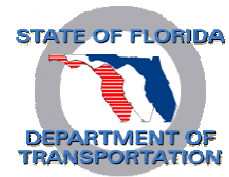


# Integrating Transit and Urban Form

## *Final Report*

Contract No. BD549, Work Order #37



December, 2008

### ***FDOT Project Manager:***

Amy Datz

### ***Report prepared by:***

Sisinnio Concas

Joseph S. DeSalvo

Center for Urban Transportation Research

University of South Florida, College of Engineering

4202 E. Fowler Ave., CUT100, Tampa, FL 33620-5375





*This research was conducted under a grant from the Florida Department of Transportation. The opinions, findings, and conclusions expressed in this report are those of the authors and not necessarily those of the Florida Department of Transportation.*

## Technical Report Documentation Page

1. Report No. BD 549 WO 37		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle <i>Integrating Transit and Urban Form</i>				5. Report Date September, 2008	
				6. Performing Organization Code	
7. Author(s) Concas, Sisinnio, DeSalvo, Joseph S.				8. Performing Organization Report No. 7716	
9. Performing Organization Name and Address National Center for Transit Research Center for Urban Transportation Research University of South Florida 4202 E. Fowler Ave., CUT 100 Tampa, FL 33620-5375				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No. BD 549 WO 37	
12. Sponsoring Agency Name and Address Florida Department of Transportation, Office of Public Transportation Transit Planning Program Florida Department of Transportation Research Center 605 Suwannee Street, MS 30 Tallahassee FL 32399-0450				13. Type of Report and Period Covered Final Report  02/01/07-12/31/08	
				14. Sponsoring Agency Code	
15. Supplementary Notes FDOT Project Manager: Amy Datz					
16. Abstract This study develops an integrated behavioral model of transit patronage and urban form. Although herein focused on transit, the framework can be easily generalized to study other forms of travel. Advanced econometric methods are used to test specific behavioral hypotheses developed in the theoretical models. Findings are then summarized in a succinct fashion showing relevance and magnitude of the impact of land use on transit demand. The empirical models also quantify these relationships in the form of point elasticity estimates that can be used as indicators of the relevance of transit supply measures. This work summarizes the study results, an exposition of the methodology and tables that lay out the findings in a readily accessible format.					
17. Key Word Transit, density, urban form, land-use, residential location, self-selection				18. Distribution Statement No restrictions	
19. Security Classif. (of this report) Not classified		20. Security Classif. (of this page)		21. No. of Pages 160	
				22. Price	

## Acknowledgements

*The authors wish to thank Dr. Xueaho Chu, Janet L. Davis, Sara J. Hendricks, Stephen L. Reich, Joel Volinski, and Philip L. Winters for their review and comments to this report; Dr. Steven E. Polzin for reviewing internal technical memoranda; and all staff members at CUTR for their critical support throughout the project.*

### **FDOT Project Manager**

Amy Datz

Florida Department of Transportation, Office of Public Transportation

### **CUTR Project Team**

Principal Investigator: Sisinnio Concas

Co-Principal Investigator: Joseph S. DeSalvo

### **Working Group**

Transportation Program Evaluation and Economic Analysis

*[This page intentionally left blank.]*

# Executive Summary

---

This study develops a behavioral model of transit patronage and urban form. Although focused on transit, the framework can be easily generalized to study other forms of travel.

The research begins with a synthesis of past and current empirical work on the relationship between urban form (i.e., employment density, population density, land-use mix, urban design) and travel behavior. The findings of this review show that there has been a shift from the study of density threshold levels that make transit cost feasible to an analysis of the effect of urban design and land-use mix on travel behavior, after controlling for density levels. The issue is no longer at what density thresholds it makes sense to implement transit, but what is the best set of policies affecting urban design and land-use mix that most influences the spatial arrangements of activity locations, so that individuals are more likely to utilize transit. This shift is reflected by an exponentially increasing number of studies dedicated to studying the relevance of transit-oriented development (TOD) policies in a context where households or individuals tend to prefer certain urban setting to others.

While early work sought to provide a generalized analytical framework that made use of aggregate data, the more recent literature consists of papers that model the simultaneous decision of location and travel in a context where individuals choose locations based on idiosyncratic travel preferences.

Finally, there is a lack of empirical work that examines the relationship between urban form and travel behavior within an analytical framework that takes into account the complexity of travel by considering trip chaining among other travel complexities.

## Behavioral Framework

This report presents a behavioral model of transit patronage in which residential location, travel behavior, the spatial dispersion of non-work activities (such as shopping and recreational activities), and urban form are all simultaneously determined. In a departure from the monocentric urban model, residential location is defined as the optimal *job-residence pair* in an urban area in which jobs and residences are dispersed. Following urban residential location theory, the location decision is assumed to be the result of a trade-off between housing expenditures and transportation costs, given income and the mode-choice set. The location decision is also based on idiosyncratic preferences for location and travel. In addition to determining optimal residential location, this approach also determines the optimal sequence of non-work trip chains, goods consumption, and transit patronage.

In this model, travel demand is considered an *indirect* demand brought about by the necessity to engage in out-of-home activities whose geographical extent is affected by urban form. Furthermore, budget-constrained utility maximizing behavior leads to an optimization of the spatiotemporal allocation of these activities and an optimal number of chained trips. Socio-demographic factors directly influence residential location, consumption, and travel behavior.

Within this framework, the report addresses questions related to the interrelation between urban form, residential location, and transit demand. How do location decisions affect travel behavior? How does urban form relate to travel behavior? Do residential location and urban form affect travel behavior? What is the impact of higher density on travel behavior? To address these questions, this research begins with a travel-demand model treating residential location and density as determined outside the model and not directly affected by travel behavior. Subsequent extensions that relax these assumptions are then introduced and expected behavioral conclusions are reached.

In its final specification, the model treats transit station proximity as influenced by residential location decisions. Transit supply is introduced through station proximity. Incorporation of transit supply and demand produces a general equilibrium model of transit travel behavior.

The research reported here is the first example of empirical work that explicitly relates location to trip chaining behavior in a context where individuals jointly decide location, the optimal trip chain, and the area of non-work activities, based on the optimal trade-off between commute time, leisure, and non-work travel activities.

## Findings

This study presents three theoretical models and their empirical estimation. It develops a comprehensive dataset that integrates travel data with land-use data. It also includes geographic measures of the spatial dispersion of out-of-home activities. The first model considers residential location and density as exogenous to the system. This means that residential location and density levels are not directly affected by travel decisions. The second model relaxes the assumption of exogenous location, and the last model considers both location and density endogenous.

Theoretical and empirical findings provide evidence of a significant causal influence of land-use patterns on transit patronage, which, in turn, affects consumption and non-work travel. It is found that gross population density does not have a large impact on transit demand and that the relative magnitude of the effect decreases when residential location is treated as



endogenous. A 20 percent increase in gross population density (or an increase of about 1,830 persons per square mile) increases transit demand by 5.4 to 9.5 percent.

The relevance of land-use policies geared at influencing transit patronage by providing a mix of residential and commercial uses is highlighted by the elasticity of travel demand with respect to changes in retail establishment density. Results show that a 20 percent increase in retail establishment density (or an increase of about 28 establishments per square mile) increases transit demand by 3.4 percent.

The importance of station proximity to the household residential unit is lessened after accounting for idiosyncratic preferences for location. When residential location and density are determined within the model, the elasticity of transit demand with respect to walking distance decreases by about 33 percent over the result of the basic model, in which density and location are determined outside the model. This decline in elasticity's magnitude is the result of explicitly accounting for residential self-selection effects.

After controlling for socio-demographic factors, it is found that households living farther from work use less transit. It is shown that trip-chaining behavior explains this observation. Households living far from work engage in complex trip chains and have, on average, a more dispersed activity space, which requires reliance on more flexible modes of transportation. Therefore, reducing the spatial allocation of activities and improving transit accessibility at and around subcenters would increase transit demand. Similar effects can be obtained by increasing the presence of retail locations in proximity to transit-oriented households.

An established central business district (CBD) is still a relevant driver of transit use, as highlighted by an elasticity of transit demand with respect to distance to the CBD of  $-1.17$ . Although subcenters play a less relevant role, they support evidence to providing services to decentralized areas to increase ridership.

Transit-oriented development has a positive and statistically significant impact on transit use. The presence of a TOD station at the point of origin increases transit demand by about 28 percent. The presence of a transit station in proximity to a workplace also has a significant positive impact on ridership, as indicated by the magnitude of the proportional changes across all three models.

## **Contributions of this Research**

The major contribution of this research effort is the development of a general equilibrium behavioral model of transit patronage and land-use that acknowledges the interrelationship between travel behavior and urban form. In particular, the framework embraces the paradigm

shift from trip generation to activity-based modeling by considering travel demand as a derived demand for out-of-home activities. In addition, this framework:

- departs from the monocentric models of residential location that do not account for increasingly decentralized urban settings by explicitly acknowledging both the presence and the relevance of subcenters;
- accounts for the trade-off between consumption and travel brought about by the finite nature of time and its allocation among household members;
- shifts the analysis from individual travel behavior to household travel behavior;
- can accommodate extensions to account for the endogeneity of time allocation across activities and households; and,
- takes advantage of the advances in geographic information systems (GIS) tools and geographic science contributions to the spatial analysis of the interactions of travel behavior and urban form.

The consequences introduced by this structure are not trivial. For example, density is no longer assumed directly to affect the demand for travel. Rather it is assumed to represent a constraint in a utility-maximization process where individuals optimally determine consumption and travel. This framework also explicitly acknowledges the changes in the urban environment that have occurred over the last 40 years due to suburbanization by adopting a polycentric, rather than monocentric, urban model.

## **Directions for Further Research**

Notwithstanding the validity of the statistical tests performed in this study, there still exists the possibility that some of the variables treated as determined outside the model are, in fact, to be considered as being shaped by travel and location decisions. For example, while this study treats vehicle ownership as predetermined and not directly influenced by the location decision, the literature review encountered studies that considered vehicle ownership as affected by the residential location process and density levels. One extension to the research would be to extend this treatment to this and other mode-choice variables. Additional extensions would also account for the joint allocation of time use among household members.

# Table of Contents

---

<b>Executive Summary .....</b>	<b>v</b>
Behavioral Framework .....	v
Findings .....	vi
Contributions of this Research .....	vii
Directions for Further Research.....	viii
<b>Chapter 1   Introduction and Summary .....</b>	<b>1</b>
Background .....	1
Objectives.....	2
Research Method.....	2
Report Organization.....	3
<b>Chapter 2   Land Use and Transit Patronage: The Role of Urban Form .....</b>	<b>5</b>
Introduction .....	5
Studies Analyzing the Influence of Urban Form on Transit Patronage .....	7
Existing Critical Literature Reviews.....	11
Studies Analyzing the Influence of Transit on Urban Form .....	13
Studies Analyzing the Contemporaneous Relationship between Transit and Urban Form .....	14
Inherent complexity: accessibility, urban design, and self-selection .....	18
Measuring accessibility .....	22
Urban Form Measures and Polycentric Cities .....	23
From Trip Generation to Activity-travel Behavior .....	25
Summary and Implications for Integrated Models of Transportation and Land-Use .....	27
<b>Chapter 3   Methodology .....</b>	<b>29</b>
Introduction .....	29

Model I: Exogenous Residential Location and Population Density .....	31
Residential location, $RL$ , and transit station proximity, $WD$ .....	32
Activity space: spatial dispersion of non-work activities .....	33
Trip chaining, $TC$ .....	33
Travel demand, $TD$ .....	35
Comparative-Static Analysis .....	35
Effects of an increase in density, $D$ .....	36
Change in residential location, $RL$ .....	37
Change in walking distance to nearest station, $WD$ .....	38
Model II: Endogenous Residential Location, Exogenous Density .....	38
Comparative-Static Analysis .....	39
Effects of an increase in density, $D$ .....	39
Change in walking distance to nearest station, $WD$ .....	39
Model III: Endogenous Residential Location, Endogenous Density .....	40
Comparative-Static Analysis .....	40
Conclusions .....	41
<b>Chapter 4 Findings .....</b>	<b>43</b>
Introduction .....	43
Summary of Data .....	43
Dependent Variables Descriptive Statistics .....	45
Measures of activity space, $AS$ .....	45
Measures of residential location, $RL$ .....	46
Measures of transit station proximity, $WD$ .....	47
Measures of density, $D$ .....	47
Explanatory Variables Descriptive Statistics .....	49
Socio-demographic variables .....	49
Travel behavior variables .....	51
Urban form variables .....	51

Transit supply variables .....	54
Method of Analysis .....	58
Model I results .....	60
Model II results .....	63
Model III results .....	67
<b>Chapter 5 Validity of Hypotheses .....</b>	<b>71</b>
Introduction .....	71
Dataset Issues .....	71
Measurement issues .....	71
Measurement scale issues .....	72
Modeling Issues: Post Estimation Tests .....	73
Tests of endogeneity and overidentification .....	74
Other Issues .....	76
Transit station proximity .....	77
Implications .....	78
<b>Chapter 6 Conclusions .....</b>	<b>81</b>
Research Contributions .....	83
Directions for Further Research .....	83
<b>References .....</b>	<b>85</b>
<b>Appendix A Table of Relevant Studies .....</b>	<b>93</b>
<b>Appendix B Comparative Static Analysis .....</b>	<b>99</b>
<b>Appendix C Equation Identification .....</b>	<b>113</b>
<b>Appendix D On the Relevance of Transit Station Proximity .....</b>	<b>127</b>
<b>Appendix E Glossary of Terms .....</b>	<b>142</b>

## List of Tables

---

<b>TABLE 2.1</b> Transit Service and Residential Density Thresholds .....	9
<b>TABLE 2.2</b> Report 16: Summary of Elasticities of Rail Station Boardings with Respect to Density .....	10
<b>TABLE 2.3</b> Elasticities of Travel with Respect to Built Environment.....	13
<b>TABLE 3.1</b> Comparative Static Results .....	41
<b>TABLE 4.1</b> Descriptive Statistics: Overall Sample Means .....	48
<b>TABLE 4.2</b> Descriptive Statistics: Sample Means of Dependent Variables and Selected Trip Measures.....	48
<b>TABLE 4.3</b> Summary of Selected Demographic Variables .....	50
<b>TABLE 4.4</b> Urban Form Variables.....	52
<b>TABLE 4.5</b> Land Use Variables by Household Type.....	55
<b>TABLE 4.6</b> Land Use Variables by Transit Station Proximity .....	55
<b>TABLE 4.7</b> List of Variables for Model Estimation .....	57
<b>TABLE 4.8</b> 2SLS Regression Results—Model I.....	62
<b>TABLE 4.9</b> Elasticity Estimates—Model I .....	63
<b>TABLE 4.10</b> 2SLS Regression Results—Model II.....	65
<b>TABLE 4.11</b> Elasticity Estimates—Model II .....	66
<b>TABLE 4.12</b> 2SLS Regression Results—Model III.....	68
<b>TABLE 4.13</b> Elasticity Estimates—Model III .....	69
<b>TABLE 5.1</b> Land Area Geographic Measures.....	73
<b>TABLE 5.2</b> Endogeneity and Overidentification Tests .....	76
<b>TABLE 6.1</b> Relevant Land-Use and Transit-Supply Elasticities.....	82

## List of Figures

---

<b>FIGURE 3.1</b> Conceptual Model of Urban Form and Travel Behavior.....	30
<b>FIGURE 3.2</b> Trip Chaining Sequence .....	34
<b>FIGURE 4.1</b> Standard Distance Circle and Standard Distance Ellipse .....	46
<b>FIGURE 5.1</b> Poverty and Transit Station Proximity .....	78

# Chapter 1 Introduction and Summary

---

*“The history and fortunes of mass transit are intricately related to developments in transportation technology and also technology’s effect on urban structure.”*

*Ian Savage*

## **Background**

Despite a significant amount of academic and practitioner-oriented research, the practice of choosing the right transit service to support desired development and the right development to support transit ridership relies on findings that no longer apply to the current urban landscape. Early studies estimated the housing and job densities necessary to support different transit modes. Such studies did not consider changes in urban structure, such as transit-oriented development, that have recently emerged. At the same time, the urban landscape has evolved from monocentricity, where the CBD is the sole employment center, to polycentricity, where multiple employment centers characterize an urban area and where households can locate anywhere in an increasingly suburban environment. Employment decentralization, coupled with the increased relevance of non-work travel, has had a profound impact on the way transit responds to urban form, making the earlier studies obsolete.

The debate has shifted from the need to determine minimum density thresholds to the need to provide reliable information to guide decision makers about what mix of land-use policies would better promote transit use. In most previous work, density is treated as exogenous and is assumed not to be impacted by transportation system changes. It is now recognized that this approach is inadequate and that what is needed is an empirically estimable behavioral model conducive to generalization and applicability.

The bulk of previous research is empirically oriented. It uses multivariate techniques to estimate the effect of measures of travel behavior (commute length, vehicle-miles of travel, mode choice) on measures of residential and employment density, while controlling for travelers’ demographic characteristics. These studies report the statistical significance, sign, and magnitude of the estimated coefficients. A statistically significant negative coefficient leads one to conclude that an inverse relationship exists between travel and density, that is, higher density leads to shorter commutes, fewer vehicle-miles of travel (VMT), or a shift from auto transportation to alternative modes, such as transit. The abundance of such studies has led to the conclusion that policy interventions to increase density would reduce automobile use. These studies have, however, undergone systematic criticism, mainly of their ad hoc

specifications and failure to recognize that the relationship between urban form and travel might entail simultaneity and endogeneity.

In addition to the widely investigated roles of employment and residential density, other factors that affect travel behavior should be examined. The goal of reducing auto travel has been sought in policies that change urban form. The rationale for these policies is that car travel can be reduced by reducing trip frequencies and travel distances. Mixed land use (where residential and commercial land uses are in close proximity) is hypothesized to reduce average travel distances, as nearby destinations will be preferred to more distant ones. Furthermore, increased public transportation choices are supposed to further reduce auto travel. These policies form what is currently defined as transit-oriented development. At the heart of transit-oriented development (TOD) effectiveness is the issue of individual self-selection to residential location. Ignoring idiosyncratic preferences regarding residential location in empirical research might lead to overestimation of the impact of TOD policies on travel behavior.

## **Objectives**

This study begins with a synthesis of academic research that examines the relationships between transit design and urban form. Conducted as a critical assessment of the methodological issues affecting past and current research, this synthesis provides the basis from which to develop behavioral models of the relationship between transit and urban form. With this as background, this study develops and estimates simultaneous equation general equilibrium models of the relationship between transit patronage and urban form. This task is accomplished acknowledging that the interactive nature of transit and urban form, while complex, can potentially be conveyed through a balance of modeling work and carefully constructed empirical investigations that look at the joint influence of transit on residential location and ridership (TCRP, 1995).

## **Research Method**

This study provides an assessment of research conducted to date on the relationship between urban form and travel behavior with a specific focus on transit patronage. The literature review is intended to provide a critical assessment of the various methodologies employed, the explicit control for relevant factors associated with transit patronage, and the general validity of findings.

The results of this critical review provide a basis from which to develop a generalized behavioral framework describing the relationship between urban form, residential location, and the demand for travel. Given the objective of this study, the focus is on the demand for transit services. Nevertheless, the framework is suited to explain the determinants of the demand for



travel in general. Advanced econometric methods are used to test specific behavioral hypotheses developed in the theoretical models. Findings are then summarized in a succinct fashion showing relevance and magnitude of the impact of land use on transit demand. The empirical models also quantify these relationships in the form of point elasticity estimates that can be used as indicators of the relevance of transit supply measures.

## **Report Organization**

Chapter 2 reviews the literature on the relationship between transit patronage and urban form. Chapter 3 presents an analytical framework of residential location, consumption, and travel behavior. Chapter 4 proceeds to test the relationships hypothesized in the previous chapter. Chapter 5 discusses the validity of the empirical work and identifies issues that might potentially affect a generalization of the findings. Chapter 6 concludes and provides direction for further research.

*[This page intentionally left blank.]*

## Chapter 2 Land Use and Transit Patronage: The Role of Urban Form

---

### Introduction

The role of urban form is essential in the way it affects the spatial extent of activities and the ensuing needs and preferences for travel. The demand for transit and the demand for travel, in general, are brought about by the necessity to engage in activities that are located outside one's place of residence and that require travel. As such, the demand for travel is considered as an indirect demand (Domencich and McFadden, 1975, McFadden, 1974). This recognition requires the study of the determinants of the demand for out-of-home activities as well as the characteristics of the environment affecting the choice of one particular mode of transport over another. In this context, urban structure affects the demand for transit in two ways. First, the location of employment affects the probability that an individual will choose transit, given the supply of this service. Second, the spatial location of activities affects transit use for non-work travel purposes.

The relevance of the urban environment on location and travel decisions has long been recognized within academic disciplines, such as urban economics, urban planning, geography, and travel behavior research. The study of the influence of urban form on travel behavior is rendered more complicated by the evolution of the built environment itself. Indeed, since early development of the monocentric-based theory of location, the urban landscape has evolved into one where multiple employment centers characterize an urban area, and where households can locate anywhere in an increasingly suburban environment. This is reflected in literature that increasingly looks at the reasons behind the formation of suburban centers or subcenters, often defined as polycentric theory of location (Anas, et al., 1997, Anas and Xu, 1999, Fujita, et al., 1997, McMillen, 2001). In this context, it becomes even more arduous to ascertain the relationship of transit and urban form. As will be seen, transit patronage is still assumed to be largely dependent upon the presence of major employment centers; although, the literature is increasingly looking at how to best implement transit services in a polycentric urban landscape (Casello, 2007, Modarres, 2003).

The purpose of this literature review is to provide an evaluation of the state of research starting from the publication of the Transit Cooperative Research Program Report 16 (henceforth Report 16) (1996a, 1996b). The literature review is concerned with studies that look at the influence of transit on urban form (such as impacts on population and employment density), the influence of land-use on transit patronage, and studies that comprehensively evaluate the relationship between transit and urban form.

In the first part of this chapter, studies that test the hypothesis that land-use influences transit patronage are reviewed. Throughout this review, the term urban form refers to various measures of land-use density and urban design. Land-use density encompasses both residential and employment densities, while the term urban design refers to both the characteristics and arrangements of land-uses that affect accessibility to both transit services and activity locations. Transit use comprises measures such as transit trips, both linked and unlinked (per person, household, acre or square mile), station boardings (at station level), or mode shares.

First, Report 16 findings as well as results from relevant literature prior to its publication are briefly summarized. To avoid duplicating the effort of Report 16, this literature review focused on a relatively smaller number of studies completed before 1995 and a more comprehensive review covering the period after 1995 on the most relevant empirical work published in the literature of travel behavior and urban form. The literature review is intended to provide a critical assessment of the various methodologies employed, the explicit control for relevant factors associated with transit patronage, and the general validity of findings. Below is a summary of the major findings of this review.

The findings of this review show that there has been a shift from the study of density threshold levels that make transit cost feasible to an analysis of the role of urban design and land-use mix, after controlling for density levels. The issue is no longer at what density threshold it makes sense to implement transit, but what is the best set of policies affecting urban design and land-use mix that best influences the spatial arrangements of activity locations so that individuals are more likely to utilize transit. This shift is reflected in the relevant literature, where a growing number of studies are dedicated to studying the relevance of transit oriented development (TOD) policies in a context where households or individuals tend to prefer certain urban settings to others. Not accounting for these inherent idiosyncratic preferences prevents the unraveling of the true impact of TOD. As it will be seen, this shift has resulted in an exponentially increasing number of papers published since 2006.

In organizing a summary of literature, several typologies were considered. For example, studies can be organized by analytical method (simulations vs. regressions, anecdotal vs. analytical, etc.), by the characteristics of urban form (urban design, land-use mix) or by the type of travel measure considered (mode shares, VMT, boardings, person trips, etc.). As recognized by Boarnet and Crane (2001), there is no single best rationale for choosing one typology or another, as all are effective formats for identifying and measuring the influence of urban form on travel. Ultimately, a typology that looks at the relationship between urban form and travel was made that takes into consideration the methodological aspects of scholar work in this topic. Therefore, the literature review distinguishes between studies that have only looked at

the influence of urban form on travel behavior, studies that have considered the impact of transit use on land-use, and those that have looked at the simultaneous nature of such relationship. This characterization also inherently considers the difference in the statistical or econometric methods herein employed.

As such, this literature review is not fully comprehensive of all work conducted in this area of research as it does not take into consideration empirical work that solely looks at anecdotal accounts or descriptive analyses, without, at a minimum, presenting an analytical framework of any sort. Descriptive studies have the benefit of assessing actual behavior without the need to a priori establish any causal links, but are limited in providing any useful perspective or guidance in the development of a theoretical or analytical model. Thus, these studies are not deemed relevant to the objectives of this research effort. An assessment and review of recent anecdotal studies has been informally conducted by Taylor and Miller (2003). For the same reason, Transit Cooperative Research Program (TCRP) reports that discuss the impact of the built environment on physical activity (Committee on Physical Activity, 2005) or report successful case studies of TOD projects (Evans, et al., 2007) are not herein summarized.

Given the large number of papers and studies reviewed and to facilitate a faster browsing of this review, Appendix A of this report succinctly summarizes papers and studies in the form of an annotated table that describe each study, the data used, the results and any associated methodological issues.

### **Studies Analyzing the Influence of Urban Form on Transit Patronage**

TCRP Report 16, published in 1996, presents results from Project H-1, *An Evaluation of the Relationships between Transit and Urban Form*. Report 16 consists of two volumes, each containing two reports. The first volume includes a comprehensive literature review of studies analyzing the relationship between transit and urban form for the period to 1995. The second volume consists of a practitioner's guidebook on patterns of development that encourage transit patronage and on mode accessibility and catchment areas for rail transit.

The most important work prior to Report 16 is Pushkarev and Zupan (1977). This publication presented "land-use thresholds" at which different types of transit were feasible investments. The methodology used single-equation ordinary least square (OLS) regression analysis. The choice of this method was dictated by the paucity of data available at the time as well as the desire to present results as nomograms. A nomogram is a graph with which one can find the value of a dependent variable given the values of two or more independent variables, with only the use of a straightedge. The nomograms were designed to facilitate a planner's choice of a feasible transit alternative, given values of current or expected density levels and other relevant variables.

The determinants of transit demand used by Pushkarev and Zupan were the size of the central business district (CBD), measured in non-residential floor space; the distance of a site from the CBD; and residential density. The study also accounted for socio-demographic characteristics affecting transit patronage, such as vehicle ownership levels, household size and income.

There are problems with this study as well as with a later one by Pushkarev and Zupan, (1982). As part of this review, CUTR researchers tried to replicate numerical examples in the study, but encountered several problems that made replication impossible. These problems are similar to those of papers reviewed later. An important example of such a problem is the lack of a formalized behavioral framework, a deficiency that in turn results in poorly specified empirical equations.

In a subsequent update of their 1977 study, Pushkarev and Zupan (1982), examined the feasibility of fixed guide-way transit under the assumption that all work travel was to the CBD. This assumption would be quite restrictive today, given the multi-centered character of many metropolitan regions.

Table 2.1 presents the relationship between residential density and different types of transit as estimated by these two studies. These results are still widely cited and employed in determining feasibility of proposed rail projects.

**TABLE 2.1** Transit Service and Residential Density Thresholds

<i>Mode</i>	<i>Service Levels</i>	<i>Density Thresholds (dwelling units per residential acre)</i>	
		<i>Pushkarev &amp; Zupan (1977)</i>	<i>Pushkarev &amp; Zupan (1982)</i>
Local Bus	Minimum service <sup>a</sup>	<b>4</b>	<b>4</b>
Local Bus	Intermediate service <sup>b</sup>	<b>7</b>	<b>7</b>
Local Bus	Minimum service <sup>c</sup>	<b>15</b>	<b>15</b>
Light Rail	5-min peak headways <sup>d</sup>	<b>9</b>	<b>9</b>
Rapid Transit	5-min peak headways <sup>e</sup>	<b>12</b>	<b>12</b>
Commuter Rail	20 trains/day <sup>f</sup>	<b>1 to 2</b>	<b>1 to 2</b>

<sup>a</sup>0.5 miles between routes, 20 buses/day. <sup>b</sup>0.5 miles between routes, 40 buses/day. <sup>c</sup>0.5 miles between routes, 120 buses/day. <sup>d</sup>Average density for a corridor of 25 to 100 square miles. <sup>e</sup>Average density for a corridor of 100 to 150 square miles. <sup>f</sup>Service only to largest downtowns, if rail exists.

Zupan et al. (1996) provide guidance on the land-use characteristics that could cost-efficiently support new fixed-guideway transit services. The authors find that, in a transport corridor, ridership rises exponentially with both CBD employment and employment density. Separate models are presented for light rail and commuter rail. For both models, the dependent variable is a natural log transformation of total daily transit boardings for 261 stations across 19 rail lines located in 11 cities. The results of the two models differ because they use different datasets and independent variables, as may be seen in Table 2.2.

**TABLE 2.2** Report 16: Summary of Elasticities of Rail Station Boardings with Respect to Density

<i>Type of Rail</i>	<i>Transit Use Measure</i>	<i>Station Area Residential Density (Persons/Acre)</i>	<i>CBD Employment Density</i>
<i>Light Rail (19 lines, 11 regions)</i>	Station Boardings	0.51	0.34
<i>Commuter Rail (47 lines, 6 regions)</i>	Station Boardings	0.18	0.64

Source: (Zupan, Cervero, Parsons Brinckerhoff Quade & Douglas and Howard/Stein-Hudson Associates, 1996), *Commuter and Light Rail Transit Corridors: The Land-use Connection*, Table 11, p. 25.

Multicollinearity impairs the reliability of these estimates, as recognized by the authors themselves. Also, lack of causality is a problem, for the estimated elasticities merely support a direct relationship between transit patronage and population density. This causality problem, which affects most findings in this research field, is discussed in a later section of this report. Finally, the authors do not employ a model that accounts for inherent, unobserved region-specific characteristics that might affect the reliability of estimates. A fixed effect model controlling for transit provider unobserved heterogeneity could provide a superior model.

Using Report 16 as a reference, Kuby et al. (2004) analyze the determinants of light rail transit ridership with a multiple regression model using weekday boardings for 268 stations in nine cities. For each city, five categories of independent variables accounting for land-use and other factors are used. The authors hypothesize that employment within walking distance of each station is the most important factor for work trips. The model also controls for the relevance of nearby airports and for city-specific unobserved effects that might affect weekly boarding, such as the presence of an international airport. The study finds that an increase of 100 persons employed within walking distance of a station increases boarding by 2.3 passengers per day while an increase of 100 persons residing within walking distance of a station increases boardings by 9.2 passengers per day. The study also finds higher residential population to be associated with higher weekly boardings and that the CBD variable is not statistically significant, indicating that centrality is no longer relevant in determining light rail ridership. This result could, however, be due to faulty test statistics produced by the high correlation between the model's measures of centrality and the CBD dummy.

Kuby et al. made some important improvements to the methodology of Report 16. First, they captured the effect of the CBD on boardings by introducing a dummy variable for CBD location. In contrast, Report 16 only examined ridership at non-CBD stations. Second, Kuby et al. included employment near non-CBD stations. Report 16 included employment within the



CBD, but it ignored employment around other stations. Third, they included accessibility to non-CBD stations. Report 16 computed distances from the stations to the CBD, but it ignored stations' accessibility to other stations. Finally, Kuby et al. used residential population within CBD as an independent variable, while Report 16 did not.

The studies previously discussed used aggregate data. The increasing availability of disaggregate (or micro) data after about 1995 provided the opportunity to study travel behavior at the individual or household level. As discussed further in this literature review, the increased amount of disaggregate data brought about a paradigm shift in travel behavior analysis.

Reilly and Landis (2002) provide an early use of micro data to study the relationship between urban form and travel. In a study of the 1996 Bay Area Travel Survey (BATS96), San Francisco, they tested the relationship between measures of urban form and mode choice. Using geographic information system (GIS) methods, they obtained small scale measurements of land-use diversity, intersection density, and average lot size. To obtain these measurements, a map of the study area was created, which was subdivided into a set of grid-cells, called rasters. A cell's dimension of 10,000 square meters was determined on an ad hoc basis. Each land-use measurement, such as the number of transit stops within a grid-cell, was obtained at the raster level. Furthermore, a buffer of a given radius (usually 0.25 or 0.5 miles) was centered on each cell, and the values of variables of the neighboring cells included within the buffer were summed. Census land-use and demographic measures at the block-group level were treated in the same fashion. For example, the authors fit gross population density and the amount of residential land area at the census block level into the grid unit level to compute density values within the buffer. The results of a multinomial logit mode-choice model show that an increase in average density of 10 persons per hectare (about four persons per acre) within one mile of an individual's residence is associated with a 7 percent increase in the probability of walking or taking transit (p. 24). As in most of the studies reviewed, this study does not determine causality between urban form and travel behavior.

## **Existing Critical Literature Reviews**

While culling the literature, relevant reviews were also considered as they summarize the most relevant empirical work covering this transportation research topic. Recent reviews of the literature are provided by Crane (2000), Boarnet and Crane (2001), Badoe and Miller (2000), and Ewing and Cervero (2001), who summarize the most relevant empirical work published in the literature of transportation research. These reviews made focusing on those papers more relevant to the objectives of this study.

In particular, Crane (2000) presents a discussion of key studies of urban form and travel behavior. He describes research methods, data, and results by dividing empirical work into two main categories: ad-hoc studies and theoretical-oriented (or demand derived) studies. The review is focused only on studies that make use of statistical techniques to uncover any relationship between travel behavior and urban form. He concludes that while most studies generally show that higher density patterns are correlated with less car travel, these findings are plagued by methodological issues all conducive to a lack of a behavioral framework. These ad-hoc studies are typically difficult to generalize and lack sufficient robustness to be used as a basis for policy. These findings are used by Crane to justify the development of a behavioral framework that is more grounded and theoretically consistent with a microeconomic theory of demand and travel, as discussed in detail in the next section of this chapter.

Badoe and Miller (2000) review the empirical literature until 2000 with the objective of pinpointing the shortcomings that lead to what are considered questionable and contradictory results. The analysis deals with studies on the relationship between land-use and travel behavior at the macro (density) and micro (design) levels. Just as Crane, Badoe and Miller uncover weaknesses either in the data used or the methodology employed. For example, some studies have worked with zonal-aggregate variables for gross spatial units that are not homogeneous with respect to neighborhood design, land-use and socioeconomic characteristics, which increase data measurement errors. Other studies have not considered relevant variables, such as measures of transit supply, falling into omitted variable bias issues.

Ewing and Cervero (2001) also summarized more than 50 empirical studies that have examined the linkages between urban form and travel behavior. They focus on presenting findings that, at a minimum, “make some effort to control for other influences on travel behavior (p. 870).” This approach leads to the classification of papers covering the period until 2000. The review does not cover papers that explicitly treat trip chaining behavior because of a lack of empirical work relating trip chaining to land-use and design variables. The relevance of accounting for trip chaining behavior in a theoretical framework that considers the joint determination of the spatiotemporal allocation of non-work activities, residential location, and travel behavior is discussed in detail in Chapter 3 of this report. Table 2.3 reports a summary of urban form elasticities summarized by Ewing and Cervero as generated by the literature they reviewed. The variables presented in this table represent composite measures obtained by principal component analysis. For example, the density measure (residents plus employees/land area) is used to represent the construct *density*.

**TABLE 2.3** Elasticities of Travel with Respect to Built Environment

	<i>Vehicle Trips (VT)</i>	<i>Vehicle Miles Traveled (VMT)</i>
Local Density (residents plus employees/ land area)	-.05	-.05
Local Diversity (Mix; job-population balance measure)	-.03	-.05
Local Design (construct for street network density; sidewalk completeness; route directness)	-.05	-.03
Regional Accessibility (gravity-based index)	--	-.20

### Studies Analyzing the Influence of Transit on Urban Form

Other research looked into the influence that transit has on urban form. In this context, the vast majority of studies looked at impacts on urban form in terms of changes in land values at station-area level (Baum-Snow and Kahn, 2000, Bollinger, 1997, Bowes and Ihlanfeldt, 2001, Cervero and Landis, 1997, Kahn, 2007, McDonald and Osuji, 1995, Nelson, et al., 2007, Zheng and Kahn, 2008). The vast majority of these studies also looks at the economic benefits of rail systems at the regional and local level. Regional growth may be impacted to the extent that transit improves productivity within a region; local impacts are related to accessibility improvements to locations of interest within a region.

Report 16, Part II, which is dedicated to ascertaining any evidence of transit influencing urban form, found evidence that rail transit impacts residential property values near stations. Furthermore, there is support that both CBDs and subcenters have benefited from transit development at the station-area level. In the case of CBDs, transit development helped centers retain their dominance. In the case of subcenters, there is evidence that regional rail systems contributed to the decentralization of both population and employment. This evidence is provided by way of anecdotal case studies, not empirical investigations based on quantitative analysis.

In an in-depth analysis of the Bay Area Rapid Transit (BART) system, Cervero and Landis (1997) found that transit investment had localized impacts on land-use that were limited to down-town San Francisco, Oakland, and a few subcenters. Some studies looked at gentrification effects associated with transit system implementation. For example, Kahn

(2007), shows that access to transit in the form of “Walk and Ride” positively impacts the gentrification trend.<sup>1</sup>

A more sophisticated analysis of the impact of rail transit on economic development is provided by Bollinger and Ihlanfeldt (1997) who present a simultaneous equation model that accounts for the simultaneity between population and employment density in proximity to rail stations of Atlanta’s MARTA (defined by a ¼ mile radius buffer). Results indicate that MARTA has had no discernible impact on total employment and population around stations.

## **Studies Analyzing the Contemporaneous Relationship between Transit and Urban Form**

Apart from the instances outlined above, the vast majority of empirical work on the relationship between transit and urban form (meaning density, urban design, and land-use) looks at this relationship as one where it is the urban landscape that influences, in a unidirectional fashion, transit implementation levels. Existing critical literature reviews have identified the shortcoming of these approaches, which fail to account for any underlying unobserved endogeneity between urban form and travel.<sup>2</sup>

There are few empirical efforts that provide an explicit analytical framework based on clearly defined behavioral assumptions (Badoe and Miller, 2000, Boarnet and Crane, 2001, Boarnet and Sarmiento, 1998, Boarnet and Crane, 2001, 2001, Boarnet and Sarmiento, 1996, Crane and Crepeau, 1998a, 1998b, Moshe and Bowman, 1998, Schimek, 1996, Voith, 1997, 1991). These analyses make use of more sophisticated statistical techniques that better account for the interrelationship between the built environment (land-use, land mix, or urban form) and travel behavior, by either employing multiple regression or more advanced discrete choice models. The most relevant examples to the objectives of this research effort are summarized next.

Using the 1990 Nationwide Personal Travel Survey (NPTS), Schimek (1996) applied a multiple regression model that accounted for simultaneity between a household’s pick of neighborhood density and the amount of travel. The model is specified as

---

<sup>1</sup> Gentrification is a phenomenon where old, deteriorated neighborhoods go through a process of renovation leading to land-value appreciation. This is brought about by population cohorts sorting themselves out in the residential clusters.

<sup>2</sup> Endogeneity refers to the situation when one of the explanatory variables appearing in a regression model is affected by, rather than affects, the dependent variable. This typically occurs because the relationship between dependent and explanatory variables is not accounted for explicitly by the model and, by default, falls into the model’s error term (its stochastic component). This results in serious estimation problems affecting both the sign and magnitude of the estimated parameter. The term *omitted variable bias* is generally employed to describe this problem.

$$V = f(\beta X) \quad (2.1)$$

$$D = f(V, \beta X) \quad (2.2)$$

where

$V$  = number of vehicles per household

$D$  = vehicle use (VMT or trips) per household

$X$  = vector of demographic and geographic characteristics (including density)

$\beta$  = column vector of parameters to be estimated

Schimek substitutes (2.1) into (2.2) to obtain a reduced form equation of (2.2) that he estimates by linear regression. To account for endogeneity between urban form and vehicle usage he assumes that density (and other factors) affects vehicle ownership levels and, in turn, vehicle ownership affects residential location. Endogeneity is controlled by using instrumental variable (IV) regression with the following instruments: race (white and Hispanic), location of household within the New York City standard metropolitan statistical area (SMSA), a dichotomous variable indicating if a household is located within an SMSA of three million or more; and a dichotomous variable indicating if a household is located within an SMSA of one million or more. Schimek justifies the choice of race by arguing that race and density are linked by spatial and housing market discrimination. He acknowledges that these variables might violate the basic IV assumption of no correlation with the exogenous variables of the reduced form equation. He does not run any test for exogeneity or over identification restrictions. The results are indeed impaired by the choice of weak instruments, as they are correlated with the other exogenous variables. The model's results show that a 10 percent increase in density leads to only a 0.7 percent reduction in household automobile travel. By comparison, a 10 percent increase in household income leads to a three percent increase in automobile travel. The results are similar when vehicle trips are used as the dependent variable.

A more robust analytical framework providing a priori specified behavioral relationships is found in Boarnet and Sarmiento (1998) with subsequent adaptations by Boarnet and Crane (2001), and Crane and Crepeau (1998b). Boarnet and Sarmiento address some of the shortcomings of previous work, namely the a priori specification of a behavioral model from which a series of hypotheses are tested. They specify a non-work trip demand function in reduced form, where trip frequencies and VMT are a function of a set of socio-economic, land-use factors, and costs of travel

$$N = f(p, y; S) \quad (2.3)$$

$$p = f(L) \quad (2.4)$$

where  $N$  = number of non-work auto trips,  $p$  = trip time cost,  $y$  = income,  $S$  = vector of socio-demographic variables, and  $L$  = vector of land-use characteristics.

The authors argue that the cost of travel is itself affected by land-use and that land-use is endogenous, as individuals tend to cluster in residential areas based on idiosyncratic preferences for residential location. This assertion is formalized by adding two equations that relate land-use  $L$  to residential location

$$L = f(ResLoc) \quad (2.5)$$

$$ResLoc_i = f(C_i, A_i) \quad (2.6)$$

where  $ResLoc_i$  indicates individual residential location;  $C_i$  = individual socio-demographic characteristics (essentially the same as  $S$ ); and  $A_i$  = characteristics of residential locations, such as amenities. In particular,  $A_i$  is a vector of IV's for  $L$  in a two-stage least squares (2SLS) regression of Equation (2.3).

The qualitative effect of each of the independent variables is expected to be indeterminate. A set of neighborhood amenity variables is used as IV's: the proportion of block group or census tract population that is black, the proportion Hispanic, the proportion of housing stock built before 1940, and the proportion built between 1940–1960. The choice of these IV's is justified by the authors based on existing evidence that neighborhood demographic composition and age of housing are determinants of residential location choice. The authors argue that these IVs play the role of good instruments, since they are correlated with land-use but not correlated with transport (VMT or trips) and are, thus, exogenous to the error term. Boarnet and Sarmiento conclude that there is limited evidence of the role of land-use on transportation behavior; the most important result is that land-use is endogenous to transportation behavior.

Several issues, related to the IV's being employed, impair the validity of these results. In this kind of analysis, good IV's must be correlated with land-use, but they must not be correlated with transportation. It is easy to show that race or minority status affects both location and transportation demand (Arnott, 1998). Race is correlated with location, since minorities' choice set might be more constrained than the choice set of whites (also with higher income). The same argument applies to their transportation choice set as, for example, when race and income are determinants of auto ownership and, therefore, impact both trips and VMT (the dependent variable employed by the authors). It can be concluded that the IV's chosen are poor, even if they pass a test for exogeneity based on over-identification, as outlined in Wooldridge (2002), and conducted by the authors.

Crane and Crepeau (1998b) introduce a set of trip demand functions (here the demand for auto trips is reported) as a function of travel time and income with the following Cobb-Douglas specification

$$a(p_a, y, \tau) = \frac{\alpha y \tau}{p_a} \quad (2.7)$$

where  $\alpha$  represents a taste parameter;  $y$  represents income;  $\tau$  represents land-use features, which serve as proxies for the cost of travel (time and distance); and  $p_a$  indicates the price of a trip. Travel time is equal to the ratio of trip length to travel speed (which are themselves choice variables). The analysis is conducted at a disaggregated level with respect to travel choice and land-use. Land-use data from the Census Bureau are merged with travel diary data using Geographic Information System (GIS) techniques to match residential location with land-use data at the tract level. GIS visual inspection of the network within 0.5 miles of the household allows measuring the characteristics of street grids and the presence of cul-de-sacs (measures of design). Land-use characteristics enter the demand function as a shift parameter. The empirical analysis looks at the impact of cross-sectional changes of  $\tau$ . The problem with this approach is that travel distance and speed are both affected by land-use and urban design, but the functions specified by the authors dismiss endogeneity between land-use and travel demand. For example, the vector of time prices is a function of speed and trip length, but trip length is also a function of location and street design. The authors acknowledge the problem and run a 2SLS regression using instruments for the price variables, although without "satisfaction with the variables available in the data (p. 233)."

Overall, the above models have been applied in the literature with minor variations to assess the impact of land-use on non-work walking trips (Greenwald and Boarnet, 2001), and to study the relationship between land-use and travel behavior in Santiago, Chile (Zegras, 2004).

A different model is developed by Voith (1991) in an analysis of transit ridership response to fair levels, where he models transit demand and transit supply in a context where changes in transit service affect residential location. In this model, the author assumes that changes in service affect location decisions around transit stations, which in turn affect transit demand and, recursively, transit supply. In an update of his earlier work, though, Voith (1997) concludes that, after controlling for prices and service attributes, demographic effects on transit are just ancillary.

Other research that attempts to model transit demand and transit supply simultaneously in an attempt to provide a better account of influencing factors is affected by methodological faults. For example, although Taylor and Miller (2003) recognize the need to model demand and supply jointly to avoid misspecification issues, they provide a poorly specified model that makes an improper use of simultaneous equation system modeling.

### **Inherent complexity: accessibility, urban design, and self-selection**

In addition to the widely investigated roles of employment and residential density, there are other dimensions that affect travel behavior. In recent years, urban policy responses intended to reduce externalities associated with employment and residential decentralization have been increasingly directed to influence the choice and amount of auto travel by manipulating urban form. The rationale behind these policies is that car travel reduction can be achieved by reducing trip frequencies and travel distances. By mixing residential with employment locations, thus expanding the choice set by clustering amenities, it is hypothesized that average travel distances tend to diminish, as nearby destinations will be preferred to more distant ones. Furthermore, offering increased public transportation choices is supposed to further reduce auto travel. These policies form what is currently defined as transit oriented development (TOD) (Cervero, et al., 2004). A relevant issue that is at the heart of the evaluation of TOD effectiveness and that has attracted the attention of transportation researchers is that of individual self-selection to residential location. In other words, individual residential idiosyncratic preferences, if not explicitly accounted for during empirical research, might lead to overestimation of the impact of TOD policies on travel behavior. Researchers have looked into aspects of the built-environment, such as the relationship between mixed land-uses (where residential and commercial land-uses are in close proximity) and accessibility measures to residential locations.

There is a vast and fast growing literature addressing the question of if and how urban form affects travel behavior and on the structural formation of such linkages. Since 2006, an increasing number of papers have been published that deals with the relationship between the built environment and travel behavior. A special issue of *Transportation* has been dedicated to this topic, which culled some of the best papers presented at the 2007 Annual Meeting of the Transportation Research Board (TRB). In the guest editor introduction, all relevant issues related to the inter-relationship between the built environment and travel behavior are outlined (Guo and Chen, 2007). Within this field of research, a topic that has been increasingly studied and debated is that of residential sorting or self-selection. This refers to the phenomenon that leads individuals or households to prefer a certain residential location due to idiosyncratic preferences for travel. In applied work, if residential self-sorting is not accounted for, findings tend to overstate the relevance of policies geared at impacting travel behavior through planned influence on the built environment. This topic of research has also been recently reviewed with respect to the shortcoming of previous work by providing new directions for research.

Mokhtarian and Cao (2008) provide a comprehensive review of empirical work spanning different analytical frameworks and econometric methods best suited to study residential self-selection. While this growing body of literature increasingly recognizes that unobserved



idiosyncratic preferences for travel affect residential location decisions, the debate hinges on the best way to treat the most common consequence of not controlling for this problem, i.e., the resulting omitted variable bias. The empirical treatment of the omitted variable bias describing self-selection spans from the use of nested logit regression by Cervero (2007), to more sophisticated error correlation models of Bhat and Guo (2004), and Pinjari et al. (2007).

Cervero (2007) estimates the degree to which residential self-selection affects transit mode choice by using conditional probability estimates that control for idiosyncratic preferences for location. He specifies a decision nest requiring the parameterization of two indirect utility functions, one function expressing residential location choice (i.e., reside within a mile of a rail stop), and one function expressing transit mode choice. Among the factors affecting location choice are workplace proximity to a rail station, job-accessibility, and household and personal attributes. The mode-choice indirect utility function is specified to include a travel time ratio (transit vs. auto), vehicle stock, personal attributes, and neighborhood density. Two issues related to the modeling technique and the choice of the observational unit cast doubts about the possibility of generalizing these findings. First the residential location utility function, although it explicitly controls for accessibility and socio-demographics, it does not include any controls for neighborhood and housing characteristics, such as poverty levels, ethnic cohorts, median housing age, size, price, nearby the residential unit to account for preferences for location. Second, it is not clear if the observational unit of analysis is the household or the individual (the subscript  $n$  in equation 1 on page 2,077 refers to the individual, but page 2,078 refers to a household). The implications of modeling household versus individual residential choice are non-trivial. For example, in a two member household, even after controlling for household characteristics, the first person might have a transit stop nearby his/her work location, while the second person does not. This results in a different travel time ratio (a control in the lower level mode-choice utility function). When estimating the nested logit regression, the predicted probabilities of residential location might differ, assigning the first person to the predicted choice of “near transit station” and the second person to the predicted choice of “far.” This results in having two individuals within the same household living in different locations.

Following the latest applications of discrete mode choice modeling developed by Bhat and Guo (2004), Pinjari et al. (2007) propose a model of joint determination of residential location and mode choice where both choices influence each other by accounting for observed and unobserved individual taste heterogeneity. Findings suggest that, after accounting for self-selection, the built environment has an impact on commute mode-choice behavior. Two indirect utility functions, one describing mode choice and one defining residential location are presented. The two functions are related by way of error-term specification. Self-selection endogeneity is captured by controlling for both observed and unobserved factors impacting

residential location and commuting mode-choice. First the mode-choice indirect utility function (*indirect* here means that the function depicts a realized choice that reflects the primitive objective function; it is not necessarily the indirect utility function of economic theory) includes a term indicating observed socio-demographic factors influencing the mode-choice decision. Then, an unobserved term is added to capture taste heterogeneity linked to the location decision but affecting mode-choice. This takes the form of an error term that is correlated to the second indirect utility function related to location choice. A final independent and identically distributed (*iid*) error term is added to the equation. The second indirect utility equation works the same way, with an error term correlated with the mode-choice utility function. The main issue with this methodology is related to the claim of simultaneously determining the choice of mode and location. This approach prompts the question “is the mode-choice decision really simultaneously determined with the location decision?” The authors seem at first to state this hypothesis, then, later, to refute it (p. 564) by admitting that, “The model structure assumes a causal influence of the residential location choice (and hence the built environment) on commute mode choice.” This apparent contradiction is probably justified by the specific econometric approach that they take. Specifically, they assume that individuals simultaneously maximize two different, although interdependent, utility functions, subject to somewhat different constraints. As in the case of Cervero (2007), this problem is the result of ad-hoc specifications of indirect utility functions without a formal understanding of the primitive objective functions, as discussed by Jara-Diaz and Martinez (1999).

Another problem in the study of self-selection arises when residential choice is modeled as a discrete variable. The treatment of the location decision as a dichotomous variable inherently presents a problem that is at the very heart of residential self-selection research. When using discrete choice modeling, one must assume that all individuals can choose among all possible locations within an urban area. The treatment of mode-choice and residential location in more sophisticated frameworks does not eliminate the need to ad-hoc determine the residential choice set. For example, both Pinjari et al. (2007) and Bhat and Guo (2004), who adopt the more sophisticated multinomial logit-ordered structure that explicitly consider the correlation of unobserved factors simultaneously affecting both choices, must a priori determine the location choice set (in that case, any individual is assumed to be able to choose among 223 different locations). This assumption dismisses the fact that, due to income and vehicle availability, some individuals have a more contrived mode choice and residential location sets, with the undesirable effects described by spatial mismatch theory (Kain, 1968). This results in not being able to fully discern the influence of idiosyncratic preferences for location on residential choice from issues related to spatial mismatch.

An alternative multivariate statistical method employed by the literature includes instrumental variable regression, with leading examples discussed earlier (Crane, Boarnet and

Sarmiento), that uses a set of properly tailored instruments. Other researchers advocate the use of simultaneous equation modeling (SEM), where additional equations are added to account for simultaneity between urban form, attitudes toward travel, and other factors. The preference for the latter approach is justified based on its inherent capability to potentially uncover any causality between travel and urban form, granted its proper use.

In many instances, research efforts that claim to uncover causality between urban design, travel behavior, and individual self-selection, do not make appropriate use of the econometric techniques therein employed. Data constraints also affect the usefulness of this statistical technique. For example, while Bagley and Mokhtarian (2002), Handy et al. (2005), and Cao et al. (2007), discuss the advantages of SEM, assuming the availability longitudinal data, they all make use of the same cross sectional dataset which employs a mix of secondary data and primary data from a travel attitude survey (the authors define this dataset as *quasi-longitudinal*). Furthermore, in the context of simultaneous equation modeling or instrumental variable regression, the validity of results hinges on the determination of the exclusion restrictions. That is, the researcher must a priori determine what explanatory variables are to be included and excluded from a given equation. The determination of the exclusion restrictions determines a model that is correctly specified in the sense that the matrix of the reduced form parameters to be estimated is unique in its representation of the more primitive structural matrix. Exclusion restrictions need to be drawn outside of the variables a researcher has available from a given dataset, i.e., they should be based on sound behavioral theory (Wooldridge, 2002).

In all studies of residential self-selection employing SEM techniques heretofore reviewed, including the work of Bagley and Mokhtarian (2002), Handy et al. (2005), and Cao et al. (2007, 2006), there is no explicit treatment of the exclusion restrictions that can be traced back to a formalized theoretical framework.

An alternative approach is presented by Vance and Hedel (2007), who employ a two-part model consisting of a Probit and OLS estimation, using the German Mobility Panel survey (MOP 2006). In the first part of the model, a Probit model that controls for socio-demographic (income, age, driving license) and urban form (commercial density, street density, commercial diversity) factors estimates the probability of owning a vehicle. The second stage, a regular OLS model, conditional on the first-stage predicted vehicle ownership, regresses vehicle use (distance traveled). The model is further enhanced by instrumenting the urban form variable using the set of instruments suggested by Boarnet and Sarmiento (1998). Although instrument validity is checked against exogeneity by applying selected diagnostics tests, the choice of instruments is limited to a set of controls for housing characteristics without the inclusion of

neighborhood characteristics controls to capture a broader set of factors affecting residential location choice.

Overall, much of the empirical work on the efficacy of such policies provides mixed evidence. This is so because these research efforts are based on ad-hoc empirical specifications and lack a formal behavioral framework that considers travel as a result of a decision-making process where activities are planned and executed through space and time.

### Measuring accessibility

Accessibility measures are widely used in transportation planning to relate the pattern of land-use and the nature of the transportation system. Various measures have been employed when analyzing the efficacy of mixed land-use or transit-oriented policies. A problem related to the use of accessibility is that its measurability is inherent to its definition and quantification. One definition is provided by “the ease and convenience of access to spatially distributed opportunities with a choice of travel” (DOE, 1996). Obviously, the main difficulty relates to quantifying the ease of accessibility. As recently summarized by Dong et al.(2006), there are essentially three measures of accessibility that have been employed to date: isochrones, gravity-based measures, and utility-based measures. The most widely employed are the gravity-based measures, which have the following generic form

$$Acc_i = \sum_j a_j f(c_{ij}) \quad (2.8)$$

where  $Acc_i$  means accessibility in zone  $i$ ;  $j$  indexes the available destination zones that can be reached from zone  $i$ ;  $a_j$  measures the activity opportunities in zone  $j$ ; and  $f(c_{ij})$  represents an impedance, or decay, function of traveling from zone  $i$  to zone  $j$ . This trip-based measure has been used in the recent work of Maat and Timmermans (2006), one of the few studies looking at the influence of land-use on activity-based travel. As pointed out by Dong et al.(2006), this measure is limited in that it neglects heterogeneity of preferences across individuals to the point that “a gravity measure of this type says that a retired grandfather and his college student grandson who live together have identical values of accessibility.” Furthermore, this measure is highly sensitive to the functional specification of the decay function. All of the models showing a relationship between increased transit usage and improvement in accessibility rely on one type of the above-mentioned measures. Analyzing the complexity of accessibility and travel behavior requires the use of accessibility measures that are strictly linked to the way activities are organized. These measures should be selected based on the relationship with the observed activity pattern. Some attempts are now appearing in the literature, although not directly related to the field of transportation research, that take into account individual heterogeneity and preferences. For example, utility-based measures of accessibility, which are based on the random utility theory as originally exposed by Domencich and McFadden (1973), provide ways

of relating accessibility measures to the characteristics of the alternative and the characteristics of the individual. The activity-based accessibility measure introduced by Dong et al. (2006), for example, is a utility-based measure. This measure is capable of capturing taste heterogeneity across individuals, combining different types of trips into a unified measure of accessibility, and of quantifying differing accessibility impacts on diverse segments of the population.

## Urban Form Measures and Polycentric Cities

As discussed in the introduction to this chapter, another problem of empirical analyses of the relationship between travel and land-use is the adoption of measures of urban form that are monocentric. Monocentric models only consider measures of the strength of the relationship between central business district (CBD) employment (and other activities located at the CBD) and travel behavior. For example, in their seminal work, Pushkarev and Zupan consider the relationship between transit service implementation and density in a context where the CBD is the main determinant of transit trips. More recently, Bento et al. (2005) examine the effects of population centrality, jobs-housing balance, city shape, road density and public transit supply on the commute-mode choices and annual vehicle-miles of travel of households living in 114 urban areas in 1990. They found that the probability of driving to work is lower the higher the population centrality and rail miles supplied and the lower the road density. Road density, in this model, is defined as miles of road multiplied by road width (for different categories of road) and divided by the size of the urbanized area.

In recent decades the process of decentralization has taken a more polycentric form, with a number of clustered employment centers affecting both employment and population distributions. The majority of these centers is subsidiary to an older CBD. Such centers are usually called subcenters or sub-regional centers. A more formal definition of subcenters is a set of contiguous tracts with significantly higher employment densities than surrounding areas (McMillen, 2001). The transportation literature has seldom looked at the influence of subcenters on travel behavior. An exception is Cervero and Wu (1998), who have examined the influence of subcenters in the San Francisco Bay area on commute distances. They conclude that employment decentralization has led to increased vehicle travel. These studies generally consider subcenters as exogenously determined either by assumption or by an empirical determination that makes use of specific density thresholds.

More recent developments in travel demand behavior and geographical science provide some insight on how to better capture the relationship between urban form and travel in a highly decentralized context. For example, Modarres (2003) proposes the use of GIS to determine subcenters using spatial clustering techniques to cluster patterns of major employers. He then considers the relevance of transit accessibility within the identified subcenters (accessibility is defined as the level of service provided by existing routes in each

census tract) to conclude that spatial accessibility is high within these subcenters. Casello (2007) analyzes the potential to and the impacts of increasing transit modal split in the polycentric metropolitan area of Philadelphia. By identifying “activity centers,” i.e., areas where transit use is likely, he models transit competitiveness and system performance. Kuby et al. (2004) update and improve previous research and find that the same factors affecting CBD boardings also influence non-CBD (subcenter based) transit ridership.

The decreasing relevance of CBD with respect to transit patronage is reduced to statistical insignificance in the recent work of Brown and Neog (2007), and Thompson and Brown (2006). In particular, Brown and Neog analyze aggregate transit ridership (not agency-specific) in 82 U.S. metropolitan statistical areas (MSA) using data from the National Transit Database as provided by the Florida Department of Transportation Transit Information System (FTIS). The authors use employment in the CBD and total metropolitan employment as urban form explanatory variables to regress a series of multivariate models. They found that transit ridership is not affected by the strength of a CBD, suggesting that improvement in ridership can be obtained by better serving decentralized urban areas.

These findings are reiterated by Brown and Thompson (2008), who employed a time series analysis of aggregate ridership data of the Metropolitan Atlanta Rapid Transit Authority (MARTA) in Atlanta, Georgia. The authors define two employment decentralization measures: number of employees within the MARTA service area located outside the Atlanta CBD (variable EMPMARTA) and the ratio of employment outside the MARTA service area to employment inside the MARTA service area (variable `RATIO_EMP`). They specify a first difference autoregressive model with annual linked passenger trips per capita as the dependent variable as a function of transit supply measures and the above-mentioned decentralization variables.<sup>3</sup> Results show that model performance is affected depending on the inclusion of a time trend variable, as reflected by the standard error estimates of the variable `RATIO-EMP` across the two models’ specifications. Notwithstanding these econometric issues, the authors conclude that there exists a positive association between decentralized employment growth (served by transit) and transit patronage. Although these conclusions favor policies geared at servicing employment rather than population concentrations, a generalization of these findings to other spatial context is not warranted.

Indeed the lack of relevance of the Atlanta CBD is due to the peculiar spatial characteristics that make it unique with respect to the rest of the U.S., and the world, as argued by Bertaud (2003). By comparing Atlanta spatial arrangement of population and employment and comparing it to other U.S. and world cities, Bertaud shows that the uniqueness of Atlanta

---

<sup>3</sup> In time series analysis the term first difference defines a variable transformation procedure used to reduce or eliminate time dependence of a given series (e.g., to transform a series from non-stationary to stationary).

(being highly polycentric) makes it cost unfeasible to implement a supply-side policy as the one prescribed by Thompson and Brown. In particular, Bertaud shows that with only 2 percent of the total jobs located at the CBD and only 8 percent within 5 km of the city center, Atlanta's dispersion of employment would require the addition of about 3,400 kilometers of metro tracks and about 2,800 new metro stations to provide the same transit accessibility to a comparable, though monocentric-based, city, requiring only 99 kilometers of tracks and 136 stations. These findings are used to justify the implementation of congestion tolling to control the negative externalities usually associated with sprawl, and the implementation of small, niche-type, transit services.

Overall the literature shows contrasting results when considering the relevance of a CBD in shaping the demand and supply of transit services. It is seen how the strength of a CBD is conditional to the spatial characteristics of the neighboring suburban areas. A more comprehensive evaluation of the relevance of CBD and subcenters should make use of spatial measures that take into account of the dispersion of activities and their influence on travel at a micro scale.

### **From Trip Generation to Activity-travel Behavior**

The literature heretofore reviewed empirically frames the relationship between travel behavior and urban form within modeling frameworks that rarely account for the fact that the demand of travel is an indirect demand spanning from the necessity to engage in activities requiring travel. The complexity of travel patterns and the recognition that trips are the result of a decision making process where activities are organized and prioritized through space and time have led to what is generally considered a paradigm shift in the study of urban travel behavior (Pass, 1985). This realization of more complex behavioral frameworks has paved the way for a new field of research, defined as activity-based modeling. Activity-based modeling is characterized by the recognition that travel is a derived demand, with a focus on constrained patterns or sequences of behavior instead of discrete trips, and the interdependence of decisions usually made within a household context (Jones, et al., 1990). Activity-based approaches are currently used to describe the activities individuals pursue, at what locations, at what times, and how these activities are scheduled within a transportation network characterized by opportunity and constraints (Bhat and Koppelman, 1999). This approach, while undergoing a continuous evolution and increased acceptance by practitioners, still lacks a comprehensive and strong behavioral framework (Davidson, et al., 2007). This is so because of the strong empirical basis upon which the approach developed, drawing mainly on advanced discrete-choice econometric modeling techniques.

Notwithstanding these shortcomings, this framework is better suited than those previously used to analyze the impact of land-use on travel patterns, as it fully acknowledges the presence



of trip-chaining behavior. Essentially, a trip chain may be defined as a sequence of trips that starts from home and/or ends at home. Different taxonomies defining trip-chaining complexity are possible depending on the purpose or mode of the trip for different classes of travelers. Sometimes called stop-making behavior, trip-chaining behavior in activity-based modeling describes the importance of multi-purpose trip-making rather than single-purpose trip-making. Numerous studies have examined trip-chaining or stop-making models using the frequency of stops on the way home and/or on the way to work as dependent variables (Bhat, 1999, Chu, 2003, Concas and Winters, 2007, Shiftman, 1998). In these studies, stop-making behavior describes stopping behavior made by a traveler, in particular a commuter, on the way to home or work. Under the assumption that a commuter follows a regular route, then stopping at a location other than home or work during the commute is treated as a deviation from the commute trip. In prior research, stop-making models were usually applied to trips linking non-work activities with work activities, including the morning commute, midday trips, evening commute, and trips before or after the commute.

The analysis of travel behavior within this context allows the recognition that trips are interrelated as opposed to the current transportation planning modeling assumptions of separation and independence of trips by purpose. Models based on microeconomic theory that explicitly treat the trade-offs involved in the choice of multiple-stop chains (i.e., the linking of several out-of-home activities and related trips into one tour) first appeared in the 1970's (Adler and Ben-Akiva, 1979). In addition to work trips, non-workers' trip-chaining as a series of out-of-home activity episodes (or stops) of different types interspersed with periods of in-home stays have also been investigated (Misra and Bhat, 2002, Misra, et al., 2003). Although travel demand forecasting models are now starting to incorporate trip-chaining behavior explicitly, only a limited number of studies exists that link the different aspects of trip-chaining behavior (trip tour frequency, complexity, duration) and urban form. There are some studies that relate trip chaining to regional accessibility or that compare trip chaining behavior across regional subareas, for example, city versus suburbs, as summarized by Ewing and Cervero (2001). A recent effort that looks at the influence of land-use on trip-chaining behavior (by way of analyzing tour complexity) is represented by Maat and Timmermans (2006).

There are some examples of location models that try to integrate the activity-based perspective into the process, by using developing discrete choice models of household residential location and travel schedules (Ben-Akiva and Bowman, 1998).

To date, no empirical work has been done that explicitly relates location to trip-chaining behavior in a context where individuals jointly decide location, the optimal trip chain, and area of non-work activities, based on the optimal trade-off between commute time and non-work travel activities. Better insight on the relationship between urban form and travel behavior



would be gained by testing the hypothesis that an individual's residential location is based on utility maximizing behavior.

## **Summary and Implications for Integrated Models of Transportation and Land-Use**

The bulk of research reviewed in this chapter is empirically oriented and based on the application of multivariate techniques that regress various measures of travel behavior (commute length, vehicle miles of travel, mode choice) on measures of residential and employment density, while controlling for the demographic characteristics of travelers. These studies examine the statistical significance, sign, and magnitude of the estimated coefficient on residential population density or employment density. A statistically significant negative coefficient leads one to conclude that a negative relationship exists between travel and density. For example, higher density leads to shorter commutes, fewer vehicle miles of travel (VMT) or a shift from auto transportation to alternative modes, such as transit. The abundance of these types of studies has led to the conclusion that policy interventions directed to influence density are capable of reducing automobile use [6-8].

The literature review uncovered the following issues that, to date, have been addressed but not completely resolved. In particular, it is widely recognized that there is a lack of a behavioral framework that can be applied to empirical work and is conducive to generalization of findings and applicability. Studies that relate density (population and employment) measures to travel behavior are monocentric and, therefore, fail to account for the employment and residential decentralization now characterizing the urban landscape. In most of this work, density is treated as exogenous and is not assumed to be impacted by transportation system changes. These studies have undergone systematic criticism due to their ad-hoc specifications and because of omitted variable bias problems due to the possibility that the relationship between urban form and travel might entail simultaneity and endogeneity. In addition, most of the work that looks at the joint estimation of transit demand, transit supply, and factors affecting both supply and demand are affected by methodological faults, ranging from misuse of simultaneous equation modeling (SEM) methods to improper functional specifications. These examples are noted in the complete literature review with detailed descriptions of how and where SEM models were misused, with explicit examples.

More recent developments in travel demand behavior and geographical science provide some insight on how better to capture the relationship between urban form and travel in a highly de-centralized context. The significance of CBD in determining transit ridership levels has been revisited and more relevance is now attributed to decentralized employment by looking at the influence of subcenters in an increasingly polycentric urban landscape.

While early work sought to provide a generalized analytical framework that made use of aggregate data, the more recent literature consists of papers that model the simultaneous decision of location and travel (as an application of improved discrete-choice modeling techniques) in a context where individuals choose locations based on specific travel preferences (for example, a preference about a specific mode) at the disaggregate level. Location decisions based on idiosyncratic preferences for travel define the term “residential self-selection behavior,” to indicate how individuals with similar tastes and preferences tend to cluster together in given locations, influences location decisions.

Finally, there is a lack of empirical work that looks at the relationship between urban form and travel behavior within an activity-based framework, which takes into account the complexity of travel (i.e., that accounts for trip chaining). Those studies that have employed activity-based modeling have failed to properly account for endogeneity and have disregarded spatial mismatch effects. In examining the relationship between urban form and travel, it is crucial to distinguish the effects of land-use from the effects of systematic socio-demographic differences of the individuals.

# Chapter 3      Methodology

---

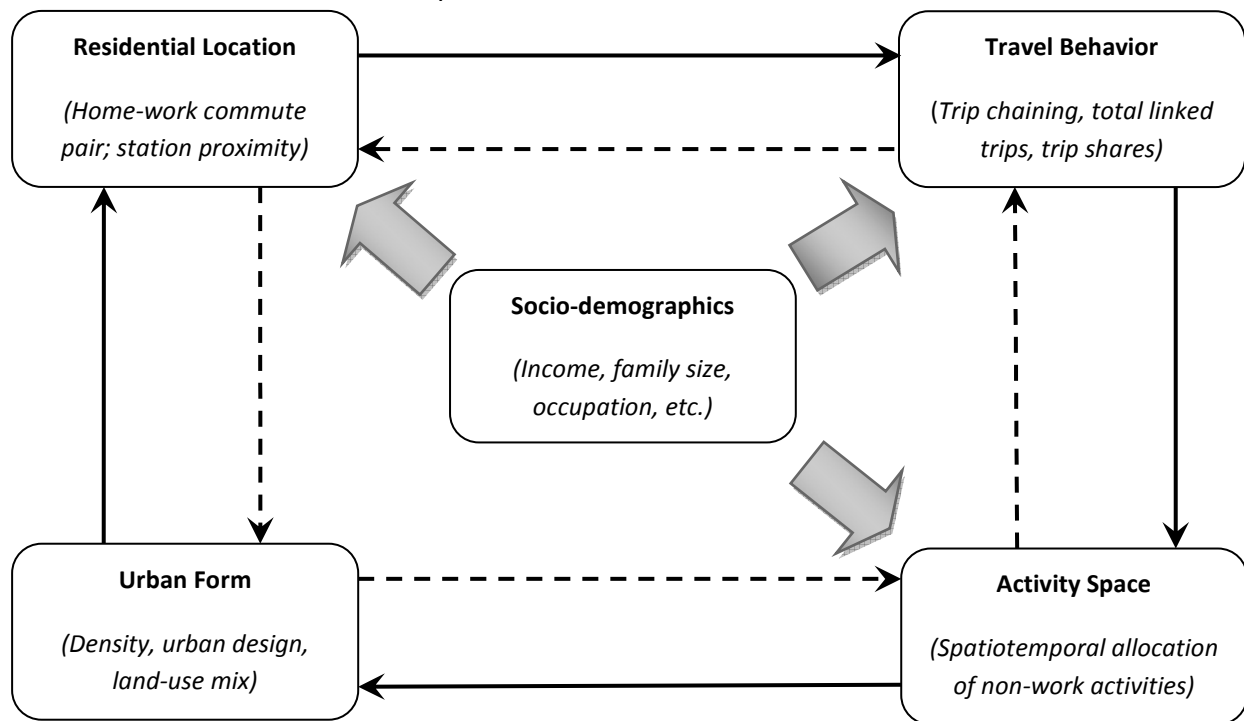
## Introduction

The main objective of this chapter is to develop an integrated approach to examine the relationship between transit and urban form. This task is intended to be accomplished in line with the recommendations provided by Report 16, which stated that “the interactive nature of transit and urban form, while complex, can potentially be conveyed through a balance of modeling work and carefully constructed empirical investigations that look at the joint influence of transit on residential location and ridership.”

Following the methodological issues previously highlighted by the literature review, the proposed analytical framework seeks to address unresolved issues as follows:

- It controls for individual idiosyncratic preferences for residential location
- It shifts the focus from monocentric-based measures of urban form to polycentric ones
- It utilizes a framework that better accounts for the spatial influence on travel patterns, by shifting the focus from a single-purpose trip-generation analysis to one that accounts for scheduling and trip chaining effects
- It accounts for the trade-off between commute time and non-work activities

Figure 3.1 presents a conceptual representation of the relationship between urban form, residential location, and the demand for travel. Given the objective of this study, the focus is on the demand for transit services. The framework, though, is suited to explain the determinants of the demand for travel in general.

**FIGURE 3.1** Conceptual Model of Urban Form and Travel Behavior

In this model, residential location, travel behavior, the activity space and urban form are all endogenously determined. In a departure from the monocentric model, the definition of residential location is taken from the polycentric model of Anas and his associates (Anas and Kim, 1996, Anas and Xu, 1999). Residential location is defined as the optimal *job-residence pair* in an urban area in which jobs and residences are dispersed. Following urban residential location theory, the location decision is assumed to be the result of a trade-off between housing expenditures and transportation costs, given income and the mode-choice set. Following Anas, the location decision is also based on idiosyncratic preferences for location and travel. In addition to determining optimal residential location, this approach also determines the optimal sequence of non-work trip chains, goods consumption, and transit patronage.

In this model, travel demand is considered an *indirect* demand brought about by the necessity to engage in out-of-home activities whose geographical extent is affected by urban form. Furthermore, budget constrained utility maximizing behavior leads to an optimization of the spatiotemporal allocation of these activities and an optimal number of chained trips. Socio-demographic factors directly influence residential location, consumption, and travel behavior. To date, no empirical work has been done that explicitly relates location to trip chaining behavior in a context where individuals jointly decide location, the optimal trip chain, and the

area of non-work activities, based on the optimal trade-off between commute time, leisure, and non-work travel activities.

It is within this framework that questions related to the interrelation between urban form, residential location, and the ensuing transit demand are addressed. How do location decisions affect travel behavior? How does urban form relate to travel behavior? Do residential location and urban form affect travel behavior? What is the impact of higher density on travel behavior? To address these questions, a basic travel demand model treating residential location and density as exogenous is first introduced. Subsequent extensions that relax these assumptions are then introduced and expected behavioral conclusions are reached. In its final specification the model also considers transit station proximity to a residential unit as endogenous. The specific treatment of station proximity takes into account supply-side measures that allow defining the model as a general equilibrium model of transit demand and transit supply.

### Model I: Exogenous Residential Location and Population Density

In this specification, residential location, transit station proximity, and density are exogenous. Given these variables, the model jointly determines the activity space and the optimal trip chain. The optimization (not explicitly modeled here) determines a travel demand function, given consumption and location decisions. The household (rather than the individual) is the unit of analysis because these decisions take place at the household level. Empirical studies on the relevance of transit station proximity to transit patronage show a strong relationship between transit use and station proximity (Cervero, 2007, Cervero and Wu, 1998). Therefore, this model includes this possibility. To include these considerations, Model I takes the following specific form, where exogeneity is indicated by a bar on the variables

$$TC = TC(AS, \overline{RL}, WD, X_{TC}) \quad (3.1)$$

$$AS = AS(TC, \overline{D}, X_{AS}) \quad (3.2)$$

$$TD = TD(TC, AS, \overline{RL}, WD, X_{TD}) \quad (3.3)$$

where

$TC$  = the demand for non-work (measured as the number of trips per commute-chain)

$AS$  = the activity space (measured as the geographic area surrounding the residence in which non-work trips are made)

$TD$  = the demand for transit trips (measured as the number of transit trips)

$\overline{RL}$  = residential location (measured as the *job-residence pair* distance)

$\overline{D}$  = a vector of residential and employment density controls

$WD$  = transit station proximity (measured as walking distance to the nearest transit station)

$X_{TC}$  = a vector of controls specific to the TC function;

$X_{AS}$  = a vector controls specific to the AS function

$X_{TD}$  = a vector of controls specific to the TD function

This model permits testing the hypothesis that individuals living farther from the workplace engage in more complex tours characterized by a higher number of non-work trips linked to the commute tour. In addition, trip chaining, as it relates to transit patronage, is directly affected by transit station proximity and by other factors summarized by the vector of controls,  $X_{TC}$ . This vector, as explained in more detail in Chapter 4, includes vehicle availability and the presence of young children among other factors likely to affect trip-chaining formation.

Trip-chaining behavior defines an activity space,  $AS$ , which is assumed to represent the optimized spatiotemporal allocation of non-work activities as affected by the built environment, summarized by the exogenous vector,  $\bar{D}$ . For example, more densely populated urban areas have more densely clustered activity locations, which shrink the size of the activity space relative to less densely populated areas. A more contrived activity space reduces trip chaining,  $TC$ , ultimately affecting the demand for travel,  $TD$ .

The model may be used to test the effect of urban design policies directly affecting travel distances and the land-use mix. In general, it may be used to test if higher density environments entail shorter travel distances, which in turn should affect the composition and complexity of trip chains and the overall amount of travel.

### **Residential location, $RL$ , and transit station proximity, $WD$**

The definitions of residential location and transit station proximity used here differ from those used in the current literature. For example, in studies of residential self-selection, the location decision is often presented as a dichotomous choice, i.e., whether to live near or far away from a transit station. Proximity is measured by a circular buffer around a station, often with a half-mile radius. The extent of this buffer is usually justified on empirical grounds. Cervero (2007), for example, used a half-mile radius in estimating a nested logit model of the joint determination of mode and location. This measure of transit proximity fails to account for barriers that prevent access to a station that lies within the half-mile radius. Some researchers have considered residential location as a choice to reside within a geographical unit, such as a traffic assignment zone (Bhat and Guo, 2004, Pinjari, Pendyala, Bhat and Waddel, 2007).

The use of transit proximity as a proxy for residential location, while dictated by the need to sort out the influence of the built environment from self-selection, is not based on any other theoretical underpinnings about the decision-making process that is at the heart of urban residential location theory. That is, it does not take into consideration the trade-off between

housing and transportation costs that, at the margin, determine where an individual decides to locate. For example, standard theory of location shows that individuals choose an optimal distance between work and home location given housing and transportation costs. In a monocentric model that only looks at travel between home and the central business district (CBD), individuals locate at a distance where the marginal cost of transportation is equal to the marginal housing cost savings obtained by a move farther out from the CBD (Moses, 1958, Muth, 1969). Recent departures from this view consider that individuals can locate anywhere in an urban area, choosing an optimal home-work distance that optimizes also the amount of non-work travel and non-work activities (Anas and Kim, 1996, Anas and Xu, 1999). Further explorations also consider the role of trip chaining behavior (Anas, 2007).

### **Activity space: spatial dispersion of non-work activities**

As argued in the introduction, the literature on self-selection rarely accounts for the fact that the demand of travel is an indirect demand derived from the necessity to engage in activities requiring travel (i.e., as fitting within activity-based theory).

The concept of activity space, although not new to behavioral sciences, is novel in terms of its application to travel behavior. The relationship between urban form and geographical patterns of activities is being studied only recently, due to the availability of specialized travel diary data and increasingly sophisticated geospatial tools. A growing field of research that looks at the relationship between urban form and the spatiotemporal allocation of activities and travel provides additional insight on the impact of the built environment. Recent research describing travel behavior and the influence of urban morphology and entire patterns of daily household activities and travel demonstrates how households residing in decentralized, lower density, urban areas tend to have a more dispersed activity-travel pattern than their urban counterpart (Buliung and Kanaroglou, 2007, 2006, Maoh and Kanaroglou, 2007).

This study explicitly accounts for the influence of the built environment in affecting the spatial dispersion of activities and how spatial dispersion affects the demand for travel and location decisions. This effect is accounted for by introducing the variable *activity space*, *AS*, into the model. The extent of the activity space is assumed to be affected by the built environment. Densely populated urban areas tend to cluster activity locations together thus shrinking the size of the activity space. This affects the spatial allocation of activities, thus affecting the demand for travel. As seen in the next chapter, there exist several ways empirically to measure the spatial dispersion of activities.

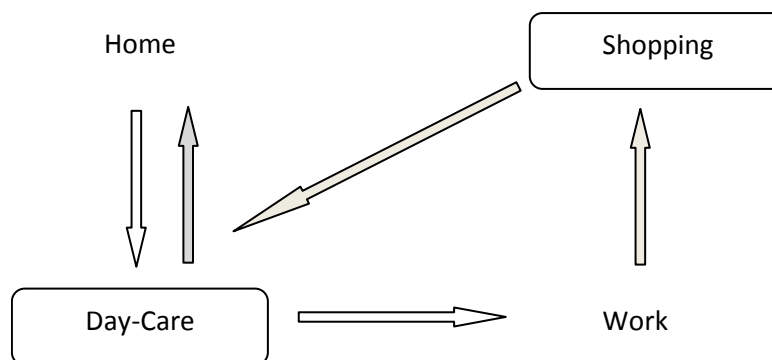
### **Trip chaining, *TC***

According to activity-based modeling practice, trip chaining describes how travelers link trips between locations around an activity pattern. In this context, a trip from home to work with an

intermediate stop to drop children off at day care is an example of a trip chain. In the literature there is not a formal definition of *trip chain*, and different terms and expectations exist as to what kind of trips should be considered as part of a chain (McGuckin and Murakami, 1999). Sometimes, the term *trip chain* is used interchangeably with the term *tour* to indicate a series of trips that start and end at home. For example, the following configuration can be considered:

- Five separate trips
- Two trip chains, one from home to work and one from work to home (this study's definition)
- One home-based trip tour

**FIGURE 3.2** Trip Chaining Sequence



In this study, it is hypothesized that trip chaining occurring on the home-job commuting pair permits saving time. This time savings in turn, can be either allocated to additional non-work travel, thus increasing the overall demand for travel (e.g., total number of trips), or be used to determine a longer commute (i.e., a farther apart home-job commuting pair). This hypothesis has recently been theoretically derived (Anas, 2007). The hypothesis of a positive relationship between more complex trip chains and home-work commute is also confirmed by empirical work. For example, in an analysis of trip chaining involving home-to-work and work-to-home trips using data from the 1995 nationwide personal transportation survey (NPTS), McGuckin and Murakami (1999) found that people are more likely to stop on their way home from work, rather than on their way to work. About 33 percent of women linked trips on their way to work compared with 19 percent of men, while 61 percent of women and 46 percent of men linked trips on their way home from work. Using the 1991 Boston Household Travel Survey, Bhat (1997) found that about 38 percent of individuals made stops during the commute trip. Davidson (1991) found similar results from her analysis of commute behavior in a



suburban setting, showing that travelers rely heavily on trip chaining in an urban context characterized by higher spatial dispersion of non-work activities. Other studies also provide empirical evidence of increased stop-making during the commute periods (Bhat, 2001) or how the ability to link trips is enhanced by the flexibility inherent in automobile use (Strathman, 1995).

### **Travel demand, $TD$**

Travel demand is herein treated as an indirect demand brought about by the need to purchase goods and services. Travel demand,  $TD$ , measures the number of work and non-work transit trips at the household level. The decision process behind the choice of the number of trips, as formalized by the above rational choice framework, considers trip generation as a function of trip chaining and exogenous residential location and socio-demographic factors. The constrained maximization problem of the joint determination of activity space and trip chaining defines an optimal vector of non-work trips, given residential location and urban form characteristics (e.g., residential and employment density levels, land-use mix). This treatment of travel demand as derived from the desire to engage in out-of-home activities departs in terms of behavioral sophistication from the treatment of trip generation as developed by Boarnet and Crane (2001) in their analysis of travel demand and urban design. In Boarnet and Crane (2001) trip demand functions are either directly affected by land use or indirectly (by influencing the cost of travel).

In contrast, in this model land use (i.e., urban form) directly affects the spatial allocation of activities. It is the budget-constrained utility maximization behavior that defines optimal travel patterns. The complexity of this mechanism is better shown in the ensuing comparative static analysis, which allows ascertaining the effect that urban form exerts on the demand for travel.

### **Comparative-Static Analysis**

The basic theoretical implications of Model I can be explored in advance empirical testing by employing comparative static analysis<sup>4</sup>. This section considers the impact of changes in exogenous density,  $\bar{D}$ , and exogenous residential location,  $\bar{RL}$ , on travel demand,  $TD$ . Basically, starting from an equilibrium state, the impacts of an increase in density and residential location on the initial equilibrium are determined. The objective is to see what happens to transit demand as density levels change. The most relevant results of the comparative static analysis

---

<sup>4</sup> Comparative static analysis is a tool commonly used in mathematical economics and microeconomic theory. It allows comparing different equilibrium states associated with different sets of values of parameters and exogenous variables. Comparative statics can be either qualitative or quantitative in nature. In this case, it allows conducting a qualitative assessment as it permits to focus on the direction of change, rather than its magnitude, of changes in location and density. (Chiang, Alpha C. 1984. *Fundamental Methods of Mathematical Economics*. New York: McGraw-Hill, Inc. )

are summarized below, while the derivation of the comparative statics is detailed in Appendix B.

### Effects of an increase in density, $\bar{D}$

The effect of an increase in density on travel demand is obtained as

$$\frac{dTD}{d\bar{D}} = \frac{\overbrace{TD_{AS}AS_D}^{\alpha \begin{smallmatrix} (-) & (-) \end{smallmatrix}} + \overbrace{TD_{TC}TC_{AS}AS_D}^{\beta \begin{smallmatrix} (+) & (+) & (-) \end{smallmatrix}}}{\underbrace{1 - AS_{TC}TC_{AS}}_{\begin{smallmatrix} (+) & (+) \\ (+) \end{smallmatrix}}} \geq 0 \quad (3.4)$$

Where

$AS_{TC} = \frac{\partial AS}{\partial TC}$  is the partial derivative of activity space with respect to trip chaining

$TC_{AS} = \frac{\partial TC}{\partial AS}$  is the partial derivative of trip chaining with respect to activity space

$TD_{AS} = \frac{\partial TD}{\partial AS}$  is the partial derivative of transit demand with respect to activity space

$TD_{TC} = \frac{\partial TD}{\partial TC}$  is the partial derivative of transit demand with respect to trip chaining

$\alpha$  = change in transit demand caused by a contraction in activity space as a result of increased density

$\beta$  = change in transit demand caused by decreasing trip chaining as a result of increased density

Based on assumed positive relationship between spatial dispersion of activities and trip chaining (see Appendix B for assumptions), the result of this analysis shows that changes in density levels exert two contrasting effects on the demand for transit trips. The result shows an ambiguous effect of an increase in density on transit demand (as measured in total linked trips per household). Indeed, for  $\frac{dTD}{d\bar{D}} > 0$  it must be that  $\alpha > -\beta$ . In other words for transit demand to be positively related to density it must be that the increase in transit demand caused by a contraction in activity space (as a result of increased density,  $\alpha > 0$ ) is greater than the reduction in transit demand caused by reduced trip chaining (as a result of increased density,  $\beta < 0$ ).

This explanation is inherent to the determinants of trip chaining behavior. In higher density environments, as the spatial extent of non-work activities reduces, trip chaining needs decrease, but individual trips increase and individuals prefer to make non-chained trips. First, increased density reduces the extent of the activity space, which directly increases the demand

for non-chained transit trips. Second, higher densities reduce the activity space, which reduces the need to chain trips (as time saving opportunities decrease) and thus the demand for transit trips. Therefore, an increase in transit trips occurs if transit demand is more sensitive to changes affecting the spatial allocation of non-work activities than to changes affecting trip chaining behavior. In other words, the above comparative static result shows that the increase in density exerts two opposite effects on transit demand.

### Change in residential location, $RL$

Next the comparative statics of an increase in residential location,  $RL$ , are derived. Note that  $RL$  is considered as predetermined in Model I. This model is suited to either describe a situation where residential location is considered as predetermined, such as a short run time frame or can be used to cross compare decision making among households at any point in time. The question to be answered is: “What happens to transit demand as the job-residence pair changes?” Using cross sectional data, this question can be translated as: “What happens to transit demand for those households facing long commutes?”

The comparative statics result describing the impact of a change in residential location on the demand for transit trips is

$$\frac{dT D}{dRL} = \frac{\overbrace{TD_{RL}}^{(\pm)} + \overbrace{TC_{RL}}^{(+)} \overbrace{TD_{TC}}^{(+)} + \overbrace{AS_{TC}}^{(+)} \left( \overbrace{TC_{RL}}^{(+)} \overbrace{TD_{AS}}^{(-)} - \overbrace{TC_{AS}}^{(+)} \overbrace{TD_{RL}}^{(\pm)} \right)}{1 - \underbrace{AS_{TC} TC_{AS}}_{\substack{(+)(+) \\ (+)}}} \geq 0 \quad (3.5)$$

where

$TD_{RL} = \frac{\partial TD}{\partial RL}$  is the partial derivative of transit demand with respect to residential location

$TC_{RL} = \frac{\partial TC}{\partial RL}$  is the partial derivative of trip chaining with respect to residential location

As previously discussed, an increase in residential location increases trip chaining ( $TC_{RL} > 0$ ), which in turn positively affects both the size of the activity space,  $AS$ , and the demand for transit services. The overall effect on transit demand hinges on the sign of  $TD_{RL}$ . To the extent that an urban area is well served by transit, then the relationship between transit demand and residential location is positive. A positive relationship is observed in older, more monocentric-type cities, with existing transit services supporting major work commute travel routes. On the other hand, if supply constraints exist, transit demand declines as the job-residence distance increases. Therefore, the overall effect on transit demand due to a change in location depends on both the sign and magnitude of  $TD_{RL}$  ( $TD_{RL} \leq 0$ ).

### Change in walking distance to nearest station, WD

A change in transit station proximity causes a change in transit demand equivalent to

$$\frac{dT D}{d W D} = \frac{\overbrace{T D_{W D}}^{(-)} + \overbrace{T D_{T C}}^{(+)} \overbrace{T C_{W D}}^{(-)} + \overbrace{A S_{T C}}^{(+)} \left( \overbrace{T C_{W D}}^{(-)} \overbrace{T D_{A S}}^{(-)} - \overbrace{T C_{A S}}^{(+)} \overbrace{T D_{W D}}^{(-)} \right)}{\underbrace{1 - A S_{T C} T C_{A S}}_{\substack{(+)(+) \\ (+)}}} < 0 \quad (3.6)$$

where

$T D_{W D} = \frac{\partial T D}{\partial W D}$  is the partial derivative of transit demand with respect to walking distance

$T C_{W D} = \frac{\partial T C}{\partial W D}$  is the partial derivative of trip chaining with respect to walking distance

An increase in distance to the nearest station directly affects transit demand and the ability to engage in trip chaining using transit. The overall effect is a decline in transit usage due to reduced accessibility negatively affecting trip chaining and transit demand, net of indirect effects.

Other changes in exogenous factors directly affecting only the demand for transit (e.g., transit station proximity at workplace) can be also be obtained. These can be summarized as

$$\frac{dT D}{d \bar{x}_i} = \frac{\partial T D}{\partial \bar{x}_i} \quad (3.7)$$

Where  $\frac{\partial T D}{\partial \bar{x}_i}$  is the direct effect of a change in exogenous variable  $x_i$  appearing only in the travel demand function. In Chapter 4, for example, the analysis looks at the impact of transit oriented development (TOD) stations on transit ridership, and the presence of a transit station at workplace.

### Model II: Endogenous Residential Location, Exogenous Density

In this model, the assumption of residential location is relaxed. Treated as a choice variable, residential location is the outcome of a trade-off between transportation and land use costs. Taking into account idiosyncratic preferences for location, households choose an optimal home-work commute pair, while at the same time optimizing goods consumption and the ensuing non-work travel behavior (optimal non-work trip chaining and activity space). This model is specified as

$$T C = T C(A S, R L, X_{T C}) \quad (3.8)$$

$$A S = A S(T C, \bar{D}, X_{A S}) \quad (3.9)$$

(3.10)

(3.11)

## Comparative-Static Analysis

The complete comparative statics are presented in Appendix B. A discussion of the findings is presented below. Note that the inclusion of the endogenous residential location equation,  $RL$ , complicates the computation of the total partial derivatives.

### Effects of an increase in density, $\bar{D}$

The effect of an increase in density on travel demand is obtained as

$$\frac{dT_D}{d\bar{D}} = \frac{\overbrace{AS_D^{(-)} \left[ \left( 1 - \overbrace{RL_{TC}^{(+)} \overbrace{TC_{RL}^{(+)}}} \right) \overbrace{TD_{AS}^{(-)} + TC_{AS}^{(+)}} \left( \overbrace{RL_{TC}^{(+)} \overbrace{TD_{RL}^{(\pm)}}} + \overbrace{TD_{TC}^{(+)}} \right) \right]}^{[I]}}}{\underbrace{1 - \overbrace{RL_{TC}^{(+)} \overbrace{TC_{RL}^{(+)}}} - \overbrace{RL_{TD}^{(\pm)} \overbrace{TD_{RL}^{(\pm)}}} + \overbrace{AS_{TC}^{(+)} \left[ -\overbrace{RL_{TD}^{(\pm)} \overbrace{TC_{RL}^{(\pm)}}} \overbrace{TD_{AS}^{(-)} + TC_{AS}^{(+)}} \left( -1 + \overbrace{RL_{TD}^{(\pm)} \overbrace{TD_{RL}^{(\pm)}}} \right) - \overbrace{RL_{TD}^{(\pm)} \overbrace{TC_{RL}^{(\pm)}}} \overbrace{TD_{TC}^{(+)}} \right]}_{[I]}}}_{\leq 0} \quad (3.12)$$

In the long run, the spatial extent of non-work activities, trip chaining and residential location are all jointly determined. Exogenous shifts in density levels affect this decision making process. An increase in density directly impacts the spatial extent of non-work activity locations in terms of an increased activity space,  $AS$  (i.e., activities are more disperse across the urban landscape). This increase affects trip chaining with feedback effects both on the demand for transit trips and residential location patterns in a looping fashion. Comparing equation (3.12) to equation (3.4), the complexity of the relationship between residential location and travel arrangements as a result of exogenous changes in density,  $D$ , increases substantially. As in equation (3.4) the magnitude of the changes depends on the trade-off between transit demand and changes in the spatial extent of non-work activities ( $AS_D$ ) and trip chaining behavior. Ultimately, the sign of equation (3.12) depends on the relationship between transit demand and residential location ( $TD_{RL} \leq 0$ ).

### Change in walking distance to nearest station, WD

A change in transit station proximity causing a change in transit demand is derived as

$$\frac{dT_D}{dWD} = \frac{\overbrace{TD_{WD}}^{(-)} + \overbrace{TD_{TC}TC_{WD}}^{(+)(-)} + \overbrace{RL_{TC}}^{(+)} \left( \overbrace{TC_{WD}TD_{RL}}^{(\pm)(-)} - \overbrace{TC_{RL}TD_{WD}}^{(+)(-)} \right) + \overbrace{AS_{TC}}^{(+)} \left( \overbrace{TC_{WD}TD_{AS}}^{(+)(-)} - \overbrace{TC_{AS}TD_{WD}}^{(+)(-)} \right)}{|J|} \quad (3.12)$$

With endogenous residential location, the sign, as well as the magnitude of  $\frac{dT_D}{dWD}$  depends on both the sign and magnitude of  $TD_{RL}$  ( $TD_{RL} \leq 0$ ), as well as the magnitude of  $TC_{WD}$ .

Other changes in exogenous factors directly affecting transit only the demand for transit can be obtained by applying equation (3.5).

### Model III: Endogenous Residential Location, Endogenous Density

In this last extension to Model I, the assumption of exogenous density is relaxed. This model translates the conceptual framework of Figure 3.1 into the following analytical model

$$TC = TC(AS, RL, X_{TC}) \quad (3.13)$$

$$AS = AS(TC, D, X_{AS}) \quad (3.12)$$

$$TD = TD(TC, AS, RL, WD, X_{TD}) \quad (3.13)$$

$$D = D(RL, AS, X_D) \quad (3.14)$$

$$RL = RL(TC, X_{RL}) \quad (3.15)$$

In the long run, the simultaneous choice of location and travel decisions are assumed to affect density levels across a given urban area. This model best describes long term equilibrium, both location and travel decisions optimized under constraint. Urban form is treated as endogenous to the process and is itself affected by household travel decisions and location behavior. Aspects of this relationship and its influences on transit patronage have been previously considered in the literature. For example, while modeling long-run transit demand responses to fare changes, Voith (1997) treats density as endogenous and being affected directly by transit patronage levels. In the long run, these levels are affected by supply-side changes. Voith (1997) assumes that as transit services improve, more people tend to live in proximity to transit stations, thus increasing the demand for transit services. Empirically, the latter can be acknowledged by treating transit station proximity as endogenous to this process.

Ideally, empirical testing of this model would rely on disaggregate travel diary data in the form of a panel that collects behavior of a same set of individuals across time. When dealing with observational data across different individuals at a point in time (i.e., a cross-sectional dataset), changes in behavior can be studied by controlling for individual heterogeneity.

### Comparative-Static Analysis

Given the endogenous treatment of density, this model can be used to test the effects of policies geared at directly affecting density, such as policy interventions intended to increase density around transit stations. Assuming an exogenous shock,  $\theta$ , positively affecting density, comparative statics can be obtained. The inclusion of two more equations complicates the

calculations to derive the relevant comparative static results. The results are basically the same as Model II, although the expected magnitudes of impacts differ. To avoid cluttering, Appendix B reports the comparative statics calculations which will be used in the empirical work of Chapter 4. Table 3.1 reports a summary of the comparative statics highlighting the expected signs from changes in the most relevant variables affecting transit demand.

**TABLE 3.1** Comparative Static Results

<i>Exogenous Variable</i>	<i>D</i>	<i>RL</i>	<i>AS<sub>φ</sub><sup>*</sup></i>	<i>WD<sup>**</sup></i>
<i>Effect on TD</i>	<i>+/-</i>	<i>+/-</i>	<i>-</i>	<i>-</i>

*\*Shift parameters affecting AS*

*\*\*Transit station proximity (walking distance)*

## Conclusions

The analytical framework herein presented seeks to strike a balance between the complexity of activity-based modeling and the more traditional discrete-choice frameworks. The failure of the traditional four step travel demand model lies in not recognizing the relevance of non-work travel as a derived demand, which has nontrivial implications on the analysis of the relationship between travel behavior and land-use patterns. The added complexity of the models herein introduced is intrinsic to the explicit consideration of non-work travel behavior and its interrelationship with the spatial extent of non-work activities.

The analytical models developed in this chapter are general and can be applied to data from any urban area. As seen in the next chapter, empirical testing of the hypotheses these models carry requires more detailed travel behavior data at the individual levels. The increased level of sophistication of activity-based travel diaries allows collecting information on activities conducted at home and out of home, as well as their spatial location. As it will be seen, the contribution of geographic information system (GIS) modeling permits the measurement of the geographic dimension of both activities and travel and to relate them to the surrounding urban landscape. Coupling GIS with econometric modeling permits conducting empirical tests of the relationships explained by the models herein developed.

*[This page intentionally left blank.]*



# Chapter 4 Findings

---

## Introduction

In this chapter, all relevant hypotheses about the relationship between urban form and transit patronage introduced in Chapter 3 are subject to empirical testing. The objectives are:

1. to test if the signs summarized by Table 3.1 are confirmed by actual data;
2. to assess the presence of endogeneity in the relationship between transit and urban form; and,
3. to assess the magnitude of this relationship.

The aim is to ascertain to what extent density matters in shaping the demand for transit, after accounting for any endogeneity or simultaneity that might be present. To test these hypotheses, an empirical dataset is used that culls travel behavior information at the disaggregate level. The dataset is described by presenting a set of descriptive statistics of both dependent and independent variables. Next, an econometric framework that parameterizes the hypothetical relationships of Model I through Model III is specified and suitable multivariate regression methods are considered. The results of regression are presented and discussed in detail at the end of the chapter.

## Summary of Data

To test the models presented in the previous chapter, the analysis relies on the use of travel diary data. Travel diaries ask respondents to compile a log of activities and travel made during a selected time frame, usually one or two days encompassing both weekday and weekend travel. The new generation of activity-based travel surveys is characterized by travel diaries that provide a high level of detail of activities, both at home and out-of-home, to obtain a comprehensive picture of all behavioral aspects at the individual and household levels affecting travel decisions. Information on activities by purpose (work, recreation, shopping, etc.) is logged by respondents. The main advantage of these surveys, as highlighted by Davidson et al. (2007), is that they are based on tour structure of travel, with travel derived within a general framework of the daily activities undertaken by households and persons.

This study uses travel diary data from the 2000 Bay Area Travel Survey (BATS2000). BATS2000 is a large-scale regional household travel survey conducted in the nine-county San Francisco Bay Area of California by the Metropolitan Transportation Commission (MTC). Completed in the spring of 2001, BATS2000 provides consistent and rich information on travel

behavior of 15,064 households with 2,504 households that make regular use of transit<sup>5</sup>. BATS2000 used the latest applications of activity and time-based survey instruments to study travel behavior. The data from BATS2000 are available online and maintained as a set of relational data files and are available as comma-separated value (CSV) and American Standard Code for Information Interchange (ASCII) text files (MTC, 2008). Each data file has a corresponding statistical analysis system (SAS<sup>®</sup>) script to read the data file and act as the data dictionary for the data file (MTC, 2007). In the dataset, 99.9 percent of home addresses and 80 percent of out-of-home activities were geocoded using geographic information systems (GIS) to the street address or street intersection level (99.5 percent to the street address level). This permits a precise geographic determination of non-work activities, job, and residential unit locations.

The choice of this dataset goes beyond its quality. Currently, most of the relevant academic and practitioner work on the relationship between transit and urban form, research on the issue of residential self-selection, and the efficacy of transit oriented development policies (TOD) have made use of BATS2000 (Commission, 2008). For example Cervero (2007) used BATS2000 and census land use data to evaluate TOD impacts on ridership and self-selection. In his analysis, he notes that between 1998 and 2002 about 13,500 apartment and condominium units were built within a half-mile of urban stations of Southern California and the San Francisco Bay Area, often using land previously occupied by park-and-ride lots; this makes the dataset suitable to also test the impact of TOD on ridership.

The final dataset is a result of combining BATS2000 travel behavior data with geographical data from the Census Bureau. The latter data source is specifically the Summary File 3, which consists of detailed tables of social, economic and housing characteristics compiled from a sample of approximately 19 million housing units (about 1 in 6 households) that received the Census 2000 long-form questionnaire (Bureau, 2007). These data were obtained at the block group level.

The dataset unit of observation is the household to reflect the higher hierarchical decision making process of both residential location and travel needs. Thus, housing and neighborhood characteristics are measured at the block group level where the residential unit is located. Referring to MTC work on transit use and station proximity (MTC, 2006), a transit household is defined as one where one or more members used transit at least once during the two-day surveying period.

---

<sup>5</sup>MTC defines a transit household as one where one or more members used transit at least once during the two-day surveying period.

## Dependent Variables Descriptive Statistics

Before proceeding to present basic descriptive statistics on the dependent variables, issues related to their measurement are first discussed. While definitions of trip chaining, *TC*, and walking distance to the nearest station (i.e., station proximity), *WD*, have already been discussed in Chapter 3, some additional explanation on their measurement is warranted.

### Measures of activity space, AS

Activity space measures the spatial dispersion of non-work activity locations. Non-work activities comprise shopping, recreational (e.g. visiting friends or dining out) and non-recreational activities (doctor visits, child rearing, recurring activities). These activities can be located in proximity to the household residential unit or be located away from it. To measure their spatial extent across the urban landscape, area-based geometric measures developed in transportation geography are used. Different metrics that describe the spatial extent of activity locations can be employed. The simplest measure is represented by the standard distance circle (SDC) (otherwise defined as standard distance deviation in spatial statistics) and is essentially a bivariate extension of the standard deviation of a univariate distribution. It measures the standard distance deviation from a mean geographic center and is computed as

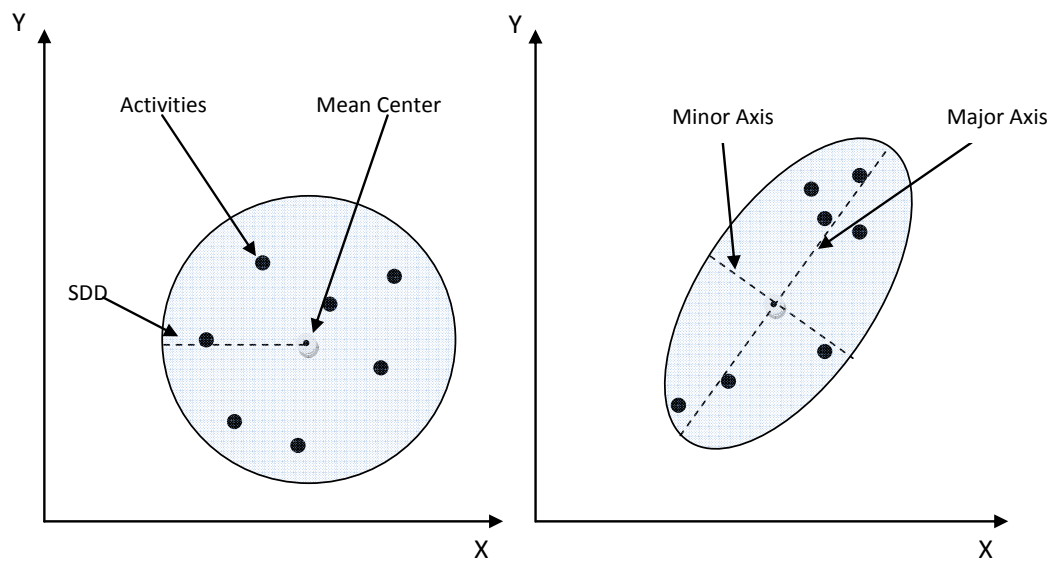
$$SD = \sqrt{\frac{\sum(x_i - \bar{x}) + \sum(y_i - \bar{y})}{n}} \equiv \sqrt{\frac{\sum d^2}{n}} \quad (4.1)$$

where  $\bar{x}$  and  $\bar{y}$  represent the spatial coordinates of the mean center of non-work activities at the household level, and the  $i$  subscript indicates the coordinates of each non-work activity. The mean activity center is analogous to the sample mean of a dataset and it represents the sample mean of the  $x$  and  $y$  coordinates of non-work activities contained in each household activity set. The coordinates represent longitude and latitude measurement of each activity and are reported in meters following the Universal Transverse Mercator (UTM) coordinate system. Household activity locations are those visited by surveyed household members during a specified time interval, in this case two representative weekdays. Thus, the standard distance of a household's activity pattern is estimated as the standard deviation (in meters or kilometers) of each activity location from the mean center of the complete daily activity pattern. Interpretation is relatively straightforward with a larger standard distance indicating greater spatial dispersion of activity locations. The area of the *SDC* is obtained as the area of a circle with a radius equal to the standard distance. It provides a summary dispersion measure that can be used to explore systematic variations of activities subject to socio-demographic, travel patterns, and patterns of land-use.

As pointed out by Ebdon (1977), this measure is affected by the presence of outliers or activities that are located furthest from the mean center. As a result of the squaring of all the

distances from the mean center, the extreme points have a disproportionate influence on the value of the standard distance. To eliminate dependency from spatial outliers, another measure of dispersion called the standard deviational ellipse (SDE) is usually employed, which uses an ellipse instead of a circle. The advantages with respect to the SDC have been discussed in the literature (Ebdon, 1977). In addition to control for outliers, it also allows accounting for directional bias of activities with respect to its mean center. The ellipse is centered on the mean center with the major axis in the direction of maximum activity dispersion and its minor axis in the direction of minimum dispersion (See Figure 4.1). This study employs the standard distance ellipse (SDE), using the formula described in Levine (2005).

**FIGURE 4.1** Standard Distance Circle and Standard Distance Ellipse



### Measures of residential location, RL

Residential location is defined as the average distance of household employment activities to the household residential unit

$$RL = \frac{\sum_{m=1}^k dist_{mj}}{k} \quad (4.2)$$

where  $dist_{mj}$  is the Euclidean distance to the residential unit located at  $j$ , from a household member work location  $m$ , and  $k$  is the total number of employed household members. An alternative specification only considers the distance between the householder work location and the residential unit. This assumes that the residential location choice puts more relevance to the location of the household “breadwinner.” This and other measures are discussed in detail in Chapter 5.

### Measures of transit station proximity, WD

This study treats transit proximity as a continuous variable measuring distance to the nearest transit station from the household residential unit. A 2006 publication from MTC made use of BATS2000 data to look at the relationship between transit use, population density, and characteristics of individuals living nearby transit stations (MTC, 2006). An appendix to this study was recently published on the MTC website which reports an updated version of the household file containing an additional variable measuring network walking distance from each household residential unit to the nearest transit station (Purvis, 2008). Walking distance is measured as actual distance based on network characteristics that take into consideration the existence of accessibility impediments.

### Measures of density, D

The dependent variable  $D$ , used in Model I through III is measured using gross population density at the Census block group level where the household residential unit is located. The Census block group area is measured in square miles. As seen in Chapter 3, other studies on transit and urban form tend to utilize number of dwelling units per square mile. Additional urban form measures, treated as exogenous to the model, are also considered and described under the exogenous variable section of this chapter.

Table 4.1 presents basic descriptive statistics of the dependent variables, split by different gross population density levels corresponding to the classification adopted by MTC to differentiate between urbanized and non-urbanized areas. As documented by other studies, there exists an underlying correlation between density levels and travel behavior. This table shows how the activity space is slightly larger for transit households (19.1 versus 17.2 square miles of non-transit households) and contracts as density increases, while trip chaining does not follow this linear relationship. Walking distance to the nearest station noticeably decreases at higher density levels. To highlight the relevance of transit patronage, sample transit trip averages are compared to auto, walk and other trips in Table 4.2. This table shows marked differences in terms of trip making and trip chaining behavior between transit and non-transit households, as well as in average travel times between home and work (51.9 minutes versus 37.4 minutes of non-transit households).

**TABLE 4.1** Descriptive Statistics: Overall Sample Means

Density (persons/mile <sup>2</sup> )	Household Activity Space (mile <sup>2</sup> )	Residential Location, RL (miles)	Residential Location, RL (min)	Trip Chaining, TC (number)	Transit Trips (number)	Auto Trips (number)	Walk Trips (number)	Walking Distance, WD (mile)
0 to 499	27.84	14.12	43.40	2.96	0.14	9.00	0.50	2.33
500 to 5,999	19.31	11.82	40.97	3.04	0.27	8.78	0.72	0.45
6,000 to 9,999	15.69	10.02	38.70	2.98	0.29	8.40	0.80	0.23
>=10,000	13.70	8.56	39.41	3.01	0.73	5.98	1.33	0.14

**TABLE 4.2** Descriptive Statistics: Sample Means of Dependent Variables and Selected Trip Measures

		Gross Population Density (persons/mile <sup>2</sup> )	Household Activity Space (mile <sup>2</sup> )	Residential Location, RL (miles)	Residential Location, RL (min)	Trip Chaining, TC (number)	Transit Trips (number)	Auto Trips (number)	Walk Trips (number)	Walking Distance, WD (mile)
Transit Household										
No	Mean	7,910.51	17.16	10.33	37.36	2.87	-	8.32	0.73	0.49
	SD	8,752.95	38.40	10.07	33.32	1.77	-	6.14	1.62	1.44
	N	12,260.00	10,548.00	9,128.00	8,353.00	11,242.00	12,260.00	12,260.00	12,260.00	12,260.00
Yes	Mean	15,172.65	19.14	11.58	51.92	3.65	2.32	5.96	1.70	0.22
	SD	17,193.12	37.84	9.76	35.35	1.73	1.29	5.77	2.38	0.38
	N	2,503.00	2,176.00	2,138.00	1,918.00	2,446.00	2,503.00	2,503.00	2,503.00	2,503.00
Overall Sample	Mean	9,141.78	17.50	10.57	40.08	3.01	0.39	7.92	0.89	0.45
	SD	11,006.88	38.31	10.03	34.18	1.79	1.02	6.14	1.81	1.33
	N	14,763.00	12,724.00	11,266.00	10,271.00	13,688.00	14,763.00	14,763.00	14,763.00	14,763.00

## Explanatory Variables Descriptive Statistics

### Socio-demographic variables

The following socio demographic variables are considered as potential exogenous explanatory variables:

- Household characteristics
- Householder gender
- Householder race
- Number of children of schooling age
- Number of persons employed full-time
- Household income
- Number of vehicles
- Number of licensed individuals
- Tenure (own versus rent)

This information is available from the BATS2000 person file. Some of these socio-demographic variables have been included in the studies reviewed in Chapter 3 dealing with the influence of land use on transit patronage, while the most current literature on self-selection considers all of them. Table 4.3 provides a summary of these variables for the overall sample. As with the vast majority of travel survey, the white population is overly represented, as well as the higher income cohort.

**TABLE 4.3** Summary of Selected Demographic Variables

<i>Variables</i>	<i>Frequency</i>	<i>% Share</i>
Householder Gender		
Male	6,901	45.8%
Female	8,163	54.2%
Householder Race		
Asian	1,223	8.1%
Black	442	2.9%
Hispanic	647	4.3%
Other	674	4.5%
White	12,078	80.2%
Children, by age group		
< 6 year	1,539	10.2%
6 to 11 year	1,973	13.1%
12 to 18 year	2,202	14.6%
Employed, Full Time (persons)		
0	876	7.0%
1	7,214	57.8%
2	4,063	32.5%
>=3	335	2.7%
Household Income (\$)		
Less than 10,000	225	1.7%
10,000 to 14,999	230	1.7%
15,000 to 19,999	322	2.4%
20,000 to 24,999	368	2.8%
25,000 to 29,999	464	3.5%
30,000 to 34,999	424	3.2%
35,000 to 39,999	514	3.9%
40,000 to 44,999	756	5.7%
45,000 to 49,999	833	6.3%
50,000 to 59,999	1,352	10.2%
60,000 to 74,999	1,660	12.6%
75,000 to 99,999	2,359	17.9%
100,000 to 124,999	1,620	12.3%
125,000 to 149,999	804	6.1%
>= 150,000	1,260	9.6%
Vehicles		
0	610	4.0%
1	4,938	32.8%
2	6,542	43.4%
3	2,238	14.9%
4	554	3.7%
>=5	182	1.2%
Tenure		
Own	10,415	69.4%
Rent	4,597	30.6%



## Travel behavior variables

Additional explanatory variables at the household level were created to control for factors affecting both the spatial extent of non-work activities and the ensuing travel behavior.

- Activity travel time
  - mean travel time to shopping trips starting at home
  - mean travel time to recreational trips starting at home
  - mean travel time to school trips starting at home
  - mean travel time to other trips not starting at home
  - mean travel time across all non-work activities
- Activity duration
  - mean time duration across all non-work activities

These variables are commonly used in the activity-based literature in modeling activity duration and scheduling (Bhat, 1999, 1997, 2001) and activity travel patterns (Kuppam and Pendyala, 2001). While transit households spend almost the same amount of time shopping as non-transit households (28.9.0 versus 30.3 minutes), they spend less time on recreational activities (161.9 versus 175.9 minutes) and less time at home (181.8 versus 210.1 minutes). The time spent travelling to reach out-of-home activities also differs, with transit households spending an average of 15.7 minutes on the road versus 12.9 minutes for non-transit households. The trade-off between leisure and work is also reflected in less time spent sleeping by transit households (243.6 versus 249.6 minutes for non-transit households). These time-use variations and the comparison between transit and non-transit households provided in Table 4.2 are indicative of the trade-offs inherent to total time available, residential location, and trip-chaining behavior discussed in Chapter 3.

## Urban form variables

Although BATS2000 does not include land-use variables, it provides exact geographical information about the location of each of the 15,064 households. GIS coordinates permit a precise allocation of each household residential unit within each Census Bureau geographical unit of reference by using GIS techniques. By linking each households' residential unit x and y geographic coordinates to GIS Census block group level maps of the San Francisco Bay area, a comprehensive set of land-use variables was merged with the travel diary dataset.<sup>6</sup> Other

---

<sup>6</sup> Detailed GIS maps and other geographical data are available online at the MTC website Metropolitan Planning Commission. *Gis Maps and Data*. Available online at accessed on Metropolitan Planning Commission.

variables related to non-residential land use were obtained from the 2000 U.S. Census Bureau County Business Patterns (CBP) data file. Table 4.4 describes these variables and data sources.

**TABLE 4.4** Urban Form Variables

<i>Variable</i>	<i>Definition</i>	<i>Source</i>
<i>Gross population density</i>	Number of persons/Census block group area size (mile)	U.S Census Bureau Summary File 3
<i>Dwelling units</i>	Number of owner occupied units	U.S Census Bureau Summary File 3
<i>Dwelling density</i>	Number of owner occupied units/ Census block group area size (mile)	U.S Census Bureau Summary File 3
<i>Number of retail establishments</i>	Total number retail establishments within zip code	U.S Census County Business Patterns: 2000
<i>Retail establishment density</i>	Total number retail establishments/zip code area	U.S Census County Business Patterns: 2000
<i>Number of wholesale establishments</i>	Total number retail establishments within a zip code	U.S Census County Business Patterns: 2000
<i>Wholesale establishment density</i>	Total number wholesale establishments/zip code area	U.S Census County Business Patterns: 2000
<i>Distance from CBD</i>	Distance from CBD	BATS2000
<i>Distance from subcenter</i>	Distance from the nearest subcenter	BATS2000

It is important to note that the last two variables are intended to be used as proxy measures of centrality (CBD distance) and polycentricity (distance from the nearest subcenter). As mentioned in Chapter 3, monocentric-based models only consider measures of the strength of the relationship between CBD employment (and other activities located at the CBD) and travel behavior. In recent decades the process of decentralization has taken a more polycentric form, with a number of clustered employment centers affecting both employment and population distributions. The majority of these centers is subsidiary to an older CBD. Such centers are usually called subcenters or sub-regional centers (a more formal definition of subcenter is a set of contiguous tracts with significantly higher employment densities than surrounding areas). The transportation literature has seldom looked at the influence of subcenters on travel behavior. For example, Cervero and Wu (1998) have examined the influence of subcenters in the San Francisco Bay Area on commute distances to conclude that employment

decentralization has lead to increased travel. These studies generally take subcenters as exogenously determined either by assumption or by an empirical determination that makes use of specific density thresholds. There are no established methods to determine the number of subcenters present in any urban area. Existing methods rely on rules of thumb based on knowledge about specific geographic areas (Giuliano and Small, 1991), while others account for an endogenous determination based on their impact on agglomeration and employment (McMillen, 2001).

To account for urban decentralization and its effect on transit use this study adopts the Census definition of cities and non-designated places to first identify subcenters and then produce a distance measure between a household residential unit and the nearest subcenter.

In addition to the above variables, a set of explanatory variables to control for household idiosyncratic preferences for location is obtained. The literature provides some insight on the choice of land-use variables as controls or instrumental variables (Boarnet and Crane, 2001, Boarnet and Sarmiento, 1998, Crane, 2000, Crane and Crepeau, 1998b). This study uses some variables employed in the literature and introduces new ones. The following variables have all been obtained at the block group level using the Summary 3 Census Bureau file:

1. Stock of housing built before 1945 (number of housing units)
2. Housing median value (dollars; occupied owners units)
3. Housing median age (years; non-rent units)
4. Housing size (median number of rooms; occupied owners units)
5. House median monthly cost (owner occupied)
6. Percent of household living below poverty line
7. Diversity index (0 = homogeneous; 1 = heterogeneous neighborhood)

The first variable has been used before as an instrumental variable in multivariate regression studies that considered travel behavior as endogenous to urban form (Boarnet and Crane, 2001, Boarnet and Sarmiento, 1998, Crane, 2000, Crane and Crepeau, 1998b), while the remaining ones are unique to this study. Additional controls for neighborhood characteristics have also been used elsewhere. For example, the proportion of block group or census tract population that is Black and the proportion Hispanic have been used as instruments by Boarnet and Sarmiento (1998) and the percent of foreigners by Vance and Hedel (2007).

In this study variables one through five are meant to be used to control for idiosyncratic preferences for housing characteristics not directly affecting travel behavior, but directly affecting the residential choice decision at the household level. Variables six and seven are intended as controls for neighborhood characteristics. In particular, the percent of households living below poverty levels (henceforth defined as poverty) serves as a proxy for crime, while the diversity index (henceforth called diversity) is used as a proxy for ethnic preferences (i.e.,

moving into a neighborhood with similar ethnic characteristics). The latter is an index of ethnic heterogeneity that varies from zero (only one race living in the neighborhood) to one (no race is prevalent), similar to the Shannon's diversity index (Begon and Towsend, 1996).<sup>7</sup> As discussed in further detail in Chapter 5, poverty and diversity serve a dual role as instrumental variables when transit station proximity is treated as endogenous to the model.

Table 4.5 presents relevant sample mean values, which are reported for households by mode choice. Transit households tend to live in highly populated areas characterized by higher than average poverty levels, as well as smaller and older housing units. One way analysis of variance tables (not herein reported) that included an interaction term between transit household and the transit station dummy variable were generated. All variables exhibited a significant difference in means, indicating that housing price, housing age, room size, neighborhood diversity and poverty levels differ across households according to their location and mode choice. To gain additional insight on the trade-off between residential location and preference for transit, Table 4.6 reports the same measures of Table 4.5, but differentiates between households living in proximity to a transit station. Proximity is measured by a Euclidean half-mile buffer around a transit rail line in existence when the BATS2000 travel survey was being conducted.

### **Transit supply variables**

To account for the relevance of transit supply in determining transit ridership, the following measures are considered as relevant:

- Presence of a transit stop at workplace
- Supply of park-and-ride within a half-mile of transit stop
- Presence of a transit-oriented development (TOD) stop within a half-mile of residential unit

---

<sup>7</sup> The Shannon Index is a measurement used to compare diversity between habitat samples. The comparison is made by taking into account the proportion of each individuals of a given species to the total number of individuals in the set.

**TABLE 4.5** Land Use Variables by Household Type

<i>Transit Household</i>	<i>Gross Population Density (persons/mile<sup>2</sup>)</i>	<i>Dwelling density (dwellings/mile<sup>2</sup>)</i>	<i>Retail Establishments Density (number/mile<sup>2</sup>)</i>	<i>Wholesale Establishment Density (number/mile<sup>2</sup>)</i>	<i>House Median Value (\$)</i>	<i>House Median Age (years)</i>	<i>Housing Stock (% built before 1949)</i>	<i>Housing Size (rooms)</i>	<i>Households Median Income</i>	<i>Households Below Poverty</i>	<i>Diversity Index</i>
No	7,910.5	3,312.8	18.4	6.9	399,818.7	34.2	19.9%	6.0	74,189.5	5.6%	0.57
Yes	15,172.7	7,198.0	43.1	12.6	399,374.0	41.8	36.0%	5.9	67,140.8	7.6%	0.62
Overall	9,144.4	3,974.3	22.5	7.9	399,591.1	35.5	22.6%	5.9	72,994.4	5.9%	0.58

**TABLE 4.6** Land Use Variables by Transit Station Proximity

<i>Within 1/2 mile of Transit Station</i>	<i>Gross Population Density</i>	<i>Dwelling density (dwellings/mile<sup>2</sup>)</i>	<i>Retail Establishments Density (number/mile<sup>2</sup>)</i>	<i>Wholesale Establishment Density (number/mile<sup>2</sup>)</i>	<i>House Median Value (\$)</i>	<i>House Median Age (years)</i>	<i>Housing Stock (% built before 1949)</i>	<i>Housing Size (rooms)</i>	<i>Households Median Income</i>	<i>Households Below Poverty</i>	<i>Diversity Index</i>
No	7,313.8	2,939.2	14.8	5.7	396,509.6	33.6	18.6%	6.0	75,050.4	5.4%	0.57
Yes	19,871.4	10,039.7	67.6	20.8	417,647.7	46.3	46.4%	5.2	60,501.5	8.9%	0.64
Overall	9,144.4	3,974.3	22.5	7.9	399,591.1	35.5	22.6%	5.9	72,994.4	5.9%	0.58

The relevance of transit station proximity to the workplace is confirmed by the literature, as seen in Chapter 3. For example, using BATS2000, Cervero (2007) showed that the presence of a station within one mile of a workplace (with good accessibility) strongly influences both residential choice decisions and transit use. The relationship gets stronger as distance to the station declines. The presence of park-and-ride lots nearby transit stops also positively influences transit ridership by improving accessibility to those households located farther than the one-mile threshold. Furthermore, as highlighted by TCRP Report 95 (2007), the presence of park-and-ride provides increased opportunities to trip chain from the residence to the transit station on the way to work (pp.17-85). The relevance of park-and-ride lots is measured by a dichotomous variable indicating the presence of a park-and-ride lot within a half-mile of a transit stop. To produce these transit-supply explanatory variables, the same GIS maps created by MTC as part of their transit station proximity study were used (MTC, 2008) (a detailed discussion of the GIS methodology is provided in Appendix G of the MTC study).

Finally, to test the relevance of urban design policies on transit patronage, a dichotomous variable qualifying a transit stop as having the characteristics of a TOD station is also introduced in the model. TOD stops are characterized by land development policies geared at facilitating transit use by improving transit station accessibility (by reducing physical barriers), and by promoting mixed land-use development (residential and commercial) in their immediate surroundings. The California Department of Transportation Transit-Oriented Database was used to identify these stations (CALTRANS, 2008).

Table 4.7 summarizes the full set of exogenous explanatory variables previously described. In addition, the set of endogenous variables, which are used in the ensuing empirical investigation, is also listed and highlighted at the bottom of the table.

**TABLE 4.7** List of Variables for Model Estimation

<i>Variable</i>	<i>Definition</i>	<i>Use</i>
<i>inc</i>	Household income	Socio-demographic
<i>sch</i>	Number of children of school age (pre-k to middle)	Socio-demographic
<i>veh</i>	Number of vehicles	Socio-demographic
<i>own</i>	Tenure (1 = owner; 0= renter)	Socio-demographic
<i>licensed</i>	Number of persons with driving license	Socio-demographic
<i>tswork</i>	Presence of a transit stop within ½ mile of workplace (1=yes, 0=otherwise)	Transit supply
<i>prkride</i>	Presence of a park-and-ride within ½ mile of a transit stop (1=yes, 0=otherwise)	Transit supply
<i>ts_tod</i>	Transit stop characterized as transit-oriented development stop (1=yes, 0=otherwise)	Transit supply
<i>cbd_dist</i>	Residential unit distance from CBD	Urban form/land use
<i>subc_dist</i>	Residential unit distance from nearest subcenter (cities and designated places)	Urban form/land use
<i>r_est</i>	Number of retail establishments, zip code level	Urban form/land use mix
<i>w_est</i>	Number of wholesale establishments, zip code level	Urban form/land use mix
<i>hprice</i>	Median house price, block group level	Residential/neighborhood characteristics
<i>hage</i>	Median house age, block group level	Residential/neighborhood characteristics
<i>room</i>	Median number of rooms owner occupied unit, block group level	Residential/neighborhood characteristics
<i>inc_blkgrp</i>	Median household income, block group level	Residential/neighborhood characteristics
<i>pov</i>	Proportion of households living below poverty line, block group level	Residential/neighborhood characteristics
<i>div</i>	Diversity index (from 0 if block group level is ethnically homogenous to 1 if heterogeneous)	Residential/neighborhood characteristics
<i>act_dur</i>	Mean non-work activity duration	Travel behavior
<i>act_tt</i>	Mean travel time to non-work activities	Travel behavior
<i>TC</i>	Trip chain; number of non-work trip stops on the job-residence commute	Trip chaining behavior
<i>AS</i>	Household activity space; standard distance ellipse area (mile <sup>2</sup> )	Spatial extent of non-work activities
<i>RL</i>	Residential location (home-work distance)	Household residential location
<i>WD</i>	Walking distance to from the residential unit to the nearest transit station	Transit station proximity
<i>D</i>	Gross population density (persons/mile <sup>2</sup> )	Urban Form

## Method of Analysis

A common problem with nontrivial consequences in terms of both consistency and bias occurs when trying to empirically test the models of Chapter 3 using linear regression (i.e., ordinary least square or OLS). When estimating an equation that contains one or more endogenous variables as explanatory variables, the error term correlation leads to biased and inconsistent estimates of the parameter of interest. Biasedness refers to the situation where an estimated parameter associated with a variable of interest is actually skewed with respect to its true, but unobserved, value. The result is to either give too much or too little significance to its actual role. Inconsistency is related to the statistical treatment of the estimated parameter (i.e., considered as a random variable). In other words, if one were to consistently gather data repeatedly, and increase sample size, and run the same regression, the biasedness would not tend to disappear. Another consequence of the improper use of regression techniques in the context of endogeneity is related to the confusion between causality and mere correlation. The statistical significance of a parameter of interest is too often interpreted as a direct causal relationship, while it may just indicate a partial linear correlation.

Given the structural framework of Chapter 3, the empirical test of the proposed hypotheses requires the use of an econometric method defined as structural equation modeling (SEM). SEM is used to capture the causal influences of the exogenous variables on the endogenous variables and the causal influences of the endogenous variables upon one another. The use of SEM in transportation research is linked to the development of activity-based modeling in travel behavior research, which explicitly points out the causal mechanisms underlying individuals' location and travel decisions. Furthermore, more recent developments in the literature looking at the efficacy of urban design policies dealing with residential sorting effects try to sort out causality links between urban form and travel behavior. To uncover causality in a context where travel behavior and urban form simultaneously affect each other, specific econometric techniques must be called upon. As the literature review of Chapter 3 highlighted, it is only recently that transportation researchers have recognized that causal relationships among travel behavior and urban form can be effectively represented in a structural equation framework (Cao, Mokhtarian and Handy, 2007, 2006, Guevara and Moshe, 2006, Mokhtarian and Cao, 2008, Peng, et al., 1997).

The structure of a simultaneous equation model is better depicted using a matrix notation

$$\underbrace{Y}_{(nxG)} \underbrace{B}_{(GxG)} + \underbrace{X}_{(nxK)} \underbrace{\Gamma}_{(KxG)} + \underbrace{\epsilon}_{(nxG)} = \underbrace{0}_{(nxG)} \quad (4.3)$$

where

$Y$  is a  $nxG$  row vector of endogenous variables



$B$  is a  $G \times G$  matrix of structural parameters associated with the right hand side endogenous variables

$X$  is a  $n \times K$  row vector of exogenous or predetermined variables

$\Gamma$  is a  $K \times K$  matrix of parameters associated with the exogenous or predetermined variables

$\varepsilon$  is a  $n \times G$  row vector associated with unobserved disturbances

$0$  is a  $n \times G$  row vector of zeros

In reality, the matrix of structural parameters of interest (of size  $G^2 + KG$ ) cannot be observed; rather, a researcher can only try to recover it by using as much information as possible from the observable vector of exogenous variables. Specific econometric techniques exist that estimate the matrix of parameters in a reduced form to recover the original structural parameters of interest. Available methods include maximum likelihood estimation (ML), generalized least squares (GLS), two-stage least squares (2SLS), three-stage least squares (3SLS), and asymptotically distribution-free estimation (ADF).

Before proceeding with the estimation, it is necessary to ensure that the model is identified. Identification is concerned with the ability to obtain unique estimates of the structural parameters. Identification refers to the condition under which a simultaneous equation system can be meaningfully estimated. It consists of a process leading to the specification of the entire system of equations, where each equation to be estimated contains a specific set of both endogenous and exogenous variables. A necessary and sufficient condition for identification of a structural equation is provided by the rank condition. In the context of simultaneous equation modeling, the validity of results hinges on the determination of the exclusion restrictions. That is, the researcher must a priori determine what explanatory variables are to be included and excluded from each equation. The determination of the exclusion restrictions defines a model that is correctly specified in the sense that the matrix of the reduced form parameters to be estimated is unique in its representation of the more primitive structural matrix. Exclusion restrictions need to be drawn outside of the variables a researcher has available from a given dataset (i.e, they should be based on sound behavioral theory). Each of the three models presented next is subject to the rank condition for identification prior to estimation and results are succinctly reported in Appendix C. The inclusion and exclusion of relevant explanatory variables from each endogenous equation are discussed.

## Model I results

The first model of Chapter 3 with exogenous residential location and density is specified as

$$TC = \alpha_0 + \alpha_1 AS + \alpha_2 RL + \alpha_3 \overline{WD} + \alpha_4 veh + \alpha_5 act\_tt + \alpha_6 act\_dur + \alpha_7 sch + \alpha_8 subc\_dist + \varepsilon_1 \quad (4.4)$$

$$AS = \beta_0 + \beta_1 TC + \beta_2 \overline{D} + \beta_3 act\_dur + \beta_4 inc + \beta_5 r\_estd + \varepsilon_2 \quad (4.5)$$

$$TD = \gamma_0 + \gamma_1 TC + \gamma_2 AS + \gamma_3 \overline{WD} + \gamma_4 \overline{RL} + \gamma_5 tswork + \gamma_6 prkride + \gamma_7 ts\_tod + \gamma_8 veh + \varepsilon_3 \quad (4.6)$$

Note that, as specified in Chapter 3, both residential location,  $RL$ , and density,  $D$ , are considered exogenous to the model. Equation (4.4) describes trip chaining behavior occurring on the commute trip to and from the work location. Trip chaining, jointly determined with activity space,  $AS$ , is affected by vehicle availability and transit station proximity, activity travel time and duration, and household structure. Vehicle ownership and transit proximity, together with household characteristics (income and children), affect the capability of engaging in complex tours. Note the exclusion restriction assumptions do not consider that either density levels or more retail opportunities directly affect trip chaining.

Equation (4.5) describes how the spatial extent of non-work activities responds to changes in urban form, being affected directly by density levels and retail establishment concentrations. Recalling that activity space is a result of a utility maximizing behavior determining travel and consumption levels, household income is expected to directly affect its determination. As income levels increase so does the need to consume more goods. Following basic microeconomic theory of individual consumption, it is assumed that individuals have preferences for heterogeneity in consumption (i.e. convexity of indifference curves indicating preference for balanced consumption bundles). As income increases, individuals prefer to visit different locations; a behavior that positively affects the size of the activity space.

Equation (4.6) describes the demand of transit trips as brought about by the necessity to engage in non-work travel (directly affected by  $AS$  and  $TC$ ) and by the relative distance of the residential unit to the work location,  $RL$ . Transit supply measures are expected to directly affect transit ridership in terms of transit station accessibility both at origin and destination. The relevance of TOD policies affecting ridership is to be tested empirically by the inclusion of the dichotomous variable  $ts\_tod$ .

All three equations pass the rank condition for identification. Equation (4.4) is overidentified, and equation (4.5) and (4.6) are classified as just identified. The results of a two-stage least square regression (2SLS) are displayed in Table 4.8.

All signs agree with the hypotheses of Chapter 3. In particular, the joint determination of trip chaining and the spatial extent of non-work activities relate to transit patronage as hypothesized. The presence of a transit stop at workplace ( $tswork$ ) positively affects transit

demand, as well as the presence of a TOD transit stop in proximity of the residence unit ( $ts\_tod$ ). As density increases, it reduces the size of activity space, which, in turn, positively affects the demand for transit. This assumption, as stated in Chapter 3, relates more compact urban environments to increased transit patronage. As locations where non-work activities are more clustered, the need to engage in long and complex journeys requiring modes other than transit decreases, resulting in increased transit usage. The converse is also true, suggesting that policy interventions related to directly affect the clustering of non-work activity locations, such as mixed-land use policies, are likely to significantly affect ridership levels. However, the relevance of this relationship is better appreciated in a context where residential location is also treated as a choice variable (i.e., endogenous).

To better appreciate the magnitude of these effects, Table 4.9 reports point elasticities of transit demand with respect to selected explanatory variables. Following Wooldridge (2002), pp.16-17), point elasticities are obtained as

$$\varepsilon_{y_i, x_i} = \frac{\partial E(y_i | x)}{\partial x_i} = \frac{\partial \mu(x)}{\partial x_i} \cdot \frac{x_i}{\mu(x)} \quad (4.7)$$

where  $\frac{\partial \mu(x)}{\partial x_i}$  is the total partial derivative measuring the change in equation  $y_i$  brought about by a change in an exogenous variable  $x_i$  holding everything else constant;  $x_i$  is the point of evaluation; in this case the sample averages of the explanatory variable, and;  $\mu(x)$  is the sample average value of the dependent variable  $y_i$ .

For example, to obtain the elasticity of travel demand with respect to changes in density, the following elasticity is applied

$$\varepsilon_{TD, x_i} = \frac{dTD}{dD} * \frac{D}{TD} \quad (4.8)$$

where  $\frac{dTD}{dD}$  is obtained by using equation (3.4) of Model I.<sup>8</sup>

---

<sup>8</sup> Note that due to a natural log transformation of  $AS$  and  $WD$ , the estimated parameters must first be brought back to their original specification (see for example, Wooldridge (2002, pp. 14-15) for the correct procedure).

**TABLE 4.8** 2SLS Regression Results—Model I

<i>Equation</i>	<i>Coefficient</i>	<i>Std. Error</i>	<i>P</i>
<b><i>Trip chaining, TC</i></b>			
<i>RL</i>	0.0096	0.0040	0.0160
<i>AS</i>	0.0648	0.1658	0.6960
<i>WD</i>	-0.0570	0.0137	0.0000
<i>veh</i>	-0.0793	0.0308	0.0100
<i>act_tt</i>	0.0014	0.0004	0.0010
<i>act_dur</i>	-0.0022	0.0003	0.0000
<i>subc_dist</i>	0.0439	0.0068	0.0000
<i>sch</i>	0.0778	0.0144	0.0000
<i>constant</i>	1.2771	0.2611	0.0000
<b><i>Activity space, AS</i></b>			
<i>TC</i>	0.5863	0.0592	0.0000
<i>D</i>	-0.0974	0.0121	0.0000
<i>act_dur</i>	0.0001	0.0002	0.6880
<i>inc</i>	0.0299	0.0050	0.0000
<i>r_estd</i>	-0.0022	0.0003	0.0000
<i>constant</i>	1.7226	0.1351	0.0000
<b><i>Transit demand, TD</i></b>			
<i>TC</i>	0.6548	0.0732	0.0000
<i>AS</i>	-0.3002	0.0920	0.0010
<i>WD</i>	-0.0800	0.0124	0.0000
<i>RL</i>	0.0057	0.0021	0.0070
<i>tswork</i>	0.3848	0.0422	0.0000
<i>prkride</i>	-0.0737	0.0514	0.1510
<i>ts_tod</i>	0.2063	0.1097	0.0600
<i>veh</i>	-0.0456	0.0221	0.0390
<i>constant</i>	-0.1256	0.1014	0.2150

N= 8,229; F<sub>TC</sub>=49.3; F<sub>AS</sub>=73.6; F<sub>TD</sub>=122.1

Table 4.9 shows that, for example, a 20 percent increase in gross population density, which is equal to about 1,830 persons per square mile, produces an approximate nine percent increase in transit demand (linked trips at household level). Transit station proximity also plays a relevant role. A doubling of the average walking distance to the nearest transit station, or an increase from 0.3 miles to 0.6 miles, decreases transit demand by 14 percent; at about one mile, transit demand declines by 28 percent.

The presence of a transit station within a half-mile of the workplace increases transit demand by 69 percent. Living in proximity to a TOD transit station increases transit demand by

about 28 percent. There seems to be a ridership bonus associated with proximity to a station characterized by accessibility features intended to promote transit use.

**TABLE 4.9** Elasticity Estimates—Model I

<i>Elasticity</i>	<i>RL</i>	<i>WD</i>	<i>D</i>	<i>subc_dist</i>	<i>r_estd</i>	<i>tswork*</i>	<i>ts_tod*</i>
<i>TC</i>	0.087	-0.007	-0.044	0.109	0.000	-	-
<i>AS</i>	0.100	-0.008	-0.066	0.125	0.000	-	-
<i>TD</i>	-0.157	-0.137	0.475	-0.388	0.001	0.687	0.279

\* Indicates a proportional change

The model reports a negative elasticity between residential location and transit use. This is consistent with the assumption that households characterized by longer commutes engage in more complex trip chains, which positively affect the spatial extent of non-work activities. As the activity space expands, transit demand declines.

The results also show that transit demand is sensitive to the presence of nearby subcenters, or, in general, to decentralization. The negative sign associated with the elasticities shows that increased polycentricity significantly affects transit demand. The farther a household lives from a subcenter, the less it uses transit. A 50 percent increase in distance to a subcenter (from 2.9 to 4.3 miles) decreases transit demand by about 14 percent. This is because households tend to rely more on other transport modes to carry out more complex trip chains. This result is consistent with the current literature on transit competitiveness and polycentric metropolitan regions. For example, in a study of transit services and decentralized centers, Casello (2007) finds that transit improvements between and within activity centers (i.e., subcenters) are necessary to realize the greatest improvements in transit performance.

Next, Model II is specified to ascertain the extent to which the above relationships are affected by treating residential location as a choice variable (i.e., endogenous and simultaneously determined).

### Model II results

As discussed in Chapter 3, residential self-selection refers to a behavioral aspect that leads individuals or households to prefer certain residential locations due to idiosyncratic preferences for travel. In applied work, if residential self-sorting is not accounted for, findings tend to overstate the relevance of policies geared at impacting travel behavior through planned influence on the built environment.

In expanding this field of research, Model II treats residential location as endogenous while retaining density as exogenous. It assumes that individuals can locate anywhere within an

urban area choosing the most utility carrying *job-residence pair*. This process is carried out in conjunction with the optimal choice of both consumption and non-work travel. A household optimally located at a distance to work engages in trip-chaining to benefit from time-savings gained by combining errands to and from work. Time savings can either be allocated to a move farther out or to engage in additional non-work travel. By treating residential location as a choice variable, Model II can better capture the behavioral elements most likely to affect the transit use and urban form relationship. Model II is specified as

$$TC = \alpha_0 + \alpha_1 AS + \alpha_2 RL + \alpha_3 WD + \alpha_4 veh + \alpha_5 act\_tt + \alpha_6 act\_dur + \alpha_7 sch + \alpha_8 subc\_dist + \varepsilon_1 \quad (4.9)$$

$$AS = \beta_0 + \beta_1 TC + \beta_2 \bar{D} + \beta_3 act\_dur + \beta_4 inc + \beta_5 r\_estd + \varepsilon_2 \quad (4.10)$$

$$TD = \gamma_0 + \gamma_1 TC + \gamma_2 AS + \gamma_3 \bar{WD} + \gamma_4 RL + \gamma_5 tswork + \gamma_6 prkride + \gamma_7 ts\_tod + \gamma_8 veh + \varepsilon_3 \quad (4.11)$$

$$RL = \delta_0 + \delta_1 TC + \delta_2 TD + \delta_3 hprice + \delta_4 hage + \delta_5 rooms + \delta_6 div + \delta_7 pov + \delta_8 own + \varepsilon_4 \quad (4.12)$$

Housing characteristics (pricing, age, size) are considered relevant factors affecting residential location, as well as neighborhood characteristics (ethnicity, crime). In terms of exclusion restrictions, equation (4.12) assumes that while residential location is affected by travel decisions (trip chaining and transit use), housing and neighborhood characteristics do not directly affect travel behavior. Other housing characteristics variables, such as the stock of housing built before 1945, are not included in equation (4.12) as they serve the same role of those just discussed (beside being highly correlated with one another and potentially causing multicollinearity).

Equation (4.10) passes the rank condition for identification and is classified as just identified. Table 4.10 displays the results of the 2SLS regression.

**TABLE 4.10** 2SLS Regression Results—Model II

<i>Equation</i>	<i>Coefficient</i>	<i>Std. Err.</i>	<i>P</i>
<b><i>Trip chaining, TC</i></b>			
RL	0.0096	0.0118	0.4130
AS	0.0725	0.1980	0.7140
WD	-0.0573	0.0142	0.0000
veh	-0.0786	0.0316	0.0130
act_tt	0.0014	0.0005	0.0020
act_m	-0.0022	0.0003	0.0000
subc_dist	0.0435	0.0070	0.0000
sch	0.0778	0.0144	0.0000
constant	1.2604	0.2673	0.0000
<b><i>Activity space, AS</i></b>			
TC	0.2357	0.0538	0.0000
D	-0.0858	0.0107	0.0000
act_m	-0.0007	0.0002	0.0000
hhinc	0.0412	0.0045	0.0000
r_estd	-0.0014	0.0003	0.0000
constant	2.0943	0.1202	0.0000
<b><i>Transit demand, TD</i></b>			
TC	0.6964	0.0753	0.0000
AS	-0.2598	0.1157	0.0250
WD	-0.0669	0.0127	0.0000
RL	-0.0090	0.0088	0.3110
tswork	0.3716	0.0446	0.0000
prkride	-0.0669	0.0524	0.2020
ts_tod	0.1304	0.1147	0.2560
veh	-0.0365	0.0221	0.0990
constant	-0.1119	0.1020	0.2720
<b><i>Residential location, RL</i></b>			
TC	3.7324	0.5009	0.0000
TD	-1.2408	0.4660	0.0080
hprice	-2.8117	0.2722	0.0000
hage	-0.0849	0.0094	0.0000
rooms	1.1279	0.1468	0.0000
div	-2.6312	0.7238	0.0000
pov	-5.9629	2.4133	0.0130
own	0.4966	0.2658	0.0620
constant	39.1808	3.3743	0.0000

N= 8,212;  $F_{TC}=42.7$ ;  $F_{AS}=72.5$ ;  $F_{TD}=118.5$ ;  $F_{RL}=57.2$ 

The relevant signs and coefficients' magnitudes of the first three equations are consistent with those of Model I. As discussed in Chapter 3, the relationship between residential location and transit demand ( $TD_{RL} \leq 0$ ) could not be established a priori. Table 4.10 reports a negative

sign ( $TD_{RL}=-1.24$ ; statistically insignificant) indicating that residential location does not exert a relevant influence on transit demand. This might be due to the transit supply characteristics where the travel survey was conducted (e.g., fairly well served commute routes). The parameter does not have a *ceteris paribus* interpretation as it changes concurrently with the other endogenous variables. Compared to Model I, changes in activity space negatively affect transit use. More dispersed activity-travel locations result in reduced transit patronage, although this effect is now less relevant.

As done with Model I, a comparative assessment can be done by producing relevant point elasticities, as summarized by Table 4.11 (only reporting statistically significant estimates).

**TABLE 4.11** Elasticity Estimates—Model II

<i>Elasticity</i>	<i>WD</i>	<i>D</i>	<i>subc_dist</i>	<i>r_estd</i>	<i>tswork*</i>
<i>TC</i>	-0.009	-0.036	0.108	-0.023	-
<i>AS</i>	-0.003	-0.069	0.041	-0.232	-
<i>TD</i>	-0.028	0.269	-0.065	0.170	0.766
<i>RL</i>	0.002	-0.027	0.052	0.405	-

\* Indicates a proportional change

Compared to Model I, the endogenous treatment of residential location reduces the magnitude of the elasticity of travel demand with respect to density elasticity by 50 percent. In a context where households can locate anywhere in an urban area and can optimize the trip chaining and commuting costs, an exogenous 20 percent increase in density produces a 5.4 percent increase in the demand for transit (household linked trips).

Transit station proximity also loses relevance. An increase from 0.3 to 0.6 miles to the nearest transit station reduces transit demand by only 2.8 percent as opposed to the 14 percent reduction of Model I. The relevance of transit station proximity at workplace is increased; the presence of a transit stop within a half-mile of workplace increases transit demand by about 76 percent.

To understand the reasons for these changes, it is sufficient to look at the behavioral links underlying Model II. Equation (4.12) outlines a behavioral process where households optimally choose location and non-work activities, which also optimally define the spatial extent of non-work activities. Taking into consideration housing and neighborhood characteristics, households can optimally locate at a relative distance to the nearest subcenter where employment and non-work activities are likely to be concentrated, trading off commuting distance (and costs) with housing costs. Households who own their residence locate farther



from work, trading lower housing costs with increased commute distance. Trip chaining optimization is part of this trade-off process, which leads to an expansion of the activity space. This in turn reduces the opportunities to use transit to engage in non-work travel. This behavior is empirically validated by the statistical significance of all housing and neighborhood characteristics controls of equation (4.12).

### Model III results

Up to this point, urban form has been treated as exogenous to this process. What happens if urban form, as measured by gross population density, is indeed affected by travel decisions? To what extent is the relationship outlined by Model I and Model II impacted by treating density as endogenous to this process? Next, Model III introduces an additional equation relating density to this process.

The endogenous treatment of density specifies the following model

$$TC = \alpha_0 + \alpha_1 AS + \alpha_2 RL + \alpha_3 WD + \alpha_4 veh + \alpha_5 act\_tt + \alpha_6 act\_dur + \alpha_7 sch + \alpha_8 subc\_dist + \varepsilon_1 \quad (4.13)$$

$$AS = \beta_0 + \beta_1 TC + \beta_2 D + \beta_3 act\_dur + \beta_4 inc + \beta_5 r\_estd + \varepsilon_2 \quad (4.14)$$

$$TD = \gamma_0 + \gamma_1 TC + \gamma_2 AS + \gamma_3 WD + \gamma_4 RL + \gamma_5 tswork + \gamma_6 prkride + \gamma_7 ts\_tod + \gamma_8 veh + \varepsilon_3 \quad (4.15)$$

$$RL = \delta_0 + \delta_1 TC + \delta_2 TD + \delta_3 hprice + \delta_4 hage + \delta_5 rooms + \delta_6 div + \delta_7 pov + \delta_8 own + \varepsilon_4 \quad (4.16)$$

$$D = \vartheta_0 + \vartheta_1 RL + \vartheta_2 AS + \vartheta_3 subc\_dist + \vartheta_4 cbd\_dist + \varepsilon_5 \quad (4.17)$$

Equation (4.17) considers gross population density around a household residential unit location as endogenous with respect to residential location and travel preferences. A variable serving as a proxy for centrality dependence (*cbd\_dist*) and one serving as a proxy for polycentricity (*subc\_dist*) are introduced as exogenous to the model. The last equation passes the identification test, and is also classified as over identified.

There is no substantial difference in the relevant explanatory variables of interest. Regarding equation (4.7) both CBD and subcenter distance are statistically significant. The sign of the CBD measure of centrality *cbd\_dist* has the expected sign. As distance to CBD increases, density decreases, while as the distance to the nearest subcenter increases so does density. This is indicative of the spatial attraction effect exerted by CBD relative influence within the polycentric context of the urban landscape where BATS2000 travel diary data were collected. The relevance of these two variables is better highlighted by the elasticities presented in Table 4.13.

TABLE 4.12 2SLS Regression Results—Model III

<i>Equation</i>	<i>Coefficient</i>	<i>Std.Error</i>	<i>P</i>
<b><i>Trip chaining, TC</i></b>			
RL	0.0777	0.0158	0.0000
AS	1.0087	0.2241	0.0000
WD	-0.6626	0.0554	0.0000
veh	-0.0292	0.0316	0.3570
act_tt	-0.0009	0.0005	0.0560
act_m	-0.0004	0.0003	0.2650
subc_dist	0.1875	0.0317	0.0000
sch	0.0570	0.0132	0.0000
constant	-2.9369	0.3204	0.0000
<b><i>Activity space, AS</i></b>			
TC	0.5389	0.0654	0.0000
D	-0.2817	0.0002	0.0000
act_m	0.0000	0.0050	0.8390
hhinc	0.0182	0.0316	0.0000
r_estd	-0.0018	0.0010	0.0000
constant	3.5109	0.2583	0.0790
<b><i>Transit demand, TD</i></b>			
TC	0.2310	0.0782	0.0030
AS	0.2130	0.1103	0.0540
WD	-0.4740	0.0405	0.0000
RL	0.0162	0.0089	0.0700
tswork	0.4463	0.0414	0.0000
prkride	-0.0788	0.0457	0.0840
ts_tod	0.1280	0.0995	0.1980
veh	-0.0641	0.0204	0.0020
constant	-1.3114	0.1379	0.0000
<b><i>Residential location, RL</i></b>			
TC	2.4695	0.4889	0.0000
TD	1.1677	0.4700	0.0130
hprice	-2.7930	0.2491	0.0000
hage	-0.0961	0.0080	0.0000
rooms	1.3432	0.1071	0.0000
div	-6.1904	0.5571	0.0000
pov	-4.5075	1.6515	0.0060
own	1.3780	0.1901	0.0000
constant	40.3105	3.0969	0.0000
<b><i>Density, D</i></b>			
RL	-0.0091	0.0108	0.4000
AS	-0.5333	0.0710	0.0000
cbd_dist	-0.0401	0.0016	0.0000
subc_dist	-0.0707	0.0301	0.0190
constant	11.7688	0.1800	0.0000

$N = 8,212$ ;  $\chi^2_{TC} = 2,512.8$ ;  $\chi^2_{AS} = 611.2$ ;  $\chi^2_{TD} = 1,712.7$ ;  $\chi^2_{RL} = 646.3$ ;  $\chi^2_D = 1,448.6$

**TABLE 4.13** Elasticity Estimates—Model III

<i>Elasticity</i>	<i>WD</i>	<i>subc_dist</i>	<i>cbd_dist</i>	<i>r_estd</i>	<i>tswork*</i>
<i>TC</i>	-0.067	-0.195	-1.066	0.014	--
<i>AS</i>	-0.060	-0.088	-0.102	-0.009	--
<i>TD</i>	-0.093	-0.522	-1.177	-0.366	0.961
<i>RL</i>	-0.023	-0.076	-0.301	0.011	--
<i>D</i>	-0.012	-0.153	-2.972	-0.002	--

\* Indicates a proportional change

Compared to Model I and Model II, the joint endogenous treatment of residential location and density produces a model whose relevant hypotheses are confirmed. All relevant signs associated with trip chaining behavior, *TC*, the determination of activity space, *AS*, and residential locations, *RL*, are confirmed.

The elasticity of travel demand with respect to walking distance is less than that of Model I, but greater (in absolute terms) than that of Model II. An increase from 0.3 to 0.6 miles to the nearest transit station reduces transit demand by 9 percent, compared to the 14 percent reduction of Model I, and 2.4 percent reduction of Model II. The relevance of transit station proximity at workplace increases; the presence of a transit stop at workplace almost doubles the demand for transit.

The sign associated with the centrality measure (*cbd\_dist*) and its statistical significance confirm the relevance of the CBD as a generator of transit ridership. Treating density endogenous results in a more elastic travel demand with respect to distance to the nearest transit center. It is relevant to note that both *cbd\_dist* and *subc\_dist*, appear as explanatory variables, but are treated as endogenous to the model. An initial specification treated these two variables as exogenous, but overidentification tests (discussed in the next chapter) revealed that this treatment led to weak instruments (a problem leading to inconsistent estimates).

The exogenous treatment of subcenters assumes that they directly affect density, *D*, without being affected by its changes. The literature on the formation of subcenters demonstrates that the exogenous treatment of subcenters presents problems related to their identification and to the role they play in affecting both employment and population density. Recent studies show that the formation of subcenters is endogenous to the process leading to urban development (i.e., subcenters are endogenous to changes in density)(McMillen, 2001).

The elasticity of transit demand with respect to distance to the CBD (-1.17) is greater than the elasticity with respect to distance to the nearest subcenter (-0.52). In other words transit patronage is more elastic with respect to a household's relative location to a centralized place than a decentralized center. This is probably due to differences in existing transit station locations near the CBD compared to suburban areas. This result contrasts recent findings looking at increased transit use in better served decentralized urban areas (Brown and Thompson, 2008, Thompson and Brown, 2006) and empirical findings showing that that transit ridership is not affected by the strength of a CBD (Brown and Nego, 2007).

In the next chapter, the models are subject to post-estimation testing to confirm their statistical validity. Factors that could potentially affect the validity of results are also discussed in detail.

## Chapter 5      Validity of Hypotheses

---

### Introduction

The validity of the empirical results hinges on factors associated with the quality of the data used and the statistical techniques employed. This chapter discusses some of the key factors that might affect the results of the empirical investigation. There are several issues that can potentially affect the empirical results of Chapter 4, namely:

- Dataset issues
  - Measurement problems
  - Scaling
- Modeling issues
  - Use of cross sectional data
  - Misspecification
  - Endogeneity not accounted for
  - Nonlinearities

### Dataset Issues

The travel behavior dataset herein used relies on the travel diary information from BATS2000. The precise geographic coordinate measurement of households' residences and their travel patterns allows computing the dependent variables residential location,  $RL$ , activity space,  $AS$ , and other measures, such as the household residential unit distances to the CBD and the nearest subcenter. In addition, the land use data obtained from the Census 2000 Summary File 3, and from the Census county business patterns survey (CBP) are drawn from different geographical units. While the former provides measures at the block group level, the CBP provides land-use data at the zip code level. Scale measurement issues can arise and are discussed below.

### Measurement issues

While residential location is measured as home-work distance, other measures could have been considered as well. For example, an alternative is represented by the average commute time length of household employment activities to the household residential unit. This measure presents the advantage of accounting for spatial characteristics as well as network ones (such as street network design and level of service). It also represents a measure of the opportunity cost of residing at a certain distance from work. This measure can be expressed as

$$RL_{min} = \frac{\sum_{m=1}^k dur_{mj}}{k} \quad (5.1)$$

where  $dur_{mj}$  is commute length (measured in minutes of travel) to the residential unit located at  $j$ , from a household member work location  $m$ , and  $k$  is the total number of employed household members.

The final models measure residential location in terms of linear distance. This is because distance is measured using precise geographical coordinates of each household residential unit and the corresponding work location, thus providing a relatively accurate measure. On the other hand, measuring residential location in terms of travel time entails using the survey reported travel time, which is inherently affected both by measurement error (under or overstatement of actual travel time on behalf of the respondents), and unobserved factors related to the time the survey was conducted (unobserved, non-random, factors affecting traffic levels during the two-day data collection period).

Chapter 3 and Chapter 4 discussed the definition and measurement of activity space, AS, and the adoption of the distance ellipse (SDE) to measure the household spatial dispersion of non-work activities. In choosing SDE, a comparison between the second best alternative, the standard distance circle (SDC) was conducted. As discussed, the advantage of SDE over SDC is the reduction of skewness introduced by outliers. Indeed, sample descriptive statistics showed outlier influence that could not be eliminated without relevant loss of information. In addition, SDE was normalized using a log transformation that improved the linearization of the relationship between the other relevant variables.

The literature provides additional measures as well. For example, while Buliung and Kanaroglou (2006) use SDE as a relevant measure, they also introduce a new ecology-based metric defined as the household activity space (HAS). HAS is an area-based geometry that defines a minimum convex polygon containing activity locations visited by a household during the course of a reference period (i.e., the travel survey period). The advantage of HAS is that it weights the activity space by the relevance of activities, such as their type (recreational, maintenance, etc.) and their relative frequencies. Although HAS reports an accurate geographical measurement of the activity space, Buliung and Remmel (2008) show that the use of the minimum convex polygon algorithm provides similar results to SDE in terms of behavioral interpretation. Other research shows that the choice of an appropriate shape representing an individual's activity space is highly dependent on the spatial distributions and frequencies of the locations visited by the person in the given time period (Rai, et al., 2007).

### **Measurement scale issues**

As described in detail in Chapter 4, land use and urban form variables herein employed are measured at two geographic levels of detail. Gross population density is measured at the Census block group level. This scale of measurement, besides being the level most

corresponding to the neighborhood, is also consistent with the literature and allows comparison of findings. Retail establishment density, a proxy for land-use mix (commercial land uses) is measured at the zip code level, which is a wider geographical area. As argued by Boarnet and Crane (2001), this scale is appropriate especially when investigating the role of non-work travel, as non-work trips are usually long in distance and encompass more than a block group from the residential unit. Zip code level scale for the geographic area of the sample dataset also coincides with the traffic analysis zone (TAZ)<sup>9</sup>. In this study, retail establishment density directly affects the activity space equation. As summarized in Table 5.1, the average size of the activity space is much larger than the average size of a block group level, while the average size of zip codes is more consistent with the extent of the activity space.

**TABLE 5.1** Land Area Geographic Measures

<i>Transit Household</i>		<i>Household Activity Space (mile<sup>2</sup>)</i>	<i>Block Group Area (mile<sup>2</sup>)</i>	<i>Zip Code Area (mile<sup>2</sup>)</i>
No	Mean	17.2	2.30	42.88
	SD	38.4	10.92	88.66
	N	10,548	12,260	12,260
Yes	Mean	19.1	0.87	18.37
	SD	37.8	4.18	51.33
	N	2,176	2,503	2,503
Overall Sample	Mean	17.5	2.06	38.72
	SD	38.3	10.11	84.02
	N	12,724	14,763	14,763

## Modeling Issues: Post Estimation Tests

The models presented in Chapter 4 explicitly deal with endogeneity of urban form and travel. Specific econometric techniques are needed to deal with this issue, which call for the application of simultaneous equation modeling. As seen, the first step requires correctly identifying a model. This step generates models that are either just identified or overidentified, based on the number of exclusion restrictions applied to each equation (See Appendix C for more details). To judge the validity of results and the overall model performance, specific statistical tests are called for.

<sup>9</sup> According to the U.S. Census Bureau, a TAZ is a special area delineated by state and/or local transportation officials for tabulating traffic-related data, especially journey-to-work and place-of-work statistics (2008).

### Tests of endogeneity and overidentification

A property of the 2SLS regression is its loss of efficiency if the explanatory variables treated as endogenous are, in fact, exogenous, making its use unnecessary when compared to regular OLS regression. It is thus useful to test the explanatory variables suspected to be endogenous to the model.

The null hypothesis of the endogeneity test states that an OLS estimator of the same equation would yield consistent estimates; that is, any endogeneity among the regressors would not have deleterious effects on OLS estimates. A rejection of the null indicates that endogenous regressors' effects on the estimates are meaningful, and instrumental variables techniques are required. The test was first proposed by Durbin (1954) and separately by Wu (1974) and Hausman (1978). The procedure to test endogeneity of multiple explanatory variables requires (i) estimating in reduced form each endogenous variable on all exogenous variables (including those in the structural equation and those used as instruments; i.e., the explanatory variable included in the other equations); (ii) adding the estimated error terms back into the structural equation; and, (iii) testing for the joint significance of these residuals in the structural equation. Joint significance indicates that at least one variable is endogenous to the model. Under the null hypothesis, the test statistic is distributed  $\chi_q^2$  (Chi-squared) with  $q$  degrees of freedom, where  $q$  is the number of regressors specified as endogenous in the original instrumental variables regression. The procedures to conduct this test are obtained in Stata® (the statistical package used in this study) using the *ivreg2* routine (specifically, by using the command *ivendog*) developed by Baum et al. (2007).

Furthermore, after verifying the presence of endogeneity, additional tests are needed to confirm the correct choice of the exclusion restrictions characterizing the system of equation. These tests are needed to confirm the proper choice of instruments and to eliminate doubts of a poor model performance (bias and inconsistency). The overidentification tests used here are conducted by regressing the residuals from a 2SLS regression on all exogenous variables (both included exogenous regressors and excluded instruments). Under the null hypothesis that all instruments are uncorrelated with the residuals, a Lagrangian multiplier (LM) statistic of  $N \times R^2$  ( $N$  = number of regressors;  $R^2$  = r-squared from the residuals' regression) form, has a large sample Chi-squared distribution,  $\chi_r^2$ , where  $r$  is the number of *overidentifying* restrictions (i.e., the number of excess instruments). If the hypothesis is rejected, there is a doubt about the validity of the instrument set; one or more of the instruments do not appear to be correlated with disturbance process. The Stata® procedure reports the Sargan (1958) overidentification test (using the *overid* command).

Finally, when dealing with a relatively large number of exclusion restrictions, a situation herein encountered in Model III, it has been shown that the power of the overidentification



tests is reduced (Baum, Schaffer and Stillman, 2007). Furthermore, there is a need to be able to test subsets of instruments to identify weak ones and thus affect the validity of results. In this context, another test statistic can be used to test a subset of instruments; the *difference-in-Sargan* test, or *C* test. The statistic is computed as the difference between two statistics; one obtained by regression using the entire set of instruments, and a second one obtained with the smaller set of restrictions (excluding the suspected variables). Under the null hypothesis that the variables are proper instruments, the *C* test statistics is distributed  $\chi_k^2$  with  $k$  degrees of freedom equal to the number of suspect instruments being tested.

Table 5.2 reports the results of the endogeneity and overidentification tests for the travel demand equation, *TD* (the same tests and same results were obtained for the other equations). The Durbin-Wu-Hausman (DWH) test is numerically equivalent to the standard Hausman endogeneity test. Results across the three models indicate the presence of endogeneity, confirming the appropriateness of 2SLS versus OLS regression.

Model III fails the overidentification test in its initial specification that treated the land-use measures (*cbd\_dist*, *subc\_dist*, *r\_estd*) as exogenous to the system (Sargan test = 24.951; *p-value* = 0.0030). After their endogenous treatment, Model III passed the overidentification test, as signaled both by the Sargan (7.1540 with *p-value* of 0.3068) and *C* tests.

Overall, the tests indicate that SEM is an appropriate technique, and that the equation specifications of Chapter 4 produces models that also pass the overidentification tests. The validity of the models allows making conclusions regarding the parameters of interest.

**TABLE 5.2** Endogeneity and Overidentification Tests

	<i>Model I</i>	<i>Model II</i>	<i>Model III</i>
Wu-Hausman <i>F</i> test	78.0733	83.3688	12.9996
p-value	0.0000	0.0000	0.0000
Durbin-Wu-Hausman $\chi^2$ test	153.4231	243.0590	90.2171
$\chi^2$ p-value	0.0000	0.0000	0.0000
Anderson canon. corr. LR statistic (identification/IV relevance test):	42.1370	27.1370	33.5240
$\chi^2$ p-value	0.0000	0.0025	0.0000
Sargan statistic (overidentification test of all instruments):	9.6380	11.3650	24.9510
$\chi^2$ p-value	0.0570	0.2515	0.0030
Sargan statistic without suspect instruments*	-	-	7.1540
$\chi^2$ p-value	-	-	0.3068
C statistic (exogeneity/orthogonality of suspect instruments)**			17.7980
$\chi^2$ p-value			0.0005

\* Test conducted after endogenous treatment of: *cbd\_dist*, *subc\_dist*, *r\_estd*

\*\* Test conducted on exclusion of instruments: *cbd\_dist*, *subc\_dist*, *r\_estd*

## Other Issues

The use of SEM is best exploited in the context of panel datasets. Panel datasets consist of data collection on the same observational units over time. It is argued that the use of SEM and panel data better uncovers underlying causality among the relationships of interest. In the transportation literature there exist several applications of SEM using cross-sectional data. For example, Pendyala (1998) uses SEM to investigate the homogeneity of causal travel behavior across a population of interest; Fuji and Kitamura (2000) and Golob (2000) develop models of trip generation developing models of activity duration and trip generation. Additional examples of applications of SEM using cross-sectional datasets are discussed by Golob (2003).

The models of this study require a substantial amount of information, not only in terms of travel behavior data from travel diaries, but also on the spatial location of residences, work, and non-work activities. This once represented a major constraint to the development of sophisticated frameworks, but it nowadays seems to be overcome by the advances in activity-based travel survey design.

The increased sophistication of communication systems that can easily track individuals' travel patterns in space and time make the data collection effort a less daunting one, allowing increased use of sophisticated models, such as the one developed in this study. For example, the recent use of GPS tracking devices reveals that human behavior results in optimized patterns of travel based on socio-demographic characteristics. These methods not only allow

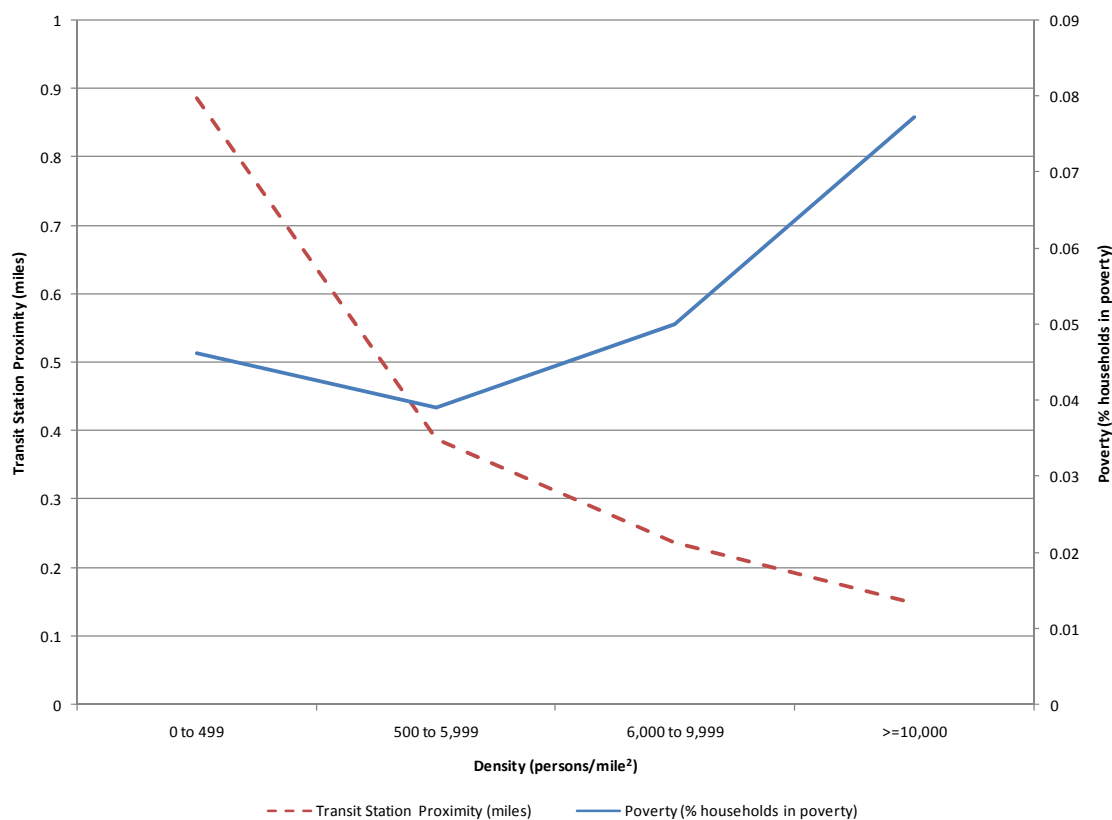
tracking travel and non-work activity locations, they also provide more accurate measures of travel itself, such as actual travel time speed based on network characteristics.

### **Transit station proximity**

Notwithstanding the validity of the above post-estimation tests, there still exists the possibility of endogeneity of some of the exogenous variables. This endogeneity although confuted by statistical tests, is not discounted by theoretical assumptions. For example, while this study treats vehicle ownership as exogenous and not directly influenced by the location decisions, the literature review encountered studies that consider vehicle ownership as a discrete choice endogenous to the residential location process and to density levels. One extension to the research might include an endogenous treatment of this variable, while overcoming the limitations imposed by ad-hoc choice set specifications.

Endogeneity also extends to transit supply measures. For example, measures of supply, such as the number of transit stations and frequency of service are treated as exogenous to the model. As discussed in several instances throughout this report, the implications of treating a variable exogenous, while being endogenous to the process, have non-trivial consequences.

One additional consideration must be made regarding the use of walking distance as a measure of transit station proximity that cannot be made when using the more traditional half-mile buffer. As density increases, the number of transit stops at the geographical unit (i.e., block group level) increase. This inherently reduces the average distance from any given household to its nearest transit station independently of location preferences. Furthermore, as shown in Figure 5.1, in densely populated areas, stations are located in neighborhoods characterized by higher than average poverty levels, and that are increasingly diverse (i.e., characterized by ethnic minorities). In other words, in higher urban density settings, an inherent supply-side spatial bias is present and correlated to relevant instrumental variables that control for neighborhood characteristics. It can be shown that, when not accounted for, this supply-side bias results in an overestimation of the relevance of station proximity to a given residential unit (a working paper dealing with this issue is presented Appendix D). For this reason, Model III, which endogenously treats residential location and density, considers walking distance as endogenous to the process.

**FIGURE 5.1** Poverty and Transit Station Proximity

## Implications

The models estimated in Chapter 4 are derived from a generalized behavioral framework of residential location and travel behavior that is innovative in many aspects; above all for its explicit incorporation of behavioral links between consumption, travel, the spatial location of non-work activities, and the ensuing interrelationship with the surrounding environment.

The empirical application of the behavioral model requires the use of simultaneous equation modeling. The biggest challenge when employing structural equation modeling lies in defining properly specified models. The necessary identification steps outlined in Chapter 4 and summarized in Appendix C are paramount to reliable estimates. The literature reviewed in this study revealed that none of the papers and studies formally follows this process. The result is the estimation and presentation of sets of parameters that are not unique, which make statistical inference unreliable. The validity of the empirical models of Chapter 4 is confirmed by the relevant endogeneity and overidentification tests presented in this chapter.

Finally, there exist more advanced econometric techniques that would allow relying on less restrictive assumptions than those required by multiple linear regression. These techniques describe an advanced field of research defined as nonparametric econometrics. These methods

permit uncovering the presence of nonlinearities among dependent and independent variables which would guide to a better parameterization of an equation of interest. Being computationally challenging, they are rarely used in applied work, especially in the field of travel behavior research. Further research that makes use of these methods is warranted.

*[This page intentionally left blank.]*

## Chapter 6 Conclusions

---

This research effort sought to develop an integrated model of transit travel behavior and urban form. A review of the current state of empirical research on the subject uncovered the main weaknesses of findings relating the built environment to travel behavior as well as noting the paradigm shift epitomized by the activity-based literature. To avoid these shortcomings and to incorporate the activity-based approach, this report developed and estimated a simultaneous general equilibrium model of transit usage and urban form.

The model presented in Chapter 3 allows household travel behavior to respond to changes in urban form, including trip-chaining for non-work travel. In the model, trip-chaining behavior results from households' desire to reduce travel time while accounting for constraints that the built environment imposes. Any travel-time saving is spent on additional non-work travel or provides inducement to reassess residential location decisions. These changes in travel behavior and residential location then affect the demand for travel.

Empirical evidence in Chapter 4 shows that lower densities define a broader activity space, which, in turn, decreases transit use. As density increases, the activity space contracts, as does the need to engage in complex trip chains. Idiosyncratic preferences for transit also affect transit demand. For example, in the absence of adequate transit, households that need to engage in complex trip chain patterns, independent of changes in the surrounding built-environment, may use the automobile. In contrast, if adequate transit services are available to accommodate their travel patterns, households would choose transit, other things equal.

To facilitate a summary of Chapter 4's findings and for ease of comparison, Table 6.1 presents elasticities from the three estimated models (only statistically significant results are reported).

The importance of station proximity is lessened after accounting for idiosyncratic preferences for location. In Model III, in which residential location and density are endogenous, the elasticity of transit demand with respect to walking distance decreases by about 33 percent over the result for Model I, in which density and residential location are exogenous. This decline in elasticity's magnitude is due to omitted bias error in a model that does not explicitly account for residential self-selection. This result is somewhat consistent with that of Cervero (2007), who found that self-selection accounts for about 40 percent of transit ridership for individuals residing near a transit station.

**TABLE 6.1** Relevant Land-Use and Transit-Supply Elasticities

<i>Elasticity</i>	<i>Model I<sup>a</sup></i>	<i>Model II<sup>b</sup></i>	<i>Model III<sup>c</sup></i>
<i>Density</i>	0.475	0.269	n/a
<i>Walking distance</i>	-0.137	-0.028	-0.093
<i>Transit station at workplace*</i>	0.687	0.766	0.961
<i>TOD station*</i>	0.279	n/a	n/a
<i>Distance to CBD</i>	n/a	n/a	-1.177
<i>Distance to nearest subcenter</i>	-0.388	-0.065	-0.522
<i>Retail establishments density</i>	0.001	0.170	n/a
<i>Residential location</i>	-0.157	n/a	n/a

<sup>a</sup> residential location exogenous; density exogenous

<sup>b</sup> residential location endogenous; density exogenous

<sup>c</sup> residential location and density endogenous

n/a = not available

\* Indicates a proportional change

Density does not have a large effect on transit demand, and the relative magnitude of the effect decreases when residential location is endogenous. A 20 percent increase in gross population density (1,830 persons per square mile) increases transit demand from a minimum of 5.4 percent to a maximum of 9.5 percent.

The importance of mixed-use development to increase transit patronage is highlighted by the elasticity of travel demand with respect to retail establishment density. Model II shows that a 20 percent increase in retail establishment density (or about 28 establishments per square mile) increases transit demand by 3.4 percent.

Households living farther from work, *ceteris paribus*, use less transit, which is due to trip-chaining behavior. Such households engage in complex trip chains and have, on average, a more dispersed activity space, which requires reliance on more flexible modes of transportation. Policies that reduce the spatial allocation of activities and improve transit accessibility at and around subcenters would increase transit demand. Similar results can be obtained by policies that increase the presence of retail locations in proximity to transit-oriented households.

Centrality and the strength of an established CBD are relevant drivers of transit use, as highlighted by the elasticity of transit demand with respect to distance from the CBD. Subcenters also play a relevant role, indicating the need to provide services in decentralized employment and residential areas to increase ridership.



In Model I, transit-oriented development near transit stations has a positive impact on transit use; a TOD stop increases transit demand by about 28 percent. Unfortunately, this finding cannot be replicated in Models II and III, probably due to the relatively low number of TOD stations in the sample. In conformity to the literature, a transit station near a workplace exerts a positive impact on ridership, as indicated by the magnitude of the proportional changes across all three models.

## Research Contributions

The major contribution of this research effort is the development of an integrated behavioral model of transit patronage and land-use, which acknowledges the interrelationship characterizing travel behavior and urban form. In particular, the framework embraces the paradigm shift from trip generation to activity-based modeling by considering travel demand as a derived demand brought about by the necessity to engage in out-of-home activities. In addition, this framework:

- departs from the monocentric models of residential location, which do not account for increased decentralized urban settings, by explicitly acknowledging both the presence and the relevance of subcenters;
- accounts for the trade-off between consumption and travel brought about by the finite nature of time and its allocations among household members;
- shifts the analysis from individual travel behavior to household travel behavior;
- can accommodate extensions to account for the endogeneity of time allocation across activities and households; and
- takes advantage of the advances in geographic information systems (GIS) tools and geographic science contributions to the spatial analysis of the interactions of travel behavior and urban form.

The consequences introduced by this structure are non-trivial as demonstrated by the comparative static analysis. For example, density is no longer assumed directly to affect the demand for travel. Rather, density is assumed to represent a constraint on individuals' ability optimally to determine consumption and travel. This framework also explicitly acknowledges suburbanization, by allowing for polycentricity versus monocentricity.

## Directions for Further Research

Notwithstanding the validity of the post-estimation tests performed in Chapter 5, there still exists the possibility that some of the variables treated as exogenous are, in fact, endogenous. For example, this study treats vehicle ownership as exogenous. The literature review, however, revealed studies that consider vehicle ownership endogenous to residential location and

density. One extension to this research, therefore, would be to include an endogenous treatment of this and other mode-choice variables.

Another extension would be to include leisure time available to households. Indeed, the behavioral model of Chapter 3 assumes that households can save time by engaging in trip chaining. Time savings are then reallocated to either more non-work travel or to an extended commute. The model does not explicitly explain what happens to leisure time. The inclusion of an identity that summarizes all time uses (in-home and out-of-home) would provide insight on time use and its effect on trip chaining.

Finally, in contrast to multiple linear regression analysis, nonparametric estimation methods would permit less restrictive assumptions. These methods can uncover the presence of nonlinearities among dependent and independent variables which could lead to a better parameterization of equations of interest. Although nonlinearity in trip-chaining formation and density levels is better captured by these methods than by more commonly used techniques, being computationally challenging, they are rarely used in applied work, especially in the field of travel behavior research and simultaneous equation modeling. Further research that makes use of these methods is warranted.

# References

---

- Adler, Thomas and Moshe Ben-Akiva. 1979. "A Theoretical and Empirical Model of Trip Chaining Behavior," *Transportation Research Part B: Methodological*, 13B, 243-257.
- Anas, Alex. 2007. "A Unified Theory of Consumption, Travel and Trip Chaining," *Journal of Urban Economics*, 62, 162-186.
- Anas, Alex, Richard Arnott and Kenneth A.Small. 1997. *Urban Spatial Structure*. Berkeley: University of California Transportation Center
- Anas, Alex and Ikki Kim. 1996. "General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration," *Journal of Urban Economics*, 40, 232-256.
- Anas, Alex and Rong Xu. 1999. "Congestion, Land Use and Job Dispersion: A General Equilibrium Analysis," *Journal of Urban Economics*, 45, 451-473.
- Arnott, Richard. 1998. "Economic Theory and the Spatial Mismatch Hypothesis," *Urban Studies*, 35, 1171--1185.
- Badoe, Daniel A. and Eric J. Miller. 2000. "Transportation--Land-Use Interaction: Empirical Findings in North America, and Their Implications for Modeling," *Transportation Research Part D*, 5, 235-263.
- Bagley, Michael N. and Patricia L. Mokhtarian. 2002. "The Impact of Residential Neighborhood Type on Travel Behavior: A Structural Equations Modeling Approach," *The Annals of Regional Science*, 36, 279-297.
- Baum-Snow, Nathaniel and Matthew E. Kahn. 2000. "The Effects of New Public Projects to Expand Urban Rail Transit," *Journal of Public Economics*, 77, 241-263.
- Baum, Christopher F., Mark E. Schaffer and Steven Stillman. 2007. "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing," *Stata Journal* 7, 465-506.
- Begon, M. J. and C.R. Towsend. 1996. *Ecology: Individuals, Populations, and Communities*. Cambridge, MA: Blackwell Science Ltd
- Ben-Akiva, Moshe and John L. Bowman. 1998. "Integration of an Activity-Based Modeling System and a Residential Location Model," *Urban Studies*, 35, 1131--1153.
- Bento, Antonio M., Maureen L. Cropper, Ahmed Mushfiq Mobarak and Katja Vinha. 2005. "The Effects of Urban Spatial Structure on Travel Demand in the United States," *The Review of Economics and Statistics*, 87, 466-478.
- Bertaud, Alain. 2003. "Clearing the Air in Atlanta: Transit and Smart Growth or Conventional Economics?," *Journal of Urban Economics*, 54, 379-400.
- Bhat, Chandra. 1999. "An Analysis of Evening Commute Stop-Making Behavior Using Repeated Choice Observations from a Multi-Day Survey," *Transportation Research Part B: Methodological*, 33, 495-510.
- . 2001. "Modeling the Commute Activity-Travel Pattern of Workers: Formulation and Empirical Analysis," *Transportation Science*, 35, 61-79.

- Bhat, Chandra and Jessica Guo. 2004. "A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling," *Transportation Research Part B: Methodological*, 38, 147-168.
- Boarnet, Marlon and Randall Crane. 2001. "The Influence of Land Use on Travel Behavior: Specification and Estimation Strategies," *Transportation Research Part A*, 35, 823-845.
- Boarnet, Marlon G. and Sharon Sarmiento. 1998. "Can Land-Use Policy Really Affect Travel Behaviour? A Study of the Link between Non-Work Travel and Land-Use Characteristics," *Urban Studies*, 35, 1155-1169.
- Boarnet, Marlon Gary and Randall Crane. 2001. *Travel by Design: The Influence of Urban Form on Travel*. Oxford ; New York: Oxford University Press
- Boarnet, Marlon Gary and Sharon Sarmiento. 1996. *Can Land Use Policy Really Affect Travel Behavior : A Study of the Link between Non-Work Travel and Land Use Characteristics*. Berkeley, Calif.: University of California Transportation Center
- Bollinger, Christopher R. 1997. "The Impact of Rapid Rail Transit on Economic Development: The Case of Atlanta's Marta," *Journal of Urban Economics*, 42, 179-204.
- Bowes, David R. and Keith R. Ihlanfeldt. 2001. "Identifying the Impacts of Rail Transit Stations on Residential Property Values," *Journal of Urban Economics*, 50, 1-25.
- Brown, Jeffrey R. and Gregory L. Thompson. 2008. "The Relationship between Transit Ridership and Urban Decentralization: Insights from Atlanta," *Urban Studies*, 45, 1119-1139.
- Buliung, Ron and Pavlos Kanaroglou. 2006. "Urban Form and Household Activity-Travel Behaviour," *Growth and Change*, 37, 174-201.
- Buliung, Ron N. and Pavlos S. Kanaroglou. 2007. "Activity-Travel Behaviour Research: Conceptual Issues, State of the Art, and Emerging Perspectives on Behavioural Analysis and Simulation Modeling," *Transport Reviews*, 37, 151-187.
- . 2006. "Urban Form and Household Activity-Travel Behavior," *Growth and Change*, 37, 172--199.
- Buliung, Ron and Tarmo K. Remmel. 2008. "Open Source, Spatial Analysis, and Activity-Travel Behavior Resear: Capabilities of the *Aspace* Package," *Journal of Geography Systems*, 10, 191-216.
- Cao, Xinyu, Patricia L. Mokhtarian and Susan L. Handy. 2007. "Do Changes in Neighborhood Characteristics Lead to Changes in Travel Behavior? Structural Equations Modeling Approach," *Transportation*, 35, 535-556.
- University of California. —. 2006. "Examining the Impacts of Residential Self-Selection on Travel Behavior: Methodologies and Empirical Findings," California: Institute of Transportation Studies, University of California
- Casello, Jeffrey M. 2007. "Transit Competitiveness in Polycentric Metropolitan Regions," *Transportation Research Part A*, 41, 19-40.
- Cervero, Robert. 2007. "Transit Oriented Development's Ridership Bonus: A Product of Self Selection and Public Policies," *Environment and Planning A*, 39, 2068-2085.
- Cervero, Robert and Kara Kockelman. 1997. "Travel Demand and the 3ds: Density, Diversity, and Design," *Transportation Research Part D*, 2, 199-219.
- Cervero, Robert and John Landis. 1997. "Twenty Years of the Bay Area Rapid Transit System: Land Use and Development Impacts," *Transportation Research Part A*, 31, 309-333.

- Transportation Research Board. Cervero, Robert, Steven Murphy, Christopher Ferrel, Natasha Goguts and Yu-Hsin Tsai. 2004. "Transit-Oriented Development in the United States: Experiences, Challenges, and Prospects," *TCRP TCRP Report 102*. Washington: Transit Cooperative Research Program, Transportation Research Board
- Cervero, Robert and Kang-Li Wu. 1998. "Sub-Centring and Commuting: Evidence from the San Francisco Bay Area, 1980-90," *Urban Studies*, 35, 1059-1076.
- Chiang, Alpha C. 1984. *Fundamental Methods of Mathematical Economics*. New York: McGraw-Hill, Inc
- Chu, You-Lian. 2003. "Empirical Analysis of Commute Stop-Making Behavior," *Transportation Research Record*, 1831, 106--113.
- Transportation Research Board, Institute of Medicine of the National Academies. Committee on Physical Activity, Health, Transportation, and Land Use. 2005. "Does the Built Environment Influence Physical Activity? Examining the Evidence," 269, ISBN 0-309-09498-4. Health Committee on Physical Activity, Transportation, and Land Use, Transportation Research Board, Institute of Medicine of the National Academies
- Concas, Sisinnio and Philip L. Winters. 2007. "The Impact of Carpooling on Trip Chaining Behavior and Emission Reductions," *Transportation Research Record*, 2010, 83-91.
- Crane, Randall. 2000. "The Influence of Urban Form on Travel: An Interpretive Review," *Journal of Planning Literature*, 15, 1-23.
- Crane, Randall and Richard Crepeau. 1998a. *Does Neighborhood Design Influence Travel? A Behavioral Analysis of Travel Diary and GIS Data*. Berkeley: University of California Transportation Center
- . 1998b. "Does Neighborhood Design Influence Travel?: A Behavioral Analysis of Travel Diary and Gis Data," *Transportation Research Part D*, 3, 225-238.
- Davidson, D. 1991. "Impact of Suburban Employee Trip Chaining on Transportation Demand Management," *Transportation Research Record*, 1321, 82-89.
- Davidson, William, Robert Donnelly, Peter Vovsha, Joel Freedman, Steve Ruegg, Jim Hicks, Joe Castiglione and Rosella Picado. 2007. "Synthesis of First Practices and Operational Research Approaches in Activity Based Travel Demand Modeling," *Transportation Research Part A*, 41, 464--488.
- Domencich, Thomas A. and Daniel McFadden. 1975. *Urban Travel Demand : A Behavioral Analysis : A Charles River Associates Research Study*. New York: North-Holland
- Dong, Xiaojing, Ben-Akiva Moshe, John L. Bowman and Joan L. Walker. 2006. "Moving from Trip-Based to Activity-Based Measures of Accessibility," *Transportation Research Part A*, 40, 163-180.
- Durbin, J. 1954. "Errors in Variables," *Review of the International Statistical Institute*, 22, 23-32.
- Ebdon, David. 1977. *Statistics in Geography a Practical Approach*. Oxford: Basil Blackwell
- Estupiñán, Nicolás and Daniel Rodríguez. 2008. "The Relationship between Urban Form and Station Boardings for Bogotá BRT," *Transportation Research Part A*, 42, 296-306.
- Transit Cooperative Research Program. Evans, John E., Richard H. Pratt, Andrew Stryker and J. Richard Kuzmyak. 2007. "Chapter 17 - Transit Oriented Development," *TCRP Report 95*

- *Traveler Response to Transportation System Changes*, 1-147, 17. Washington: Transit Cooperative Research Program
- Ewing, Reid and Robert Cervero. 2001. "Travel and the Built Environment - a Synthesis," *Transportation Research Record*, 87-114.
- Fuji, S. and R. Kitamura. 2000. "Evaluation of Trip-Reducing Effects of New Freeways Using a Structural Equations Model System of Commuters' Time Use and Travel," *Transportation Research Part B: Methodological*, 34, 339-354.
- Fujita, Masahisa, Jacques-Francois Thisse and Yvese Zenou. 1997. "On the Endogenous Formation of Secondary Employment Centers in a City," *Journal of Urban Economics*, 41, 337-357.
- Giuliano, Geneveve and Kenneth A. Small. 1991. "Subcenters in the Los Angeles Region," *Regional Science and Urban Economics*, 21, 163-182.
- Golob, Thomas F. 2000. "A Simultaneous Model of Household Activity Participation and Trip Chain Generation," *Transportation Research Part B: Methodological*, 34, 355-376.
- . 2003. "Structural Equation Modeling for Travel Behavior Research," *Transportation Research Part B: Methodological*, 37, 1-25.
- Greenwald, Michael J. and Marlon Boarnet. 2001. "Built Environment as Determinant of Walking Behavior: Analyzing Nonwork Pedestrian Travel in Portland, Oregon," *Transportation Research Record*, 1780, 33-42.
- Guevara, Cristian Angelo and Ben-Akiva Moshe. 2006. "Endogeneity in Residential Location Models," *Transportation Research Record*, 1977, 60-66.
- Guo, Jessica and Cynthia Chen. 2007. "The Built Environment and Travel Behavior: Making the Connection," *Transportation*, 34, 529-533.
- Handy, Susan, Xinyu Cao and Patricia Mokhtarian. 2005. "Correlation or Causality between the Built Environment and Travel Behavior? Evidence from Northern California," *Transportation Research Part D*, 10, 427-444.
- Hausman, J.A. 1978. "Specification Tests in Econometrics," *Economics Letters*, 8, 239-245.
- Jara-Diaz, Sergio R. and Francisco J. Martinez. 1999. "On the Specification of Indirect Utility and Willingness to Pay for Discrete Residential Location Models," *Journal of Regional Science*, 39, 675--688.
- Kahn, Matthew. 2007. "Gentrification Trends in New Transit-Oriented Communities: Evidence from 14 Cities That Expanded and Built Rail Transit Systems," *Real Estate Economics*, 35, 155-182.
- Kain, John F. 1968. "Housing Segregation, Negro Employment, and Metropolitan Decentralization," *Quarterly Journal of Economics*, 82, 175-197.
- Kuby, Michael, Anthony Barranda and Christopher Upchurch. 2004. "Factors Influencing Light-Rail Station Boardings in the United States," *Transportation Research Part A*, 38, 223-247.
- Kuppam, Arun K. and Ram M. Pendyala. 2001. "A Structural Equation Analysis of Commuters' Activity and Travel Patterns," *Transportation*, 28, 35-54.
- National Institute of Justice. Levine, Ned. 2005. "Crimestat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations " Washington D.C. : National Institute of Justice, National Institute of Justice



- Levinson, David M. and Ajay Kumar. 1997. "Density and the Journey to Work," *Growth and Change*, 28, 147-172.
- Maat, Kees and Harry J.P. Timmermans. 2006. "Influence of Land Use on Tour-Based Travel in the Netherlands," *Transportation Research Record*, 234--241.
- . 2006. "Influence of Land Use on Tour-Based Travel in the Netherlands," *Transportation Research Record*,
- Maoh, H and Pavlos Kanaroglou. 2007. "Geographic Clustering of Firms and Urban Form: A Multivariate Analysis," *Journal of Geographic Systems*, 9, 229-252.
- McDonald, John F. and Clifford I. Osuji. 1995. "The Effect of Anticipated Transportation Improvement on Residential Land Values," *Journal of Regional Science and Urban Economics*, 25, 261-278.
- McFadden, Daniel. 1973. *Conditional Logit Analysis and Qualitative Choice Behavior*. New York: Academic Press
- . 1974. *Conditional Logit Analysis of Quantal Choice Behavior*. New York: Academic Press
- McGuckin, Nancy and Elaine Murakami. 1999. "Examining Trip-Chaining Behavior: Comparison of Travel by Men and Women," *Transportation Research Record*, 1693, 79-85.
- McMillen, Daniel P. 2001. "Nonparametric Employment Subcenter Identification," *Journal of Urban Economics*, 50, 448-473.
- McNally, Michael G. and Anup Kulkarni. 1997. "Assessment of Influence of Land Use-Transportation System on Travel Behavior" *Transportation Research Record*, 1607, 105-115.
- Messenger, T and R Ewing. 1996. "Transit-Oriented Development in the Sun Belt," *Transportation Research Record*, 1552, 145-153.
- Misra, R and Chandra Bhat. 2002. "Activity Travel Patterns of Non-Workers in the San Francisco Bay Area: Exploratory Analysis," *Transportation Research Record*, 1718, 43-51.
- Misra, R, Chandra Bhat and S Srinivasan. 2003. "Continuous Time Representation and Modeling Framework for Analysis of Nonworker Activity-Travel Patterns: Tour Episodes and Attributes," *Transportation Research Record*, 1831, 11-20.
- Modarres, Ali. 2003. "Polycentricity and Transit Service," *Transportation Research Part A*, 37, 841-864.
- Mokhtarian, Patricia L. and Xinyu Cao. 2008. "Examining the Impacts of Residential Self-Selection on Travel Behavior: A Focus on Methodologies," *Transportation Research Part B: Methodological*, 42, 204-228.
- MORPACE International. MORPACE. 2002. "Bay Area Travel Survey 2000 - Final Report," Oakland, CA: MORPACE International
- Moses, Leon N. 1958. "Location and the Theory of Production," *Quarterly Journal of Economics*, 78, 259-272.
- Moshe, Ben-Akiva and John L. Bowman. 1998. "Integration of an Activity-Based Model System and a Residential Location Model," *Urban Studies*, 35, 1131-1153.
- Metropolitan Transportation Commission. MTC. 2006. "Characteristics of Rail and Ferry Station Area Residents in the San Francisco Bay Area: Evidence from the 2000 Bay Area Travel Survey," San Francisco, CA: Planning Section, Metropolitan Transportation Commission

- Muth, Richard F. 1969. *Cities and Housing; the Spatial Pattern of Urban Residential Land Use*. Chicago,: University of Chicago Press
- Nelson, Peter, Andrew Baglino, Winston Harrington, Elena Safirova and Abram Lipman. 2007. "Transit in Washington, D.C.: Current Benefits and Optimal Level of Provision," *Journal of Urban Economics*, 62, 231-251.
- Noland, Robert B. and John V. Thomas. 2007. "Multivariate Analysis of Trip-Chaining Behavior," *Environment and Planning B: Planning and Design*, 34, 953--970.
- Federal Transit Administration. Parsons Brinckerhoff Quade & Douglas, Inc. 1996a. "Transit, Urban Form, Volume I," *Transit and Urban Form*, Report 16. Transit Cooperative Research Program, Federal Transit Administration
- Federal Transit Administration, *Transit Cooperative Research Program*. ——. 1996b. "Transit, Urban Form, Volume II," *Transit and Urban Form*, Report 16. Federal Transit Administration, *Transit Cooperative Research Program*
- Pass, Eric I. 1985. "State of the Art and Research Opportunities in Travel Demand: Another Perspective," *Transportation Research Part A*, 19, 460--464.
- Peng, Zhong-Ren, Kenneth J. Duecker, James Strathman and Janet Hopper. 1997. "A Simultaneous Route-Level Transit Patronage Model," *Transportation* 24, 159-181.
- Pinjari, Abdul Rawoof, Ram M. Pendyala, Chandra R. Bhat and Paul A. Waddel. 2007. "Modeling Residential Sorting Effects to Understand the Impact of the Built Environment on Commute Mode Choice," *Transportation*, 34, 557-573.
- Pushkarev, Boris S. and Jeffrey M. Zupan. 1982. *Urban Rail in America: An Exploration of Criteria for Fixed-Guideway Transit*. Ontario, Canada: Fitzhenry & Whiteside
- Pushkarev, Boris, Jeffrey M. Zupan and Regional Plan Association (New York N.Y.). 1977. *Public Transportation and Land Use Policy*. Bloomington: Indiana University Press
- Rai, R.K., Michael Balmer, Marcel Rieser, V.S. Vaze, Stefan Schonfelder and Kay W. Axhausen. 2007. "Capturing Human Activity Spaces," *Transportation Research Record*, 2021, 70-80.
- Rajamani, Jayanthi, Chandra R. Bhat, Susan Handy, Gerritt Knaap and Yan Song. 2003. "Assessing Impact of Urban Form Measures on Nonwork Trip Mode Choice after Controlling for Demographic and Level-of-Service Effects " *Transportation Research Record*, 1831,
- Reilly, Michael and John Landis. 2002. "The Influence of Built-Form and Land Use on Mode Choice - Evidence from the 1996 Bay Area Travel Survey," Institute of Urban and Regional Development
- Sargan, J. 1958. "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393-415.
- Savage, Ian. 2003. *Economic History of Mass Transit*. Oxford University Press
- Schimek, Paul. 1996. "Household Motorvehicle Ownership and Use: How Much Does Residential Density Matter?," *Transportation Research Record*, 120-125.
- Shiftman, Yoram. 1998. "Practical Approach to Trip Model Chaining " *Transportation Research Record*, 1645, 17--23.
- U.S. Department of Transportation. Strathman, James G. 1995. "Understanding Trip Chaining, Special Report on Trip and Vehicle Attributes," *1990 NPTS Report Series*, Washington, D.C.: U.S. Department of Transportation



- Taylor, Brian D. and Douglas Miller. 2003. "Analyzing the Determinants of Transit Ridership Using a Two-Stage Least Square Regression on a National Sample of Urbanized Areas," 2004 Transportation Research Board 83rd Annual Meeting, Washington, D.C. .
- Vance, Colin and Ralf Hedel. 2007. "The Impact of Urban Form on Automobile Travel: Disentangling Causation from Correlation," *Transportation*, 34, 575-588.
- Voith, Richard. 1997. "Fares, Service Levels, and Demographics: What Determines Commute Rail Ridership in the Long Run?," *Journal of Urban Economics*, 41, 176-197.
- . 1991. "The Long-Run Elasticity of Demand for Commuter Rail Transportation," *Journal of Urban Economics*, 30, 360-372.
- Wallace, Brett, Jennifer Barnes and G. Scott Rutherford. 2000. "Evaluating the Effects of Traveler and Trip Characteristics on Trip Chaining, with Implications for Transportation Demand Management Strategies," *Transportation Research Record*, 1718, 97--106.
- Wooldridge, Jeffrey M. 2002. *Introductory Econometrics: A Modern Approach*. Mason, OH: Thomson South-Western
- . 2002. *Introductory Econometrics: A Modern Approach*, 2e. Mason: South-Western
- Wu, D.M. 1974. "Alternative Tests of Independence between Stochastic Regressors and Disturbances: Finite Sample Results," *Econometrica*, 42, 529-546.
- Zheng, Siqi and Matthew E. Kahn. 2008. "Land and Residential Property Markets in a Booming Economy: New Evidence from Beijing," *Journal of Urban Economics*, 63, 743-757.
- Transportation Research Board National Research Council. Zupan, Jeffrey, Robert Cervero, Inc Parsons Brinckerhoff Quade & Douglas and Inc Howard/Stein-Hudson Associates. 1996. "Commuter and Light Rail Transit Corridors: The Land Use Connection," *TCRP Project H-1 Transit and Urban Form* TCRP 16. Transit Cooperative Research Program, Transportation Research Board National Research Council

*[This page intentionally left blank.]*

# Appendix A Table of Relevant Studies

Study	Data and Methodology	Dependent Variable (s)	Explanatory Variables	Significant Relationships	Issues/Weaknesses
Schimek (1996)	Disaggregate: Travel diary data from the 1990 NPTS; 15,916 households  2SLS with gross population density regressed against household head race and a dummy variable indicating residential location within large standard metropolitan statistical areas	Household vehicle ownership Household vehicle distance traveled (VMT)	Household size; Workers in household; Household size; Household income; Transit stop within 3 blocks of residence; household members' age; residential location at city center	10 % increase in density leads a 0.7 % reduction in household automobile travel	Choice of weak instruments that lead to unreliable estimates of parameters of interest
Boarnet and Sarmiento (1998)	Disaggregate: Travel diary data from the 1990-1994 Panel Study of Southern California Commuters  OLS regression of a reduce form equation for travel demand (Base Model) IV regression with residential location as endogenous. Instruments considered are race at the block group level (proportions), housing stock (proportion of houses by age cohort)	Trip frequencies Person miles of travel	Socio-demographic (age, race, income, education) Retail density(Retail employment/census tract land area); Service density (service employment/census tract land area) Employment Population (Total Employment/Total Population) Population density (block group) Percent of street grid within ¼ mile radius of residence	Land-use variables are endogenous to residential location choice;  The influence of land-use variables is weak and do not have expected signs Persons living in more mixed areas take more non-work trips  Percent of street grid is endogenous and instruments pass standard tests	Choice of weak instruments that lead to unreliable estimates of parameters of interest

Study	Data and Methodology	Dependent Variable (s)	Explanatory Variables	Significant Relationships	Issues/Weaknesses
Kuby et al. (2004)	Aggregate data: weekday boardings for 268 stations in 9 cities  OLS regression	Average weekday boardings	Employment; population; airport (dummy); international border (dummy); college enrollments; accessibility to central city; CBD (dummy); park-and-ride spaces; bus connections; other rail lines;	An increase of 100 persons employed within walking distance to a station increase boarding by 2.3 passengers;  An increase of 100 persons residing within walking distance to a station is associated with an increase of 9.2 boardings;  CBD no longer relevant	Simple linear model with no recognition of simultaneity or endogeneity between urban form and transit patronage
Reilly and Landis (2002)	Disaggregate data: travel diary data from BATS96  Multinomial discrete choice model	Mode choice	Raster-size measures of land-use and urban design: land-use diversity, intersection density, and average lot size	An increase in average density of 10 persons per hectare (about four persons per acre) within a one mile of an individual's residence is associated with a 7 percent increase in the probability of walking or taking transit	No attempt to uncover any causality between urban form and travel behavior

Study	Data and Methodology	Dependent Variable (s)	Explanatory Variables	Significant Relationships	Issues/Weaknesses
Messenger and Ewing (1996)	Aggregate: 698 traffic analysis zones of Metro-Dade County, FL  Full information maximum likelihood estimation (FIML)	Bus share of work trips (home zones)  Bus share of work trips (work zones)  Proportion of households with 0 to 1 vehicles	Socio-demographic; Overall density measures	Bus share of work trips increases with density (indirectly through the effects of density on ownership and parking fees)  8.4 dwellings/acre are needed support 25-min headways at the transit operator's minimum productivity level and 19.4 dwellings/acre at the systemwide average productivity	Results' validity severely hinges on correct system of equation specification due to the use of chosen estimation method. Misspecification of one equation "contaminates" all estimates in the system.  Ad-hoc land-use variables, such as the natural log of overall density measures as total population + total employment/land area (miles <sup>2</sup> )
McNally and Kulkarni (1997)	Aggregate: land-use and travel data from 20 neighborhoods in Orange County, California  Analysis of variance (comparisons of means)  OLS regression	Household trip generation	Three neighborhoods types: traditional, planned, and mixed	Neighborhood type not statistically significant ; relationship between urban form and land-use is weak	

Study	Data and Methodology	Dependent Variable (s)	Explanatory Variables	Significant Relationships	Issues/Weaknesses
Rajamani et al. (2003)	Disaggregate: 1995 Portland Metropolitan Activity Survey  Multinomial logit mode choice	Mode choice for non-work trips	Household and individual socio-demographic; gravity-based accessibility index; Land-use diversity index; population density;	Mixed uses promote walking behavior for non-work activities The coefficient on the land-use mix diversity index specific to transit is not significant, indicating that the impacts of transit-oriented development on transit ridership may be limited  Denser neighborhoods decrease the likelihood of driving alone and increase the likelihood of transit use	Relatively simple model that does not take into account unobserved idiosyncratic preferences for location. Parameter estimates are likely to be biased. Reported elasticities are not easy to interpret and not suitable for generalization.
Levinson and Kumar (1997)	Disaggregate: 1990-1991 NPTS  OLS regression	Trip distance and speed by mode	Socio-demographic; population density of the residential zip code from the NPTS data set; metropolitan size, urbanized and metropolitan population density in 1980 and 1990 ; number of edge cities, representing the degree of polycentricity.	Declining travel times by transit and increasing travel times by auto as density rises above 10,000 persons/mile <sup>2</sup> result in higher transit mode shares  Density is relatively less relevant than socio-demographic factors	Ad-hoc density cut offs that defined two variables that describe density levels likely to affect transit patronage: low density ( $\geq 10,000$ persons/mile <sup>2</sup> ; high density (0 if $< 10,000$ , 1 if $\geq 10,000$ persons/mile <sup>2</sup>  No endogeneity between density speed and travel time is accounted for.

Study	Data and Methodology	Dependent Variable (s)	Explanatory Variables	Significant Relationships	Issues/Weaknesses
Handy, Cao, and Mokhtarian (2005)	Self-administered mailed survey of eight neighborhoods in northern California  Analysis of variance multiple regression, and ordered probit	Vehicle miles driven (VMD); Likelihood of changing driving habits	Neighborhood characteristics (safety, spaciousness, attractiveness), measured on a 5-point scale	VMD are higher in suburban vs. traditional neighborhoods A move to a more accessible, walking friendly neighborhood is associated with less auto travel	Models have poor explanatory power. Variable selection obtained using stepwise regression, with a high probability of a false null hypothesis caused by multicollinearity issues between the built environment and self-selecting variables being employed, thus leading to weak conclusions.
Cao, Mokhtarian, and Handy (2007)	Self-administered mailed survey of eight neighborhoods in northern California  Simultaneous equation system or structure equation modeling (SEM)	Changes in driving, accessibility, auto ownership, spaciousness.	Socio-demographic, neighborhood characteristics, travel attitudes, residential preferences.	Residential self-selection has significant direct and indirect impacts on travel behavior. Changes in the built environment affect travel behavior.	Authors claim a “quantum improvement in terms of modeling”, but, there is no explicit behavioral model, only a general matrix notation of a simultaneous equation system. The study is not longitudinal, contrary to what claimed.
Bagley and Mokhtarian (2002)	Self-administered mailed survey of eight neighborhoods in northern California  Structure equation modeling (SEM)	Residential location (two variables) ; Attitudes (three variables); Travel Demand (three variables); Job location (commute distance)	Socio-demographic; attitudinal	When attitudinal, lifestyle, and socio-demographic variables are accounted for, neighborhood type has little influence on travel behavior	There are two endogenous variables for residential location; three endogenous variables for attitudes; and three endogenous variables for travel demand. The model is not behavioral at best.
Greenwald and Boarnet (2001)	Disaggregate: 1994 travel diary data from Portland, Oregon  OLS and IV regression	Non-work walking trips	Socio-demographic; neighborhood land-use (population, retail density at block group level); regional land-use (population retail density at zip code level); trip cost variables; instrumental variables (race, housing stock)	Impacts of land-use take place at the neighborhood level which suggests adopting a proper geographic scale.	Same as in Boarnet and Crane and Boarnet and Sarmiento

Study	Data and Methodology	Dependent Variable (s)	Explanatory Variables	Significant Relationships	Issues/Weaknesses
Zegras (2004)	Disaggregate: 1991 origin-destination travel survey of Santiago, Chile  Ordered Probit regression	Non-work and non-school trips	Socio-demographic; land use (population density, amount of vacant land area per hectare)	An increased share of commercial and service uses increases the likelihood of making trips;  Population density does not have a significant effect	Poor explanatory power; no IV regression (such as the use of IV Probit)
Estupiñán and Rodríguez (2008)	Aggregate: route level station boardings, Bogotá, Colombia  SEM of factors scores using GMM estimator	Transit boardings (transit demand) and number of vehicles per day (transit supply)	Station attributes (physical and perceived characteristics); Neighborhood characteristics (crime, education levels, GINI coefficient, unemployment) Factor analysis is employed to group the many variables in a set of factors scores for subsequent 2SLS regression	Built environments (walking barriers) affects travel behavior;  High density and land-use mix are supportive of BRT.	Essentially an extension of the model developed by Peng et al. (1997), with the inclusion of built environment variables.  Comprehensive set of neighborhood attributes that control for idiosyncratic preferences for location  The major limitation is the exogenous treatment of density around stations.
Vance and Hedel (2007)	Disaggregate: household travel panel data, German Mobility Panel (2006)  Two-part model (2PM) consisting of OLS and Probit estimation (similar to IV regression)	Vehicle ownership (first stage)  Distance traveled (second stage)	Socio-demographic; commercial density, street density, commercial diversity.	Urban form statistically significant determinant of auto travel;  Zip code level land-use variable might not be appropriate spatial scale	The use of instruments originally presented by Boarnet and Sarmiento (1998), although properly tested, is measured at the much broader zip code level. There are no instruments that control for neighborhood characteristics affecting residential location preferences.



## Appendix B Comparative Static Analysis

The basic theoretical implications of Model I can be explored in advance of empirical testing by employing comparative-static analysis<sup>10</sup>. The impact of changes in exogenous density,  $\bar{D}$ , and exogenous residential location,  $\bar{RL}$ , on travel demand,  $TD$ , is considered. Basically, starting from an equilibrium state, the impact of an increase in density and residential location on the initial equilibrium is considered. The objective is to see what happens to transit demand as density levels change. This appendix details the most relevant comparative-static results, namely the impact of changes in density,  $D$ , and residential location,  $RL$ .

To conduct comparative-static analysis, a set of basic assumptions related to trip chaining behavior, activity space, and urban form must first be introduced. Also, although trips are integers in reality, they are herein treated as a continuous non-negative variable for analytical purposes.

### *Assumption A.1*

Residential location is defined as the optimal *job-residence pair* in an urban area in which jobs and residences are dispersed. Following urban residential location theory, the location decision is assumed to be the result of a trade-off between housing expenditures and transportation costs, given income and the mode-choice set. Following Anas and his associates (Anas and Kim, 1996, Anas and Xu, 1999), the location decision is also based on idiosyncratic preferences for location and travel. As the distance between these two locations increases, the need to engage in trip-chaining also increases. Trip chaining, as shown in Anas (2007), allows saving time, which, in turn can be allocated either to a farther move away from work (more commute time), to be spent as leisure time, or to be used for more non-work travel. As the distance defining the *job-residence pair* increases, then the need to chain non-work trips increases

$$TC_{RL} = \frac{\partial TC}{\partial RL} > 0 \quad (\text{a.1})$$

But this happens at a decreasing rate

$$\frac{\partial^2 TC}{\partial RL^2} < 0 \quad (\text{a.2})$$

<sup>10</sup> Comparative-static analysis is a tool commonly used in mathematical economics and microeconomic theory. It allows comparing different equilibrium states associated with different sets of values of parameters and exogenous variables. Comparative-statics can be either qualitative or quantitative in nature. In this case, it allows conducting a qualitative assessment as it permits to focus on the direction of change, rather than its magnitude, of changes in location and density. (Chiang, Alpha C. 1984. *Fundamental Methods of Mathematical Economics*. New York: McGraw-Hill, Inc. )

*Assumption A.2*

If density,  $\bar{D}$ , increases, then non-work activity locations, such as shopping or recreational locations, tend to be more clustered together, thus reducing the household activity space

$$AS_D = \frac{\partial AS}{\partial D} < 0 \quad (a.3)$$

*Assumption A.3*

If the household activity space gets more dispersed (its size increases) then trip chaining increases:

$$TC_{AS} = \frac{\partial TC}{\partial AS} > 0 \quad (a.4)$$

with

$$\frac{\partial^2 TC}{\partial AS^2} < 0 \quad (a.5)$$

This is reciprocal to A.1. As the household activity space grows or gets more dispersed the need to engage in trip chaining increases. Empirical evidence that confirms this hypothesis is found in Thomas and Noland (2006) who, in a multivariate analysis of trip chaining behavior, find a positive relationship between lower densities and the complexity of trip chaining behavior. Noland and Thomas found that density environments lead to both a greater reliance upon trip chaining and tours that involve more stops.

*Assumption A.4*

As trip chaining increases, the activity space increases:

$$AS_{TC} = \frac{\partial AS}{\partial TC} > 0 \quad (a.6)$$

This assumption means that factors that directly affect the trip chaining function,  $TC$ , result in feedback effects on activity space,  $AS$ . These feedback effects are less intense and marginally decreasing

$$\frac{\partial^2 AS}{\partial TC^2} < 0 \quad (a.7)$$

### Model I Comparative-static Results

Now, consider Model I. Equations (3.1), (3.2), and (3.3) can be written as specific functions in the form  $F^j(TC, AS, RL, D, X_{TC}, X_{AS}, X_{TT})$ . With continuous partial derivatives and with the relevant assumptions A.3 and A.4, the following nonzero *Jacobian determinant*<sup>11</sup> is obtained

$$|J| = \begin{vmatrix} \frac{\partial F^1}{\partial TC} & \frac{\partial F^1}{\partial AS} & \frac{\partial F^1}{\partial TD} \\ \frac{\partial F^2}{\partial TC} & \frac{\partial F^2}{\partial AS} & \frac{\partial F^2}{\partial TD} \\ \frac{\partial F^3}{\partial TC} & \frac{\partial F^3}{\partial AS} & \frac{\partial F^3}{\partial TD} \end{vmatrix} = \begin{vmatrix} 1 & -TC_{AS} & 0 \\ -AS_{TC} & 1 & 0 \\ -TD_{TC} & -TD_{AS} & 1 \end{vmatrix} = 1 - \overbrace{AS_{TC}}^{(+)} \overbrace{TC_{AS}}^{(+)} \neq 0 \quad (a.8)$$

Therefore,  $TC$ , and  $AS$  can be considered implicit functions of  $(RL, X_{TC}, D, X_{AS})$  at and around any point that satisfies Equations (3.1), (3.2), and (3.3), which would then be an equilibrium solution,  $\overline{TC}$ ,  $\overline{AS}$ , and  $\overline{TD}$ . Hence the implicit function theorem justifies writing

$$\overline{TC} = f^1(RL, D, WD, X_{TC}, X_{AS}, X_{TD}) \quad (a.9)$$

$$\overline{HAS} = f^2(RL, D, WD, X_{TC}, X_{AS}, X_{TD}) \quad (a.10)$$

$$\overline{TD} = f^3(RL, D, WD, X_{TC}, X_{AS}, X_{TD}) \quad (a.11)$$

indicating that the equilibrium values of the endogenous variables are implicit functions of the exogenous variables and parameters. The partial derivatives of the implicit functions are in the nature of comparative-static derivatives. To find these, the partial derivative of the  $F$  functions, evaluated at the equilibrium state of the model, are needed.

Next, the comparative-static analysis is conducted to ascertain the effect of changes brought about by changes in density, residential location and transit station proximity (i.e., changes in walking distance)

### Effects of an Increase in Density, $\overline{D}$

The general form for the comparative-static analysis of Model I is given by

<sup>11</sup> The *Jacobian determinant* (or a *Jacobian*, for short), is a determinant of a matrix of partial derivatives, which tests functional dependence among a set of functions. Given the equation system, partial derivatives needed for comparative-static analysis (see previous footnote) can be obtained if the Jacobian,  $J$ , is non-zero.

$$\begin{vmatrix} 1 & -TC_{AS} & 0 \\ -AS_{TC} & 1 & 0 \\ -TD_{TC} & -TD_{AS} & 1 \end{vmatrix} \begin{bmatrix} \partial \overline{TC} / \partial D \\ \partial \overline{AS} / \partial D \\ \partial \overline{TD} / \partial D \end{bmatrix} = \begin{bmatrix} 0 \\ AS_D \\ 0 \end{bmatrix} \quad (\text{a.12})$$

### Density Effect on Trip Chaining

First, the effect of increased density on trip chaining is considered. By applying Cramer's rule,<sup>12</sup> the total partial derivative is computed as

$$d\overline{TC}/dD = \frac{\begin{vmatrix} 0 & -TC_{AS} & 0 \\ AS_D & 1 & 0 \\ 0 & -TD_{AS} & 1 \end{vmatrix}}{|J|} = \frac{\overbrace{TC_{AS}AS_D}^{(-)}}{\underbrace{1 - \underbrace{AS_{TC}TC_{AS}}_{(+)}}_{(+)}} < 0 \quad (\text{a.13})$$

The results of this comparative-static show that an increase in density causes a clustering of activities which contracts the activity space, which, in turn, reduces the need to engage in trip chaining. The total reduction in trip chaining also accounts for the feedbacks into Equation (3.1) coming from Equation (3.2) by way of  $TC_{AS}$ . This outcome has been confirmed in the literature on trip chaining behavior, which shows that lower density environments increase the need to engage in trip chaining (Noland and Thomas, 2007, Wallace, et al., 2000).

### Density Effect on Activity Space

The effect of an increase in  $D$  on trip chaining is obtained in the same manner

$$d\overline{AS}/dD = \frac{\begin{vmatrix} 1 & 0 & 0 \\ -AS_{TC} & AS_D & 0 \\ -TD_{TC} & 0 & 1 \end{vmatrix}}{|J|} = \frac{\overbrace{AS_D}^{(-)}}{\underbrace{1}_{(+)}} < 0 \quad (\text{a.14})$$

Note, by assumption A.1, we have that  $\partial \overline{AS} / \partial D < 0$ . Therefore, an increase in density contracts the activity space both directly and indirectly through feedback effect coming by way of the Equation (3.1) ( $AS_{TC}TC_{AS}$ ).

<sup>12</sup> Cramer's rule is a method of matrix inversion that enables a convenient, practical way of solving a linear-equation system.

### Effect on Transit Demand

The effect of an increase in density on transit demand, the most relevant comparative-static analysis in the context of this study, is obtained as

$$\frac{d\overline{TD}}{d\overline{D}} = \frac{\begin{vmatrix} 1 & -TC_{AS} & 0 \\ -AS_{TC} & 1 & AS_D \\ -TD_{TC} & -TD_{AS} & 0 \end{vmatrix}}{|J|} = \frac{\overbrace{\alpha}^{(+)} \overbrace{\beta}^{(-)}}{\overbrace{TD_{AS}AS_D}^{(-)(-)} + \overbrace{TD_{TC}TC_{AS}AS_D}^{(+)(+)(-)}} \geq 0 \quad (\text{a.14})$$

where

$\alpha$  = change in transit demand caused by a contraction in activity space as a result of increased density

$\beta$  = change in transit demand caused by decreasing trip chaining as a result of increased density

The result shows an ambiguous effect of density on transit demand (as measured in total linked trips per household). Indeed, for  $\frac{d\overline{TD}}{d\overline{D}} > 0$  it must be that  $\alpha > -\beta$ . In other words for transit demand to be positively related to density it must be that the increase in transit demand caused by a contraction in activity space (as a result of increased density,  $\alpha > 0$ ) is greater than the reduction in transit demand caused by reduced trip chaining (as a result of increased density,  $\beta < 0$ ).

This explanation is inherent to the determinants of trip chaining behavior. In higher density environments, as the spatial extent of non-work activities reduces, trip chaining needs decrease but individual trips increase and individuals prefer to make non-chained trips. First, increased density reduces the extent of the activity space, which directly increases the demand for non-chained transit trips. Second, higher densities reduce the activity space, which reduces the need to chain trips (as time savings opportunities decrease) and thus the demand for transit trips. Thus an increase transit trips occurs if transit demand is more sensitive to changes affecting the spatial allocation of non-work activities than to changes affecting trip chaining behavior. In other words, the above comparative-statics result shows that the increase in density exerts two opposite effects on transit demand. This result relies on two additional assumptions, namely

$$\frac{\partial TD}{\partial TC} > 0 \quad (\text{a.15})$$

The demand for transit trips increases as trip chaining increases. An increase in the number of chained trips overall increases the demand for transit (or any other mode). This is brought about by the initial model specification which presents a transit trip demand as a function of trip chaining.

$$\frac{\partial TD}{\partial AS} < 0 \quad (\text{a.16})$$

This means that increased spatial dispersion of non-work activities cannot be accommodated by additional transit trips. Given the characteristics of transit service supply (being fixed at least in the short to medium run), increased spatial dispersion is accommodated by substituting transit travel with other, more flexible, modes, such as auto travel. The latter is more flexible in terms of allowing serving a more dispersed activity space. This assumption confirms that transit and auto are non-perfect substitutes.

### Living Farther from a Transit Station (change in walking distance, $WD$ )

Next, the comparative-statics of an increase in walking distance,  $WD$ , are derived. Next, the effect of an increase in distance from the nearest transit station is considered. The empirical literature provides unequivocal evidence of a negative relationship between distance to transit stops and the demand for transit services (Cervero, 2007, Cervero and Kockelman, 1997). The debate is mostly centered on the magnitude of this relationship, as highlighted by the growing body of literature on residential self-selection. All else equal, being located farther away from a transit station results in reduced transit demand

$$\frac{d\overline{TD}}{dWD} = \frac{\begin{vmatrix} 1 & -TC_{AS} & TC_{WD} \\ -AS_{TC} & 1 & 0 \\ -TD_{TC} & -TD_{AS} & TD_{WD} \end{vmatrix}}{|J|} = \frac{\overbrace{TD_{WD}}^{(-)} + \overbrace{TD_{TC}TC_{WD}}^{(+)} + \overbrace{AS_{TC}}^{(-)} \left( \overbrace{TC_{WD}TD_{AS}}^{(-)} - \overbrace{TC_{AS}TD_{WD}}^{(+)} \right)}^{(-)}}{|J|} < 0 \quad (\text{a.17})$$

### A Move Farther Away from Work (change in residential location)

Next the comparative-statics of an increase in residential location,  $RL$ , are derived. Note that  $RL$  is considered as exogenous in Model I, to indicate short run equilibrium. The resulting comparative-static follows

$$\begin{vmatrix} 1 & -TC_{AS} & 0 \\ -AS_{TC} & 1 & 0 \\ -TD_{TC} & -TD_{AS} & 1 \end{vmatrix} \begin{bmatrix} \partial \overline{TC} / \partial RL \\ \partial \overline{AS} / \partial RL \\ \partial \overline{TD} / \partial RL \end{bmatrix} = \begin{bmatrix} TC_{RL} \\ 0 \\ TD_{RL} \end{bmatrix} \quad (\text{a.18})$$

### Effect on Trip Chaining, $TC$

First, the effect of a move farther away from work on trip chaining is considered. By Assumption A.1, this has a positive impact on trip chaining. The new equilibrium results in a

higher number of trips per chain. When testing this hypothesis empirically and using cross-sectional data, individuals with a more extended commute are expected to engage in a higher number of trips per chain (or in more complex tours characterized by more stops).

$$d\overline{TC}/dRL = \frac{\begin{vmatrix} TC_{RL} & -TC_{AS} & 0 \\ 0 & 1 & 0 \\ TD_{RL} & -TD_{AS} & 1 \end{vmatrix}}{|J|} = \frac{\overbrace{TC_{RL}}^{(+)}}{|J|} > 0 \quad (\text{a.19})$$

In a longitudinal context, a move farther out entails more distance and more time spent commuting, which increases the propensity to engage in trip chaining to save overall time.

### Effect on Activity Space, AS

The effect of an increase in  $RL$  on the activity space is given by

$$d\overline{AS}/dRL = \frac{\begin{vmatrix} 1 & TC_{RL} & 0 \\ -AS_{TC} & 0 & 0 \\ -TD_{TC} & TD_{RL} & 1 \end{vmatrix}}{|J|} = \frac{\overbrace{\overbrace{AS_{TC}TC_{RL}}^{(+)}^{(+)}}^{(+)}}{(+)} > 0 \quad (\text{a.20})$$

A move farther away from work increases trip chaining, which in turn increases the activity space. This increase is indirect as it comes by way of Equation (a.9). The empirical work will reveal information on its magnitude.

### Effect on Transit Demand, TD

The change in transit demand caused by a change in residential location is given by:

$$d\overline{TD}/dRL = \frac{\begin{vmatrix} 1 & -TC_{AS} & TC_{RL} \\ -AS_{TC} & 0 & 0 \\ -TD_{TC} & -TD_{AS} & TD_{RL} \end{vmatrix}}{|J|} = \frac{\overbrace{TD_{RL}}^{(\pm)} + \overbrace{AS_{TC}}^{(+)} \left( \overbrace{TC_{RL}TD_{AS}}^{(+)(-)} - \overbrace{TC_{AS}TD_{RL}}^{(+)(\pm)} \right) + \overbrace{TC_{RL}TD_{TC}}^{(+)(+)}}{(+)} \geq 0 \quad (\text{a.21})$$

The overall effect on transit demand hinges on the sign of  $TD_{RL}$ . To the extent that an urban area is well served by transit, then the relationship between transit demand and residential location is positive. A positive relationship is observed in older, more monocentric-type cities, with existing transit services supporting major work commute travel routes. On the other hand, if supply constraints exist, transit demand declines as the job-residence distance increases. Therefore, the overall effect on transit demand due to a change in location depends on both the sign and magnitude of  $TD_{RL}$  ( $TD_{RL} \leq 0$ ).

### Exogenous shift in Trip Chaining, $TC\phi$

This comparative-static allows computing the total derivative of transit demand with respect to exogenous changes directly affecting only the trip chaining equation. Let  $\phi$  be a change in an exogenous variable appearing only in the trip chaining equation, then a change in  $\phi$  has the following effect on transit demand

$$d\overline{TD}/dTC_\phi = \frac{\begin{vmatrix} 1 & -TC_{AS} & TC_{RL} \\ -AS_{TC} & 0 & 0 \\ -TD_{TC} & -TD_{AS} & TD_{RL} \end{vmatrix}}{|J|} = \frac{\overline{TC}_\phi \left( \overbrace{TD_{TC}}^{(+)} + \overbrace{AS_{TC}}^{(+)} \overbrace{TD_{AS}}^{(-)} \right)}{|J|} \quad (a.22)$$

This result, for example is used to assess the effect a change in distance to the nearest subcenter under the empirical specification of equation (4.5) in Chapter 4.

### Exogenous shift in Activity Space, $AS\phi$

This comparative-static allows computing the total derivative of transit demand with respect to exogenous changes directly affecting only the activity space equation. Let  $\phi$  be a change in an exogenous variable appearing only in the activity space equation, then a change in  $\phi$  has the following effect on transit demand

$$d\overline{TD}/dAS_\phi = \frac{\begin{vmatrix} 1 & -TC_{AS} & TC_{RL} \\ -AS_{TC} & 0 & 0 \\ -TD_{TC} & -TD_{AS} & TD_{RL} \end{vmatrix}}{|J|} = \frac{AS_\phi \left( \overbrace{TD_{AS}}^{(-)} + \overbrace{TC_{AS}}^{(+)} \overbrace{TD_{TC}}^{(+)} \right)}{|J|} \quad (a.23)$$

This result is used to evaluate the effect of a change in retail establishment density, as appearing in equation (4.6) on transit demand.

## Model II

In this model, the assumption of residential location is relaxed. Treated as a choice variable, residential location is the outcome of a trade-off between transportation and land use costs. Taking into account idiosyncratic preferences for location, households choose an optimal home-work commute pair, while at the same time optimizing goods consumption and the ensuing non-work travel behavior (optimal non-work trip chaining and activity space). The specification of the model is given by equations (3.8) to (3.11). The inclusion equation (3.11) requires deriving a new Jacobian determinant



$$|J| = \begin{vmatrix} 1 & -TC_{AS} & 0 & -TC_{RL} \\ -AS_{TC} & 1 & 0 & 0 \\ -TD_{TC} & -TD_{AS} & 1 & -TD_{RL} \\ -RL_{TC} & 0 & -RL_{TD} & 1 \end{vmatrix} =$$

$$1 - \underbrace{RL_{TC}}_{(+)} \underbrace{TC_{RL}}_{(+)} - \underbrace{RL_{TD}}_{\pm} \underbrace{TD_{RL}}_{\pm} + \underbrace{AS_{TC}}_{+} \left[ -\underbrace{RL_{TD}}_{\pm} \underbrace{TC_{RL}}_{+} \underbrace{TD_{AS}}_{-} + \underbrace{TC_{AS}}_{+} \left( -1 + \underbrace{RL_{TD}}_{\pm} \underbrace{TD_{RL}}_{\pm} \right) \right] - \underbrace{RL_{TD}}_{\pm} \underbrace{TC_{RL}}_{+} \underbrace{TD_{TC}}_{+} \quad (a.24)$$

Given the complexity introduced by adding equation (a.24) to the model, the following comparative-static analysis focuses only on impacts of changes affecting the demand for transit trips.

### Effect of an Increase in Density on Transit Demand, D

As shown by this result, the ultimate effect on transit demand when exogenous density levels change is relatively larger than model I. Using Cramer's rule, the change is computed as

$$\frac{d\overline{TD}}{d\overline{D}} = \frac{\begin{vmatrix} 1 & -TC_{AS} & 0 & -TC_{RL} \\ -AS_{TC} & 1 & AS_D & 0 \\ -TD_{TC} & -TD_{AS} & 0 & -TD_{RL} \\ -RL_{TC} & 0 & -RL_{TD} & 1 \end{vmatrix}}{|J|} = \frac{\overline{AS_D} \left[ \left( 1 - \underbrace{RL_{TC}}_{(+)} \underbrace{TC_{RL}}_{(+)} \right) \underbrace{TD_{AS}}_{(-)} + \underbrace{TC_{AS}}_{(+)} \left( \underbrace{RL_{TC}}_{(+)} \underbrace{TD_{RL}}_{(\pm)} + \underbrace{TD_{TC}}_{(+)} \right) \right]}{|J|} \geq 0 \quad (a.25)$$

In the long run, the spatial extent of non-work activities, trip chaining and residential location are all jointly determined. Exogenous shifts in density levels affect this decision making process. An increase in density directly impacts the spatial extent of non-work activity locations in terms of an increased activity space, AS (i.e., activities are more disperse across the urban landscape). This increase affects trip chaining with feedback effects both on the demand for transit trips and residential location patterns in a looping fashion. The effect of density on transit demand is the same as the one assessed under Model I specification.

### Effect of an Increase in Walking Distance, WD

A change in density affects both transit station and activity space directly and indirectly, as specified by

$$\frac{dT\bar{D}}{d\bar{W}\bar{D}} = \frac{\begin{vmatrix} 1 & -TC_{AS} & TC_{WD} & -TC_{RL} \\ -AS_{TC} & 1 & 0 & 0 \\ -TD_{TC} & -TD_{AS} & TD_{WD} & -TD_{RL} \\ -RL_{TC} & 0 & 0 & 1 \end{vmatrix}}{|J|} = \frac{\overbrace{TD_{WD}+TD_{TC}TC_{WD}+RL_{TC}}^{(-)} \left( \overbrace{TC_{WD}TD_{RL}}^{(\pm)} - \overbrace{TC_{RL}TD_{WD}}^{(-)} \right) + \overbrace{AS_{TC}}^{(+)} \left( \overbrace{TC_{WD}TD_{AS}}^{(-)} - \overbrace{TC_{AS}TD_{WD}}^{(-)} \right)}{|J|} \quad (a.26)$$

The direct effect of density on transit proximity is due to two separate causes. First, higher densities improve transit proximity by reducing average walking distance to the nearest station. This can be identified as a supply side effect, in that more stations are likely to be located at higher densities. Second, at any given home-work commute pair arrangement individuals are more likely to utilize transit services to engage in both work and non-work activities. A change in density also indirectly affects transit station proximity, given its treatment as a choice variable. An increase in density results in more accessible non-work activities which reduce the need to engage in trip chaining, thus decreasing transit patronage. This in turn reduces the need to be located closely to a transit station.

### Exogenous shift in Trip Chaining, $TC_\phi$

En exogenous shift,  $\phi$ , has the following impact on the demand for transit

$$\frac{dT\bar{D}}{dTC_\phi} = \frac{\begin{vmatrix} 1 & -TC_{AS} & TC_\phi & -TC_{RL} \\ -AS_{TC} & 1 & 0 & 0 \\ -TD_{TC} & -TD_{AS} & 0 & -TD_{RL} \\ -RL_{TC} & 0 & 0 & 1 \end{vmatrix}}{|J|} = \frac{\overbrace{TC_\phi}^{(\pm)} \left( \overbrace{AS_{TC}TD_{AS}}^{(+)} + \overbrace{RL_{TC}TD_{RL}}^{(+)} + \overbrace{TD_{TC}}^{(\pm)} \right)}{|J|} \quad (a.27)$$

### Exogenous shift in Activity Space, $AS_\phi$

An exogenous change,  $\phi$ , affecting  $AS$ , has the following effect on transit demand

$$\frac{dT\bar{D}}{dAS_\phi} = \frac{\begin{vmatrix} 1 & -TC_{AS} & 0 & -TC_{RL} \\ -AS_{TC} & 1 & AS_\phi & 0 \\ -TD_{TC} & -TD_{AS} & 0 & -TD_{RL} \\ -RL_{TC} & 0 & 0 & 1 \end{vmatrix}}{|J|} = \frac{\overbrace{AS_\phi}^{(\pm)} \left[ \left( \overbrace{1-RL_{TC}TC_{RL}}^{(+)} \right) \overbrace{TD_{AS}}^{(-)} + \overbrace{TC_{AS}}^{(+)} \left( \overbrace{RL_{TC}TD_{RL}}^{(+)} + \overbrace{TD_{TC}}^{(\pm)} \right) \right]}{|J|} \quad (a.28)$$

### Exogenous shift in Residential Location, $RL_\phi$

An exogenous change,  $\phi$ , affecting residential location decisions, has the following impact on transit demand

$$\frac{d\overline{TD}}{dRL_\varphi} = \frac{\begin{vmatrix} 1 & -TC_{AS} & 0 & -TC_{RL} \\ -AS_{TC} & 1 & 0 & 0 \\ -TD_{TC} & -TD_{AS} & 0 & -TD_{RL} \\ -RL_{TC} & 0 & RL_\varphi & 1 \end{vmatrix}}{|J|} = \frac{\overbrace{RL_\varphi}^{(\pm)} \left[ \overbrace{TD_{RL}}^{(\pm)} + \overbrace{AS_{TC}}^{(+)} \left( \overbrace{TC_{RL}}^{(+)} \overbrace{TD_{AS}}^{(-)} - \overbrace{TC_{AS}}^{(+)} \overbrace{TD_{RL}}^{(\pm)} \right) + \overbrace{TC_{RL}}^{(+)} \overbrace{TD_{TC}}^{(+)} \right]}{|J|}}{\quad} \quad (a.29)$$

### Model III

In this model, the assumption of density as being exogenous to the model is relaxed. At any given home-commute pair arrangement, the decision to locate in proximity of a station is dependent upon transit patronage levels and factors related to density. Density in proximity to transit stations is affected by patronage levels. This relationship is described by equations (3.13) to (3.15). The computation of the Jacobian is further complicated by the addition of the density equation and is equal to

$$J = \begin{vmatrix} 1 & -TC_{AS} & 0 & -TC_{RL} & 0 \\ -AS_{TC} & 1 & 0 & 0 & -AS_D \\ -TD_{TC} & -TD_{AS} & 1 & -TD_{RL} & 0 \\ -RL_{TC} & 0 & -RL_{TD} & 1 & 0 \\ 0 & -D_{AS} & 0 & -D_{RL} & 1 \end{vmatrix} = \quad (a.30)$$

$$\begin{aligned} |J| = & 1 - \overbrace{RL_{TC}}^{(+)} \overbrace{TC_{RL}}^{(+)} - \overbrace{RL_{TD}}^{(\pm)} \overbrace{TD_{RL}}^{(\pm)} - \overbrace{AS_{TC}}^{(+)} \left\{ \overbrace{RL_{TD}}^{(\pm)} \overbrace{TC_{RL}}^{(+)} \overbrace{TD_{AS}}^{(-)} + \overbrace{TC_{AS}}^{(+)} \left( 1 - \overbrace{RL_{TD}}^{(\pm)} \overbrace{TD_{RL}}^{(\pm)} \right) \right\} - \\ & \overbrace{RL_{TD}}^{(\pm)} \overbrace{TC_{RL}}^{(+)} \overbrace{TD_{TC}}^{(+)} + \overbrace{AS_D}^{(-)} \left[ -\overbrace{D_{RL}}^{(-)} \left( \overbrace{RL_{TC}}^{(+)} \overbrace{TC_{AS}}^{(+)} + \overbrace{RL_{TD}}^{(\pm)} \left( \overbrace{TD_{AS}}^{(-)} + \overbrace{TC_{AS}}^{(+)} \overbrace{TD_{TC}}^{(+)} \right) \right) \right] + \overbrace{D_{AS}}^{(-)} \left[ -1 + \right. \\ & \left. \overbrace{RL_{TC}}^{(+)} \overbrace{TC_{RL}}^{(+)} + \overbrace{RL_{TD}}^{(\pm)} \left( \overbrace{TD_{RL}}^{(\pm)} + \overbrace{TC_{RL}}^{(+)} \overbrace{TD_{TC}}^{(+)} \right) \right] \right\} \neq 0 \end{aligned}$$

This model best describes long term equilibrium, where both location and travel decisions are optimized under constraint. Ideally, empirical testing of this model would rely on disaggregate travel diary data in the form of a panel that collects behavior of a same set of individuals across time. When dealing with observational data across different individuals at a point in time (i.e., a cross-sectional dataset), changes in behavior can be studied by controlling for individual heterogeneity.

The comparative-static analysis focuses on changes affecting the demand for transit trips.

### Effect of Exogenous Shift in Density

Given the endogenous treatment of density, this model can be used to test the effects of policies geared at directly affecting density, such as policies that are intended to increase density around transit stations. Assuming an exogenous shock,  $\theta$ , positively affecting density the following comparative-statics is obtained

$$\frac{d\bar{D}}{dD_\phi} = \frac{\begin{vmatrix} 1 & -TC_{AS} & 0 & -TC_{RL} & 0 \\ -AS_{TC} & 1 & 0 & 0 & -AS_D \\ -TD_{TC} & -TD_{AS} & 0 & -TD_{RL} & 0 \\ -RL_{TC} & 0 & 0 & 1 & 0 \\ 0 & -D_{AS} & D_\phi & -D_{RL} & 1 \end{vmatrix}}{|J|} = \frac{AS_D D_\phi [(1 - RL_{TC} TC_{RL}) TD_{AS} + TC_{AS} (RL_{TC} TD_{RL} + TD_{TC})]}{|J|} \geq 0 \quad (a.31)$$

### Effect of an Increase in Walking Distance, $WD$

An increase in walking distance causes the following change in transit demand

$$\begin{aligned} \frac{d\bar{D}}{dWD} &= \frac{\begin{vmatrix} 1 & -TC_{AS} & TC_{WD} & -TC_{RL} & 0 \\ -AS_{TC} & 1 & 0 & 0 & -AS_D \\ -TD_{TC} & -TD_{AS} & TD_{WD} & -TD_{RL} & 0 \\ -RL_{TC} & 0 & 0 & 1 & 0 \\ 0 & -D_{AS} & 0 & -D_{RL} & 1 \end{vmatrix}}{|J|} = \\ &= \frac{RL_{TC} TC_{WD} TD_{AS} + TC_{WD} TD_{TC} + TD_{WD} - RL_{TC} TC_{RL} TD_{WD} + AS_{TC} (TC_{WD} TD_{AS} - TC_{AS} TD_{WD}) +}{|J|} \\ &\quad + AS_D \{D_{RL} RL_{TC} (TC_{WD} TD_{AS} - TC_{AS} TD_{WD}) - D_{AS} [TC_{WD} TD_{TC} + TD_{WD} + RL_{TC} (TC_{WD} TD_{RL} - TC_{RL} TD_{WD})]\} \end{aligned} \quad (a.31)$$

### Effect of an Exogenous Change in Trip Chaining and Residential Location

This comparative-static is used to assess the magnitude of an exogenous change affecting both trip chaining and density. In particular, it is used to assess the extent of the impact of a change in distance to the nearest subcenter, an exogenous variable appearing on equation (4.13) and equation (4.17) of Model III. This change is measured as

$$\begin{aligned} \frac{d\overline{TD}}{dsubc\_dist} &= \frac{\begin{vmatrix} 1 & -TC_{AS} & TC_{\phi} & -TC_{RL} & 0 \\ -AS_{TC} & 1 & 0 & 0 & -AS_D \\ -TD_{TC} & -TD_{AS} & 0 & -TD_{RL} & 0 \\ -RL_{TC} & 0 & 0 & 1 & 0 \\ 0 & -D_{AS} & D_{\phi} & -D_{RL} & 1 \end{vmatrix}}{|J|} = \\ &= \frac{TC_{\phi}(AS_{TC}TD_{AS}+RL_{TC}TD_{RL}+TD_{TC})+AS_D\{TC_{\phi}[D_{RL}RL_{TC}TD_{AS}-D_{AS}(RL_{TC}TD_{RL}+TD_{TC})]+ \\ &\quad +D_{\phi}[(1-RL_{TC}TC_{RL})TD_{AS}+TC_{AS}(RL_{TC}TD_{RL}+TD_{TC})]\}}{|J|} \end{aligned} \quad (a.32)$$

### Effect of an Exogenous Change in Activity Space

An exogenous change affecting the activity space has the following impact on the demand for transit

$$\frac{d\overline{TD}}{dAS_\varphi} = \frac{\begin{vmatrix} 1 & -TC_{AS} & 0 & -TC_{RL} & 0 \\ -AS_{TC} & 1 & AS_\varphi & 0 & -AS_D \\ -TD_{TC} & -TD_{AS} & 0 & -TD_{RL} & 0 \\ -RL_{TC} & 0 & 0 & 1 & 0 \\ 0 & -D_{AS} & 0 & -D_{RL} & 1 \end{vmatrix}}{|J|} = \frac{AS_\varphi[(1-RL_{TC}TC_{RL})TD_{AS}+TC_{AS}(RL_{TC}TD_{RL}+TD_{TC})]}{|J|} \quad (a.33)$$

This comparative-static is used to measure the change in transit demand due a change in retail establishment density ( $r\_estd$ ), an exogenous variable appearing in equation (4.14) in Model III.

## Appendix C Equation Identification

### Identification

In the context of simultaneous equation modeling, the validity of results hinges on the determination of the exclusion restrictions. That is, the researcher must a priori determine what explanatory variables are to be included and excluded from each equation. The determination of the exclusion restrictions defines a model that is correctly specified in the sense that the matrix of the reduced form parameters to be estimated is unique in its representation of the more primitive structural matrix. Exclusion restrictions need to be drawn outside of the variables a researcher has available from a given dataset (i.e, they should be based on sound behavioral theory).

A necessary and sufficient condition for identification of a structural equation is provided by the rank condition. The rank condition assures that the exclusion restrictions are sufficient and are unique. The following steps are required to obtain the rank condition for a given structural equation.

1) Let  $\Delta$  be a matrix of all the structural parameters

$$\Delta = \begin{bmatrix} B \\ \Gamma \end{bmatrix} \quad (c.1)$$

and let  $R_i$  be the matrix of exclusion restrictions defining structural equation  $i$

$$R = [1 \quad \dots \quad 0] \quad (c.2)$$

2) Premultiply (c.1) by (c.2) to obtain the list of variables excluded from equation  $i$

$$R\Delta = [1 \quad \dots \quad 0] \begin{bmatrix} \beta_{1i} \\ \vdots \\ \gamma_{ki} \end{bmatrix} = \beta_{1i} + \dots \quad (c.3)$$

3) Compute the rank of  $R\Delta$

4) Equation ( $i$ ) is identified (overidentified) if the rank is equal (greater) to  $G-1$ ; where  $G$  is equal to the number of endogenous variables

Next, each of the four models presented next is subject to the rank condition for identification prior to estimation and results are reported below. Note that the size of  $R$  depends on the number of exogenous and endogenous structural parameters excluded by each

equation. The following notation is used to denote exogenous and endogenous appearing or being excluded by each equation

$G$  = total number of endogenous variables

$K$  = total number of exogenous variables

$g_i$  = number of endogenous variables included in equation  $i$

$g_i^*$  = number of endogenous variables excluded from equation  $i$

$k_i$  = number of exogenous variables included in equation  $i$

$k_i^*$  = number of exogenous variables excluded from equation  $i$

## Model I

Recalling the specification of Chapter 4, the following rank conditions for identification are obtained.

### *Trip Chaining Equation*

<i>Inclusions/Exclusions</i>	<i>Number</i>
$G$	3
$K$	13
$g_i$	1
$g_i^*$	1
$k_i$	7
$k_i^*$	6

The rank condition is given by



$$\begin{pmatrix}
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix}
 \begin{pmatrix}
 -1 & b_{1,2} & b_{1,3} \\
 b_{21} & -1 & b_{2,3} \\
 0 & 0 & -1 \\
 \gamma_{1,1} & 0 & \gamma_{1,3} \\
 \gamma_{2,1} & 0 & \gamma_{2,3} \\
 \gamma_{3,1} & 0 & \gamma_{3,3} \\
 \gamma_{4,1} & 0 & 0 \\
 \gamma_{5,1} & \gamma_{5,2} & 0 \\
 \gamma_{6,1} & 0 & 0 \\
 \gamma_{7,1} & \gamma_{7,2} & 0 \\
 0 & \gamma_{8,2} & 0 \\
 0 & \gamma_{9,2} & 0 \\
 0 & \gamma_{10,2} & 0 \\
 0 & 0 & \gamma_{11,3} \\
 0 & 0 & \gamma_{12,3} \\
 0 & 0 & \gamma_{13,3}
 \end{pmatrix}
 =
 \begin{pmatrix}
 0 & 0 & -1 \\
 0 & \gamma_{8,2} & 0 \\
 0 & \gamma_{9,2} & 0 \\
 0 & \gamma_{10,2} & 0 \\
 0 & 0 & \gamma_{11,3} \\
 0 & 0 & \gamma_{12,3} \\
 0 & 0 & \gamma_{13,3}
 \end{pmatrix}$$

Rank  $\underset{7 \times 16}{R} \underset{16 \times 3}{\Delta} = 2$ ;  $(G - 1) = 2$  The trip chaining equation is just identified

### Activity Space Equation

Inclusions/Exclusions	Number
$G$	3
$K$	13
$g_i$	1
$g_i^*$	1
$k_i$	4
$k_i^*$	9

The rank condition is given by

$$\begin{pmatrix}
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix}
 \begin{pmatrix}
 -1 & b_{1,2} & b_{1,3} \\
 b_{21} & -1 & b_{2,3} \\
 0 & 0 & -1 \\
 \gamma_{1,1} & 0 & \gamma_{1,3} \\
 \gamma_{2,1} & 0 & \gamma_{2,3} \\
 \gamma_{3,1} & 0 & \gamma_{3,3} \\
 \gamma_{4,1} & 0 & 0 \\
 \gamma_{5,1} & \gamma_{5,2} & 0 \\
 \gamma_{6,1} & 0 & 0 \\
 \gamma_{7,1} & \gamma_{7,2} & 0 \\
 0 & \gamma_{8,2} & 0 \\
 0 & \gamma_{9,2} & 0 \\
 0 & \gamma_{10,2} & 0 \\
 0 & 0 & \gamma_{11,3} \\
 0 & 0 & \gamma_{12,3} \\
 0 & 0 & \gamma_{13,3}
 \end{pmatrix}
 =
 \begin{pmatrix}
 0 & 0 & -1 \\
 \gamma_{1,1} & 0 & \gamma_{1,3} \\
 \gamma_{2,1} & 0 & \gamma_{2,3} \\
 \gamma_{3,1} & 0 & \gamma_{3,3} \\
 \gamma_{4,1} & 0 & 0 \\
 \gamma_{6,1} & 0 & 0 \\
 \gamma_{7,1} & \gamma_{7,2} & 0 \\
 0 & 0 & \gamma_{11,3} \\
 0 & 0 & \gamma_{12,3} \\
 0 & 0 & \gamma_{13,3}
 \end{pmatrix}$$

Rank  $\underset{10 \times 16}{R} \underset{16 \times 3}{\Delta} = 3$ ;  $(G - 1) = 2$  the activity space equation is overidentified

### Transit Demand Equation

Inclusions/Exclusions	Number
$G$	3
$K$	13
$g_i$	2
$g_i^*$	0
$k_i$	4
$k_i^*$	7

The rank condition is given by

$$\begin{pmatrix}
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
 \end{pmatrix}
 \begin{pmatrix}
 -1 & b_{1,2} & b_{1,3} \\
 b_{21} & -1 & b_{2,3} \\
 0 & 0 & -1 \\
 \gamma_{1,1} & 0 & \gamma_{1,3} \\
 \gamma_{2,1} & 0 & \gamma_{2,3} \\
 \gamma_{3,1} & 0 & \gamma_{3,3} \\
 \gamma_{4,1} & 0 & 0 \\
 \gamma_{5,1} & \gamma_{5,2} & 0 \\
 \gamma_{6,1} & 0 & 0 \\
 \gamma_{7,1} & \gamma_{7,2} & 0 \\
 0 & \gamma_{8,2} & 0 \\
 0 & \gamma_{9,2} & 0 \\
 0 & \gamma_{10,2} & 0 \\
 0 & 0 & \gamma_{11,3} \\
 0 & 0 & \gamma_{12,3} \\
 0 & 0 & \gamma_{13,3}
 \end{pmatrix}
 =
 \begin{pmatrix}
 \gamma_{4,1} & 0 & 0 \\
 \gamma_{5,1} & \gamma_{5,2} & 0 \\
 \gamma_{6,1} & 0 & 0 \\
 \gamma_{7,1} & \gamma_{7,2} & 0 \\
 0 & \gamma_{8,2} & 0 \\
 0 & \gamma_{9,2} & 0 \\
 0 & \gamma_{10,2} & 0
 \end{pmatrix}$$

Rank  $\underset{7 \times 16}{R} \underset{16 \times 3}{\Delta} = 2$ ;  $(G - 1) = 2$  the transit demand equation is just identified





**Transit Demand Equation**

<i>Inclusions/Exclusions</i>	<i>Number</i>
$G$	4
$K$	18
$g_i$	3
$g_i^*$	0
$k_i$	5
$k_i^*$	13

The rank condition is given by

$$\begin{pmatrix}
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix}
 \begin{pmatrix}
 -1 & b_{1,2} & b_{1,3} & b_{1,4} \\
 b_{2,1} & -1 & b_{2,3} & 0 \\
 0 & 0 & -1 & b_{3,4} \\
 b_{4,1} & 0 & b_{4,3} & -1 \\
 \gamma_{1,1} & 0 & \gamma_{1,3} & 0 \\
 \gamma_{2,1} & 0 & \gamma_{2,3} & 0 \\
 \gamma_{3,1} & 0 & 0 & 0 \\
 \gamma_{4,1} & \gamma_{4,2} & 0 & 0 \\
 \gamma_{5,1} & 0 & 0 & 0 \\
 \gamma_{6,1} & 0 & 0 & 0 \\
 0 & \gamma_{7,2} & 0 & 0 \\
 0 & \gamma_{8,2} & 0 & 0 \\
 0 & \gamma_{9,2} & 0 & 0 \\
 0 & 0 & \gamma_{10,3} & 0 \\
 0 & 0 & \gamma_{11,3} & 0 \\
 0 & 0 & \gamma_{12,3} & 0 \\
 0 & 0 & 0 & \gamma_{13,4} \\
 0 & 0 & 0 & \gamma_{14,4} \\
 0 & 0 & 0 & \gamma_{15,4} \\
 0 & 0 & 0 & \gamma_{16,4} \\
 0 & 0 & 0 & \gamma_{17,4} \\
 0 & 0 & 0 & \gamma_{18,4}
 \end{pmatrix}
 =
 \begin{pmatrix}
 \gamma_{3,1} & 0 & 0 & 0 \\
 \gamma_{4,1} & \gamma_{4,2} & 0 & 0 \\
 \gamma_{5,1} & 0 & 0 & 0 \\
 \gamma_{6,1} & 0 & 0 & 0 \\
 0 & \gamma_{7,2} & 0 & 0 \\
 0 & \gamma_{8,2} & 0 & 0 \\
 0 & \gamma_{9,2} & 0 & 0 \\
 0 & 0 & 0 & \gamma_{13,4} \\
 0 & 0 & 0 & \gamma_{14,4} \\
 0 & 0 & 0 & \gamma_{15,4} \\
 0 & 0 & 0 & \gamma_{16,4} \\
 0 & 0 & 0 & \gamma_{17,4} \\
 0 & 0 & 0 & \gamma_{18,4}
 \end{pmatrix}$$

Rank  $\underset{13 \times 22}{R} \underset{22 \times 4}{\Delta} = 3$ ;  $(G - 1) = 3$  the transit demand equation is just identified











### ***Residential Location Equation***

<i>Inclusions/Exclusions</i>	<i>Number</i>
$G$	5
$K$	18
$\mathbf{g}_i$	2
$\mathbf{g}_i^*$	2
$\mathbf{k}_i$	6
$\mathbf{k}_i^*$	12

The rank condition is given by

[illegible]

Rank  $\underbrace{R}_{14 \times 23} \underbrace{\Delta}_{23 \times 5} = 4$ ;  $(G - 1) = 4$  the residential location equation is just identified



*[This page intentionally left blank.]*

# Appendix D On the Relevance of Transit Station Proximity

---

## Introduction

This working paper revisits the issue of residential sorting and travel behavior by looking at the relationship between transit demand and residential location choice. An analytical model of residential location and transit patronage is presented, where the location decision is jointly determined with the demand for transit services. Simultaneity and endogeneity between residential location and mode-choice are dealt with by applying instrumental probit regression. Findings show that not accounting for location idiosyncratic preferences results in underestimation of the relevance of transit station proximity. The treatment of transit proximity as an endogenous continuous variable reveals the presence of an omitted variable bias to date not explicitly treated by the current literature. In particular, in higher urban density settings, an inherent supply-side spatial bias is present and correlated to relevant instrumental variables that control for neighborhood characteristics. When not accounted for, this supply-side bias exerts a downward bias that results in an overestimation of station proximity to a given residential unit. Alternative model structures and estimation methods that are employed to confirm these results are also presented. An explanatory variable that considers the spatiotemporal dispersion of non-work activities is also introduced and its role discussed. Recommendations to shift the study of residential self-selection from a trip-based approach to one that fits into the new paradigm of activity-based travel behavior research are also provided.

## Analytical Framework

Within the literature of residential sorting, the location decision is often presented as a dichotomous choice, i.e., live near or far away from a transit station. Proximity is defined by applying a circular buffer around a station, usually a Euclidean (linear) radius of a half mile. The extent of this buffer is usually justified on empirical grounds. Cervero (Cervero, 2007), for example, applies a nested logit model of the joint decision of mode and location based on a half-mile buffer. Use of a linear distance dismisses the possibility of accounting for barriers that might prevent access a station.

The use of transit proximity as a proxy for residential location, while dictated by the need to sort out the influence of the built environment from self-selection, is not based on any other theoretical underpinnings about the decision making process that is at the heart of urban residential location theory. That is, it does not take into consideration the trade-off between

housing and transportation costs that, at the margin, determine where an individual decides to locate. For example, standard theory of location shows that individuals choose an optimal distance between work and home location given housing and transportation costs. In a monocentric model that only looks at travel between home and the central business district (CBD), individuals locate at a distance where the marginal cost of transportation is equal to the marginal housing cost savings obtained by a move farther out from the CBD (Moses, 1958, Muth, 1969). Recent departures from this view consider that individuals can locate anywhere in an urban area, choosing an optimal home-work distance that optimizes also the amount of non-work travel and non-work activities (Anas and Kim, 1996, Anas and Xu, 1999). Further explorations also consider the role of trip chaining behavior (Anas, 2007).

In a departure from the monocentric model, the definition of residential location is taken from the polycentric model of Anas and his associates (Anas and Kim, 1996, Anas and Xu, 1999). Residential location is defined as the optimal *job-residence pair* in an urban area in which jobs and residences are dispersed. Following urban residential location theory, the location decision is assumed to be the result of a trade-off between housing expenditures and transportation costs, given income and the mode-choice set. Following Anas, the location decision is also based on idiosyncratic preferences for location and travel. This framework allows for the optimal determination of residential location and, given the home-work commuting distance, of non-work trip, goods consumption, and mode choice. Thus, the joint determination of mode choice and residential location choice can be expressed as

$$TD = f(HAS, RL, X_{TD}) \quad (d.1)$$

$$RL = f(TD, X_{RL}) \quad (d.2)$$

Where

RL= residential location – optimal home-work distance, or optimal “commute pair arrangement”

TD= Transit Demand – dichotomous choice of transit versus other modes

HAS= Household Activity Space – a spatiotemporal measure of activity-location dispersion

$X_{RL}$  = vector of location controls, such as housing and location characteristics

$X_{TD}$  = vector of controls, comprising mode-specific and socio-demographic controls

Substituting equation (d.2) into equation (d.1), the following reduced form equation is obtained

$$TD = f(X_{TD}, X_{RL}, HAS) \quad (d.3)$$

## Spatial dispersion of non-work activities and travel behavior

The literature on self-selection empirically frames this issue within modeling frameworks that rarely account for the fact that the demand of travel is derived from the activities requiring travel (i.e., as fitting within activity-based theory).

The study of the interaction between activity-travel requirements and features of the urban landscape is not new, although it requires the availability of specialized data and geospatial tools. A growing field of research that looks at the relationship between urban form and the spatiotemporal allocation of activities and travel provides additional insight into the impact of the built environment. Recent research describing travel behavior and the influence of urban morphology and entire patterns of daily household activities and travel demonstrates how households residing in decentralized, lower density urban areas tend to have a more dispersed activity-travel pattern than their urban counterpart (Buliung and Kanaroglou, 2006).

This paper explicitly accounts for how the built environment affects the spatial dispersion of activities, the demand for travel, and location decisions. To account for the influence of the built environment on travel demand, the explanatory variable defining the household activity space (HAS) is used. The term household activity space (HAS) was first introduced by Kanaroglou and Bouliung as a measure of spatial dispersion of activities at the household level. This paper, while retaining the same acronym, measures HAS with a different metric.

In this study, the extent of the activity space is assumed to be affected by the built environment. Densely populated urban areas tend to cluster activity locations together thus shrinking the size of the activity space. This affects the spatial allocation of activities, thus affecting the demand for travel. This effect is accounted for by introducing the variable *HAS* into equation (d.1), and, in its simplest specification, by treating it as exogenous to the model. Note that this framework does not explicitly treat the influence that trip chaining might have in affecting the extent of the household activity space (the author is currently working on an extension that looks at the joint determination of household activity space, residential location and non-work travel in a context of trip-chaining behavior).

Equation (d.3) considers transit mode-choice as a function of both the independent variables affecting it directly ( $X_{TD}$ ), the independent variables affecting equation (d.2) and specified by the vector  $X_{RL}$ , and the extent of the household activity space, *HAS*. The empirical treatment of this relationship requires careful consideration of the relationship between these two equations, as discussed in the next section.

## Econometric Model Specification

Often, empirical estimation of equation (d.3) treats the vector  $X_{RL}$  as exogenous. While a growing body of literature increasingly recognizes that unobserved idiosyncratic preferences affect  $X_{RL}$ , the debate hinges on the best way to treat the most common consequence of not controlling for this problem, i.e., the resulting omitted variable bias (Mokhtarian and Cao, 2008).

The empirical treatment of omitted variable bias in the study of self-selection ranges from nested logit regression (Cervero, 2007) to error correlation models (Bhat and Guo, 2004, Pinjari, Pendyala, Bhat and Waddel, 2007). Alternative methods include instrumental variable regression, with leading examples discussed earlier (Crane, Boarnet and Sarmiento), that use a set of properly tailored instruments. Whatever the method employed, residential choice, as discussed earlier, is usually modeled as a dichotomous variable. This allows specifying the residential location decision as one where the location choice set is not exogenous and must ad-hoc be determined by the researcher. The treatment of the location decision as a dichotomous variable inherently presents a problem that is at the very heart of residential self-selection research. When using discrete choice modeling, one must assume that all individuals can choose among all possible locations within an urban area (see for example, Cervero and Pinjari et al.). The treatment of mode-choice and residential location in more sophisticated frameworks does not eliminate the need of an ad-hoc determination of the residential choice set. For example, both Pinjari et al. (Pinjari, Pendyala, Bhat and Waddel, 2007), and Bhat and Guo (Guo and Chen, 2007), who adopt the more sophisticated multinomial logit-ordered structure that explicitly consider the correlation of unobserved factors simultaneously affecting both choices, must a priori determine the location choice set (in that case any individual is assumed to be able to choose among 223 different locations). This assumption ignores the fact that, due to income and vehicle availability, some individuals have more contracted mode choice and residential location sets. This results in not being able fully to discern if residential choice is actually a choice that accounts for idiosyncratic preferences or a result of spatial mismatch.

It can be shown that treating residential location as a continuous choice variable allows specifying an estimation model that does not require an ad-hoc determination of the location choice set. This allows equation (d.1) and equation (d.2) to be specified using the following parameterization

$$y_1^* = \mathbf{z}_1\delta_1 + \alpha_1 y_2 + u_1 \quad (\text{d.4})$$

$$y_2 = \mathbf{z}_1\delta_{21} + \mathbf{z}_{22}\delta_{22} + v_2 = \mathbf{z}\delta_2 + v_2 \quad (\text{d.5})$$



$$y_1 = 1[y_1^* > 0] \quad (d.6)$$

Where

$y_1^*$  = TD, mode choice (transit versus other)

$y_2$  = RL, residential location choice

$\mathbf{z}$  = vector of controls ( $X_{RL}, X_{TD}, HAS$ )

Given the treatment of residential location as a continuous variable, and of transit choice as a dichotomous variable, equations (4) and (5) allow for the explicit treatment of a case of omitted variable bias in the context of a limited dependent variable model. Assuming that  $u_1$ , and  $v_2$  have zero means, bivariate normal distributions and independent of  $\mathbf{z}$ , equation (d.4) and equation (d.6) represent a structural equation with equation (d.5) representing a reduced form equation. As discussed in Wooldridge (2002), the reduced form equation can be estimated by applying an instrumental variable regression using a binary probit model that leads to an average partial effect interpretation. In addition, a simple test of endogeneity of  $y_2$  based on Rivers and Vuong (1988) can be conducted, as presented in the next section.

In the context of simultaneous equation modeling or instrumental variable regression, the validity of results hinges on the determination of the exclusion restrictions. That is, the researcher must a priori determine what explanatory variables are to be included and excluded from a given equation. The determination of the exclusion restrictions determines a model that is correctly specified in the sense that the matrix of the reduced form parameters to be estimated is unique in its representation of the more primitive structural matrix. Exclusion restrictions need to be drawn outside of the variables a researcher has available from a given dataset, i.e, they should be based on sound behavioral theory.

The general form of equation (d.1) reports a vector of controls represented by  $X_{TD}$ , which in its parameterized version is defined as the vector  $\mathbf{z}_1$  of equation (d.4). The literature on transit patronage provides insight as to which variables are relevant to transit choice. Among the socio-demographic variables to be explicitly included as controls are vehicle availability, income, and household structure. The inclusion of these controls is the natural consequence of considering the demand for travel as derived from the demand for goods and services. Additional controls are directly related to transit as a mode of choice. These include the presence of a transit stop at workplace, measured using a half-mile buffer, the presence of park-and-ride facilities, and, most important, walking distance from residence unit to the nearest transit station.

The general form of equation (d.2) reports the vector of controls,  $X_{RL}$ . This vector includes controls for household socio-demographic characteristics and controls that account for idiosyncratic preferences. The latter constitute the set of exclusion restrictions from equation

(d.1) and represent the instrumental variable for residential location when estimating the probability of choosing transit as a mode. These exclusion restrictions are discussed in the section that reports the model empirical results.

### **Model Data**

The dataset is a result of combining data from travel behavior at the individual and household level with geographical data from the Census Bureau. The latter data source is specifically Summary File 3, which consists of detailed tables of social, economic and housing characteristics compiled from a sample of approximately 19 million housing units (about 1 in 6 households) that received the Census 2000 long-form questionnaire (reference here). These data were obtained at the block group level.

Travel behavior data at the micro level were obtained from 2000 Bay Area Transportation Survey (BATS2000). BATS2000 is a large-scale regional household travel survey conducted in the nine county San Francisco Bay Area of California by the Metropolitan Transportation Commission (MTC). Completed in the spring of 2001, BATS 2000 provides consistent and rich information on trip chaining behavior of 15,064 households, including 2,503 household that make regular use of transit.

BATS 2000 uses the latest applications of activity and time-based survey instruments to study travel behavior. The literature on self-selection has relied on this dataset in several instances and to study other aspects of travel behavior (Commission, 2008). The data from BATS2000 are available online and maintained as a set of relational data files and are available as comma-separated value (CSV) and straight ASCII text files. In the dataset, 99.9 percent of home addresses were geocoded to the street address or street intersection level (99.5 percent to the street address level). For activity locations, 86 percent are geocoded to the street address level and 8 percent to street intersection (MORPACE, 2002). This level of detail allows computing geographic measures of the household activity space.

The unit of observation is the household, to reflect the higher hierarchical decision making process of both residential location and the ensuing travel needs. Thus, housing and neighborhood characteristics are measured at the block group level where the residential unit is located.

### **Measures of Household Activity Space**

Household activity space is measured using area-based geometric measures used in transportation geography. Different metrics that describe the spatial extent of activity locations can be employed. The simplest measure is represented by the standard distance circle (SDC) (otherwise defined as standard distance deviation in spatial statistics) and is

essentially a bivariate extension of the standard deviation of a univariate distribution. It measures the standard distance deviation from a mean geographic center and is computed as

$$SD = \sqrt{\frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{n}} \equiv \sqrt{\frac{\sum d^2}{n}} \quad (d.7)$$

where  $\bar{x}$  and  $\bar{y}$  represent the spatial coordinates of the mean center of non-work activities at the household level, and the  $i$  subscript indicates the coordinates of each non-work activity. The mean center is analogous to the sample mean of a dataset and it represents the sample mean of the  $x$  and  $y$  coordinates of non-work activities contained in each household activity set. The coordinates represent longitude and latitude measurement of each activity and are reported in meters, following the Universal Transverse Mercator (UTM) coordinate system. Household activity locations are those visited by surveyed household members during a specified time interval, in this case two representative weekdays.

Thus, the standard distance of a household's activity pattern is estimated as the standard deviation (in meters or kilometers) of each activity location from the mean center of the complete daily activity pattern. Interpretation is relatively straightforward with a larger standard distance indicating greater spatial dispersion of activity locations. The area of the *SDC* is obtained as the area of a circle with a radius equal to the standard distance. It provides a summary dispersion measure that can be used to explore systematic variations of activities subject to socio-demographic, travel patterns, and patterns of land-use. As pointed out by Ebdon (1977), this measure is affected by the presence of outliers or activities that are located farthest from the mean center. As a result of the squaring of all the distances from the mean center, the extreme points have a disproportionate influence on the value of the standard distance.

To eliminate dependency on spatial outliers, another measure of dispersion is usually employed, called the standard deviational ellipse (SDE). The advantages of SDE over SDC have been discussed in the literature. In addition to control for outliers, SDE also allows accounting for directional bias of activities with respect to its mean center. The ellipse is centered on the mean center with the major axis in the direction of maximum activity dispersion and its minor axis in the direction of minimum dispersion. This study uses the standard distance ellipse (*SDE*), as described in Levine (2005).

### **Measures of Residential Location, *RL***

In this empirical model, residential location is defined as the average distance of household employment activities from the household residential unit. This average distance can be expressed as

$$RL = \frac{\sum_{m=1}^k dist_{mj}}{k} \quad (d.8)$$

Where  $dist_{mj}$  is the Euclidean distance to the residential unit located at  $j$  from a household member work location  $m$ , and  $k$  is the total number of employed household members.

### Transit Station Proximity

As previously discussed, this paper does not treat transit station proximity as a proxy for residential location; instead residential location is the optimal home-work commuting pair arrangement between  $i$  and  $j$ . As such, transit station proximity is itself endogenous to the choice of mode and affected by idiosyncratic preferences for location. Walking distance is herein used to indicate proximity to the nearest transit station. As discussed in the next section, walking distance is measured as actual distance based on network characteristics to take into consideration the existence of accessibility impediments.

### Instrumental Variables

The literature provides some insight on the use of variables as instruments. This paper uses some instruments employed in the literature as well as some new ones. The following instrumented variables have all being obtained at the block group level using the Summary 3 Census Bureau file:

1. Stock of housing built before 1945 (number of housing units)
2. Housing median value (dollars; occupied owners units)
3. Housing median age (years; non-rent units)
4. Housing size (median number of rooms; occupied owners units)
5. Percent of households living below poverty line
6. Diversity index (0 = homogeneous; 1 = heterogeneous neighborhood)

The first variable has been used before (Boarnet and Crane, 2001, Boarnet and Sarmiento, 1998, Crane, 2000, Crane and Crepeau, 1998b), while the remaining ones are unique to this study. Additional controls for neighborhood characteristics have also been used elsewhere. For example, the proportion of block group or census tract population that is Black, and the proportion Hispanic have been used as instruments by Boarnet and Sarmiento (1996), and the percent of foreigners by Vans and Hedel (2007).

In this paper, instruments one through four are meant to be used to control for idiosyncratic preferences for housing characteristics, not directly affecting travel behavior, but directly affecting the residential choice decision at the household level. Variables five and six are intended as controls for neighborhood characteristics. In particular, the percent of households living below poverty levels (henceforth defined as poverty) serves as a proxy from

crime, while the diversity index (henceforth called diversity) is used as a proxy for ethnic preferences (i.e., moving into a neighborhood with similar ethnic characteristics). The latter is an index of ethnic heterogeneity that varies from zero (only one race living in the neighborhood) to one (no race is prevalent). As discussed in further detail in the next section, poverty and diversity serve a dual role as instrumental variables when transit station proximity is treated as endogenous to the model.

In this paper, the location decision is intended as a process where the household decides an optimal home-work commute (the one yielding the highest utility among all possible combinations) that takes into consideration the trade-off of housing costs and transportation costs, while at the same time accounting for idiosyncratic preferences for both location and travel. Thus, it is assumed that housing and neighborhood characteristics, while not directly affecting travel, indirectly affecting residential location. This is a somewhat different role than the one in previous studies. In those studies, residential sorting affects land-use variables, which in turn affects the cost of travel; the latter, in turn, affects travel behavior.

One way analysis of variance tables (not herein reported) that included an interaction term between transit household and the transit station dummy variable were generated. All variables exhibited a significant difference in means indicating that housing price, housing age, room size, neighborhood diversity and poverty levels differs across households according to their location and mode choice.

## Results

Next, the empirical model is specified to include a set of relevant explanatory variables for equation (d.4) and the set of instruments for equation (d.5) discussed in the previous section. The vector of controls entering equation (d.4) is the same as the vector of controls  $X_{TD}$  in its generalized form. The following explanatory variables are considered to affect transit choice and thus are entered in equation (d.4), which closely follows the literature on residential sorting behavior

### *Household Characteristics*

- Number of employed persons in the household
- Number of children by age group
- Household combined income
- Householder ethnicity
- Spatial dispersion of non-work activities (household activity space)

*Mode Choice Controls*

- Presence of a transit stop within a half-mile of work location for householder
- Presence of a park-and-ride facility within a half-mile of residential unit
- Transit stop proximity to residential unit
- Number of vehicles

Table D.1 reports the regression results. The base model is a binary probit mode-choice model that treats both residential location and transit station proximity as exogenous. The second column reports the results that are obtained by running the instrumental probit regression that treats residential location transit station proximity as endogenous (IV Probit). This constitutes the model of interest from which conclusions are drawn and comparison made. The third column reports the results that treat transit station proximity as endogenous, residential location exogenous, and removes the poverty and diversity instruments (IV Probit\*).

An initial comparison of the sign and magnitude of the parameters of interests reveals that, the endogenous treatment of both residential location and proximity results in significant differences with respect to a regular probit regression. First, the negative signs of residential location and transit stop proximity at work and the residence unit have the expected sign, with home walking distance much more relevant in the IV Probit than regular probit. Park and ride has a statistically significant negative sign, not expected. This might be due to that fact that transit choice is modeled at the household level rather the individual level.

TABLE D.1 Model Results

	Probit		Transit Use (1,0) IV Probit		IV Probit *	
	Coefficient	P>  z	Coefficient	P>  z	Coefficient	P>  z
Residential location (home-work distance; miles) <sup>a</sup>	0.049	(0.000)	-0.271	(0.016)	0.054	(0.000)
Walking distance to nearest transit station (miles)	-0.220	(0.000)	-1.187	(0.000)	-1.870	(0.000)
Transit stop at work (1/2 mile buffer)	0.825	(0.000)	0.745	(0.000)	0.592	(0.000)
Park and ride facility (1/2 mile buffer)	-0.161	(0.030)	-0.394	(0.004)	-0.305	(0.011)
Number of Vehicles	-0.458	(0.000)	-0.291	(0.000)	-0.266	(0.000)
Household income (\$,0000) <sup>b</sup>	-0.019	(0.485)	-0.046	(0.341)	-0.130	(0.009)
Employed ( number of household members)	0.224	(0.000)	0.309	(0.000)	0.124	(0.028)
Householder age	-0.004	(0.023)	-0.007	(0.027)	0.001	(0.642)
Householder race (base level=Non-hispanic white)						
Black	0.446	(0.000)	0.316	-0.073	0.067	(0.715)
Hispanic	-0.068	(0.403)	-0.031	(0.821)	-0.234	(0.085)
Asian	-0.019	(0.756)	0.028	(0.794)	-0.148	(0.147)
Other	0.054	(0.587)	0.064	(0.690)	0.030	(0.850)
Children (number)						
Under 6	-0.138	(0.001)	-0.095	(0.160)	-0.170	(0.007)
6 to 11	-0.089	(0.008)	-0.035	-0.518	-0.013	(0.804)
12 to 19	0.222	(0.000)	0.159	(0.003)	0.226	(0.000)
over 19	0.406	(0.000)	0.342	(0.000)	0.443	(0.000)
San Francisco County (1,0)	0.772	(0.000)	0.142	(0.496)	0.074	(0.000)
Household Activity Space (miles <sup>2</sup> )	0.055	(0.000)	0.091	(0.000)	0.465	(0.000)
Constant	-1.668	(0.000)	0.172	(0.778)	-0.929	(0.001)
Log-likelihood	-3377.09					
(Pseudo) R <sup>2</sup>	0.204		n/a		n/a	
Wald test of exogeneity Chi2	n/a		82.07		65.92	
Prob> chi2	n/a		0.000		0.000	
Number of observations	8,669		8,669		8,669	
<i>p</i> -values in parentheses						

\* Poverty and diversity instruments excluded; residential location treated as exogenous

a,b: squared terms used but not reported

n/a: not applicable

Model comparisons can also be conducted by looking at the marginal effects (or partial derivatives over the predicted probability of choosing transit) as they provide a quantitative measure of influence on the predicted probability of choosing transit. Given that the model specification treats income and age as non linear (not reported in Table A.1), and household activity space as natural log, the marginal effects were evaluated at sample mean values

following Wooldridge (page 466). The marginal effects were obtained by measuring changes in the predicted probability

$$\hat{P}(y = 1) = \Phi(\bar{x}\hat{\beta}) \quad (\text{d.10})$$

where

$\Phi$  = standard normal cumulative distribution function

$\bar{x}$  = vector of sample averages of independent variables

$\hat{\beta}$  = vector of estimated parameters

Table D.2 presents a marginal effect comparison across models, obtained by utilizing average sample values of the independent variables to evaluate changes in predicted probabilities due to the presence of a transit stop at work place and to evaluate the relevance transit station proximity near the residential unit. For example, for a non-Hispanic White household with a median income of \$110,000 and residing at about 11 miles from work and at 0.45 miles from a transit station, the Probit model estimates that the presence of a transit station in proximity of work (1/2 mile buffer) increases the predicted probability by about .18. This effect is relatively smaller once the inherent endogeneity of residential location and transit choice is accounted for, as reflected in a marginal effect of .173 (IV Probit).

**TABLE D.2** Marginal Effects

<i>Change in Probabilities (dy/dx)</i>	<i>Probit</i>	<i>IV Probit</i>	<i>IV Probit*</i>
Transit stop at work (1/2 mile buffer)	0.185	0.173	0.126
Walking distance (from 0.5 to 0.75 mile)	-0.021	-0.099	-0.121
Vehicle ownership (0 to 1)	-0.181	-0.112	-0.096

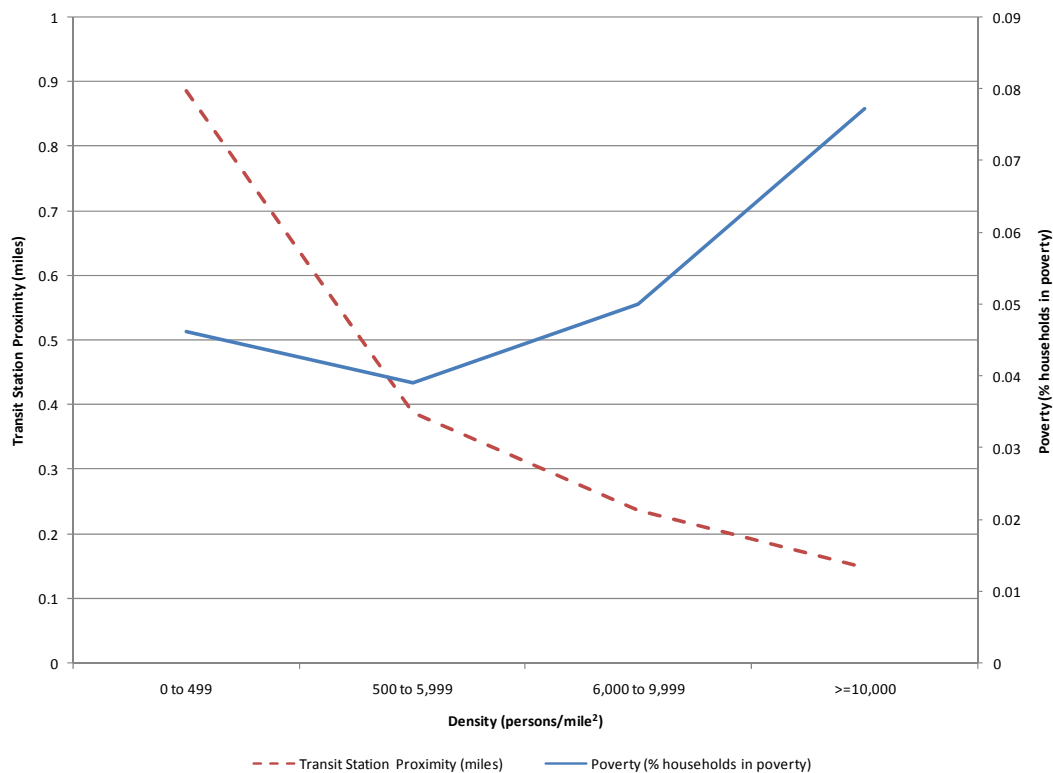
A further discussion ensues when considering transit station proximity itself as endogenous (while retaining residential location as exogenous). In fact, the choice of instruments such as poverty and diversity uncovers an unobserved effect that the literature on self-selection so far has not explicitly considered. Indeed, when both residential location and station proximity are endogenized in the model using the same set of instruments, the inclusion of poverty and diversity as instruments raises a question related to their relationship with proximity. Pearson



partial correlations (not reported here) show that poverty is negatively related to proximity, and positively correlated with diversity; a relationship that becomes stronger in more densely populated areas, as shown in Figure D.1.

In highly populated density areas the average distance to the nearest station is negatively correlated with poverty levels. As poverty levels increase, the average distance to the nearest station decreases (i.e., transit station proximity increase). This occurs because a higher proportion of the transit stations are located in urban areas characterized by higher than average poverty levels. A similar relationship exists between transit proximity and diversity. As transit proximity increases so does diversity, and indication that increased minority groups are more present at the block group level as density increases. This might be due to the fact that in densely populated urban areas, the supply of transit is more likely to be clustered around neighborhoods characterized by higher than average poverty levels. Essentially, as density increases, the likelihood of living in proximity to a station increases. Not being able to explicitly measure this phenomenon, the use of instrumental variables intended to control for neighborhood characteristics serve a dual role.

**FIGURE D.1** Poverty and Transit Station Proximity

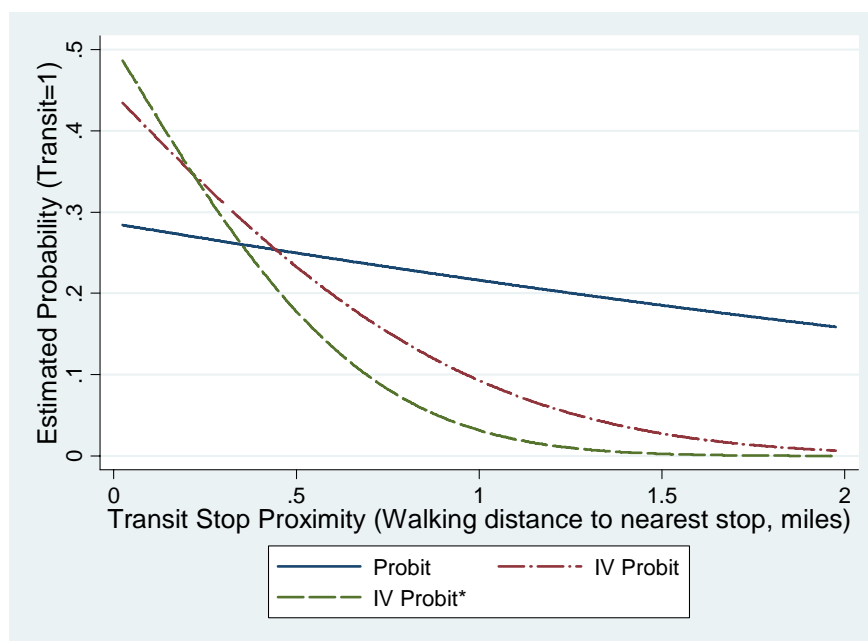


To see what happens when both poverty and diversity are omitted, a second IV-Probit was regressed without these instruments. The results are reported in the last column of Table 2

(IVProbit\*). It can be seen that the marginal effects of walking distance, transit stop at workplace location and park and ride are smaller. These effects are indicative of a downward bias given due to omitted variable bias (i.e., not accounting for this spatial bias). *Ceteris paribus* (holding socio-demographic factors constant) once this inherent bias is accounted for, transit station proximity becomes more relevant. This does not explicitly imply causality, a situation such that the supply of transit might be affected by the need to cater disadvantage cohorts of the population. It might just indicate a spatial correlation that warrants attention and further consideration.

The extent of the bias direction due to this phenomenon is illustrated in Figure 2, which graphs the estimated probabilities of the three models against transit station proximity using probability estimates. In all cases, the explanatory variables were set at their sample averages. For simplicity of exposition, the graph is sized to only report a maximum walking distance of two miles. At the sample mean walking distance of 0.45 miles, the probit model estimated probability of choosing transit is about .25, the instrumented Probit estimate is about .17 (IV Probit), and the estimate of the model with the exclusions of the diversity and poverty instruments is .20 (IV Probit\*). The biggest difference between the marginal effects estimates of the models, as indicated in Figure D.2. For example, an increase of walking distance from 0.5 miles to 0.75 reduces transit choice by 2.1 percent if using regular probit, about 9.9 percent if using IV Probit, and about 12.1 percent with the excluded instruments version IV Probit\*.

**FIGURE D.2** Estimated Probabilities



## Conclusions

This paper revisits the issue of residential sorting and travel behavior by looking at the relationship between transit demand and residential location choice. An analytical model of residential location and transit patronage is presented, where the location decision is jointly determined with the demand for transit services. Simultaneity and endogeneity between residential location and mode-choice are dealt with by applying instrumental probit regression.

Findings show that not accounting for location idiosyncratic preferences results in underestimation of the relevance of transit station proximity. The treatment of residential location as endogenous to the choice of transit shows that, after controlling household for socio-demographic characteristics, accessibility to transit services is more relevant than the case when idiosyncratic preferences for location are not considered.

As argued in the introduction, the literature on self-selection empirically frames this issue within modeling frameworks that rarely account for the fact that the demand of travel is derived from the demand for activities requiring travel. This paper considers the demand for travel deriving from out-of-home non-work activities at the household level, directly recognizing the joint nature of this decision-making process. In this study, the spatial extent of activity travel location is treated as exogenous to the model and not jointly determined with both location and the demand for travel. An extension of the model is one that would consider the determination of an optimal household activity space that is the result of a constrained optimization process (i.e., treating HAS as endogenous).

Finally, the treatment of transit proximity as an endogenous continuous variable revealed the presence of an omitted variable bias, not explicitly treated by the current literature. In particular, in higher urban density settings, an inherent supply-side spatial bias is present and correlated to relevant instrumental variables that control for neighborhood characteristics. When not accounted for, this supply-side bias exerts a downward bias that results in an overestimation of the relevance station proximity to a given residential unit. This downward bias due to omitted variables does not explicitly imply causality. Further study is warranted to test if this endogenous relationship persists as density levels vary, an indication that the supply of transit is partly a dependent upon serving the needs of a specific population group.

*[This page intentionally left blank.]*

## Appendix E Glossary of Terms

---

**Activity-based theory:** A travel behavior theory that is characterized by the recognition that travel is a derived demand, with a focus on constrained patterns or sequences of behavior instead of discrete trips, and the interdependence of decisions usually made within a household context.

**Ad-hoc:** A Latin phrase which means “for this [purpose].” In econometric and statistical modeling, it refers to a model specified to fit a particular dataset or a situation, which cannot be used for generalization of findings.

**A priori:** In behavioral models, it refers to laying out a set of behavioral relationships based on hypothesis (or derived by logic) rather than experiment.

**Bias:** The difference between the expected value of an estimator and the population value that the estimator is supposed to be estimating.

**Central business district (CBD):** The commercial and often geographical center of a city. The U.S. Census Bureau defines a central business district as an area of very high land valuation characterized by a high concentration of retail businesses, service businesses, offices, theaters, and hotels, and by a very high traffic flow.

**Ceteris paribus:** Is a Latin phrase meaning “with other things the same.” In statistics, it refers to the situation when all other relevant factors are held constant while the factor of interest is left to change. In economics, it refers to the assumption that all other relevant factors are held constant when examining the influence of one particular variable in an economic model; reflected in mathematical models by the use of partial differentiation.

**Continuous variable:** A variable that can assume an infinite set of values. In this report, transit station proximity to a household residence is a continuous variable measuring actual walking distance (measured in miles) to the nearest transit station.

**Dependent variable:** A variable that is determined by other variables, also known as an endogenous variable. In econometric theory, the dependent variable is a variable whose value changes (i.e., depends upon) in response to changes in the independent variable(s). In statistics and econometrics, the term response variable can also be used.

**Derivative:** A mathematical expression used to measure how a function changes when the value of its inputs vary.

**Discrete variable:** A variable that takes a finite number of values. The terms dichotomous variable and dummy variable can also be used to describe a discrete variable.

**Dummy variable:** A variable that can assume a finite set of values. In statistical models a dummy variable can either be equal to 1 or zero. For example, in this report the presence of a transit stop at workplace is represented by a dummy variable with a value of 1 and its absence indicated by a value of zero. The terms dichotomous variable and discrete variable can also be used to describe a dummy variable.

**Elasticity:** The percentage change in a dependent (y or endogenous) variable for a given percentage change in an independent (x or exogenous) variable.

**Endogenous variable:** In economics, a variable whose value is determined within the model in which it appears. For example, in a supply and demand model, the price of a good is considered as endogenous and determined by changes in exogenous or predetermined variables. In statistics and econometrics, it refers to a variable appearing in a simultaneous equation model whose value is determined by the equations in the system (i.e., a dependent variable).

**Endogeneity:** A general term to describe a situation in which a right-hand side variable, or dependent variable, is not independent of the error term.

**Error term:** An element of a linear regression equation which contains unobserved factors that affect the dependent variable.

**Explanatory variable:** in regression analysis, a variable that is used to explain variation in the dependent variable.

**Exogeneity:** The assumption that the independent variables of a linear regression model are not correlated with the error term.

**Exogenous variable:** a variable whose value is determined outside the model. In econometric theory, the term refers to a variable that is uncorrelated with the error term in the model of interest.

**Function:** A mathematical concept to express the dependence between two quantities, one of which is given (the independent variable, or its input), and the other produced (the dependent variable, or value of the function).

**General equilibrium model:** In economics, a model that portrays the operation of many markets simultaneously; for example, the functioning of supply and demand markets.

**GIS:** Geographic information systems; the set of tools that allow the query and analysis of spatial information.

**Identification:** The condition under which a simultaneous equation system can be meaningfully estimated.

**Independent variable:** A variable whose value is deliberately manipulated to invoke a change in a dependent variable(s). In other words, “if  $x$  is given, then  $y$  occurs,” where  $x$  represents the independent variable (s) and  $y$  represents the dependent variable(s). In statistics and econometrics, the term explanatory variable is also used.

**Instrumental variable:** In statistics and econometrics it refers to an estimation method that allows consistent estimation when there are relevant explanatory variables suspected to be endogenous or in the presence of omitted variable bias.

**Instrumental variable regression:** A statistical method that addresses the omitted variable bias problem by introducing an instrumental variable into the regression function.

**Linear regression:** In statistic, it is a method that allows modeling the relationship between a dependent variable and one or more independent variables by assuming the presence of a linear relationship. A linear regression with one independent variable originates a straight line. The estimated parameter of interest measures the inclination of the straight line.

**Matrix:** In mathematics it denotes an array of numbers. Matrices are also used to describe system of equations.

**Monocentric:** In urban economics, it refers to the presence of one unique center of activities, defined as the central business district.

**Monocentric model:** In urban economics, a theoretical model that only looks at travel between home and the central business district (CBD), and where individuals locate at a distance where the marginal cost of transportation is equal to the marginal housing cost savings obtained by a move farther out from the CBD.

**Multicollinearity:** A term that refers to correlation among independent variables in a multiple regression model; this correlation leads to imprecise estimates of the parameters of interest.

**Multinomial logit model:** A regression model where the dependent variable is a discrete variable (for example, the choice of riding a bus, driving a car or walking).

**Multiple linear regression:** A model linear in its parameters, where the dependent variable is a function of two or more independent variables plus an error term. The term multivariate regression is also used a synonym.

**Omitted variable:** One (or more) variable(s) which we would like to control for, but that have been omitted from the regression model.

**Omitted variable bias:** The bias that arises in the OLS estimators when one relevant variable is omitted from the model.

**Ordinary least squares regression (OLS):** A statistical method for estimating the parameters of a linear regression model by minimizing the sum of squared residuals.

**Overidentification:** In a simultaneous equation system, the condition where the number of instrumental variables is strictly greater than the number of endogenous explanatory variables.

**Polycentric:** An urban economics term that describes an urban area as characterized by multiple employment and residential subcenters.

**Polycentric model:** A modeling approach that assumes that individuals can locate anywhere in an urban area, choosing an optimal home-work distance that optimizes also the amount of non-work travel and non-work activities.

**Residual:** The actual (or observed) value minus the predicted value of a regression model.

**Simultaneity:** A term that means that at least one explanatory variable in a multiple linear regression model is determined jointly with the dependent variable.

**Simultaneous equation model:** A model that jointly determines two or more endogenous variables, where each endogenous variable can be a function of other endogenous variables as well as of exogenous variables and an error term.

**Subcenter:** A set of contiguous tracts with significantly higher employment densities than surrounding areas.

**Trip-chaining:** Nomenclature that describes how travelers link trips between locations around an activity pattern. The transportation literature does not provide a formal definition of trip chain.

**Two-stage least squares regression (2SLS):** An instrumental variable regression technique where the instrumental variable for an endogenous explanatory variable is obtained as the fitted value from regressing all endogenous variables on all exogenous variables. This regression technique is used in the presence of simultaneity or endogeneity.

**Urban form:** Refers to a set of spatial and land-use attributes affecting travel behavior, such as employment density, population density, land-use mix, urban design, accessibility, and subcenter and central business district size.

**Vector:** A one-dimensional matrix often representing the solution of a system of linear equations.