Title
Commute Mode Choice, Parking Policies, and Social Influence

Author
Khrodagui, Nagwa

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Commute Mode Choice, Parking Policies, and Social Influence

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Nagwa Khordagui

Dissertation Committee:
Professor Jan Brueckner, Chair
Professor David Brownstone
Professor Matthew Freedman

2019

DEDICATION

To my parents, Elsayeda and Hosny,
and to my husband, Moustafa,

for their continuous and unfaltering
support, love, encouragement and sacrifice.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my dissertation committee members, Professors Jan Brueckner, David Brownstone, and Matthew Freedman, for their invaluable guidance and assistance. This dissertation would not have been possible without their support and insights.

I am grateful to Professor Jan Brueckner for his continuous support and mentoring. His courses spurred my interest in the field and provided me with a strong foundation. His discussions always inspired me to advance my thinking, and his meticulous reading and constructive feedback pushed me to keep improving my work and polishing my writing. I am grateful to Professor David Brownstone for introducing me to discrete choice analysis, and offering valuable advice on commute mode choice and on how to advance my work. I owe a lot of what I know about road transportation to him. I thank Matthew Freedman for his discussions of my research in general and identification strategies in particular. His conversations pushed me to think harder about my inferences.

I would also like to thank Kevin Roth for his guidance, and for emphasizing writing a convincing motivation and focusing on the economic question. Special thanks to Sofia Franco, the co-author of Chapter 2, for providing me with unique data for that chapter, great insight, and long hours of discussion. Thanks to Dhari Aljutaili for welcoming the many conversations and offering helpful feedback and editing.

I would like to acknowledge funding provided by the UCI Department of Economics, the UCI School of Social Sciences, and the University of California Transportation Center.

I am grateful to my family for their unwavering support and continuous encouragement. This would not have been possible without my mother's dedication and patience. I also thank my father and husband for their faith in me and their sacrifices.

CURRICULUM VITAE

Nagwa Khordagui

EDUCATION

PhD, Economics, **University of California, Irvine** 2019
MA, Economics, **University of California, Irvine** 2015
MA, Financial Economics, **American University of Beirut** 2005
BA, Economics, *summa cum laude*, **American University of Beirut** 2003
(MINOR: Computer Science)

RESEARCH INTERESTS

Transportation Economics, Urban Economics, Public Economics, Environmental Economics, Health Economics

WORKING PAPERS

"Determinants of Public Transit Use and the Role of Social Influence"
"Parking Prices and the Decision to Drive" *(Revise & Resubmit* at *Transportation Research Part A)*
"Commute Mode Choices and the Role of Parking Prices, Parking Availability and Urban Form: Evidence from Los Angeles County" (in collaboration with *Sofia Franco*)

WORK IN PROGRESS

"Pollution and Child Mortality: Evidence from Egypt"
"Does Crime Affect Commute Mode Choice?"

PROFESSIONAL EXPERIENCE

University of California, Irvine (UCI)

Teaching Assistant 2014 – 2019

Arab Academy for Science and Technology (AAST), Alexandria, Egypt

Instructor and Teaching Assistant, 2008 – 2011
College of Management and Technology

Adjunct Co-Instructor, *MBA program* 2010 – 2011

Adjunct Instructor and Teaching Assistant, 2005 – 2011
London School of Economics (LSE) External Program

American University of Beirut (AUB), Beirut, Lebanon

Graduate Assistant 2003 – 2005

TECHNICAL SKILLS

Stata, R, Matlab, ArcGIS, QGIS, Eviews, Python, LaTeX

TEACHING EXPERIENCE

UCI - TEACHING ASSISTANT: Intermediate Economics II, Intermediate Economics III, Corporate Governance, Money and Banking, Economics of Accounting, Probability and Statistics

LSE External Program - INSTRUCTOR: Econometrics

LSE External Program - TEACHING ASSISTANT: Introduction to Economics, Intermediate Microeconomics, Intermediate Macroeconomics, International Economics

AAST - INSTRUCTOR: Principles of Microeconomics, Principles of Macroeconomics, Economic Analysis (Managerial Economics), International Business

AAST - CO-INSTRUCTOR: Contemporary Management (*graduate course*)

AAST - TEACHING ASSISTANT: International Trade, Economic Analysis (Managerial Economics), Principles of Microeconomics

CONSULTING & RESEARCH PROJECTS

Compiling a database of non-tariff trade barriers in Egypt - World Bank Project	2010
Estimating non-tariff trade barriers in Egypt - World Bank Project	2010
Estimating the demand for the AAST	2009
Preparing a proposal for a marine feasibility study for the Roads & Transportation Authority of the Government of Dubai, UAE	2008
Financial Development and Economic Growth in Egypt – *Masters Thesis*	2005

AWARDS & GRANTS

Economics Department Block Grant, UCI	2018
Economics Merit Fellowship, UCI	2018
Economics Summer Research Fellowship, UCI	2018
Middle East Economics Association Travel Grant	2018
School of Social Sciences Travel Grant, UCI	2017
Economics Summer Research Fellowship, UCI	2017
Economics Department Block Grant, UCI	2017
International Transportation Economics Association Summer School Scholarship	2017
University of California Transportation Center Summer Fellowship, UCI	2016
Economics Summer Research Fellowship, UCI	2016
Social Sciences Teaching Assistantship, UCI	2014 - 2019
Social Sciences Non-Resident Tuition Fellowship, UCI	2014 - 2015
Graduate Assistantship, AUB	2003 - 2005

WORKSHOPS & CERTIFICATES

Computable General Equilibrium Economic Modeling Workshop, North American Regional Sciences Conference, San Antonio, TX	2018
International Transportation Economics Association Summer School, Barcelona, Spain	2017
Berkeley Summer School in Environmental & Energy Economics, Berkeley, CA	2016
ArcGIS web course	2016

CONFERENCES & PRESENTATIONS

Environmental Protection Agency, National Center for Environmental Economics	2019
US Census Bureau	2019
MITRE Corporation	2019
Federal Communication Commission	2019
University of Wisconsin Whitewater	2019
US Treasury, Office of Economic Policy	2019
Analysis Group	2019
California State University, Long Beach	2018
65^{th} North American Regional Sciences Conference (*Presenter, Discussant, Chair*)	2018
Graduate Poster Presentation, Economics Department, UCI	2018
Transportation, Urban and Environmental Seminar, Economics Department, UCI	2018
38^{th} Annual Meeting of the Middle East Economic Association (*Presenter* and *Discussant*)	2018
Graduate Poster Presentation, Economics Department, UCI	2017
92^{nd} Western Economic Association International (*Presenter, Discussant, Chair*)	2017
International Transportation Economics Association Conference	2017
Transportation, Urban and Environmental Seminar, Economics Department, UCI	2017
Graduate Poster Presentation, Economics Department, UCI	2016
Emerging Research Paradigms in Business and Social Sciences	2013
11^{th} Middle East Economic Association International Conference	2012

LANGUAGES

English (proficient), Arabic (proficient)

ABSTRACT OF THE DISSERTATION

Commute Mode Choice, Parking Policies, and Social Influence

By

Nagwa Khordagui

Doctor of Philosophy in Economics

University of California, Irvine, 2019

Professor Jan Brueckner, Chair

This dissertation examines the impact of parking policies and social influence on commute mode choice using discrete choice analysis. A key feature of the dissertation is overcoming the problem of insufficient data by using unique datasets, building unique datasets, or exploring appropriate estimation strategies and assumptions.

Chapter 1 studies the impact of parking prices on the decision to drive to work using the California Household Travel Survey. The chapter tackles estimation challenges posed by insufficient parking information. The first challenge is the estimation of parking prices for those who do not drive, which is addressed by using a sample selection model. The second challenge is to understand the effect of the extent of the prevalence of Employer-Paid parking coupled with incentive programs offered in-lieu of parking. To address this challenge, two extreme scenarios are examined, and a range for the marginal effects of parking prices is estimated; one scenario assumes everyone receives Employer-Paid parking coupled with in-lieu of parking incentives, and the second assumes that no one is offered such incentives. The results suggest that higher parking prices reduce driving, regardless of the followed approach. It is estimated that a 10% increase in parking prices leads to a 1 - 2 percentage point decline in the probability of driving to work. Moreover, there seems to be no evidence of

sample selection bias. The evidence suggests that parking pricing can indeed be an effective transportation demand management tool.

Chapter 2 extends the analysis of Chapter 1 to simultaneously estimate the impact of parking pricing, parking availability, and urban form on commute mode choice. The joint role of these three factors is examined using a dataset that is constructed by merging three major different data sources. The California Household Travel Survey data are matched to two unique datasets on parking for Los Angeles County; one for prices and the other availability. Chapter 2 first examines how these three factors affect the binary decision of whether to drive, while controlling for a rich set of covariates. The analysis then becomes more specific and examines how these factors affect particular commute modes in a multinomial context. The results indicate that parking prices have a significant negative impact on the decision to drive to work, where a 10% increase in parking prices is associated with a 1.1% drop in the probability of driving to work. Both on-street and off-street parking availability at home, as well as urban form measures of the workplace tract, are found to significantly affect commute mode choices. These findings have important policy implications in terms of minimum parking requirements, maximum parking standards, employer-paid parking, and parking pricing policies.

Chapter 3, on the other hand, examines the impact of a number of fundamental determinants of commute mode choice on transit use, and introduces the role of social influence. The determinants explored cover socioeconomic characteristics, built environment and neighborhood characteristics, transit accessibility, and trip characteristics. Social interactions have been found to affect many of the decisions of economic agents, and are likely to play a role in the decision to use transit. A unique dataset is built to conduct this analysis across a number of major US cities and examine the effects in both the residence and workplace neighborhoods, where a neighborhood is defined as a census tract. Social influence is explored along three different dimensions: space (neighborhood), income, and race. A novel instrumental variable

is constructed in order to identify spatial social influence, and an alternative identification strategy is devised to identify income-group and racial social influence. The evidence suggests that spatial social influence exists among both coworkers and residential neighbors, and that peer effects among coworkers are larger than those among residential neighbors. Moreover, income-group social influence, among both coworkers and residential neighbors, plays a significant role in the rich commuter's decision to use transit. However, racial social influence does not affect a commuter's decision to use transit, regardless of race.

Chapter 1

Parking Prices and the Decision to Drive to Work: Evidence from California

1.1 Introduction

Parking in the United States, in many cases, is either underpriced or provided free of charge, which leads to several market distortions. The cost of provision of parking, comprised of land, construction, and maintenance costs, varies widely from place to place, and from one form to another (surface, underground, structure, etc.), but is nevertheless a substantial figure. For example, Cutter et al. (2016) estimate construction costs to be over $5,000 for a surface parking space in Los Angeles, and over $36,000 for an underground space. By offering free or underpriced parking, the incidence of the cost of parking is shifted from the users to the rest of society (Shoup, 2005). Parking costs are passed on to everyone through higher commodity or service prices (in the case of shopping centers), lower salaries (in the case of employer-paid parking), and higher taxes (in the case of on-street and publicly-provided parking). Moreover, free and underpriced parking in many cases translate to a classic case of excess demand for parking spots, leading to what is known in the literature as cruising, which contributes significantly to traffic congestion, and has been estimated to represent about 30% of downtown congestion (Shoup, 2005). Furthermore, the availability of free and underpriced parking distorts people's travel mode choices, encouraging them to drive instead of considering other available alternatives. As a result, free and underpriced parking amplify the negative externalities associated with vehicle usage; namely, traffic congestion and pollution.

The purpose of this paper is to empirically examine whether parking prices can significantly influence the decision to drive alone to work. There is growing evidence that parking price influences travel mode choices, and thus can be used as an effective tool for travel demand management. The empirical evidence, however, is not well-developed, and although there are extensive empirical travel mode choice studies, only a few examine the impact of parking prices. Prior empirical studies find consistent evidence that higher parking prices lead to less driving. Willson (1992) finds that driving to work drops significantly in downtown Los Angeles when commuters have to pay for parking, while Brownstone and Golob (1992) find that parking incentives to rideshare reduce driving alone in Southern California. Miller and Everett (1982) find that, in a sample of 15 worksites in Washington DC, driving alone drops after the elimination of existing parking subsidies. Other studies use stated preferences to investigate the relationship in Sydney (Hensher and King, 2001) and Vancouver (Washbrook et al., 2006), and some use revealed preferences to investigate the relationship in Toronto (Gillen, 1977), Portland (Peng et al., 1996; Hess, 2001), and Seattle (Su and Zhou, 2012).

Lack of parking data is a major hindrance to investigating the impact of parking prices on travel mode choice. This paper relies on reported parking data from the 2012 California Household Travel Survey (CHTS); a cross-sectional dataset covering all of California, thus offering an opportunity to address the question at a state level, which has not been done before. The focus of the study is on work commutes, which allows the examination of recurrent behavior and ensures that the commuter is aware of the parking price he will face at the destination. Moreover, the study focuses on the impact of parking prices on the decision to drive versus not driving, and not on the various travel mode choices. The main objective of the paper is to establish that parking prices can reduce vehicle usage, regardless of the alternative modes chosen. If so, then parking prices can be used as a policy tool to reduce the negative externalities associated with vehicle usage as long as there is not a large shift to ride-sharing services like Uber or Lyft.

The nature of the dataset gives rise to questions that were not necessarily relevant in previous studies. Parking prices are only reported for those who drive, but unobserved for those who do not. To address this problem, different econometric techniques are used to impute parking prices for non-drivers. This study thus offers a way to estimate parking prices, noting that whether parking prices are observed is conditional on not driving, which may give rise to sample selection bias. Therefore, a sample selection model is estimated, and there seems to be no evidence of sample selection bias. Alternatively, parking prices are imputed by averaging the reported parking prices in the neighborhood.

Another issue arises from the lack of information on whether the commuter is offered employer-paid (EP) parking, and if so, whether any incentives are offered in lieu of free parking. There is evidence that EP parking encourages vehicle usage on work commute trips. EP parking is widespread in the United States. Brueckner and Franco (2018) develop a theoretical model that shows how road usage increases as a result of EP parking, and several case studies investigate the impact of EP parking on driving to work. For example, Willson and Shoup (1990) provide a survey of five case studies that examine the impact of EP parking on the commute mode choice. The evidence consistently suggests that a greater share of employees drive alone when offered EP parking. The fact that whether commuters receive EP parking is unobservable does not in itself pose a problem for the paper's empirical strategy. Those who receive EP parking are typically either reporting parking for free at work or do not report on parking prices at all. Since EP parking is perceived by commuters as free parking, it is reasonable to assume they face a cost of $0 for parking.

On the other hand, the fact that whether a commuter is offered an incentive in-lieu of parking is unobservable represents a challenge to the current empirical investigation. Incentives, such as the cash-out program, translate into a parking opportunity cost. Indeed, evidence suggests that commuters alter their commute mode choice in response to such incentives. Shoup (1997a) assesses the effectiveness of California's cash-out parking law, which requires employers to provide employees with the option to receive cash in lieu of free parking. The findings for eight case studies in Southern California indicate that, on average, driving alone to work dropped from 76% to 63%, whereas the other commute modes all rose. Ignoring the arising parking opportunity cost because of lack of information in the CHTS dataset could potentially lead to an upward bias of any estimates of the impact of parking prices on the decision to drive. This study, therefore, extends the analysis to explore the relationship in question under the assumption that all commuters are offered EP parking as well as an alternative compensation in-lieu of free parking, i.e., there is always an opportunity cost of parking. This is then compared to the results obtained under an initial set of assumptions of absence of in-lieu of parking incentives, i.e., absence of an opportunity cost of parking for those who do not report paying for parking.

Therefore, this paper makes a number of contributions to the literature. First, it empirically estimates the effects of parking prices on the decision to drive to work in the state of California using revealed preferences data, and improves our understanding of the effectiveness of parking prices for travel demand management. Second, it addresses the issue of missing parking prices conditional on not driving and finds that sample selection bias is not a concern. Third, the paper overcomes the issue of lack of information on the prevalence of cash-out programs by estimating parking prices for two extreme situations and then estimating the impact of parking prices on the decision to drive, producing a range for the effect. Fourth, a large number of covariates are controlled for and are constructed by combining data from various sources using Geographic Information System (GIS) tools and Google Maps, exploiting the availability of home and work locations.

It is important to note that this paper does not make any assumptions on the prevalence of EP or in-lieu of parking incentives. The findings indicate a statistically significant effect of parking prices on the decision to drive, regardless of the adopted approach, and in any scenario. The results are robust to the inclusion of various socioeconomic and trip-related variables. Average marginal effects of parking prices are in the range of -0.1 to -0.2, which implies that an increase of parking prices by 10% leads to a 1 - 2 percentage point decrease in the probability of driving.

A key assumption of this paper is that commuters distinguish between parking prices and other monetary costs of the commute trip. A trip-related monetary cost is defined as the cost incurred to travel from the origin to the destination, whereas the parking price is an additional cost that is independent of the trip itself and any trip-related costs, and arises only if there is a need to park a vehicle, i.e., if the commute mode is driving alone or carpooling. Traditional models, on the other hand, typically aggregate all monetary costs. However, the assessment of the impact of parking prices using such a model would only be valid under the unrealistic assumptions that parking cost represents a fixed proportion of the total monetary cost, or that the impact of parking costs is identical to that of the other monetary trip-related costs (Feeney, 1989). In fact, Feeney (1989) lists a number of studies that model parking prices distinctly and argues that a separate measure for parking prices is needed in order to assess the impact of the parking on travel mode choice. Moreover, Gillen (1977) finds that the coefficients of parking prices and other monetary costs are significantly different. Ideally, one would estimate both models and compare the results. However, due to the nature of the dataset as well as the fact that it is cross-sectional, it is not possible to adequately account for trip-related monetary costs, and so the impact of parking prices is modeled separately.

The remainder of the paper is organized as follows. Section 1.2 describes the construction of the sample data used while section 1.3 introduces the basic model. Section 1.4 describes the estimation methods used under the assumption that parking price is known for all drivers but missing for all non-drivers, i.e., commuters not reporting paying for parking face $0 in terms of opportunity cost. This section includes the Sample Selection model estimation as well as an alternative average price estimation. Results for these estimations are presented and briefly discussed. Section 1.5 describes the estimation method under the assumption that parking price is missing for a lot of drivers in addition to all non-drivers, i.e., commuters not reporting paying for parking face an opportunity cost. Again, results are presented and briefly discussed. Section 1.6 brings all the findings together and discuss the main implications, whereas Section 1.7 concludes.

1.2 Data

The 2012 CHTS dataset is obtained from the National Renewable Energy Laboratory (NREL). The dataset includes information on work and home locations, trip purpose, mode of transportation, whether parking prices are paid and how much is paid, parking location type, among other things. The dataset does not offer any information on whether EP parking is provided, nor on whether a cash-out option is offered (or any other incentive packages discouraging driving alone to work). Moreover, no data are collected on the price a non-driver would face had he been driving. The location of each destination is available at the census tract level in the publicly available dataset; however, exact locations (geocoded addresses) are available through restricted access to the NREL secured data portal.

Table 1.1: Summary Statistics

	Full Sample	**Drive Alone**	**Alternative Mode**
Observations	6,793	5,194	1,599
Work Counties	26		
Work Zipcodes	216		
Pay for Parking	3.40%		
Socioeconomic Characteristics			
Household Income ($)	115,503	114,704	118,113
	(72,558)	(72,296)	(73,372)
Household Size	3	3	3
	(1)	(1)	(1)
Average Income ($)	43,475	43,023	44,960
	(35,811)	(35,709)	(36,118)
Age	45	46	44
	(16)	(15)	(16)
Female	46%	46%	46%
High School	97%	97%	95%
College Degree	62%	59%	69%
White	74%	74%	72%
Trip Characteristics			
Distance (miles)	15.61	15.48	16.01
	(15.77)	(15.52)	(16.58)
Driving Time (mins)	23	23	23
	(16)	(16)	(15)
Transit time (mins)	35	36	32
	(21)	(21)	(21)
Transit Time Cost (ratio)	3.19	3.38	2.59
	(4.2)	(4.41)	(3.34)
Transit-Work Distance (miles)	0.31	0.35	0.19
	(0.43)	(0.45)	(0.3)
Transit-Home Distance (miles)	0.54	0.58	0.4
	(0.52)	(0.54)	(0.44)
Peak Traffic Hours	54%	53%	59%
Chain	50%	50%	48%
Neighborhood Characteristics			
Density of Work Tract (person/sq mile)	8,200	7,141	11,682
	(9,427)	(8,103)	(12,227)
Density of Home Tract (person/sq mile)	8,533	7,683	11,324
	(8,839)	(7,773)	(11,235)

Standard errors in parentheses

1.2.1 Sample

The sample is composed of 6,793 work trips to 216 zipcode areas. Table 1.1 provides summary statistics for the main variables used. The analysis is carried out at the zipcode level. Households that do not have a vehicle are eliminated since driving alone is not a viable option. Destination zipcodes where no one pays for parking, whether for work or non-work purposes, are excluded. The sample coverage is thus reduced to 26 counties from a total of 58. A list of the counties included in the study can be found in the appendix. Restricting the sample to work commutes permits the analysis of recurrent frequent behavior rather than occasional trips where individuals may or may not know of parking prices in advance. Furthermore, non-work trips could also be subject to venue-choice modifications.

1.2.2 Variables

Commute Mode

The variable of interest is a binary variable indicating whether the worker drives alone to work. Therefore, a value of 0 includes transit, carpooling, walking, cycling, and passengers. Over 76% of commuters drive alone.

Parking

Overall, 3.4% of those who drive alone report paying for parking. All of those who park at the destination location either report paying nothing for parking or do not report at all, and those who report paying are either parking on-street, or at an off-site parking facility. Figure 1.1 shows the share of those who pay for parking in each county, which is computed as the ratio of those who drive and pay to those who drive and do not pay (or do not report on paying). Table 1.2 presents summary statistics for parking prices. The first 4 rows present summary statistics for parking prices in the sample of interest; namely, commuters.

The last 2 rows, on the other hand, present summary statistics for parking prices in the sample made up of both commuters and non-commuters traveling to the 216 zipcodes of the sample. The table shows statistics for parking prices including zero prices, i.e., accounting for free parking, and also parking prices excluding zero prices, i.e., including paid parking only (non-zero prices).

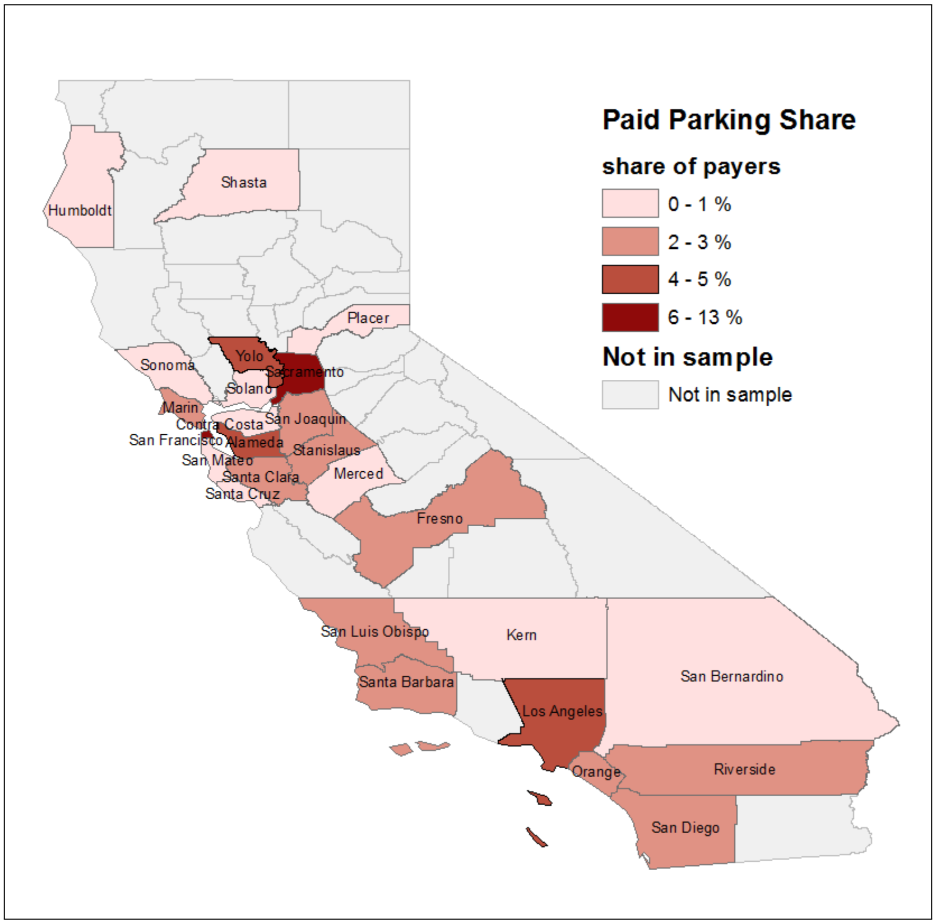

Figure 1.1: Share of Drivers who Pay for Parking

Socioeconomic Characteristics

Several socioeconomic variables are included in the model as controls. To control for income, the household income is divided by the household size to obtain the average income of a household member. (Household income and size are included separately as robustness

Table 1.2: Parking Prices Summary Statistics

	Mean	**Observations**
Parking Prices Among Commuters		
Parking Price (including 0 prices)	0.27	5,194
	(2.67)	
Non-zero Parking Prices	7.8	177
	(12.29)	
Zipcode Average Parking Price (including 0 prices)	0.46	6,793
	(2.5)	
Zipcode Average Non-zero Parking Price	5.7	3,764
	(8.82)	
Parking Prices Among All Drivers in Sample Areas		
Zipcode Average Parking Price (including 0 prices)	0.42	6,793
	(2.42)	
Zipcode Average Non-zero Parking Price	7.83	6,793
	(7.54)	

Standard errors in parentheses

checks, and the results do not change). All other variables are obtained directly from the dataset.

Trip Characteristics

The Euclidean distance between home and work locations is constructed using the geocoded addresses in NREL. This measure is independent of commute mode choice. Another important aspect of commute mode choice is the relative time cost of the different modes. To account for time cost, the ratio of transit time to driving time is constructed. Transit time and driving time are constructed by using Google Maps . The time costs are constructed for the departure time reported by each commuter on a given weekday. The only drawback of using Google Maps is that the time costs constructed are for the year 2017, i.e., 5 years after the CHTS dataset. However, using Google Maps is the best option for constructing such variables and would be a reasonable estimate of the average time cost of going from one neighborhood to another.

Almost 20% of commuters stop somewhere else on the way to work for one reason or another. A binary variable indicating such chain trips is created, taking the value of 1 if there is at least one stop on the way to work. Peak traffic time is also controlled for and is included as a binary variable that takes the value 1 if the commuter leaves home to work during peak traffic hours, and 0 otherwise. Peak traffic hours are defined here as weekdays 5.00 - 9.00 am and 3.00 – 7.00 pm.

Neighborhood Characteristics

Population density of both the census tract of the work location and that of home location are controlled for, since there is evidence that people are more likely to drive to less dense neighborhoods and from less dense neighborhoods. The density measures used are population per square mile. Another variable of interest is the distance to the nearest transit stop and is constructed using GIS tools for everyone whether they take transit or not. In fact, two variables are constructed: one to indicate the distance between home and the nearest transit station, and another to indicate the distance between work and the nearest transit station.

These variables are constructed using the geocoded locations of both individuals and transit stations. The exact home and destination locations of individuals are available in the NREL secured portal. The exact locations of the transit stations are obtained from several sources; some are obtained as GIS data from the corresponding transit system or county portals, others are obtained from a compilation of geocoded addresses of transit stations, and yet others are digitized from maps using GIS techniques. A list of the source of the data can be found in the appendix. GIS techniques are then used to find the distance to the nearest transit station for each home and work location.

1.3 Empirical Model

The decision to drive is modeled as a function of parking cost along with a large set of control variables:

$$D_i^* = \beta_P \ln(P_i^*) + \boldsymbol{X_i \beta_x} + \boldsymbol{Y_i \beta_y} + u_i \tag{1.1}$$

$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0 \\ 0 & \text{if } D_i^* \leq 0 \end{cases} \tag{1.2}$$

where D_i^* represents a commuter's unobservable propensity to drive, D_i is a binary variable that takes on the value of 1 if the commuter drives and 0 otherwise, P_i^* represents the parking price charged at the destination, $\ln(P_i^*)$ represents the natural logarithm of the parking price (adding 1 to avoid loss of 0 prices), X_i is a vector of socioeconomic variables, and Y_i a vector of trip-related variables and neighborhood characteristics. Y_i includes variables such as trip distance, distance to nearest transit station, transit time cost ratio, census tract population density, time of day, and whether the trip is part of a trip-chain, among others.

The difficulties that arise for this estimation are the result of both missing prices and missing information on EP parking and in-lieu of parking incentives. Since only 177 commuters report paying for parking, parking prices have to be estimated for the remaining commuters. The estimation procedure for parking prices will depend on the assumptions made about EP parking and in-lieu of parking incentives, because such assumptions will dictate whether free parking is truly free or whether there is a hidden opportunity cost that must be accounted for. Therefore, the analysis is divided into two parts. The first part of the analysis, which will be referred to as the "Free Parking Scenario", is carried out under the assumption that drivers who do not report parking prices indeed face a price of \$0. The second part, which will be referred to as the "Parking Opportunity Cost Scenario", is carried out under the assumption that drivers who do not report parking prices all receive both EP parking and in-lieu of parking incentives, and hence, face a non-zero opportunity cost of parking.

A potential source of bias in the estimation of β_P is that commuters planning to drive to work may be choosing jobs located in areas with no or low parking prices, or may be choosing jobs that offer free parking. Although in principle it is possible that drivers may be choosing jobs based on parking prices, this paper assumes that in practice it is highly unlikely that parking prices play a significant role in job choice. In fact, it is also assumed that, unlike residential neighborhood sorting phenomena, sorting into particular workplace neighborhoods is highly unlikely because the neighborhood is defined at the zipcode level and workers are unlikely to be able to choose jobs, and hence sort into workplace neighborhoods, based on the zipcode. Although the model includes a wide array of covariates, it certainly does not include all possible confounding variables. Therefore, a potential source of bias emerges from the possibility of unobserved heterogeneity, and this is not accounted for in this paper.

1.4 Free Parking Scenario

The major assumption for this part of the analysis is that drivers who do not report paying for parking are parking for free. Therefore, parking price is always known for drivers. On the other hand, individuals who do not drive do not report any parking price and hence parking price is a missing variable for that group of commuters. One approach to estimate the parking prices these non-drivers would face if driving is to compute the average parking price in the destination zipcode and assign it to them. However, parking price is not missing at random, but is rather missing conditional on not driving. That is, parking price is always missing when the commuter is not driving to work, and is known only when the commuter drives to work. In this case imputing missing values is not straightforward. Therefore, the following situation holds:

$$\ln(P_i) = \begin{cases} \ln(P_i^*) & \text{if } D_i = 1 \\ - & \text{if } D_i = 0 \end{cases} \tag{1.3}$$

where P_i is observed parking price (and P_i^* actual parking prices)[1]. Figure 1.2 illustrates the situations.

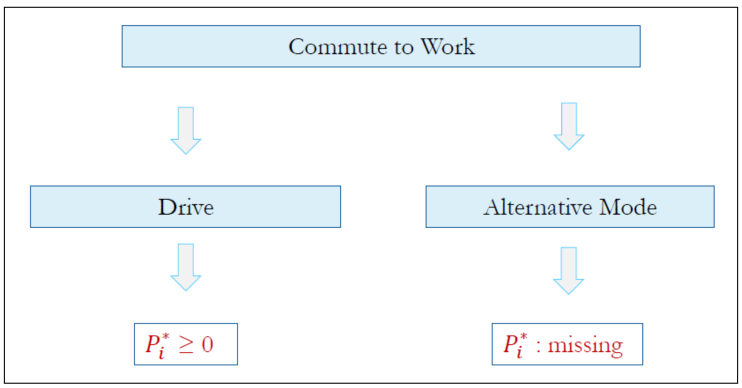

Figure 1.2: Free Parking Scenario

The approach followed, thus, is to model the selection and estimate parking prices given that they are observed conditional on driving to work alone. Therefore, this approach is a Bivariate Selection Model that can be modeled as a Type 2 Tobit model (Amemiya, 1985), where equation 1.2 is the participation equation and equation 1.3 is the outcome equation. The approach is discussed further below.

[1]The distribution of actual parking prices is not observed in its entirety; hence the distinction between P_i and P_i^*.

1.4.1 Selection Model

The standard Type 2 Tobit model is one where the censored outcome variable (parking prices - $\ln(P_i)$) is observed conditional on a decision represented by a binary variable (the decision to drive - D_i). There is additional complexity in this paper's model, though, arising from the fact that the decision to drive itself is dependent on the censored outcome variable, parking prices (as modeled in equation 1.1), and so parking prices are endogenously determined in such a case. This issue has been addressed by Lee (1978, 1979). Details on how this works can be found in the appendix.

Interestingly, the Likelihood Ratio test of independent equations in the Selection Model indicates that the null hypothesis of independent equations cannot be rejected, which suggests that modeling parking prices as a Type 2 Tobit model is not necessarily advantageous, and hence, selection bias is not a valid concern for the problem at hand. Nevertheless, the estimated parking prices resulting from the Selection Model are used to estimate equation 1.1, and the results are presented in Table 1.3.

Average marginal effects are calculated by computing marginal effects for each commuter and then averaging over all commuters. The estimates for the parking price marginal effect beyond specification (2) decrease in magnitude gradually as more confounding variables are controlled for, but remain negative and significant. Adding household characteristics does not seem to impact the marginal effect of parking prices, but adding trip-related costs such as distance and transit time cost along with distance from the work location to the nearest transit station does. The more drastic drop in the magnitude of the estimate of the parking price marginal effect takes place when county fixed effects are added. Based on the results of specification (5), a 10% increase in parking prices results in a 0.74 percentage point drop in the probability of driving to work. Further discussion will follow in section 1.6.

Table 1.3: Model Estimation (Selection Model)

Specification	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
ln(Price)	-0.194***	-0.197***	-0.165***	-0.150***	-0.074***
	(0.038)	(0.038)	(0.033)	(0.031)	(0.015)
ln(Income)		0.005	0.004	0.003	0.021***
		(0.01)	(0.009)	(0.009)	(0.008)
Age		0.001***	0.001***	0.001***	0.001***
		(0)	(0)	(0)	(0)
Female		-0.001	0.01	0.015	0.002
		(0.011)	(0.011)	(0.011)	(0.01)
High School		0.130***	0.108***	0.109***	0.090***
		(0.029)	(0.028)	(0.029)	(0.028)
College Degree		-0.082***	-0.067***	-0.065***	-0.043***
		(0.016)	(0.015)	(0.015)	(0.014)
White		0.019	0.012	-0.003	-0.005
		(0.014)	(0.013)	(0.013)	(0.012)
ln(Distance)			0.016**	0.016**	0.022***
			(0.007)	(0.008)	(0.006)
Mode Time Ratio			0.001**	0.001**	0.001*
			(0)	(0)	(0)
Transit-Work Distance			0.005***	0.004***	0.005***
			(0.002)	(0.002)	(0.002)
Transit-Home Distance			0.003***	0.001	0
			(0.001)	(0.001)	(0.001)
ln(Density at Work)				-0.025***	-0.003
				(0.009)	(0.007)
ln(Density at Home)				-0.022***	-0.016***
				(0.007)	(0.006)
Peak				-0.046***	-0.043***
				(0.011)	(0.01)
Chain				0.017	0.034**
				(0.015)	(0.015)
County Fixed Effects	No	No	No	No	Yes
Observations	6628	6050	5650	5503	5503
Pseudo R^2	0.0334	0.0447	0.0658	0.0823	0.1549

Average Marginal Effects are reported
Average Marginal Effect of price has been adjusted to reflect that of price itself not 1+price
Standard errors are reported in parentheses, and are clustered at the zipcode level
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$
Estimations include a constant

1.4.2 Alternative Approach: Average Parking Prices

Another approach would be to ignore potential selection bias and assign non-drivers the average parking price computed based on the parking prices of drivers in the corresponding zipcode area. This approach is reasonable since the Likelihood Ratio test for independent equations from the Selection Model failed to reject the null hypothesis of independent equations and, hence, no selection bias exists. Since under this approach the assumption is that all drivers not reporting paying for parking face no opportunity cost (do not receive any cash-out offers), and hence are facing a parking price of \$0, then non-drivers are expected to face the same probability of facing \$0 because of the absence of selection bias. Therefore, the expected value of parking prices for non-drivers is the weighted average of all the free parking as well as the paid parking. The average parking price is then used in the estimation of equation 1.1.

The zipcode average parking price is computed using parking prices reported by all those who drive to the destination zipcode regardless of the trip purpose. The alternative is to compute the zipcode average parking price using parking prices reported by commuters only. The advantage of including all drivers regardless of trip purpose is that there are over 900 reported (positive) parking prices in such a case, relative to only 177 reported (positive) parking prices by commuters. Basing the computations on the larger pool of reported parking prices is likely to result in a more reliable estimate. The only advantage of using the 177 parking prices reported by commuters is that such an estimate would more likely reflect the parking prices that are faced by commuters, which is likely to be lower than those faced by other drivers due to possible monthly discounts or simply more familiarity with parking options and their price ranges. Results using average parking prices based on commuters only are also reported, but only for the preferred specification, for comparison and robustness purposes.

Table 1.4: Parking Price = Average Value in Zipcode

	All Drivers					Commuters
Specification	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
ln(Price)	-0.303***	-0.296***	-0.275***	-0.265***	-0.204***	-0.176***
	(0.022)	(0.022)	(0.021)	(0.021)	(0.017)	(0.013)
ln(Income)		0.016**	0.012	0.012	0.023***	0.024***
		(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
Age		0.001***	0.001**	0.001***	0.001***	0.001***
		0.000	0.000	0.000	0.000	0.000
Female		-0.001	0.009	0.012	0.004	0.002
		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
High School		0.112***	0.098***	0.098***	0.086***	0.086***
		(0.026)	(0.025)	(0.026)	(0.026)	(0.026)
College Degree		-0.057***	-0.046***	-0.046***	-0.032***	-0.037***
		(0.013)	(0.013)	(0.013)	(0.012)	(0.013)
White		0.012	0.011	0	-0.001	-0.002
		(0.012)	(0.012)	(0.012)	(0.011)	(0.012)
ln(Distance)			0.023***	0.024***	0.026***	0.025***
			(0.006)	(0.007)	(0.006)	(0.006)
Mode Time Ratio			0.001**	0.001**	0.001*	0.001*
			0.000	0.000	0.000	0.000
Transit-Work Distance			0.003***	0.003**	0.004***	0.004***
			(0.001)	(0.001)	(0.001)	(0.001)
Transit-Home Distance			0.002**	0.001	-0.001	-0.001
			(0.001)	(0.001)	(0.001)	(0.001)
ln(Density at Work)				-0.01	0	0.001
				(0.006)	(0.006)	(0.006)
ln(Density at Home)				-0.019***	-0.017***	-0.016***
				(0.006)	(0.006)	(0.006)
Peak				-0.042***	-0.041***	-0.041***
				(0.011)	(0.010)	(0.010)
Chain				0.013	0.025*	0.028**
				(0.014)	(0.014)	(0.014)
County Fixed Effects	No	No	No	No	Yes	Yes
Observations	6793	6050	5650	5503	5503	5503
Pseudo R^2	0.1442	0.1473	0.1572	0.1684	0.2049	0.1949

Average Marginal Effects are reported
Average Marginal Effect of price has been adjusted to reflect that of price itself not 1+price
Standard errors are reported in parentheses, and are clustered at the zipcode level
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$
Estimations include a constant

Results are reported in Table 1.4. The specifications are identical to the ones described for Table 1.3. Specification (6) corresponds to the average parking prices based on commuters only. The results are very much in line with those in Table 1.3. Marginal effects of parking prices are consistently negative and significant at the 1% significance level.

1.5 Parking Opportunity Cost Scenario

The second major issue is the unobservability of whether EP parking is offered, and if so, whether any incentive to give it up is being offered as well. Many employers provide free parking to their employees as a fringe benefit, which would typically come out of all employees' wages. Nevertheless, as long as the employee is not offered the choice between free parking and equivalent cash, parking will be perceived as free due to the lack of incentive to change the commute mode, i.e., the opportunity cost of parking is zero.

On the other hand, if an employee is offered some sort of incentive to change his commute mode, then the resulting opportunity cost must be regarded as the price of parking. Incentives could take the form of transit subsidies, carpooling subsidies, or cash-out programs where the employee is offered the choice to receive the cash equivalent of the parking price (or the portion of parking price paid by the employer if not fully subsidized) in lieu of free parking. If incentives are provided, then the value of the incentive is the opportunity cost of parking and thus should be set as the parking price. Because no information on incentives is available in the dataset, it is virtually impossible to distinguish between the following 3 groups of drivers:

- Drivers who are not provided parking by employers: the observed parking price is correct.

- Drivers who have EP parking but are not offered an incentive: the observed parking price is correct, and the parking price should be set to zero.
- Drivers who have EP parking and are also offered an incentive: parking price should be set to the value of the incentive foregone.

This is further illustrated in Figure 1.3. For this part of the analysis, the opportunity cost of free parking is assumed to be equal to the prevalent parking price in the neighborhood. All those who drive and do not report paying for parking are assumed to have been offered both EP parking and incentive programs. Similarly, non-drivers are also assumed to have been offered incentive programs. This implies that whether parking prices are observed is no longer conditional on driving, and all missing parking prices and all free parking are equivalent. This is an extreme case since in practice only some commuters will be offered such incentives and not everyone as assumed here. An average parking price is computed for each zipcode area, where only positive parking prices are included in the computation, and is then used in the estimation of the model presented by equations 1.1 and 1.2.

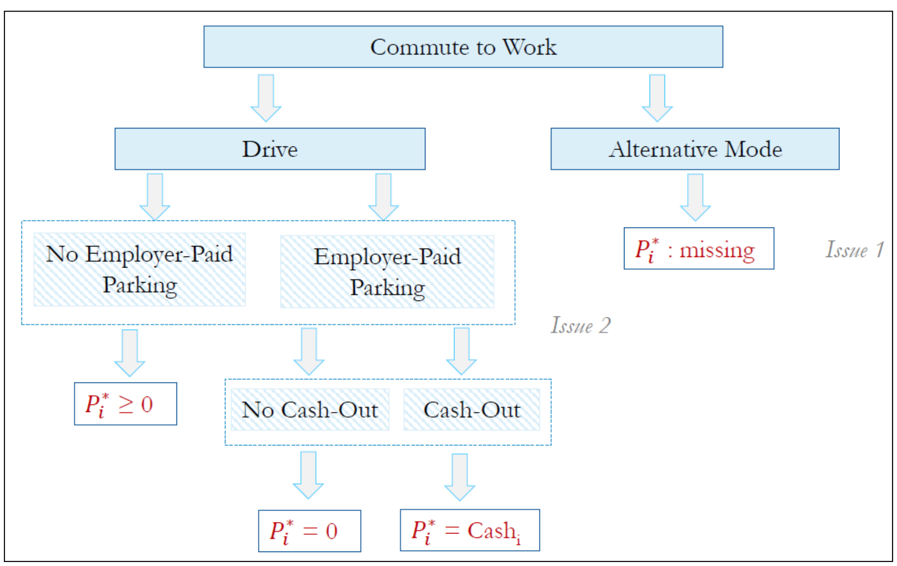

Figure 1.3: Parking Opportunity Cost Scenario

The difference between the average parking price computed in this section and the one computed in the previous section is that here only positive parking prices are averaged and $0 parking prices are dropped from the computation. This method corresponds to the assumption that no one enjoys free parking and everyone is facing a parking opportunity cost, and so for those who do not report paying for parking and those who do not drive, the expected value of the parking price they are likely to face is the average of that actually paid in the destination zipcode.

The zipcode average price is again computed twice, once based on the commuter parking price only, and another time based on prices paid by anyone parking at the destinations of interest regardless of the trip purpose. Using the latter computation becomes even more advantageous in this part of the analysis because of the structure of the sample. Doing so enables us to compute positive zipcode average parking prices, although commuters are not observed paying for parking there, which increases the sample size. Table 1.5 presents the results.

The results correspond to average parking prices based on the pool of all drivers, and only specification (6) correspond to parking prices based on commuters only. Results are still in line with the previous ones, revealing a negative significant impact of parking prices on the decision to drive to work. A 10% increase in parking prices results in a 1 percentage point reduction in the probability of driving to work. The marginal effect of parking prices is much lower in absolute value than the estimate obtained using average prices in the first part of the analysis. Further discussion of the results is presented in the following section.

Table 1.5: Parking Price = Average Non-Zero Price

			All Drivers			Commuters
Specification	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
ln(Price)	-0.143***	-0.139***	-0.127***	-0.122***	-0.104***	-0.140***
	(0.023)	(0.022)	(0.019)	(0.018)	(0.011)	(0.016)
ln(Income)		0.007	0.005	0.005	0.021***	0.016
		(0.010)	(0.009)	(0.009)	(0.008)	(0.010)
Age		0.001***	0.001***	0.002***	0.001***	0.001**
		0.000	0.000	0.000	0.000	0.000
Female		-0.016	-0.004	0.001	-0.006	-0.006
		(0.010)	(0.010)	(0.011)	(0.010)	(0.014)
High School		0.119***	0.099***	0.100***	0.092***	0.04
		(0.029)	(0.028)	(0.029)	(0.028)	(0.042)
College Degree		-0.083***	-0.070***	-0.066***	-0.044***	-0.042**
		(0.015)	(0.014)	(0.015)	(0.013)	(0.020)
White		0.01	0.007	-0.007	-0.009	-0.014
		(0.013)	(0.013)	(0.012)	(0.011)	(0.015)
ln(Distance)			0.019***	0.019**	0.025***	0.027***
			(0.007)	(0.008)	(0.006)	(0.009)
Mode Time Ratio			0.001**	0.001**	0.001**	0
			0.000	0.000	0.000	0.000
Transit-Work Distance			0.005***	0.003*	0.005***	0.004
			(0.002)	(0.002)	(0.002)	(0.003)
Transit-Home Distance			0.002**	0.001	-0.001	0
			(0.001)	(0.001)	(0.001)	(0.001)
ln(Density at Work)				-0.024***	-0.005	-0.002
				(0.009)	(0.007)	(0.009)
ln(Density at Home)				-0.016**	-0.013***	-0.008
				(0.007)	(0.005)	(0.007)
Peak				-0.044***	-0.041***	-0.034***
				(0.010)	(0.010)	(0.011)
Chain				0.022	0.036**	0.024
				(0.014)	(0.014)	(0.019)
County Fixed Effects	No	No	No	No	Yes	Yes
Observations	6793	6050	5650	5503	5503	2970
Pseudo R^2	0.0825	0.0924	0.1104	0.1255	0.1904	0.2611

Average Marginal Effects are reported
Average Marginal Effect of price has been adjusted to reflect that of price itself not 1+price
Standard errors are reported in parentheses, and are clustered at the zipcode level
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$
Estimations include a constant

1.6 Discussion

Overall, parking prices have a significant impact on the decision to drive, and the results are robust to the inclusion of various control variables and their variations and to the estimation under various assumptions. The two estimation approaches presented in sections 1.4 and 1.5 present extreme scenarios, and therefore the true impact of parking prices is expected to lie within the range produced. The model estimated under the "Free Parking Scenario" correctly predicts the decision to drive 82.4% of the time whereas the model under the "Parking Opportunity Cost Scenario" does so 81.3% of the time.

The average marginal effects for the preferred specification (Specification (5)) for the "Free Parking Scenario" and the "Parking Opportunity Cost Scenario" are -0.2 and -0.1 respectively. The findings suggest that increasing the parking price by 10% leads to a decrease of the probability to drive to work somewhere between 1 and 2 percentage points. The estimates indicate the extent to which parking prices can be used as an effective transportation demand management tool. The marginal effects vary with prices and the relationship is not a linear one. Figures 1.4 and 1.5 depict the relationship between marginal effects and prices under the different scenarios. The figures illustrate how effective raising parking prices are depending on the initial prices. For example, a 10% increase in parking prices results in a drop in the probability of driving by 1.1 – 2.8 percentage points if the parking prices are initially \$5, and a drop by 1.2 – 2.1 percentage points if the parking prices are initially \$10.

Elimination of free parking or raising the prices of underpriced parking thus have the potential to reduce vehicle usage on commute trips. The findings also suggest that offering incentives to give up EP parking, such as cash-out programs, would also reduce driving alone to work. An elimination or a reduction of EP parking would reduce driving alone to work as well. However, offering cash-out alternatives and the elimination of EP parking will have the anticipated effect only if parking in the workplace neighborhood is not free of charge.

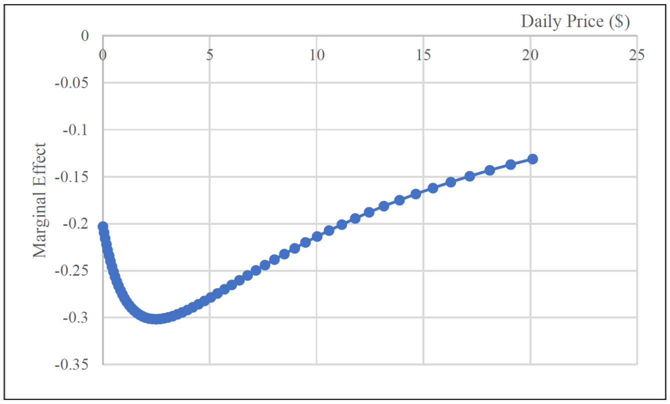

Figure 1.4: Free Parking Scenario

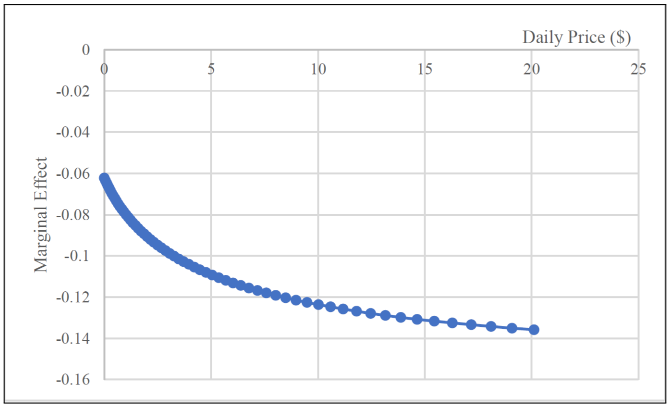

Figure 1.5: Parking Opportunity Cost Scenario

The model also explains how other covariates impact the decision to drive. The estimates and significance for these variables seem to be consistent over all specifications . Under all approaches higher income seems to increase the probability of driving, as does age. Race and gender, on the other hand, do not seem to influence the probability of driving to work. The effect of education varies, depending on the level of education. High School graduates drive more, whereas college degree holders tend to drive less.

The distance between home and work is significant, indicating that commuters tend to drive more the farther away they live from work. Moreover, the greater the ratio of transit time to driving time, the more likely driving is. These results are all intuitive.

The distance between the work location and the nearest transit station is significant, indicating that the farther away a transit station is from the work location the more likely a commuter is to drive. The same effect is present for the distance between home and the nearest transit station, but is significant only in the specification (3). Once population density is controlled for the effect disappears. The findings for transit station distance is in line with those of Kwoka et al. (2015) for Denver, where proximity to transit at the work location is more important than proximity at home. Population density is significant and indicates, as the literature suggests, that higher population densities encourage alternative commute mode choices. The effect of population density at the work location disappears, however, once county fixed effects are controlled for.

Whether the work trip is part of a chain trip, and whether departure time from home is during peak hours are both significant and indicate that commuters are more likely to drive if they have another stop to make on the way to work, and are less likely to drive if they depart during rush hours. Although the effect of these variables is intuitive, they may seem to be prone to an endogeneity problem since they are choice variables. Therefore, it may be argued that people may choose the time of departure based on their travel mode choice, or that they may make additional stops on the way to work just because they have the car. Moreover, since departure time and number of stops on the way to work are choice variables, they both may be driven by unobserved heterogeneity. However, since the study is limited to work commutes, it is reasonable to assume that the majority of commuters are not very flexible with the time of departure to work. Similarly, it is reasonable to assume that the majority of commuters will not make additional stops on the way to work just because they have the car. Commuters tend to be rushing to work, or are on a tight schedule and are unlikely to squeeze in unnecessary stops just because of their chosen mode of transportation. Generally speaking, it is reasonable to argue that for work commutes, departure time and stops on the way to work are typically not flexible choices and therefore may not be susceptible to typical concerns associated with choice variables. At any rate, the variables are not of focus in this paper and are only used for control purposes and do not seem to affect the robustness of the results for parking prices.

1.7 Conclusion

This study examines the impact of parking prices on the decision to drive to work in California, and finds a statistically significant impact. The study offers different approaches to the estimation of parking prices that are unobserved conditional on not driving, i.e., are unobserved for non-drivers. The paper also raises the question of the impact of incentive packages offered to encourage drivers to switch to an alternative mode; in particular, incentives offered in lieu of parking. The analysis offers a way to address such a concern, and deals with the extreme case of assuming that everyone is offered some form of incentive. The estimates obtained thus provide a range of average marginal effects in which the true marginal effect lies, and suggest that a 10% increase in parking prices could potentially reduce the probability of driving alone to work by 1 - 2 percentage points.

The results imply that parking pricing policies can be effective transportation demand management tools, yet the effectiveness will vary depending on the initial parking price charged. Charging the market price for parking can reduce vehicle usage, leading to less traffic congestion, energy consumption, and pollution. Moreover, the results suggest that elimination of EP parking, or offering incentives in lieu of parking, can be effective in reducing driving to work alone, as long as the parking prices in the workplace neighborhoods are positive. A major limitation of the paper is that the dataset is cross-sectional, limiting our ability to learn more about the dynamics of the effectiveness of parking policies over time. It also limits the ability to address identification threats adequately. Future research could improve on the estimation of parking prices by collecting actual parking data and matching it to the work locations of commuters. Furthermore, future research could extend the binary analysis to mode choice analysis, exploring the effects of parking prices on the alternative commute modes, and not only on the decision to drive.

Chapter 2

Commute Mode Choices and the Role of Parking Prices, Parking Availability and Urban Form: Evidence from Los Angeles County

2.1 Introduction

There is growing evidence that inexpensive or free parking is a major incentive for solo driving [Willson and Shoup, 1990; Willson, 1992; Hess, 2001; Shoup, 2005]. When commuters are shielded from the cost of parking, they drive alone more often, frustrating public policy goals to reduce solo driving. This implies that employer-paid parking, a popular fringe benefit, incentivizes workers to drive to work alone. By directly or indirectly subsidizing parking, employers reduce the cost of the commute trip by car while requiring the employee to pay only the driving cost. Thus, parking subsidies at the workplace can work at cross purposes with public policies designed to reduce traffic congestion, energy consumption, and air pollution.

In addition, all car trips start and end in a parking space. When parking is scarce or hard to find at either or both ends of a trip, the relative advantage of transit exceeds the advantages of automobile use. There is evidence that guaranteed off-street parking at home also results in a larger share of car owners choosing to drive to work, even in areas that are well served by public transit (Weinberger, 2012). Zoning regulations that require residential buildings to include off-street parking for some or all residents therefore contribute to increases in driving to work [Weinberger et al., 2008a;2008b; 2009] and to losses in urban green spaces. In fact, the additional off-street parking required by residential parking minima comes often at the expense of green space, as front and rear yards are converted to alleys or driveways to accommodate vehicles. Moreover, parking can influence destination choice, trip timing, and car occupancy [Feeney, 1989; Inci, 2015], as well as car ownership [Guo, 2013a; 2013b].

A good understanding of how parking costs and different types of parking availability at both ends of a trip affect commute mode choices is thus necessary if a shift away from the private car towards more sustainable methods of transport is to be achieved. Comprehension of such a matter will result in the development of appropriate land use planning strategies and infrastructure provision to increase the use of sustainable transport modes (e.g., public transport, walking and cycling) to places of employment.

In this paper we study the work commute mode-choice effects of the interactions of parking price, parking space availability (both at work and at home locations), and urban form, in Los Angeles County. We split our analysis into three research questions. First, we explore how parking prices, parking space availability, and urban form jointly affect the decision to drive. We use a conditional binary logit model to estimate the effects of the factors of interest. Second, we explore how parking prices, parking space availability, and urban form jointly affect the commuting mode choice, where commute mode choice is broken down to three alternatives: driving, public transit, and non-motorized. This question is addressed using a conditional multinomial logit model. Third, we predict the response of the probability of driving to work to changes in parking prices, and parking space availability, while also examining the effect of simultaneously changing parking prices and other variables of interest.

Los Angeles County offers a good setup to study our questions. Los Angeles is widely recognized for its automobile dependence and issues associated with traffic congestion, which may be due to the decentralized nature of the county's density along freeway corridors (Sorensen et al., 2008). Los Angeles County currently has more lane-miles of arterials, highways, and interstates per square mile than any other US metro area (Federal Highway Statistics, 2013). However, the area occupied by roads is only a fraction of the land devoted to automobiles, since the total land area dedicated to on- and off-street parking is 40% larger than the 140 square miles dedicated to the roadway system (Chester et al., 2015). Manville and Shoup (2005) find that Los Angeles County offers more abundant and free parking than many other cities in the United States. This infrastructure is scattered throughout the metropolitan area in on-street parking spaces and off-street parking lots and structures.

Our results suggest that parking prices are an important determinant of commute mode choice to work in Los Angeles County. Parking prices have a significant negative impact on the decision to drive to work, and our analysis indicates that a 10% increase in parking prices may decrease the probability of driving to work by 1.1%. The multinomial logit results indicate further that higher parking prices shift people away from driving alone to using public transit and non-motorized modes. Another interesting finding is that parking space availability at home, both off-street and on-street, increases the probability of driving to work. On-street parking availability at home, however, seems to matter more than residential off-street parking availability at home, since a 10% increase in on-street parking would increase the probability of driving by 1.3% whereas a similar increase in residential off-street parking would increase the probability of driving by only 0.6%.

Our evidence thus suggests that parking pricing can be an effective transportation demand management tool, and that elimination of employer-paid parking and provision of in-lieu parking subsidies can reduce driving to work. Moreover, parking space availability at home cannot be overlooked in parking policy decisions. Based on our analyses, it is evident that removal of minimum parking requirements may help reduce vehicle usage. The remainder of this paper is organized as follows. Section 2.2 starts with a brief literature review looking at the link between land use, parking charges, parking availability and mode choice. Section 2.3 outlines the empirical framework which is then followed by a description of the data used, the data sampling processes, and the selected variables. Section 2.4 discusses the estimated results, while section 2.5 discusses the predictions produced by our model. Finally, a conclusion wraps up the paper, with special attention given to the practical implications of our research.

2.2 Literature Review

The literature on land use, parking pricing and availability, and mode choice is vast. This section provides a brief review of the main elements of these research streams which have been studied using different approaches, cities, and data sources.

Parking costs have been shown to be an important factor in travel mode choice. For example, Willson and Shoup (1990) review empirical studies of car parking subsidies and find that eliminating free car parking at work reduces single-occupancy vehicle commuting between 19% and 81%. Another study examining parking subsidies in Los Angeles finds that between 25% and 34% fewer automobiles were driven to workplaces when workers had to pay to park their cars (Willson, 1992). Shoup (1997b) reviewed the effects of “cash out” programs and finds that single-occupancy vehicle commuting fell by 17% among employees in eight case study firms after they complied with California’s cash-out requirement. An analysis of parking subsidies in Portland, Oregon, estimated that a daily car parking charge of $6 reduced single-occupancy vehicle commuting by 16% (Hess, 2001), whereas Khordagui (2017) finds that a 10% increase in parking prices reduces the probability of driving to work by 1 - 2 percentage points, depending on whether in-lieu of parking packages are offered.

There are also studies that have examined how parking space availability in workplace and residential settings affect work commute mode choices. While parking space availability at employment sites has attracted considerable research efforts for decades (Inci, 2015), parking space availability at home sites, has recently become subject to increasing research interest [Christiansen et al., 2017; Guo, 2013b; Marsden, 2006; Weinberger, 2012]. The association between features of the built environment at the employment site or residence and commute choices has also been studied by several researchers. Examples include Cervero (2018, 1996), Kockelman (1995), Messenger and Ewing (1996), Cervero and Wu (1997), Levtnson and Kumar (1997), Ewing and Cervero (2001), and Ewing and Cervero (2010). In general, land use characteristics have been found to significantly impact mode choice decisions.

A central challenge for empirical analyses of parking and car use is access to sufficient amount of valid and reliable data. Most prior research depended on a limited number of case study areas [Weinberger et al., 2008b; Transport for London, 2012], manual observations [Christiansen and Hanssen, 2014; Hanssen et al., 2014], stated preference surveys (Guo and McDonnell, 2013) and parking revenue data (Kobus et al., 2013). A few studies have combined different data sources, such as residential prices and residential parking zones (van Ommeren et al., 2011) or, household travel data and parking information retrieved using Google Street View (Guo, 2013b).

Our work contributes to this existing literature by investigating the impact of parking strategies (parking charges and parking space availability at employment and residence sites) on work trip mode choice, using a rich dataset for Los Angeles County: the household travel data from the 2012 California Household Travel Survey (CHTS), parking prices retrieved using a unique dataset on commercial parking facilities for the Southern California Association of Governments (SCAG) region, and parking availability from a dataset on on-street and off-street parking in LA County. We also explore the disincentive effects for auto-commuting of daily parking charges with secured parking space at work under alternative urban forms and residential parking space availability scenarios. Moreover, our conditional binomial and multinomial logit models allow us to jointly identify both commuter and travel mode specific effects on the commute mode choice.

Modeling of mode choice is done by means of a discrete choice model. Most of the mode choice studies have used binary and multinomial logit models. However, the conditional logit regression model, known as McFadden's choice model, is also well suited for the work-trip travel mode choice framework because it exploits detailed information on transport modes while allowing for multiple mode choices. While the form of the likelihood function of the conditional logit is similar to that of the unconditional logit, the variables of the conditional logit are choice-specific attributes rather than individual-specific characteristics. Therefore, variation in the attributes of the transport modes in a commuter's choice set drives the estimates. Our conditional logit model builds on McFadden's conditional logit but is more general as it allows for the two types of independent variables: mode-specific and commuter-specific variables. Mode-specific variables vary across both individuals and travel modes, whereas commuter-specific variables vary only across individuals.

2.3 Data and Selected Variables

2.3.1 Probabilistic Choice Model

The first discrete choice framework models the decision to drive as follows:

$$D_i^* = \beta_P P_i + \boldsymbol{SPACE_i \beta_{SP}} + \boldsymbol{URBAN_i \beta_{URB}} + \boldsymbol{HH_i \beta_{HH}} + \boldsymbol{T_i \beta_T} + Z_{ij}\gamma + \; u_i \quad (2.1)$$

$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0 \\ 0 & \text{if } D_i^* \leq 0 \end{cases} \quad (2.2)$$

where D_i^* represents a commuter's unobservable propensity to drive, and D_i is a binary variable that takes the value of 1 if the commuter drives alone or carpools and 0 otherwise. P_i represents the daily parking price, $\boldsymbol{SPACE_i}$ represents the various parking space measures, $\boldsymbol{URBAN_i}$ is a vector of urban form measures, $\boldsymbol{HH_i}$ is a vector of household controls, $\boldsymbol{T_i}$ is a vector of trip characteristics, and Z_{ij} is the speed of the alternative travel modes. We estimate equation 2.2 as a conditional binary logit model, where P_i, $\boldsymbol{SPACE_i}$, $\boldsymbol{URBAN_i}$, $\boldsymbol{HH_i}$ and $\boldsymbol{T_i}$, are commuter-specific, i.e., vary across individuals (i), and Z_{ij} is mode-specific, i.e., vary across travel modes (j) for each individual.

Next, we examine the impact of our explanatory variables on the choice of three travel modes, in a conditional multinomial logit framework. Typically, we would like to explore the effects on driving alone, public transit, carpooling, non-motorized, and passenger. However, due to the small number of observations of both carpooling and passenger groups in our sample, we cannot rely on estimates of such a model. Therefore, we limit our next analysis to the subsample where there are only three mode choice alternatives: drive alone, public transit or active commute such as walking or biking. In the model presented below, driving alone is treated as the reference mode. As such, our second mode choice framework models the commute mode decision as follows:

$$M_i^* = \beta_P P_i + \boldsymbol{SPACE_i \beta_{SP}} + \boldsymbol{URBAN_i \beta_{URB}} + \boldsymbol{HH_i \beta_{HH}} + \boldsymbol{T_i \beta_T} + Z_{ij}\gamma + \ u_i \quad (2.3)$$

$$M_i = \begin{cases} 1 & \text{if } DriveAlone \\ 2 & \text{if } Transit \\ 3 & \text{if } Non-motorized \end{cases} \quad (2.4)$$

and M_i^* is the unobserved propensity of mode choice.

2.3.2 Data Sources

Commute mode choice, socioeconomic characteristics, and trip-related characteristics are obtained from the 2012 California Household Travel Survey (CHTS). The 2012 CHTS dataset is a cross-sectional dataset that includes 42,431 households from all over California and is obtained from the Transportation Secure Data Center of the National Renewable Energy Laboratory (NREL). Addresses are provided at the census tract level in the public use dataset, but we are able to work with geocoded addresses available in the NREL secured data portal. We limit our dataset to Los Angeles County commuters, looking at work trips that start and end within the county. We also limit our study to households that have at least one vehicle.

The 2012 CHTS has data on whether the commuter pays for parking at work if driving, and the amount paid. However, two major drawbacks emerge when dealing with the 2012 CHTS reported parking prices. First, those who do not drive to work do not report on parking prices and, therefore, the parking price they would have paid had they driven is unobservable. Second, the 2012 CHTS does not provide information on whether the commuter receives employer-paid parking and whether an in-lieu of parking subsidy, such as a cash-out program, is offered. Therefore, we do not observe whether the reported free parking is indeed free or conceals an unobserved opportunity cost of parking. To address such concerns, we make some assumptions.

We assume that all commuters failing to report a positive parking price are in fact facing some cost of parking, that is, we assume that they are offered both employer-paid parking and in-lieu-of-parking subsidies. This will provide conservative estimates for the effect of a daily parking price on mode choice because we assume that all those who drive to work face a parking charge, which may not necessarily be the case. In addition, we proxy daily parking charges with the work subarea market average price paid for parking.

To calculate our parking charge measure we use a unique dataset that provides data on market parking prices and parking supply at commercial off-street parking facilities that are open to the public in the SCAG region. This dataset is constructed from field work and information from websites such as Parkopedia, BestParking, Parkme, LADOT, and the ie511 to account for commercial parking spaces open to the public in each census tract and for market parking prices. Market parking prices vary depending on location, since parking facility costs are based on land costs, which itself varies. Costs are likely higher toward central business districts, where land values are higher, and lower in suburban areas. Additional sources, such as press articles, the Yellow Pages, and local government's websites, are also used in the compilation of these parking data. Further investigations were carried out for some of the parking facilities through on-site inspections, phone calls, or Google Street View for either clarifications or validation purposes. More information on the parking facilities dataset can be found in Franco (2016b).

We geocode the addresses of all the identified commercial off-street parking facilities and match them to the commuter's work location from 2012 CHTS using GIS techniques. The weighted average parking price within a radius of the work location is then computed for

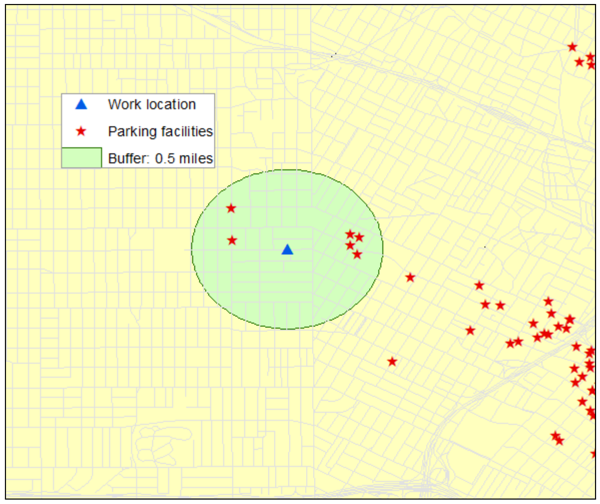

Figure 2.1: Illustration of Parking Prices Calculation

each commuter using the daily price charged and the number of parking spaces in each of the parking facilities located within the work subarea.

Since data on the search time and walking time for parking are also not available, we assume that a commuter would not walk more than 10 minutes from a parking location to the workplace. This corresponds to a half mile radius around the employment site. As such, the weighted average market price for parking within a half-mile radius of the work location is calculated and used to measure the price disincentive to driving (acknowledging nevertheless that there may be different tradeoffs between cost and walking time that are not measured). This is illustrated in Figure 2.1.

As shown in Figure 2.1, only parking facilities located within the half-mile buffer are considered when calculating the weighted average parking price. Commuters who report positive parking prices on the 2012 CHTS are assigned their reported parking prices. On the other hand, all those who report free parking or do not report paying for parking or do not report what their parking cost would have been had they driven to work, are assigned our matched weighted average market parking price.

Parking availability of on-street and off-street parking at the origin (home) and destination (employment site) of the work trip are calculated based on the 2010 parking supply census tract data for Los Angeles County assembled by Chester et al. (2015).

Finally, we use the countywide zoning and the Los Angeles census tract GIS shapefiles, available from the Los Angeles County GIS Data Portal, to construct our land use mix measure at employment sites. The dataset includes zoning codes for each city, their general plan codes, and a generalized general plan code developed by SCAG to support regional analysis.

Because we restrict our sample to observations for which there are no missing values and to those where a half-mile radius average parking price can be computed, our final sample for the analysis includes 927 individual home-based work trips.

In the first part of the analysis, where we investigate the binary decision of whether to drive, both solo drivers and carpoolers are grouped in the driving group. Drivers comprise 79.6% of the sample (1.3% are carpoolers whereas 78.3% are solo drivers). Transit users, those who use non-motorized means, and others are grouped in the alternative group. Transit riders comprise 11.4% of the full sample, whereas the non-motorized commuters represent 6.5%.

When we move to the multinomial analysis part, we have tighter definitions for the commute modes, where the first group now comprises only solo drivers, the second group comprises of transit users, and the third group includes non-motorized or active commuting users. The carpoolers and those using other commute modes are dropped in this part of the analysis because there are too few observations to comprise a group. We further note that a commuter is grouped in the transit class as long as public transit is part of the commute trip. Therefore, commuters who park and drive are also included in the public transit group. The non-motorized class includes commuters who take an active commute such as walking and cycling.

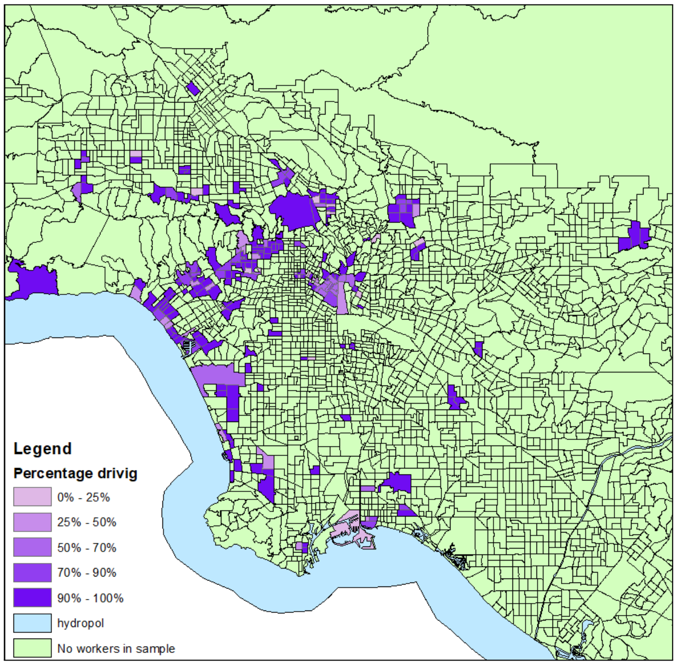

Figure 2.2: Percentage of driving commuters

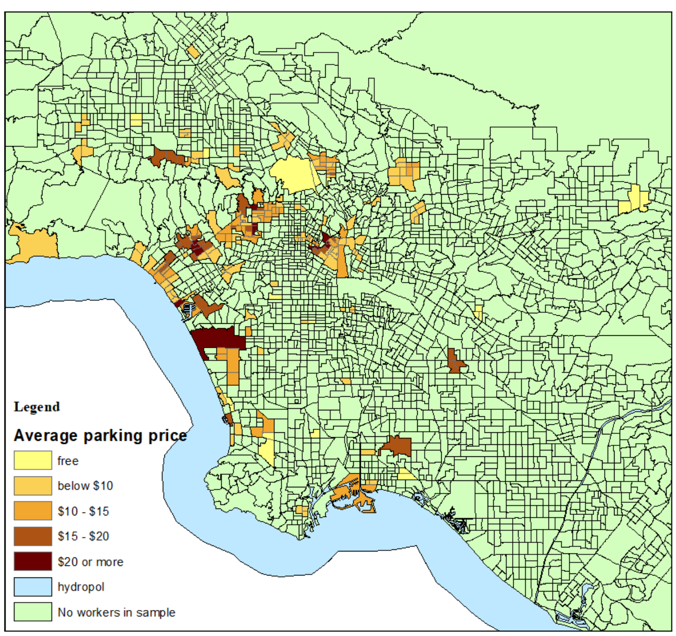

Figure 2.3: Distribution of Average Parking Prices at workplace census tract

Figure 2.2 shows the census tracts of Los Angeles County and the percentage of commuters driving to work in our sample. The color gradient represent tracts where these commuters work. Figure 2.3, on the other hand, shows the distribution of average parking prices assigned to commuters in our sample.

2.3.3 Discussion of Selected Variables

Parking Charges

Table 2.1 shows that the average reported daily parking price paid among those who drive and pay for parking is \$8.39. Yet, only 5.18% of those who drove to work reported paying for parking. This is not surprising given that a large number of LA employers provide free parking as a fringe benefit to their employees (Shoup, 2005). Our average constructed daily parking price is \$11.95.

Table 2.1: Descriptive Statistics of Parking Variables

	Full Sample	**Drive**	**Not drive**
Parking Charges			
Drivers Reported Paying for Parking (%)	5.18		
Average Reported Daily Parking Price (\$)	8.39		
Average Constructed Daily Parking Price (\$)	11.95	11.5	13.62
Parking Availability			
Average Nonresidential Off-street Parking Spaces for Work Tract (per square mile)	50,240	45,340	68,747
Average On-street Parking Spaces for Home Tract (per square mile)	4,360	4,345	4,418
Average Residential Off-street Parking Spaces for Home Tract (per square mile)	4,091	3,982	4,502

Parking Availability

Our parking space availability measures include three types of parking supply variables: the number of non-residential off-street parking spaces at the work census tract, the number of on-street parking spaces available at the residence census tract, and the number of non-residential off-street parking spaces at the residence census tract. On-street parking spaces refer to both metered and unmetered parking spaces at the curbside. Non-residential off-street parking spaces refer to both parking lots and parking structures, whereas residential off-street parking spaces refer to driveways, and parking garages allotted for residential use. The natural logarithms of these variables are used in the conditional logit regression.

Parking availability variables can be proxies for parking convenience at origin and destination locations. Different types of parking (on-street versus off-street) can be used as proxies for parking certainty to a household. For instance, for residential on-street parking, certainty is reflected by the combination of search time and distance between parking location and home, which may vary by time of day and week. For residential off-street parking, parking certainty may depend on the agreement between the landlord and the tenants. However, when a household chooses a commute mode, the household considers parking certainty from all available parking options at the residence location, not just the one they eventually use. Parking location (on-site, off-street and on-street) can also proxy for parking ease, assuming each location has a set of parking features. Table 2.2 shows the distribution of those who park for free and those who pay to park by parking location type.

Table 2.2: Workplace Parking Location

Location of Parking for Drivers Who:	**On-site**	**On-street**	**Off-street**
Pay to park	0%	13.70%	86.30%
Do not report paying	80%	6%	12.50%

As shown in Table 2.2, 80% of those who drive yet do not report paying for parking park on-site as opposed to parking on-street or off-street at a commercial lot or garage. Most of those who pay to park (86.3%) are parking off-street. It should be noted, however, that our model neither predicts commuters' choice of parking location, nor controls for parking location. We do not control for parking location because this information is missing for non-drivers.

A rich set of additional covariates capturing other features of the commuting cost, urban form and commuter's characteristics are included in our analysis to control for other factors that affect mode choice. Table 2.3 provides descriptive statistics for the urban form variables of interest, whereas 2.4 provides descriptive statistics for both household and trip-related controls.

Table 2.3: Urban Form Summary Statistics (Mean)

	Full Sample	**Drive**	**Not drive**
Land Use Mix Measure			
Herfindahl Index	0.25	0.25	0.26
Public Transit Accessibility Measures			
Distance Between Home Location and Nearest Rail Transit Station (miles)	1.37	1.4	1.3
Distance Between Work Location and Nearest Rail Transit Station (miles)	0.97	1	0.85
Total Distance to Nearest Rail Transit Station (miles)	2.34	2.4	2.15
Density Measures			
Population Density for Work Tract (Person / Sq.Mile)	11,673	11,477	12,413
Population Density for Home Tract (Person / Sq.Mile)	12,362	11,651	15,048
Job Density for Work Tract (Person / Sq.Mile)	60,035	50,785	94,969
Job Density for Home Tract (Person / Sq.Mile)	4,365	3,946	5,947

Urban Form Variables

As shown in Table 2.3, our urban form measures belong to three categories: land use mix, public transit accessibility and density. The first category comprises a land use mix diversity variable calculated as a Herfindahl index (HHI). The index is the sum of squares of the proportion of different component parts. If there is only one land use type present in an area, HHI will equal 1, and if all land use types are equally present, then HHI $= \frac{1}{k}$, with k equal to the number of land use types. Therefore, the higher values of HHI correspond to less land use mix. The HHI is symmetric with respect to land uses and sensitive to the size of the most prevalent land use.

The second category of urban form measures represents public transit accessibility. In the analyses that follow, we use total distance to the nearest rail transit station. This is calculated as the sum of the distance from residence to the nearest rail transit station and the distance from the workplace to the nearest rail transit station. The rationale for this approach is that a commuter would typically think of rail accessibility as the total distance that needs to be traveled to and from the rail station, rather than the separate accessibility from home or from work. The separate rail transit distance variables are also used for robustness checks. The two distance variables are constructed using GIS tools, the geocoded addresses of residences and workplaces from the secured 2012 CHTS dataset, and the geocoded addresses of rail stations obtained from the LA Metro website. The two variables represent the distance in bins of 0.05 miles, and are top-coded at 1.5 miles. As shown in Table 2.3, the average distance between residence and the nearest rail transit station is 1.37 miles, whereas the distance between the work site and the nearest rail transit station is 0.97 miles. Both average distances tend to be slightly longer for drivers than those using an alternative mode.

Density is the last land use measure included in our model specification. Density reflects how intensively land is used for housing, employment, and other uses. We construct population density (per square-mile in the residence census tract) and job density (per square-mile in the work census tract) using data from the Census Bureau (American Community Survey). Residential population density could also relate to differences in street connectivity and urban design [Saelens et al., 2003; Ewing and Cervero, 2010].

In our robustness checks, we also use a dummy variable denoted as 'CBD' which takes the value 1 if the census tract is classified as a CBD and 0 otherwise. This variable is intended to capture whether a commuter lives and/or works in a CBD as opposed to the urban and inner suburbs. As such, this variable is constructed for both residence and work census tracts. The designation of a particular census tract as 'CBD' resulted from the overlap of two layers using GIS techniques: a map layer with the classification of 99 zones of the SCAG region as CBD, urban and suburbs (Franco, 2016a), and a census tract layer[1].

[1] Franco (2016a) classified 99 zones of the SCAG region as CBD, urban and suburbs based on the square foot land values for each zone given by Zhang and Arnott (2011). The square foot values were used to

Household and Commuter Specific Characteristics

Our work trip data was matched with the appropriate demographic characteristics of the individual pursuing the trip and his/her household. The commuter characteristic variables are based on the socio-economic-demographic data reported by the respondents in the 2012 CHTS. These variables include household income, age (in years), gender (1=female) and race (1= white). The average annual household income is $97,782 and the average household size is 3 persons. All households in our sample are car owners, with the average number of vehicles in the household equal to 2 cars [2]. The median commuter age is 44 years. The sample is composed of 46% females, and 62% white commuters. We interact the income and household size measures to get a single variable representing income per household member, which has an average value of $41,755.

Other Work-trip related Characteristics

We distinguish between vehicle running costs and parking costs. We calculate the vehicle running money costs by multiplying gas prices by driving distance divided by 22 miles-per-gallon. Driving distance is calculated as the distance travelled in miles from the home census tract centroid to the workplace census tract centroid (using Google Maps API). Gas prices are computed as the lowest price of the home and workplace zipcodes gasoline prices. The zipcode gasoline prices are calculated as the average of the gas price charged at the three

calculate per acre values, which were then used to make the classifications. Zones with per acre land values less than $500,000 were listed as suburban, zones with per acre land values of $500,000 to less than $3,000,000 were listed as urban, and zones with per acre land values of $3,000,000 and above were listed as CBD.

[2] Most studies of the determinants of mode choice identify income and car ownership as important drivers in mode choice, though both variables are closely correlated. Increasing incomes make owning and maintaining a car feasible and also increase the opportunity costs of travel time making faster travel modes, such as the car, more attractive. However, some studies speculate that socio-economic factors may be less important in developed countries, where most households own a car (Lipps and Kunert, 2005). This suggests that demographic variables, such as household composition, gender, and age may be more important drivers of mode choice in wealthy countries. In addition, auto ownership is an endogenous variable. Given that all households in our sample are car owners we have not included this variable explicitly in the analysis. Yet our models in a sense capture a worker's work choice of mode conditional on auto ownership.

Table 2.4: Household and Trip Summary Statistics (Mean)

	Full Sample	**Drive**	**Not Drive**
Household Characteristics			
Household Income ($)	97,782	100,777	86,468
Household Size	3	3	3
Household income per person ($)	41,755	43,273	36,019
Age	44	45	43
Female (%)	0.46	0.46	0.47
White (%)	0.62	0.62	0.63
Trip Characteristics			
Distance from home to work (miles)	18.23	20.02	11.37
Running Cost ($)	1.97	2.02	1.76
Speed of drive mode (miles/hr)	29.6	30.07	27.79
Speed of transit mode (miles/hr)	9.84	9.9	9.58
Rush Hour (%)	0.57	0.56	0.62

nearest gas stations within each zipcode provided by gasbuddy.com. We assume an average miles-per-gallon of 22 in our computation of the running cost.

We also include the distance between home and work locations. The CHTS 2012 reports for each commuter the distance traveled on each trip. However, this measure would be unsuitable for our use because it is a function of the travel mode. Instead, we use the straight-line distance between the home and work geocoded addresses in its logarithmic form. This variable is independent of mode choice, and is used to measure how far the work location is, instead of the actual distance traveled.

Another covariate is the time of departure. We construct a binary variable 'peak' that takes the value of 1 if the commuter leaves home to work during peak traffic hours, and 0 otherwise. Peak traffic hours are defined as weekdays from 5.00 - 9.00 am and 3.00 – 7.00 pm.

We also control for the speed of the commute mode, which is constructed by dividing the distance traveled by the trip duration. The 2012 CHTS includes data on the actual duration of the trip. However, this is not a measure that we would be able to use because, like trip distance, it is a function of the mode choice. Therefore, we construct travel distances and trip durations from Google Maps API. The preferred method would be to use the geocoded addresses of home and workplace locations to compute the individual travel mode times. However, this is not possible because such data are secured and cannot be used with internet access, and so we resort to using the centroid of the relevant tracts which are publicly available. Travel mode speeds are constructed for the commute modes of interest. In the binary analysis, transit speed is used for the non-driving commute mode alternative, and in the multinomial analysis the bicycling speed is used for the non-motorized mode. This way, the convenience of each mode is accounted for in the analysis. The average speed for the driving mode is 29.6 miles per hour, as shown in Table 2.4, and that for transit is 9.8 miles per hour. These speeds factor in traffic, as well as the waiting time or access time in the public transit mode case.

2.4 Estimation Results

The results of our estimated conditional logit models are presented in Tables 2.5 (the binomial model) and 2.6 (the multinomial model). For ease of interpretation, we present the results as odds ratios and elasticities. The odds ratio gives the proportionate change in the relative risk of choosing a given alternative. In the case of our multinomial logit model, they represent the likelihood of choosing to commute by public transportation or active commuting (walking or cycling) relative to the base category of driving, while controlling for other variables in the analysis. An odds ratio greater than 1 implies that an increase in the explanatory variable increases the likelihood that Y=1 and, vice-versa.

2.4.1 Conditional binary logit: to drive or not to drive?

We first examine the choice between driving and not driving, where driving refers to solo drivers and carpoolers and non-drivers to transit riders, non-motorized mode commuters, and other mode users. Results are presented in Table 2.5. As one would expect, an increase in any of the mode cost attributes reduces the chances of that mode being chosen for the commute to work. Parking and running costs have the largest effect of the cost attributes on mode choice. However, between the two, running cost has a relatively smaller effect on the likelihood of driving. The odds ratios for the daily parking price and parking space availability are particularly revealing when addressing questions of public policy. The model produces a significant odds ratio for the daily parking price variable which is less than 1, indicating that raising a commuter's daily parking price at work decreases the likelihood that he or she would drive to work. Based on the elasticity results, a 1% increase in the daily parking charge at work reduces the probability of driving to work by 0.11%, while increasing the probability of choosing the alternative mode by 0.36%. Similarly, a 1% increase in running costs reduce the probability of driving to work by 0.08%, while increasing the probability of the alternative mode by 0.38%.

On the other hand, increasing on-street and off-street parking space availability at the residence location produces significant odds ratios that are greater than 1. This suggests that parking availability at home increases the likelihood of commuting to work by car, with a stronger effect for on-street parking. A 1% increase in the number of on-street parking spaces increases the probability of driving to work by 0.12%, while decreasing the probability of the alternative mode by 0.49%. Likewise, a 1% increase in the number of off-street parking supply at home increases the probability of driving to work by 0.06%, while decreasing the probability of the alternative mode by 0.24%.

Interestingly, the odds ratio for off-street parking availability at the work location is also greater than 1 though insignificant. Our finding may be due to the composition of our sample since most of the commuters who drive to work (74%) park on-site at their employment site, which may result in their lack of concern about off-street parking availability at the work location[3].

Regarding the other variables, our odds ratios reveal that the farther away work is from home, the more likely the commuter is to drive. Furthermore, the longer the distance traveled to and from the nearest rail transit station, the more likely the commuter is to drive. Income also has a significant odds ratio greater than 1, indicating that an increase in income increases the likelihood of driving to work.

Another interesting result is the effect of rush hour on the likelihood of choosing a mode choice. The odds ratio associated with our peak variable reveals that the likelihood of driving to work tends to decrease during rush hours. This is not surprising since traffic congestion costs occur primarily during peak periods. The probability of driving to work tends to decrease by 7%, while the probability of not driving to work tends to increase by 35% as a result of leaving to work during a rush hour.

Next, our conditional binary logit model is extended to multiple modes. Parking, land use mix, income and distance variables are still used in their natural logarithmic form. All estimations include a constant.

2.4.2 Conditional multinomial logit: which commute mode?

Results for our conditional multinomial logit specification are shown in Table 2.6. The reference mode for this part of the analysis is driving. Our conditional multinomial and

[3]In contrast to on-street parking, for off-street parking, the available space, location and time are often guaranteed without uncertainty. For example, street cleaning makes street parking temporarily unavailable. Even though most cities in CA own a small amount of off-street parking supply, most city officials, especially in large cities, consider provision of parking as a role of the private sector. Moreover, parking regulations that mandate a minimum number of parking spaces for a given floor area for each possible use of the property are also common, which ends up securing parking spaces at home and at work. In addition, employer paid parking is also a common fringe benefit, which secures a free or partially subsidized parking space at the work location, regardless of whether the spot is located on the premises of the work facility or rented by the employer in a nearby commercial parking facility for employee use.

Table 2.5: Conditional Binomial Logit Results

Dependent variable: D_i	**Odds Ratio** ($D_i = 1$)	**Elasticity** ($D_i = 1$)	**Elasticity** ($D_i = 0$)
Daily parking price	0.623***	-0.106	0.36
	(0.095)		
Off-street parking at work	1.143	0.027	-0.107
	(0.142)		
Residential off-street parking at home	1.357**	0.062	-0.243
	(0.212)		
On-street parking at home	1.850*	0.125	-0.491
	(0.598)		
Land Use Mix	1.201	0.037	-0.146
	(0.156)		
Rail station distance	1.019***	0.159	-0.705
	(0.007)		
Population density at work	0.980**	-0.054	0.187
	(0.010)		
Job density at work	0.996***	-0.069	0.15
	(0.001)		
Population density at home	0.965***	-0.116	0.348
	(0.013)		
Job density at home	0.988	-0.015	0.041
	(0.009)		
Income per person	1.187**	0.035	-0.137
	(0.092)		
Age	1.006	0.13	-0.51
	(0.005)		
Female	0.94	-1.246	5.066
	(0.167)		
White	0.782	-4.717	21.771
	(0.125)		
Distance	1.944***	0.135	-0.53
	(0.310)		
Peak	0.690**	-6.976	34.729
	(0.110)		
Running cost	0.784**	-0.082	0.377
	(0.081)		
Speed	1.007	0.04	-0.058
	(0.015)		
Observations	927		
Wald Chi^2	112.41		

Null hypothesis: odds ratio is 1
Standard errors are reported in parentheses, and are clustered at the tract level
$^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$
Estimations include a constant

binary logit results are aligned. Raising a commuter's daily parking price at work increases the likelihood of the commuter travel to work by public transport or by a non-motorized mode (walking or cycling). A 1% increase in daily parking prices reduces the probability of driving to work by 0.09%, while it increases the probability of using public transit and active commuting modes by 0.37% and 0.3%, respectively. As in the case of the conditional binary logit model, the elasticity of driving with respect to daily parking price is low, which is to be expected in a short-run model of travel demand. Another possible explanation may be the existence of car parking fringe benefits in the form of free parking spaces at work.

Regarding parking availability at the work and residence sites, our odds ratios further reveal that increases in on-street parking availability at home, as well as off-street parking availability both at work and home, significantly decrease the likelihood of commuting by non-motorized modes. On the other hand, the likelihood of commuting by public transit seems only to be significantly decreased by an increase in off-street parking at home. This is an interesting correlation suggesting that parking features at the work location may impact physical activity as they also affect the likelihood of choosing to use an active commuting mode.

Regarding urban form, it is also interesting to note that an increase in the commuter's overall distance to the nearest rail transit station or of a decrease in land use mix at the commuter's employment site (which corresponds to an increase in the land use mix HHI) only significantly decrease the likelihood of commuting by public transit over driving. Higher population density at the residence census tract also increases the likelihood of taking public transit over driving. Our correlations may be related to most commuters in our sample living in urban and suburban areas of LA County with poor public transport service at home. Another possible explanation for these results is the jobs-housing imbalance that characterizes the County [Franco, 2013; SCAG, 2001][4].

[4]Jobs-housing balance refers to the distribution of employment relative to the distribution of resident workers within a geographic area. The central concern of jobs-housing balance as it relates to transportation

Table 2.6: Conditional Multinomial Logit Results

Dependent variable: M_i	**odds ratio**		**Elasticity**		
	$(M_i = 2)$	$(M_i = 3)$	$(M_i = 1)$	$(M_i = 2)$	$(M_i = 3)$
Daily parking price	1.584**	1.476*	-0.088	0.372	0.302
	(0.342)	(0.316)			
Off-street parking at work	1.059	0.683**	0.019	0.077	-0.362
	(0.176)	(0.105)			
Residential off-street parking at home	0.728*	0.631*	0.067	-0.25	-0.393
	(0.133)	(0.152)			
On-street parking at home	0.601	0.364*	0.126	-0.383	-0.886
	(0.202)	(0.211)			
Land Use Mix	0.710*	0.979	0.04	-0.303	0.019
	(0.125)	(0.287)			
Rail station distance	0.974**	0.985	0.162	-1.063	-0.522
	(0.011)	(0.011)			
Population density at work	1.013	1.041***	-0.059	0.09	0.418
	(0.015)	(0.013)			
Job density at work	1.004***	1.002**	-0.06	0.167	0.056
	(0.001)	(0.001)			
Population density at home	1.053***	1.021	-0.12	0.545	0.15
	(0.016)	(0.017)			
Job density at home	1.020*	1.001	-0.011	0.078	-0.007
	(0.010)	(0.012)			
Income per person	0.733***	1.353**	0.015	-0.296	0.317
	(0.070)	(0.189)			
Age	1	0.981**	0.124	0.165	-1.778
	(0.007)	(0.009)			
Female	1.256	0.579*	1.393	27.38	-41.313
	(0.316)	(0.173)			
White	1.143	2.147**	-5.751	7.731	102.344
	(0.229)	(0.738)			
Distance	1.705*	0.235***	0.037	0.571	-1.408
	(0.512)	(0.060)			
Peak	1.703**	2.000**	-9.486	54.133	80.837
	(0.361)	(0.587)			
Running cost	0.907	1.503***	-0.005	-0.19	0.761
	(0.188)	(0.231)			
Speed	1.03	1.03	0.064	0.112	0.13
	(0.040)	(0.040)			
Observations	899				
Wald Chi^2	369.65				

Null hypothesis: odds ratio is 1
Standard errors are reported in parentheses, and are clustered at the tract level
$^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$
Estimations include a constant

Thus, our conditional logit estimates suggest that strategies that promote densification, increase land use mix, and improve transit accessibility in transit catchment areas may positively influence work transit commuting. This in turn aligns with the idea that public transport demand is to some extent conditioned by the built environment. The integration of transport and land use planning is therefore also key to divert commuters from private car to public transport and to attain sustainable development. Our results further suggest that increasing daily parking prices or reducing off-street residence parking accessibility may encourage work commuting by public transit.

2.5 Further Discussion

We now use our conditional binary logit model to make predictions for the Los Angeles County commuters' mode choice. This is accomplished by using the conditional binary logit estimates to compute probabilities for a range of values for one particular variable (e.g. daily parking charges) while using average values for the remaining variables. Table 2.7 describes our base scenario when all the variables are set at their mean values. Tables 2.8 and 2.9 show the effects on the probability of driving to work under different scenarios and a range of daily parking charges.

---

policy is the journey to work. The concept implicitly assumes that workers choose to work as close to home as possible (or that workers choose homes as close to their jobs as possible). If a given region has a much greater concentration of employment than resident workers, workers must be drawn from other areas, leading to longer commutes. Similarly, if resident workers greatly outnumber job opportunities, they must seek jobs in other more distant areas. LA County is a job-rich area within the Greater Los Angeles region and the ratio workers-employed residents varies substantially across the county. For example, Franco (2013) shows that the ratio in the year 2000 ranges from 0.27 (Lake Los Angeles) to 3.26 (Westwood). The ratio for Downtown Los Angeles is 2.77.

2.5.1 Effects of Daily Parking Prices

Table 2.7 shows how the probability of driving falls as the daily parking price is gradually increased (even if still parking on-site). The results indicate that if parking is free of charge, while all other factors are equal to the corresponding mean values, then the probability of driving is 95%. The drop of the probability to drive in response to changes in daily parking prices is very gradual and not pronounced, holding all other variables constant at the mean. One possible explanation for the observed low diversion from driving to non-driving is that most drivers in the sample park on-site at work and, thus, are guaranteed a spot. Therefore, even though free or inexpensive parking has a bearing on a commuter's choice of driving versus using public transport or active commuting, a guaranteed parking space at work or abundant availability of workplace parking may be relevant in the choice of travel mode for the journey to work. In theory, we expect workplace parking charges to be more effective when guaranteed workplace parking capacity (on-site parking) is reduced. While a parking charge secures an efficient allocation of scarce spaces, limited parking capacity restricts the overall use of the car as a possible travel mode.

Unfortunately our model does not allow to test for the case where on-site parking at work (that is, parking availability on-site) is reduced because we do not have data on on-site parking for non-drivers. Table 2.7 also provides an overview of changes of the likelihood to drive as parking availability at home changes, while evaluating parking prices at work at the mean value. When the average daily parking price is combined with no parking availability at home, the probability of driving to work drops to almost 57% if no on-street parking is available, and to 75% if no off-street parking at home exists (that is, not having an own dedicated parking space with certainty).

Table 2.7: Probability of driving in response to varying significant parking variables

		Home Parking			
parking price	$\Pr(D_i = 1)$	**Off-street**	$\Pr(D_i = 1)$	**On-street**	$\Pr(D_i = 1)$
0	0.95	0	0.75	0	0.57
1	0.93	1,000	0.79	1,000	0.67
2	0.91	2,000	0.82	2,000	0.76
3	0.9	3,000	0.84	3,000	0.8
4	0.89	4,000	0.85	4,000	0.83
5	0.88	5,000	0.86	5,000	0.85
6	0.87	6,000	0.87	6,000	0.86
7	0.87	7,000	0.87	7,000	0.87
8	0.86	8,000	0.88	8,000	0.88
9	0.85	9,000	0.88	9,000	0.89
10	0.85	10,000	0.88	10,000	0.89
11	0.84				
12	0.84				
13	0.83				
14	0.83				
15	0.82				

For purposes of these computations, all other variables are evaluated at the mean

2.5.2 Effects of Daily Parking Charges and Home Parking Availability

Table 2.8, on the other hand, presents the effects of joint changes in daily parking prices at work and parking availability at home. When daily parking charges at work are set to zero under the scenario of home parking availability restrictions, the probability of driving to work drops from 95% either to 90% in the case of no off-street parking (that is, no parking certainty), or to 80% in the case of no on-street parking (that is, parking ease at home) at home. The effect of no residential parking is even more pronounced when combined with higher parking prices at the workplace.

Assessing the scenario of limited residential parking spaces (1000 spaces per square mile) reveals that coupling increases in daily parking prices at work with limited residential parking, particularly on-street parking, can lead to the sought-after reduction in driving. For exam-

Table 2.8: Probability of driving in response to joint variation of parking variables

	Home Parking					
	Off-street			**On-street**		
parking price	**0**	**1000**	**2000**	**0**	**1000**	**2000**
0	0.9	0.92	0.93	0.8	0.86	0.91
1	0.87	0.89	0.91	0.75	0.82	0.87
2	0.84	0.87	0.89	0.71	0.79	0.85
3	0.83	0.85	0.88	0.68	0.76	0.83
4	0.81	0.84	0.87	0.66	0.74	0.82
5	0.8	0.83	0.86	0.64	0.73	0.8
6	0.78	0.82	0.85	0.62	0.71	0.79
7	0.77	0.81	0.84	0.6	0.7	0.78
8	0.76	0.8	0.83	0.59	0.69	0.77
9	0.76	0.79	0.82	0.58	0.68	0.76
10	0.75	0.78	0.82	0.57	0.67	0.75
11	0.74	0.78	0.81	0.56	0.66	0.75
12	0.73	0.77	0.81	0.55	0.65	0.74
13	0.73	0.76	0.8	0.54	0.64	0.73
14	0.72	0.76	0.8	0.53	0.63	0.73
15	0.71	0.75	0.79	0.52	0.63	0.72

ple, setting on-street parking at home to only 1000 spaces per square mile while charging only $10 a day for parking at work can lead to a reduction of the probability of driving to work to 67%. A more pronounced effect of the same on-street parking availability at home combined with a daily parking price of $15 at the workplace can be seen in Table 2.8, where the probability of driving drops even further to 63%.

Therefore, limited access to parking at home affects vehicle usage, even when employee parking subsidies (e.g. free parking and/or secured parking at work) are in place. The magnitude of the shift in mode choice from driving to non-driving commuting can, therefore, be large when travel demand programs rely both on residential parking supply restrictions and workplace parking charges. Our results' implications are thus also aligned with recent studies that have shown that limited access to home parking is an effective way of reducing vehicle use on work trips [Weinberger, 2012; Christiansen et al., 2017].

2.5.3 Effects of Daily Parking Charges and Urban Form

Land use attributes in terms of density and mixture of different uses are surrogate measures of the built environment. These measures capture additional influences of some concealed factors affecting people's travel mode choices, to a certain degree. For example, increased density results in reduced spatial separation, enhancing travel by all modes. However, non-motorized modes benefit more from reduced spatial separation than motorized modes because the former are more sensitive to distance changes than the latter.

Table 2.9: Probability of driving in response to joint variation of parking price and urban form measures

Parking price	**Land Use Mix**		**Job Density**		**Population Density**		**Distance to Rail Station**	
	0.1	**0.2**	**80,000**	**120,000**	**15,000**	**22,000**	**0.5**	**1**
0	0.94	0.95	0.94	0.93	0.94	0.93	0.9	0.91
1	0.92	0.93	0.92	0.91	0.92	0.91	0.86	0.88
2	0.9	0.91	0.9	0.89	0.91	0.89	0.84	0.86
3	0.89	0.9	0.89	0.88	0.89	0.88	0.82	0.85
4	0.88	0.89	0.88	0.87	0.88	0.87	0.8	0.83
5	0.87	0.88	0.87	0.85	0.87	0.86	0.79	0.82
6	0.86	0.87	0.86	0.85	0.87	0.85	0.78	0.81
7	0.85	0.87	0.86	0.84	0.86	0.84	0.77	0.8
8	0.84	0.86	0.85	0.83	0.85	0.83	0.76	0.79
9	0.84	0.85	0.84	0.82	0.84	0.83	0.75	0.78
10	0.83	0.85	0.84	0.82	0.84	0.82	0.74	0.77
11	0.82	0.84	0.83	0.81	0.83	0.81	0.73	0.76
12	0.82	0.84	0.83	0.8	0.83	0.81	0.72	0.76
13	0.81	0.83	0.82	0.8	0.82	0.8	0.72	0.75
14	0.81	0.83	0.82	0.8	0.82	0.8	0.71	0.75
15	0.8	0.82	0.81	0.79	0.81	0.79	0.7	0.74

Table 2.9 presents the effects of daily parking prices at work on the probability of driving to work under different land use mix, density (employment and population) and overall accessibility to public transit scenarios at the workplace destination census tract. We remind the reader that our land use mix measure is an HHI index summing the squares of the proportion of different land use component parts. Our land use components include residential use,

open space and different types of non-residential uses (e.g. industrial, office, commercial, parking, etc). Therefore, the higher values of HHI correspond to less land use mix. Our HHI average value is 0.24. Above-average values of HHI result in probabilities of driving to work higher than those computed at the mean. So we focus on the cases where the HHI is below average. Similarly, only above average values of the density measures are explored.

Overall, improvements in land use mix, increased job density, and increased population density at the work location all lead to similar effects on the probability of driving, when combined with changes in parking prices. Transit accessibility seems to be more effective in reducing the likelihood to drive to work, which can bring down the chances of driving to work to 72% in the case of a total distance of 0.5 miles to the rail transit station (again, this is the sum of distance from home to the rail transit station and from work to the rail transit station) when combined with a parking price of $12 (the average constructed parking price for the sample).

Our results thus suggest that there are features of the land use that are important to a commuter's short-term mode choices that are not fully captured by just travel times or costs. This finding is consistent with past evidence on how urban form affects travel mode choices (see for example Cervero and Gorham (1995), Cervero and Kockelman (1997), and Cervero (2002)). In fact, urban form has diverse impacts on travel distance and mode. Compact development patterns, featured by relatively high population density, mixed land use, and easily accessible public transit facilities, are beneficial for walking, cycling and public transportation. The disincentive effect of daily parking charges at the workplace on driving can thus be reinforced if combined with land use and zoning policies that create more balanced land use, better overall public transport accessibility and more compacted built environments.

2.5.4 Effects of Daily Parking Charges and Household Income

We also examine the effects of daily parking prices under alternative annual household income scenarios on the probability of driving to work in Table 2.10. Consider a parking cost set at the daily average parking charge for the prediction sample ($12). Our predicted probabilities show that commuters in higher-income households have a more inelastic demand for driving when there is a parking charge. In particular, the probability ranges from an 80% chance of driving to work with an annual household income per person of $5,000 to an 86% chance with an annual household income per person of $80,000. The fact that the magnitude of the drop in probability of driving resulting from the drop in income does not match expectations for such a huge drop in income may be explained by the combined existence of secure parking at work and the dispersed land use pattern and overall low accessibility to public transit that characterizes LA County.

Table 2.10: Probability of driving in response to joint variation of parking price and income

Parking price	**Income per person**	
	5,000	**80,000**
0	0.93	0.95
1	0.9	0.94
2	0.89	0.93
3	0.87	0.92
4	0.86	0.91
5	0.85	0.9
6	0.84	0.89
7	0.83	0.89
8	0.82	0.88
9	0.81	0.88
10	0.81	0.87
11	0.8	0.87
12	0.8	0.86
13	0.79	0.86
14	0.78	0.85
15	0.78	0.85

Thus, our overall results show that the disincentives of parking charges at work may be hindered by the existence of secured parking at the workplace and by the lack of viable travel mode alternatives to driving when choosing the travel mode for the journey to work. As a result, the probability of commuting to work by car is still high when workplace parking prices increase, regardless of the household income bracket.

2.6 Conclusion

The worst traffic congestion in urban areas, and in particular in core-oriented cities, occurs during the periods of travel to and from work. Congestion exists partly because a lot of car owners find it more convenient to travel to work by car than by public transit or by active commuting, even in congested situations. Pricing the use of road space, parking policy and subsidizing public transit are just a few policy measures that have been proposed to reduce urban traffic congestion by increasing either the disincentives of car usage or the attractiveness of public transit.

This paper examines how parking prices and parking availability affect the probability of choosing to drive to work, using LA County as the context of our analysis. Moreover, and for completeness, we also explore the role urban form may play in the choice of travel mode for the journey to work. Our results show that daily parking prices have a significant negative impact on the decision to drive alone to work, where a 10% increase in daily parking prices reduces the probability of driving alone by 1.1%. Our conditional multinomial logit results further indicate that higher daily parking prices shift people away from driving to public transit. Another interesting finding is that work-related commuting appears to be in part a product of convenience and accessibility constraints: when public transit is proximal, workers are more likely to travel via that mode, whereas having parking available at the worksite is positively associated with work-related vehicle usage. In particular, parking at home, whether on-street or off-street, also has a significant impact on the decision to drive. The policy-relevant implication of these results is that maximum parking standards, and removal of parking and of eligible parking expense payment fringe benefits may reduce vehicle use on trips to work.

Furthermore, good accessibility to public transit at origin (residence) and destination (work) locations is also found to reduce vehicle usage. While limited number of parking spaces restricts the overall possibility of car use, parking prices help secure efficient allocation of scarce parking resources. Since the majority of workers in Los Angeles County are guaranteed a parking space by their employer, any move towards fewer and paid-for parking spaces can have a significant impact on commute mode choice. This, in turn, can also reduce congestion and urban pollution and the need for large investments in road capacity.

Like almost every empirical study of parking and vehicle use, endogeneity and selection bias are also limitations of our analysis. Residents are aware of parking availability when they decide to rent or buy a dwelling and when deciding where to work. It is then reasonable to assume that individuals with lower car use or car ownership are more likely to choose to live with reduced home parking availability and therefore also choose to work in locations where parking availability is smaller. In addition, individuals with unobserved preferences for using public transport, walking or cycling may choose to live in central areas at higher rates than in the suburbs. Thus, the potential for endogeneity and selection bias suggest caution should be taken in interpreting our results.

Finally, our study is based on Los Angeles County, which is widely recognized for its automobile dependence and associated issues with traffic congestion due to the decentralized nature of the county's density along freeway corridors. Thus, our results may not necessarily be representative of the rest of the country. Studies from other US counties could help solidify the results of our research.

Chapter 3

Determinants of Public Transit Use and the Role of Social Influence

3.1 Introduction

The majority of trips in the United States are made using a vehicle, and although the benefits of reducing vehicle usage are salient, the trend does not seem to be reversing. Public transportation is an important alternative to vehicle usage, and is generally thought of as a more sustainable mode of transportation, conserving energy and reducing emissions. However, ridership in many American cities falls short of what the system was built for. Although transit ridership has been increasing nationally over the years for a while, the increases over the past decade were largely due to increases in ridership in the New York region, while ridership of most of the other top transit areas declined (Mallett, 2018). National ridership, as well as New York's, has been also declining since 2014. Some of the proposed reasons for the drop in ridership include rising vehicle ownership, gentrification, the rise of services such as Uber and Lyft, to mention a few. However, careful examination of these factors is required before we can claim causality. Furthermore, apart from New York, transit ridership has been low even when it was on the rise.

There is a vast literature examining the factors affecting a commuter's preferences for one mode over the other, covering an array of built-environment measures, trip characteristics including financial costs, transit service accessibility and quality, and demographics. One thing is evident: public transit is not a perfect substitute for vehicle use. Even though transit is more affordable, and in some instances could be relatively convenient, commuters' preferences are still in favor of driving. To address transit use and understand the relative importance of underlying factors promoting it, it is important to examine these factors simultaneously, and across a number of regions to learn whether there are elements that are region-specific. It is also necessary to go beyond the baseline fundamental determinants that are well-studied in the literature and consider alternative factors, such as the role of social influence, that can indirectly, but significantly, impact transit use and how commuters and policymakers alike view it.

The goal of this paper is threefold. First, the paper builds a binary discrete choice model that estimates public transit use, which includes both bus and rail transit, incorporating a rich set of commute mode choice determinants. What this paper does differently is that it jointly examines these potential determinants at the tract level, in the workplace and residence neighborhoods simultaneously, across a number of major US cities. In order to do so, a unique dataset is built by drawing on data from various sources, and constructing variables using Geographic Information System (GIS) tools. The dataset combines household travel surveys from California, New York, New Jersey, and Massachusetts. The travel surveys provide tract level locations of both the residences and workplaces, which are then exploited in the construction of transit-access route distances using GIS tools, the construction of transit and driving trip durations and distances using Google Maps API, the construction of parking prices and gas expenses at the zipcode level through webscraping, and in matching other tract-level data.

Second, the paper introduces social influence to the model. Social influence arises when an individual's utility is a function of the actions and behaviors of the social group the individual is affiliated with. The role of social influence in commute mode choice in general, and in public transit use in particular, remains understudied, even though the importance of social interactions in decision making is increasingly recognized by economists. Of particular relevance are Manski (1993), Brock and Durlauf (2001), and Walker et al. (2011), and their discussions of social effects and estimation issues, as well as McFadden's (2010) discussion of how social interactions affect perceptions and preferences. Although the social reference group can be defined along many lines, ranging from immediate networks, such as family, friends, and coworkers, to broader definitions of social groups that an individual identifies with, this paper is specifically interested in the effect of the broader social group, and defines it as the social group encompassed within the tract neighborhood, whether residence or workplace. This social group is referred to as the spatial social group, and is measured by the local field effect, which is the share of transit users in a census tract. The definition of a social group is refined further by introducing both an income and a racial dimension to produce what are referred to hereafter as the income-group and the racial group. The income-group social influence is measured by the median transit user income, whereas the racial group social influence is measured by the percentage of white transit users.

Third, the paper makes a methodological contribution while accounting for potential endogeneity of social influence. A novel instrumental variable (IV) is constructed to identify the spatial social influence. The IV is the weighted average distance of commute trips of individuals in the relevant tract, and is computed for both the residence tract, and the workplace tract. The weighted average distance of trips for a given tract satisfies the exclusion criterion of IVs because the average distance of trips of a single commuter's neighbors or coworkers would only affect his decision to use transit through its effect on their decision to use transit. The effect of the neighbors' and coworkers' decision to use transit are the social influence effects of the residence and workplace respectively. The only potential correlation with the error term may arise because of potential collinearity between the commuter's actual commute distance and the weighted average distance of commute trips for the relevant tract. This possibility is overcome, however, by the inclusion of the individual's actual commute distance in the model. Furthermore, the weighted average commute distance is a relevant instrument because commute distance is found to be associated with public transit use, and the evidence that emerges in the paper rules out the possibility of it being a weak instrument. On the other hand, to identify income-group and racial social influence, an alternative identification strategy is employed. Both income and racial group fixed effects are controlled for, as well as the percentage of transit users in a tract, which is argued to control for tract-specific heterogeneity that would impact transit use. An interaction of the income and racial group fixed effects with the percentage of transit users in a tract further controls for tract-specific heterogeneity that would impact transit use and is income- or race-specific.

The evidence that emerges from the analysis confirms that white commuters and richer commuters are less likely to use transit. A $10,000 increase in household income reduces transit use probability by 0.2 percentage points, whereas being white reduces it by 2.8 percentage points. Both population densities and job densities positively increase the likelihood of transit use, whereas a longer access route decreases it. The rail access route seems to have a larger impact in the workplace tract. The farther the workplace is from the residence the more likely the commuter is to use transit, and for every additional 10 minutes spent on the trip when using transit (relative to the time spent driving) the probability of transit use drops by 2.2 percentage points. Higher parking prices encourage transit use, and so do higher gas expenses.

The findings suggest that spatial social influence exist in both the residence and workplace tracts, implying that a 1 percentage point increase in transit use among neighbors is associated with a 0.15 percentage point increase in the probability of a commuter's transit use, whereas the same increase among coworkers leads to a 0.4 percentage point increase in the probability of the commuter's transit use. The result for the workplace persists even when using the IV, but the effect of the residence network is no longer significant. Exogeneity tests, however, indicate that, given the wide array of control variables, we cannot reject the null hypothesis of exogeneity of spatial social influence. Taken together with the large first-stage F-statistics rejecting the weakness of the IVs, and the fact that the estimates of the marginal effects do not change when using the IV, the failure to reject the null hypothesis of exogeneity implies that the model adequately controls for major sources of heterogeneity and that endogeneity of spatial social influence is not a concern. Findings also indicate that income-group social influence is also significant, implying that the rich are more likely to use transit as the median income of transit users rises. The effect exists for both the residence and workplace, but is of a larger magnitude for the workplace tract, implying once more that the effect of coworkers is more important than that of neighbors. Overall, racial social influence has no or very little impact on the decision to use transit.

This analysis reveals that ridership may benefit by encouraging unconventional groups to use transit. For example, if the rich increase their use of transit, the effect will go beyond that one-time increase, encouraging more commuters to use transit through social influence. Policymakers tend to focus on the poor, but evidence may suggest that it would be more effective to target the rich, whose choice will cascade into increased use by both rich and poor. Modeling social influence at both the residence and workplace allows us to gauge the relative effect of neighbors and coworkers, and hence determine the more effective social group to target for a greater impact.

This study contributes to the literature in many respects. First, it builds a unique dataset that spans many US cities, and simultaneously models the determinants in both the workplace and residence at the census tract level. Second, the dataset includes a rich set of variables, constructed by synthesizing data from various sources, and exploiting the availability of both workplace and residence census tract information to compute specific commute trip characteristics. The model presented is one that jointly estimates a wide range of determinants, rather than limit the focus to built-environment factors or to trip characteristics. Third, social influence is examined and both neighbors and coworkers are accounted for. Furthermore, alternative definitions of social groups are explored; namely, spatial groups, income groups, and racial groups. Fourth, a novel IV is constructed to address potential endogeneity in the case of spatial social influence, and an alternative identification strategy is devised in the cases of income-group and racial social influence.

The paper is organized as follows. Section 3.2 gives a brief overview of the literature and discusses further the motivation for the paper. Section 3.3 describes the data used, the construction of the dataset, and the variable definitions. Section 3.4 presents the model to be estimated and lays out the identification strategy for social influence and the construction of the IV used, while section 3.5 reports and interprets the results. Section 3.6 discusses the implications of the findings, whereas section 3.7 concludes, describing the study limitations and suggesting potential extensions for future research.

3.2 Background

The determinants explored in this paper can be grouped into socioeconomic characteristics, neighborhood characteristics, transit accessibility, and trip characteristics. Socioeconomic characteristics are included to gauge for factors such as income and race, among other characteristics. Neighborhood characteristics include built environment measures as well as neighborhood demographics. Several studies confirm the role of built environment measures in travel mode choice, such as Cervero (1996), Cervero and Kockelman (1997), and Ewing and Cervero (2001), whereas a few fail to find a significant impact, such as Crane and Crepeau (1998). Standard built environment measures, such as population density and job density are used in this paper, as well as land use measures and pedestrian-oriented intersection density for robustness checks. Transit accessibility is gauged by rail transit station access-route distance. Trip characteristics refer to financial costs, time costs, and distance costs. Financial costs include both gas expenses and parking prices, and both are expected to predict an increase in transit use probability. The expectation for gas expenses arises from the fact that higher gas prices reduce vehicle use [Bento et al.,2009; Gillingham, 2014] and so may increase transit use, and that for parking prices from evidence on a significant impact of parking pricing on driving [Feeney, 1989; Shoup, 2005; Khordagui, 2017; Franco and Khordagui, 2018]. The difference between the trip duration using transit and that using a vehicle is also accounted for, while accounting for the distance of the commute trip separately.

Several reasons motivate the interest in the role of social influence in transit use. First, there is a wealth of literature on the importance of social effects in characterizing people's behavior. Applications in the economics literature include labor market outcomes and hiring [Bayer et al., 2008; Mas and Moretti, 2009; Brown and Laschever, 2012; Cornelissen et al., 2017], education [Sacerdote, 2001; Hoxby, 2000; Hanushek et al., 2003; Angrist and Lang, 2004; Ammermueller and Pischke, 2009], health [Powell et al., 2005; Nakajima, 2007; Trogdon et al., 2008; Fortin and Yazbeck, 2015], crime [Glaeser et al., 1996; Patacchini and Zenou, 2009], finance [Bursztyn et al., 2014; Foucault and Fresard, 2014; Kaustia and Rantala, 2015], and consumption [Grinblatt et al., 2008; Moretti, 2011], to mention a few. In the commute mode choice literature, Walker et al. (2011) find evidence using a discrete choice framework and a dataset from the Netherlands, that the socially-defined reference group significantly affects the transportation mode choice of an individual. Similarly, Kim et al.

(2017) find evidence that inter- and intra-household interactions have a strong impact on transportation mode choice in Cincinnati, while Goetzke (2008) finds that network effects play a role in transportation mode choice in New York. Pike and Lubell (2018) collect data from University of California Davis students and also find that social networks impact transportation mode choice.

Second, understanding the role of social influence in people's decisions is important because such a role may lead to social multipliers (Glaeser et al., 2003), potentially generating multiple equilibria (Glaeser and Scheinkman, 2000). Such an analysis sheds light on the indirect effects of shifting demographics that go beyond the explicit effects. Possible explanations for the existence of such an effect may be contagion, conformity, or social learning (Young, 2009). According to contagion models, commuters may be more likely to use public transit because they met public transit users, whereas according to conformity models a commuter's decision to use transit is a function of the number or proportion of the reference social group using transit. Social learning models, the more prevalent models in the economics literature, would imply that commuters learn about the transit service by observing transit use of the social group, which emphasizes the informational role of social interactions.

Third, social influence is of particular relevance to transit use because many view public transportation as welfare that exists to serve the poor. Indeed, poor commuters are more likely to use transit, and previous research shows that the poor choose to locate where transit is accessible [LeRoy and Sonstelie, 1983; Glaeser et al., 2008]. Therefore, there is reason to believe that some commuters may avoid transit use because of dominance by the poor or by a racial or ethnic minority group. Focusing on a related phenomenon, Schelling (1971) illustrates how residential patterns change as a result of individual choices responding to the location redistribution of the reference group, and he introduces the notion of tipping points for segregation of neighborhoods. Card et al. (2008) use census tract data between 1970 and 2000 to confirm that there is "strong evidence that white population flows exhibit tipping-like behavior in most cities." Moreover, some studies find evidence that migration leads to native-flight from neighborhoods receiving migrants [Accetturo et al., 2014; Mussa et al., 2017]. Therefore, if such impacts are observed for residential choices, then they may potentially be observed for commute mode choice as well.

3.3 Data

3.3.1 Data Sources

The dataset is built by matching data from different sources and constructing variables using GIS techniques. Household travel surveys represent the foundation of the dataset. The household travel surveys used are the California Household Travel Survey (CHTS) for 2012, the Massachusetts Travel Survey (MTS)[1] for 2011, and the New York and New Jersey Regional Household Travel Survey (RHTS) for 2011. The CHTS sample will be referred to hereafter as the CA sample, the MTS as the MA sample, and the RHTS as the NY-NJ sample. The travel surveys are cross-sectional datasets based on travel diaries. Table B.1 in the appendix lists all the Metropolitan Statistical Areas (MSAs) included in the sample. Although the RHTS includes observations from a number of MSAs, the data used for the purposes of this analysis are restricted to only one MSA. The same can be said of the MTS data. Therefore, the NY-NJ and MA samples are each made up of only one MSA (and the observations of the remaining MSAs of the RHTS and the MTS data are added back later on for robustness checks). On the other hand, the CA sample is made up of 5 different MSAs. Household characteristics, commute mode choice, and work and home locations are obtained from the travel surveys. The CHTS and RHTS have work and home locations at the census tract level, and so whenever the work or home locations are used in any GIS analysis for the

[1] Not available online. Acquired by contacting the Massachusetts Department of Transportation.

CA and NY-NJ samples, the geocoded address of the census tract centroid is used. On the other hand, the MTS has work and home locations at the block level, and so the geocoded address of the block centroid is used whenever the work or home locations are used for GIS analysis for the MA sample.

Other sources of data are the American Community Survey (ACS), where some of the demographic variables are obtained, and the Census Transportation Planning Products (CTPP), where the commuting mode breakdowns for both the workplace and residence tracts are obtained. The number of jobs, as well as origin-destination data, are obtained from the Longitudinal Employer-Household Dynamics (LEHD). GIS shapefiles of the rail stations for the relevant years are obtained from the National Transportation Atlas Database (NTAD) published by the Bureau of Transportation Statistics, and supplemented by shapefiles from the Berkeley Library Geodata for the NY-NJ sample. The street maps are obtained from the Topologically Integrated Geographic Encoding and Referencing (TIGER) shapefiles. Time duration for commute trips for different modes are obtained from Google Maps API, and parking prices and gas prices from ParkMe.com and GasBuddy.com respectively. Table B.2 in the appendix lists all variables, and their sources.

3.3.2 Variables

The model explores the impact of a rich set of variables on public transit use. The dependent variable is whether an individual uses transit or not, whether rail or bus. The binary variable takes on a value of 1 if the commuter uses transit and 0 if he drives. Therefore, the sample is limited to transit users and drivers, and so any further reference to commuters refers to transit users and drivers. Along with the socioeconomic characteristics obtained from travel surveys, such as household income and size, and commuter age, sex and race, there are neighborhood characteristics, transit access measures, trip characteristics including time and financial costs, and social influence measures. Neighborhood characteristics accounted for are population density, job density, white percentage of commuters, and the median household income of commuters. All neighborhood characteristics are for both the workplace and residence tracts. For the workplace tract, the measures for the White percentage of commuters and the median household income of commuters are those of workers in the tract of interest, not residents.

Transit access is measured by transit access distance. Transit access distance measures are constructed using GIS tools and distinguish between rail stations and bus stops. A street network is constructed using the TIGER street shapefiles. The exact access routes between either the residence or workplace location (geocoded centroids of census tracts) and the geocoded locations of the nearest rail station and bus stop are constructed. Figure 3.1 shows an illustration of the construction of the access routes. The commute trip characteristics include a variety of measures. One measure is the distance of the trip, which is measured by the Euclidean distance between the residence and workplace. The objective of using this measure is to learn how the proximity of the workplace to the residence impacts transit use rather than learn how the distance traveled using a particular commute mode impacts transit use. Using driving distance, from Google Maps API, for example, although highly correlated, would incorporate a driving-specific cost. Furthermore, the trip distance reported in some travel surveys is not used because that measure is a function of the commute mode, which would render the estimate highly endogenous.

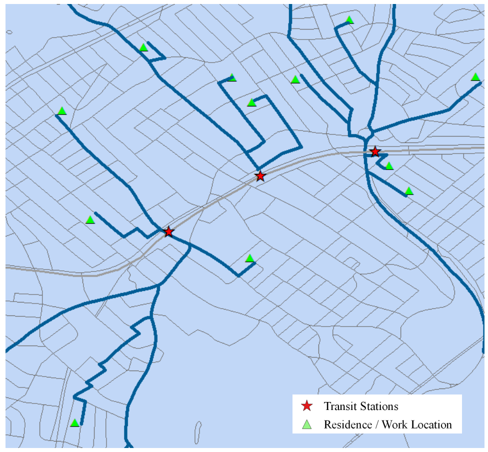

Figure 3.1: Transit Access Route Illustration

Another trip characteristic examined is the additional time cost incurred when using public transit. This is measured by computing the difference between the trip duration using transit and that when driving. This measure is mostly positive but not necessarily always so, because in some cases it may take less time to arrive to work using transit than driving. Trip duration are obtained from Google Maps API, which offers only current (or future) measures. Therefore, an important assumption made while using this measure is that the relative trip durations of 2018 closely mimic those of 2011 and 2012.

Financial costs associated with a trip are also measured. However, only costs associated with driving are measured for lack of accurate and prevalent data on transit fares from residence to workplace. The costs considered are parking prices and gas expenses. Parking prices are obtained for the workplace location at the zipcode level. An average of the daily price is computed for every zipcode using posted daily prices when available, while computing them from hourly prices otherwise. For areas with no parking prices, the daily parking price is set to $0. Gas prices are obtained for both the workplace and residential zipcode. The average gas price for all zipcodes is computed, and then the workplace and residential zipcode prices are compared to determine the lowest price for each commuter. The lowest gas price is then used to compute gas expenses, using driving distance obtained from Google Maps API. An average of 22 miles per gallon is assumed for the computation of the gas expense. Once more, parking price, gas prices, and Google Maps API are current data (2018), and the assumption is that the magnitudes and distributions of the measures in 2018 closely reflect those of 2011 and 2012.

Measures for social influence are divided into three different groups: spatial, income-group, and racial. These measures are all constructed for both the workplace and residence tracts, and are based on CTPP data. The CTPP data are special tabulations of the ACS that facilitate transportation analyses. One interesting feature is that certain breakdowns of transportation modes are given at the workplace as well as at the residence. The data at the workplace give the transportation mode of people who work in the specific census tract, whereas those at the residence give the transportation mode choice of people living in the tract of interest. This is advantageous because the analysis is not only exploring the impact in the residential tract, but the impact in the workplace as well. Therefore, these data are matched to the travel survey data such that the workplace tabulations are matched to the workplace tracts, and residence tabulations to the residence tracts. Subsequently, the CTPP residence measures are used for the purpose of calculating the social influence variables at the residence census tract, whereas the CTPP workplace measures are used for the workplace calculations.

The spatial social influence is measured as the percentage of commuters in a given tract who are public transit users. Spatial influence is going to be referred to in the model as SN_i^R and SN_i^W where the R and W superscripts correspond to the residence and workplace tracts respectively. Income group influence is referred to in the model as YN_i^R and YN_i^W. The median household income of public transit users is used as a measure for income-group influence, and is referred to as Median Transit Income. This is used to determine whether the income composition of the transit users affects the decision of a particular individual to take public transportation. Median Transit Income is expected to have a positive impact on

Table 3.1: Social Influence Variables

Variable Name	Variable Measure
Spatial Social Influence	
Transit Use Rate^R (SN_i^R)	$\frac{Transit\ Users\ in\ Residence\ Tract}{Commuters\ in\ Residence\ Tract} \times 100\%$
Transit Use Rate^W (SN_i^W)	$\frac{Transit\ Users\ in\ Workplace\ Tract}{Commuters\ in\ Workplace\ Tract} \times 100\%$
Income-Group Social Influence	
Median Transit Income^R (YN_i^R)	*Median Income of Transit Users in Residence Tract*
Median Transit Income^W (YN_i^W)	*Median Income of Transit Users in Workplace Tract*
Racial Social Influence	
Transit White Rate^R (RN_i^R)	$\frac{White\ Transit\ Users\ in\ Residence\ Tract}{All\ Transit\ Users\ in\ Residence\ Tract} \times 100\%$
Transit White Rate^W (RN_i^W)	$\frac{White\ Transit\ Users\ in\ Workplace\ Tract}{All\ Transit\ Users\ in\ Workplace\ Tract} \times 100\%$

high income commuters. Interaction terms are utilized to distinguish the different effects. Interaction terms interact the Median Transit Income with a binary variable, $RICH_i^R$ or $RICH_i^W$, that takes on the value 1 if the commuter has a household income above the median of his tract. The racial social influence is referred to in the model as RN_i^R and RN_i^W. The racial measures are in terms of whites and minorities (nonwhites). The percentage of white transit users, which is referred to as the Transit White Rate, is used as a measure for the racial influence effect. This measure is interacted with a binary variable, $WHITE_i$, indicating whether the commuter is white, in order to identify the different effects. Table 3.1 summarizes all the social influence variables discussed.

3.3.3 Descriptive Statistics

Tables 3.2 and 3.3 presents the overall descriptive statistics of the dataset, while Table 3.4 presents those of the social influence measures. The full sample has over 20,000 commuters, and is made up of only those who either drive or use transit. Out of all these commuters, around 22% use transit, while the rest drive. The percentage of transit users varies from one sample to another, with the lowest percentage in the CA sample at 10.5% and the highest in NY-NJ at 38%. The average household income is $116,000, and although 68% of the commuters in the full sample are White, the percentage ranges from 55% in CA to 87% in MA.

The average route distance to the nearest rail transit station is 2.8 miles for the residence tract and 3.3 miles for the workplace tract, with the shortest being in NY-NJ at 1.9 miles for the residence tract and 1.4 for the workplace. The average Euclidean distance from residence to workplace is 7.9 miles, while on average, commuters would spend 46 additional minutes commuting if they use transit. Note that this time includes transit access and egress times, while the driving time does not include any time needed to get from a parking location to work. The average daily parking price in the full sample is $8.5, and ranges from as low as $3.3 in CA to $14.6 in NY-NJ. Note that these averages are of parking prices the commuters in the sample would face, and do not necessarily reflect the average price of parking over all the MSAs. Gas prices in the full sample are $3.2 on average, ranging from $2.9 in MA and NY-NJ up to $3.6 in CA. The gas expense is estimated to be $1.6 on average.

Table 3.4 shows that transit use rate in the workplace tract (20%) is higher than that in the residence tract (16.7%). Moreover, the Median Transit Income is consistently higher in the residence tract than in the workplace tract. Although the Median Transit Income ($104,309) is lower than the overall Median Commuter Income ($107,425) in the residence tract, the difference is much larger between the Median Transit Income ($73,430) and the Median Commuter Income ($94,381) in the workplace tract. The Transit White Rate is also higher in the residence tract (50.7%) than in the workplace tract (33.5%), and the deviation from the Commuter White Rate seems to be more pronounced for the workplace tract where there is almost a 20 percentage point difference compared to a 7.4 percentage point difference

Table 3.2: Descriptive Statistics: Sample & Socioeconomic and Neighborhood Characteristics

Variable	**Full**	**CA**	**NY-NJ**	**MA**
Sample Size	20,068	8,075	6,452	5,541
Transit Use (%)	21.9	10.5	38	19.5
	(41.4)	(30.9)	(48.6)	(39.7)
Socioeconomic Characteristics				
Household Income ($)	116,004	115,474	107,377	126,821
	(72,837)	(74,448)	(65,219)	(77,340)
Household Size	2.9	3	2.7	3
	(1.3)	(1.4)	(1.3)	(1.3)
Age	46.4	45.8	46.7	47
	(13.8)	(14.7)	(12.1)	(14.4)
White (%)	68.1	54.7	69	86.6
	(46.6)	(49.8)	(46.2)	(34.1)
Female (%)	49.3	46.2	50.1	53
	(50.0)	(50.0)	(50.0)	(50.0)
Above Median Income^R	48.2	49.7	44.2	50.7
	(50.0)	(50.0)	(50.0)	(50.0)
Above Median Income^W	54.6	58.5	50.2	54.1
	(49.8)	(49.3)	(50.0)	(50.0)
Neighborhood Characteristics				
Population Density^R (per m^2)	0.006	0.004	0.010	0.003
	(0.009)	(0.003)	(0.014)	(0.004)
Population Density^W (per m^2)	0.006	0.003	0.010	0.004
	(0.008)	(0.003)	(0.012)	(0.004)
Job Density^R (per m^2)	0.002	0.001	0.003	0.001
	(0.015)	(0.004)	(0.024)	(0.013)
Job Density^W (per m^2)	0.031	0.010	0.053	0.038
	(0.081)	(0.026)	(0.106)	(0.092)
Commuter White Rate^R (%)	58.1	44.3	58.8	77.2
	(28.9)	(25.4)	(29.2)	(21.3)
Commuter White Rate^W (%)	53.6	39.8	51.7	75.9
	(21.3)	(17.2)	(16.9)	(10.3)
Median Commuter Income^R ($)	107,425	104,822	107,649	110,956
	(36,944)	(39,335)	(35,364)	(34,788)
Median Commuter Income^W ($)	94,381	87,683	97,058	101,024
	(19,305)	(22,361)	(17,005)	(12,906)

Standard errors in parentheses

Table 3.3: Descriptive Statistics: Transit Access & Trip Characteristics

Variable	**Full**	**CA**	**NY-NJ**	**MA**
Transit Access				
Distance to Rail^R (mi)	3.3	5.6	1.9	1.7
	(4.9)	(6.7)	(2.4)	(1.8)
Distance to Rail^W (mi)	2.8	4.8	1.4	1.4
	(4.6)	(6.2)	(2.2)	(1.8)
Trip Characteristics				
Distance (mi)	7.9	8.7	8.5	6.3
	(7.4)	(7.5)	(8.1)	(6.0)
Time Difference (min)	46.4	53.7	42.5	40.3
	(33.5)	(32.1)	(34.2)	(32.6)
Parking Prices ($/day)	8.5	3.3	14.6	9.1
	(15.9)	(9.8)	(20.4)	(14.2)
Gas Price ($/gallon)	3.2	3.6	2.9	2.9
	(0.4)	(0.1)	(0.1)	(0.1)
Gas Expense ($)	1.6	2.0	1.6	1.2
	(1.4)	(1.6)	(1.4)	(1.1)

Standard errors in parentheses

Table 3.4: Social Influence Descriptive Statistics

Variable	**Full**	**CA**	**NY-NJ**	**MA**
Spatial Social Influence				
Transit Use Rate^R (%)	16.7	8.4	28.9	14.8
	(22.7)	(11.3)	(30.9)	(17.1)
Transit Use Rate^W (%)	20.0	9.1	35.9	17.3
	(26.3)	(13.1)	(34.8)	(19.1)
Income-Group Social Influence				
Median Transit Income^R ($)	104,309	93,332	110,705	111,306
	(54,672)	(56,086)	(55,372)	(49,010)
Median Transit Income^W ($)	73,430	70,789	73,864	77,514
	(32,751)	(36,406)	(27,934)	(31,521)
Racial Social Influence				
Transit White Rate^R (%)	50.7	36.5	49.8	72.2
	(35.5)	(34.5)	(33.4)	(28.2)
Transit White Rate^W (%)	33.5	21.4	29.1	54.0
	(29.4)	(29.6)	(22.2)	(26.4)

Standard errors in parentheses

in the residence tract. The overall Commuter White Rate and Income in both the residence and workplace tracts are controlled for in the model to account for these differences.

3.4 Empirical Model

3.4.1 Fundamental Model

Public transit use is modeled as a function of social influence $\boldsymbol{S}_i^R$ and $\boldsymbol{S}_i^W$ (where the superscripts indicate *residence* and *workplace* respectively), socioeconomic characteristics $\boldsymbol{X_i}$, neighborhood characteristics $\boldsymbol{B}_i^R$ and $\boldsymbol{B}_i^W$, transit accessibility A_i^R and A_i^W, trip characteristics $\boldsymbol{C}_i$, and MSA fixed effects ν_i. Parking prices and gas expenses are modeled in logarithmic form. Let T_i^* be the commuter's unobserved propensity to use public transit, and T_i be a binary variable taking the value 1 if the commuter uses transit, and 0 otherwise (drives). The latent variable T_i^* and its relationship with the discrete binary variable T_i can be represented as follows:

$$T_i^* = f(\boldsymbol{S}_i^R, \boldsymbol{S}_i^W, \boldsymbol{X}_i, \boldsymbol{B}_i^R, \boldsymbol{B}_i^W, A_i^R, A_i^W, \boldsymbol{C}_i, \nu_i) \tag{3.1}$$

$$T_i = \begin{cases} 1 & \text{if } T_i^* > 0 \\ 0 & \text{if } T_i^* \leq 0 \end{cases} \tag{3.2}$$

A binary probit model is used to estimate the relationship. In other words, the model to be estimated is specified as:

$$Pr[T_i = 1 \mid \boldsymbol{S}_i^R, \boldsymbol{S}_i^W, \boldsymbol{X}_i, \boldsymbol{B}_i^R, \boldsymbol{B}_i^W, A_i^R, A_i^W, \boldsymbol{C}_i, \nu_i] = \Phi(\boldsymbol{S}_i^R, \boldsymbol{S}_i^W, \boldsymbol{X}_i, \boldsymbol{B}_i^R, \boldsymbol{B}_i^W, A_i^R, A_i^W, \boldsymbol{C}_i, \nu_i) \tag{3.3}$$

The first binary model to be estimated will be one that examines the roles of all the factors of interest except social influence. The purpose of this estimation is to offer an overview of the impact of the fundamental factors believed to directly affect transit use. Therefore, the first model to be estimated models the latent variable T_i^* as follows:

$$T_i^* = \boldsymbol{X}_i\boldsymbol{\beta}_{\boldsymbol{X}} + \boldsymbol{B}_i^R\boldsymbol{\beta}_{\boldsymbol{B}^R} + \boldsymbol{B}_i^W\boldsymbol{\beta}_{\boldsymbol{B}^W} + \beta_{A^R}A_i^R + \beta_{A^W}A_i^W + \boldsymbol{C}_i\boldsymbol{\beta}_{\boldsymbol{C}} + \nu_i + \varepsilon_i \tag{3.4}$$

3.4.2 Spatial Social Influence

Model

The second model to be estimated builds on equation 3.4, and adds the spatial social influence variables SN_i^R and SN_i^W for both the residence and workplace. Let

$$\boldsymbol{H_i} = \begin{bmatrix} \boldsymbol{X}_i & \boldsymbol{B}_i^R & \boldsymbol{B}_i^W & A_i^R & A_i^W & \boldsymbol{C}_i \end{bmatrix}$$

$$\boldsymbol{\beta_H} = \begin{bmatrix} \boldsymbol{\beta_X} & \boldsymbol{\beta}_{\boldsymbol{B}^R} & \boldsymbol{\beta}_{\boldsymbol{B}^R} & \beta_{A^R} & \beta_{A^W} & \boldsymbol{\beta_C} \end{bmatrix}'$$

T_i^* can now be modeled as:

$$T_i^* = \alpha^R SN_i^R + \alpha^W SN_i^W + \boldsymbol{H}_i\boldsymbol{\beta}_H + \nu_i + \varepsilon_i \tag{3.5}$$

where α^R and α^W are hypothesized to be positive, reflecting positive neighborhood peer effects.

Endogeneity

There are several sources of potential endogeneity. First, unobserved heterogeneity at the tract level could result in biased estimates if it happens to be correlated with the social influence measures. A particular unobserved feature of the neighborhood may be responsible for increasing public transit use of everyone in the neighborhood, which would be reflected in a higher transit use rate as well as a greater probability of an observed commuter being a public transit user. In such a case, the estimated coefficients would be partially reflecting the effect of these unobserved factors rather than the effect of social influence. Second, neighborhood sorting may also lead to biased estimates. People with similar preferences or characteristics may choose to live or work in the same neighborhoods, and these preferences and characteristics may be responsible for their commute mode choices rather than social influence. Third, exploring social influence gives rise to the reflection problem discussed by Manski (1993).

The first two sources of potential endogeneity are the ones of concern for this analysis. The third source, the reflection problem, is not a concern because the social influence variable of interest is a field variable which is estimated for the all residents or workers in a given tract. Therefore, it is reasonable to assume that it is unlikely that the commute mode choice of a single commuter would affect the choices of everyone living or working in the neighborhood. Although the analysis controls for a rich set of neighborhood and household characteristics, the first two sources of potential endogeneity are still a concern because the control variables may not capture everything. To address the potential endogeneity, an IV is constructed.

Instrumental Variable

The IV used is the weighted average of the Euclidean commute trip distance for every neighborhood. In this paper, as in previous studies, the commute trip distance is a significant determinant of public transit use. It is reasonable to believe that the commute distance of an individual's neighbors or coworkers does not influence the individual's choice of commute mode except through its influence on the neighbors' or coworkers' choice of commute mode. However, the commute distance itself of an individual may be correlated with that of the neighbors or coworkers. Nevertheless, this does not affect the exclusion restriction of the IV because the individual's commute distance is in fact controlled for in the model. Therefore, the weighted average of the commute distance of neighbors or coworkers is a suitable IV for their public transit use.

The IV is constructed by using the LEHD Origin-Destination (OD) data. The LEHD OD files provide the number of workers in each pair of home and work neighborhoods, where each neighborhood is a census block. To construct the IV, the data are first aggregated to the census tract level since this is the definition of the neighborhood in this study, and then the destination tracts for each origin tract are identified. The Euclidean distance is then computed between the centroid of the origin tract and each of the centroids of all the associated destination tracts. Therefore, for every origin tract j, where $j = 1, ..., J$ and $J =$ number of origin tracts, a Euclidean distance ED^O_{j,k_j} is computed, where $k_j = 1, ..., K_j$ and $K_j =$ number of destination tracts paired with tract j. Thus, for each origin tract, there are K_j associated Euclidean distances, ED^O_{j,k_j}.

A weighted average of these distances is then computed, with the weights being the corresponding fraction of total commuters living in the origin tract and commuting to a given destination tract. That is, for every origin tract j, weights, w^O_{j,k_j}, for all destination tracts k are computed, such that

$$w^O_{j,k_j} = \frac{workers^O_{j,k_j}}{workers^O_j}$$

where $workers^O_j$ is the total number of workers commuting from tract j, and $workers^O_{j,k_j}$ is the number of workers commuting from tract j to tract k_j. The weighted average Euclidean distance for origin tracts, ED_j^O is then computed as:

$$ED_j^O = \sum_{k_j=1}^{K_j} w_{j,k_j}^O \times ED_{j,k_j}^O$$

Each origin tract now has a weighted average distance variable, which is used as an IV for the residence tract transit use rate. Hence, Z_i^R, the IV for SN_i^R, is matched to the corresponding ED_j^O, such that

$$Z_i^R = ED_j^O \quad for \quad R = O$$

The algorithm is repeated to construct an IV for the workplace social influence variable, where the origin neighborhoods, j_k, for each destination neighborhood, k, are identified, and then the Euclidean distances, ED_{k,j_k}^D, between each destination tract centroid and the associated origin tracts centroids are computed. The weighted average of these distances, ED_k^D, are then calculated, where the weights, w_{k,j_k}^D reflect the corresponding fraction of total commuters working in the destination tract and commuting from a given origin tract. Therefore,

$$w_{k,j_k}^D = \frac{workers_{k,j_k}^D}{workers_k^D}$$

$$ED_k^D = \sum_{j_k=1}^{J_k} w_{k,j_k}^D \times ED_{k,j_k}^D$$

$$Z_i^W = ED_k^D \quad for \quad W = D$$

where $workers_k^D$ is the total number of workers commuting to tract k, $workers_{k,j_k}^D$ is the number of workers commuting from tract j_k to tract k, and Z_i^W is the IV for SN_i^W.

The model to be estimated is now:

$$T_i^* = \boldsymbol{SN}_i\boldsymbol{\alpha} + \boldsymbol{H}_i\boldsymbol{\beta}_H + \nu_i + \varepsilon_i \tag{3.6}$$

$$\boldsymbol{SN}_i^{'} = \boldsymbol{Z}_i\boldsymbol{\Pi}_Z + \boldsymbol{H}_i\boldsymbol{\Pi}_H + \nu_i + \boldsymbol{\mu}_i \tag{3.7}$$

where $\boldsymbol{SN}_i = \begin{bmatrix} SN_i^R & SN_i^W \end{bmatrix}$, $\boldsymbol{\alpha} = \begin{bmatrix} \alpha^R & \alpha^W \end{bmatrix}'$, $\boldsymbol{Z}_i = \begin{bmatrix} Z_i^R & Z_i^W \end{bmatrix}'$, Z_i^R and Z_i^W represent the exogenous variables (the constructed average Euclidean commute distance) for the residence and workplace tracts respectively, and $\boldsymbol{\Pi}$ represents the reduced form coefficients. The error terms, ε_i and $\boldsymbol{\mu}_i$, are assumed to be jointly normally distributed, and the model is estimated using Maximum Likelihood Estimation.

3.4.3 Income-Group Social Influence

Model

To estimate the effect of income-group social influence, the model builds on equation 3.5, adding the income-group influence variables for the residence and workplace tracts YN_i^R and YN_i^W, binary variables indicating whether the commuter has a household income above the median household income of commuters in the corresponding tract and denoted by $RICH_i^R$ and $RICH_i^W$ respectively, and interaction terms of the income-group influence variables and the corresponding $RICH_i$ denoted by $YN_i^R \times RICH_i^R$ and $YN_i^W \times RICH_i^W$. Therefore, to explore the existence of income-group social influence in public transit use decisions, T_i^* is modeled as:

$$T_i^* = \boldsymbol{SN}_i\boldsymbol{\alpha} + \boldsymbol{YN}_i\boldsymbol{\gamma} + \boldsymbol{YNRICH}_i\boldsymbol{\gamma}_r + \boldsymbol{RICH}_i\boldsymbol{\delta}_r + \boldsymbol{SNRICH}_i\boldsymbol{\alpha}_r + \boldsymbol{H}_i\boldsymbol{\beta}_H + \nu_i + \varepsilon_i \quad (3.8)$$

where:

$\boldsymbol{YN}_i = \begin{bmatrix} YN_i^R & YN_i^W \end{bmatrix}$,

$\boldsymbol{YNRICH}_i = \begin{bmatrix} YN_i^R \times RICH_i^R & YN_i^W \times RICH_i^W \end{bmatrix}$,

$\boldsymbol{RICH}_i = \begin{bmatrix} RICH_i^R & RICH_i^W \end{bmatrix}$,

$\boldsymbol{SNRICH}_i = \begin{bmatrix} SN_i^R \times RICH_i^R & SN_i^W \times RICH_i^W \end{bmatrix}$,

$$\boldsymbol{\gamma}_r = \begin{bmatrix} \gamma_r^R & \gamma_r^W \end{bmatrix}',$$

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma^R & \gamma^W \end{bmatrix}',$$

$$\boldsymbol{\delta}_r = \begin{bmatrix} \delta_r^R & \delta_r^W \end{bmatrix}', \text{ and}$$

$$\boldsymbol{\alpha}_r = \begin{bmatrix} \alpha_r^R & \alpha_r^W \end{bmatrix}'.$$

Identification Strategy

The income-group social influence variable refers to the income-group within a specific tract, whether the residence or workplace tract. Therefore, the income-group social influence variable can be thought of as two-dimensional; one dimension being spatial, and the other income-group-related. To ensure that $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_r$ are in fact reflecting income-group social influence, rather than also reflecting spatial social influence, the spatial social influence variables, $\boldsymbol{SN}_i$, are also included in the regression to control for social influence specific to the neighborhoods.

There are three potential sources of identification threats: unobserved heterogeneity that is income-group-specific, unobserved heterogeneity (or sorting) that is neighborhood-specific, or unobserved heterogeneity that is specific to income groups in particular neighborhoods. The third source could be thought of as the intersection of income-group-specific heterogeneity with neighborhood-specific heterogeneity, since neighborhood-specific heterogeneity could vary across the income groups.

To control for unobserved heterogeneity across income groups as well as potential bias that may arise because of sorting into certain income groups, an income-group fixed effect is included in the regression, which is represented by $\boldsymbol{RICH}_i$ in equation 3.8. The presence of $\boldsymbol{SN}_i$ plays an additional role in that it controls for any unobserved heterogeneity across tracts that impacts transit use, and captures potential neighborhood sorting bias promoting transit use. Therefore, $\boldsymbol{SN}_i$ can be thought of as controlling for potential neighborhood-specific heterogeneity promoting transit use. To account for the potential intersection of neighborhood-specific and income-group-specific heterogeneity, an interaction term of the income-group fixed effect and $\boldsymbol{SN}_i$ is included in the estimation ($\boldsymbol{SNRICH}_i$). As a result, the estimates for $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_r$ are purged from neighborhood-related bias impacting transit users, income-group-related bias, and bias specific to the combined effect of income and neighborhood transit use. Hence, it is reasonable to assume that $\boldsymbol{\gamma}$ in fact reflects income-group influence.

Interpretation

In equation 3.8, $\boldsymbol{\gamma}$ reflects the income-group influence for the poor group, defined as those whose household income lies below the median household income of commuters in the relevant tract. A higher median transit user income does not necessarily reflect increasing rich transit use, but reflects increasing rich transit use relative to poor transit use. The distinction between the impact on the rich and the poor is made using an interaction term. Note that $RICH_i$ is defined twice: once relative to the median household income in the residence tract, and another relative to that in the workplace tract. These two $RICH_i$ variables are correlated, but are not necessarily equal. Moreover, the variable $RICH_i$ is not defined relative to the median household income of transit users, which means that the definition of $RICH_i$ does not vary as the median income of transit users varies. The income-group influence for the rich commuters, $\frac{\partial T_i^*}{\partial YN_i^R}$ and $\frac{\partial T_i^*}{\partial YN_i^W}$, are represented by $\gamma^R + \gamma_r^R$ and $\gamma^W + \gamma_r^W$. The $\boldsymbol{\gamma}_r$ reflects whether the effect on the rich group is significantly different from that on the poor group. Therefore, $\boldsymbol{\gamma}_r$ can be negative or 0 and there would still be positive income-group influence for the rich if $\boldsymbol{\gamma} + \boldsymbol{\gamma}_r$ is still positive. In such a case, a negative $\boldsymbol{\gamma}_r$ would only indicate that rich commuters are less responsive to the income-group influence than the poor group.

The expectation is that $\frac{\partial T_i^*}{\partial YN_i^R} \geq 0$ and $\frac{\partial T_i^*}{\partial YN_i^W} \geq 0$ for rich commuters. That is, rich commuters are expected to prefer to have more rich commuters relative to poor ones using transit (as reflected by a higher Median Transit Income), and are unlikely to be inspired to use transit when the poor transit user population increases in relative size (as would be reflected by a lower Median Transit Income). However, the income-group influence for the poor group may not necessarily work that way. It may be the case that poor commuters are more likely to use transit the greater the percentage of transit users who are poor. In other words, $\frac{\partial T_i^*}{\partial YN_i^R} \leq 0$ and $\frac{\partial T_i^*}{\partial YN_i^W} \leq 0$ for the poor. That is, as the median income falls (in other words, more poor relative to rich), more poor commuters are likely to use transit. On the other hand, the poor may favor a more integrated transit system and respond to a higher median transit user income by using transit as well. This would be reflected by $\frac{\partial T_i^*}{\partial YN_i^R} > 0$ and $\frac{\partial T_i^*}{\partial YN_i^W} > 0$ for the poor. For the poor commuter, γ^R and γ^W represent $\frac{\partial T_i^*}{\partial YN_i^R}$ and $\frac{\partial T_i^*}{\partial YN_i^W}$ respectively.

3.4.4 Racial Social Influence

Model

The model for the racial social influence also builds on equation 3.5. The racial influence variables RN_i^R and RN_i^W, for the residence and workplace respectively, are added to equation 3.5, and so are the interaction terms of these racial influence variables with the binary variable $WHITE_i$ that indicates whether a commuter is white. The interaction terms are denoted by $RN_i^R \times WHITE_i$ and $RN_i^W \times WHITE_i$. T_i^* can, thus, be modeled as:

$$T_i^* = \boldsymbol{SN}_i\boldsymbol{\alpha} + \boldsymbol{RN}_i\boldsymbol{\lambda} + WHITE_i \times \boldsymbol{RN}_i\boldsymbol{\lambda}_{wh} + WHITE_i \times \boldsymbol{SN}_i\boldsymbol{\alpha}_{wh} + \boldsymbol{H}_i\boldsymbol{\beta}_H + \nu_i + \varepsilon_i \quad (3.9)$$

where:

$\boldsymbol{RN}_i = \begin{bmatrix} RN_i^R & RN_i^W \end{bmatrix}$, $\boldsymbol{\lambda} = \begin{bmatrix} \lambda^R & \lambda^W \end{bmatrix}'$, $\boldsymbol{\lambda}_{wh} = \begin{bmatrix} \lambda_{wh}^R & \lambda_{wh}^W \end{bmatrix}'$, and $\boldsymbol{\alpha}_{wh} = \begin{bmatrix} \alpha_{wh}^R & \alpha_{wh}^W \end{bmatrix}'$.

Identification Strategy

The identification strategy for racial social influence is identical to that for the income-group social influence. There are three potential sources of identification threats: unobserved heterogeneity that is race-group-specific, unobserved heterogeneity (or sorting) that is neighborhood-specific, or unobserved heterogeneity that is specific to race groups in particular neighborhoods. Neighborhood-specific heterogeneity promoting transit users is controlled for by the inclusion of the spatial influence measure, $\boldsymbol{SN}_i$, and race-group-specific heterogeneity is controlled for by $WHITE_i$, which can be thought of as a race-group fixed effect. $WHITE_i$ is included in $\boldsymbol{H}_i$. Heterogeneity specific to the intersection of the race-group and the neighborhood is controlled for by the interaction of the race-group fixed effect, $WHITE_i$, and the spatial social influence measure, $\boldsymbol{SN}_i$. Therefore, $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}_{wh}$ represent consistent estimates of racial social influence.

Interpretation

In general, the white commuters are expected to have preferences that are increasing in white transit use if racial social influence exists. Therefore, the expectations are $\frac{\partial T_i^*}{\partial RN_i^R} \geq 0$ and $\frac{\partial T_i^*}{\partial RN_i^W} \geq 0$ for whites. The nonwhite commuters may also be expected to have preferences that are increasing in nonwhite transit use, in which case $\frac{\partial T_i^*}{\partial RN_i^R} \leq 0$ and $\frac{\partial T_i^*}{\partial RN_i^W} \leq 0$ for the nonwhite, but may also have preferences that are increasing in white transit use, in which case $\frac{\partial T_i^*}{\partial RN_i^R} > 0$ and $\frac{\partial T_i^*}{\partial RN_i^W} > 0$. The λ^R and λ^W in equation 3.9 represent the nonwhite commuter's $\frac{\partial T_i^*}{\partial RN_i^R}$ and $\frac{\partial T_i^*}{\partial RN_i^W}$ respectively. On the other hand, the white commuter's $\frac{\partial T_i^*}{\partial RN_i^R}$ and $\frac{\partial T_i^*}{\partial RN_i^W}$ are computed as $\lambda^R + \lambda_{wh}^R$ and $\lambda^W + \lambda_{wh}^W$ respectively. As with income-group influence, a negative $\boldsymbol{\lambda}_{wh}$ does not necessarily indicate a negative racial influence effect on the white commuter.

3.5 Results

3.5.1 Fundamental Determinants

The model is estimated with the fundamental determinants only first, before adding the second-order effects of social influence. This step is done to gauge the model before introducing the new factors, and to take a closer look at the effect of variables that have not been examined across several major US cities all at once, such as parking prices, gas expenses, commute time difference, and rail transit access route distance. A binary probit model is estimated, where the binary depended variable is defined as in equation 3.2 and T^* by equation 3.4. The average marginal effects are presented in Table 3.5.

Household income is a statistically significant determinant of public transit use. The results indicate that a \$10,000 increase in household income reduces the probability of transit use by 0.17 percentage points. This is consistent with the literature and expectations about people's response to higher income. As income increases, commuters tend to switch to the more expensive mode. An increase in household size also reduces a commuter's probability of using transit. The probability of transit use drops by 0.77 percentage points with every additional household member. The effect does not exist in CA, though, but is larger in NY-NJ and MA, where an additional household member reduces the probability of public transit use by 1.4 percentage points. This is likely to be the case because a bigger household is more likely to involve children who may need to be driven around.

Age is also statistically significant, and an additional year of age reduces the probability of using public transit use by 0.08 percentage points. Examining the individual samples, though, it seems that age is insignificant in the CA sample. White commuters are 2.8 percentage points less likely to use transit, holding everything else constant. This effect is absent in the CA sample, but is stronger in NY-NJ where the marginal effect is 4.2 percentage

Table 3.5: Determinants of Transit Use (Eq. 3.4)

Dependent Variable: T_i	**Full**	**CA**	**NY-NJ**	**MA**
Socioeconomic Characteristics, $\boldsymbol{X}_i$				
Income	-0.0017***	-0.0019***	-0.0027***	-0.0014**
Household Size	-0.0077***	0.0029	-0.0141***	-0.0137***
Age	-0.0008***	-0.0004	-0.0005*	-0.0017***
White	-0.0282***	-0.0106	-0.0420***	-0.0257*
Female	-0.0034	0.0084	0.0058	-0.0298***
Neighborhood Characteristics, $\boldsymbol{B}_i$				
Population Density^R	4.7227***	4.0244***	3.8132***	6.3553***
Population Density^W	2.3899***	3.8219***	1.4180***	3.5031***
Job Density^R	0.4190**	0.2688	1.2801***	0.3646*
Job Density^W	0.3697***	1.0094***	0.3747***	0.1940***
Commuter White Rate^R	-0.0001	0.0003	-0.0004**	0.0006**
Commuter White Rate^W	-0.0005**	0.0002	-0.0005	-0.0016***
Median Commuter Income^R	-0.0001	-0.0023	0.0024*	-0.0025
Median Commuter Income^W	0.0040***	0.0006	-0.0003	0.0266***
Transit Access A_i				
Distance to Rail^R	-0.0026**	-0.0020**	-0.0061**	-0.0200***
Distance to Rail^W	-0.0059***	-0.0020*	-0.0214***	-0.0800***
Trip Characteristics, $\boldsymbol{C}_i$				
Distance	0.0065***	0.0029***	0.0051***	0.0088***
Time Difference	-0.0022***	-0.0017***	-0.0022***	-0.0023***
Parking Price	0.0202***	0.0096***	0.0238***	0.0172***
Gas Expense	0.0199***	0.0522***	0.0241***	0.0081*
MSA Fixed Effects	Yes	Yes	No	No
Observations	20068	8075	6452	5541
Pseudo-R^2	0.4785	0.3177	0.5798	0.4109
Wald χ^2	4523	910	2286	1026
Log Likelihood	-5515	-1874	-1803	-1615

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

A constant is included

points instead. Women are less likely to use public transit in MA only, where the probability of transit use drops by 3 percentage points. Overall, though, gender is not a significant determinant of transit use.

Population density, a standard urban form measure, is significant at both the workplace and residence. For every additional person per 100 square-meter in the residence tract, the probability of transit use increases by 4.7 percentage points, whereas the same increase in the workplace tract leads to an increase in probability of 2.4 percentage points. The results are significant across all samples, but the size of the marginal effect varies, going to as high as 6.4 percentage points for residence tracts in MA. Furthermore, job density is a significant determinant of transit use. For every extra job per 100 square-meter in the residence tract, the probability of transit use increases by 0.42 percentage points, whereas the same increase in the workplace tract raises the probability of using transit by 0.37 percentage points. For workplace job density, the results are significant across the individual samples, and the marginal effects range from 0.19 percentage points in MA to 1.01 in CA for every additional job per 100 square-meter. Job density in the residence tract, however, is insignificant in CA.

For the full sample, racial composition of the commuting population in the residence tract is not a significant determinant of public transit use. However, this results does not hold for all individual samples. For example, a 1 percentage point increase of white commuters in the residence tract reduces the probability of public transit use in NY-NJ by no more than 0.04 percentage points, while increasing the probability in MA by 0.06 percentage points. The negative marginal effect for the NY-NJ sample implies that in neighborhoods with a greater percentage of white commuters, an individual commuter is less likely to use public transit regardless of his or her own race. This is not surprising since we have already seen that white commuters are less likely to use public transit. However, the estimates for the MA sample may imply that, in MA, a more racially homogeneous residential neighborhood encourages public transit use. On the other hand, a 1 percentage point increase in the workplace tract reduces transit use probability in the full sample by 0.05 percentage points. The result does not seem to hold for the CA and NY-NJ samples, though.

The findings on the impact of overall income level of commuters in a neighborhood on public transit use are mixed. For the residential neighborhood, overall income of commuters does not affect public transit use. The only exception shows up in the NY-NJ results, where an additional $10,000 in median commuter household income raises the probability of using transit by 0.2 percentage points. Examining the results for the workplace neighborhood commuter income level reveals that, overall, a higher income level increases the probability of public transit use by 0.4 percentage points for every additional $10,000. This result, however, does not hold across the individual samples. The MA sample is the only individual sample for which commuter income in the workplace tract is significant. In fact, a $10,000 increase in commuter income raises the probability of transit use by 2.7 percentage points.

The longer the access route to the nearest rail transit station, the less likely a commuter is to use public transit. The evidence suggests that for every additional half-mile from the residence, the probability of using transit drops by 0.13 percentage points. Although the results are significant across all samples, the marginal effect can go up to as high as 1 percentage points (MA) for the same increase in access route distance. The workplace rail transit access distance seems to matter more than the residence one, with probability falling by 0.27 percentage point with every additional half-mile. The effect holds across all samples. The marginal effect for the MA sample is largest in magnitude, and is in fact 4 percentage points for an additional half-mile of rail transit station access distance.

Distance to work is highly significant in every sample. The evidence suggests that for every additional mile, the probability of using public transit increases by 0.65 percentage points. The magnitude of the marginal effect varies across the individual samples, with a minimum of a 0.29 percentage points in CA and a maximum of 0.88 in MA. On the other hand, every extra minute on public transit (relative to driving time) reduces the probability of transit use by 0.2 percentage points. The result is significant and consistent across all samples, which signifies the importance of the time cost associated with many public transit trips.

Both parking prices and gas expenses are significant across all samples. The evidence suggests that a 10% increase in either gas expenses or daily parking prices leads to a 0.2 percentage point increase in the probability of using public transit. Nevertheless, the magnitudes of the marginal effects of the two costs vary across samples. Gas expenses seem to have the largest effect in the CA sample, with a marginal effect associated with a 10% increase goes as high as 0.52 percentage points, and the smallest in the MA sample, with a marginal effect of just 0.08 percentage points. Parking prices, on the other hand, have the largest marginal effect in the NY-NJ sample, where a 10% increase raises the probability of transit use by 0.24 percentage points, and the lowest marginal effect in the CA sample, with an increase of transit use probability by only 0.1 percentage points. The discrepancy between the results for parking prices marginal effects in the CA and NY-NJ sample is largely due to the fact that the average daily parking price in NY-NJ is much higher than that in CA.

This paper does not attempt to address potential endogeneity that may be associated with some of the examined variables. Some of the estimates may be biased because of sorting, but one way of looking at the results is that they represent short-run marginal effects. Another way would be viewing the magnitudes of these estimates as upper limits to the effect of these variables, assuming that the bias would lead to the overestimation of these effects. In any case, these results are not necessarily meant to determine causality, but are meant to give an overview of the predictors of public transit use prior to the introduction of social influence effects.

3.5.2 Spatial Social Influence

To explore the role of spatial influence in transit use, a binary probit model is estimated, where the binary depended model is defined as in equation 3.2 and T^* as in equation 3.5. The results are presented in Table 3.6. The marginal effects of spatial social influence measures are significant for both the workplace and residential tracts, as can be seen in Table 3.6. The spatial social influence measure is defined as the percentage of transit users in a given tract. The results suggest that the impact of the workplace social influence is larger than that of the residential one. A 1 percentage point increase in transit use in the workplace tract is associated with a 0.4 percentage point increase in transit use probability, whereas the same increase in the residential tract is associated with a 0.15 percentage point increase. The results are highly statistically significant (0.1% level), and consistent throughout all individual samples.

The estimates for the other determinants are generally robust, except for the density measures and financial costs associated with the commute trip. The magnitude of the marginal effects drop across all samples. The marginal effect of parking prices, although still statistically significant, is now less than half what it was before the inclusion of the spatial network effect. Now a 10% increase in parking prices is associated with a 0.08 percentage point increase in the transit use probability instead of the 0.2 percentage point increase estimated by equation 3.4. Gas prices, on the other hand, become insignificant in the MA sample, but are still statistically significant for all other samples. For the full sample, a 10% increase in gas expenses is associated with a 0.1 percentage point increase in the probability of transit use.

To address potential endogeneity the model is estimated once more but using an IV as described in the previous section. The results are reported in Table 3.7. The table also reports the results for the exogeneity test, where the null hypothesis is that spatial influence

Table 3.6: Spatial Social Influence: Transit Use Rate (Eq. 3.5)

Dependent Variable: T_i	**Full**	**CA**	**NY-NJ**	**MA**
Spatial Social Influence, $\boldsymbol{SN}_i$				
Transit Use Rate^R	0.0015***	0.0015***	0.0012***	0.0018***
Transit Use Rate^W	0.0039***	0.0035***	0.0036***	0.0047***
Socioeconomic Characteristics, $\boldsymbol{X}_i$				
Income	-0.0021***	-0.0021***	-0.0031***	-0.0019***
Household Size	-0.0062***	0.002	-0.0121***	-0.0107***
Age	-0.0007***	-0.0004*	-0.0004	-0.0016***
White	-0.0264***	-0.0084	-0.0455***	-0.0246*
Female	0.0005	0.0112*	0.0069	-0.0256***
Neighborhood Characteristics, $\boldsymbol{B}_i$				
Population Density^R	1.1208***	-0.2688	1.2763***	1.084
Population Density^W	-0.6638**	0.0405	-0.6435**	0.7558
Job Density R	0.1183	-0.4794	0.5986*	0.1624
Job Density^W	0.0888***	0.3292**	0.0608	0.0795**
Commuter White Rate^R	0	0.0003*	-0.0003	0.0007***
Commuter White Rate^W	0.0004*	0.0004	0.0005	0.0003
Median Commuter Income^R	-0.0002	-0.0029**	0.0030**	-0.0017
Median Commuter Income^W	0.0040***	0.0007	0.0047*	0.0148***
Transit Access, $\boldsymbol{A}_i$				
Distance to Rail^R	-0.0022**	-0.0014*	-0.0054**	-0.0152***
Distance to Rail^W	-0.0033***	-0.0015	-0.0101***	-0.0381***
Trip Characteristics, $\boldsymbol{C}_i$				
Distance	0.0053***	0.0025***	0.0043***	0.0074***
Time Difference	-0.0013***	-0.0012***	-0.0013***	-0.0014***
Parking Price	0.0075***	0.0042***	0.0076***	0.0068***
Gas Expense	0.0109***	0.0458***	0.0131***	0.0023
MSA Fixed Effects	Yes	Yes	No	No
Observations	20068	8075	6452	5541
Pseudo-R^2	0.5216	0.3524	0.6188	0.4473
Wald χ^2	5143	1031	2486	1220
Log Likelihood	-5059	-1779	-1636	-1515

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

A constant is included

Table 3.7: IV Results (Eq.3.7)

Dependent Variable: T_i	(1) **Full**	(2) **CA**	(3) **NY-NJ**	(4) **MA**
Spatial Social Influence, $\boldsymbol{SN}_i$				
Transit Use Rate^R	0.0015	0.0015	0.0011	0.0018
Transit Use Rate^W	0.0038***	0.0033*	0.0036***	0.0046
Socioeconomic Characteristics, $\boldsymbol{X}_i$				
Income	-0.0021***	-0.0021***	-0.0031***	-0.0019***
Household Size	-0.0061***	0.002	-0.0121***	-0.0105***
Age	-0.0007***	-0.0004**	-0.0004	-0.0016***
White	-0.0267***	-0.0088	-0.0456***	-0.0249**
Female	0.0007	0.0118**	0.0069	-0.0250***
Neighborhood Characteristics, $\boldsymbol{B}_i$				
Population Density^R	1.1034	-0.3406	1.2701	0.971
Population Density^W	-0.5885	0.4691	-0.6275	0.7734
Job Density^R	0.1191	-0.5019	0.6136	0.1585
Job Density^W	0.0878	0.2816	0.0599	0.0847
Commuter White Rate^R	0	0.0003*	-0.0003	0.0007*
Commuter White Rate^W	0.0003	0.0003	0.0005	0.0001
Median Commuter Income^R	-0.0002	-0.0029***	0.0030**	-0.0017
Median Commuter Income^W	0.0040***	0.0007	0.0044	0.0149**
Transit Access, $\boldsymbol{A}_i$				
Distance to Rail^R	-0.0022**	-0.0013	-0.0053**	-0.0149***
Distance to Rail^W	-0.0033***	-0.0016*	-0.0103**	-0.0390***
Trip Characteristics, $\boldsymbol{C}_i$				
Distance	0.0052***	0.0025***	0.0043***	0.0075***
Time Difference	-0.0013***	-0.0012***	-0.0013***	-0.0014***
Parking Price	0.0075***	0.0041	0.0076	0.007
Gas Expense	0.0109***	0.0454***	0.0130***	0.0018
MSA Fixed Effects	Yes	Yes	No	No
Observations	20068	8075	6452	5541
Wald χ^2	6328.26	1390.41	2561.77	1749.16
Wald Exogeneity Test				
χ^2	3.7	3.37	0.78	2.69
p-value	0.157	0.1855	0.6786	0.2603
First-Stage: RESIDENCE				
F-stat	2088.44	497.22	1045.73	543.28
Adjusted-R^2	0.7374	0.6058	0.7728	0.6727
First-Stage: WORKPLACE				
F-stat	3527.51	646.31	2259.6	894.72
Adjusted-R^2	0.8259	0.6665	0.8803	0.7721

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

A constant is included

Table 3.8: IV Results for Full sample (all vs. limited controls)

	(1)	(2)
Dependent Variable: T_i	**All controls**	**Limited controls**
Spatial Social Influence, $\boldsymbol{SN}_i$		
Transit Use Rate^R	0.0015	0.0015**
Transit Use Rate^W	0.0038***	0.0058***
Socioeconomic Characteristics, $\boldsymbol{X}_i$		
Income	-0.0021***	-0.0016***
Household Size	-0.0061***	-0.0079***
Age	-0.0007***	-0.0009***
White	-0.0267***	-0.0246***
Female	0.0007	-0.0056
Neighborhood Characteristics, $\boldsymbol{B}_i$		
Population Density^R	1.1034	
Population Density^W	-0.5885	
Job Density^R	0.1191	
Job Density^W	0.0878	
Commuter White Rate^R	0	
Commuter White Rate^W	0.0003	
Median Commuter Income^R	-0.0002	
Median Commuter Income^W	0.0040***	
Transit Access, $\boldsymbol{A}_i$		
Distance to Rail^R	-0.0022**	
Distance to Rail^W	-0.0033***	
Trip Characteristics, $\boldsymbol{C}_i$		
Distance	0.0052***	
Time Difference	-0.0013***	
Parking Price	0.0075***	
Gas Expense	0.0109***	
MSA Fixed Effects	Yes	Yes
Observations	20068	20542
Wald χ^2	6328.26	3257.26
Wald Exogeneity Test		
χ^2	3.7	17.04
p-value	0.157	0.0002
First-Stage: RESIDENCE		
F-stat	2088.44	575.9
Adjusted-R^2	0.7374	0.2668
First-Stage: WORKPLACE		
F-stat	3527.51	726.76
Adjusted-R^2	0.8259	0.3147

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

A constant is included

(Transit Use Rate) is exogenous, as well as the first-stage F-statistics and adjusted-R^2. The findings reveal several things. The first result to note is that the first-stage F-statistics indicate that the instruments cannot be classified as weak instruments. The F-statistic is 2088 for the full sample, and is the lowest for the CA sample at 497. The second result to note is that, whereas the residence tract spatial influence is no longer significant (Transit Use Rate^R), the workplace tract one is still statistically significant for all the samples except the MA sample. For the full sample, a 1 percentage point increase in transit use by coworkers results in a 0.38 percentage point increase in the probability of transit use. The third result to note is that the marginal effects are identical to the non-IV estimates. This suggests that the non-IV estimates are in fact consistent estimates, since the IV estimation produces the same marginal effects.

The fourth result to note is the exogeneity test. According to the Wald exogeneity test we fail to reject the null hypothesis of exogeneity for all the samples. This indicates that the loss of efficiency associated with the IV estimation is outweighing any benefit in terms of bias. Together with the third results noted, this finding suggests that endogeneity may not be a serious concern, given the rich set of factors that are jointly controlled for, and that there are no gains associated with IV estimation even though the IV used is a relevant instrument. Specification (2) in Table 3.8 elaborates on this point.

Table 3.8 presents the results with a complete set of control variables for the full sample in specification (1). This is the same as specification (1) from Table 3.7. Specification (2) in Table 3.8, on the other hand, estimates the model using the IV but keeping out most of the control variables. The only variables kept are the socioeconomic ones. As ca be seen, the exogeneity test reveals that the null hypothesis of exogeneity is rejected. Both the residence and workplace spatial influence measures are now significant. Comparing the evidence from specifications (1) and (2) in Table 3.8 it seems that the rich set of control variables included in the model succeeds in capturing most heterogeneity that would be correlated with spatial influence.

3.5.3 Income-Group Social Influence

To explore the role of income-group social influence in transit use, a binary probit model is estimated, where the binary depended model is defined as in equation 3.2 and where T^* is defined by equation 3.8. The average marginal effects are presented in Table 3.9. The results strongly support the hypothesis of a positive impact of Median Transit Income on a rich commuter's decision to use transit. Although both are statistically significant, the size of the marginal effect of income-group influence in the workplace tract is larger than that of the residence. The evidence suggests that, whereas the poor's propensity to use transit is unaffected by the Median Transit Income, a $10,000 increase in Median Transit Income in the residence tract raises the probability of the rich commuter by 0.2 percentage points, and the same increase in the workplace tract raises the probability by 0.7 percentage points.

The results of the individual samples vary to some extent. For example, the Median Transit Income has no significant effect on neither rich nor poor commuters' decision to use transit in CA. In MA, the results hold only for the workplace, and fail to hold for the residence, implying that income-group social influence may exist only among coworkers over there. The NY-NJ results tell a more complex story. It is not merely the case that poor commuters are unaffected by the Median Transit Income, but are rather negatively influenced by it in the residence tracts. A $10,000 increase in Median Transit Income in the residence tract reduces a poor commuter's probability of using transit by 0.48 percentage points. This finding points to peer effects among the poor. The interaction term for the richer commuters is significant and thus the effect different from that among the poor commuters. Adding up the marginal effects to compute that of Median Transit Income on the rich commuter in the residence

Table 3.9: Income-Group Social Influence: Median Transit Income (Eq. 3.8)

Dependent Variable: T_i	**Full**	**CA**	**NY-NJ**	**MA**
Spatial Social Influence, $\boldsymbol{SN}_i$				
Transit Use Rate^R	0.0022***	0.0022***	0.0016***	0.0031***
Transit Use Rate^W	0.0038***	0.0029***	0.0032***	0.0060***
INTERACTIONS: $\boldsymbol{SNRICH}_i$				
$SN_i^R \times RICH_i^R$	-0.0008***	-0.0010*	-0.0003	-0.001
$SN_i^W \times RICH_i^W$	0.0008***	0.0013**	0.0013***	-0.0006
Income-Group Social Influence, $\boldsymbol{YN}_i$				
Transit Median Income^R	-0.0005	0.0019	-0.0048***	0.0036
Transit Median Income^W	-0.0001	0.0013	-0.0037	-0.0037
INTERACTIONS: $\boldsymbol{YNRICH}_i$				
$YN_i^R \times RICH_i^R$	0.0022*	0.0019	0.0065***	-0.0054*
$YN_i^W \times RICH_i^W$	0.0070***	0.0015	0.0114***	0.0101**
$\boldsymbol{RICH}_i$				
$RICH_i^R$	-0.0360**	-0.0318	-0.0897***	0.0666
$RICH_i^W$	-0.0565***	-0.0162	-0.1192***	-0.0463
Socioeconomic Characteristics, $\boldsymbol{X}_i$				
Income	-0.0020***	-0.0015	-0.0030**	-0.0030**
Household Size	-0.0070***	0.0037	-0.0123***	-0.0138***
Age	-0.0008***	-0.0005*	-0.0004	-0.0021***
White	-0.0299***	-0.007	-0.0491***	-0.0254
Female	0.0044	0.0168**	0.0111	-0.0330***
Neighborhood Characteristics, $\boldsymbol{B}_i$				
Population Density^R	1.3146***	0.1495	1.3648***	2.4133
Population Density^W	-0.6439*	0.0969	-0.6617*	0.4831
Job Density^R	0.1958	-0.0386	0.8075*	0.2615
Job Density^W	0.1088***	0.3254**	0.0719	0.0962**
Commuter White Rate^R	0.0001	0.0007***	-0.0002	0.0010***
Commuter White Rate^W	0.0006**	0.0005	0.0006	0.0005
Median Commuter Income^R	-0.0028*	-0.0091***	0.0036	-0.0028
Median Commuter Income^W	0.003	-0.0007	0.0051	0.0199***
Transit Access, $\boldsymbol{A}_i$				
Distance to Rail^R	-0.0027**	-0.0013	-0.0054**	-0.0228***
Distance to Rail^W	-0.0034**	-0.0016	-0.0128***	-0.0598***
Trip Characteristics, $\boldsymbol{C}_i$				
Distance	0.0059***	0.0025***	0.0053***	0.0105***
Time Difference	-0.0016***	-0.0014***	-0.0015***	-0.0024***
Parking Price	0.0090***	0.0053***	0.0083***	0.0087***
Gas Expense	0.0181***	0.0511***	0.0125**	0.0144*
MSA Fixed Effects	Yes	Yes	No	No
Observations	14937	5930	5303	3704
Pseudo-R^2	0.5038	0.3642	0.5938	0.3935
Wald χ^2	4268	958	2052	943.7831
Log Likelihood	-4387	-1462	-1483	-1308

$^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

A constant is included

tract reveals that a \$10,000 increase raises the probability that a rich commuter uses transit by 0.2 percentage points. On the other hand, the same increase in the workplace tract leads to a 1.1 percentage point increase in transit use probability by the rich.

The $\boldsymbol{RICH_i}$ variables, which are income-group specific fixed effects, are statistically significant in the full sample and indicate that having an income exceeding the median income level of commuters in the tract reduces the probability of using transit by 3.6 percentage points in the residence tract, and by 5.7 percentage points in the workplace tract. This result, however, does not hold across all samples. The NY-NJ sample is the only sample that also produces statistically significant estimates for $\boldsymbol{RICH_i}$, and the marginal effects are of a larger magnitude. The $\boldsymbol{SNRICH_i}$ interaction terms indicate that there are some differences in the responsiveness of the rich to the Transit Use Rate in the corresponding tract. The rich are less responsive to the spatial influence in the residence tract, yet more responsive in the workplace tract. This result holds for the CA sample as well. For the NY-NJ sample, the same results hold for the workplace tract, but for the residence tract, the rich seem to be as responsive to spatial influence as the poor. These interaction terms capture any further tract-specific heterogeneity that affects transit use and that is income-group-specific.

3.5.4 Racial Social Influence

To explore the role of racial social influence in transit use, a binary probit model is estimated, where the binary depended variable is defined as in equation 3.2 and where T^* is defined by equation 3.9. The average marginal effects are presented in Table 3.10. The results that emerge indicate that there are peer effects among the nonwhite group in the workplace, where the probability of transit use by a nonwhite commuter increases as the percentage of white transit users falls, i.e., the percentage of nonwhite commuters rises. Since the interaction term of the $WHITE_i$ variable with the Transit White Rate of the workplace is positive

Table 3.10: Racial Social Influence: Transit White Rate (Eq. 3.9)

Dependent Variable: T_i	**Full**	**CA**	**NY-NJ**	**MA**
Spatial Social Influence, $\boldsymbol{SN}_i$				
Transit Use Rate^R	0.0018***	0.0016***	0.0013***	0.0013*
Transit Use Rate^W	0.0036***	0.0037***	0.0027***	0.0048***
INTERACTIONS: $WHITE_i \times \boldsymbol{SN}_i$				
$WHITE_i \times SN_i^R$	-0.0003	-0.0003	-0.0001	0.0008
$WHITE_i \times SN_i^W$	0.0009***	0.0001	0.0018***	0.0007
Racial Social Influence, $\boldsymbol{RN}_i$				
Transit White Rate^R	-0.0001	-0.0002	0	0.0005
Transit White Rate^W	-0.0004**	-0.0002	-0.0004	-0.0005
INTERACTIONS: $WHITE_i \times \boldsymbol{RN}_i$				
$WHITE_i \times RN_i^R$	0.0002	0.0003	0.0003	-0.0002
$WHITE_i \times RN_i^W$	0.0003*	-0.0001	-0.0001	0.0005
Socioeconomic Characteristics, $\boldsymbol{X}_i$				
Income	-0.0023***	-0.0022***	-0.0032***	-0.0025***
Household Size	-0.0071***	0.0041	-0.0125***	-0.0109***
Age	-0.0008***	-0.0004	-0.0003	-0.0018***
$WHITE_i$	-0.0693***	-0.0145	-0.1346***	-0.0857*
Female	0.0023	0.0163**	0.0099	-0.0291***
Neighborhood Characteristics, $\boldsymbol{B}_i$				
Population Density^R	1.1948***	-0.4762	1.2120***	1.6236
Population Density^W	-0.6322**	-0.1883	-0.4589	0.3885
Job Density^R	0.1249	-0.2052	0.6107*	0.1819
Job Density^W	0.1022***	0.3154**	0.0698	0.0954**
Commuter White Rate^R	0	0.0004	-0.0004	0.0004
Commuter White Rate^W	0.0007***	0.0005	0.0010**	0.0004
Median Commuter Income^R	0	-0.0029	0.0031**	-0.0018
Median Commuter Income^W	0.0037**	-0.0014	0.004	0.0160***
Transit Access, $\boldsymbol{A}_i$				
Distance to Rail^R	-0.0047***	-0.0031**	-0.0056**	-0.0183***
Distance to Rail^W	-0.0046***	-0.0026*	-0.0111***	-0.0469***
Trip Characteristics, $\boldsymbol{C}_i$				
Distance	0.0060***	0.0037***	0.0047***	0.0088***
Time Difference	-0.0015***	-0.0014***	-0.0014***	-0.0016***
Parking Price	0.0086***	0.0048***	0.0081***	0.0077***
Gas Expense	0.0116***	0.0418***	0.0129***	0.0047
MSA Fixed Effects	Yes	Yes	No	No
Observations	16338	5648	5990	4700
Pseudo-R^2	0.5172	0.3505	0.6114	0.4228
Wald χ^2	4528	775	2318	1077
Log Likelihood	-4487	-1337	-1574	-1451

$^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

A constant is included

and almost equal in magnitude, then the white commuters are not affected in any way by the percentage of white coworkers or white neighbors using transit. Both the effects for white and nonwhite commuters are statistically insignificant across all individual samples. The results from the full sample shoud be itnerpreted with caution since the definition of nonwhite is different across the various MSAs, and is therefore not necessarily referring to one homogeneous group of people. It is also worth noting that it seems that the white commuters are more responsive to spatial social influence than the nonwhite commuters; a result that does not hold for either CA or MA.

3.5.5 Robustness

The results are largely robust to several variations. The results are robust to the inclusion of extra variables such as land use mix, number of rail transit stations, and pedestrian-oriented intersection density, for both the workplace and residence. Other neighborhood characteristics, such as the unemployment rate, gini coefficient, and the percentage of residents below the age of 18, have been added to the model for robustness but did not affect the findings. The dataset has been further restricted to observations that had an access to rail transit in the residence tract of less than 5 miles, and the results still hold. The dataset has been expanded to include all observations from the RHTS and MTS datasets regardless of the MSA, and the results are still robust.

3.6 Implications

Overall, the evidence that emerges from the analysis implies that there exists significant spatial social influence, and that spatial social influence among coworkers, or coworker peer effects, are consistently larger than those among neighbors. These findings have several implications. One implication is that exogenous increases in other determinants that increase ridership would have indirect cascading effects through spatial social influence. For example, if gas taxes or parking prices increase, public transit use will increase. The effect would not stop at the direct effect of gas expenses or parking prices, but rather would have indirect effects as well through the spatial social influence where the individual commuter is motivated further to use transit because other neighbors or coworkers are using it. Another implication is that exogenous changes or policies that target the workplace would have a bigger impact than the ones that would target the residence. All workers are also residents, and when public transit use in the workplace rises then so does public transit use in the residence tract, which would amplify the effects. Nevertheless, targeting workplaces is more effective because the policy or change would have a greater initial effect which would then propagate to the residence locations. A third implication is that any changes initiating from social influence of income-groups would be amplified further by the spatial social influence.

The evidence further suggests that income-group social influence plays a significant role for the rich. The results indicate that rich commuters prefer to use public transit when other rich commuters do so, everything else constant. Such evidence suggests that policymakers should focus on setting policies that would encourage the rich to use transit. As more rich commuters use public transit, the median transit user income rises, and as a result more rich commuters opt for public transit because of the income-group social influence channel. As a result, the median transit user income rises further. The rise in median transit user income does not reduce the transit use of the poor, but increases transit use by the rich. Because the rich are not crowding out the poor, the transit use rate must increase, everything else constant. As a result, more poor and rich increase their transit use because of the spatial social influence channel this time, amplifying the effect of the policy that initially targeted the rich. Therefore, a multiplier effect plays out through both the income-group and spatial social influence channels.

On the other hand, a policy that targets rich and poor equally, would likely have an indirect multiplier effect through the spatial social influence channel only. Alternatively, consider a policy that would only target the poor, since they are the group more likely to use transit, as evidenced by the estimates for the Income and $\boldsymbol{RICH}_i$ variables. As the transit use rate rises initially there will be positive spatial peer effects. However, since the poor transit user population increases relative to the rich one, the median transit user income drops. As a result, rich commuters become less likely to use public transit because of the income-group social influence channel. Depending on the relative sizes of the changes, a policy targeting only the poor could end up reducing overall transit use rate.

The evidence also provides insight on what exogenous changes in the median income of all commuters in the residence or workplace neighborhoods would lead to. These changes could be the result of shifting demographics, where neighborhoods are becoming richer, for instance. An increase in the median income of all commuters in the residence tract is likely to be associated with a drop in the probability of transit use through the direct effect presented by the neighborhood characteristic variable, Median Commuter Income. The exogenous increase in Median Commuter Income is likely to be accompanied by an increase in Median Transit Income. The increase of the Median Transit Income will only be a fraction of the increase in Median Commuter Income. The results in Table 3.9 show that the estimate for the Median Commuter Income in the residence tract is roughly equal in magnitude, but opposite in direction, to the income-group social influence effect for the residence (0.22 percentage points). Since the change in the Median Transit Income is less than that in Median Commuter Income, the overall initial effect will be negative. The total effect may still be positive if the social multiplier effect outweighs the negative direct impact of a higher Median Commuter Income. Thus, an increase in Median Commuter Income, holding everything else constant, is likely to, but not necessarily, lead to a lower probability of a given commuter using transit, whether rich or poor.

On the other hand, an exogenous increase in Median Commuter Income in the workplace has no direct effect on the decision to use transit, but an accompanying increase in Median Transit Income would lead to higher probabilities of transit use by the rich through the income-group peer effects for the rich. Therefore, we can conclude that a higher Median Commuter Income in the residence tract may reduce the probability of transit use, holding everything else constant, while a higher Median Commuter Income in the workplace tract would increase the probability of transit use, holding everything else constant.

Another important implication of the results on social influence is that there is no evidence of the presence of racial social influence. This statement holds across samples and regardless of race. Although the full sample shows some racial social influence effects among minority coworkers, these results may be questionable because of the varying constituency of the minority population across states. Such findings indicate that commuters are really influenced in their decisions regarding transit use by income-groups rather than racial groups, and what seems to some that commuters may be avoiding transit because of dominance by minorities may really just be income-group social influence in action, but is confused with racial social influence because the minority transit users usually also happen to also be poor transit users.

Some other implications of the findings relate to other determinants. For instance, the marginal effect of the population density falls once the spatial social influence measures are included. This may imply that part of the density's importance in commute mode choice really spurs from it's correlation with spatial social influence measures. In models that do not account for social influence the effects are being picked up to an extent by the density measures. Another implication is related to parking prices. This study is the first to look at parking prices across many cities. Although statistically significant, the marginal effect of parking prices is relatively small. However, there is reason to believe that this estimate is a lower bound estimate for the marginal effect of parking prices due to the presence of Employer-Paid parking fringe benefits which are unaccounted for in this study. Khordagui

(2017) discusses this further and explains why we should think of this estimate as a lower bound in this case. Nevertheless, the estimates obtained here still imply that parking pricing can be used as a policy that targets workplaces in particular. Other regulations aimed at employers, such as those designed to incentivize transit use or discourage driving in general, would be policies specifically targeting workplaces and hence benefit from the added influence of peer effects.

3.7 Conclusion

The paper explores the underlying fundamental determinants of public transit use and introduces the role of social influence, where a social group is defined across three different dimensions: space, income, and race. The analysis makes use of a unique dataset specifically built for the purpose of testing for all the effects across major US cities and surrounding areas, and drawing on data from various sources, using various techniques such as GIS and webscraping. The paper exploits the availability of census tract locations to construct variables and explore the effects in both the workplace and residence locations. To overcome potential endogeneity of spatial social influence, a novel IV is constructed, and is found to be a suitable instrument. The IV is the weighted average distance of commute trip distances of a particular tract, whether residence or workplace, computed from OD data. The first-stage F-statistics indicate that the IV is not weak. The exogeneity test indicates that the null hypothesis of exogeneity cannot be rejected when using the IV method. Estimating the model once more while leaving out most of the control variables results in rejecting the null hypothesis of exogeneity. The evidence suggests that the wide array of control variables included in the model account for heterogeneity correlated with the spatial influence measures. An alternative identification strategy is employed for the estimation of income-group and racial social influence, where income and race group fixed effects control for income- and race-specific heterogeneity, tract transit use rate controls for tract-specific heterogeneity impacting transit use, and an interaction term of both controls for tract-specific heterogeneity impacting transit use and is income- or racial-specific.

The results of the fundamental determinants model are largely in line with expectations and previous findings. The findings also suggest that social influence plays a role in the decision to use transit, after controlling for a rich set of transit use determinants, and after addressing endogeneity. Social influence among coworkers is relatively more influential than that among neighbors. Furthermore, income-group social influence is particularly relevant for the richer commuters. Theses findings allow policymakers to understand how changing demographics could have cascading effects on transit use, and sheds light on groups that can be targeted to encourage increased ridership. According to the findings of this paper, it is best to target workplace social groups, and target richer commuters. Policies that encourage transit use by these groups would eventually increase transit use by all groups through an indirect multiplier effect. Furthermore, the findings introduce a new dimension to the implications of changing demographics within a city. For instance, as residential neighborhoods become richer it is likely that public transit use will fall, but as workplace neighborhoods become richer the probability that a rich commuter uses transit increases, which would lead to a cascading effect through both income-group and spatial social influence.

A major limitation of this study is that the dataset is cross-sectional, and that limits the ability to model how these effects could play out over time. Nevertheless, it highlights the importance of social influence, and compares the fundamental factors associated with transit use across different cities, which is a good starting to point to encourage further research in the area. Further research could examine the dynamics of the impact of social influence on transit use, and examine the social multiplier effect. Tipping points are likely to exist as well, and this could be an area of investigation. Furthermore, future research could examine racial social influence for nonwhite subgroups to better understand its dynamics in transit use. Other definitions of social influence could also be investigated, while also exploring the existence of education-based or occupation-based social influence. Another line of research could distinguish between, the effects on bus and rail transit use, possibly using a multinomial discrete choice model. Overall, social influence is understudied in the commute mode choice literature and there is a lot of room to expand on the current research to develop a more solid understanding of the topic and its implications.

Bibliography

Accetturo, A., Manaresi, F., Mocetti, S. & Olivieri, E. (2014), 'Don't stand so close to me: The urban impact of immigration', *Regional Science and Urban Economics* **45**, 45 – 56.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

American Association of State Highway and Transportation Officials (2017), Census transportation planning products, 2006-2010.

Ammermueller, A. & Pischke, J. (2009), 'Peer effects in european primary schools: Evidence from the progress in international reading literacy study', *Journal of Labor Economics* **27**(3), 315–348.

Angrist, J. D. & Lang, K. (2004), 'Does school integration generate peer effects? evidence from boston's metco program', *American Economic Review* **94**(5), 1613–1634.

Bayer, P., Ross, S. L. & Topa, G. (2008), 'Place of work and place of residence: Informal hiring networks and labor market outcomes', *Journal of Political Economy* **116**(6), 1150–1196.

Bento, A. M., Goulder, L. H., Jacobsen, M. R. & von Haefen, R. H. (2009), 'Distributional and efficiency impacts of increased us gasoline taxes', *American Economic Review* **99**(3), 667–99.

Brock, W. A. & Durlauf, S. N. (2001), 'Discrete choice with social interactions', *The Review of Economic Studies* **68**(2), 235–260.

Brown, K. M. & Laschever, R. A. (2012), 'When they're sixty-four: Peer effects and the timing of retirement', *American Economic Journal: Applied Economics* **4**(3), 90–115.

Brownstone, D. & Golob, T. F. (1992), 'The effectiveness of ridesharing incentives: Discrete-choice models of commuting in southern california', *Regional Science and Urban Economics* **22**(1), 5 – 24.

Brueckner, J. K. & Franco, S. F. (2018), 'Employer-paid parking, mode choice, and suburbanization', *Journal of Urban Economics* **104**, 35 – 46.

Bureau of Transportation Statistics (2017), National transportation atlas database, 2012.

Bursztyn, L., Ederer, F., Ferman, B. & Yuchtman, N. (n.d.), 'Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions', *Econometrica* **82**(4), 1273–1301.

California Environmental Protection Agency (2009), 'California parking cash-out program: Informational guide for employers'.

Card, D., Mas, A. & Rothstein, J. (2008), 'Tipping and the Dynamics of Segregation*', *The Quarterly Journal of Economics* **123**(1), 177–218.

Cervero, R. (1996), 'Mixed land-uses and commuting: Evidence from the american housing survey', *Transportation Research Part A: Policy and Practice* **30**(5), 361–377.

Cervero, R. (2002), 'Induced travel demand: Research design, empirical evidence, and normative policies', *Journal of Planning Literature* **17**(1), 3–20.

Cervero, R. (2018), *America's suburban centers: the land use-transportation link*, Routledge.

Cervero, R. & Gorham, R. (1995), 'Commuting in transit versus automobile neighborhoods', *Journal of the American planning Association* **61**(2), 210–225.

Cervero, R. & Kockelman, K. (1997), 'Travel demand and the 3ds: Density, diversity, and design', *Transportation Research Part D: Transport and Environment* **2**(3), 199 – 219.

Cervero, R. & Wu, K.-L. (1997), *Influences of land-use environments on commuting choices: an analysis of large US metropolitan areas using the 1985 American Housing Survey*, number 683, University of California at Berkeley, Institute of Urban and Regional Development.

Chester, M., Fraser, A., Matute, J., Flower, C. & Pendyala, R. (2015), 'Parking infrastructure: A constraint on or opportunity for urban redevelopment? a study of los angeles county parking supply and growth', *Journal of the American Planning Association* **81**(4), 268–286.

Christiansen, P., Engebretsen, Ø., Fearnley, N. & Hanssen, J. U. (2017), 'Parking facilities and the built environment: Impacts on travel behaviour', *Transportation Research Part A: Policy and Practice* **95**, 198–206.

Christiansen, P. & Hanssen, J. U. (2014), Park and ride - results from surveys among users, TOi-Report 1367/2014, Institute of Transportation Economics.

Cornelissen, T., Dustmann, C. & Schönberg, U. (2017), 'Peer effects in the workplace', *American Economic Review* **107**(2), 425–56.

Crane, R. & Crepeau, R. (1998), 'Does neighborhood design influence travel?: A behavioral analysis of travel diary and gis data', *Transportation Research Part D: Transport and Environment* **3**(4), 225 – 238.

Cutter, W. B., Franco, S. & Lewis, W. S. (2016), The shadow cost of parking minimums: Evidence from los angeles county.

Ewing, R. & Cervero, R. (2001), Travel and the built environment: A synthesis, Vol. 1780.

Ewing, R. & Cervero, R. (2010), 'Travel and the built environment: a meta-analysis', *Journal of the American Planning Association* **76**(3), 265–294.

Federal Highway Administration (2013), Highway statistics, Report.

Feeney, B. P. (1989), 'A review of the impact of parking policy measures on travel demand', *Transportation Planning and Technology* **13**(4), 229–244.

Fortin, B. & Yazbeck, M. (2015), 'Peer effects, fast food consumption and adolescent weight gain', *Journal of Health Economics* **42**, 125 – 138.

Foucault, T. & Fresard, L. (2014), 'Learning from peers' stock prices and corporate investment', *Journal of Financial Economics* **111**(3), 554 – 577.

Franco, S. F. (2013), Validation of the RELU-TRAN L.A. Model, Technical Report 2013-5, MRPI program of the University of California.

Franco, S. F. (2016*a*), Parking costs in los angeles county, Technical Report 2016-2-June, MRPI program of the University of California.

Franco, S. F. (2016*b*), Vehicle Parking Facilities Data Collection for the RELU-TRAN L.A.-Parking Model, Technical Report 2016-1-May, MRPI program of the University of California.

Franco, S. & Khordagui, N. (2018), Commute mode choices and the role of parking prices, parking availability and urban form: Evidence from los angeles county.

Gillen, D. W. (1977), 'Estimation and specification of the effects of parking costs on urban transport mode choice', *Journal of Urban Economics* **4**(2), 186 – 199.

Gillingham, K. (2014), 'Identifying the elasticity of driving: Evidence from a gasoline price shock in california', *Regional Science and Urban Economics* **47**, 13 – 24. SI: Tribute to John Quigley.

Glaeser, E. L., Kahn, M. E. & Rappaport, J. (2008), 'Why do the poor live in cities? the role of public transportation', *Journal of Urban Economics* **63**(1), 1 – 24.

Glaeser, E. L., Sacerdote, B. & Scheinkman, J. A. (1996), 'Crime and social interactions', *The Quarterly Journal of Economics* **111**(2), 507–548.

Glaeser, E. L. & Scheinkman, J. (2000), Non-market interactions, Working Paper 8053, National Bureau of Economic Research.

Glaeser, E. L., Scheinkman, J. A. & Sacerdote, B. I. (2003), 'The social multiplier', *Journal of the European Economic Association* **1**(2/3), 345–353.

Goetzke, F. (2008), 'Network effects in public transit use: Evidence from a spatially autoregressive mode choice model for new york', *Urban Studies* **45**(2), 407–417.

Grinblatt, M., Keloharju, M. & Ikäheimo, S. (2008), 'Social influence and consumption: Evidence from the automobile purchases of neighbors', *The Review of Economics and Statistics* **90**(4), 735–753.

Guo, Z. (2013*a*), 'Does residential parking supply affect household car ownership? the case of new york city', *Journal of Transport Geography* **26**, 18 – 28.

Guo, Z. (2013*b*), 'Home parking convenience, household car usage, and implications to residential parking policies', *Transport Policy* **29**, 97 – 106.

Guo, Z. & McDonnell, S. (2013), 'Curb parking pricing for local residents: An exploration in new york city based on willingness to pay', *Transport Policy* **30**, 186 – 198.

Hanssen, J. U., Tennøy, A., Christiansen, P. & Øksenholt, K. V. (2014), Which kinds of p&r can contribute to reduced greenhouse gas emissions?, TOi-Report 1366/2014, Institute of Transportation Economics.

Hanushek, E. A., Kain, J. F., Markman, J. M. & Rivkin, S. G. (n.d.), 'Does peer ability affect student achievement?', *Journal of Applied Econometrics* **18**(5), 527–544.

Hensher, D. A. & King, J. (2001), 'Parking demand and responsiveness to supply, pricing and location in the sydney central business district', *Transportation Research Part A: Policy and Practice* **35**(3), 177 – 196.

Hess, D. B. (2001), 'Effect of free parking on commuter mode choice: Evidence from travel diary data', *Transportation Research Record* **1753**(1), 35–42.

Hoxby, C. (2000), Peer effects in the classroom: Learning from gender and race variation, Working Paper 7867, National Bureau of Economic Research.

Inci, E. (2015), 'A review of the economics of parking', *Economics of Transportation* **4**(1), 50 – 63.

Kaustia, M. & Rantala, V. (2015), 'Social learning and corporate peer effects', *Journal of Financial Economics* **117**(3), 653 – 669.

Khordagui, N. (2017), Parking prices and the decision to drive: Evidence from california.

Kim, C., Parent, O. & vom Hofe, R. (2018), 'The role of peer effects and the built environment on individual travel behavior', *Environment and Planning B: Urban Analytics and City Science* **45**(3), 452–469.

Kobus, M. B., i Puigarnau, E. G., Rietveld, P. & Ommeren, J. N. V. (2013), 'The on-street parking premium and car drivers' choice between street and garage parking', *Regional Science and Urban Economics* **43**(2), 395 – 403.

Kockelman, K. (1995), Which matters more in mode choice: Density or income?, Compendium of Technical Papers 844-867, Institute of Transportation Engineers.

Kwoka, G. J., Boschmann, E. E. & Goetz, A. R. (2015), 'The impact of transit station areas on the travel behaviors of workers in denver, colorado', *Transportation Research Part A: Policy and Practice* **80**, 277 – 287.

Lee, L.-F. (1978), 'Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables', *International Economic Review* **19**(2), 415–433.

Lee, L.-F. (1979), 'Identification and estimation in binary choice models with limited (censored) dependent variables', *Econometrica* **47**(4), 977–996.

Levtnson, D. & Kumar, A. (1997), 'Density and the journey to work', *Growth and Change* **28**, 147 – 172.

Lipps, O. & Kunert, U. (2005), Measuring and explaining the increase of travel distance: A multilevel analysis using repeated cross sectional travel surveys, DIW Discussion Papers 492, Berlin.

Mallett, W. (2018), Trends in public transportation ridership: Implications for federal policy, Report R45144, Congressional Research Service.

Manski, C. F. (1993), 'Identification of endogenous social effects: The reflection problem', *The Review of Economic Studies* **60**(3), 531–542.

Manville, M. & Shoup, D. (2005), 'Parking, people, and cities', *Journal of Urban Planning and Development* **131**(4), 233–245.

Marsden, G. (2006), 'The evidence base for parking policies—a review', *Transport Policy* **13**(6), 447 – 457. Parking.

Mas, A. & Moretti, E. (2009), 'Peers at work', *American Economic Review* **99**(1), 112–45.

Massachusetts Department of Transportation (2018), 'Massachusetts travel survey, 2010-2011'.

McFadden, D. (2010), *Sociality, Rationality, and the Ecology of Choice*, chapter 1, pp. 1–17.

Messenger, T. & Ewing, R. (1996), 'Transit-oriented development in the sun belt', *Transportation Research Record* **1552**(1), 145–153.

Miller, G. K. & Everett, C. T. (1982), 'Raising commuter parking prices — an empirical study', *Transportation* **11**(2), 105–129.

Moretti, E. (2011), 'Social Learning and Peer Effects in Consumption: Evidence from Movie Sales', *The Review of Economic Studies* **78**(1), 356–393.

Mussa, A., Nwaogu, U. G. & Pozo, S. (2017), 'Immigration and housing: A spatial econometric analysis', *Journal of Housing Economics* **35**, 13 – 25.

Nakajima, R. (2007), 'Measuring Peer Effects on Youth Smoking Behaviour', *The Review of Economic Studies* **74**(3), 897–935.

New York Metropolitan Transportation Council (2017), 'Regional household travel survey, 2010-2011'.

Patacchini, E. & Zenou, Y. (2009), 'Juvenile Delinquency and Conformism', *The Journal of Law, Economics, and Organization* **28**(1), 1–31.

Peng, Z., Dueker, K. J. & Strathman, J. G. (1996), 'Residential location, employment location, and commuter responses to parking charges', *Transportation Research Record* **1556**(1), 109–118.

Pike, S. & Lubell, M. (2018), 'The conditional effects of social influence in transportation mode choice', *Research in Transportation Economics* **68**, 2 – 10. SI: Frontiers in Transportation.

Powell, L. M., Tauras, J. A. & Ross, H. (2005), 'The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior', *Journal of Health Economics* **24**(5), 950 – 968.

Sacerdote, B. (2001), 'Peer Effects with Random Assignment: Results for Dartmouth Roommates', *The Quarterly Journal of Economics* **116**(2), 681–704.

Saelens, B. E., Sallis, J. F., Black, J. B. & Chen, D. (2003), 'Neighborhood-based differences in physical activity: an environment scale evaluation', *American journal of public health* **93**(9), 1552–1558.

SCAG (2001), The new economy and jobs/housing balance in southern california, Report, Southern California Association of Governments.

Schelling, T. C. (1971), 'Dynamic models of segregation', *The Journal of Mathematical Sociology* **1**(2), 143–186.

Shoup, D. (2005), *The High Cost of Free Parking*, Planners Press Chicago.

Shoup, D. C. (1997*a*), 'Evaluating the effects of cashing out employer-paid parking: Eight case studies', *Transport Policy* **4**(4), 201 – 216.

Shoup, D. C. (1997*b*), 'The high cost of free parking', *Journal of Planning Education and Research* **17**(1), 3–20.

Sorensen, P., Wachs, M., Min, E. Y., Kofner, A. & Ecola, L. (2008), *Moving Los Angeles: Short-term policy options for improving transportation*, Rand Corporation.

Su, Q. & Zhou, L. (2012), 'Parking management, financial subsidies to alternatives to drive alone and commute mode choices in seattle', *Regional Science and Urban Economics* **42**(1), 88 – 97.

Transport for London (2012), Residential parking provision in new developments, Travel in london research report.

Transportation Secure Data Center (2015), 'California household travel survey, 2012'. National Renewable Energy Laboratory. Available from www.nrel.gov/tsdc.

Trogdon, J. G., Nonnemaker, J. & Pais, J. (2008), 'Peer effects in adolescent overweight', *Journal of Health Economics* **27**(5), 1388 – 1399.

United States Census Bureau (2017*a*), American Community Survey.

United States Census Bureau (2017*b*), LEHD Origin-Destination Employment Statistics Data.

van Ommeren, J., Wentink, D. & Dekkers, J. (2011), 'The real price of parking policy', *Journal of Urban Economics* **70**(1), 25 – 31.

Walker, J. L., Ehlers, E., Banerjee, I. & Dugundji, E. R. (2011), 'Correcting for endogeneity in behavioral choice models with social influence variables', *Transportation Research Part A: Policy and Practice* **45**(4), 362 – 374. Special Issue: Transportation and Social Interactions.

Washbrook, K., Haider, W. & Jaccard, M. (2006), 'Estimating commuter mode choice: A discrete choice analysis of the impact of road pricing and parking charges', *Transportation* **33**(6), 621–639.

Weinberger, R. (2012), 'Death by a thousand curb-cuts: Evidence on the effect of minimum parking requirements on the choice to drive', *Transport Policy* **20**, 93 – 102.

Weinberger, R., Seaman, M. & Johnson, C. (2008a), Suburbanizing the city: how new york city parking requirements lead to more driving, Transportation alternatives.

Weinberger, R., Seaman, M. & Johnson, C. (2009), 'Residential off-street parking impacts on car ownership, vehicle miles traveled, and related carbon emissions: New york city case study', *Transportation Research Record* **2118**(1), 24–30.

Weinberger, R., Seaman, M., Johnson, C. & Kaehny, J. (2008b), Guaranteed parking–guaranteed driving: Comparing jackson heights, queens and park slope, brooklyn shows that a guaranteed parking spot at home leads to more driving to work, Transportation alternatives.

Willson, R. (2013), *Parking Reform Made Easy*, Island Press.

Willson, R. W. & Shoup, D. C. (1990), 'Parking subsidies and travel choices: Assessing the evidence', *Transportation* **17**(2), 141–157.

Wilson, R. W. (1992), 'Estimating the travel and parking demand effects of employer-paid parking', *Regional Science and Urban Economics* **22**(1), 133 – 145.

Young, H. P. (2009), 'Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning', *American Economic Review* **99**(5), 1899–1924.

Zegras, C. (2010), 'The built environment and motor vehicle ownership and use: Evidence from santiago de chile', *Urban Studies* **47**(8), 1793–1817.

Zhang, H. & Arnott, R. (2011), Computing value per square foot of vacant parcels and aggregating to model zone level, Technical report, MRPI program of the University of California.

Appendix A

Chapter 1 Appendix

A.1 Data & Sources

Table A.1: List of counties

Alameda	Marin	San Bernardino	Santa Barbara	Stanislaus
Contra Costa	Merced	San Diego	Santa Clara	Yolo
Fresno	Orange	San Francisco	Santa Cruz	
Humboldt	Placer	San Joaquin	Shasta	
Kern	Riverside	San Luis Obispo	Solano	
Los Angeles	Sacramento	San Mateo	Sonoma	

A.2 Selected Variables

Most of those who pay for parking in this sample live in 4 counties: Los Angeles (32%), San Francisco (21%), Alameda (15%), and Sacramento (8%). Overall, 3.4% of those who drive alone report paying for parking. Sacramento tops the list with 13% of commuting drivers paying for parking, followed by San Francisco, where 11% of commuting drivers pay

Table A.2: Transit station / stop sources

County	Source
Alameda	Metropolitan Transportation Commission
Contra Costa	Metropolitan Transportation Commission
Fresno	TransitLand
Humboldt	TransitLand
Kern	TransitLand
Los Angeles	Los Angeles Metro Developer & TransitLand
Marin	Metropolitan Transportation Commission
Merced	TransitLand
Orange	Orange County Transportation Authority
Placer	Sacramento Area Government of Councils
Riverside	TransitLand
Sacramento	Sacramento Area Government of Councils
San Bernardino	TransitLand
San Diego	San Diego's Regional Planning Agency
San Francisco	Metropolitan Transportation Commission
San Joaquin	TransitLand (San Joaquin RTD)
San Luis Obispo	Digitized from map on SLOcity.org
San Mateo	Metropolitan Transportation Commission
Santa Barbara	TransitLand
Santa Clara	Metropolitan Transportation Commission
Santa Cruz	TransitLand
Shasta	TransitLand
Solano	Metropolitan Transportation Commission
Sonoma	Metropolitan Transportation Commission
Stanislaus	TransitLand
Yolo	Sacramento Area Government of Councils

for parking. Of those who drive to work, 81% report parking at the destination location, followed by a 7.25% parking on-street, and 10% parking at an off-site parking facility.

Parking prices reported in the dataset vary in units; 49% pay parking on a monthly basis, and 31% report paying on a daily basis. Therefore, a daily price is constructed for each individual based on the reported data. The typical number of hours per day of work is computed from the reported number of work-hours per week and the number of days worked per week, which is then used to determine daily parking price for each individual. The distance to the nearest transit station is binned into 0.05 mile bins, and top-coded at 1.5 miles. Therefore, there are 30 bins for this variable and the 31st represents all those who are more than at 1.5 miles away from a transit stop. The variable can thus be thought of us whether there is a transit stop within a certain radius from a given location.

A.3 Sample Construction

The elimination process leads to the inclusion of destination zipcodes in the sample where no commuter pays for parking. This can be illustrated with the help of Figure A.1. Suppose we start off with destination zipcodes in the CHTS that look like one of the hypothetical areas in panel (a). Zipcode A represents a destination where some commuters and some non-commuters pay for parking. Zipcode B, on the other hand, represents a destination where no commuter pays for parking and only some non-commuters pay for parking; whereas zipcode C represents one where no one pays for parking regardless of the trip purpose. Dropping destination zipcodes where no one pays for parking eliminates zipcodes C from the sample. Therefore, all trips made to area C are eliminated, regardless of purpose. The destination zipcodes remaining in the sample are zipcodes A and B. The following round of elimination then removes non-commuters from zipcodes A and B, and hence only commuters going to either areas A or B remain in the sample. Notice now that in panel (b) no one pays for

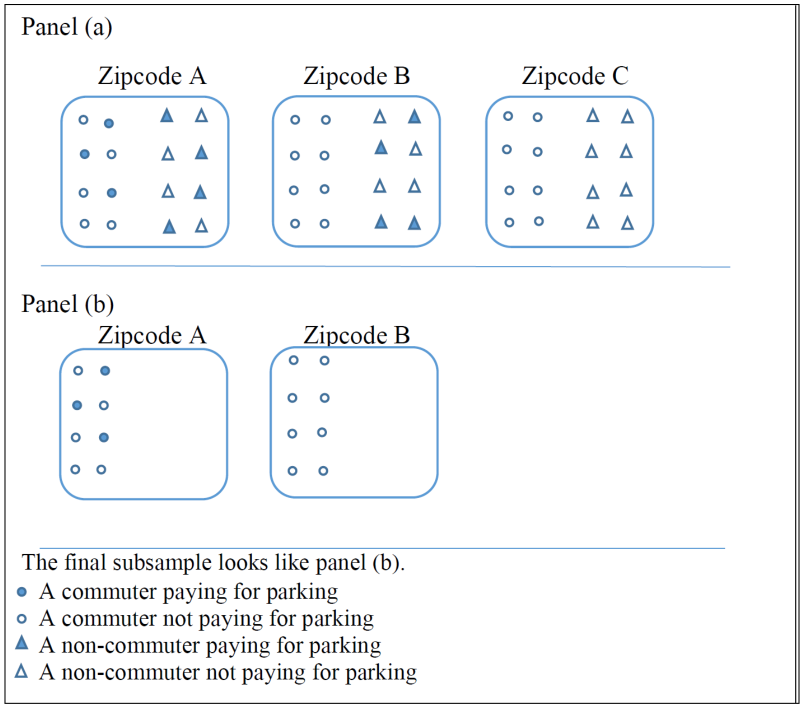

Figure A.1: Sample Construction

parking in zipcode B although the area itself survives the elimination process. Zipcode B is not eliminated from the sample because although commuters happen not to pay for parking, a price for parking exists in that area and is paid by some non-commuters, implying that the true market price for parking is a positive value.

A.4 Sample Selection Model

To see how this works, let us first assume that parking prices are a function of location, represented by zipcodes, and of the characteristics of commuters. Parking prices can thus be modeled as follows:

$$\ln(P_i^*) = \boldsymbol{Z_i \alpha_z} + \boldsymbol{X_i \alpha_x} + \nu_i \tag{A.1}$$

where $\boldsymbol{Z_i}$ is a matrix of dummy variables for the destination zipcodes. Therefore, parking price is modeled as a function of location characteristics, represented by $\boldsymbol{Z_i}$, and individual characteristics, represented by $\boldsymbol{X_i}$. Equation A.1 is used to impute $\ln(P_i^*)$ values.

Following Lee (1979), equation 1.3 can be re-written as such:

$$\ln(P_i) = \begin{cases} \boldsymbol{Z_i\alpha_z} + \boldsymbol{X_i\alpha_x} + \nu_i & \text{if } \boldsymbol{Z_i\alpha_z} + \boldsymbol{X_i\alpha_x} + \boldsymbol{Y_i\beta_y} + u_i > 0 \\ - & \text{if } \boldsymbol{Z_i\alpha_z} + \boldsymbol{X_i\alpha_x} + \boldsymbol{Y_i\beta_y} + u_i \leq 0 \end{cases} \tag{A.2}$$

The error terms, ν_i and u_i, are assumed to be serially independent, and normally distributed with mean 0, such that:

$$\begin{pmatrix} \nu_i \\ u_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\nu^2 & \sigma_{\nu,u} \\ \sigma_{\nu,u} & \sigma_u^2 \end{pmatrix}\right]$$

and σ_ν^2 is normalized to 1. The simultaneous equation system presented above can be written in a reduced form by substituting equation A.1 into equation 1.1. Hence, equation 1.1 can be written in reduced form as:

$$D_i^* = \boldsymbol{Z_i}\boldsymbol{\gamma_z} + \boldsymbol{X_i}\boldsymbol{\gamma_x} + \boldsymbol{Y_i}\boldsymbol{\gamma_y} + \epsilon_i \tag{A.3}$$

where $\boldsymbol{\gamma_z} = \dfrac{\beta_P\boldsymbol{\alpha_z}}{\sigma}$, $\boldsymbol{\gamma_x} = \dfrac{\beta_P\boldsymbol{\alpha_x} + \boldsymbol{\beta_x}}{\sigma}$, $\boldsymbol{\gamma_y} = \dfrac{\beta_y}{\sigma}$, $\epsilon_i = \dfrac{u_i + \ \beta_P\nu_i}{\sigma}$, and $\sigma^2 = E(\ u_i + \beta_P\nu_i)^2$.

Equation A.2 can now be presented in reduced form as follows:

$$\ln(P_i) = \begin{cases} \boldsymbol{Z_i\alpha_z} + \boldsymbol{X_i\alpha_x} + \nu_i & \text{if } \boldsymbol{Z_i}\boldsymbol{\gamma_z} + \boldsymbol{X_i}\boldsymbol{\gamma_x} + \boldsymbol{Y_i}\boldsymbol{\gamma_y} + \epsilon_i > 0 \\ - & \text{if } \boldsymbol{Z_i}\boldsymbol{\gamma_z} + \boldsymbol{X_i}\boldsymbol{\gamma_x} + \boldsymbol{Y_i}\boldsymbol{\gamma_y} + \epsilon_i \leq 0 \end{cases} \tag{A.4}$$

Since ϵ_i is truncated, then standard least squares estimation would lead to inconsistent estimates as is the case with standard Tobit models. On the other hand, Maximum Likelihood (ML) estimation produces estimates that are both consistent and efficient[1]. The error terms, ϵ_i and nu_i, are assumed to be jointly normally distributed and homoscedastic. The likelihood function to be maximized is as follows:

$$\mathcal{L} = \prod_{i}^{n} Pr(D_i^* \leq 0)^{1-D_i}[f(\ln(P_i)|D_i^* > 0)Pr(D_i^* > 0]^{D_i} \tag{A.5}$$

where $f(\ln(P_i)|D_i^* > 0)$ is the conditional density of $\ln(P_i)$ given $D_i = 1$, $Pr(D_i^* \leq 0)^{1-D_i}$ represents the likelihood over the observations where $D_i = 0$, and $f(\ln(P_i)|D_i^* > 0)Pr(D_i^* > 0]^{D_i}$ represents the likelihood over the observations where $D_i = 1$[2].

The ML results for both the parking price estimation equation and the selection equation, equations A.2 and A.4 respectively, are presented in Table A.3.

[1]The model could also be estimated using Heckman's two-step estimation, but the estimates would be inefficient, and, therefore, ML estimation is preferred.

[2]The density of parking prices is equal to the probability of the commuter driving, i.e. $Pr(D_i = 1)$, multiplied by the conditional probability of parking prices given that the commuter drives, i.e. $f(\ln(P_i)|D_i = 1)$ This results in the $f(\ln(P_i)|D_i = 1)Pr(D_i = 1)$ part of the likelihood function. This is raised to the power D_i because this part holds only if the commuters drive, i.e., this part holds only if $D_i = 1$. If the commuter does not drive, then $D_i = 0$, and so $f(\ln(P_i)|D_i = 1)Pr(D_i = 1)$ reduces to 1. Furthermore, if the commuter does not drive then that fact is the only thing observed. That is, if $D_i = 0$, then the density for parking prices is just the probability of the commuter not driving, which is represented by $Pr(D_i = 0)$. This is raised to the power $(1 - D_i)$ because it holds only when $D_i = 0$, and reduces to 1 if $D_i = 1$. Therefore, the likelihood is just the probability of the commuter not driving if the commuter does not drive, and is the product of the probability of driving and the condition probability of parking prices on driving if the commuter drives.

Table A.3: Selection Model Estimation)

Dependent Variable:	D_i	$\ln(Price)$
	(Eq A.3)	(Eq A.4)
ln(Income)	0.123***	0.006
	(0.030)	(0.007)
Age	0.005***	-0.001***
	(0.001)	0.000
Female	-0.048	0.033***
	(0.043)	(0.010)
High School	0.419***	0.002
	(0.112)	(0.032)
College Degree	-0.119**	0.01
	(0.050)	(0.012)
White	-0.001	0.025**
	(0.050)	(0.012)
ln(Distance)	0.142***	
	(0.019)	
Transit-Work Distance	0.014**	
	(0.006)	
Transit-Home Distance	0	
	(0.003)	
Peak	-0.175***	
	(0.043)	
Chain	0.153***	
	(0.055)	
Observations	5816	
Wald Chi^2	789.387	

Standard errors are reported in parentheses, and are clustered at the zipcode level
$*p < 0.1$, $**p < 0.05$, $***p < 0.01$
Estimations include a constant and zipcode fixed effects

Appendix B

Chapter 3 Appendix

Table B.1: MSAs

MSA Code	**MSA**
CA Sample	
31100	Los Angeles-Long Beach-Santa Ana, CA
40900	Sacramento–Roseville-Arden-Arcade, CA
41740	San Diego-Carlsbad, CA
41860	San Francisco-Oakland-Hayward, CA
41940	San Jose-Sunnyvale-Santa Clara, CA
NY-NJ Sample	
35620	New York-Northern New Jersey-Long Island, NY-NJ-PA
MA Sample	
14460	Boston-Cambridge-Quincy, MA-NH

Table B.2: Data Sources

Variable	Source
Transit Use	Travel Survey
Socioeconomic Characteristics	
Income	Travel Survey
Household Size	Travel Survey
Age	Travel Survey
White	Travel Survey
Female	Travel Survey
Neighborhood Characteristics	
Population Density	Constructed by Author
Population	ACS
Area	TIGER shapefile
Job Density	Constructed by Author
Jobs	LEHD WAC files
Area	TIGER shapefile
Commuter White Rate	CTPP
Median Commuter Income	CTPP
Transit Accessibility	
Distance to Rail Station	Constructed by Author
Workplace Location	Travel Survey
Residence Location	Travel Survey
Rail Station Location	NTAD, Transportation Agency shapefiles
Trip Characteristics	
Distance	Constructed by Author
Workplace Location	Travel Survey
Residence Location	Travel Survey
Time Difference	Constructed by Author
Workplace Location	Travel Survey
Residence Location	Travel Survey
Transit Duration	Google Maps API Distance Matrix
Driving Duration	Google Maps API Distance Matrix
Parking Prices	Constructed by Author
Workplace Location	Travel Survey
Parking Prices	ParkMe.com
Gas Expense	Constructed by Author
Workplace Location	Travel Survey
Residence Location	Travel Survey
Gas Price	GasBuddy.com
Driving Distance	Google Maps API Distance Matrix
Social Influence Measures	
Transit Use Rate	CTPP
Median Transit Income	CTPP
White Transit User Rate	CTPP

Table B.3: Data Sources for IV and Robustness Checks

Variable	**Source**
IV	
Weighted Average Commute Distance	CONSTRUCTED BY AUTHOR
Workplace Location	Travel Survey
Residence Location	Travel Survey
Jobs	LEHD OD Files
Robustness Checks	
Pedestrian-Oriented Intersections	EPA
Land Use Mix	EPA
Rail Station Count	CONSTRUCTED BY AUTHOR
Workplace Location	Travel Survey
Residence Location	Travel Survey
Rail Station Location	NTAD, Transportation Agency shapefiles
Gini Coefficient	ACS
Unemployment Rate	ACS
Population below 18	ACS