected. My results show that a 1% increase in network access distance to the nearest major road for a multi-family dwelling results on average in a 0.0239% decrease in rent. Hence, a multi-unit dwelling with a monthly rent of 2,650 BDT that is located 100 m from a major road will be rented 6.33 BDT (~US $0.10) more than an identical property located 110 m from a major road. For reference, for Bangkok (Thailand) Chalermpong (2007) found a price discount of 0.0132% (687 baht/5.2 million baht; see page 117) for each additional meter from a major arterial, which is smaller in percentage for dwellings close to arterials, but larger in absolute value (after assuming that the value of a property is the present value of all rents). This is not surprising because per capita income in Bangladesh is lower than in Thailand and Rajshahi's per capita income is lower than the national average.

Second, the coefficient of the variable indicating that a multi-unit property is within 400 m of a primary school is also statistically significant and positive, which suggests that rents are higher close to primary schools. From Equation (2.11), the rent of a multi-unit dwelling would increase by 3.53% for being within 400 m of a primary school, which represents 93.55 BDT or $1.40 for a monthly rent of 2,650 BDT.

Third, the coefficient of the variable indicating that a multi-unit property is within 400 m of a healthcare facility is positive and significant, so Rajshahi residents value living close to these facilities. A couple of reasons may explain this result. First, healthcare facilities are typically co-located with pharmacies and small retail outlets, and second patients often have to rely on walk-in doctor appointment due to the absence of over the phone appointment services. From Equation (2.11), the rent of a multi-family unit within 400 m of a healthcare facility is 4.13% higher than if it were further away, which represents a premium of 109.45 BDT ($1.64).

Other transportation accessibility variables are not statistically significant. This is the case for network access distance to the nearest sub-center (p-value=0.735). This may seem surprising given that several studies (e.g., see Andersson *et al.*, 2010; Ibeas *et al.*, 2012; Efthymiou and Antoniou, 2013) find that property values increase with proximity to the CBD, although others disagree (see Cervero and Kang 2011; Noland *et al*., 2012; Diao, 2015). I conjecture that a gravity-based accessibility measure could perform better here (Ahlfeldt, 2011) but as mentioned above, data for implementing this approach are not available.

Since only 60% of access roads are paved in Rajshahi City (DDC, 2004) I expected that dwellings with a paved access road would command a higher rent, but this is not the case. Although the coefficient of the variable indicating paved access is positive, it is not statistically significant (p=0.276), possibly because Rajshahi residents rely mostly on walking, rickshaws, and biking for their transportation needs.

The other accessibility variables (shopping center within 400 m and network access distances to the nearest small industry employment area) are not significant possibly because these facilities are relatively uniformly accessible from the properties in my sample.

Likewise, my four disamenity variables (proximity to railroad tracks, bus terminals, highways, and wholesale markets) are not statistically significant. I conjecture that noise and air pollution have multiple sources (including indoors for the latter) and are ubiquitous in least-developing country cities so the contribution of specific sources is difficult to isolate.

Table 2.7 Impact of Estimation Method on Results (N=526)

	Linear-Linear SARAR		Log-Linear SARAR		Log-Log SARAR	
Variables	**ML**	**GS2SLS**	**ML**	**GS2SLS**	**ML**	**GS2SLS**
Structural Variables						
Usable living area	4.2683***	4.0720***	0.0018***	0.0016***	0.2626***	0.2295***
Number of bedrooms	201.4987***	169.3708***	0.0743***	0.0683***	0.0777***	0.0689***
Number of bathrooms	87.4216*	90.4426**	0.0347*	0.0319**	0.0241	0.0219
Building age in 2006	-3.5803**	-4.0510**	-0.0012*	-0.0014**	-0.0209	-0.0267*
Distance to nearest drainage network	1.5079**	0.9053**	5.22E-04***	3.39E-04**	0.0083	0.0065
Neighborhood Variables						
23 Ward binary variables			*Omitted for brevity*			
Accessibility Variables						
Network access distance to the nearest of 5 subcenters	-0.0764	0.0067	-0.12E-04	7.48E-06	-0.0019	0.0054
Network access distance to nearest major roads	-0.7126***	-0.4901***	-1.91E-04**	-1.50E-04**	-0.0250**	-0.0215**
Binary: 1 if access road is paved	66.4047	45.7663	0.0293	0.0246	0.025	0.0188
Network access distance to nearest railway station	0.0297	-0.0187	-6.62E-06	-0.12E-04	-0.0088	-0.0170
Network access distance to nearest bus stop	-0.0367	0.0888	-3.67E-06	0.25E-04	-0.0228	-0.0064
Binary: 1 if primary school within 400 m	104.4166	92.9511**	0.0348	0.0329**	0.0355	0.0310**
Binary: 1 if healthcare facility within 400 m	155.0233**	115.2989**	0.0500**	0.0442**	0.0460**	0.0361**
Binary: 1 if shopping center within 400 m	-26.8828	-38.4265	-0.0338	-0.0205	-0.0417	-0.0317
Network access distance to nearest small industry	0.0399	0.0151	-0.22E-04	-0.14E-04	-0.0149	-0.0126
Disamenity Variables						
Binary: 1 if within 0.25 km from railway tracks	-24.4255	-24.1497	-0.0108	-0.0082	-0.0196	-0.0130
Binary: 1 if within 0.30 km of a bus terminal	-2.00E+02	-91.9153	-0.0645	-0.0407	-0.0814	-0.0544
Binary: 1 if within 0.25 km from a highway	-68.8353	-55.3578	-0.0144	-0.0160	-0.0243	-0.0239
Binary: 1 if within 0.30 km of a wholesale market	-1.50E+02	-1.90E+02	-0.0826	-0.0745	-0.0834	-0.0739
Constant	1.6e+03**	-36.7018	5.4975***	2.7571***	4.7369***	1.8800***
Spatial lag coefficient (λ)	-0.0198	0.5837***	0.2420	0.5917***	0.2568*	0.6311***
Spatial error coefficient (ρ)	0.1116	-0.5270***	-0.2349	-0.6025***	-0.2542	-0.6444***
Dependent var. corr(predicted, observed)	0.6939	0.6825	0.7181	0.7114	0.7249	0.7123

Notes: 1. *, **, and *** indicate significance at 10%, 5%, and 1%. 2. For the log-log models (last two columns above) continuous explanatory variables are log-transformed. 3. ML = maximum likelihood; GS2SLS = generalized spatial two-stage least squares.

Table 2.8 Results for Preferred Model (N=526)

Variables	Coefficient	Log-Log GS2SLS SARAR ADI	AII	ATI	*OLS Robust Std. Err.*
Structural Variables					
Log of usable living area	0.2295***	0.2565***	0.3811***	0.6376***	0.2611***
Number of bedrooms	0.0689***	0.0804***	0.1158***	0.1962***	0.0832***
Number of bathrooms	0.0219	0.0252	0.0377	0.0630	0.0234
Building age in 2006	-0.0267*	-0.0301*	-0.0461*	-0.0762*	-0.0205
Log of distance to nearest drainage network	0.0065	0.0072	0.0109	0.0181	0.0073
Neighborhood Variables					
23 Ward binary variables			*Omitted for brevity*		
Accessibility Variables					
Log of network access distance to the nearest of 5 subcenters	0.0054	0.0060	0.0093	0.0153	-0.0089
Log of network access distance to nearest major roads	-0.0215**	-0.0239**	-0.0357**	-0.0596**	-0.0260**
Binary: 1 if access road is paved	0.0188	0.0213	0.0311	0.0524	0.0310
Log of network access distance to nearest railway station	-0.0170	-0.0190	-0.0292	-0.0482	-0.0030
Log of network access distance to nearest bus stop	-0.0064	-0.0073	-0.0096	-0.0169	-0.0346
Binary: 1 if primary school within 400 m	0.0310**	0.0355*	0.0525*	0.0881*	0.0371*
Binary: 1 if healthcare facility within 400 m	0.0361**	0.0412**	0.0598**	0.1010**	0.0582**
Binary: 1 if shopping center within 400 m	-0.0317	-0.0341	-0.0526	-0.0867	-0.0360
Log of network access distance to nearest small industry	-0.0126	-0.0141	-0.0212	-0.0353	-0.0158
Disamenity Variables					
Binary: 1 if within 0.25 km from railway tracks	-0.0130	-0.0139	-0.0224	-0.0363	-0.0210
Binary: 1 if within 0.30 km of a bus terminal	-0.0544	-0.0575	-0.0871	-0.1446	-0.1039
Binary: 1 if within 0.25 km from a highway	-0.0239	-0.0263	-0.0412	-0.0675	-0.0247
Binary: 1 if within 0.30 km of a wholesale market	-0.0739	-0.0767	-0.1201	-0.1968	-0.0689
Constant	1.8800***				6.8371***
Spatial lag coefficient (λ)	0.6311***				
Spatial error coefficient (ρ)	-0.6444***				

Notes: 1. *, **, and *** respectively indicate significance at 10%, 5%, and 1%. 2. ADI, AII, ATI = Average Direct, Indirect, and Total Impact.

2.5.3 Comparison with OLS

Finally, it is interesting to compare SARAR results with OLS estimates (last column of Table 2.8). In terms of statistical significance, OLS results mostly agree with SARAR results, with one exception: in the OLS model, building age is not significant (they are significant in the SARAR model). In terms of magnitude, the main difference between the two models is that significant OLS coefficients tend to be close to SARAR ADI values so they do not capture indirect impacts, which are at least at large as direct impacts. Relying on OLS to assess the consequences of accessibility improvements in an area would therefore substantially underestimate the value of these improvements, which are best captured by total effects.

2.6 CONCLUSIONS

To my knowledge, my paper is the first to estimate the impacts of transportation accessibility on the housing market of a least developed country city - here on the rents of multi-unit residential property in Rajshahi City, Bangladesh. To deal with spatial autocorrelation while accounting for heteroskedasticity, I estimated a spatial autoregressive model with spatial-autoregressive disturbances (SARAR) using generalized spatial two-stage least squares (GS2SLS). A comparison between coefficients estimated via OLS, maximum likelihood (ML) and GS2SLS illustrated the perils of using OLS and ML in the presence of spatial autocorrelation and heteroscedasticity respectively. This suggests that it would be useful to revisit results obtained to-date on the value of transportation accessibility because most published papers did not have the benefit of the results of Arraiz *et al.* (2010).

I found that multi-unit dwelling rents decrease by 0.0239% for every 1% increase in network access distance to the nearest major road. Conversely, proximity (within 400 m) of a primary school and a healthcare facility commands rent premiums of 3.53% and 4.13%, which corresponds to 93.55 BDT ($1.40) and 109.45 BDT ($1.64) respectively. Surprisingly, whether access roads are paved or not is not statistically significant, probably because of the dominance of walking, rickshaws use, and biking and the rarity of personal cars (as of January 2015, Bangladesh had fewer than 270,000 registered private cars for over 166 million people). Likewise, proximity to bus stops and to train stations is not statistically significant, likely because they only provide regional and national service.

Although the capitalization of accessibility is statistically significant, it is small. This is not surprising, given the relatively low household monthly income (the median value is roughly 9000 BDT or $137) of most Rajshahi households, and it echoes results obtained in some Western (e.g., see Adair *et al.*, 2000, Hess & Almeida, 2007) and Asian cities (Andersson *et al.*, 2010). It is important to note, however, that the cost of building roads in Bangladesh (Collier *et. al*., 2013) is one order of magnitude cheaper than in the United States, thanks to much cheaper labor.

Nevertheless, knowing the level of capitalization of accessibility in the rental market should be useful to Rajshahi planners and policy makers who are currently considering area-dependent property taxes (current property tax rates there are uniform) to improve local transportation services. Direct value capture instruments that reflect a fraction of actual gains to landowners could take the form of a capital gains tax on real estate, an annual property tax tuned to changes in property values, or an “unearned increment tax” when a title is transferred (Alterman, 2012). Alternatively, if political support is lacking for direct value capture, indirect measures such as general property taxes or impact fees may also be considered if they can be properly justified for mitigating impacts, providing needed public service or enhancing social justice. Implementing any of these measures, however, would require putting in place an equitable, professional, and transparent appraisal mechanism (Alterman, 2012) in addition to removing the distortion between actual and declared real estate transactions (Mitra *et al.*, 2005).

One potential limitation of this study is the relatively small size of my dataset, which reflects the difficulty of collecting reliable housing market data in most least-developed country cities. Data limitations also precluded me from using gravity-based accessibility measures for capturing the impact on rents of Rajshahi's sub-centers, which is unfortunate because underestimating labor market accessibility effects on rents could help quantify willingness-to-pay for transportation infrastructure improvements. This is left for future work. It would also be of interest to examine the potential for value capture in larger cities (particularly in Dhaka, which ranks in the top 10 largest cities in the world), both for residential and commercial properties, and to quantify the capitalization of bus services (which are lacking in Rajshahi City but are available in Dhaka), provided of course that reliable property rent or sale data can be collected.

2.7 REFERENCES

Adair, A., McGreal, S., Smyth, A., Cooper, J., & Ryley, T. (2000). House prices and accessibility: The testing of relationships within the Belfast urban area. *Housing studies*, *15*(5), 699-716.

Ahlfeldt, G., (2011). If Alonso was right: modeling accessibility and explaining the residential land gradient. *Journal of Regional Science, 51*(2), 318-338.

Ahlfeldt, G. M., & Wendland, N. (2011). Fifty years of urban accessibility: The impact of the urban railway network on the land gradient in Berlin 1890–1936. *Regional Science and Urban Economics*, *41*(2), 77-88.

Ahlfeldt, G. M. (2013). If we build it, will they pay? Predicting property price effects of transport innovations. *Environment and Planning A*, *45*(8), 1977-1994.

Alam, J. B., Sikder, S. H., & Goulias, K. G. (2004). Role of transportation in regional economic efficiency in Bangladesh: Data Envelopment Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *1864*(1), 112-120.

Allen, M. T., Rutherford, R. C., & Thomson, T. A. (2009). Residential asking rents and time on the market. *The Journal of Real Estate Finance and Economics*, *38*(4), 351-365.

Alonso, W. (1964). Location and land use. Toward a general theory of land rent. *Location and land use. Toward a general theory of land rent.*

Alterman, R. (2012). Land Use Regulations and Property Values: The 'Windfalls Capture' Idea Revisited. *Chapter in: 'The Oxford Handbook of Urban Economics and Planning' (Nancy Brooks, Kieran Donaghy, and Gerrit-Jan Knaap, eds.) pp*. 755-786.

Andersson, D. E., Shyr, O. F., & Fu, J. (2010). Does high-speed rail accessibility influence residential property prices? Hedonic estimates from southern Taiwan. *Journal of Transport Geography*, *18*(1), 166-174.

Andersson, D. E., Shyr, O. F., & Lee, A. (2012). The successes and failures of a key transportation link: accessibility effects of Taiwan's high-speed rail. *The Annals of Regional Science*, 48(1), 203-223.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models* (Vol. 4). Springer.

Anselin, L. (2005). Exploring spatial data with GeoDaTM: a workbook. *Urbana*.

Armstrong, R. J., & Rodriguez, D. A. (2006). An evaluation of the accessibility benefits of commuter rail in eastern Massachusetts using spatial hedonic price functions. *Transportation*, *33*(1), 21-43.

Arraiz, I., Drukker, D. M., Kelejian, H. H., & Prucha, I. R. (2010). A spatial Cliff-Ord-type model with heteroskedastic innovations: small and large sample. *Journal of Regional Science*, *50*(2), 592-614.

Babalola, S. J., Umar, A. I., & Sulaiman, L. A. (2013). An economic analysis of determinants of house rents in the university environment. *European Scientific Journal*, *9*(19), 99-111.

Bajari, P., & Benkard, C. L. (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of Political Economy*, 113(6), 1239-1275.

Banfi, S., Filippini, M., & Horehájová, A. (2008). Valuation of environmental goods in profit and non-profit housing sectors: evidence from the rental market in the City of Zurich. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, *144*(4), 631-653.

Bangladesh Bureau of Statistics (BBS). (2011). Report of the household income & expenditure survey 2010, Statistics Division, Ministry of Planning. Government of the People's Republic of Bangladesh.

Bangladesh Bureau of Statistics (BBS). (2012). Bangladesh Population and Housing Census 2011, National Report, Volume-4, Statistics and Information Division, Ministry of Planning. Government of the People's Republic of Bangladesh.

Bangladesh Bureau of Statistics (BBS). (2013). District Statistics 2011, Rajshahi, Statistics and Information Division, Ministry of Planning. Government of the People's Republic of Bangladesh.

Bangladesh Road Transport Authority (BRTA) (2015). Year Wise report for vehicles registration. Available from http://www.brta.gov.bd/index.php/statistics.

Banister, D., & Thurstain-Goodwin, M. (2011). Quantification of the non-transport benefits resulting from rail investment. *Journal of Transport Geography*, *19*(2), 212-223.

Baranzini, A., Schaerer, C., Ramirez, J. V., & Thalmann, P. (2008). Do foreigners pay higher rents for the same quality of housing in Geneva and Zurich?. *Swiss Journal of Economics and Statistics*, *144*(4), 703-730.

Baranzini, A., Schaerer, C., & Thalmann, P. (2010). Using measured instead of perceived noise in hedonic models. *Transportation Research Part D: Transport and Environment*, *15*(8), 473-482.

Baranzini, A., & Schaerer, C. (2011). A sight for sore eyes: Assessing the value of view and land use in the housing market. *Journal of Housing Economics*, *20*(3), 191-199.

Casella, G., & Berger, R. L. (1990). *Statistical inference*, The Wadsworth & Brooks/Cole Statistics/Probability Series, Wadsworth & Brooks.

Celik, H. M. and Yankaya, U. (2006). The impact of rail transit investment on the residential property values in developing countries: the case of Izmir subway, Turkey. *Property Management*, *24(4)*, 369-382.

Central Intelligence Agency (CIA) (2014). The World Factbook. South Asia – Bangladesh. Available from: https://www.cia.gov/library/publications/the-world-factbook/geos/bg.html#People

Cervero, R., & Murakami, J. (2009). Rail and property development in Hong Kong: Experiences and extensions. *Urban Studies*, *46*(10), 2019-2043.

Cervero, R., & Kang, C. D. (2011). Bus rapid transit impacts on land uses and land values in Seoul, Korea. *Transport Policy*, *18*(1), 102-116.

Chalermpong, S. (2007). Rail transit and residential land use in developing countries: hedonic study of residential property prices in Bangkok, Thailand. *Transportation Research Record: Journal of the Transportation Research Board*, *2038*(1), 111-119.

Chatman, D. G., Tulach, N. K., & Kim, K. (2012). Evaluating the economic impacts of light rail by measuring home appreciation a first look at New Jersey's River Line. *Urban Studies*, *49*(3), 467-487.

Cliff, A. D., & Ord, J. K. (1981). *Spatial Processes: Models & Applications* (Vol. 44). London: Pion.

Collier, P., Kirchberger, M. & Soderbom, M. (2013). The cost of road infrastructure in developing countries. www.soderbom.net/13_05_02_Unit%20cost%20paper.pdf.

Concas, S. (2013). Accessibility and Housing Price Resilience. *Transportation Research Record: Journal of the Transportation Research Board*, *2357*(1), 66-76.

Creightney, C. D. (1993). Transport and economic performance: a survey of developing countries. *World Bank Technical Paper (ISSN 0253-7494, 232).*

Crespo, R., & Grêt-Regamey, A. (2012). Spatially explicit inverse modeling for urban planning. *Applied Geography*, *34*, 47-56.

Crespo, R., & Grêt-Regamey, A. (2013). Local hedonic house-price modelling for urban planners: advantages of using local regression techniques. *Environment and Planning B: Planning and Design*, *40*(4), 664-682.

Debrezion, G., Pels, E., & Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: a meta-analysis. *The Journal of Real Estate Finance and Economics*, *35*(2), 161-180.

Debrezion, G., Pels, E., & Rietveld, P. (2011a). The Impact of rail transport on real estate prices an empirical analysis of the Dutch housing market. *Urban Studies*, *48*(5), 997-1015.

Debrezion, G., de Graaff, T., & Rietveld, P. (2011b). The Impact of accessibility on the value of offices. In *Transitions Towards Sustainable Mobility* (pp. 41-57). Springer Berlin

Heidelberg.

Development Design Consultants Limited (DDC). (2004), “ Draft final report: structure plan, master plan and detailed area development plan for Rajshahi Metropolitan City”, Prepared for Rajshahi Development Authority, Ministry of Housing and Public Works, Government of the People’s Republic of Bangladesh, Dhaka.

Diao, M. (2015). Selectivity, spatial autocorrelation and the valuation of transit accessibility. *Urban Studies*, *52*(1), 159-177.

Donovan, G. H., & Butry, D. T. (2011). The effect of urban trees on the rental price of single-family homes in Portland, Oregon. *Urban Forestry & Urban Greening*, *10*(3), 163-168.

Dowall, D. & Monkkonen, P., 2007. Consequences of the Plano Piloto: the urban development and land markets of Brasilia. *Urban Studies,* 44 (10), 1871–1887.

Drukker, D. M., Prucha, I. R., & Raciborski, R. (2013). Maximum-likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial autoregressive disturbances. *Stata Journal 13,* 221–241.

Du, H., & Mulley, C. (2006). Relationship between transport accessibility and land value: local model approach with geographically weighted regression. *Transportation Research Record: Journal of the Transportation Research Board*, *1977*(1), 197-205.

Du, H., & Mulley, C. (2007). The short-term land value impacts of urban rail transit: Quantitative evidence from Sunderland, UK. *Land Use Policy*, *24*(1), 223-233.

Dubé, J., Thériault, M., & Des Rosiers, F. (2013). Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992–2009.*Transportation Research Part A: Policy and Practice*, *54*, 49-66.

Dubé, J., Legros, D., Thériault, M., & Des Rosiers, F. (2014). A spatial Difference-in-

Differences estimator to evaluate the effect of change in public mass transit systems on house prices. *Transportation Research Part B: Methodological*, 64, 24-40

Duncan, M. (2011). The impact of transit-oriented development on housing prices in San Diego, CA. *Urban Studies*, *48*(1), 101-127.

Efthymiou, D., & Antoniou, C. (2013). How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research Part A: Policy and Practice*, *52*, 1-22.

Efthymiou, D., & Antoniou, C. (2015). Investigating the impact of recession on transportation cost capitalization: a spatial analysis. *Journal of Transport Geography*, *42*, 1-9.

Fahrländer, S. S., Gerfin, M., & Lehner, M. (2015). The influence of noise on net revenue and values of investment properties: evidence from Switzerland. Working paper, University of Bern. Available from http://www.vwl.unibe.ch/papers/dp/dp1502.pdf.

Fischer, M. M., & Wang, J. (2011). Spatial data analysis: models, methods and techniques. Springer Science & Business Media.

Füss, R., & Koller, J. (2015). The Role of Spatial and Temporal Structure for Residential Rent Predictions. *University of St. Gallen, School of Finance Research Paper*, (2015/23).

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

Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*, *57*(1), 148-169.

Giuliano, G., Gordon, P., Pan, Q., & Park, J. (2010). Accessibility and residential land values: some tests with new measures. *Urban Studies*, *47*(14), 3103-3130.

Grimes, A., & Young, C. (2013). Spatial effects of urban rail upgrades. *Journal of Transport Geography*, *30*, 1-6.

Habib, M. A., & Miller, E. J. (2008). Influence of transportation access and market dynamics on property values: multilevel spatiotemporal models of housing price. *Transportation Research Record: Journal of the Transportation Research Board*, *2076*(1), 183-191.

Halden, D. (2002). Using accessibility measures to integrate land use and transport policy in Edinburgh and the Lothians. *Transport Policy*, 9(4), 313-324.

Haque, A. (2014). *Transport situation in Rajshahi.* Prepared for United Nations Economic and Social Commission for Asia and the Pacific (UN ESCAP).

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: an exploration of issues and alternatives. *Environment and Planning A*, 29(7), 1175-1194.

Hess, D. B., & Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. *Urban Studies*, *44*(5-6), 1041-1068.

Ibeas, Á., Cordera, R., Coppola, P., & Dominguez, A. (2012). Modelling transport and real-estate values interactions in urban systems. *Journal of Transport Geography*, *24*, 370-382.

Islam, I., Mitra, S.K., Nayeem, A., Rahman, A. (2007) “Land Price in Dhaka City: Distribution, Characteristics and Trend of Changes.” In Jahan, S. Maniruzzaman K.M. (ed.), *Urbanization in Bangladesh Patterns, Issues and Approaches to Planning,* Bangladesh Institute of Planners, Dhaka, 2007.

Jun, M. J. (2012). Redistributive effects of bus rapid transit (BRT) on development patterns and property values in Seoul, Korea. *Transport Policy*, *19*(1), 85-92.

Kay, A. I., Noland, R. B., & DiPetrillo, S. (2014). Residential property valuations near transit stations with transit-oriented development. *Journal of Transport Geography*, *39*, 131-140.

Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*(1), 53-67.

Kim, K., & Lahr, M. L. (2013). The impact of Hudson-Bergen Light Rail on residential property appreciation. *Papers in Regional Science*. DOI: 10.1111/pirs.12038.

LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL: Chapman &Hall/CRC.

Lee, J., & Sohn, K. (2013). Identifying the Impact on Land Prices of Replacing At-grade or Elevated Railways with Underground Subways in the Seoul Metropolitan Area. *Urban Studies*, 0042098013484543.

Li, W., & Joh, K. (2016). Do Residents Value the Integration of Bicycling and Transit? Assessing Their Revealed Preferences Based on Property Sale Transactions. In \ *Transportation Research Board 95th Annual Meeting* (No. 16-5429).

Löchl, M., & Axhausen, K. W. (2010). Modelling hedonic residential rents for land use and transport simulation while considering spatial effects. *Journal of Transport and Land Use*, *3*(2), 39-63.

MacLennan, D. (1977). Some thoughts on the nature and purpose of house price studies. *Urban Studies*, *14*(1), 59-71.

Martinez, L. M., & Viegas, J. M. (2009). Effects of transportation accessibility on residential property values. *Transportation Research Record: Journal of the Transportation Research Board*, *2115*(1), 127-137.

Mathur, S., & Ferrell, C. (2013). Measuring the impact of sub-urban transit-oriented developments on single-family home values. *Transportation Research Part A: Policy and Practice*, *47*, 42-55.

McIntosh, J. R., Newman, P., Trubka, R., & Kenworthy, J. (2015). Framework for land value capture from investments in transit in car-dependent cities. *Journal of Transport and Land Use*.

Medda, F. (2012). Land value capture finance for transport accessibility: a review. *Journal of Transport Geography*, *25*, 154-161.

Mikelbank, B. A. (2004). Spatial analysis of the relationship between housing values and investments in transportation infrastructure. *The Annals of Regional Science*, *38*(4), 705-726.

Mills, E. S. (1972). *Studies in the Structure of the Urban Economy.* Johns Hopkins University Press, Baltimore, Md.

Mitra, S.K., Nayeem, A., Rahman, A. (2005). *A study on land price of Dhaka City* (BURP Thesis, Department of Urban and Regional Planning, Bangladesh University of Engineering and Technology, Dhaka).

Mohammad, S. I., Graham, D. J., Melo, P. C., & Anderson, R. J. (2013). A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice*, *50*, 158-170

Munoz-Raskin, R. (2010). Walking accessibility to bus rapid transit: Does it affect property values? The case of Bogotá, Colombia. *Transport Policy*, *17*(2), 72-84.

Muth, R. (1969). Cities and housing: The spatial patterns of urban residential land use. *University of Chicago, Chicago*.

Noland, R. B., DiPetrillo, S., & Lahr, M. L. (2012). Residential Property Values and New Jersey Transit Village Program. *Transportation Research Record: Journal of the Transportation Research Board*, *2276*(1), 78-88.

Osland, L., & Thorsen, I. (2008). Effects on housing prices of urban attraction and labor-market accessibility. *Environment and Planning A*, 40(10), 2490-2509.

Osland, L., & Pryce, G. (2012). Housing prices and multiple employment nodes: is the relationship non-monotonic? *Housing Studies*, 27(8), 1182-1208.

Osland, L., & Thorsen, I. (2013). Spatial impacts, local labour market characteristics and housing prices. *Urban Studies*, 1-21.

Pagliara, F., & Papa, E. (2011). Urban rail systems investments: an analysis of the impacts on property values and residents' location. *Journal of Transport Geography*, *19*(2), 200-211.

Pan, H., & Zhang, M. (2008). Rail transit impacts on land use: Evidence from Shanghai, China. *Transportation Research Record: Journal of the Transportation Research Board*, *2048*(1), 16-25.

Rodriguez, D. A., & Mojica, C. H. (2009). Capitalization of BRT network expansions effects into prices of non-expansion areas. *Transportation Research Part A: Policy and Practice*, *43*(5), 560-571.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy*, *82*(1), 34-55.

Schläpfer, F., Waltert, F., Segura, L., & Kienast, F. (2015). Valuation of landscape amenities: A hedonic pricing analysis of housing rents in urban, suburban and periurban Switzerland. *Landscape and Urban Planning*, *141*, 24-40.

Shyr, O., Andersson, D. E., Wang, J., Huang, T., & Liu, O. (2013). Where Do Home Buyers Pay

Most for Relative Transit Accessibility? Hong Kong, Taipei and Kaohsiung Compared. *Urban Studies*, *50*(12), 2553-2568.

Smith, J. J., & Gihring, T. A. (2006). Financing transit systems through value capture. *American Journal of Economics and Sociology*, *65*(3), 751-786.

Sohail, M., Maunder, D. A. C., & Cavill, S. (2006). Effective regulation for sustainable public transport in developing countries. *Transport Policy*, *13*(3), 177-190.

UN-OHRLLS (2013). State of the least developed countries, 2013. UN Under-Secretary-General and High Representative for LDCs, LLDCs and SIDS (UN-OHRLLS).

Vadali, S. (2008). Toll roads and economic development: exploring effects on property values. *The Annals of Regional Science*, *42*(3), 591-620.

Valente, J., Wu, S., Gelfand, A., & Sirmans, C. (2005). Apartment rent prediction using spatial modeling. *Journal of Real Estate Research*, *27*(1), 105-136.

Wilson, B., & Frew, J. (2007). Apartment rents and locations in Portland, Oregon: 1992-2002. *Journal of Real Estate Research*, *29*(2), 201-218.

Xiao, Y., Orford, S., & Webster, C. J. (2016a). Urban configuration, accessibility, and property prices: a case study of Cardiff, Wales. *Environment and Planning B: Planning and Design*, *43*(1), 108-129.

Xiao, Y., Webster, C., & Orford, S. (2016b). Identifying house price effects of changes in urban street configuration: An empirical study in Nanjing, China. *Urban Studies*, *53*(1), 112-131.

Wu, N., Zhao, S., & Zuo, Z. (2013). A Study of the impacts of inner-city and inter-city accessibility on house prices. *Journal of the Eastern Asia Society for Transportation Studies*, 10(0), 970-988.

Yiu, C. Y., & Wong, S. K. (2005). The effects of expected transport improvements on housing prices. *Urban Studies*, *42*(1), 113-125.

Zhang, M., & Wang, L. (2013). The impacts of mass transit on land development in China: The case of Beijing. *Research in Transportation Economics*, *40*(1), 124-133.

Chapter 3. Exploring the Unnoticed: An Analysis of Carless Households in California

3.1 INTRODUCTION

Widespread automobile ownership gives us the freedom to travel and it shaped our society, but this comes at the cost of accidents, congestion, and air/noise pollution. Moreover, it entails millions of acres of asphalt for parking lots and roads, as well as urban sprawl and its related social problems (Banister, 2005). With current engine technology, motor vehicles are also a major contributor to global climate change (Curtis, 2009), which has at last received universal attention after Pope Francis' urgent appeal (The Washington Post, June 18, 2015). In California, Governor Jerry Brown issued an executive order on April 29, 2015 that requires a 40% cut in greenhouse gas emissions below 1990 levels by 2030, thus speeding up the goals set by Assembly Bill 32 and Senate Bill 375 (The New York Times, 2015). These laws and executive orders have turned reducing vehicle-miles traveled and auto dependency into an important policy goal but the path to transitioning away from an auto-dependent society is still quite unclear. One possible starting point is to learn from the households who currently live without a motor vehicle. Note that in this dissertation I call "carless" households who do not own a motor vehicle (including cars, pickups, or SUVs) for their personal transportation needs.

From 2007 to 2012, the number of carless households in the United States increased from 8.7 % to 9.2% (Sivak, 2014). Approximately 10.5 million U.S. households currently do not own a motor vehicle (2008-12 American Community Survey). These households, who are often forgotten in transportation policy discussions, can be organized into two groups: involuntary carless households who are forced to live without motor vehicles, and voluntary carless households who simply chose to live without cars. Understanding the characteristics of households who voluntarily decided to forgo cars is important because it could inform policies that attempt to reduce our dependency on cars and reduce greenhouse gas emissions. Understanding the characteristics of involuntary carless households is no less important, however, because these households face physical isolation, poor access, and social exclusion. Unfortunately, our knowledge of carless households appears to be severely lacking. My goal in this paper is therefore to start filling this gap. I analyze data from the 2012 California Household Travel Survey (CHTS) to characterize voluntary and involuntary carless households in California and explore basic features of their travel patterns to understand how they meet their daily life needs. I also assess the effects of various socio-economic, life-cycle stage, and land use variables on the likelihood that a household is carless, voluntarily or not. California is a fascinating place to study carless households because it offers a wide range of contexts, from San Francisco where many policies tend to discourage the use of private motor vehicles to Southern California suburbs where cars are indispensable.

The rest of this chapter is organized as follows. Section 3.2 provides a brief review of the relevant studies to inform my modeling choices. Section 3.3 presents my data and Sections 3.4 outlines my methodology. Results are discussed in Section 3.5 and Section 3.6 summarizes my main findings, discusses some limitation of my approach, and proposes directions for future research.

3.2 LITERATURE REVIEW

In this section, I first discuss studies interested in carless households. I then review the variables used in car ownership models estimated on cross sectional datasets and finish with selected studies dealing with the life-oriented approach related to car ownership.

3.2.1 Carless Households

Despite of the increasing number of U.S. households without a car (Sivak, 2014), only a handful of published papers have analyzed this group of households, and many of these papers highlight the disadvantages of being carless.

In an early study, Paaswell and Recker (1976) surveyed over 400 Buffalo, New York, residents to understand the problems faced by carless people. They found some heterogeneity among carless people, although they were predominantly unemployed, female, and low-income, with a mean age lower than the population average, which is unsurprising because many younger people do not have a vehicle. They also reported that people with access to cars tend to had better jobs than those without, even with equal skills.

On the West Coast of United States, using 1976 Urban and Rural Travel Survey for Los Angeles County, Marquez (1980) found that the elderly and women are the most disadvantaged groups among the carless residents of Los Angeles County, California, including unemployed or retired people with no driver's licenses.

Bromley and Thomas (1993) studied the shopping behavior of carless households to better understand the relationship between being carless and social disadvantage. They found that carless people are constrained to make greater use of local, smaller, and more expensive stores as they are unable to take advantage of newer, larger, but also more distant shops that offer a wider range of products at better prices.

Changes in car ownership practices such as car sharing may improve access to mobility for carless households, however. In a recent paper, Lovejoy (2012) analyzed data from the 2009 National Household Travel Survey and focus groups of recent California immigrants to understand car use, with a particular interest for carless and households with more drivers than cars. She developed measures of mobility fulfillment for given demographic profiles and elucidated circumstances where innovative car sharing might be adopted more readily.

In summary, I was not able to find published research that analyzes differences between voluntary and involuntary carless households. To the best of my knowledge, this study is the first to investigate this question, in addition to characterizing how voluntary carless households differ from involuntary carless households, but also from conventional households who rely on motor vehicles to fulfill their transportation needs.

3.2.2 Car Ownership Studies with Cross Sectional Data

This study also relates to the literature that analyzes factors that foster vehicle ownership using cross sectional data, although the focus in these papers is typically on households with one or more vehicles (see Table 3.1). In general, these studies show that income is one of the primary determinants of car ownership. Moreover, they suggest that life-cycle stage, number of household workers, number of household members with a driver's license (Chu, 2002; Whelan, 2007, Potoglou and Kanaroglou, 2008), the availability of other means of transport (which, in turn, depends on the built environment; Kim and Kim, 2004), and urbanization (Oakil *et al.*, 2016) influence car ownership. Another strand of the literature shows that travel attitudes and lifestyles play a significant role in vehicle ownership (Van Acker *et al*., 2014), and so can residential self-selection, backyard size, and off-street parking facilities (Cao and Cao, 2014).

The influence of the built environment on car ownership has sparked a lot of interest around the world. Influential variables include measures of density (Giuliano and Dargay, 2006; Goetzke and Weinberger, 2012), diversity (Potoglou and Kanaroglou, 2008; Zegras, 2010), design (Bhat and Guo, 2007; Zegras, 2010), accessibility (Matas *et al.*, 2009; Van Acker and

Witlox, 2010), transit availability (Pinjari *et al.*, 2011; Goetzke and Weinberger, 2012; Huang *et al.*, 2016), and commute distance/time (Bhat and Guo, 2007; Potoglou and Kanaroglou, 2008). Possibly because of its diversity, results from this literature appear to be inconsistent about the importance of built environment variables. In their recent analysis of the relationship between historical exposure to the built environment and current vehicle ownership patterns, Macfarlane *et al.* (2015) concluded that models with both current and past neighborhood attributes can provide a better understanding of the built environment's causal influences on vehicle ownership decisions. However, like other studies in this literature, they do not consider carless households.

Review of the literature suggests that understanding factors that lead households to voluntarily forgo motor vehicles has received very limited attention to-date, but it suggested variables to consider in my models.

3.2.3 Car Ownership in the Context of a Life-Oriented Approach

Most published studies on car ownership are based on cross-sectional travel surveys. However, several studies argued that this single temporal point cannot adequately capture the dynamics of households' car-ownership decisions over their lifetime (Nolan, 2010; Zhang *et al.*, 2014; Oakil, 2014; Khan and Habib, 2016). To understand the vehicle ownership decision of households over their lifetime, travel behavior researchers have been using longitudinal approaches for last two decades, and most of them focused on vehicle transaction studies (e.g, Gilbert, 1992; Yamamoto *et el.*, 1999; Mohammadian and Miller, 2003; de Jong and Kitamura, 2009), but none of them focused on carless households. There are another body of literatures where pseudo-panel approaches have been used in the absence of panel data (see Dargay, 2002; Anowar *et al.*, 2015).

Table 3.1 Summary of Selected Car Ownership Studies with Cross Section Data (2007-2016)

Authors (Year Published)	Data [Period Analyzed] Models	Socio-demographic Variables Life-Cycle Stage Variables Built Environment and Other Variables	Key Results
Oakil *et al.* (2016).	• 861 households from vehicle registration data, Netherlands [2012/2013]. • Binary logit model.	• HH reference person: age, ethnicity; HH income; HH composition: young couples, young singles, young two-parent families, young single-parent families; HH employment. • NA. • Urbanization level: very high to very low densities.	Urbanization level and household composition are essential factors influencing car ownership. The influence of urbanization level on car ownership is much stronger for young couples than for young families or singles.
Huang *et al.* (2016).	• 1,442 households in 21 Guangzhou communities, China [2011/2012]. • Random effect ordered probit model.	• HH income; HH size; having a driver's license; occupation: white collar worker, private business owner, retired. • NA. • Local transit access; density of road network.	Transit access is negatively associated with auto ownership, after controlling for demographics and other built environment variables.
Houston *et al.* (2015)	• 7,889 HH sample from California Household Travel Survey and Neighborhood Travel and Activity Study, California, US [2012] • Binary logit model.	• # of HH: workers, non-workers; HH income; race; highest educational attainment. • NA. • Near residence land use; intersections; population density; employment/destination access; transit service: low, moderate, highest; distance to light rail station; dwelling type.	Higher population density and employment/destination access were associated with a lower probability of vehicle ownership for both the countywide and study area sample.
Macfarlane *et al.* (2015).	• 227,830 households moving history data of the 13-county metro Atlanta region, USA. • Multinomial logit model.	• HH age, ethnicity, income. • Number of children in the HH. • Current neighborhood: density, non-vehicle share; past neighborhood: density.	Models with current but not past neighborhood attributes (controlling for socioeconomic variables) can forecast vehicle ownership ~ well.

Authors (Year Published)	**Data [Period Analyzed]** **Models**	**Socio-demographic Variables** **Life-Cycle Stage Variables** **Built Environment and Other Variables**	**Key Results**
Cao and Cao (2014).	• 1,303 sample information from a self-administrative survey, Minneapolis-St. Paul, USA [04/2011]. • Ordered logit model	• Number of drivers; gender; income; driver's license. • NA • Number of businesses within 1/4 mi; attitudes: pro-driving, pro-transit; perception of 'large back yards'; preference for: large back yards, plenty of off-street parking, easy access to transit stop/station.	Residential self-selection influences car ownership; backyard size, off-street parking and business density marginally impact it, but light rail transit does not affect it directly.
Van Acker *et al.* (2014).	• Information about 1,878 persons gathered via an internet survey, Ghent, Belgium [05-10/2007]. • Structural equation model.	• Gender. • Student living at home; older, younger family members. • Local, regional center; accessibility: local, regional; density; life styles characteristics; residential attributes: car alternatives, open space and quietness, safety and neatness, accessibility, social contact; travel attitudes; travel mode attitudes.	Controlling for residential and travel attitudes and lifestyles, residential land use may still have expected impact on car ownership. Car availability lower in dense, accessible neighborhoods close to regional centers.
Goetzke and Weinberger (2012).	• 3,322 household, New York City, USA [2004]. • OLS and binary probit model.	• HH income, ethnicity, number of workers. • Presence of children in the HH. • Tract density and accessibility; city fixed-effects; tract education levels, poor HHs, HH size; building types.	Households have a higher probability of possessing a vehicle if they are surrounded by other motorized households.
Pinjari *et al.* (2011).	• 5,147 household observations from San Francisco Bay Area Travel Survey, San Francisco, USA [2000]. • Integrated simultaneous multi-dimensional choice model.	• HH # of: employed individuals, disabled individuals; HH income; age; gender; ethnicity. • HH # of: active adults, seniors, children; single parent HH; single individual HH. • HH density, total commute time/cost, number of commuters, block densities, number of zones accessible via transit (<30 min) or biking (<6 mi); Employment areas: biking facility density, street block density; single family housing; dwelling ownership.	Residing in a zone with higher housing or employment density is associated with lower levels of auto ownership, especially in the case of lower income households residing in high employment locations.

Authors (Year Published)	**Data [Period Analyzed]** **Models**	**Socio-demographic Variables** **Life-Cycle Stage Variables** **Built Environment and Other Variables**	**Key Results**
Van Acker and Witlox (2010).	• 5,500 persons from the Ghent Travel Behavior Survey, Belgium [2000]. • Structural Equations.	• HH income; age; marital status; driving license. • NA. • Built up index; land use diversity; distance to: railway station, CBD; accessibility by car (15 min).	Ignoring car ownership as a mediating variable likely misspecifies impacts of some built-environment variables on car use.
Li *et al*. (2010).	• Data from 36 megacities and 1,200 (Beijing) and 1,001 (Chengdu) observations from household survey data, China [2006]. • OLS and binary logit.	• HH head: age, marital status, education, having bus pass, took bus; home ownership; owning a bike, an electric bike, a motor bike; HH size, HH income. • HH with children. • Population density; distance to: CBD, nearest bus stop; living within fourth ring road (R4); family lives closed to husband/wife's workplace.	Urban affluence, urban scale, and road infrastructure supply have significant positive effects on city level ownership of private cars. Population density (sub-district level) has a significant negative effect on private car ownership.
Zegras (2010).	• 14,729 observations from household origin–destination survey, Santiago de Chile [2001]. • Multinomial logit.	• HH income; # of workers. • Number of children in HH. • Dwelling unit density; diversity index; 4-way intersection per km; distance to CBD; accessibility ratio; metro station within 500 m; live in an apartment.	Income dominates household vehicle ownership decision; relationship between several built environment variables and likelihood of car ownership.
Matas *et al*. (2009).	• 52,375 (Barcelona) and 63,903 (Madrid) HHs from Spanish micro-census, Spain [2001]. • Ordered probit.	• Head of HH: age, years of education, gender, marital status, employment status, occupation, citizenship. • Number of adults and working adults in the HH. • Unemployment rate; dummy for central city; Job accessibility of working adults; housing tenure, size.	Time cost to access jobs by public transport is a determinant of car ownership. The elasticity for average car ownership is -0.25 for Barcelona and -0.19 for Madrid.
Potoglou and Kanaroglou (2008).	• 774 household data from an internet-survey, Hamilton, Canada [04/2005]. • Multinomial logit.	• HH number of: full-time workers, part-time workers; HH income; number of driver's licenses/HH size. • HH type; type of dwelling. • Mixed density index; land-use entropy index; number of bus stops within 500 m form dwelling; number of individuals working > 6 km away.	Household life-cycle stage, socio economic factors, mixed density and land-use diversity within walking distance from residence influence household car ownership.

Authors (Year Published)	Data [Period Analyzed] Models	Socio-demographic Variables Life-Cycle Stage Variables Built Environment and Other Variables	Key Results
Potoglou and Susilo (2008).	• NHTS for the Baltimore Metro Area, Maryland, USA [2001]; Dutch National Travel Survey [2005]; Osaka Metro Person Trip Data, Japan [2000]. • Ordered model, multinomial logit/probit.	• # of HH workers; HH income; race. • HH life cycle: single, couple, single parent, couple with children, retired. • Residential density; reside in: highly urbanized area, high urbanized area, moderately urbanized area, low urbanized area, non-urbanized area; type of dwelling: single-family house.	Baltimore couples are less likely to own one car, Dutch and Osaka couples are more likely to own 1 and 3 cars than 0 cars. Single-parent households in Osaka are less likely to own a car. The multinomial logit model is best for modeling the level of household car ownership.
Bhat and Guo (2007).	• 2,954 household data from San Francisco Bay Area travel survey, CA, USA [2000]. • Ordered response car ownership model.	• HH income; ethnicity; ownership of dwelling. • HH structure. • HH and employment density; drive commute time and cost; street block density; transit availability and access time; multifamily housing unit.	Marginally significant to insignificant negative impacts of HH and employment density on car ownership, low income HHs in high employment density areas less likely to own cars.
Whelan (2007).	• 46,137 households from the family expenditure survey [1971-1996] and the national travel survey [1991], Great Britain. • Binary logit.	• HH income; motoring cost; company car; # of employments; license holding. • HH structure (combinations of number of adults and children, with retirement status). • Binary variables for different areas.	The models are successfully validated at the household level and the model forecasts compare favorably with actual ownership information extracted from the 2001 Census.

Note: HH stands for household

Table 3.2 Summary of Selected Life Oriented/Course/Cycle Studies of Car Ownership (2006-2016)

Authors (Year Published)	Data [Period Analyzed] Models	Life Course/Cycle/Events Variables Socio-demographic Variables Built Environment and Other Variables	Key Results
Khan and Habib (2016)	• Vehicle ownership and life history of 446 households, Halifax, Canada [2012-2013]. • Parametric hazard-based duration model of household ownership states.	• Birth of a child; death of a member; member move-in and move-out; increase & decrease in employment. • Age of the primary worker; household income; dwelling type; household size; primary mode of commuting. • Neighborhood level: land-use index, dwelling density, participation rate, employment rate, percentage of rental house and own house.	Birth of a child, move-in of members, and increase in employment in the household exhibit high probability of shorter no-car ownership state, leading to the first-time vehicle ownership.
Oakil (2016)	• Retrospective questionnaire survey of 1200 individuals, Utrecht, Netherlands [1990-2010]. • Mixed logit model.	• Birth of the first child; cohabitation; divorce; employer change; residential relocation. • Age; university graduate; income; living with partner; working full-time. • NA.	Life events such as birth of the first child, divorce, residential relocation and employer change significantly impact the decision to get full access to a car only for the female respondents. No significant effect from life event is observed for the decision to sacrifice full access of a car for both men and women.

Authors (Year Published)	Data [Period Analyzed] Models	Life Course/Cycle/Events Variables Socio-demographic Variables Built Environment and Other Variables	Key Results
Clark *et al.* (2015)	• 19,334 households from the UK Household Longitudinal Study (UKHLS) [2009-2001]. • Binary logistic regression model.	• Changed employer; entered/lost employment; residential relocation; gained driving license; birth of a child; gained/lost partner; retired; more/fewer household adults; change in income, proximity to bus stops/rail station, travel time to nearest employment center, food stores accessible by walking, population density. • Household composition: size, child, older person; income; education level; job. • Location: inner/outer London, metropolitan areas, large /medium/small/very small urban, rural; # of bus stops; rail station; accessible by walking/public transit: # of employment centers, # of food stores; travel time to nearest town center; living environment index; population density.	Changes to composition of households (people arriving and leaving) and to driving license availability are the strongest predictors of car ownership level changes, followed by employment status and income changes.
Zhang *et al.* (2014)	• Retrospective web based life history survey of 1000 households in major Japanese cities [11/2009]. • Chi Square Automatic Interaction Detection (CHAID) analysis.	• Residential relocation; household structure change; employment/education change; car-ownership change. • Age; HH income. • NA.	Results confirm two-way cause–effect relationships over a lifetime between residential and car ownership histories, which are influenced by household structure and employment/education events.
Oakil *et al.* (2014)	• Retrospective questionnaire survey of 1200 individuals, Utrecht, Netherlands [1990-2010]. • Mixed logit model.	• Birth of the first child; cohabitation; divorce; employer change; residential relocation; retired. • Age; education; living with partner; dual working household; HH income. • NA.	Strong and simultaneous relationships between car ownership changes and household change processes. Childbirth and residential relocation linked to car ownership changes.

Authors (Year Published)	Data [Period Analyzed] Models	Life Course/Cycle/Events Variables Socio-demographic Variables Built Environment and Other Variables	Key Results
Yamamoto (2008)	• Panel survey of 3,638 households, France [1984-1998] and retrospective survey of 1,849 households in recent three years, Kofu, Japan [2005-2006]. • Hazard-based duration model and multinomial logit model	• Increased and decreased in adults; increased children; decreased income; increased and decreased drivers; increased workers; moving. • Number of adults; number of children; household income; number of vehicles; elderly household; number of drivers; farmers. • Binary: largest city, second largest cities; large cities; distance to station; distance to station for elderly household; bus frequency: for family with kids, family with infants, for increased drivers.	The explanatory power of the life course events on vehicle ownership dynamics, in terms of contribution to the improvement in the goodness of-fit statistics of the model to the data, is found to be small.
Prillwitz *et al.* (2006)	• The German socioeconomic panel (GSOEP) survey of 4,698 households [1998-2003]. • Bivariate analysis and binomial probit model.	• Relocation within research period; moved and change in characteristics of residential area; difference in: number adults in HH, number of children (<18); first child in HH within research period; change in weighted monthly income; Head of HH changed to: fully employed, unemployed, pensioner; increase in education level of HH head. • Number of cars per HH in 1998; age and squared age of HH head in 1998; weighted monthly income in 2003. • NA.	A strong influence of four household key events on car ownership growth: (*a*) a changing number of adults per household; (*b*) the birth of a first child; (*c*) a change in the weighted monthly income per household; and (*d*) the moving of residence from a regional core to a (possibly the same) regional core area.

Note: HH stands for household

More recently life oriented approach introduced by Zhang *et al.,* (2014), have been getting attention to capture the temporal dynamics of households' car ownership decision. For Zhang (2015), the life-oriented approach strives to capture dependences between different life domains. These "biographical interdependences," as he calls them, have four components: residential, household structure, employment, and car ownership. Here a biography tracks key events of interest in a person's life (Zhang *et al.*, 2014, Zhang, 2015).

Using a web-based life history survey data, Zhang *et al.* (2014) explored the interdependences between residential and car ownership biographies in various Japanese cities. They found two-way cause–effect relationships between residential and car ownership decisions that are further influenced by household structure and employment/education events. In their study, household structure and employment/education biographies were found to be more influential on residential decisions than car ownership choices, although the latter play an important role in explaining car ownership mobility decisions.

Recently, similar approaches such as life-cycle/life-event (Oakil *et al.*, 2014; Clark et al., 2015; Oakil, 2016) and life-course (Prillwitz *et al.*, 2006; Yamamoto, 2008; Khan and Habib, 2016) have also been getting attention to capture the temporal dynamics of households' car ownership decision. The life cycle approach has been defined as the birth-to-death sequence of stages in the life of an individual or a family (Zimmerman, 1982). The life-course approach is based on the concept that people's travel behavior has some common features over a lifetime but is shaped by specific events such as marriage, residential relocation, the birth of a child, or changing jobs (Lanzendorf, 2003; Yamamoto, 2008). The main contrast between these studies and the life-oriented approach proposed by Zhang *et al.* (2014) is that the former assume that life-cycle and/or life-event variables affect residential and travel behavior, whereas Zhang *et al.*

(2014) argue that the life-oriented approach considers two-way relationships between residential decisions/travel behavior and other life domains.

Life-cycle/life-course/life-oriented studies suggest that the following factors play an important role in car ownership and travel decisions: birth of the first child (Yamamoto, 2008; Oakil *et al.*, 2014; Clark *et al.*, 2015; Khan and Habib, 2016), divorce (Oakil, 2016), residential relocation (Yamamoto, 2008; Oakil *et al.*, 2014; Oakil, 2016), income changes (Prillwitz *et al.*, 2006; Clark *et al.*, 2015), changes in employment status (Oakil *et al.*, 2014; Oakil, 2016; Khan and Habib, 2016), and changes in the number of household adults (Prillwitz *et al.*, 2006; Yamamoto, 2008). These studies are summarized in Table 3.2.

Although a number of life-cycle/life-course/life-oriented studies examine ownership levels, few discuss carless households. One exception is Clark *et al.* (2015). Using a panel dataset from the U.K., they analyzed predictors of different types of car ownership level change (zero to one car, one to two cars and vice versa). They found that households are more likely to relinquish a vehicle following an income reduction than they are to acquire one after an income gain. They also emphasized the importance of the spatial context, as poorer access to public transport increases the probability that a carless household would acquire a car, and lowers the probability that a household with a single vehicle would relinquish it.

Another exception is Khan and Habib (2016), who applied a life-course approach to model vehicle ownership and choice of a vehicle type using retrospective survey data from Halifax, Canada. They confirmed that events such as the addition of a household member or of a job in the household accelerate the acquisition of a first household vehicle.

From published life-cycle/life-course/life-oriented studies, it is not clear, however, what the main causes are for living without a car in an auto-oriented society or how major life-cycle events, socio-economic, or built environment variables influence the decision to be voluntarily carless, which is my focus here.

3.3 DATA

3.3.1 Survey Data

This paper analyzes geocoded data from the 2012 California Household Travel Survey (CHTS), which was a unique statewide, collaborative effort to gather travel information from households in all of California's 58 counties. Data were collected using various tools, including diaries, computer assisted telephone interviews (CATI), a website, and three types of global positioning systems (GPS) devices - wearable, in-vehicle, and in-vehicle with an on-board diagnostic (OBD) unit. After a pretest in late fall, 2011, the survey was fielded in January, 2012. Participating households were asked to record their travel in a diary for a pre-assigned 24-hour period. Households who participated in the GPS assisted survey wore GPS devices for three days, and data were collected from instrumented vehicle for seven days.

A total of 42,431 households completed the survey, including 5,717 households who provided GPS information. Of the GPS households, 3,855 were assigned wearable GPS, 422 used in-vehicle GPS only, and 1,440 had an in-vehicle GPS plus an OBD unit. The 2012 CHTS provides information on household car ownership, in addition to detailed information on the socio-economic characteristics of individuals and households (such as income, education, and household composition), as well as the latitude and longitude of each household location.

3.3.2 Definition of Voluntary and Involuntary Carless Households

To understand whether carless households chose to live without cars voluntarily or not, I analyzed the CHTS question that asks about reasons for not owning a vehicle; answers are summarized in Table 3.3.

Table 3.3 Classification of Carless Households

No.	Reasons for not owning a motor vehicle	Classification
1	Do not need a car - can do what I need and want to without a motor vehicle	Voluntary
2	Concerned about impact on environment	Voluntary
3	Can't drive and (1 or 2)	Voluntary
4	No driver's license and (1 or 2)	Voluntary
5	Get rides from other people and (1 or 2)	Voluntary
6	Use public transit and (1 or 2)	Voluntary
7	Too expensive to buy	Involuntary
8	Too expensive to maintain (gas/insurance/repairs)	Involuntary
9	Health/age related reasons	Involuntary
10	Cannot get insurance	Involuntary
11	Can't drive and (7 or 8 or 9)	Involuntary
12	No driver's license and (7 or 8 or 9)	Involuntary
13	Get rides from other people and (7 or 8 or 9)	Involuntary
14	Use public transit and (7 or 8 or 9)	Involuntary
15	Other	Unknown
16	Mentioned both reasons for voluntary and involuntary	Unknown
17	No answer	Unknown

Note: Table 3.3 relies on CHTS Question: HHNOV (CHTS Code): "Please let us know the reasons why you/your household does not own a motor vehicle."

Respondents who selected either "want to be without a car" or "concerned about impact on environment" (items 1 and 2 in Table 3.3) were assumed to have voluntarily chosen to forgo motor vehicles provided they did not select any other answer suggesting that their choice was constrained.

Conversely, households who stated that they cannot afford a vehicle, cannot get insurance, or have health/age constraints were classified as involuntary carless households provided they did not also select any of the two reasons that characterize voluntary carless households. Finally, households who selected at least one item from the voluntary and one from the involuntary answers were assigned to an "unknown" group.

After an exhaustive investigation and after removing observations with missing information, I obtained a sample of 1,972 carless households (including 303 voluntary, 831 involuntary, and 838 "unknown" households) and 30,839 households with vehicles. Figure 3.1 shows the home location of these households.

3.3.3 Explanatory Variables

I categorized explanatory variables into three groups: (1) socio-economic and demographic, (2) life-cycle stage, and (3) dwelling type and built environment.

3.3.3.1 Socio-Economic and Demographic Characteristics

Following a number of published car ownership studies (See Table 3.1 and Table 3.2), my explanatory variables include the following household characteristics: size, ratio of household bicycles to household size, number of workers, income, Hispanic or Latino status, ethnicity, and education level.

Household size and number of household workers are count variables. For the ratio of household bicycles to household size I considered the number of household members over 5 years old. For household income, I used the midpoint of each CHTS category (note that my sample does not include households with an annual income over $250,000). To capture education, I created binary variables to indicate that at least one household member reached a specific education level. Likewise, Hispanic status and ethnicity are binary variables (with Caucasian as the baseline category for the latter).

3.3.3.2 Life-Cycle Stage

Different life-course/life-cycle/life-event variables have been found to be significant in explaining household car ownership decisions (see Table 3.1 and Table 3.2). Like Potoglou and Kanaroglou, (2008) and Giuliano and Dargay, (2006), I created the following variables as proxy for life-cycle stage variables: binary variables for the number of children (households with one, two, and three or more children) with childless households as the baseline, and presence of at least one household member over 65 years. My starting hypothesis is that the presence of children and elderly people may decrease the probability of being voluntarily carless because of their needs for additional non-work trips.

3.3.3.3 Dwelling Type and Land Use Variables

Published papers based on both life history (Khan and Habib, 2016) and cross-sectional travel data (e.g., see Potoglou and Kanaroglou, 2008 or Houston *et al.*, 2015) indicate that dwelling type is linked to vehicle ownership and travel decisions. In the US, single-family residences offer free parking in garages and driveways, and tend to be physically separated from other types of land uses, which creates the need for more car trips (e.g., to shop). I therefore hypothesized that households who reside in single-family (detached or semi-detached) houses are less likely to be voluntarily carless and my models include a binary variable that indicates whether or not a household lives in a single-family detached or semi-detached dwelling.

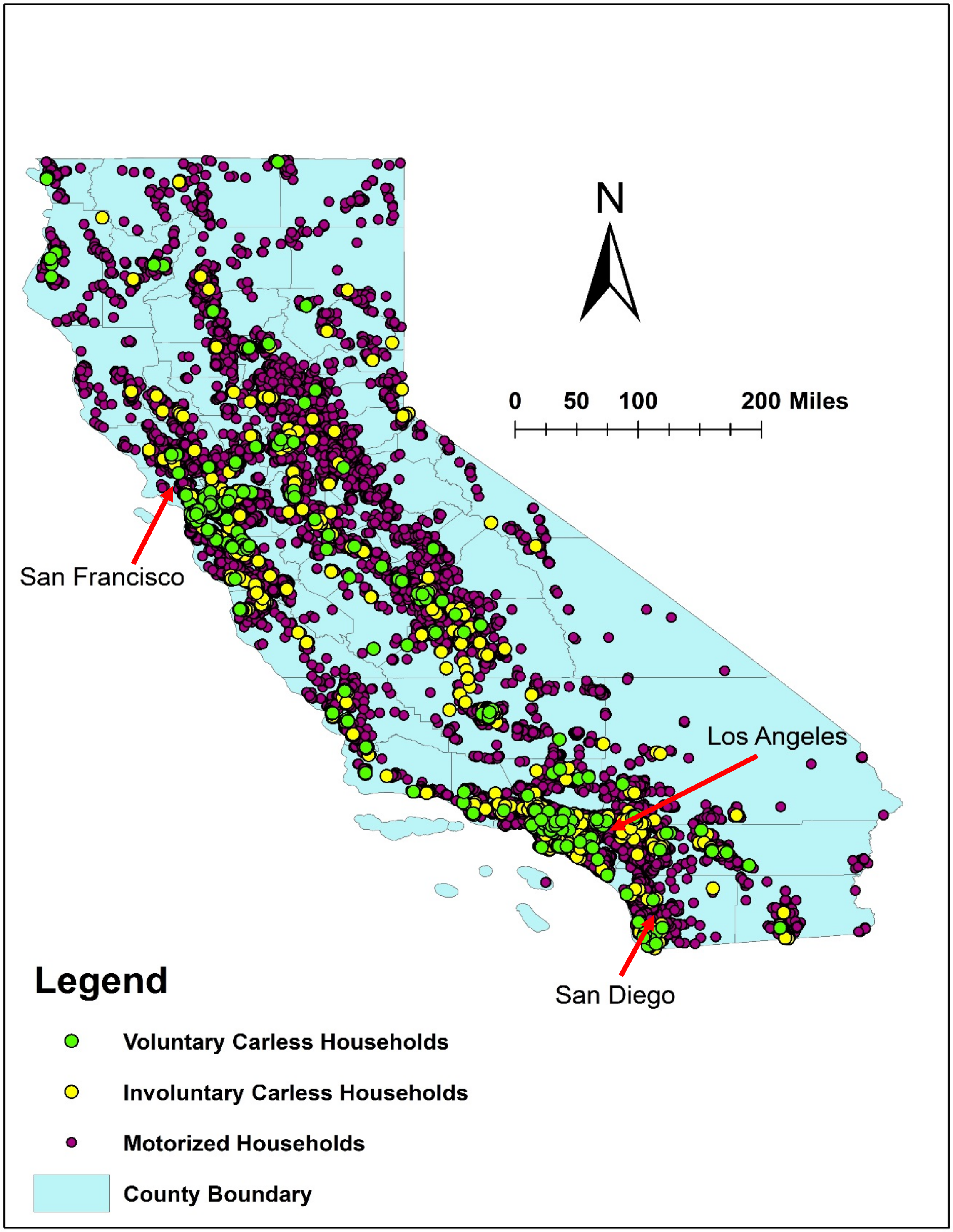


Figure 3.1 Home Location of CHTS Respondents

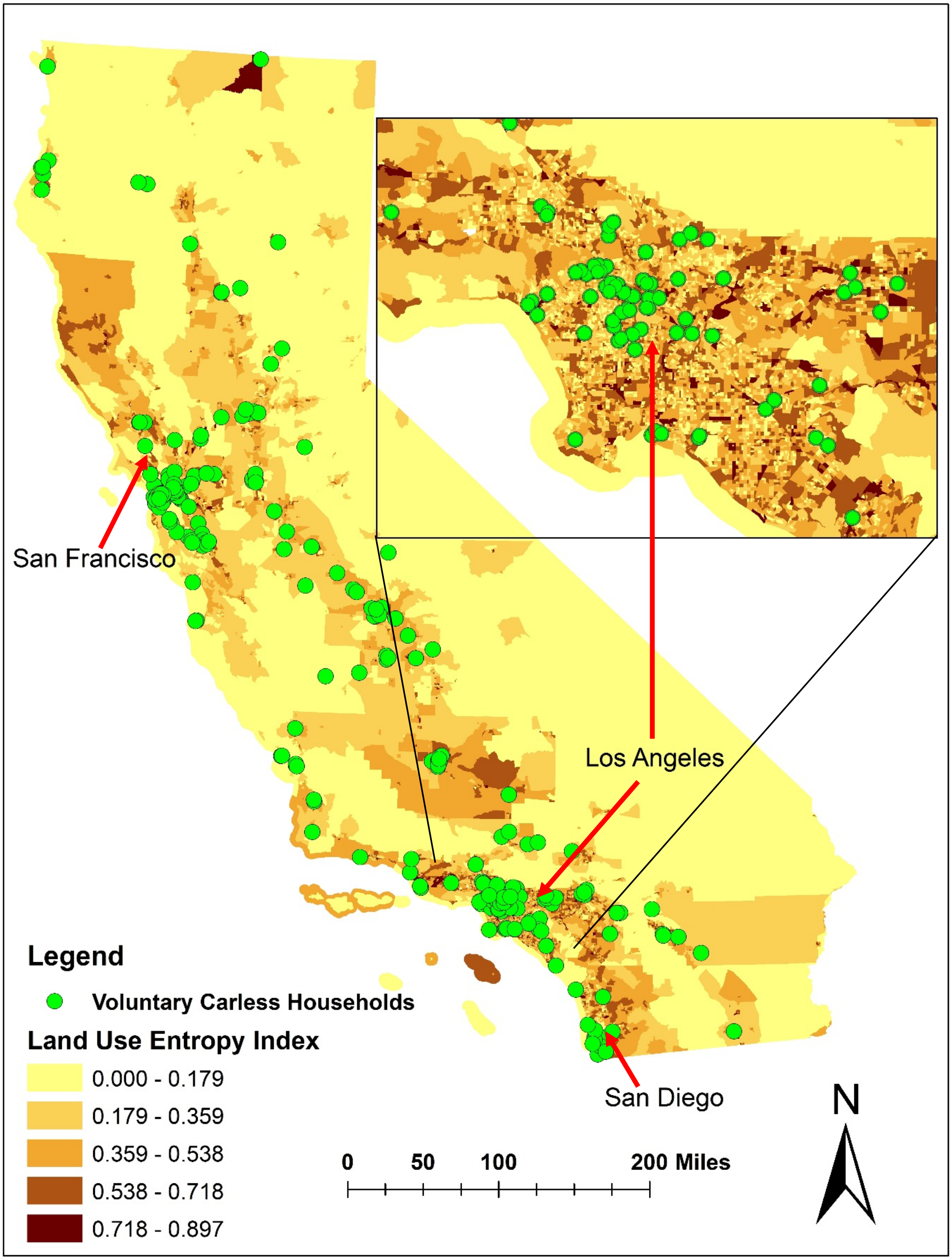


Figure 3.2 Location of Voluntary Carless Households and Land Use Entropy Index

Land use variables have been found to be important determinants of households' car ownership decisions in both life-oriented and cross-sectional studies (see Table 3.1 and Table 3.2). Some life-oriented studies used changes in built environment variables following a residential relocation (e.g. Prillwitz *et al.*, 2006; Clark *et al.*, 2015), and others used only land use characteristics of a household's current location (e.g. Yamamoto, 2008; Khan and Habib, 2016). Because of data availability, I follow the latter.

To better capture the land use variables, I include the "three Ds" groups of variables (Cervero and Kockelman, 1997): density, diversity, and design. I used population density and the mixed density index (MDI_j) as my density variables. I relied on 2010 census data to measure population density at the block group level. Following Chu (2002) and Potoglou and Kanaroglou (2008), the MDI in block group *j* is defined by:

$$MDI_j = \frac{ED_j RD_j}{ED_j + RD_j}, \tag{3.1}$$

where ED_j is employment density (number of workers per square mile) and RD_i is residential density (number of housing units per square mile). *MDI* is a proxy for employment accessibility and an indicator of job-housing balance, which has been shown to result in lower car ownership (Potoglou and Kanaroglou, 2008), so a higher value of *MDI* is expected to motivate more households to voluntarily forgo owning a motor vehicle.

Likewise, mixed land use is expected to reduce car dependence and car ownership (Cervero and Kockelman, 1997; Soltani, 2005; Potoglou and Kanaroglou, 2008; Zegras, 2010), which has been captured by different variables in both life-oriented (Yamamoto, 2008; Clark *et al.*, 2015; Khan and Habib, 2016) and cross-sectional studies (see Table 3.1). To reflect the diversity of land uses, I calculated a block-group entropy index (LEI) based on parcel-level GIS land use data from the Southern California Association of Governments (SCAG), the California Geoportal, the California Atlas, and the California Department of Water Resources. I calculated the LEI from (e.g., see Cervero and Kockelman, 1997; e Silva *et al.*, 2006):

$$LEI_j = \frac{-1}{\ln(k)} \sum_{n=1}^{k} P_{jn} \ln(P_{jn}), \tag{3.2}$$

where LEI_j is the land use entropy index of block group j; k is the number of land use types (k=8 here: residential, commercial, industrial, public facilities, open space and recreational, mixed development, agriculture, and other land uses); and P_{jn} is the areal percentage of land use of type n in block group j. *LEI* varies between 0 to 1, where 0 indicates a single land-use type and 1 implies a perfect balance between all land use types. At the outset, I expected higher LEI value to be associated with a higher probability that a household is voluntarily carless. As shown on Figure 3.2, households who are voluntarily carless live in block groups with higher LEI values.

A variety of network design measures have been used in the literature (e.g., see Bhat and Guo, 2007; Zegras, 2010; Pinjari *et al.*, 2011; Huang *et al.*, 2016). Here, to capture walkability I used network density calculated as facility miles of pedestrian-oriented links per square mile (extracted from the smart location database of the U.S. Environmental Protection Agency), as I expected a higher network density value to be associated with a higher likelihood to be carless.

In addition to the 3 'Ds', transit service availability has been used in both life-course (Yamamoto, 2008; Clark *et al.*, 2015) and cross-sectional car ownership studies (Zegras, 2010; Van Acker and Witlox, 2010; Goetzke and Weinberger, 2012; Cao and Cao, 2014; Huang *et al.*, 2016). As a proxy for transit service availability, I used the percentage of the regional population that can be accessed from block group j within 45-minute via transit and walking (also extracted from the smart location database of the U.S. EPA.)

Using ArcGIS, I created my land use variables and assigned them to each household based on residential location. Table 3.4 presents summary statistics for the variables considered in my models.

3.4 METHODOLOGY

I analyzed my data in two steps: 1) simple univariate analyses to gauge basic differences between groups of households defined by their ownership of motor vehicles; and 2) multivariate analyses to tease out what factors contribute to the decision to forgo cars voluntarily or not.

3.4.1 Univariate Analysis

First, I contrasted the characteristics (socio-economic and demographic, life-cycle stage, type of dwelling and built environment) and travel patterns of voluntary carless, involuntary carless, and motorized households. I did not include "unknown" carless households in my analyses. To characterize travel behavior, I calculated the average number of trips per day, average travel distance, modal share, and trip purpose for each group and contrasted them. This comparison allows testing the assumption that involuntary carless households experience greater levels of transport disadvantage than their voluntary counterparts which in turn may negatively affect their quality of life.

I used one-way analysis of variance (ANOVA) to test the statistical significance of differences in selected continuous and count variables for these three groups. For the one way analysis of variance, I calculated F statistics from (Wabed and Tang, 2010):

$$F^* = \frac{\dfrac{\sum_{k=1}^{g} n_k \left(\overline{x}_k - N^{-1}\sum_{j=1}^{g}\sum_{i=1}^{n_j} x_{ij}\right)^2}{k-1}}{\dfrac{\sum_{j=1}^{g}\sum_{i=1}^{n_j}\left(x_{ij} - \overline{x}_j\right)^2}{N-k}}, \quad (3.3)$$

where n_i is the number of observations in group i, g is the number of groups, $\overline{x}_i$ is the mean of group i, and $N = \sum_{k=1}^{g} n_k$ is the total number observations. Under the null hypothesis that the means of these groups do not differ, F^* has an $F(k-1, N-k)$ distribution.

I then performed post hoc tests to see which pairs of means are significantly different using the Tukey-Kramer post hoc test (Ramsey, 2010) for pairwise testing of means in a one-way analysis of variance with unequal sample sizes. For that procedure, a single critical difference, CD, is calculated for each pair of means as given by (Ramsey, 2010):

$$CD = q_{1-\alpha}(k-1, df_E)\sqrt{\frac{MSE}{2}\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}, \quad (3.4)$$

where:

- $q_{1-\alpha}(k-1, df_E)$ is the 100 (1-α) percentage point of the Studentized range distribution with parameters k and df_E;
- df_E is the error degrees of freedom from ANOVA;
- *MSE* is the error mean sum of squares from ANOVA; and
- n_i and n_j are the sample sizes of groups i and j (with $i \neq j$).

In addition, I used χ^2 tests (Washington *et al*., 2010) to test the statistical significance of differences in categorical variables for the three groups since χ^2 tests are most appropriate for analyzing relationships among nominal variables (Connor-Linton, 2010).

Table 3.4 Descriptive Statistics of Variables for Each Model

Variables	**Model 1** (N=32,811)				**Model 2** (N=31,142)				**Model 3** ((N=1,134))			
	Mean	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
Socio-Economic and Demographic Characteristics												
Household (HH) size	2.60	1.39	1	8	2.64	1.38	1	8	1.77	1.24	1	8
Ratio of HH bicycle and HH size(>5 years)	0.61	0.74	0	15	0.63	0.73	0	15	0.31	0.71	0	15
Number of HH workers	1.25	0.88	0	6	1.29	0.88	0	6	0.52	0.65	0	3
Midpoint of annual HH income ($ in thousand)	86.01	62.41	5	250	89.10	62.08	5	250	24.60	28.49	5	225
Binary: 1=Hispanic or Latino	0.24	0.43	0	1	0.24	0.43	0	1	0.37	0.48	0	1
Binary:1=Caucasian	0.65	0.48	0	1	0.66	0.47	0	1	0.55	0.50	0	1
Binary:1=African American	0.04	0.19	0	1	0.03	0.18	0	1	0.13	0.34	0	1
Binary:1= Other ethnicity	0.31	0.46	0	1	0.30	0.46	0	1	0.32	0.47	0	1
Binary: 1= No high school degree	0.04	0.19	0	1	0.03	0.17	0	1	0.19	0.39	0	1
Binary: 1= High school graduate	0.10	0.30	0	1	0.09	0.29	0	1	0.25	0.43	0	1
Binary: 1= Some college credit but no degree	0.15	0.35	0	1	0.14	0.35	0	1	0.20	0.40	0	1
Binary: 1= Associate’s degree	0.11	0.32	0	1	0.11	0.32	0	1	0.10	0.31	0	1
Binary: 1= Bachelor degree	0.29	0.45	0	1	0.30	0.46	0	1	0.14	0.35	0	1
Binary: 1= Graduate degree	0.32	0.46	0	1	0.33	0.47	0	1	0.11	0.31	0	1
Life-Cycle Stage												
Binary: 1= If HH has no child	0.74	0.44	0	1	0.74	0.44	0	1	0.85	0.36	0	1
Binary: 1= If HH has 1 child	0.11	0.31	0	1	0.11	0.32	0	1	0.07	0.25	0	1
Binary: 1= If HH has 2 children	0.10	0.30	0	1	0.10	0.31	0	1	0.06	0.23	0	1
Binary: 1= If HH has 3 or more children	0.05	0.21	0	1	0.05	0.21	0	1	0.03	0.18	0	1
Binary: 1= If HH has members >65 years	0.30	0.46	0	1	0.30	0.46	0	1	0.29	0.45	0	1
Dwelling Type												
Binary:1=Single family detach housing	0.73	0.44	0	1	0.75	0.43	0	1	0.24	0.42	0	1
Land Use Variables												
Population density per square mile(thousand)	7.82	9.54	0	204.07	7.31	8.35	0	204.07	17.98	20.61	0.00	204.07
Mixed density index (MDI)	0.88	1.85	0	90.01	0.80	1.60	0	90.01	2.66	5.08	0.00	90.01
Land use entropy index (LEI)	0.34	0.19	0	0.90	0.34	0.19	0	0.90	0.38	0.21	0.00	0.87
Network density (facility miles of pedestrian-oriented links per square mile)	13.68	7.17	0.007	64.31	13.55	7.15	0.007	64.31	16.42	7.23	0.22	47.16
% of population accessible by transit	0.03	0.06	0	0.49	0.03	0.06	0	0.49	0.09	0.09	0.00	0.42

3.4.2 Multivariate Analysis

In order to characterize carless households in a multivariate framework, I estimated three binary logit models, where the dependent variable Y_i is defined respectively by:

- Model 1: Y_i =1 if a household is carless and 0 if it owns one or more vehicles; the goal is to contrast households with and without motor vehicles; the sample size is 32,811;
- Model 2: Y_i =1 if a household is voluntarily carless and 0 if it owns at least one motor vehicle; the goal here is to understand how voluntary carless households differ from households who own motor vehicles; the sample size is 31,142 (both involuntary carless households and households who were not unambiguously carless were excluded)
- Model 3: Y_i =1 if a household is voluntarily carless and 0 if it is involuntary carless; the sample size is 1,134 (carless households who could not be unambiguously classified were excluded);

For each of these models, the probability that the dependent variable equals 1 for household *i* is given by (Greene, 2008):

$$\Pr(Y_i = 1 \mid \mathbf{X_i}) = \frac{\exp(\mathbf{X_i}\boldsymbol{\beta})}{1 + \exp(\mathbf{X_i}\boldsymbol{\beta})}, \tag{3.5}$$

where $\mathbf{X_i}$ is a matrix of explanatory variables; and $\boldsymbol{\beta}$ is a vector of unknown coefficients estimated via maximum likelihood.

A convenient way of interpreting results from a logit model is to report odds ratios (Greene, 2008). The odds of observing Y_i =1 versus Y_i =0 are

$$\Omega(\mathbf{X_i}) = \frac{\Pr(Y_i = 1 \mid \mathbf{X_i})}{\Pr(Y_i = 0 \mid \mathbf{X_i})} = \exp(\mathbf{X_i}\beta), \tag{3.6}$$

so if I denote by $\Omega(\mathbf{X_i}, x_j{+}1)$ the odds obtained by adding 1 to explanatory variable x_j $j \in \{1,\ldots,k\}$ in Equation (3.6), the odds ratio for variable x_j is given by

$$OR_j = \frac{\Omega(\mathbf{X_i}, x_j + 1)}{\Omega(\mathbf{X_i}, x_j)} = \exp(\beta_j), \quad (3.7)$$

which does not depend on the characteristics of household *i*. Hence, increasing x_j raises the likelihood that Y_i=1 if and only if β_j>0.

To assess the fit and adequacy of my models, I performed common diagnostics (e.g., see Long and Freese, 2006). First, to see if my models have any explanatory power, I performed a likelihood ratio (LR) tests with the null hypothesis that coefficient estimates except the constant are jointly zero (Cameron and Trivedi, 2010).

To check model specification, I conducted link tests (Bruin, 2011); link tests are based on the idea that for a properly specified model I should not be able to find any additional predictor that is statistically significant except by chance. I also performed Hosmer-Lemeshow (HL) goodness-of-fit tests, which compare predicted and observed frequencies, with the idea that for a well specified model they should match closely (Cameron and Trivedi, 2010).

Finally, I examined model residuals to find outliers (i.e., observations with large residuals) and looked for influential observations.

3.5 RESULTS AND DISCUSSION

My statistical work was performed with Stata 13. Results are presented in Table 3.5, Table 3.6 and Table 3.7.

Table 3.5 Household Characteristics by Vehicle Ownership Group

Variables	Category	Voluntary carless HH	Involuntary carless HH	HH with vehicles	Statistical test
Socio-Economic and Demographic Characteristics					
Household size	Average HH size	1.59	1.84	2.65	F(2,31970)=227.25***
Bike ratio	Average Ratio of HH bicycle and HH Size(>5years)	0.43	0.27	0.63	F(2,31970)=110.08***
Number of workers	Average Number of household workers	0.62	0.48	1.30	F(2,31970)=440.34***
Income	Average midpoint of annual HH income ($)	38.25	19.62	89.60	F(2,31970)=628.17***
Hispanic status	Hispanic Household (%)	30.69	39.47	23.72	$\chi^2(2)$=116.49***
Ethnicity	Caucasian (%)	61.39	52.50	66.08	$\chi^2(2)$=64.46***
	African American (%)	10.23	13.96	3.41	$\chi^2(2)$=285.44***
	Other ethnicity (%)	28.38	33.09	30.51	$\chi^2(2)$=3.22
Highest household educational attainment	No high school degree (%)	12.87	21.30	3.01	$\chi^2(2)$=862.29***
	High school graduate (%)	20.79	26.96	9.53	$\chi^2(2)$=349.29***
	Some college credit but no degree (%)	19.47	19.98	14.20	$\chi^2(2)$=28.25***
	Associate or technical school degree (%)	10.89	10.23	11.33	$\chi^2(2)$=1.09
	Bachelor degree (%)	18.48	13.00	29.72	$\chi^2(2)$=126.48***
	Graduate degree (%)	17.49	8.54	32.76	$\chi^2(2)$=241.61***
Life-Cycle Stage	No child in the household (%)	88.12	83.39	73.49	$\chi^2(2)$=73.23***
	HH with 1 child (%)	7.26	6.26	11.40	$\chi^2(2)$=26.31***
	HH with 2 children (%)	3.63	6.26	10.49	$\chi^2(2)$=30.42***
	HH with 3 children (%)	0.99	4.09	4.62	$\chi^2(2)$=9.52***
	HH with older member (>65) (%)	29.04	28.76	29.54	$\chi^2(2)$=0.289
Dwelling Type	Single family detach housing (%)	24.42	23.23	75.85	$\chi^2(2)$=1.6E+03***
Land Use Variables	Average population density per square mile (thousand)	21.30	16.76	7.17	F(2,31970)=878.40***
	Average Mixed density index	3.96	2.19	0.77	F(2,31970)=833.37***
	Average Land use entropy index	0.44	0.36	0.34	F(2,31970)=42.57***
	Average Network density (facility miles of pedestrian-oriented links per square mile)	17.68	15.97	13.50	F(2,31970)=97.93***
	Average % of population accessible by transit	0.11	0.08	0.03	F(2,31970)=636.02***
Sample Size		303	831	30,839	

Note: *, **, and *** indicate significance respectively at 10%, 5%, and 1%.

3.5.1 Univariate Results

3.5.1.1 Household Characteristics

The last column of Table 3.5 displays results of the statistical tests that compare the means of explanatory variables for the three groups of households considered: voluntarily carless, involuntarily carless, and motorized households. It shows that the latter are more likely to have a higher household income and to be more educated. However, the voluntary carless group is more likely to have a higher average household income and a higher education level than the involuntary group. Moreover, voluntary carless households tend to be smaller than other households and they have a smaller number of workers than motorized households but higher than involuntary carless households. The number of older people in the household does not differ statistically across groups but voluntary carless households have fewer children than other households. As expected, motorized households are more likely to live in single family detached houses than their carless counterparts. I also note that the percentage of voluntary carless households in single family detached houses is marginally higher than for the involuntary group.

Looking at land use variables, I see that the average population density and the mixed density index of the location of voluntary carless households are significantly higher than for the other two groups of households. Moreover, voluntary carless households are more likely to live in areas with a higher land use entropy index (which is also supported by Figure 3.2), a higher average network density (pedestrian-oriented links per square mile) and better public transport service coverage. Hence, voluntarily carless households may have higher mobility than involuntary carless households since Lovejoy (2012) found that among carless people, mobility fulfillment was generally greatest among those living in high density environments with better car-free alternatives such as walking and transit and with a rich set of proximate destinations.

Table 3.6 Travel Patterns of Voluntary and Involuntary Carless Households

Variables	Category	Voluntary carless HH	Involuntary carless HH	HH with vehicles	Statistical test
Number of trips	Average HH person trips on travel day	6.89 [a]	7.36 [a]	8.71	F(2,31970)=19.43***
	% of HH with no trip on travel day	20.79	26.11	11.62	F(2,31970)=91.52***
Change type of transportation/transfer	Percent of trips with one or more stops	34.83	40.77	4.39	F(2,342655)=11731.26***
Travel distance	Average daily person miles traveled (PMT)	25.64	27.52	92.42	F(2,31970)=12.97***
	Average daily vehicle miles traveled (VMT)	5.19	6.21	49.74	F(2,31970)=35.53***
	Average travel distance per trip	7.05 [ab]	3.99 [a]	12.38	F(2,31970)=5.72***
Travel mode (percent of trips)	By car	9.46 [a]	9.93 [a]	64.55	F(2,360431)=6611.63***
	By public transit	18.6	21.79	2.11	F(2,360431)=7182.27***
	By walking	47.16	45.00	8.33	F(2,360431)=8467.79***
	By cycling	4.12	2.11	1.29	F(2,360431)=96.14***
Travel duration	Average daily travel time	98.85 [a]	120.81 [a]	179.98	F(2,31970)=56.98***
	Average travel duration per trip	13.93 [a]	14.41 [a]	22.02	F(2,31970)=30.69***
Number of Activities	Average number of activities on travel day	11.22 [a]	10.24 [a]	12.23	F(2,28108)=8.47***
	Average activity duration per trip	257.66	273.83	319.05	F(2,360433)=90.73***
Activity purpose (percent of trips)	For work purpose	12.67 [b]	6.59	11.45	F(2,324370)=48.50***
	For school purpose	2.87 [b]	4.78	3.86	F(2,324370)=6.71***
	For shopping purpose	10.42 [ab]	12.3 [a]	8.85	F(2,324370)=31.87***
	For social/recreational purpose	14.79 [b]	11.66	16.01	F(2,324370)=30.08***
	For Personal business	5.31 [a]	5.39 [a]	3.60	F(2,324370)=24.94***
	For medical/dental purpose	2.43 [a]	2.92 [a]	1.35	F(2,324370)=43.28***
	For religious activities	2.06 [ab]	2.94 [a]	1.72	F(2,324370)=18.08***
	In home activities	41.70 [b]	45.96	41.57	F(2,324370)=16.37***
	Others	7.74 [a]	7.47 [a]	11.58	F(2,324370)=44.66***

Notes: 1.[ab] indicate values that differ using Tukey-Kramer post-hoc test, 2.[a] statistically not significant between voluntary and involuntary carless group, 3. [b] statistically not significant between one of the carless groups and car-owned group. 4. *, **, and *** indicate significance at 10%, 5%, and 1%.

Table 3.7 Logit Models Results

Variables	Model 1 (Carless HH vs. HH with cars)		Model 2 (Voluntary carless vs. HH with cars)		Model 3 (Voluntary carless vs. involuntary carless HH)	
	Coefficient	OR	Coefficient	OR	Coefficient	OR
Socio-Economic and Demographic Characteristics						
Household (HH) size	-0.37***	0.69***	-0.66***	0.52***	-0.17	0.84
Ratio of HH bicycles to HH Size (>5 years)	-0.03	0.97	0.06	1.06	0.19**	1.21**
Number of HH workers	-0.74***	0.48***	-0.64***	0.52***	0.07	1.07
Annual HH income ($)	-0.02***	0.98***	-0.01***	0.99***	0.02***	1.02***
Binary: 1=Hispanic or Latino	0.24***	1.27***	0.42**	1.52***	-0.01	0.99
HH ethnicity (baseline=Caucasian)						
Binary:1=African American	0.61***	1.84***	0.51**	1.67**	-0.35	0.71
Binary:1= Other	-0.01	0.99	-0.29*	0.75	-0.11	0.89
Highest HH educational attainment (baseline=graduate degree)						
Binary: 1= No high school degree	1.11***	3.02***	0.96***	2.62***	-0.11	0.90
Binary: 1= High school graduate	0.66***	1.93***	0.66***	1.94***	0.00	1.00
Binary: 1= Some college credit but no degree	0.15	1.16	0.03	1.03	0.09	1.10
Binary: 1= Associate degree	0.05	1.05	0.02	1.02	0.23	1.26
Binary: 1= Bachelor degree	-0.21**	0.81	-0.23	0.80	-0.03	0.97
Life cycle stage (baseline=no child)						
Binary: 1= If HH has 1 child	-0.17	0.85	0.58**	1.78**	0.70**	2.01**
Binary: 1= If HH has 2 children	0.10	1.11	0.61	1.84	0.16	1.17
Binary: 1= If HH has 3 or more children	0.48**	1.62**	0.52	1.68	-0.27	0.76
Binary: 1= If HH has older member (>65 years)	-0.39***	0.68***	-0.28*	0.76*	0.23	1.26
Type of Dwelling						
Binary:1=Single family housing	-0.98***	0.37***	-0.85***	0.43***	0.28	1.33
Land Use variables						
Population density (thousands per square mile)	0.02***	1.02***	0.03***	1.03***	-0.001	0.99
Mixed density index (MDI)	0.08***	1.08***	0.02	1.03	0.04*	1.04*
Land use entropy index (LEI)	0.95***	2.58***	3.95***	51.77***	2.59***	13.37***
Network density (facility miles of pedestrian-oriented links per square mile)	0.02***	1.02***	0.04***	1.04***	0.03**	1.03***
Percentage of population accessible by transit	5.52***	250.72***	8.51***	4949.89***	2.38**	10.78***
Constant	-1.26***		-4.87***		-3.21***	
Pseudo R-square	0.38		0.34		0.13	
Number of observations	32,811		31,142		1,134	

Notes: 1. *, **, and *** indicate significance at 10%, 5%, and 1%.

3.5.1.2 Travel Patterns

As shown in Table 3.6, travel patterns are markedly different between carless and motorized households. The average number of household trips is highest for motorized households at 8.71 person trips on travel day. Voluntary carless households make 6.89 person trips daily on average - less than involuntary carless households but the difference is not statistically significant based on a Tukey-Kramer post hoc test. However, 26.11 % of involuntary carless households did not travel on the survey day, which is significantly higher than for voluntary carless (20.79%) and motorized households (11.62%). This suggests that involuntary carless households may be less mobile than voluntary carless households, which could potentially contribute to a more isolated lifestyle, with a higher degree of isolation and a lower degree of well-being.

Moreover, approximately 41 % of trips by involuntary carless households have more than one stop for transferring to another mode, which is significantly higher than for voluntary carless (34.83%) and especially motorized (4.39%) households. One possible explanation is that involuntary carless households have fewer transportation options than other households. This result also supports my hypothesis that involuntary carless households experience greater levels of transport disadvantage than their voluntary counterparts, which, in turn, reduces their ability to participate in various activities, erodes their social support networks, and ultimately lowers their well-being.

Daily distance traveled also differs significantly between the three groups of households. Voluntary carless households travel significantly fewer miles, and fewer miles by vehicle than involuntary carless households and motorized households, possibly because voluntary carless households live in neighborhoods with mixed land uses and high densities as indicated above.

However, there is no statistically significant difference between the average distance per trip of voluntary and involuntary carless households.

While voluntary and involuntary carless households made up for this by using other modes, motorized households mostly relied on their vehicles for approximately 65% of trips. The use of transport mode also differs significantly between voluntary and involuntary groups. The use of motor vehicles for daily trips is marginally (but not significantly) higher for involuntary carless (9.93%) than for voluntary carless (9.46%) households. However, more than 50% of trips by voluntary carless households rely on walking and cycling, which is significantly higher than for involuntary carless households.

In general, voluntary carless households live in areas with a higher than average pedestrian oriented network density (more pedestrian street crossings and smaller block sizes) than involuntary carless households. However, involuntary carless households use public transit (21.79% of trips) more than their voluntary counterparts (18.6% of trips). Conversely, motorized households spend more time traveling daily than carless households, and as expected there is no statistical difference in average daily travel time and average travel duration per trip between voluntary and involuntary carless households.

From Table 3.6, I see that the average number of activities completed by carless households on travel day is lower than for the motorized group, with no significant difference between voluntary and involuntary carless households. However, the average activity duration per trip of involuntary carless households (273.83 minutes) is significantly higher than for voluntary carless households (257.66 minutes), and the distribution of trip purposes is significantly different. The involuntary carless group has a lower percentage of work (6.59%) and social trips (11.66%) than voluntary carless households (12.67% and 14.79%). Moreover, the percentages of work and social trips of voluntary and motorized households do not differ significantly, but they are significantly lower for involuntary carless households. However, there is no statistically significant difference in shopping, personal business, medical and religious trips between voluntary and involuntary carless households. Viewed in conjunction with the higher percentage of no trip on travel day, involuntary carless households appear to live more restricted lives, which may negatively impact their quality of life. This is not the case for voluntary carless households, who appear to have similar opportunities for jobs, social connections, and entertainment as motorized households.

3.5.2 Multivariate Results

3.5.2.1 Model Diagnostics and Fit

Variance inflation factors (VIF) for my explanatory variables indicated that multicollinearity is not a problem here (the largest VIF is <4). All three models in Table 3.7 passed the link test and their explanatory variables are jointly significant based on likelihood ratio tests. Models 2 and 3 passed the Hosmer-Lemeshow test but not model 1; an exploration of model 1 using a lowess graph (Long and Freese, 2006) did not reveal any problem, however. Moreover, the pseudo-R-square values of the three models are 0.38, 0.34 and 0.13 respectively, which suggests that my models fit the data reasonably well. Investigations of model residuals did not reveal any additional problems.

3.5.2.2 Socio-Economic and Demographic Characteristics

From Table 3.7, I see that model 1 (carless households vs. households with cars) and model 2 (voluntarily carless households vs. households with cars) have much in common. Many of the same variables are statistically significant and many of their odds ratios are comparable. Four out of the five socio-economic/demographic variables are significant. In line with my expectations, larger households are less likely to be carless (OR=0.69*** for model 1 and 0.52*** for model 2), in agreement with several life-cycle/life-course studies (Clark *et al.*, 2015; Khan and Habib, 2016). The two most important variables in term of odds ratio are Hispanic/Latino status (OR=1.27*** for model 1 and 1.52** for model 2) and number of household workers (OR=0.48*** for model 1 and 0.52*** for model 2), which shows that Hispanic/Latino households are more likely to be carless and that a household is less likely to be carless when the number of workers is larger. Likewise for ethnicity, African Americans are more likely to be carless (OR=1.84*** for model 1 and 1.67** for model 2) than Caucasians.

As expected (and as reported in Clark et *al.*, 2015), households where educational attainment is low are also much more likely to be carless: OR=3.02*** for model 1 and 2.62*** for model 2 in the case of people with no high school degree, with slightly lower odd ratios (1.94***) for high school graduates.

Let us now focus on model 3, which contrasts voluntary and involuntary carless households. Among socio-economic and demographic variables, only 2 variables are significant: income (OR=1.02***) but its impact is small with an odds ratio so close to one, and bicycle availability (OR=1.21**). Ethnicity does not come into play and neither does education.

3.5.2.3 Life-Cycle Stage

The agreement between models 1 and 2 is not as good for life-cycle stage variables but differences make sense. For model 1, households with over 3 children are more likely (OR=1.62**) to be carless but this does not hold for model 2, possibly because a number of larger and poorer California families are involuntarily carless. This result is consistent with Clark *et al.* (2015), who found that households with children are also less likely to acquire a car than those without children, possibly because those households with children might have been voluntary carless. At the other end of the spectrum, families with fewer children are more likely to be carless for model 2 (OR=1.78**) but not for model 1, which suggests that this does not hold for involuntary households. This finding echoes the life-cycle/life-course studies of Oakil *et al.* (2014) and Khan and Habib (2016), who reported that the birth of a child is likely to trigger the acquisition of a first household vehicle, possibly because many of the carless households they analyzed might have been involuntarily carless. Moreover, households with members over 65 are less likely to be carless (OR=0.68*** for model 1 and 0.76* for model 2).

3.5.2.4 Dwelling Type and Land Use Variables

As I hypothesized, dwelling type is important as households who live in single family dwellings are less likely to be carless (OR=0.37*** for model 1 and 0.43*** for model 2). This result agrees with previous cross sectional (Bhat and Pulugurta, 1998; Chu, 2002; Potoglou and Kanaroglou, 2008) and life-course (Khan and Habib, 2016) car ownership studies. Dwelling type is not significant for model 3.

Most land use variables are statistically significant for all three models but two have larger odds ratio: land use entropy (LEI) and especially percentage of population accessible by transit. In contrast, population density, mixed density index, and network density have either odds ratios close to one or they are not statistically significant so I do not discuss them further. As in Khan and Habib (2016), I see that higher LEI values are associated with a higher likelihood to be carless (OR=2.58*** for model 1), especially voluntarily (OR=51.77*** for model 2 and OR=13.37*** for model 3). As in Zergas (2010), Houston *et al.* (2015), and Huang *et al.* (2016), accessibility via transit (specifically, the percentage of population accessible by transit) is even more important with odds ratios of 250.72***, 4949.89***, and 10.78** for models 1, 2, and 3 respectively. This suggests that good transit accessibility is paramount for fostering the abandonment of motor vehicles by households.

3.6 CONCLUSIONS

The purpose of this paper was to improve our understanding of carless households by characterizing voluntary and involuntary carless households. In particularly, I analyzed the joint influence of various socio-economic, life cycle-stage, and land use variables on households' decision to be voluntarily carless. I analyzed data from the 2012 California Household Travel Survey using both simple tests and binary logit models.

Results of my univariate analyses show that both voluntary and involuntary carless households are more likely to be low-income than motorized households. However, compared to their involuntary counterparts, voluntary carless households tend to have a higher household income, a better education, a higher number of workers, and a lower number of children. They also live in higher density areas with better transit service and they walk and bike more to satisfy their daily travel needs. This and the lower percentage of social/recreational trips of involuntary carless households suggest that the latter are at a transportation disadvantage with more limited access, lower social participation, and possibly lower well-being.

My logit models show that households are more likely to be carless if they are Hispanic or African American, have a lower education level, and a larger family. Conversely, they are less likely to be carless if the number of workers is higher, elderly are present, and if they live in a single family dwelling. Although all land use measures are statistically significant in the model that contrasts carless with motorized households, only two land use variables stand out (and help households forgo their cars): land use diversity (via the land use entropy index) and even more so transit accessibility. Interestingly, very similar variables help households voluntarily forgo their cars, with a couple of notable exceptions: life cycle-stage variables like households with a single child are more likely to be voluntarily carless but having a larger family does not matter. Moreover, land use diversity and transit accessibility are even more important.

A policy implication of my findings is that to encourage households to live voluntarily without cars, planners and policy makers should focus on increasing land use diversity and transit accessibility. Different land use strategies such as transit-oriented development and urban villages (Delbosc and Currie, 2012) may also help households if they find themselves transitioning to an involuntary carless state. Second, increasing population density or providing more pedestrian-friendly neighborhoods seems to have only a very minor effect on a household's decision to live voluntarily without motor vehicles.

My study is not without limitations. First, a household's decision to forgo cars is likely connected to the type of neighborhood it wants to live in. Although recent studies suggest that a rich set of socio-demographic controls (which I have here) can reduce the residential self-selection bias (e.g., see Brownstone, 2008; Cao *et al.*, 2010; Boarnet, 2011; Cao and Cao, 2014), the best way to address this problem with cross-sectional data would be to construct a joint model of residential urban form and car-ownership (Boarnet and Crane, 2001; Kim and Brownstone, 2013).

Second, data restrictions prevented me from understanding the dynamics of switching in or out of vehicle ownership, which could be possible with longitudinal information (e.g., see

Zhang *et al.*, 2014; Clark *et al.*, 2015; Khan and Habib, 2016); this is left for future work. In particular, following Zhang *et al.*, (2014) and Zhang (2015), it would be of interest to develop integrated models that analyze biographical interdependences related to residential and car ownership choices with a focus on carless households.

Third, I defined carless households based on an indirect question asked in the CHTS. It would have been preferable to directly ask households if they chose to live voluntarily without motor vehicles or not.

Finally, I would like to mention that new car sharing arrangements, which are rapidly gaining in popularity, combined with the emergence of automated vehicles may change the relationship between car ownership and vehicle miles traveled (not to mention land use for parking). If these services become sufficiently affordable, they may provide greater accessibility to currently disadvantaged households who are unable to drive for a variety of reasons, while possibly increasing the miles traveled by voluntarily carless households, who currently drive little. The future will tell how much these exciting innovations will revolutionize the link between driving and vehicle ownership.

3.7 REFERENCES

American Community Survey (2008-2012), https://www.census.gov/programs-surveys/acs/.

Anowar, S., Eluru, N., & Miranda-Moreno, L. F. (2015). Analysis of vehicle ownership evolution in Montreal, Canada using pseudo panel analysis. *Transportation*, 1-18.

Banister, D. (2005). *Unsustainable transport: city transport in the new century*. Taylor & Francis.

Bhat, C. R., & Pulugurta, V. (1998). A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. *Transportation Research Part B: Methodological*, 32(1), 61-75.

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.

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. G. (2011). A broader context for land use and travel behavior, and a research agenda. *Journal of the American Planning Association*, 77(3), 197-213.

Bromley, R. D., & Thomas, C. J. (1993). The retail revolution, the carless shopper and disadvantage. *Transactions of the Institute of British Geographers*, 18(2), 222-236.

Brownstone, D. (2008). Key Relationships between the Built Environment and VMT. Paper Prepared for the Committee on the Relationships Among Development Patterns, Vehicle Miles Traveled, and Energy Consumption. *Transportation Research Board and the Division on Engineering and Physical Sciences*.

Bruin, J. (2011). newtest: command to compute new test @ONLINE. Feb. 2011. url: http://www.ats.ucla.edu/stat/stata/ado/analysis/.

Cameron, A. C., and Trivedi, P.K. (2010). Micro econometrics using Stata.

Cao, X., 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.

Cao, J., & Cao, X. (2014). The Impacts of LRT, Neighborhood Characteristics, and Self-selection on Auto Ownership: Evidence from Minneapolis-St. Paul. *Urban Studies*, *51*(10), 2068-2087.

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.

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

Clark, B., Chatterjee, K., & Melia, S. (2015). Changes in level of household car ownership: the role of life events and spatial context. *Transportation*, 1-35.

Connor-Linton, J. (2010). Chi-Square test. *Encyclopedia of research design*. Edited by: Neil J Salkind, Sage Publications, Inc. Thousand Oaks.

Curtis, F. (2009). Peak globalization: Climate change, oil depletion and global trade. *Ecological Economics*, *69*(2), 427-434.

Dargay, J. M. (2002). Determinants of car ownership in rural and urban areas: a pseudo-panel analysis. *Transportation Research Part E: Logistics and Transportation Review*, *38*(5), 351-366.

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

Delbosc, A., & Currie, G. (2012). Choice and disadvantage in low-car ownership households. *Transport Policy*, *23*, 8-14.

e Silva, J., Golob, T., & Goulias, K. (2006). 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), 121-131.

Gilbert, C. C. (1992). A duration model of automobile ownership. *Transportation Research Part B: Methodological*, *26*(2), 97-114.

Giuliano, G., & Dargay, J. (2006). Car ownership, travel and land use: a comparison of the US and Great Britain. *Transportation Research Part A: Policy and Practice, 40*(2), 106-124.

Goetzke, F., & Weinberger, R. (2012). Separating contextual from endogenous effects in automobile ownership models. *Environment and Planning A*, *44*(5), 1032-1046.

Greene, W. H. (2008). *Econometric Analysis*. 6th edition, Prentice Hall, Englewood Cliffs.

Houston, D., Boarnet, M. G., Ferguson, G., & Spears, S. (2015). Can compact rail transit corridors transform the automobile city? Planning for more sustainable travel in Los Angeles. *Urban Studies*, *52*(5), 938-959.

Huang, X., Cao, X., & Cao, J. (2016). The association between transit access and auto ownership: evidence from Guangzhou, China. *Transportation Planning and Technology*, 1-15.

Khan, N. A., & Habib, M. A. (2016). Life Course-oriented Approach of Modeling Vehicle Ownership State and Vehicle Type Choice Using Panel Data. In *Transportation Research Board 95th Annual Meeting* (No. 16-2832).

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.

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

Lanzendorf, M. (2003). Mobility biographies. A new perspective for understanding travel behaviour. In *10th International Conference on Travel Behaviour Research (IATBR), Lucerne* (Vol. 1015).

Li, J., Walker, J., Srinivasan, S., & Anderson, W. (2010). Modeling private car ownership in China: investigation of urban form impact across megacities. *Transportation Research Record: Journal of the Transportation Research Board*, (2193), 76-84.

Long, J. S., and Freese, J. (2006). *Regression models for categorical dependent variables using Stata*. Stata press.

Lovejoy, K. (2012). Mobility fulfilment among low-car households: Implications for reducing auto dependence in the United States. Research report (UCD-ITS-RR-12-27), Institute of Transportation Studies, University of California, Davis, 2012.

Macfarlane, G. S., Garrow, L. A., & Mokhtarian, P. L. (2015). The influences of past and present residential locations on vehicle ownership decisions. *Transportation Research Part A: Policy and Practice*, 74, 186-200.

Marquez, M. (1980). The travel behavior of Carless households: a case study for Los Angeles County. *Master's Thesis*, University of California, Los Angeles, 1980.

Matas, A., Raymond, J. L., & Roig, J. L. (2009). Car ownership and access to jobs in Spain. *Transportation Research Part A: Policy and Practice*, *43(*6), 607-617.

Mohammadian, A., & Miller, E. (2003). Dynamic modeling of household automobile transactions. *Transportation Research Record: Journal of the Transportation Research Board*, (1831), 98-105.

Nolan, A. (2010). A dynamic analysis of household car ownership.*Transportation research part A: policy and practice*, *44*(6), 446-455.

Oakil, A. T. M., Ettema, D., Arentze, T., & Timmermans, H. (2014). Changing household car ownership level and life cycle events: an action in anticipation or an action on occurrence. *Transportation*, 41(4), 889-904.

Oakil, A. T. M. (2016). Securing or sacrificing access to a car: Gender difference in the effects of life events. *Travel Behaviour and Society*, *3*, 1-7.

Oakil, A. T. M., Manting, D., & Nijland, H. (2016). Determinants of car ownership among young households in the Netherlands: The role of urbanization and demographic and economic characteristics. *Journal of Transport Geography*, *51*, 229-235.

Paaswell, R. E., & Recker, W. W. (1978). *Problems of the Carless*. Praeger Publishers.

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

Prillwitz, J., Harms, S., & Lanzendorf, M. (2006). Impact of life-course events on car ownership. *Transportation Research Record: Journal of the Transportation Research Board*, (1985), 71-77.

Potoglou, D., & Kanaroglou, P. S. (2008). Modelling car ownership in urban areas: a case study of Hamilton, Canada. *Journal of Transport Geography*, *16*(1), 42-54.

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

Ramsey, P.H. (2010). Post hoc analysis. *Encyclopedia of research design*. Edited by: Neil J Salkind, Sage Publications, Inc. Thousand Oaks.

Sivak, M. (2014). *Has motorization in the U.S. peaked? Part 4: Household without a light duty vehicles (*Report No. UMTRI-2014-5). Ann Arbor: The University of Michigan

Transportation Research Institute. Available at: http://deepblue.lib.umich.edu/bitstream/handle/2027.42/102535/102988.pdf

Soltani, A. (2005). Exploring the impacts of built environments on vehicle ownership. In *Proceedings of the Eastern Asia Society for Transportation Studies* (Vol. 5, pp. 2151-2163).

The New York Times. (2015). California Governor Orders New Target for Emissions Cuts, by Adam Nagourney, April 29, 2015.

The Washington Post (2015). Release of encyclical reveals pope's deep dive into climate science. by Anthony Faiola, Michelle Boorstein and Chris Mooney , June 18, 2015. Available from https://www.washingtonpost.com/local/how-pope-franciss-not-yet-official-document-on-climate-change-is-already-stirring-controversy/2015/06/17/ef4d46be-14fe-11e5-9518-f9e0a8959f32_story.html.

Van Acker, V., & Witlox, F. (2010). Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship. *Journal of Transport Geography*, *18*(1), 65-74.

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.

Wabed, A. S. and Tang, X. (2010). Analysis of variance. *Encyclopedia of research design*. Edited by: Neil J Salkind, Sage Publications, Inc. Thousand Oaks.

Washington, S. P., Karlaftis, M. G., & Mannering, F. L. (2010). *Statistical and econometric methods for transportation data analysis*. CRC press.

Whelan, G. (2007). Modelling car ownership in Great Britain. *Transportation Research Part A: Policy and Practice*, *41*(3), 205-219.

Yamamoto, T., Kitamura, R., & Kimura, S. (1999). Competing-risks-duration model of household vehicle transactions with indicators of changes in explanatory variables. *Transportation Research Record: Journal of the Transportation Research Board*, (1676), 116-123.

Yamamoto, T. (2008). The impact of life-course events on vehicle ownership dynamics: The cases of France and Japan. *IATSS research*, *32*(2), 34-43.

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

Zhang, J., Yu, B., & Chikaraishi, M. (2014). Interdependences between household residential and car ownership behavior: a life history analysis. *Journal of Transport Geography*, *34*, 165-174.

Zhang, J. (2015). The life-oriented approach and travel behavior research. A discussion paper for the Workshop "Life-Oriented Approach for Transportation Studies," 14th *International Conference on Travel Behaviour Research*, Windsor, UK, July 19-23.

Zimmerman, C. A. (1982). The life cycle concept as a tool for travel research. *Transportation*, *11*(1), 51-69.

Chapter 4. Determinants of Long-distance Commuting: Evidence from the 2012 California Household Travel Survey

4.1 INTRODUCTION

The pressing need to curb the environmental impacts of transportation has turned reducing daily vehicle-miles traveled into an important policy goal. One way to achieve this goal is to reduce the length of commuting trips and/or to increase the use of alternative modes of transportation (Schiller *et al.*, 2010; Modarres, 2013; Motte *et al.*, 2016). Although many studies have investigated the determinants of commuting (see Table 4.2), I could not find a study that explicitly focused on long-distance commuting in the US (in this work, long-distance trips are 50 miles or more one-way). Since long-distance commuting involves more time and out-of-pocket costs, they differ substantially from short-distance commuting trips, so people are likely to treat these trips differently (Jin and Horowitz, 2008) and the conclusions reached in the commuting literature likely do not apply.

Long-distance commuting has a number of important dimensions. First, its environmental impacts, which include local air pollution, noise, and the emission of greenhouse gases, could be substantial as long-distance commuting by car increasingly takes place in congested conditions on roads nearing capacity (FHWA, 2006; Van Nostrand *et al.*, 2013; Outwater *et al.*, 2016). Second, commuting has a social dimension especially for lower wage employees who are accepting long commutes in order to access more affordable housing (Lukas, 2011; Motte *et al.*, 2016). Finally, long-distance commuting has economic impacts as it is well known that more accessible regions are ceteris paribus more successful economically (Krugman, 1991). In summary, understanding the determinants of long-distance commuting behavior is important for a range of economic, environmental, and social reasons. Unfortunately, there is a dearth of research on long-distance commuting, especially in the US, partly because of data availability. The 2012 California Household Travel Survey (CHTS) offers a rare opportunity to study long-distance travel so the purpose of this study is to analyze CHTS data to gain a better understanding of long-distance commuting in California.

In spite of the importance of the topic, a review of the literature (see Tables 4.1 and 4.2) suggests that several issues have not been fully addressed in published studies. For instance, travel behavior and residential self-selection have rarely been jointly studied. In addition, few published papers have considered car-ownership as a mediating variable between residential location decision and commuting or long-distance travel behavior even though ignoring the endogeneity of this relationship may bias econometric results (Van Acker and Witlox, 2010; e Silva *et al.*, 2012). Furthermore, previous studies offer few insights about how residential land value influences residential long-distance commuting decisions. By contrast, this study strives to fill this gap by assessing the effects of different socio-economic, land use and land value variables on the likelihood that households commute long-distance by explicitly considering car-ownership as a mediating variable and by incorporating self-selection effects due to socio-economic characteristics.

The remainder of this chapter is organized as follows. Section 4.2 provides a review of selected papers dealing with long-distance travel and the factors that influence commuting behavior. Section 4.3 describes my datasets, discusses my explanatory variables, and gives a short account of long-distance travel in California. Section 4.4 presents my methodology. It is followed by a discussion of my results in Section 4.5. Finally, Section 4.6 offers conclusions and presents some suggestions for future research.

4.2 LITERATURE REVIEW

This section starts with an overview of the different ways long-distance travel has been measured in previous studies and then focuses on selected long-distance travel papers. I then review the variables used in studies dealing with commuting behavior.

4.2.1 Methodological Issues

As shown in Table 4.1, there is a great deal of variation in the definition of long-distance travel: while most studies I reviewed used distance-based measures, a few relied on travel time measures, and the rest adopted a combination of both.

In the U.S., the minimum distance for a long-distance journey varies between 50 and 100 miles one way: studies based on the 2001 National Household Travel Survey and State level travel surveys commonly adopt a 50 mile threshold (e.g. Erhardt *et al*., 2007; LaMondia and Bhat, 2011; LaMondia *et al.*, 2016) whereas studies that analyze the 1995 American Travel Survey use at least a 100 mile threshold. In Europe, these measures range from 30 km to 100 km one way, with at least 30 km (Euclidian distance) for Sandow and Westin (2010) in Sweden, 50 km for Rohr *et al.* (2013) in the U.K. and Limtanakool *et al.* (2006a-b) in the Netherlands, and 100 km in Germany for Reichert and Holz-Rau (2015). Measures based on travel time or on combination of both distance and travel time have also been used: for example, Cassel *et al*. (2013) considered a 40 minute commuting threshold in Sweden while Jin and Horowitz (2008) adopted a joint threshold of 50 miles or 60 minutes one way in the United States.

Second, the methodology used to collect long-distance travel data also varies from country to country. In some countries researchers used mail surveys while others relied on telephone or in-person visits (Axhausen, 2001; Dargay and Clark, 2012).

A third issue is whether to conduct dedicated long-distance travel surveys. A number of European countries have been collecting local and long-distance travel data in the same survey, whereas, the United States has gone back and forth: the 1995 American Travel Survey (ATS) focused on long-distance travel whereas the Nationwide Personal Transportation Survey (NPTS) collected local travel data (Dargay and Clark, 2012), but then the 2001 National Household Travel Survey (NHTS) collected both local and long-distance travel data, before the 2009 National Household Travel Survey reverted to focusing on local travel. This may explain the relative dearth of long-distance travel studies in the United States in recent years.

A fourth issue is underreporting, which is reflected in how far back respondents are asked to recall their long-distance trips (this ranges from a couple of weeks to 3 months.) For example, in the United States the 1995 ATS used a 3 months recall period while the 2001 NHTS used 4 weeks. In Europe, the UK National Travel Survey (NTS) had a 3 week threshold whereas the 2008 Mobility in Germany (MiD) Survey adopted a 3 month recall period.

In summary, the measures of long-distance travel as well as the data collection methodologies have been varying substantially from study to study and over time.

4.2.2 Previous Long-distance Travel Studies

As shown in Table 4.1, compared to Europe (e.g. Creemers *et al.*, 2012; Cassel *et al.*, 2013; Reichert and Holz-Rau, 2015) there are relatively few empirical studies of long-distance travel in the U.S., possibly because of data limitation. Since 2009 National Household Travel Survey did not incorporate long-distance travel questions, most U.S. studies analyze the 1995 American Travel Survey (e.g. Van Nostrand *et al.*, 2013; Sivaraman *et al.*, 2016), a few rely on data from the 2001 National Household Travel Survey (e.g. Jin and Horowitz, 2008; LaMondia and Bhat, 2011), and others analyze statewide surveys (e.g. Erhardt *et al*., 2007; LaMondia *et al.*, 2016).

Published long-distance travel studies can be broadly divided into three categories. The first category comprises studies that model demand for long-distance travel. It can be further divided in two groups: general-purpose models of long-distance travel (e.g. Erhardt *et al*., 2007; Jin and Horowitz, 2008; Rohr *et al.*, 2013; Outwater *et al.*, 2016) and long-distance vacation and leisure trips (e.g., see Van Nostrand *et al.*, 2013; LaMondia *et al.*, 2014). Outwater *et al.* (2016) is an example of the former: they developed a tour-based, multimodal micro-simulation model of annual long-distance passenger travel for all households and trip purposes in the U.S. Van Nostrand *et al.* (2013) focused instead on long-distance trips for leisure. Using a multiple discrete–continuous extreme value structure, they formulated an annual vacation destination choice and time allocation model for the U.S. to jointly predict the different vacation destinations that a household could visit during a year and the time allocated to each.

Papers in the second category of long-distance travel studies are concerned with mode choice. These studies report that a number of socio-economic variables significantly influence mode choice for long-distance travel, including household size (LaMondia *et al.*, 2016), number of workers in the household (LaMondia *et al.*, 2016), income (Reichert and Holz-Rau, 2015; LaMondia *et al.*, 2016), car accessibility and education level (Reichert and Holz-Rau, 2015), and gender (Limtanakool *et al.*, 2006a). Other influential variables include land use variables such as population density (Limtanakool *et al.*, 2006a), travel cost and travel time (Limtanakool *et al.*, 2006a; Creemers *et al.*, 2012), as well as travelers' attitude (Creemers *et al.*, 2012) and trip purpose (LaMondia *et al.*, 2016).

The third category of papers examines the determinants of long-distance travel (Limtanakool *et al.*, 2006b; Dargay and Clark, 2012; LaMondia *et al.*, 2014; Holz-Rau *et al.*, 2014). Let us review these papers more in detail since it is the focus on my work here.

Limtanakool *et al.* (2006b) analyzed the 1998 National Travel Surveys for the UK and the Netherlands to understand the factors that influence the decision to undertake specific long-distance trips with particular modes. They concluded that gender, role in the household, and income are important determinants of long-distance commuting and business travel. Their analysis further suggests that urban structure, population size, and local population density play a part in long-distance travel participation.

This topic has also received attention in Great Britain where Dargay and Clark (2012) examined the determinants of long-distance travel based on data from the 1995–2006 National Travel Surveys using a reduced form model that captures the joint relationship between long-distance travel and car ownership. Apart from the importance of gender, age, employment status, and household composition, they found that long-distance travel is strongly related to income. Moreover, air travel is most income-elastic, followed by rail, car and bus.

In the sole recent U.S. study I could find, LaMondia *et al.* (2014) analyzed 1200 self-reported retrospective questionnaires using an ordered probit model to tease out the factors that impact the frequency of different types of long-distance travel. Their results show that the type of a long-distance trip matters and is influenced by the presence of a spouse and children. Moreover, a better education and a higher income increase most types of long-distance travel.

To my knowledge, Holz-Rau *et al.* (2014) is the only paper to specifically examine the impacts of urban form on distance travelled during long-distance trips. Their analysis of German data from 2008 using Heckman models and ordinary least squares shows that socio-demographic variables influence long-distance and daily trips in the same way while urban form impacts them in opposite ways. For example, residents of small municipalities and low-density neighborhoods make fewer and/or shorter long-distance trips than those living in large cities and high‑density neighborhoods, although the latter travel shorter distances in their daily lives.

In summary, this review suggests that the socio-economic characteristics of travelers are important determinants of long-distance travel, but offers few insights as to how different land use variables influence the decision to undertake long-distance journeys. Moreover, although several of these studies include car ownership to explain long-distance travel behavior, only Dargay and Clark (2012) endogenized car ownership. Another limitation is that most published papers use the same framework for all long-distance trips, even though commuting, business, and vacation long-distance trips likely have different determinants, as reported by LaMondia *et al.* (2014).

4.2.3 Commuting Travel Behavior Studies

This study also relates to the literature that analyzes factors affecting commuting travel behavior. Selected papers published over the past 10 years are summarized in Table 4.2. My review suggests that the following socio-economic variables play in important role in commuting travel behavior: income (Van Ommeren and Dargay, 2006; Marion and Horner, 2008; Bergantino and Madio, 2015), gender (Axisa *et al.*, 2012; Oakil *et al.*, 2015), age (Axisa *et al.*, 2012; Maoh and Tang, 2012), education level (Sandow, 2008), number of workers in the household (Surprenant-Legault *et al.* 2013), marital status (Neto *et al.*, 2014), life events (Clark *et al.*, 2016), having informal jobs (Motte *et al.*, 2016), and length of employment (Bergantino and Madio, 2015). In addition, land use variables found to be important for commuting distance and time include population density (Zhao *et al.*, 2011; Dai *et al.*, 2015), job housing balance (Zhao *et al.*, 2011; Dai *et al.*, 2015), residential location (Elldér, 2014; Hjorthol and Vågane, 2014), land use diversity (Maoh and Tang, 2012), and job proximity (Cervero and Duncan, 2006; Kawabata and Shen, 2007; Watts, 2009). Furthermore, housing value (Plaut, 2006); commuting mode (Dai *et al.*, 2015), reward to avoid rush hour (Ben-Elia and Ettema, 2011) and transport accessibility (Dai *et al.*, 2015) also influence commuting behavior.

Only a few commuting behavior studies have focused on long-distance travel. I found only two published over the last 10 years and they are both concerned with Sweden. The first one (Sandow and Westin, 2010) analyzed the duration of long-distance commuting trips (30 km or more) and the characteristics of commuters using a dataset that spans 1995 to 2005. Findings show that previous experience with long-distance commuting influences commuting behavior, along with income (which is positively correlated with the persistence of long-distance commute over time) and gender (males commuters benefit more from long-distance commuting than female commuters.)

The second paper (Cassel *et al.*, 2013) examined survey data of unemployed job seekers in Dalarna County, Sweden, to predict the probability that an individual is willing to commute for more than 40 minutes. Their linear probability model shows that gender, level of education, and the presence of children in the household influence the willingness to commute. In addition, men are more prone to commute long-distance than women, and age interactions with length of unemployment, educational level, and gender are statistically significant.

This brief review indicates that the determinants of long-distance commuting have received very limited attention to-date, but it suggested variables to consider in my models.

Table 4.1 Summary of Selected Long-distance (LD) Travel Studies (2006-2016)

Authors (Year Published)	Data [Period Analyzed] Measures of LD Travel Models	Socio-Economic Variables Land-Use Variables Other Variables	Key Results
Long-distance Travel Modeling Framework			
Outwater *et al.* (2016)	• 1995 ATS, 2003 Ohio Statewide HTS, 2010 Colorado Front Range TS, 2012 CHTS. • Outbound and return trip > 50 miles from home, with or without stops. • Tour based micro simulation model.	• Income; family composition; working status; vehicle ownership. • HH residential location. • NA.	A disaggregate tour-based approach is feasible to predict annual LD passenger travel demand for all HHs in the U.S.
Sivaraman *et al.* (2016)	• 36,401 HH data from 1995 ATS, U.S. [1995]. • One way trips >100 miles. • Regression and Multinomial models.	• Age; education; ethnicity; income. • Region • HH location: metropolitan, mid-west, south, west, northeast.	Low income HHs are more likely and HHs with full time employed members are less likely to travel to visit friends.
Rohr *et al.* (2013)	• Data from 116,039 NTS respondents and 3 week recall survey of LD trips [2002-2006] + 65,357 HH interviews [2009], UK. • One way trips > 50 miles • Frequency, mode and destination choice model	• HH income; gender; part time/full time worker; students. • Destination region; • Mode destination components.	The frequency of LD trips is strongly related to income; car and air usage increase with income while coach use decreases.
Van Nostrand *et al.* (2013)	• 6,715 randomly sampled HH data from the 1995 ATS, US [1995]. • One way trips >100 miles. • Multiple discrete-continuous extreme value.	• Retired HH. • Destination: MSA, leisure employment density, land area. • Distance and level of service; destination characteristics: total lodging and non-lodging cost/night, length of coastline, highway distance to destination, summer & winter T°.	Travel times & costs, lodging costs, leisure activity opportunities, length of coastline, and weather conditions at the destinations => influence HHs' destination choices for LD vacations.
LaMondia and Bhat (2011)	• Data from 28,294 households from the 2001 NHTS, U.S. [2001]. • One-way trips ≥ 50 miles. • Copula based model.	• HH: income, size; ownership: home, vehicle; phone access; life-cycle: children, adults; # of: drivers, workers. • Home: MSA pop., city size, census region. • HH travel season; travel day; impact of 9/11.	HHs appear to have more emotional attachment to activities associated with long-distance travel, relative to those they pursue on a daily basis.

Authors (Year Published)	Data [Period Analyzed] Measures of LD Travel Models	Socio-Economic Variables Land-Use Variables Other Variables	Key Results
Jin and Horowitz (2008)	• 3,322 LD trip records from the 2001 NHTS, USA [2001]. • One way trip ≥ 50 mi or ≥ 60 min. • Multinomial logit.	• Age; gender; education; HH: income, size, car ownership, presence of child. • NA. • Trip: purpose, mode, travel time, travel companions; activity-related factors: duration.	Trip duration, activity duration, travel day type, whether traveling with other persons, and the presence of young children impact departure time choice for LD trips.
Erhardt *et al*. (2007)	• 8,000 HH data from LD travel survey, Ohio, USA [2002-03]. • One way travel >50 miles • Binary logit.	• HH: # of workers, # of autos, size, income, # of students; occupation; age, gender. • NA. • Dwelling type.	HHs with more automobiles and higher incomes are more likely to travel LD.
Long-distance Mode Choice			
LaMondia *et al.* (2016)	• 4,330 respondents from State LD travel Survey, Michigan, US [2009]. • One way trips > 50 miles • Negative binomial and binary logit.	• HH size; # of vehicles; # of HH workers. • County population density; % employment in education; % of HH income > 60k. • Travel time.	Larger HH sizes, workers and income take more LD trips. Trip purpose and income dominate mode choice decisions.
Reichert and Holz-Rau (2015)	• 25,922 HH information from MiD Survey, Germany [2008]. • One-way trips >100 km. • Logit model.	• Employment; education; age; gender, HH: income, type; # of cars. • Population: size, density; rail access. • NA	Urban dwellers take more LD trips, in particular by train and air (controlling for income, car accessibility, and education).
Creemers *et al.* (2012)	• 492 respondents from a stated preference survey, Flander, Belgium [2010]. • 10-40 km (medium/long-distance) • Logistic regression	• Age; gender; # of cars; frequency of public transport use • NA. • Transport variables: cost, time, transfer/wait time, seat availability; punctuality; attitude/perception: modes, cost, comfort.	Transport system specific factors, socio-economic variables, attitudinal factors, perceptions and the frequency of using public transport => preference of light rail transit for medium/LD trip.
Limtanakool *et al.* (2006a)	• 6,330 individual observations from the NTS, Netherlands [1998]. • One way trips > 50 km. • Binary logit model for mode choice.	• Gender; HH: type, income; education. • Origin/destination: population density; type of municipality; train station. • NA.	Land use attributes and travel time impact variations in mode choice for medium/LD travel (controlling for socioeconomic characteristics).
Monzon and Rodríguez-Dapena (2006)	• 3,446 HHs, Madrid-Barcelona corridor survey, Spain, 1992. • NA • Weighted estimator for mode choice.	• Income • NA • Transport supply: time, frequency, price.	Result allows to achieve a more flexible cheaper survey procedure for interurban transport planning activities.

Authors (Year Published)	Data [Period Analyzed] Measures of LD Travel Models	Socio-Economic Variables Land-Use Variables Other Variables	Key Results
Determinants of Long-distance Travel			
Holz-Rau *et al.* (2014)	• 25,922 HH information from MiD Survey, Germany [2008]. • One-way trips >100 km. • OLS and Heckman model.	• Employment; gender; age; education; HH: income, type. • Population: size, density; land use mix. • NA	LD trips for residents of small municipalities and low-density neighborhoods < those living in large cities and high-density areas.
LaMondia *et al.* (2014)	• 1200 individuals recruited to an online retrospective survey, US [02/2013]. • No distance based threshold used. • Ordered probit model.	• Age; education; gender; years in current residence; employment status; income; HH: size, vehicles, spouse, children. • Region; home: straight line & road distance. • NA.	Education and income => increase most types of LD travel & having a spouse or children => decreases some types of LD travel.
Cassel *et al.* (2013)	• Questionnaire survey of 151 individual, Dalarna, Sweden [12/2009-03/2010]. • Commute time > 40 minutes • Linear probability model.	• Age; age squared; presence of children; gender; access to a car; education; employment status. • NA. • NA.	Factors influencing the willingness to commute: gender, level of education, and the presence of children in the HH.
Dargay and Clark (2012)	• A sample of 147,826 individuals from the NTS, Great Britain [1995-2006]. • One-way trips ≥ 50 miles. • Reduced form equation.	• HH income; gender; age; employment status; company car; # of: adults, children. • Region of residence; metropolitan area. • Dwelling type; length of residence.	LD travel is strongly related to income; air is most income-elastic, followed by rail, car and finally coach.
Sandow and Westin (2010).	• 178,662 commuters' information from a longitudinal register data, Sweden [1995-2005]. • One way travel ≥ 30 km (Euclidian) • Multilinear regression.	• Gender; age; income; education level; employment sector; having children; partner is a long-distance commuter. • Living in city region. • NA.	Economic incentives (higher income) => continuing to LD commuting more than a few years. Male commuters ~ more economic outcome of LD commuting.
Limtanakool *et al.* (2006b)	• 20,773 HHs from the NTS, Netherlands [1998] and 30,150 from the NTS, UK [1998]. • One way trip ≥ 50 km • Binary logit	• Age; gender; HH income; HH composition: single worker, one-worker couple, two-worker couple, one worker-family, two-worker family, family more than two workers, other HH type. • Population density. • NA.	Factors influencing medium/LD travel: the overall structure of the urban system combined with the size of the country and the local population density.

Note: LD: long-distance, HH: household, OLS: Ordinary Least Square, ATS: American Travel Survey, HTS: Household Travel Survey, MiD: Mobility in Germany, NHTS: National Household Travel Survey, NTS: National Travel Survey.

Table 4.2 Summary of Selected Studies on Factors Affecting Commuting (2006-2016)

Authors (Year Published)	Data [Period Analyzed] Models	Socio-Economic Variables Land-Use Variables Other Variables	Key Results
Clark *et al.* (2016)	• 15,200 workers from the UKHLS, UK [2009/10-2010/11]. • Binary logit model.	• HH: size, income, child presence; live with a partner; gender; age; education; employment type. • Settlement type; population density; # of: employment centers, food stores; change in residential context. • Travel time to: employment & town center; # of bus stops; rail stations; social environment; attitudes; life events: relocation, had child, starting/stopping cohabitating, switched employer, acquired driving license; commuting distance.	Switching to non-car commuting becomes over 9 times more likely as the distance to work drops below 3 miles. It is also 1.3 times more likely for those with pro-environmental attitudes.
Motte *et al.* (2016)	• 34,000 HHs from HTS, Rio De Janeiro, Brazil [2003]. • SEM.	• Gender; position in HH; education; sector of activity. • Distance to CBD; place of work; transport mode. • NA	Commuting distances and times are shorter for the informal sector.
Bergantino and Madio (2015)	• 77,029 individuals from QLFS, UK [2004-2011]. • Multinomial logit	• Age; gender; income; job type; couple; presence of children. • NA. • Length of employment; homeownership type.	Earnings and length of employment are important in explaining commuting behavior.
Dai *et al.* (2015)	• 816 respondents from 36 communities, Guangzhou City, China [2014]. • Multilevel logit.	• Gender; age; education; employment; occupation; HH: size, income, # of employed people; # of cars. • Transport accessibility; population density. • Housing source; home based job opportunity; commuting mode.	Population density, jobs-housing balance, transport accessibility and commuting mode significantly affect commuting time and distance.
Oakil *et al.* (2015)	• 925 respondents from TBO and NTS, Netherlands [2006]. • Binary logit model.	• Gender; education; working status; company car. • NA • Commute duration; daily activity before/after commute: childcare, child related travel, HH work, works at home; attitudes.	Women commute more during morning rush hours but less during evening rush hours.
Owen and Levinson (2015).	• 2,082 block group data from ACS, US [2007-20011]. • Logit model.	• Mean (HH): income, size, vehicles; % of population: white, non-Hispanic, 25+ with B.A. /B.S. or higher. • Transit accessibility: maximum, average, variance; auto accessibility.	Increases in both maximum and average accessibility are associated with increases in transit mode share.

Authors (Year Published)	Data [Period Analyzed] Models	Socio-Economic Variables Land-Use Variables Other Variables	Key Results
Elldér (2014).	• 785,369 people from a longitudinal dataset, Sweden [1990-2010]. • Multilevel model.	• Gender; income; education; life course. • Regional location; urban area; # of job opportunities/# of gainfully employed residents. • Branch of industry.	Results show a growing variation in home-work distance for workers living in the same neighborhoods.
Hjorthol and Vågane (2014).	• 7,174 individuals from NTS, Norway [2009]. • OLS.	• Age; gender; education; working hours; children; occupation. • Place of residence. • NA.	Women do not commute as far as men in comparable groups.
Neto *et al.* (2014)	• 549,867 census individuals, Sao Paulo, Brazil [2010]. • Probit model.	• Income; HH size; status: marital, work; children; inactive senior; education; age; occupation. • NA. • Property characteristics: ownership, # of rooms.	Marital status has a stronger influence on the commuting time of working women.
Surprenant-Legault *et al.* (2013)	• 43,267 HH from O-D survey of AMT, Montreal, Canada [2003]. • OLS.	• Gender; age; mean age (two workers); children; HH income. • Accessibility of jobs; home-work distance: longest, shortest, partners; home location; home-work network distance/home-work Euclidian distance • Modal characteristics.	For every 1% increase in a partner's commuting distance, total commute distance increases by less than 1%.
e Silva *et al.* (2012).	• 7, 227 workers from OD survey, Montreal, Canada [2003]. • SEM.	• Age; gender; average age: HH, adult; HH: income, size, # of workers, teens, one/two members. • Work/home: density, accessibility: car/transit, distance to CBD, index: entropy & compactness, km road/person; % of people within: 500 m of subway, 1 km of freeway node. • NA	Commuting distance is negatively influenced by the residence in a central denser and accessible area and by residence in a mixed area well served by freeways.
Axisa *et al.* (2012)	• 357,164 census individuals, Canada [2006]. • OLS.	• Status: migration, marital; employment: status & type; age; gender; HH: income, structure; youngest child age; children. • Geographic context (census): metropolitan area, agglomeration area, metropolitan & agglomeration influence zones.	Longer commuting distance for recent migrants who reside in accessible rural areas.

Authors (Year Published)	**Data [Period Analyzed]** **Models**	**Socio-Economic Variables** **Land-Use Variables** **Other Variables**	**Key Results**
Maoh and Tang (2012).	• 22,309 census individuals, Windsor, Canada [2006]. • OLS.	• Gender; age; occupation; status: employment & migration. • Entropy index; location quotient (particular occupation type). • Transportation mode.	Workers living in mixed land use neighborhoods (tracts) have a shorter commute distance.
Ben-Elia and Ettema (2011).	• Survey data from 339 commuters, Netherlands [2006]. • Mixed logit.	• Gender; education. • NA • Attitude/perception of: drive early, drive late, no driving.	Rewards reduce the shares of rush-hour driving, shift driving to off-peak times.
Zhao *et al.* (2011)	• 712 respondents from HHIS, Beijing, China [2001]. • OLS model.	• HH income; occupation; employment. • Population density; job-housing balance; accessibility. • Transport mode.	Jobs-housing balance has significant association with a worker's commuting time.
Manaugh *et al.* (2010).	• 31,997 trips from O-D survey of AMT, Montreal, Canada [2003]. • OLS, simultaneous two-equation model.	• Age; gender; income; # of vehicles; with children. • Origin: park, big box, highway, train, waterfront, urban mixed use, commercial streets, single family; destination: job center, isolated sub, waterfront, mixed use, office park. • NA.	While single-family origins generate much longer trips, suburban destinations are not shown to be associated with longer commutes.
Wang & Chai (2009).	• 736 HHs from HHIS, Beijing, China [2001]. • SEM	• Gender; age; occupation; working unit; income; marital status; education. • Jobs-housing relation. • Sources of housing; transport mode; commuting time.	Commuters who live in work housing units have fewer commuting trips than those who live in market houses.
Watts (2009).	• 468 journeys to work, Sydney, Australia [2001]. • OLS, spatial error & SAR models.	• Socio-economic status; relative real wage; % of : >15 years did not complete Year 10, employees work<=15 hrs/week, speak only English; % of HH : with no car, own houses; • Distance: min. commute, centroid to centroid; job proximity. • NA.	Both minimum commute distance and job proximity have better explanatory power jointly with socioeconomic variables.
Marion and Horner (2008).	• 133,715 individuals from PUMS, U.S. [2000]. • Binary logit model.	• Age; gender; education; marital status; presence of children; ethnicity; HH income; dwelling type. • Central area; suburban area. • Carpool; work hours; depart time; recently moved.	A decrease in total HH income increases the odds of extreme commuting.

Authors (Year Published)	**Data [Period Analyzed]** **Models**	**Socio-Economic Variables** **Land-Use Variables** **Other Variables**	**Key Results**
Sandow (2008).	• 72,586 individuals from a longitudinal dataset, Sweden [1985-2003]. • Logit model.	• Age; income; status: marital, employment; children; education. • Population density; employment opportunities: low degree, medium degree, high degree.	Workers with higher education & higher income have a higher probability of longer commuting distance.
Kawabata and Shen (2007)	• 2143 commuters from Census TP Package, Bay Area, US [1990 & 2000]. • Spatial Lag & Error	• HH income. • Job accessibility; income; density: employment, population; % of: female headed HH, female labor force, no high school diploma, foreign born, different races, different jobs. • NA	Greater job accessibility => shorter commuting time for driving alone & for public transit.
Helminen & Ristimäki (2007).	• 19,068 respondents from LFS, Finland [2001]. • Logit model.	• NA. • NA. • Length of commuting trip.	The probability of working at home increases with commuting distance.
Cervero and Duncan (2006).	• 16,503 HHs from BTS, San Francisco, US [2000]. • OLS.	• Motor vehicles/licensed drivers; income; employment type; full/part time; age; gender; Latino. • Within 4 miles: occupationally matched jobs, total jobs, retail and service jobs.	Availability of many jobs within four miles of home significantly reduces VMT and VHT for work trips.
e Silva *et al.* (2006).	• 7,849 individuals from mobility survey, Lisbon, Portugal [1993-94]. • SEM.	• Age; gender; HH: teens, size, # of workers; income; work time. • Density; distance to CBD; mix of jobs; index: entropy & compactness; trunk roads/person; % of people (400 m): bus stops; train, metro, ferry station, freeway junction (1000m). • NA	Land use patterns of residence and employment have respectively negative and positive effects on commuting distance.
Plaut (2006)	• 7,595 HH from AHS, US [2001]. • Seemingly unrelated regression (SUR).	• Income; age; race; # of cars; HH size; receive dividend income. • Living in: central/secondary city, rural; within ½ block: green space, apartments; shopping service; property value/rent. • Gated community; police protected; dwelling type.	Commute distance is sensitive to housing value, but more for women owners than for men.
Shearmur (2006).	• 290,000 census individual, Montreal, Canada [2001]. • OLS.	• Gender; employment sector; occupation; income; contribution to HH income; presence of young child. • Distance to CBD from: work, residence; job location. • NA.	Women travel farther to access job than men.

Authors (Year Published)	Data [Period Analyzed] Models	Socio-Economic Variables Land-Use Variables Other Variables	Key Results
Van Ommeren and Dargay (2006).	• 9,361 HHs from NTS, UK [1989-1991, 1999-2001]. • OLS (reduced form).	• Income; gender; age; part-time employment; presence of: adults, children. • Population density; municipality size. • NA.	The income elasticity of commuting speed is ~ 0.13.

Note: HH: Household, UKHLS: UK Household Longitudinal Survey, HTS: Household Travel Survey, QLFS: Quarterly Labor Force Survey, ACS: American Community Survey, TBO: Dutch Time Use Survey, NTS: National Travel Survey, AMT: Agence métropolitaine de transport, BTS: Bay Area Transport Survey, PUMS: Public-Use Microdata Samples. TP Package: Transportation Planning Package, HHIS: Household Interview Survey.

Although a handful of studies summarized in Table 4.2 controls for residential self-selection bias, none of them focused on long-distance commuting, which is one of my contributions here in the U.S. context.

4.3 DATA

4.3.1 Survey Data

This study analyzes data from the 2012 California Household Travel Survey (CHTS). The CHTS collected long-distance data via an optional travel recall survey that requested information about long-distance travel during the eight weeks preceding a respondent's assigned travel day. Note that only 5% of the long-distance trips in the CHTS were reported in the daily travel diaries (NuStats, 2013). A total of 68,193 long-distance trips were collected from 18,012 households, which represents 42% of all households who completed both recruitment and retrieval successfully. 41, 902 of these trips are one way long-distance trips and the rest are return trips. I do not include return trips in my analysis, because return trips have not been assigned any trip purposes in the CHTS.

The 2012 CHTS also provides detailed information on the socio-economic characteristics of individuals and households (such as income, education, and household composition), as well as the latitude and longitude of each household location. Note that in the CHTS long-distance travel is defined as journeys or trips of 50 miles or more one-way, which is also the definition employed here.

4.3.2 Long-distance Commute Trips in California

Let us now briefly examine some characteristics of long-distance commute trips collected from the long-distance travel log of the 2012 CHTS. As shown in Figure 4.1, commuting accounts for only 6.35% of reported long-distance trips, while vacation trips accounts for highest share of the long-distance trips (30.16%), which is marginally higher than visiting friends and relatives (VFR) trips (29.29%), followed by business (15.96%), personal business (10.55%), with others accounting for the rest.

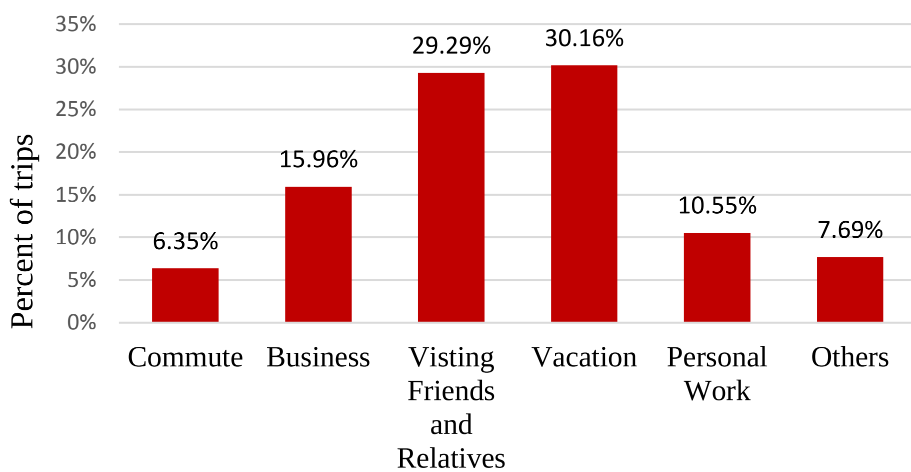


Figure 4.1 Long-distance Travel Purpose

The percent of long-distance commuting trip in the 2012 CHTS is surprisingly fewer than expected. This might be due to the under reporting issue of respondents for multiple long-distance trips to the same location via the same travel mode (mostly for commuting trips) (Bierce and Kurth, 2014). Because, the 2012 CHTS long-distance survey did not include a “repetition frequency” question which would have allowed respondents to quickly report the multiple similar trips and it was also triggered by the respondents’ fatigue combined with a lack of understanding of the need for respondents to report all long-distance travel (Bierce and Kurth, 2014). An adjustment of 2012 CHTS long-distance survey, performed by Bierce and Kurth, (2014), increased the percent of commuting trips to 16 percent of total long-distance trips.

Figure 4.2 shows the percentage of long-distance commute trips by destinations. Excluding returning home trips, 94.09% of commuters' destinations were within California and 5.34% had destination outside. Only 0.57% of reported long-distance commuting trips ended outside of the U.S.

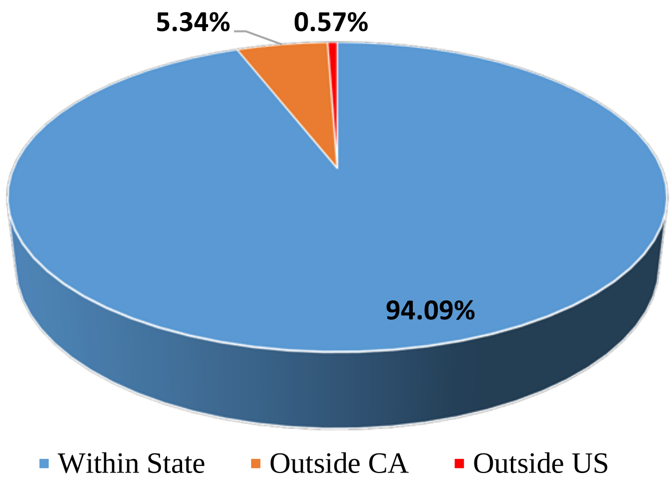


Figure 4.2 Long-distance Commute Trips by Destinations

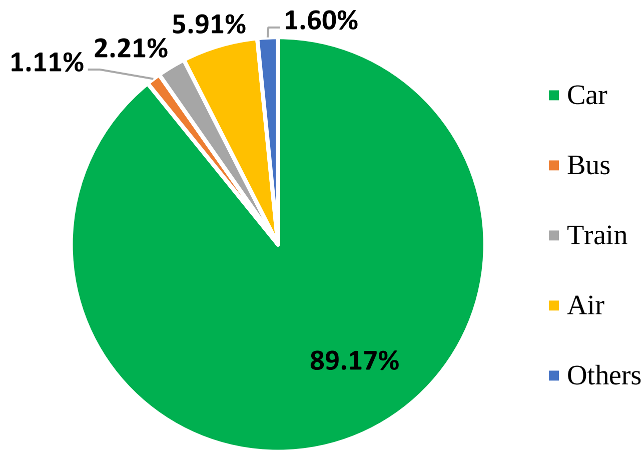


Figure 4.3 Modal Share of Long-distance Commute

A look at Figure 4.3 clearly shows that car driving dominates long-distance commuting with 89.17% of trips, followed by air (5.91%), train (2.21%), and bus (1.11%).

4.3.3 Model Explanatory Variables

Since the main objective of this study is to tease out the determinants of long-distance commuting, let us discuss the selection of variables that may influence household long-distance commuting. As shown in Table 4.3, I categorized explanatory variables into two broad categories: (1) Socio-economic and demographic characteristics, and (2) Land-use and land value related variables.

4.3.3.1 Socio-Economic and Demographic Characteristics

I include socio-economic and demographic characteristics of the household or household head in my model following previous research on long-distance travel (see Table 4.1) and published commuting behavior studies (Table 4.2). Age, gender, ethnicity, Hispanic status and highest educational attainment all refer to the household head, whereas household level variables are income, number of cars, length of residence at current address and household composition.

Age of the household head is a count variable. For household income, I used the midpoint of each CHTS category. Since the last category was open ended, I include a binary indicator for those households having income USD 250,000 or more. Length of residence is represented by categorical variables with residence more than 10 years at current address as baseline. I included the number of cars in the household in my model, because owning a car enables people to travel longer distances compared to people who must rely on slower modes (Bagley and Mokhtarian, 2002). Since the presence of children, household size, number of household workers and marital status of individual were found to be influential variables for commuting as well as long-distance travel (see Table 4.1 and Table 4.2), I created household composition categorical variables based on four household characteristics: marital status, household size, the number of household workers, and the presence of children.

I also created binary variables to indicate the highest educational attainment of the household head with graduate degree as the baseline, with the expectation that more educated people are more likely to commute long-distance (Sandow, 2008) because they tend to be more mobile than others (Eliasson *et al.*, 2003). Likewise, gender (baseline female), Hispanic status and ethnicity of household head are binary variables (with Caucasian as the baseline category for the latter).

4.3.3.2 Land-Use and Land Value related Variables

The few long-distance studies that used land use variables emphasized their importance for understanding long-distance travel. To capture the effect of land use on long-distance commuting I created a mixed density index (MDI_j), which for block group j is defined by:

$$MDI_j = \frac{ED_j \cdot RD_j}{ED_j + RD_j}, \tag{4.1}$$

where ED_j is employment density (number of workers per square mile) and RD_i is residential density (number of housing units per square mile). *MDI* is a proxy for employment accessibility and an indicator of the job-housing balance, which has been shown to be associated with lower commuting times and distances (Verhetsel and Vanelslander, 2010; Zhao *et al.*, 2011; Dai *et al.*, 2015), so I expect a higher value of *MDI* to be associated with less long-distance commuting.

Table 4.3 Descriptive Statistics of Model Variables

Variables	Mean	Std. Dev.	Min	Max
Endogenous Variables				
Long-distance Commuting				
Binary: 1= If HH is commuting long-distance	0.06	0.24	0	1
Car-Ownership				
Number of motor vehicles in the HH	2.11	0.94	0	8
Land-Use and Land Value related Variables				
Mixed density index (MDI) at the block group level	0.69	1.34	0	29.52
Median home value at census tract level (in $1,000)	459.71	249.74	11.7	1000
Exogenous Variables (Socio-Economic and Demographic Characteristics)				
Household Level				
Income				
Midpoint of annual HH income (in thousand $)	104.46	62.54	5	250
Binary: 1 = If HH income > 250 (in thousand $)	0.05	0.22	0	1
Length of residence at current address				
Binary:1 = Residence < 5 years	0.19	0.39	0	1
Binary:1 = Residence 5 to 10 years	0.19	0.40	0	1
Binary:1 = Residence > 10 years	0.62	0.49	0	1
Household composition	0.19	0.39	0	1
Binary: 1= Single worker	0.14	0.34	0	1
Binary: 1= One-worker couple	0.15	0.36	0	1
Binary: 1= Two-worker couple	0.18	0.39	0	1
Binary: 1= One-worker family	0.20	0.40	0	1
Binary: 1= Two-worker family	0.24	0.43	0	1
Binary: 1= More than two worker family	0.08	0.27	0	1
Household Head				
Age	52.07	11.92	18	94
Gender: Binary: 1= If Male and 0 otherwise	0.48	0.50	0	1
Hispanic Status: Binary: 1=Hispanic or Latino	0.13	0.34	0	1
Ethnicity				
Binary:1=Caucasian	0.83	0.38	0	1
Binary:1=African American	0.02	0.14	0	1
Binary:1= Other ethnicity	0.15	0.36	0	1
Educational Attainment				
Binary: 1= No high school degree	0.02	0.14	0	1
Binary: 1= High school graduate	0.08	0.27	0	1
Binary: 1= Some college credit but no degree	0.16	0.37	0	1
Binary: 1= Associate's degree	0.11	0.31	0	1
Binary: 1= Bachelor degree	0.32	0.47	0	1
Binary: 1= Graduate degree	0.31	0.46	0	1
Observations (N)		12,623		

Note: HH stands for household

Neo-classical urban theory suggests that household residential location choice results from a trade-off between preferred housing and commuting costs (Alsonso, 1964; Muth, 1969; Mills, 1972) where monetary and mental costs of commuting are compensated for by the housing market (Sandow and Westin, 2010). This argument is also supported by Renkow and Hower, (2000). To capture this effect, I included a census tract level median home value variable in my model with the expectation that a higher median home value is associated with a lower probability that an individual chooses to commute long-distance.

I relied on 2010 census data to measure population density at the block group level and on the 2012 American Community Survey for median home value at the census tract level. Excluding households without workers, my final dataset has 12,623 observations. Table 4.3 presents summary statistics for the variables considered in my model.

4.4 METHODOLOGY

I analyzed my data in two steps. First, I performed simple univariate analyses for different socio-economic characteristics of long-distance commuters and for the land use characteristics of their residential location. Second, I conducted multivariate analyses to tease out what factors contribute to the decision to commute long-distance.

4.4.1 Univariate Analysis

First, I analyzed the socio-economic and demographic characteristics of long-distance commuters. Then, I contrasted the land use and land value variables around the residences of long-distance commuters with the same variables for the rest of the workers in my dataset.

I used χ^2 tests (Washington *et al*., 2010) to test the statistical significance of differences between different categorical variables and long-distance commuting, since χ^2 tests are most appropriate for analyzing relationships among nominal variables (Connor-Linton, 2010).

I also used one-way analysis of variance (ANOVA) to test the statistical significance of differences in selected continuous and count variables for two groups of workers: long-distance commuters and workers who do not commute long-distance. For the one way analysis of variance, I calculated an *F* statistics from (Wabed and Tang, 2010):

$$F^* = \frac{\dfrac{\sum_{k=1}^{g} n_k \left(\overline{x}_k - N^{-1}\sum_{j=1}^{g}\sum_{i=1}^{n_j} x_{ij}\right)^2}{k-1}}{\dfrac{\sum_{j=1}^{g}\sum_{i=1}^{n_j}\left(x_{ij} - \overline{x}_j\right)^2}{N-k}}, \qquad (4.2)$$

where n_i is the number of observations in group i, g is the number of groups, $\overline{x}_i$ is the mean of group i, and $N = \sum_{k=1}^{g} n_k$ is the total number observations. Under the null hypothesis that the means of these groups do not differ, F^* has an $F(k\text{-}1, N\text{-}k)$ distribution.

4.4.2 Multivariate Analysis

In this section, I first discuss my conceptual model for addressing the relation between long-distance commuting and my explanatory variables (land use, land value and socio-economic characteristics) before outlining my structural equation model (SEM).

4.4.2.1 Conceptual Model

Figure 4.4 shows my conceptual model and the hypothesized links among variables. This model intends to address the relation between land use, land value and socio-economic characteristics, residential location, car-ownership and long-distance commuting behavior. To account for residential self-selection, land use around a residence is assumed to be influenced by the socio-economic characteristics of residents. This model also capture my initial hypothesis that residential land values influence the decision to commute long-distance, and since it is influenced by household income, it is also endogenous.

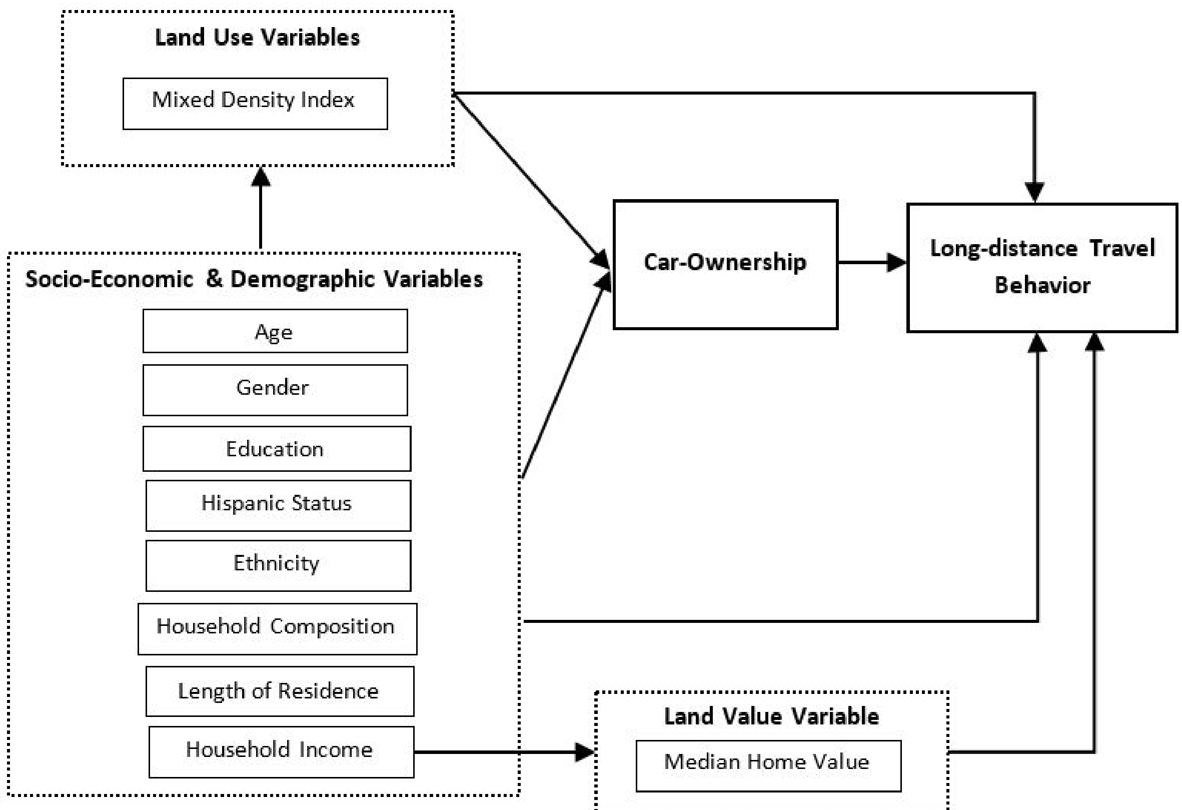


Figure 4.4 Conceptual Model

In this model I also consider car-ownership as a mediating variable between land use variables and long-distance commuting behavior because car ownership is a medium-term decision influenced by longer term decisions such as place of employment and residential locational choices (Van Acker and Witlox, 2010). Long-distance commuting behavior is then directly determined by socio-economic variables, car ownership, land use, and land value variables, while car ownership itself is influenced by land use and socio-economic variables. This results in indirect effects of socio-economic and land use variables on long-distance commuting behavior via car ownership, which serves as a mediating variable. Car ownership is therefore an outcome (or endogenous) variable in one set of relationships and a predictor (or explanatory variable) of long-distance commuting. A theoretical justification for this structure is given by Ben-Akiva and Atherton (1977) while Van Acker and Witlox (2010) discuss the consequences of ignoring this effect.

4.4.2.2 SEM Approach

I selected structural equation modeling (SEM) to estimate my conceptual model because it allows to easily parameterize endogenous relationships (Golob, 2003), which allowed me to account for self-selection effects and for car-ownership endogeneity. Moreover, extensions of SEM (i.e., generalized SEM) allow estimating SEM models with discrete (which is the case here since my model attempts to explain who will engage in long-distance commuting) and censored dependent variables (Bollen, 1989; e Silva *et al.*, 2012) in a utility theoretic framework (Golob, 2003), which can be useful for interpreting results.

SEM has been used for some time in transportation (e.g. see Aditjandra *et al.*, 2012; Wang, 2013; or Dillon *et al.*, 2015). A SEM model can have two components: a measurement component and a structural component (Kline, 2015). The measurement model defines the relationships between a latent variable and its indicators. Since all variables in my conceptual model are directly observed, following Van Acker and Witlox (2010) and Etminani-Ghasrodashti and Ardeshiri (2016), my model includes only a structural component.

The structural component is a simultaneous equation system where variables are divided into two sets: endogenous and exogenous (Golob, 2003; Kline, 2015) and where the direction of causality is explicitly specified. Exogenous variables are not caused by any other variable in the model. Instead, exogenous variables influence endogenous variables. In a graphical representation of a SEM (see Figure 4.4), no paths (symbolized by arrows) point towards exogenous variables and paths only depart from exogenous variables towards other variables. Endogenous variables are influenced by exogenous variables, either directly or indirectly through other endogenous variables (Kline, 2015).

Since my conceptual model has a multilevel data structure with generalized response variables, it can only be fitted by Generalized Structural Equation Model (GSEM) (Rabe-Hesketh *et al.*, 2004). GSEM has two features that are not available in SEM: the ability to fit models containing generalized linear response variables and the ability to fit multilevel data structures; these two features can be used separately or together (Rabe-Hesketh et al., 2004).

Since the main dependent variable of interest (long-distance commuting) of my multilevel model is a binary response variable, I use GSEM to estimate a binary logit model where dependent variable $Y_i = 1$ if a household commute long-distance commute or 0 otherwise. The probability that the dependent variable equals 1 for household *i* is given by (Rabe-Hesketh *et al.*, 2004; Skrondal and Rabe-Hesketh, 2005):

$$\Pr(Y_i = 1 | \boldsymbol{\eta_i}) = \frac{\exp(\boldsymbol{\eta_i \lambda})}{1 + \exp(\boldsymbol{\eta_i \lambda})} \tag{4.3}$$

where**:**

$$\begin{cases} \boldsymbol{\eta} = \beta_{11}\boldsymbol{c} + \beta_{12}\boldsymbol{l} + \beta_{13}\boldsymbol{v} + \Gamma_1\boldsymbol{X} + \varepsilon_1, \\ \boldsymbol{c} = \beta_{22}\boldsymbol{l} + \Gamma_2\boldsymbol{X} + \varepsilon_{2,} \\ \boldsymbol{l} = \Gamma_3\boldsymbol{X} + \varepsilon_3, \\ \boldsymbol{v} = \Gamma_4\boldsymbol{X} + \varepsilon_4, \end{cases} \tag{4.4}$$

In the above:

- $\boldsymbol{\eta}$ is an assumed latent continuous variable;
- $\boldsymbol{c}$ is an n × 1 vector of the number of cars owned by each household, where n is the sample size;
- $\boldsymbol{l}$ is an n × 1 vector of block group level residential mixed density indexes;
- $\boldsymbol{v}$ is an n × 1 vector of census tract level residential median home values;
- $\boldsymbol{x}$ is an n × p matrix of p household explanatory variables; For the third equation in Equation (4.3), I only used household income as an explanatory variable for home value ($\boldsymbol{v}$)v);
- $\boldsymbol{\Gamma_1}$, $\boldsymbol{\Gamma_2}$, $\boldsymbol{\Gamma_3}$ and $\boldsymbol{\Gamma_4}$ are n × np matrices. $\boldsymbol{\Gamma_j} = (\delta_{ji}I_n \ldots\ldots.. \delta_{jk}I_n)$ for $j \epsilon \{1, 2, 3, 4\}$, with δ_{ji} s are unknown coefficients, and I_n is the n × n identity matrix;
- $\varepsilon_1, \varepsilon_2$ ε_3 and ε_4 are n × 1 error vectors assumed to be uncorrelated; and
- $\boldsymbol{\beta_{11}}, \boldsymbol{\beta_{12}}, \boldsymbol{\beta_{13}}$, β_{21}, β_{31} $\boldsymbol{\beta_{22}}$, and δ_{ji} s ($j \epsilon \{1, 2, 3, 4\}$, $I \epsilon \{1, \ldots\ldots, p\}$, in $\boldsymbol{\Gamma_j}$) are unknown model parameters to estimate.

The four equations in Equation (4.4) represent the structural component of my model. Their structure reflects causal paths (see Figure 4.5). Here, $\boldsymbol{\eta}$, $\boldsymbol{c}$, $\boldsymbol{l}$, and $\boldsymbol{v}$ are vectors of endogenous variables and $\boldsymbol{X}$ is a matrix of exogenous socio-economic and demographic variables. In the fourth Equation of (4.4), median home value ($\boldsymbol{v}$) is explained by household income and in the third equation mixed density index ($\boldsymbol{l}$) is explained by households' socio-economic and demographic characteristics. In the second equation, car-ownership ($\boldsymbol{c}$) is explained by land use, land value and households' socio-economic and demographic characteristics. Finally, in the first equation long-distance commuting ($\boldsymbol{\eta}$) is explained by land use, land value, car-ownership and household socio-economic and demographic characteristics.

Model Estimation

To estimate unknown model parameters, GSEM (like SEM) minimizes the difference between the sample covariance and the covariance predicted by the model (Bollen, 1989). Following Bollen (1989, p.80-88), the model-replicated covariance matrix $\sum(\theta)$ can be written:

$$\sum(\theta) = \begin{bmatrix} (\mathrm{I}-\mathrm{B})^{-1}(\Gamma\,\Phi\,\Gamma' + \Psi)[(\mathrm{I}-\mathrm{B})^{-1}]' & (\mathrm{I}-\mathrm{B})^{-1}\,\Gamma\,\Phi \\ \Phi\,\Gamma'(\mathrm{I}-\mathrm{B})^{-1} & \Phi \end{bmatrix} \tag{4.5}$$

where: $\mathrm{B} = \begin{pmatrix} \beta_{11}\mathrm{I_n} & \beta_{12}\mathrm{I_n} & \beta_{13}\mathrm{I_n} \\ \mathrm{K_n} & \beta_{22}\mathrm{I_n} & \mathrm{K_n} \\ \mathrm{K_n} & \mathrm{K_n} & \mathrm{K_n} \end{pmatrix}$ and $\Gamma = \begin{pmatrix} \delta_{11}\mathrm{I}_n & \cdots & \delta_{1p}\mathrm{I}_n \\ \delta_{21}\mathrm{I}_n & \cdots & \delta_{2p}\mathrm{I}_n \\ \delta_{31}\mathrm{I}_n & \cdots & \delta_{3p}\mathrm{I}_n \end{pmatrix}$.

In the above, Φ is the covariance matrix of household explanatory variables in X; Ψ is the covariance matrix of error terms $\boldsymbol{\varepsilon_i}$, with $j \epsilon \{1, 2, 3, 4\}$; and k_n and I_n are respectively the $n \times n$ zero and identity matrices. Here, the $\boldsymbol{\varepsilon_j}$ s are uncorrelated, all causal paths are directed to $\boldsymbol{y}$, **B** is upper triangular, and Ψ is diagonal, so the model is identifiable recursively (Bollen, 1989, p.82-84).

Model Interpretation

SEM decomposes the mediating effect of car-ownership and residential self-selection on long-distance commuting behavior by estimating direct, indirect and total effects of endogenous and exogenous variables.

Direct effects quantify the impact of one variable on another variable without mediation via another variable, so here, direct effect refers to how socio-economic and demographic variables directly influence different endogenous variables (car-ownership, mixed density index, land value and long-distance commuting) as well as how other endogenous variables directly influence the decision to undertake long-distance commuting. It can be shown that for an identified SEM or GSEM (which is the case here), the direct effect of the exogenous variables and endogenous variables on the endogenous variables are $\boldsymbol{\Gamma}$ and **B** respectively (Bollen, 1989, p. 376-383).

Indirect effects are mediated by at least one other variable which can be expressed by: $(\mathrm{I}-\mathrm{B})^{-1}\Gamma-\Gamma,$ and $(\mathrm{I}-\mathrm{B})^{-1}-\mathrm{I}-\mathrm{B},$ as the indirect effects of exogenous and endogenous variables respectively.

Finally, total effects are the sum of direct and indirect effects. It should be noted that interpreting a model by using the direct effects alone could be misleading (e Silva *et al.*, 2006).

Since my dependent variable is binary, I report odds ratio because it is convenient way of interpreting results from a logit model. The odds of observing Y_i =1 versus Y_i =0 are

$$\Omega(\eta_i)=\frac{\Pr(Y_i=1\,|\,\eta_i)}{\Pr(Y_i=0\,|\,\eta_i)} \tag{4.6}$$

The odds ratio (OR) indicates the relative amount by which the odds of an outcome (long-distance commuting) increases (OR >1) or decreases (OR<1) when the value of the corresponding independent variable increases by one unit, or indicates that the corresponding independent variable does not affect (OR=1) the odds of outcome.

4.5 RESULTS AND DISCUSSION

My statistical work was performed with Stata 14. Results are presented in Table 4.4, Table 4.5, Table 4.6 and Table 4.7.

4.5.1 Univariate Results

The last column of Table 4.4 displays results of the statistical test that investigates how categorical socio-economic variables impact long-distance commuting. First, I find that there is

Table 4.4 Socio-Economic and Demographic Characteristics of Long-distance Commuters

Variables	Category	Commute Long-distance (%)	Statistical test
Gender	Male	64.46	χ2(1)=52.61***
	Female	35.54	
Race	White	80.17	χ2(2)=11.63**
	African American	3.25	
	Others	16.59	
Hispanic Status	Hispanic	13.32	χ2(1)=0.3814
	Non-Hispanic	86.68	
Age	16-24	4.42	χ2(3)=10.18**
	25-39	15.95	
	40-64	70.92	
	65+	8.71	
Educational Attainment	No high school degree	2.17	χ2(5)=23.83***
	High school graduate	11.84	
	Some college credit but no degree	15.7	
	Associate's degree	10.87	
	Bachelor degree	31.04	
	Graduate degree	28.38	
Length of Residence	Residence < 5 years	20.79	χ2(2)=5.92*
	Residence 5 to 10 years	22	
	Residence > 10 years	57.21	
HH Composition	Single worker	11.9	χ2(6)=36.91***
	One-worker couple	10.46	
	Two-worker couple	23.2	
	One-worker family	19.59	
	Two-worker family	25.84	
	More than two worker family	9.01	
HH Income (in thousand $)	< 49	15.5	χ2(3)=4.23
	50-74	18.03	
	75-99	18.27	
	> 100	48.2	
Number of HH Vehicles	No vehicle	1.08	χ2(2)=9.59**
	One vehicle	18.15	
	Two or more vehicles	80.77	

a significant relationship between gender and long-distance commuting: in line with previous studies (Hjorthol and Vågane, 2014), men (64.46%) engage more in long-distance commuting than women (35.54%). Second, results show that most of the long-distance commuters are white

(80.17%) and a significant relationship is found between ethnicity and long-distance commuting. However, there is no statistically significant relationship between long-distance commuting and Hispanic status. Third, I find that most of long-distance commuters are between 35 and 44 years of age (70.92%), possibly because many in this age group have financial commitments, resulting in more long-distance commuting.

In agreement with previous studies (Cassel *et al.*, 2013), most of the long-distance commuters have at least a bachelor's degree (bachelor: 31.04% and graduate: 28.38%). Results also show a significant relationship between long-distance commuting, length of residence, and household composition. More than 50% of long-distance commuters have lived at the same address for at least last 10 years and commuters from more than two workers family participate in long-distance commuting less than other comparable groups. Results also show that the share of households with income over $100,000 per year is almost half (48.2%) of all the long-distance commuters, although there is no statistically significant association between long-distance commuting and household income. I also note that most of the long-distance commuters have at least two vehicles in their households (80.77%).

Table 4.5 Land Use and Land Value Variables of Long-distance Commuters' Residence

Variables	**Commute Long-distance**		**Statistical test and significance**
	Yes	**No**	
Average Median Home Value (in thousand $)	409.00	473.73	F(1,11893)=51.73***
Average Mixed Density Index	0.53	0.75	F(1,11934)=10.80**

As shown in Table 4.5, land use and land value variables related to workers' place of living have a statistically significant impact on long-distance commuting. As expected, those living in areas with a lower average mixed density index participate more in long-distance commuting. Likewise, the average median home value of long-distance commuters' residential location is significantly lower than that of other workers.

4.5.2 Multivariate Results

Following Dillon *et al.* (2015), I estimated my GSEM model using quasi-maximum likelihood where the variance covariance matrix of the estimators is calculated using the Huber-White-Sandwich estimator to relax the assumption that errors are identically and normally distributed since my model includes several not-normally distributed variables.

The modelling process consisted of two phases. Following Van Acker *et al.*, (2014), during the first phase, all variables mentioned in Table 4.3 were included in the models. However, only those variables that significantly influence long-distance commuting, car ownership, and land use variables were retained in the second modelling phase during which the final models were estimated. Direct effects between variables that were insignificant were constrained to zero to eliminate their influence on the significance of indirect and total effects. Note that a significant direct effect may be weakened and result into an insignificant total effect due to mediating variables (Van Acker *et al.*, 2014).

Common fit statistics developed for SEM are not available for GSEM, because common fit statistics depend on the assumptions that observed endogenous, observed exogenous, and latent endogenous variables are jointly normally distributed, which does not hold here since my model contains a number of binary variables (Rabe-Hesketh *et al.*, 2004).

As mentioned earlier SEM decomposes the effects of one variable on another into direct, indirect and total effects. The direct effects between all the variables are discussed first, followed by the indirect and total effects. Figure 4.5 shows the final model structure and links among the variables based on the model results.

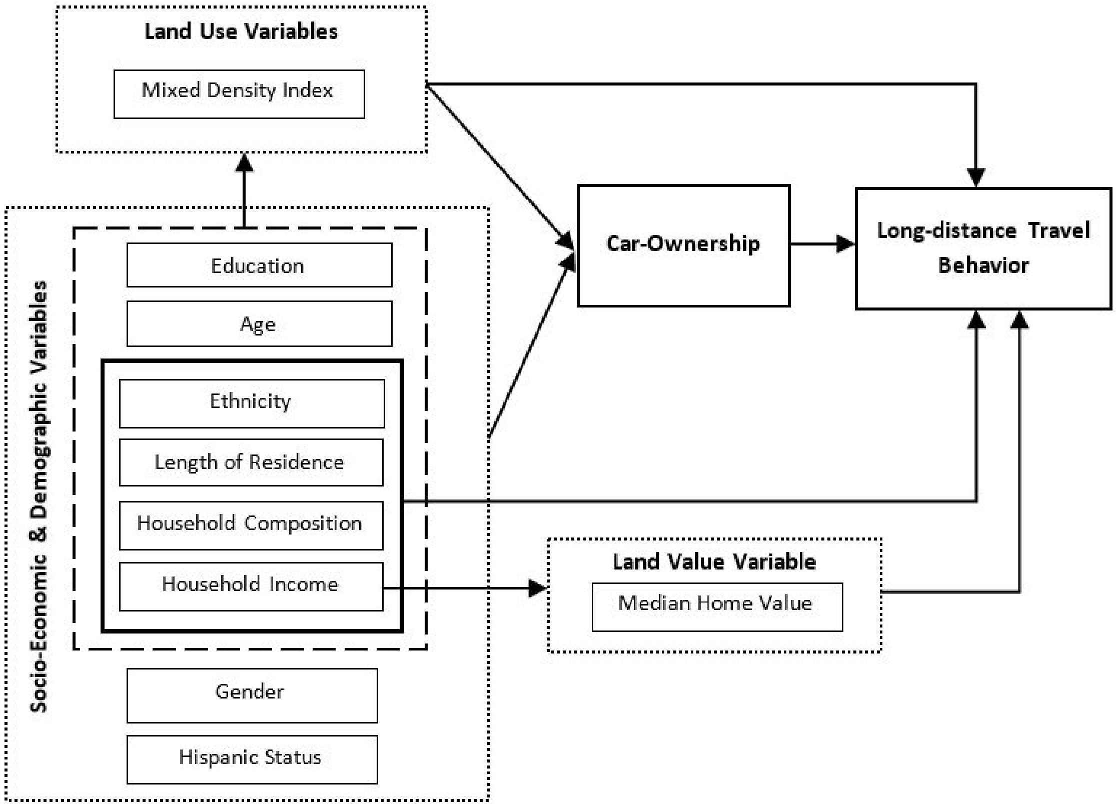


Figure 4.5 Final Model Structure

4.5.2.1 Direct Effects

Table 4.6 reports GSEM structural model coefficients, which are also direct effects. Let us first discuss the first equation of Equation (4.4) which explains long-distance commuting behavior. The direct effects on long-distance commuting behavior conform to what I expected. Households with more cars are more likely to commute long-distance (OR=1.10**) which is consistent with previous studies (Manaugh *et al.*, 2010; e Silva *et al.*, 2012). In line with my expectation, households who live in areas with a higher mixed density index are less likely to commute long-distance (OR=0.91*) which suggest that a better jobs-housing balance reduces workers' commuting distance, in agreement with previous studies (Zhao *et al.*, 2011; e Silva *et al.*, 2012; Dai *et al.*, 2015), and consistent with the co-location hypothesis according to which maintaining a jobs-housing balance shortens commuting distance in the context of suburbanized developed countries, and especially in the US (Cervero, 1989). As in Renkow and Hower (2000), median residential home value is significant (although not practically important because OR=0.99*** is close to 1), which implies that households prefer to locate further from work and have greater commuting costs in exchange for lower housing costs. This result also weakly supports neo-classical spatial-economic urban models which suggest that households determine their residential locations based on the trade-off between commuting cost and land rent (Alsonso, 1964; Muth, 1969; Mills, 1972).

Most of the socio-economic and demographic variables have the expected effect on long-distance commuting behavior. As in Sandow (2008) and Limtanakool *et al.* (2006b) households income is significant, although there is no practical effect of household income on long-distance commuting (OR=1.00***). However, this does not seem to hold for households with annual income over $250,000 (OR=0.68*), which suggests that these households are less likely to commute long-distance, possibly because they can afford houses close to their job locations or because they prefer to live in downtown areas where median home values are higher. As expected, households who have been living in their current residence for under 5 years or from 5 to 10 years are more likely to commute long-distance (OR=1.21** and OR=1.21*) compared to longer term residents. This finding echoes results of Dargay and Clark (2012), who found a significant effect of length of residence at current address on long-distance travel, with long-distance travel generally declining the longer an individual lives at the same address.

Household composition is important as one-worker couples are less likely to commute long-distance than single worker households (OR=0.73*). Ethnicity also plays a role in explaining long-distance commuting behavior. African Americans are more likely to commute long-distance (OR=1.77***) than Caucasians. Surprisingly the educational attainment of households head is not significant for long-distance commuting behavior, while previous studies found significant relationship between education and commuting (Cassel *et al.*, 2013; Motte *et al.*, 2016). One possible explanation is that the impact of education on commuting is indirect only and is manifested through residential self-selection and car-ownership. It is therefore important to explain this relationship based on total effects (I discuss them below). Results from previous studies might be influenced by the fact that they did not account for residential self-selection and car-ownership endogeneity. The same explanation may hold for age and gender of the household head, since direct effects of these variables are not significant here even though they were found to be important determinants of commuting distance in previous studies (Axisa *et al.*, 2012; Maoh and Tang, 2012).

The coefficients of most of the variables in the car-ownership equation are statistically significant and have expected sign (third column of Table 4.6 reports coefficients) and also in line with previous studies. According to direct effects, car ownership is lower among people living in areas with higher mixed density (-0.09***) as expected, which means that people who prefer to own fewer cars tend to choose denser locations. Car ownership is positively related to household income (0.002***) and length of residence (-0.18*** for residence < 5 years; -0.17*** for residence 5 to 10 years; baseline: residence > 10 years). As expected, household composition is significant, with lower car ownership for single worker compared to other households. It is also positively related to age of the household head (0.005***), being male

(0.09***), and having bachelor (0.11***) or other degree compared to having a graduate degree, possibly because graduate degree holder might be voluntary low car households which is also supported by my results in Section 3.5.2. However, car-ownership is negatively related with being Hispanic (-0.05**) and being African American (-0.15**) compared to being Caucasian.

Direct effects show that land use variables are significantly influenced by some of the socioeconomic variables (second from last column in Table 4.6 reports coefficients), thus revealing the existence of self-selection effects due to socioeconomic differences between individuals. First, we see that households with higher incomes tend to choose neighborhoods with a higher mixed density index. Although the effect is mild (0.001***), it is surprising because households with more annual income tend to live in the neighborhoods with less residential and job densities; they are more likely to reside in places with longer distance to working places (Etminani-Ghasrodashti and Ardeshiri, 2016) and this is seems to depart from their long-distance commuting behavior. However, this finding is consistent with the results of past studies which concluded that higher income people are more likely to live in central, denser, accessible (e Silva *et al.*, 2012) and more diverse (Van Acker and Witlox, 2010) neighborhoods.

Households with longer residence times tend to dwell in neighborhoods with a higher mixed density index, in accordance with their commuting behavior. Household composition seems especially important as working couples and families (-0.61*** and -0.55*** for 1 and 2 worker couples; -0.69***, -0.70***, and -0.70*** for families with 1, 2 and more than 2 workers) prefer neighborhood with a lower mixed density index compared to single workers.

When looking at the land value equation (last column of Table 4.6 reports coefficients), as expected households with higher incomes tend to choose neighborhood with higher median home values (1.58***). The magnitude of this effect is even higher for households with an annual income over $250,000 (43.86***).

4.5.2.2 Indirect and Total Effects

Table 4.7 reports indirect and total effects of socio-economic and demographic variables and endogenous variables on long-distance commuting and on car-ownership. Indirect effects refer to how socio-economic variables affect long-distance commuting through residential self-selection and car-ownership. Results suggest that focusing on direct effects only would lead to inconsistent conclusions in some cases.

For example, educational attainment of household head is not significantly associated with long-distance commuting based on direct effects only. However, long-distance commuting is influenced by educational attainment but mainly in indirectly through the interaction with car-ownership and land use variables, leading to a significant total effect. Total effects suggest that household heads with less than a graduate degree are more likely to commute long-distance (OR=1.02** for bachelor degree, OR= 1.05** for associate's degree, OR=1.04*** for some college credit but not degree and OR=1.04** for high school degree). This indicates that the long-distance commuting behavior of educated people is not necessarily caused by their education but rather by their higher car-ownership and location of residence preference which is not clearly stated in previous studies (Sandow and Westin, 2010; Cassel *et al.*, 2013). Likewise, the total effects of age and gender of the household head is significant, although there is no practical effect of these variables on long-distance commuting (in both cases, OR=1.00**).

Another example relates to the influence of household composition on long-distance commuting. Based on direct effect, I found conflicting behavior of one-worker couple between their residential preference and commuting decision. However, the total effect of one-worker couples is not significant while it is significant for two-worker couples. This results suggest that two-worker couples are more likely to commute long-distance than single worker couples (OR=1.29*) which is in line with their residential preference, because two-worker couple tend to live in neighborhoods with lower residential and job densities than their single counterparts.

Although the total effects of age and gender of the household head is significant, there is no practical effect of these variables on long-distance commuting (in both cases, OR=1.00**).

4.6 CONCLUSIONS

The purpose of this study was to examine the determinants of long-distance commuting in California. More specifically, I analyzed the joint influence of different socio-economic, land use and land value variables on the likelihood that households commute long-distance by using long-distance data from the 2012 California Household Travel Survey. So far, empirical studies on commuting and long-distance travel behavior have rarely controlled for residential self-selection bias or considered the endogeneity issue of car-ownership. To address these both issues, I estimated a Generalized Structural Equation Model (GSEM) that treats car-ownership as a mediating variable and incorporated self-selection effects due to socio-economic characteristics.

Results of my univariate analyses show that more men commute long-distance than women and most of them are white and are more highly educated. In addition, most of long-distance commuters have two or more cars in their households. Long-distance commuters also tend to live in lower mixed density areas, with lower median home value.

Results of my GSEM model show a negative relationship between mixed density index and long-distance commuting which confirms published results (Zhao *et al.*, 2011; Dai *et al.*, 2015). As in Manaugh *et al.*, 2010 and e Silva *et al.*, 2012, I found that households with more cars are more likely to commute long-distance, while car-ownership is negatively related to the mixed density index. Furthermore, the results from my land value factor weakly support the neo-classical urban theory of trade-off between greater commuting cost and housing value (Alonso, 1964; Muth, 1969; Mills, 1972). Likewise, my model confirms the effects of residential self-selection as land use and land value variables are influenced by some of the socio-economic and demographic characteristics.

Unlike some other studies (e.g. Sandow, 2008; Axisa *et al.*, 2012; Maoh and Tang, 2012; Hjorthol and Vågane 2014), my model results did not show any significant direct relationship between long-distance commuting and gender, age and educational attainment of the household head. However, total and indirect effects suggest that the long-distance commuting behavior of educated people is caused by their higher car-ownership and residential location preferences. This result agrees with e Silva *et al.* (2006), who suggest that focusing on only direct effects may lead to inconsistent conclusions. It also emphasizes the importance of controlling for residential self-section bias while considering car-ownership as a mediating variable in order to correctly assess transportation and planning polices to curb long-distance commuting and reduce car-ownership.

The empirical evidence provided in this study confirms that some current planning strategies for promoting sustainable urban transportation are moving in the right direction. For instance, the desired effect of mixed density index on long-distance commuting in my models suggest that smart growth via higher job-housing balance could help curb long-distance commuting. Although it is clearly impossible to increase the mixed density of every neighborhood (Maoh and Tang, 2012), it may be feasible to target specific centers to promote polycentrism through a higher job-housing balance, thus decreasing the need to commute long-distance in California.

This study is not without limitations. First, data restrictions prevented me from understanding the effects of attitudes/perceptions and different life events on long-distance commuting (e.g., see Ben-Elia and Ettema, 2011; Creemers *et al.*, 2012; Clark *et al.*, 2016). Second, although previous studies found a significant relationship between commuting and land uses at work locations (Manaugh *et al.*, 2010; e Silva *et al.*, 2012), I was not able to include these variables in my model due to the data restrictions. Furthermore, it would be of interest to apply the approach proposed in this study to model the determinants of long-distance commute in other US urban area to shed additional light on the key factors of long-distance commuting in the US. Finally, it would also be of interest to study the determinants of other types of long-distance travel, such business, weekend, and vacation long distance trips because they likely have different determinants (LaMondia *et al.*, 2014). All of this is left for future work.

Table 4.6 Generalized SEM Structural Model Coefficients/Direct Effects.

Variables	**Long-distance Commuting Behavior**		**Car-Ownership**	**Land-Use and Land Value related Variables**	
Exogenous ↓ | Endogenous →	***HH commute long-distance***		***Number of cars in the HH***	***MDI***	***Median home value***
	Coefficients	**OR**	**Coefficients**	**Coefficients**	**Coefficients**
Car-Ownership					
Number of cars in the HH	0.10**	1.10**	-	-	-
Land-Use and Land Value related Variables					
Mixed Density Index (MDI)	-0.10*	0.91*	-0.09***	-	-
Median home value (in thousand $)	-0.001***	0.99***	-	-	-
Socio-Economic and Demographic Characteristics					
Household Level					
Income					
Midpoint of annual HH income (in thousand $)	0.003***	1.00***	0.002***	0.001***	1.58***
Binary: 1 = If HH income > 250 (in thousand $)	-0.39*	0.68*	-0.07*	-0.007	43.86***
Length of residence (Baseline: Residence > 10 years)					
Binary:1 = Residence < 5 years	0.19**	1.21**	-0.18***	0.11***	-
Binary:1 = Residence 5 to 10 years	0.19**	1.21*	-0.17***	0.10***	-
Household composition (Baseline: Single worker)					
Binary: 1= One-worker couple	-0.32**	0.73**	0.73***	-0.61***	-
Binary: 1= Two-worker couple	0.13	1.13	0.80***	-0.55***	-
Binary: 1= One-worker family	-0.06	0.95	0.85***	-0.69***	-
Binary: 1= Two-worker family	-0.02	0.98	1.11***	-0.70***	-
Binary: 1= More than two worker family	-0.12	0.88	1.94***	-0.70***	-
Household Reference Person (Household Head)					
Age	-	-	0.005***	-0.01***	-
Gender: Binary: 1= If Male or Female	-	-	0.09***	-	-
Hispanic Status: Binary: 1=Hispanic or Latino	-	-	-0.05**	-	-
Ethnicity (Baseline: Caucasian)					

Variables	Long-distance Commuting Behavior		Car-Ownership	Land-Use and Land Value related Variables	
Exogenous ↓ | Endogenous →	***HH commute long-distance***		***Number of cars in the HH***	***MDI***	***Median home value***
	Coefficients	**OR**	**Coefficients**	**Coefficients**	**Coefficients**
Binary:1=African American	0.57***	1.77***	-0.15***	0.10	-
Binary:1= Other ethnicity	-0.05	0.95	-0.02	0.15***	-
Educational Attainment (Baseline: Graduate degree)					
Binary: 1= No high school degree	-	-	-0.11**	-0.04	-
Binary: 1= High school graduate	-	-	0.17***	-0.24***	-
Binary: 1= Some college credit but no degree	-	-	0.18***	-0.19***	-
Binary: 1= Associate’s degree	-	-	0.18***	-0.28***	-
Binary: 1= Bachelor degree	-	-	0.11***	-0.08**	-

Notes: 1. *, **, and *** indicate significance at 10%, 5%, and 1%.

Table 4.7 Generalized SEM Indirect and Total Effects.

Variables	**HH commute long-distance**				**Number of cars in the HH**	
	Indirect Effects		***Total Effects***		***Indirect Effects***	***Total Effects***
Exogenous ↓ | Endogenous →	**Coefficients**	**OR**	**Coefficients**	**OR**	**Coefficients**	**Coefficients**
Car-Ownership						
Number of cars in the HH	-	-	0.09**	1.10**	-	-
Land-Use and Land Value related Variables						
Mixed Density Index (MDI)	-0.009**	0.99**	-0.10*	0.90*	-	-0.09***
Median home value (in thousand $)	-	-	-0.001***	0.99***	-	-
Socio-Economic and Demographic Characteristics						
Household Level						
Income						
Midpoint of annual HH income (in thousand $)	-0.001***	0.99***	0.002**	1.00**	-0.0001***	0.002***
Binary: 1 = If HH income > 250 (in thousand $)	-0.05***	0.95***	-0.44**	0.65**	0.0006	-0.07*
Length of residence (Baseline: Residence > 10 years)						
Binary:1 = Residence < 5 years	-0.03***	0.97***	0.16*	1.17*	-0.01***	-0.19***
Binary:1 = Residence 5 to 10 years	-0.03***	0.97***	0.16*	1.17*	-0.01***	-0.18***
Household composition (Baseline: Single worker)						
Binary: 1= One-worker couple	0.13***	1.14***	-0.18	0.84	0.06***	0.79***
Binary: 1= Two-worker couple	0.13***	1.14***	0.26*	1.29*	0.05***	0.85***
Binary: 1= One-worker family	0.15***	1.16***	0.09	1.10	0.06***	0.91***
Binary: 1= Two-worker family	0.17***	1.19***	0.16	1.17	0.06***	1.17***
Binary: 1= More than two worker family	0.25***	1.29***	0.13	1.14	0.06***	2.01***
Household Reference Person (Household Head)						
Age	0.001**	1.00**	0.001**	1.00**	0.001***	0.01***
Gender: Binary: 1= If Male or Female	0.01**	1.00**	0.01**	1.00**	-	0.09***
Hispanic Status: Binary: 1=Hispanic or Latino	-0.004	0.99	-0.004	0.99	-	-0.05*
Ethnicity (Baseline: Caucasian)						
Binary:1=African American	-0.02**	0.98**	0.55***	1.73***	-0.009	-0.15***
Binary:1= Other ethnicity	-0.02*	0.98*	-0.07	0.93	-0.01***	-0.04*

Variables	HH commute long-distance				Number of cars in the HH	
Exogenous ↓ | Endogenous →	*Indirect Effects*		*Total Effects*		*Indirect Effects*	*Total Effects*
	Coefficients	**OR**	**Coefficients**	**OR**	**Coefficients**	**Coefficients**
Educational Attainment (Baseline: Graduate degree)						
Binary: 1= No high school degree	-0.009	0.99	-0.01	0.99	0.003	-0.11*
Binary: 1= High school graduate	0.04**	1.04**	0.04**	1.04**	0.02***	0.19***
Binary: 1= Some college credit but no degree	0.04***	1.04***	0.04***	1.04***	0.02***	0.19***
Binary: 1= Associate’s degree	0.05**	1.05**	0.05**	1.05**	0.03***	0.20***
Binary: 1= Bachelor degree	0.02**	1.02**	0.02**	1.02**	0.01**	0.12***

Notes: 1. *, **, and *** indicate significance at 10%, 5%, and 1%

4.7 REFERENCES

Alsonso, W., 1964. Location and Land Use. Toward a General Theory of Land Rent. Harvard University Press, Cambridge, MA.

Axhausen, K. W. (2001). Methodological research for a European survey of long-distance travel. In *Personal Travel: the Long and Short of It. Conference Proceedings June 28–July 1, 1999 Washington, DC* (pp. 321-342).

Axisa, J. J., Scott, D. M., & Newbold, K. B. (2012). Factors influencing commute distance: a case study of Toronto's commuter shed. *Journal of Transport Geography*, *24,* 123-129.

Aditjandra, P. T., Cao, X. J., & Mulley, C. (2012). Understanding neighbourhood design impact on travel behaviour: An application of structural equations model to a British metropolitan data. *Transportation research part A: policy and practice*, *46*(1), 22-32.

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

Ben-Akiva, M., & Atherton, T. J. (1977). Methodology for short-range travel demand predictions: Analysis of carpooling incentives. *Journal of Transport Economics and Policy*, 224-261.

Ben-Elia, E., & Ettema, D. (2011). Rewarding rush-hour avoidance: A study of commuters' travel behavior. *Transportation Research Part A: Policy and Practice*, *45*(7), 567-582.

Bergantino, A. S., & Madio, L. (2015). The Travel-to-work. Which factors matter? An analysis on regional labor market in UK. In *ERSA conference papers* (No. ersa15p888). European Regional Science Association.

Bierce, E., & Kurth, D. (2014). The Use of three surveys for long-distance travel estimates in California. In *Transportation Research Board 93rd Annual Meeting* (No. 14-5674).

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Cassel, S. H., Macuchova, Z., Rudholm, N., & Rydell, A. (2013). Willingness to commute long-distance among job seekers in Dalarna, Sweden. *Journal of Transport Geography*, *28*, 49-55.

Cervero, R. (1989). Jobs-housing balancing and regional mobility. *Journal of the American Planning Association*, *55*(2), 136-150.

Cervero, R., & Duncan, M. (2006). 'Which reduces vehicle travel more: Jobs-housing balance or retail-housing mixing? *Journal of the American planning association*, *72*(4), 475-490.

Clark, B., Chatterjee, K., & Melia, S. (2016). Changes to commute mode: The role of life events, spatial context and environmental attitude. *Transportation Research Part A: Policy and Practice*, *89*, 89-105.

Creemers, L., Cools, M., Tormans, H., Lateur, P. J., Janssens, D., & Wets, G. (2012). Identifying the determinants of light rail mode choice for medium-and long-distance trips: results from a stated preference study. *Transportation Research Record: Journal of the Transportation Research Board*, (2275), 30-38.

Dai, D., Zhou, C., & Ye, C. (2015). Spatial-temporal characteristics and factors influencing commuting activities of middle-class residents in Guangzhou City, China. *Chinese Geographical Science*, 1-19.

Dargay, J. M., & Clark, S. (2012). The determinants of long-distance travel in Great Britain. *Transportation Research Part A: Policy and Practice*, *46*(3), 576-587.

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.

Eliasson, K., Lindgren, U., & Westerlund, O. (2003). Geographical labour mobility: migration or commuting?. *Regional Studies*, *37*(8), 827-837.

e Silva, J., Golob, T., & Goulias, K. (2006). 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), 121-131.

e Silva, J. D. A., Morency, C., & Goulias, K. G. (2012). Using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal. *Transportation Research Part A: Policy and Practice*, *46*(8), 1252-1264.

Elldér, E. (2014). Commuting choices and residential built environments in Sweden, 1990–2010: a multilevel analysis. *Urban Geography*, *35*(5), 715-734.

Erhardt, G., Freedman, J., Stryker, A., Fujioka, H., & Anderson, R. (2007). Ohio long-distance travel model. *Transportation Research Record: Journal of the Transportation Research Board*.

Etminani-Ghasrodashti, R., & Ardeshiri, M. (2016). The impacts of built environment on home-based work and non-work trips: An empirical study from Iran. *Transportation Research Part A: Policy and Practice*, *85*, 196-207.

Federal Highway Administration, (2006). U.S. Department of Transportation, NPTS Brief, Retrieved from http://nhts.ornl.gov/briefs/Long%20Distance%20Travel.pdf

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

Greene, W. H. (2008). *Econometric Analysis*. 6th edition, Prentice Hall, Englewood Cliffs.

Helminen, V., & Ristimäki, M. (2007). Relationships between commuting distance, frequency and telework in Finland. *Journal of Transport Geography*, *15*(5), 331-342.

Hjorthol, R., & Vågane, L. (2014). Allocation of tasks, arrangement of working hours and commuting in different Norwegian households. *Journal of Transport Geography*, *35,* 75-83.

Holz-Rau, C., Scheiner, J., & Sicks, K. (2014). Travel distances in daily travel and long-distance travel: What role is played by urban form? *Environment and Planning A*, *46*(2), 488-507.

Jin, X., & Horowitz, A. (2008). Time-of-day choice modeling for long-distance trips. *Transportation Research Record: Journal of the Transportation Research Board*, (2076), 200-208.

Kawabata, M., & Shen, Q. (2007). Commuting inequality between cars and public transit: The case of the San Francisco Bay Area, 1990-2000. *Urban Studies*, *44*(9), 1759-1780.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy*, *99*(3), 483-499.

LaMondia, J. J., & Bhat, C. R. (2011). A conceptual and methodological framework of daily and long-distance leisure activity-travel behavior. In *Transportation Research Board 90th Annual Meeting* (No. 11-2867).

LaMondia, J., Aultman-Hall, L., & Greene, E. (2014). Long-Distance work and leisure travel frequencies: ordered probit analysis across non-distance based definitions. *Transportation Research Record: Journal of the Transportation Research Board*, (2413), 1-12.

LaMondia, J. J., Fagnant, D. J., Qu, H., Barrett, J., & Kockelman, K. (2016). Long-distance travel mode-shifts due to automated vehicles: A statewide mode-shift simulation experiment and travel survey analysis. In *Transportation Research Board 95th Annual Meeting* (No. 16-3905).

Limtanakool, N., Dijst, M., & Schwanen, T. (2006a). The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium-and longer-distance trips. *Journal of transport geography*, *14*(5), 327-341.

Limtanakool, N., Dijst, M., & Schwanen, T. (2006b). On the participation in medium-and long-distance travel: A decomposition analysis for the UK and the Netherlands. *Tijdschrift voor economische en sociale geografie*, *97*(4), 389-404.

Lucas, K. (2011). Making the connections between transport disadvantage and the social exclusion of low income populations in the Tshwane Region of South Africa. *Journal of Transport Geography*, *19*(6), 1320-1334.

Manaugh, K., Miranda-Moreno, L. F., & El-Geneidy, A. M. (2010). The effect of neighbourhood characteristics, accessibility, home–work location, and demographics on commuting distances. *Transportation*, *37*(4), 627-646.

Maoh, H., & Tang, Z. (2012). Determinants of normal and extreme commute distance in a sprawled midsize Canadian city: Evidence from Windsor, Canada. *Journal of Transport Geography*, *25*, 50-57.

Marion, B., & Horner, M. (2008). Comparison of socioeconomic and demographic profiles of extreme commuters in several US metropolitan statistical areas. *Transportation Research Record: Journal of the Transportation Research Board*.

Mills, E.S., 1972. Studies in the Structure of the Urban Economy. John Hopkins University Press, Baltimore, MD.

Modarres, A. (2013). Commuting and energy consumption: toward an equitable transportation policy. *Journal of Transport Geography*, *33*, 240-249.

Monzon, A., & Rodríguez-Dapena, A. (2006). Choice of mode of transport for long-distance trips: Solving the problem of sparse data. *Transportation Research Part A: Policy and Practice*, *40*(7), 587-601.

Motte, B., Aguilera, A., Bonin, O., & Nassi, C. D. (2016). Commuting patterns in the metropolitan region of Rio de Janeiro. What differences between formal and informal jobs?. *Journal of Transport Geography*, *51*, 59-69.

Muth, R.H., 1969. Cities and Housing. University of Chicago, Chicago.

Neto, R. S., Duarte, G., & Páez, A. (2014). Gender and commuting time in São Paulo Metropolitan Region. *Urban Studies*, *52*(2), 298-313.

NuStats. (2013). 2010-2012 California Travel Survey final report, http://www.dot.ca.gov/hq/tpp/offices/omsp/statewide_travel_analysis/files/CHTS_Final_Report_June_2013.pdf

Oakil, A. T. M., Nijland, L., & Dijst, M. (2015). Rush hour commuting in the Netherlands: Gender-specific household activities and personal attitudes towards responsibility sharing. *Travel Behaviour and Society*.

Owen, A., & Levinson, D. M. (2015). Modeling the commute mode share of transit using continuous accessibility to jobs. *Transportation Research Part A: Policy and Practice*, *74*, 110-122.

Outwater, M. L., Bradley, M., Ferdous, N., Bhat, C., Pendyala, R., Hess, S., & LaMondia, J. (2015). Tour-based national model system to forecast long-distance passenger travel in the United States. In *Transportation Research Board 94th Annual Meeting* (No. 15-4322).

Plaut, P. O. (2006). The intra-household choices regarding commuting and housing. *Transportation Research Part A: Policy and Practice*, *40*(7), 561-571.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*(2), 167-190.

Reichert, A., & Holz-Rau, C. (2015). Mode use in long-distance travel. *Journal of Transport and Land Use*, *8*(2).

Renkow, M., & Hoover, D. (2000). Commuting, Migration, and Rural-Urban Population Dynamics. *Journal of regional science*, *40*(2), 261-287.

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S., & Tsang, F. (2013). Modelling long-distance travel in the UK. *2013 Transportation Research Board Annual Meeting*.

Sandow, E. (2008). Commuting behaviour in sparsely populated areas: Evidence from northern Sweden. *Journal of Transport Geography*, *16*(1), 14-27.

Sandow, E., & Westin, K. (2010). The persevering commuter–duration of long-distance commuting. *Transportation Research Part A: Policy and Practice*, *44*(6), 433-445.

Schiller, P. L., Bruun, E., & Kenworthy, J. R. (2010). *An Introduction to Sustainable Transportation: Policy, Planning and Implementation*. Routledge.

Shearmur, R. (2006). Travel from home: An economic geography of commuting distances in Montreal. *Urban Geography*, *27*(4), 330-359.

Sivaraman, V., Pinjari, A. R., & Polzin, S. E. (2016). Long-distance travel for visiting friends and relatives (VFR) and leisure purposes empirical analysis and modeling framework. In *Transportation Research Board 95th Annual Meeting* (No. 16-6751).

Skrondal, A., & Rabe-Hesketh, S. (2005). *Structural equation modeling: categorical variables*. John Wiley & Sons, Ltd.

Surprenant-Legault, J., Patterson, Z., & El-Geneidy, A. M. (2013). Commuting trade-offs and distance reduction in two-worker households. *Transportation Research Part A: Policy and Practice*, *51*, 12-28.

Van Acker, V., & Witlox, F. (2010). Car ownership as a mediating variable in car travel behavior research using a structural equation modelling approach to identify its dual relationship. *Journal of Transport Geography*, *18*(1), 65-74.

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.

Van Nostrand, C., Sivaraman, V., & Pinjari, A. R. (2013). Analysis of long-distance vacation travel demand in the United States: A multiple discrete–continuous choice framework. *Transportation*, *40*(1), 151-171.

Van Ommeren, J., & Dargay, J. (2006). The optimal choice of commuting speed: Consequences for commuting time, distance and costs. *Journal of Transport Economics and Policy*, 279-296.

Verhetsel, A., & Vanelslander, T. (2010). What location policy can bring to sustainable commuting: an empirical study in Brussels and Flanders, Belgium? *Journal of Transport Geography*, *18*(6), 691-701.

Wang, D., & Chai, Y. (2009). The jobs–housing relationship and commuting in Beijing, China: the legacy of Danwei. *Journal of Transport Geography*, *17*(1), 30-38.

Wang, F. (2001). Explaining intraurban variations of commuting by job proximity and workers' characteristics. *Environment and Planning B: Planning and Design*, *28*(2), 169-182.

Wang, K. (2013). Causality Between Built Environment and Travel Behavior: Structural Equations Model Applied to Southern California. *Transportation Research Record: Journal of the Transportation Research Board*, (2397), 80-88.

Watts, M. J. (2009). The impact of spatial imbalance and socioeconomic characteristics on average distance commuted in the Sydney metropolitan area. *Urban Studies*, *46*(2), 317-339.

Zhao, P., Lü, B., & De Roo, G. (2011). Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. *Journal of transport geography*, *19*(1), 59-69.

Chapter 5. Summary and Conclusions

In this dissertation, I presented three essays on accessibility, carless households, and long-distance travel, which address important gaps in transportation-land use interaction research.

In Chapter 2, I provided empirical evidence about the impact of transportation accessibility on land value in a South Asian developing country city. This is a unique case study since most published studies focus on housing markets in developed countries, which typically differ from those of South Asian Countries. In addition, few published papers rely on the most recent spatial econometric techniques that address the risk of maximum likelihood estimation in the presence of heteroscedasticity. To address this gap, I estimated a spatial autoregressive model with spatial-autoregressive disturbances (SARAR) using generalized spatial two-stage least squares (GS2SLS) by analyzing a dataset collected via in-person interview in Rajshahi City, Bangladesh.

Results of my model indicate that the rent of a multi-unit dwelling decreases by 0.0239% for every 1% increase in network access distance to the nearest major road in Rajshahi City. Moreover, proximity (within 400 m) to a primary school and to a healthcare facility commands rent premiums of respectively 93.55 BDT ($1.40) and 109.45 BDT ($1.64). Surprisingly, whether access roads are paved or not does not statistically impact rents, probably because of the dominance of walking, rickshaws use, and biking, combined with the rarity of personal cars. Likewise, proximity to bus stops and to train stations is not reflected in rents of multi-family dwellings, likely because buses and trains in Rajshahi City only provide regional and national service. Differences in estimates of my spatial models between maximum likelihood (ML) and generalized spatial two-stage-least-squares illustrate the danger of relying on ML in the presence of heteroscedasticity and also suggests to revisit results obtained to-date on the value of transportation accessibility. These results should be useful for planning transportation infrastructure funding measures in least developed country cities like Rajshahi City.

In this study, data limitations precluded me from using gravity-based accessibility measures for capturing the impact on rents of Rajshahi's sub-centers, which is unfortunate because underestimating labor market accessibility effects on rents could help quantify willingness-to-pay for transportation infrastructure improvements. This might be of interest for future research. It would also be of interest to examine the potential for value capture in larger cities of Bangladesh as well as in other South Asian cities, both for residential and commercial properties.

In Chapter 3, I characterize carless households in California based on the 2012 California Household Travel Survey (CHTS) using simple statistical tests and logit models. More specifically, I assess the effects of various socio-economic, life-cycle stage, and land use variables on the likelihood that a household is carless, voluntarily or not. Understanding why some households decided to voluntarily forgo cars could inform policies to reduce our dependency on cars and the resulting greenhouse gas emissions but understanding the plight of households who do not have access to cars is no less important as these households are at greater risk of social exclusion.

My results show that voluntary carless households are more likely to have a higher household income, a better education, a higher number of employed members, and a lower number of children than their involuntary counterparts. Results of my binary logit models show the importance of land use diversity (via a land use entropy index) and of good transit service to help households voluntarily forgo their vehicles and they downplay the impact of population density and pedestrian-friendly facilities. Results from this study should help planners and policy makers formulate land use measures that could encourage households to live voluntarily without cars and thus help reducing automobile dependency.

Since a household's decision to forgo cars is likely connected to the type of neighborhood it wants to live in, constructing a joint model of residential urban form and car-ownership (focusing on carless households) might be of interest for future research. In this study, I defined carless households based on an indirect question asked in the CHTS. It would have been preferable to directly ask households if they chose to live voluntarily without motor vehicles or not. In addition, it would be of interest to develop integrated models that analyze biographical interdependences related to residential and car ownership choices with a focus on carless households.

In my third essay (Chapter 4), I examine the determinants of long-distance commuting in California. In particular, I analyze the joint influence of different socio-economic, land use and land value variables on the likelihood that households commute long-distance by analyzing long-distance data from the 2012 California Household Travel Survey. Thus far, long-distance commuting has only received limited attention from transport researchers despite of its importance. Moreover, empirical studies on commuting and long-distance travel behavior have rarely controlled for residential self-selection or considered the endogeneity of car-ownership. To fill this gap, I estimated a Generalized Structural Equation Model (GSEM) that considers car-ownership as a mediating variable and incorporates residential self-selection.

Results of my univariate analyses show that more men commute long-distance than women and most of them are white and have more education. In addition, most long-distance commuters have two or more cars in their households. Results of my GSEM model show that long-distance commuting is negatively associated with mixed density and with residential home values, but positively related with car-ownership. My results also confirm the presence of residential self-selection as land use and land value variables are influenced by some of the socio-economic and demographic characteristics. Total and indirect effects illustrate the perils of ignoring the residential self-selection bias and car-ownership endogeneity. They also suggest that focusing on only direct effects may lead to inconsistent conclusions. The empirical evidence of this study provides some justifications for existing land use planning strategies to curb long-distance commuting and thus help promote more sustainable urban transportation.

In future, it would be of interest to further evaluate the influence of land use variables on long-distance commuting by incorporating land uses at work locations in the model. Another direction of future research might be to analyze the effects attitudes/perceptions and different life events on long-distance commuting. Furthermore, it would be of interest to apply the approach proposed in this study to model the determinants of long-distance commute in other US urban area to shed additional light on the key factors of long-distance commuting in the US.