Sustainable Metropolitan Growth Strategies:
Exploring the Role of the Built Environment

By

Mi Diao

Bachelor of Architecture, Tsinghua University (1996)
Master of Architecture, Tsinghua University (2002)
Master of City Planning, Massachusetts Institute of Technology (2006)

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Urban and Regional Planning

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author____

Department of Urban Studies and Planning
September 6, 2010

Certified by ____

Joseph Ferreira, Jr.
Professor of Urban Planning and Operations Research
Dissertation Supervisor

Accepted by____

Professor Eran Ben-Joseph
Chair, PhD Committee
Department of Urban Studies and Planning

Sustainable Metropolitan Growth Strategies:
Exploring the Role of the Built Environment

By

Mi Diao

Submitted to the Department of Urban Studies and Planning
on September 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Urban and Regional Planning

Abstract:

The sustainability of metropolitan areas has been considered one of the most significant social challenges worldwide. Among the various policy options to achieve sustainable metropolitan growth, smart-growth strategies attract increasing interests due to their financial and political feasibility. Leveraging the interconnection between land use and transportation, smart-growth strategies aim to improve urban life and promote sustainability by altering the built environment with such mechanisms as transit-oriented development, mixed-use planning, urban-growth boundary, etc. My focus in this study is to understand the role that the built environment can play in sustainable metropolitan growth. Unlike previous studies that rely primarily on household survey data in the land use-transportation research, I explore the potential for utilizing spatially detailed administrative data to calibrate urban models and support metropolitan planning.

I structure this study in three separate essays. In these essays, with several newly available fine-grained administrative datasets and advanced Database Management System (DBMS) and Geographic Information Systems (GIS) tools, I compute a set of improved indicators to characterize the built environment at disaggregated level and incorporate these indicators into quantitative models to investigate the relationships between the built environment, household vehicle usage and residential property values. I select the Boston Metropolitan Area as the study area.

The focus of the first essay is to understand the built-environment effect on household vehicle usage as reflected by the millions of odometer readings from annual vehicle safety inspections for all private passenger vehicles registered in the Boston Metropolitan Area. By combining the safety inspection data with fine-grained GIS data layers of common destinations, land use, accessibility, and demographic characteristics, I develop an extensive and spatially detailed analysis of the relationship between annual vehicle miles traveled (VMT) and built-environment characteristics. The empirical results suggest that there are significant associations between built-environment factors and household vehicle usage. In particular, distance to non-work destinations, connectivity, accessibility to transit and jobs play significant roles in explaining the VMT variations. The research findings can help analysts understand the environmental implications of alternative regional development scenarios, and facilitate the dialogue among regional planning agencies, local government and the public regarding sustainable regional development strategies.

In the second essay, I investigate the built-environment effect on residential property values with a cross-sectional analysis. The major dataset is the single-family housing transaction records from city and town assessors in the Boston Metropolitan Area assembled by the Warren Group. I use factor analysis to extract several built-environment factors from a large number of built-environment variables, and integrate the factors into hedonic-price models. Spatial econometric techniques are applied to address the spatial autocorrelation. The empirical results suggest that the transaction price of single-family properties is positively associated with accessibility to transit and jobs, connectivity, and walkability, and negatively related to auto dominance. The built-environment effects depend on neighborhood characteristics. In particular, households living in neighborhoods with better transit accessibility tend to pay a higher premium for smart-growth type built-environment features. The research findings suggest that most smart-growth strategies are positively associated with residential property values. Although built-environment characteristics advocated by smart-growth analysts do not have universal appeal to households, they no doubt satisfy an important market segment.

In the third essay, I examine the role that selectivity and spatial autocorrelation could play in valuing the built environment. Using transaction and stock data for single-family properties in the City of Boston from 1998 to 2007, I integrate a Heckman-selection model and spatial econometric techniques to account for sample selection and spatial autocorrelation, and estimate the willingness-to-pay for built-environment attributes. The empirical results suggest that the built environment can influence both the probability of sale and transaction price of properties. Failing to correct for sample selection and spatial autocorrelation leads to significant bias in valuing the built-environment. The bias might misguide policy recommendations for intervening urban development patterns and distort estimations of the value-added effect of infrastructure investment for land-value-capture programs.

Thesis Supervisor: Joseph Ferreira, Jr.
Title: Professor of Urban Planning and Operations Research

Thesis Committee Member: Karen R. Polenske
Title: Peter de Florez Professor of Regional Political Economy

Thesis Committee Member: Lynn M. Fisher
Title: Associate Professor of Real Estate

Thesis Committee Member: P. Christopher Zegras
Title: Associate Professor of Transportation and Urban Planning

Acknowledgement

I am grateful to many people that make this possible.

My heartfelt thanks go to Professor Joseph Ferreira. Joe has been mentor, teacher, friend, and incredible source of wisdom and confidence. Joe guided me through my life at MIT for seven years. I am truly honored to have had the chance to learn from him, work with him, and know him. Professor Karen Polenske has been a great source of knowledge, inspiration and motivation for me. I greatly appreciate her invaluable instructions, support, and help ever since I joined MIT in 2003. Despite her heavy research and teaching schedule, she always has time to listen to my problems and help me whenever she can. I am grateful to Professor Lynn Fisher for her generous support and informative guidance during my dissertation research and job search. I am thankful to Professor Chris Zegras. His insights helped me improve the quality of this dissertation and get a deeper understanding of the underlying land use and transportation issues.

This dissertation would not have been possible without the generous data support from MassGIS, the Warren Group and the Suffolk County Registry of Deeds. I also wish to thank Dr. Henry Pollakowski at the MIT Center for Real Estate and George Young at the Suffolk County Registry of Deeds, for their valuable advice and data assistance. Partial support for this dissertation work has come from University Transportation Center (Region One) grant, "MITR21-4: New Data for Relating Land Use and Urban Form to Private Passenger Vehicle Miles," from Martin Family Society of Fellows for Sustainability, and from MIT Portugal Program transportation focus area work on modeling transportation, land use, and environmental interactions.

I am also indebted to the DUSP staff who helped me throughout my stay at MIT. Sandy Wellford and Kirsten Greco offered distinguished administrative support to the DUSP community, from which I benefited greatly. CRON provided excellent technical and computing support, and Sue Delaney was always helpful and encouraging.

I am grateful to my friends at MIT. Their care and help are always an important source of power for me to move forward. To name just a few, Guo Zhan, Li Weifeng and Xia Jie, Zhu Yi and Ye Zi, Song Hailin and Deng Hui, Zhao Jinhua and Tan Zhengzhen, Jiang Shan and Jin Tao, Gao Lu and Lu Yu, and Kyung-Min Nam and Li Xin.

Lastly, I would like to thank my family. Without their love and patience, this dissertation would not have been possible. This dissertation is dedicated to them.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AIC: Akaike Info Criterion
BE: Built Environment
BRT: Bus Rapid Transit
CBD: Central Business District
DBMS: Database Management System
DEM: Demographic
GHG: Greenhouse Gas
GIS: Geographic Information Systems
GNP: Gross National Product
HH: Household
MAUP: Modifiable Areal Unit Problem
MBTA: Massachusetts Bay Transportation Authority
OLS: Ordinary Least Square
SC: Schwarz Criterion
TAZ: Traffic Analysis Zone
VMT: Vehicle Miles Traveled
WTP: Willingness-to-Pay

CHAPTER ONE: INTRODUCTION

In the last few decades, the growing concentration of greenhouse gas (GHG) in the atmosphere and associated negative effects of global warming are causing increasing concerns all over the world. Meanwhile, the world is undergoing the largest wave of urban growth in history. In 2008, one half of the world's population (about 3.35 billion) lives in urban areas (PRB 2008). This number is projected to swell to about 5 billion by 2030 (PRB 2008). The rapid growth of urban population underscores the critical role of metropolitan areas in global sustainability. The transportation sector represents roughly one-quarter of the world's energy-related GHG emissions (Price et al. 2006). Transportation-related challenges, such as congestions, emissions, and the exhaustion of non-renewable resources are imposing tremendous pressure on the sustainability of metropolitan areas. Various policy options aiming to reduce travel demand and achieve sustainable metropolitan growth are currently being discussed. Technology-driven approach, such as biofuel, hybrids and electric cars, can improve the fuel-efficiency of driving and reduce its carbon contribution, but it takes time and efforts. Financial (dis)incentive, such as fuel tax and congestion tolls, has proven to be an efficient tool in influencing household travel behavior, but it often faces political barriers to be implemented. In addition, many municipalities have adopted smart-growth strategies, trying to alter the physical environment that requires households to drive. None of these policy options is sufficient. We will likely need a suite of technology, policy and pricing approaches to adequately reduce transportation emissions and achieve sustainable metropolitan growth (Zegras et al. 2009).

Among these policy options, smart-growth strategies invite special interest due to their financial and political feasibility, and the potential long-term effects as they are slowly implemented and produce changes over time. Smart growth aims to improve urban life and promote sustainability by leveraging the land use – transportation interconnections and altering the built environment via such mechanisms as urban growth boundary, mixed-use planning and transit-oriented development. The major goals of these planning initiatives concentrate on two aspects: first, to promote sustainable transportation through land use planning, and, second, to encourage efficient urban development through strategic transportation investment. The coordination of land use and transportation planning is crucial in smart growth. Yet the mixed success of smart-growth strategies highlights the importance of fully understanding the complex interactions between land use and transportation, and, more generally, understanding the role that the built environment can play in sustainable metropolitan growth.

Previous studies on the land use-transportation interconnection tend to focus on two complementary relationships: the impact of the built environment on travel behavior and the impact of transportation (as part of the built environment) on development patterns. The former relationship is widely investigated in the transportation field. Most researchers find that many built-environment characteristics can significantly influence household travel behavior. However, there are still extensive debates regarding the magnitude of the built-environment effect, and whether or not it is feasible to tap this effect to reduce travel demand. For a detailed review, see Crane (2000); Ewing and Cervero (2001); Frank and Peter (2001); and Handy (1996). The latter relationship has its origin in urban economics and location theory. The classical monocentric city model developed by von Thunen (1966), Alonso (1964), Muth (1969), and Mills (1972), describes the equilibrium land-use pattern in a monocentric city. In this model, all land users benefit from increased accessibility, thus bid to be closer to the city center to save transportation cost, which leads to a zonal distribution of land uses around the center. Analysts widely believe that the transportation system could influence urban development in terms of location choice, property value, or characteristics of development.

The majority of previous studies on the land use-transportation interconnection, especially those focusing on the built-environment effect on travel behavior, rely on household surveys to carry out empirical analyses, because survey data provide detailed description of demographic, residence and travel attributes to support modeling. However, this approach has several drawbacks. The high expense of individual surveys tends to limit the sample size and frequency of surveys – commonly they are limited to a few thousands observations and are updated every 5-10 years. Privacy concerns often limit the geographic specificity with which details about residence and trips can be revealed. Accordingly, in planning practice, planning agencies have lacked the data and the analytic techniques needed to make informed decisions in both long-term planning to achieve sustainable metropolitan growth and short-term reaction to make the city more responsive to real time changes.

Thanks to the rapid development of spatial data infrastructure, planning agencies are entering an era in which a large volume of administrative data with spatial details are available, for example, vehicle safety-inspection records from the Registry of Motor Vehicles, housing-transaction records from the Registry of Deeds, housing-assessment records from the Assessing Department, transit-fare card information from the transit agency, and cell phone-usage records from mobile companies. These administrative datasets have distinct advantages over the traditional survey data used in land-use transportation research: large temporal and spatial coverage, continuous data flow, low marginal cost, accuracy, automatic collection and central storage, etc. Due to these unique features, there exists a great potential for utilizing such novel datasets to support metropolitan planning and promote sustainable growth. Meanwhile, advanced data manipulation and analysis methodologies and techniques are required before the full value of administrative data can be realized.

My primary objective in this study is to investigate the bidirectional relationships between land use and transportation, and understand the role that the built environment can play in sustainable metropolitan growth. Unlike previous studies relying on household survey data, I explore the potential of utilizing administrative data to calibrate urban models and support metropolitan planning with the help of advanced information technologies such as Database Management System (DBMS) and Geographic Information System (GIS) tools.

The main body of the study comprises three separate essays. The first essay focuses on the impact of the built environment on household vehicle usage. The second and third essays focus on the impact of the built environment on residential property values. In these essays, with several newly-available, fine-grained administrative datasets and advanced DBMS and GIS tools, I compute a set of improved indicators to characterize the built environment at a disaggregated level and incorporate these indicators into quantitative models to investigate the relationships between the built environment, household vehicle usage and residential property values. I select the Boston Metropolitan Area as the study area.

The first essay examines the built-environment effect on household vehicle usage using the millions of odometer readings from annual vehicle safety inspections for all private passenger vehicles registered in the Boston Metropolitan Area. By combining the safety inspection data with fine-grained GIS data layers of common destinations, land use, accessibility, and demographic characteristics, I develop an extensive and spatially detailed analysis of the (cross-sectional) relationship between annual vehicle miles traveled and built-environment characteristics. The research findings of the first essay could help us understand the environmental implications of alternative regional development scenarios and facilitate the dialogue between regional planning agency, local government, and the public regarding sustainable metropolitan growth.

In the second essay, I investigate the built-environment effect on residential property values with a cross-sectional analysis. The major dataset is the single-family housing transaction records from city and town assessors in the Boston Metropolitan Area assembled by the Warren Group. I use factor analysis to extract five built-environment factors from a large number of built-environment variables, and integrate the factors into hedonic-price models. I apply spatial econometric techniques to account for spatial autocorrelation effects. This study can help understand the property-value effect of land use change and assess the impact of smart growth on local neighborhoods.

In the third essay, I explore the impact of selectivity and spatial autocorrelation in valuing the built environment, using the transaction records from the Suffolk County Registry of Deeds and the assessing records from the Boston Assessing Department for single-family properties in the City of Boston. I apply the Heckman two-step procedure (Heckman 1976) to correct for sample selection bias and integrate spatial econometric techniques into the Heckman-selection model to solve for spatial autocorrelation. I further investigate the magnitude of the bias caused by sample selection and spatial autocorrelation by comparing the willingness-to-pay for the same built-environment attribute computed from conventional hedonic-price model and Heckman-selection models. This bias might misguide policy recommendations for impacting urban development patterns and distort estimations of the value-added effect of infrastructure investment for land value capture programs.

In Chapter 2, I describe the methodology and outcomes to quantify the built environment in the Boston Metropolitan Area, which will be used in all three essays. To avoid redundancy, I make this part a separate chapter. I use Chapter 3, 4, and 5 to present the three research essays respectively. Finally, in Chapter 6, I summarize the research findings and discuss policy implications and future research directions.

CHAPTER TWO: MEASURING THE BUILT ENVIRONMENT IN THE BOSTON METROPOLITAN AREA

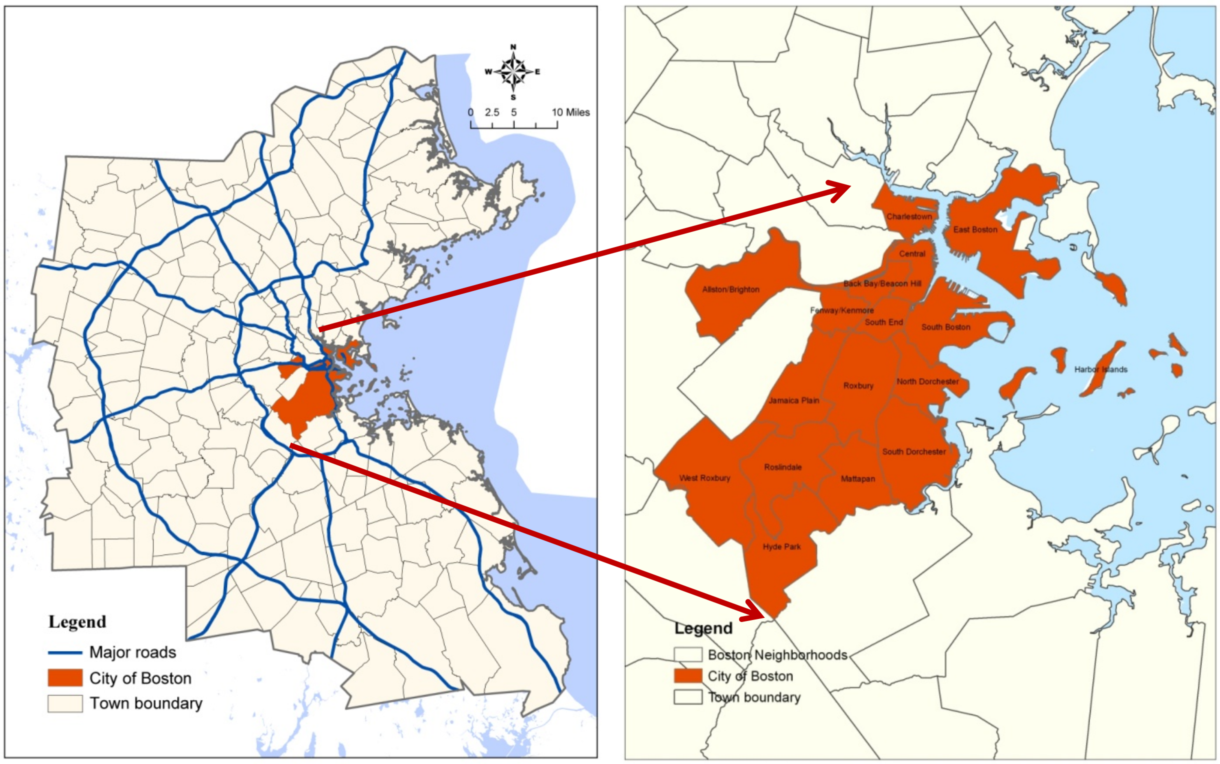
One prerequisite to model the built environment is to quantify it. In this chapter, I present the datasets, methodology, and variables to describe the built environment in the Boston Metropolitan Area. To deal with the potential multicollinearity among built-environment variables, I apply factor analysis to reduce the large set of built-environment variables to several factors to explain source of spatial differentiation within the Metro.

2.1 BUILT-ENVIRONMENT DATASETS AND SPATIAL UNIT OF ANALYSIS

I select the Boston Metropolitan Area as the study area. Metro Boston exhibits a rich set of built-environment characteristics, which makes it a compelling case for empirical analyses. Figure 1 maps Metro Boston and the City of Boston.

Metro Boston

City of Boston



Source: The author

Figure 1: Metro and City of Boston

In describing the built environment of Metro Boston, I benefit from a set of built-environment datasets with exceptional spatial detail, which are mainly from MassGIS, the State's Office of Geographic and Environmental Information. MassGIS utilized Dun and Bradstreet business location database to locate household non-work destinations, and geocoded these businesses to a point layer, which were then aggregated by business category into business counts within each 250x250m grid cell. Institutional destinations, such as schools, hospitals, and parks, exist as independent data layers developed and maintained by MassGIS. The road-inventory database with detailed information on road networks in the region is from the Massachusetts Department of Transportation. MassGIS generated population and household data from the 2000 Census, constrained them to those areas identified as residential by the 2000 land use dataset, and assigned them to 250mx250m grid cells.

The Modifiable Areal Unit Problem (MAUP) is a well-known challenge in studies on spatial phenomena, which may lead to inconsistency in measurement results and statistical analyses. Zhang and Kukadia (2005) summarize three commonly recognized approaches to resolve the MAUP issues: (a) use disaggregate data where possible; (b) report scope and magnitude of the MAUP; and (c) use behavior-based selection of scale and areal unit definition. Robsen identifies the grid cell approach as a possible means to mitigate the MAUP (Robsen 1969). To deal with the MAUP, the spatial unit used in this study is a 250x250m grid cell layer developed by MassGIS. A grid cell contains an area just over 15.4 acres, which is sufficiently small to capture spatial details and neighborhood effects. Meanwhile, using the grid cell as a basic study unit, I can take advantage of powerful raster analysis tools in GIS software. For each grid cell, I define a catchment area (neighborhood) as the 3x3 nearest grid cells, compute the variable of interest for the catchment area, and assign the value to the grid cell in the middle. The 750x750m catchment area has a size that is close to the "transportation impact area", which is conventionally defined as a circle with a 1⁄4-mile radius, a size that has been backed by behavioral and empirical research (Untermann 1984). The employment of a catchment area also helps create a smooth surface, reducing noise in the raw data.

Compared with previous research, my study is performed at a much more fine-grained scale. Table 1 compares the grid cells I use and some spatial units that are widely used in land use and transportation research for Metro Boston.

Table 1: Comparison of Spatial Units for Metro Boston

	Grid Cell	TAZ	Block Group	Census Tract
No. of observations	119,834	2,727	3,323	894
No. of observations with population	73,714	2,606	3,319	894
Vehicle count for populated units				
Min	0	0	1	1
Max	3,117	3,022	11,593	13,631
Mean	32	941	744	2,764
Std. Dev.	49	603	514	1,514
Household count for populated units				
Min	0	0	0	0
Max	1,624	2,318	2,211	4260
Mean	22	631	495	1,839
Std. Dev.	48	391	246	713
Individual count for populated units				
Min	1	1	2	70
Max	3,673	4,969	6,131	12,051
Mean	58	1,654	1,297	4,817
Std. Dev.	112	992	626	1,825

Source: Calculated by the author.

2.2 BUILT-ENVIRONMENT VARIABLES

For this study, I computed 27 built-environment variables. Because spatial distribution of destinations can significantly influence travel costs, accessibility to common destinations is an important determinant of vehicle usage and properties values. I compute a gravity-type measure of job accessibility at the TAZ level to represent work distance, which takes the following form known as the Hansen accessibility model (1959). I assign each grid cell the value of the TAZ that it belongs to.

- Job accessibility: $A_j = \sum_j O_j f(C_{ij})$, where $f(C_{ij}) = \exp(-\beta * C_{ij})$; O_j is the number of jobs in TAZ j; $f(C_{ij})$ is an impedance function; C_{ij} is the network distance between

TAZ i and j; β is set to 0.1, based on Zhang's calibration using an Activity–Travel Survey conducted by the Central Transportation Planning Staff for the Boston region (2005).

MassGIS computed distances to a variety of non-work destinations at a 250m*250m grid cell level using GIS tools. I select eight types of the most important non-work destinations based on average trip rate from the 2001 National Household Transportation Survey, including:

- Distance to shopping mall: Euclidian distance to the nearest shopping mall
- Distance to grocery store: Euclidian distance to the nearest grocery store
- Distance to school: Euclidian distance to the nearest school
- Distance to hardware store: Euclidian distance to the nearest hardware store
- Distance to restaurant: Euclidian distance to reach at least 4 restaurants
- Distance to church: Euclidian distance to reach at least 4 churches
- Distance to dentist: Euclidian distance to reach at least 4 dentists
- Distance to gym: Euclidian distance to reach at least 4 gyms

Other built-environment variables describe density, land-use mix, road networks, transit proximity, and pedestrian environment, respectively. They also have the potential to affect travel costs for different travel modes. Among them, I computed distance-related variables directly for the target grid cell. I computed other measures for the 9-grid-cell catchment area and then assigned the value to the target grid cell.

- Population density: population/residential area
- Land-use mix: the land-use mix measure is based on the concept of entropy — a measure of variation, dispersion or diversity (Turner, Gardner and O'Neill, 2001). In

the first step, I compute it for each grid cell, using $-\sum_j P_j * \ln(P_j)/\ln(J)$, where P_j is the proportion of land in the jth land-use category and J is the total number of land-use categories considered. In this study, J=5: single family, multi-family, commercial, industrial, and recreation and open space. A value of 0 means the land in the grid cell is exclusively dedicated to a single use, while a value of 1 suggests perfect mixing of the five land uses. Then, I assign each grid cell the average value of the nine grid cells in the catchment area.

- Intersection density: number of intersections / area
- Density of 3-way intersections: number of 3-way intersections / area
- Density of 4-way intersections: number of 4-way intersections / area
- Road density: total length of road / area
- Percent of 4-way intersections: number of 4-way intersections / number of intersections
- Percent of roads with access control: total length of road with access control / total road length
- Average road width: ∑(width of road segment * length of road segment) / total road length
- Percent of roads with over 30-mph speed limit: total length of road segment with over 30-mph speed limit / total road length
- Distance to highway exit: Euclidian distance to the nearest highway exit
- Percent of roads with curbs: total length of road segment with curbs / total road length

- Percent of roads with sidewalks: total length of road segment with sidewalks / total road length
- Average sidewalk width: ∑(sidewalk width of road segment * length of road segment) / total road length
- Distance to subway station: Euclidian distance to the nearest subway station
- Distance to commuter rail station: Euclidian distance to the nearest commuter rail station
- Distance to MBTA bus stop: Euclidian distance to the nearest MBTA bus stop
- Distance to MBTA parking lot: Euclidian distance to the nearest MBTA parking lot

I use GIS techniques and database management tools extensively in the computation of these built-environment variables.

2.3 FACTOR ANALYSIS FOR BUILT-ENVIRONMENT VARIABLES

Due to the multi-dimensional nature of the built environment, one central issue in studies of the built environment is the selection of relevant variables from a large set of potentially important variables. Furthermore, many built-environment variables tend to be closely correlated. For example, relatively dense neighborhoods tend to have a greater variety of land uses, smaller blocks, and so on. A regression model with highly correlated variables is likely to result in numerous insignificant or incorrectly-signed coefficients. To deal with the multicollinearity, I use factor analysis to reduce the total number of built-environment variables to a small set of factors and include factor scores in regression models. The idea is that the multicollinearity between variables exists because they are indicators of common factors, and that these underlying factors are important determinants. As linear combinations of the built-environment variables, built-environment factors represent these latent underlying forces. For example, factor analysis allows variables like “average sidewalk width”, “percent of roads with curbs”, and “percent of roads with sidewalks” to be linearly combined to represent a dimension that we might call “walkability”.

I perform a principle component analysis with Varimax rotation using the 27 built-environment variables. The top 5 factors with initial eigenvalues greater than 1 explain 69.8% of variance in original variables. In other words, there is only a 30% loss in information incurred by the 82% reduction in the number of built-environment variables from 27 to 5. Factor loadings for built-environment variables are presented in Table 2.

TABLE 2: Factor Loadings of Built-Environment Factors

	Variables	Factor 1 Distance to non-work destinations	Factor 2 Connectivity	Factor 3 Inaccessibility to transit and jobs	Factor 4 Auto dominance	Factor 5 Walkability
1	Distance to restaurant	0.784				
2	Distance to mall	0.764				
3	Distance to hardware store	0.746				
4	Distance to grocery	0.733				
5	Distance to dentist	0.688		0.398		
6	Distance to gym	0.676				
7	Distance to church	0.674				
8	Distance to school	0.645				
9	Land-use mix	-0.480				
10	Density of 4-way intersections		0.872			
11	Intersection density		0.849			
12	Density of 3-way intersections		0.809			
13	Population density		0.785			
14	Road density	-0.353	0.765			
15	Percent of 4-way intersections		0.609			
16	Distance to bus stop			0.833		
17	Distance to commuter rail station			0.810		
18	Distance to subway station			0.801		
19	Distance to MBTA parking lot			0.775		
20	Job accessibility		0.486	-0.636		
21	Percent of roads with access control				0.910	
22	Average road width				0.875	
23	Percent of roads with 30+ speed limit				0.856	
24	Distance to highway exit				-0.362	
25	Percent of roads with sidewalks					0.910
26	Percent of roads with curbs					0.908
27	Average sidewalk width		0.583			0.602

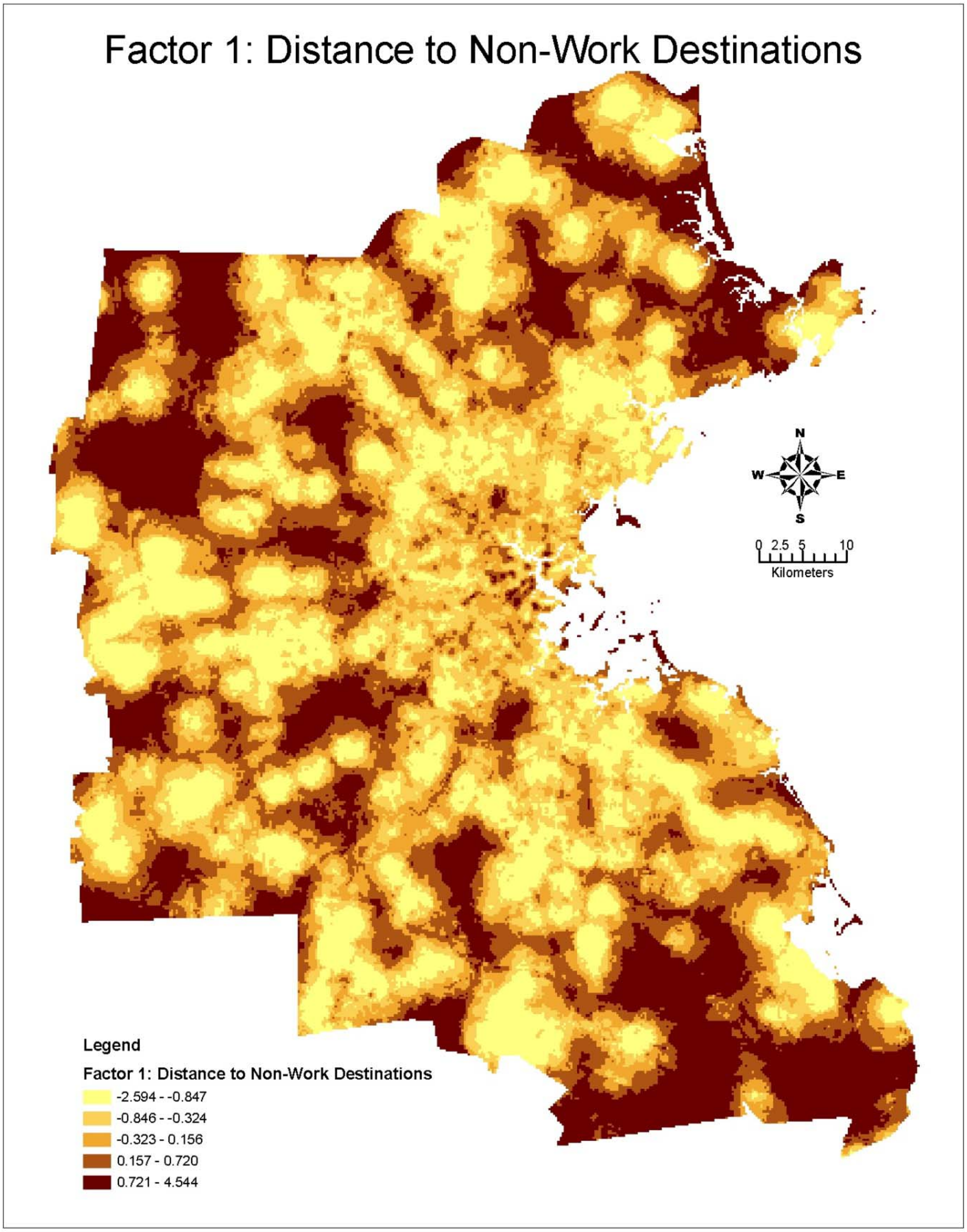
* I suppress factor loadings with absolute value less than 0.35 for interpretation convenience.
Source: Calculated by the author using SPSS.

Factor 1 has high loadings on variables for distance to non-work destinations and land-use mix, and therefore describes primarily “distance to non-work destinations”. Grid cells with higher scores in factor 1 tend to have longer distance to non-work destinations, and thus are hypothesized to have higher VMT (others factors held constant). Factor 2 places the highest weights on street network layout and population density. I label it as "connectivity". Good connectivity can improve the connection of people and places and shorten local trips (Crane 1996), thereby reducing household vehicle usage. Factor 3 describes the difficulty of accessing transit systems and jobs, with positively high loadings on distance to transit variables and negatively high loading on job accessibility. Factor 3 could be positively associated with VMT. Factor 4 leans to the traffic management side, representing the degree of auto dominance, that is, the extent to which automobile movement is facilitated in the locality. It could decrease travel costs of the auto mode, thus increasing vehicle usage. The fifth factor "walkability" describes the pedestrian environment, which can reduce the travel costs of walking, thus decreasing VMT. Figures 2 - 6 show the spatial patterns of built-environment factors. Compared with grid cells in the suburbs, grid cells in urban centers have better accessibility to non-work destinations, jobs, and transit systems, better connectivity, and better pedestrian environment as expected[1]. Grid cells with higher scores in the "auto dominance" factor tend to be located along major transportation corridors. Note the extent to which the factors vary from one another and spatially at different local and regional scale.

The built-environment indicators computed in this chapter will be integrated into quantitative models in the following three chapters to investigate the impact of the built environment on household vehicle miles traveled and residential property values.

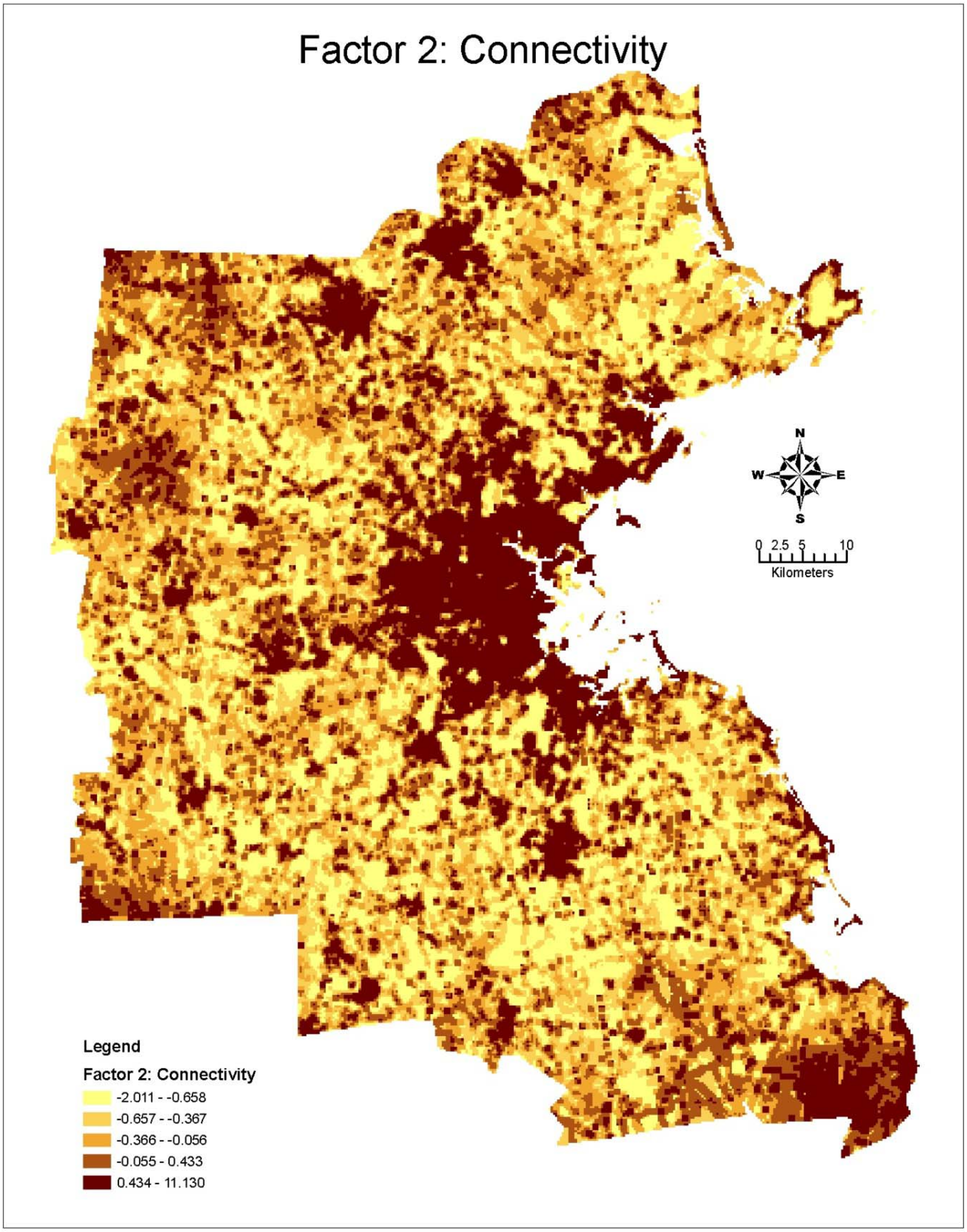
---

[1] It should be noted that Figure 4 shows some boundary effect in the "inaccessibility to transit and jobs" factor. The boundary effect may influence statistical results and will be further discussed in Chapter 3.



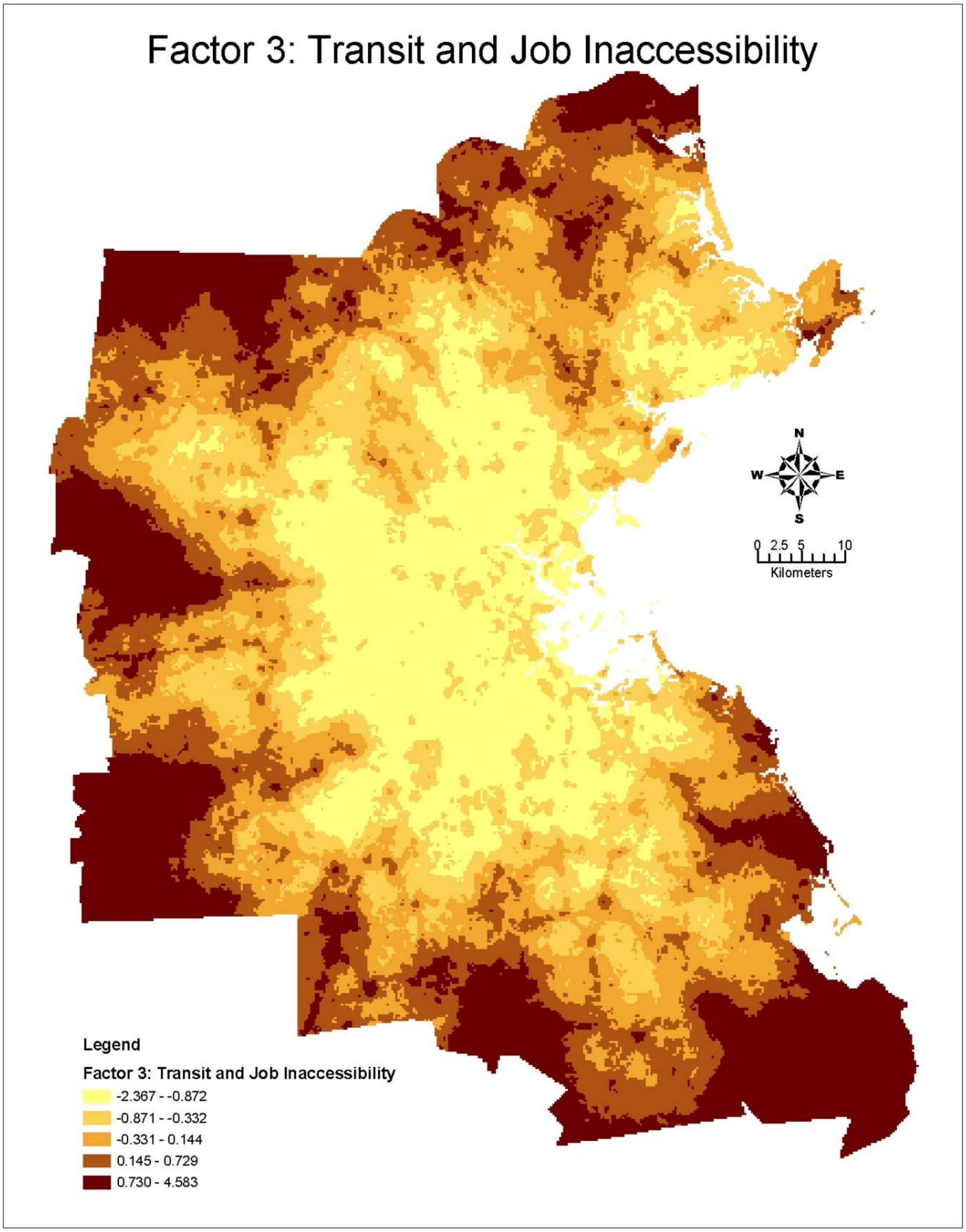
Source: The author

Figure 2: Metro Boston Built-Environment Factors – Distance to Non-Work Destinations



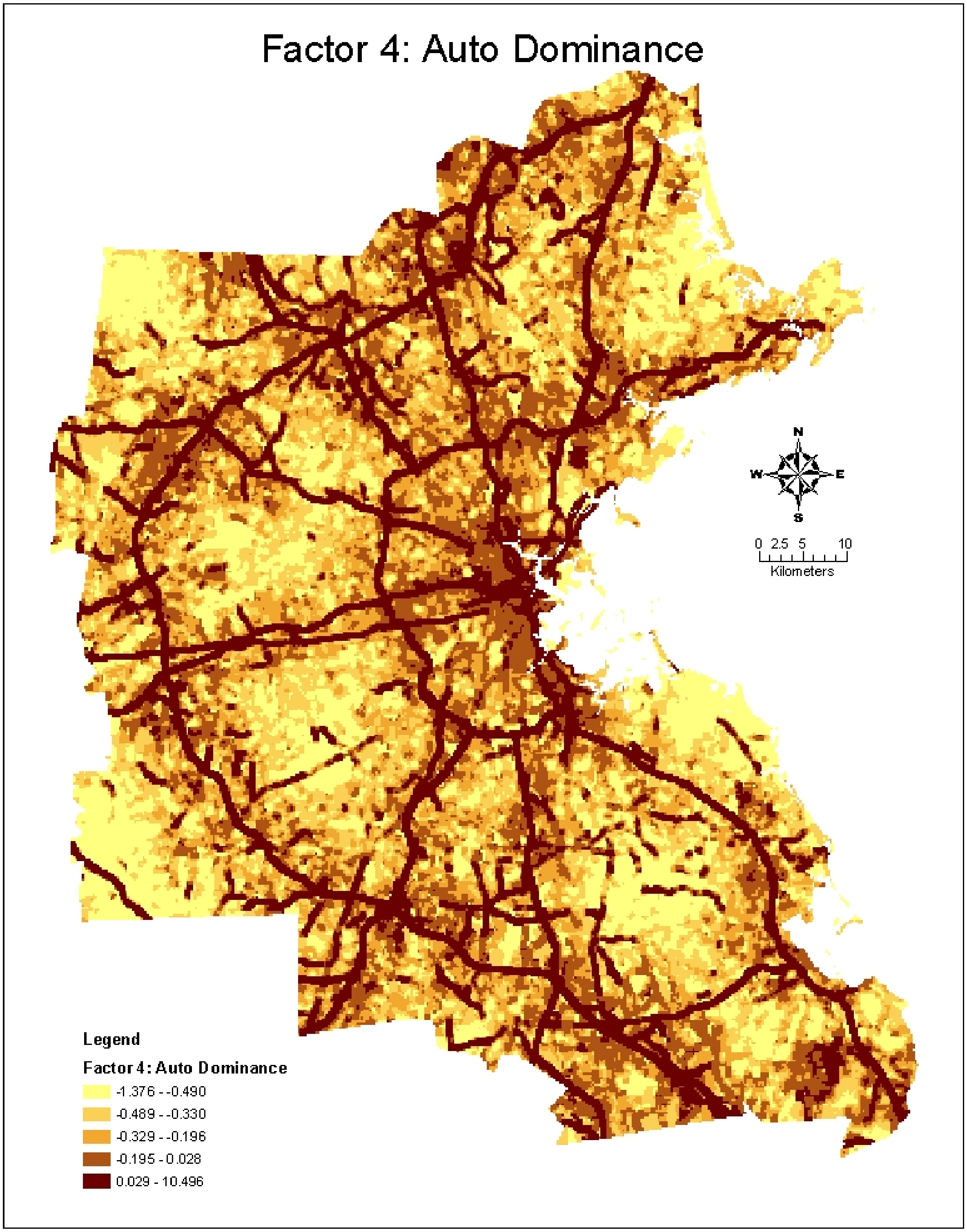
Source: The author

Figure 3: Metro Boston Built-Environment Factors - Connectivity



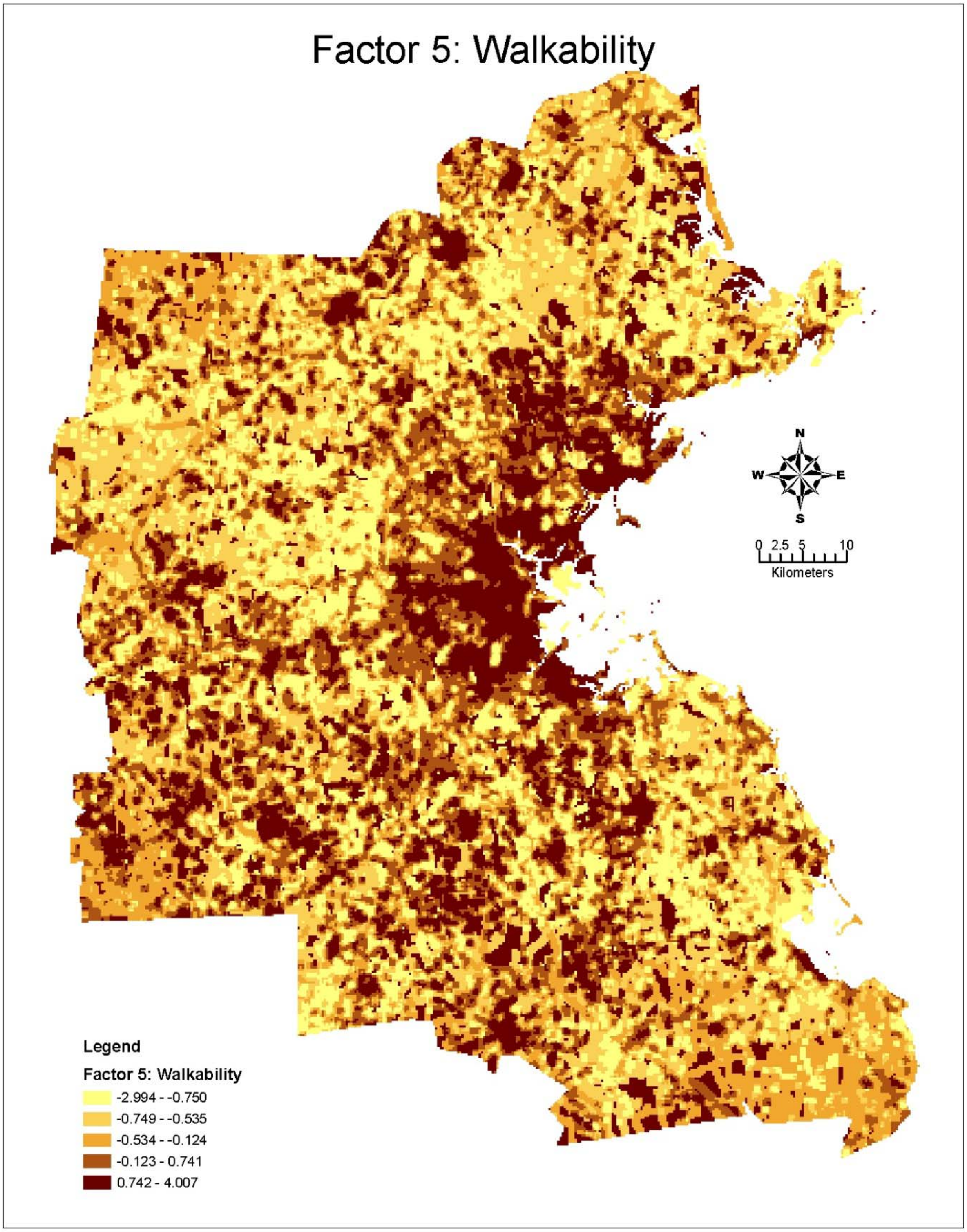
Source: The author

Figure 4: Metro Boston Built-Environment Factors – Inaccessibility to Transit and Jobs



Source: The author

Figure 5: Metro Boston Built Environment Factors – Auto Dominance



Source: The author

Figure 6: Metro Boston Built-Environment Factors - Walkability

CHAPTER THREE: VEHICLE MILES TRAVELED AND THE BUILT ENVIRONMENT: EVIDENCE FROM VEHICLE SAFETY INSPECTION DATA

3.1 INTRODUCTION

In the last few decades, the rapid growth of Greenhouse Gas (GHG) concentration in the atmosphere and associated negative effects of global warming are causing increasing concerns about the sustainability of the world. The transportation sector is currently responsible for one-quarter of the world's energy-related GHG emissions (Price et al. 2006), and personal mobility consumes about two thirds of the total transportation energy use (IEA 2004). As an important source of GHG emissions, transportation plays a critical role in the global efforts to achieve sustainable development. Multiple strategies to reduce transportation energy use and emissions are currently explored by analysts and policy makers, such as fuel-efficient vehicles, financial (dis)incentives, and various smart-growth policies. Among these policy options, smart-growth policies invite special interests due to their financial and political feasibility.

Central to smart-growth strategies is leveraging the interconnection between the built environment and travel behavior to reduce travel demand. The built environment comprises urban design, land use, and the transportation system, and encompasses patterns of human activity within the physical environment (Handy et al. 2002). Smart-growth policies try to reshape household travel behavior by changing the built environment via such mechanisms as regional planning, zoning, and provisions of alternative transportation modes.

The relationship between transportation and the built environment has long been studied and is recognized as complex, as reviewed in Handy (1996), Boarnet and Crane (2000), Crane (2000), Ewing and Cervero (2001), and Frank and Engelke (2001). There continues to be debates about whether the relationship is “strong” or “weak” (Krizek 2005). Household or individual-based survey data (for sampled individuals and households) are the preferred instrument for empirical analysis of travel behavior since the unit of analysis, an individual, can be readily associated with the mode availability, travel cost, demographic factors, and built-environment measures. However, the high expense of individual travel surveys tends to limit sample sizes, and privacy concerns often limit the geographic specificity with which trip origins and destinations can be revealed. These issues constrain the capability of survey-based studies in providing confidence in statistical accuracy at the neighborhood level.

Another line of research characterizes both the built environment and travel using aggregate measures. Newman and Kenworthy (1999) analyze the relationship between density and energy use for an international sample of cities and find significant negative correlation between density and energy use. However, besides the fundamental problem of comparing places with different cultural, political, and historical contexts, their study is also criticized for its use of simple measures of urban form and travel (Handy 1996). Holtzclaw (1994) uses odometer reading data from biennial auto emission inspections to derive estimates of total travel for 28 zip code zones in California and relate them to built-environment measures. The result shows that annual vehicle miles traveled is significantly associated with neighborhood density. Miller and Ibrahim (1998) carry out an empirical investigation into the relationship between the built environment and automobile travel at traffic analysis zone (TAZ) level in the Greater Toronto Area. They find that zonal VMT per worker increases with increasing distance from the CBD, and/or other major employment zones within the urban area. Holtzclaw et al. (2002) use socio-demographic variables to control for population differences across different zones and find that auto ownership and mileage per car are functions of neighborhood urban design and socio-economic characteristics in the Chicago, Los Angles, and San Francisco.

The aggregate approach has provided promising evidence of the potential effectiveness of smart-growth policies in reducing travel demand (Handy 1996). However as many researchers have suggested, this approach also has significant shortcomings: (1) It does not allow for an exploration of underlying factors and the mechanisms by which the built environment influences individual decisions; (2) The zones used in previous aggregate studies are usually very large in size. For example, Newman and Kenworthy (1999) use city-level data in their study and Holzclaw et al. (2002) use zip-code-zone as their unit of analysis. At such an aggregated level, the intra-zone variations of the built environment and demographic measure could be too large to ignore; (3) Previous studies either omit or include very few demographic variables in their statistical analyses, thus make limited effort to control the residential self-selection problem and construct causal relationships (Brownstone 2008); and (4) spatial autocorrelation may affect the results significantly but analysts neglected this effect.

In this study, I take advantage of a newly-available unique dataset, the odometer readings from annual safety inspections for all private passenger vehicles registered in Metro Boston to develop an extensive and spatially-detailed analysis of the built environment and household vehicle usage. I use Vehicle Miles Traveled (VMT) as the primary variable of interest, which is a convenient measure that reduces the multi-dimensional travel demand (number of trips, the spatial distribution of these trips, the modes and routes chosen to execute these trips) to a single variable (Miller and Ibrahim 1998). The basic spatial unit for my analysis is a statewide 250 meter (m) by 250m grid-cell layer developed by MassGIS, the State's Office for Geographic and Environmental Information. I perform multivariate regression analyses at the grid cell level to identify built-environment and demographic factors that are significantly associated with household vehicle usage. Spatial econometric techniques are applied to account for potential spatial autocorrelation.

Given the nature of my analysis, I raise two cautions at the outset. First, my objective is not to project the impact of a given policy on vehicle usage, which requires a dynamic model of land use-transportation interaction (Miller and Ibrahim 1998). My more modest objective is to examine the spatial distribution of travel behavior within a metropolitan area, which can be seen as the outcome of this dynamic land use-transportation process, and to clarify the irreducible spatial components of household travel behavior. The second issue concerns the ecological fallacy. In particular, I focus on the spatial patterns of the relationship between the built environment and household vehicle usage. Even though I use small grid cells (of 15.4 acres each) as the basic spatial unit, they measure behavior aggregated across multiple households in the grid cell. Hence, the underlying factors and the behavior mechanisms by which the built environment influences individual decisions cannot be revealed by my study.

3.2 STUDY AREA AND DATA

I select the Boston Metropolitan Area as the study area for the empirical analyses. Metro Boston exhibits a variety of built-environment characteristics, which makes it a compelling case for the study.

In this study, I use a unique VMT dataset, the annual vehicle safety inspection records from the Registry of Motor Vehicles (RMV) to estimate annual mileage for every private passenger vehicle registered in Metro Boston. Safety inspection is mandated annually beginning within one week of registering a new or used vehicle. The safety inspection utilizes computing equipment that records vehicle identification number (VIN) and odometer reading and transmits these data electronically to the RMV where they can be associated with the street address of the place of residence of the vehicle owner. MassGIS obtained access to the safety inspection records from the RMV for a “Climate Roadmap” project that details possible plans for significant reductions in GHG emissions for 2020-2050 in Massachusetts. MassGIS compared the two recent vehicle inspection records for all private passenger vehicles, calculated the odometer reading difference, and pro-rated it based upon the time period between inspection records so as to reflect the estimated annual mileage traveled. MassGIS then geocoded each vehicle to an XY location approximating the owner's address using GIS tools, and tagged each VIN with the 250x250m grid cell ID containing that address. MassGIS then provided the VINs, XY locations, and grid cell IDs, to MIT for use in our research. Overall, 2.47 million private passenger vehicles are included in this dataset. Among them, 2.10 million vehicles (84.9%) have credible odometer readings. For the remaining 0.37 million vehicles, I know their places of garaging but do not have reliable odometer readings, either because the reported reading was determined to be in error or because two readings sufficiently far apart were not available.

Although this dataset lacks individual trip details, it does provide a very high percentage sample of total passenger vehicle miles traveled. Furthermore, unlike travel surveys, this dataset does not depend on the subjects' willingness or ability to remember and report their driving. The Energy Information Administration (EIA)'s 1994 Residential Transportation Energy Consumption Survey shows that self-reported VMT values are 13 percent greater than odometer-based VMT in urban areas. EIA suggests that odometer-based VMT should be obtained if possible (Schipper and Moorhead 2000). Holtzclaw et al. (2002) use a similar dataset in their study, odometer readings from auto emission inspections (smog check), but since California exempts new vehicles from smog checks for the first two years, their measure systematically biases VMT downwards for zones with large numbers of new vehicles (Brownstone 2008).

My study also benefits from built-environment data with exceptional spatial detail, which are mainly from MassGIS. Detailed descriptions about the datasets and the spatial unit to compute built-environment measures can be found in Chapter 2.

3.3 METHODOLOGY

In this section, I present the methodology employed in this study.

3.3.1 Model Specifications

In the base model, I specify VMT as a function of built-environment and demographic factors.

$$VMT_i = \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i \qquad (1)$$

where VMT_i is the zonal average VMT per vehicle, per household or per capita for the catchment area of grid cell i; BE_i is a vector of built-environment variables of grid cell i, and DEM_i is a vector of demographic variables of the block group that grid cell i falls in.

Many previous analysts (e.g., Ewing and Cervero 2001) suggest that built environment can influence travel behavior. This effect can be partitioned into direct influences associated with the characteristics of the neighborhood where the household locates and indirect influences associated with the travel behavior and built-environment characteristics of neighboring areas. I estimate both spatial lag model and spatial error models (Anselin 1993) to capture this spatial effect. Spatial lag suggests a possible diffusion process -- VMT of one place is affected by the independent variables of this place as well as neighboring areas. With spatial lag in an OLS regression, the estimation result will be biased and inefficient. Spatial error is indicative of omitted independent variables that are spatially correlated. With spatial error in an OLS regression, the estimation result will be inefficient. The spatial lag model can be specified as:

$$VMT_i = \rho W_{VMT_i} + \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i \tag{2}$$

where ρ is a spatial-lag correlation parameter, and ε is an Nx1 vector of i.i.d. standard normal errors. The spatial error model can be specified as:

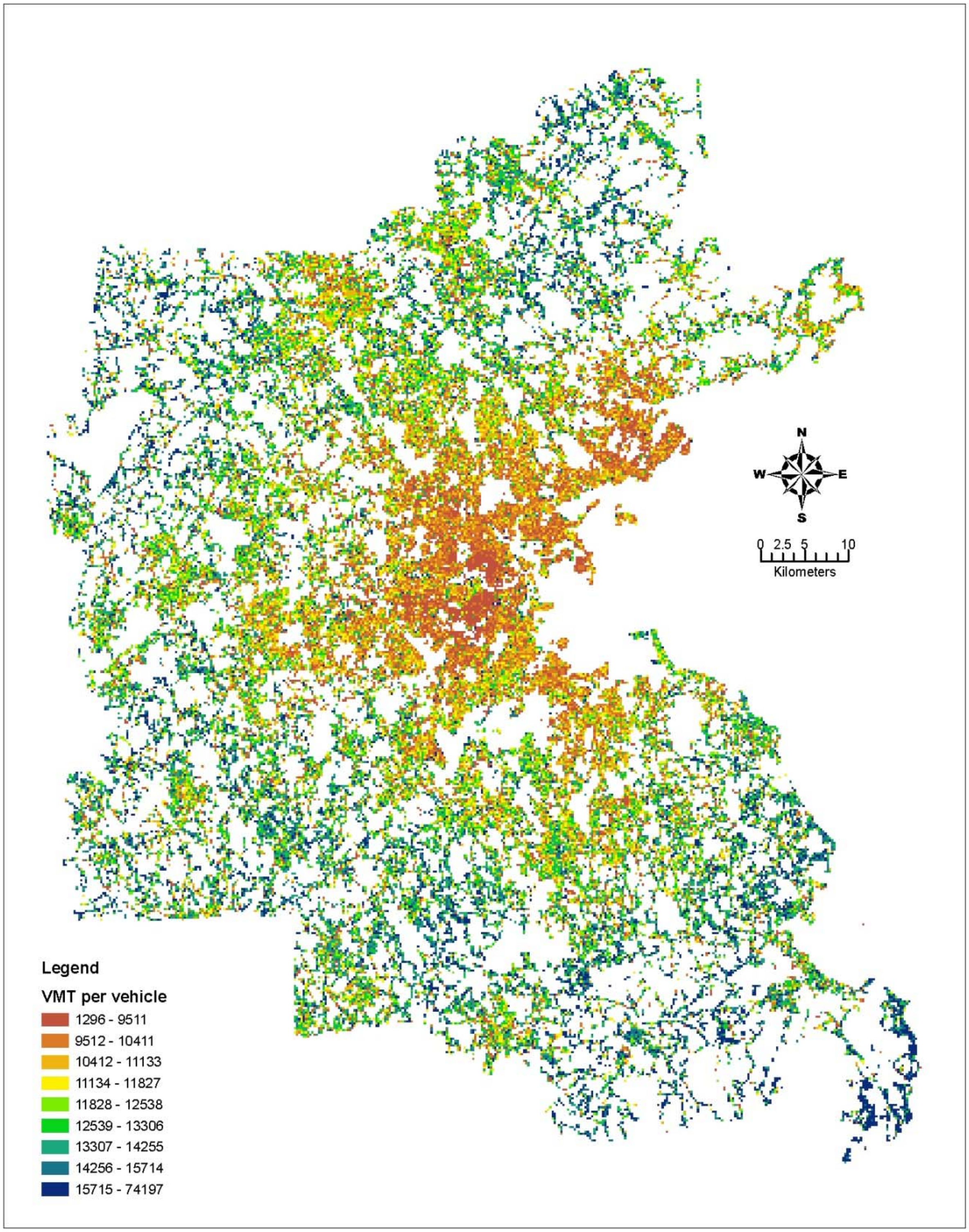
$$\begin{aligned} VMT_i &= \sum \alpha_j BE_{ij} + \sum \beta_k DEM_{ik} + \varepsilon_i \\ \varepsilon_i &= \lambda W_{\varepsilon_i} + \mu_i \end{aligned} \tag{3}$$

where λ is a spatial-error correlation parameter, and μ is an Nx1 vector of i.i.d. standard normal errors.

In Equations (2) and (3), W is the NxN matrix of spatial weights, which I developed assuming a constant spatial dependence among grid cells up to a maximum distance. I used the maximum Euclidean distance of 750m. Both models can be estimated by maximum likelihood.

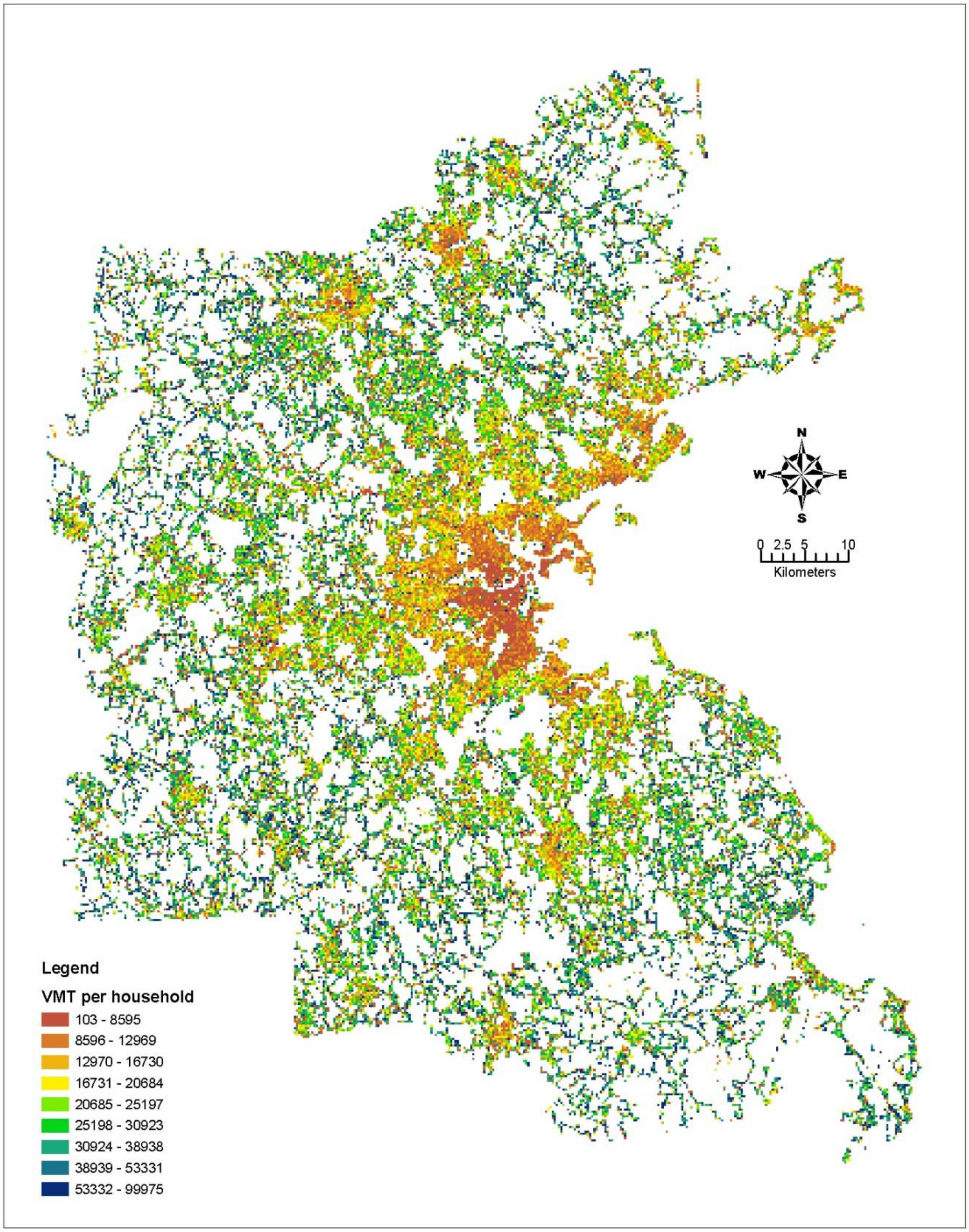
3.3.2 VMT Variables

In this study, I explore the built-environment effects on three VMT measures: (1) VMT per vehicle, (2) VMT per household, and (3) VMT per capita. VMT per vehicle is a single indicator of individual car usage, while VMT per household and VMT per capita are also influenced by auto ownership. I compute the VMT per vehicle for each grid cell based on vehicle-level annual mileage estimates from MassGIS. Some grid cells have very few vehicles. I apply the spatial interpolation function of GIS software to overcome issues related to sparse cells. For grid cells that have at least 12 vehicles with credible odometer readings (denoted as “good” cars), I assign the zonal average annual mileage of all “good” cars to the grid cell. For grid cells with 1-11 “good” cars, I assign the inverse distance weighted average of 12 closest “good” annual mileages to the grid cell. I compute VMT per household (VMT per capita) for each grid cell by multiplying the estimated VMT per vehicle within the grid cell by total number of vehicles within the grid cell then dividing by number of households (individuals). These odometer-readings-based VMT estimates provide a more accurate and reliable picture of household vehicle usage than survey-based self-report VMT estimates, establishing a baseline for tracking future changes in vehicle usage and associated energy consumptions and emissions for Metro Boston. Figures 7 - 9 plot VMT per vehicle, VMT per household and VMT per capita across grid cells in Metro Boston respectively, using quantile classification method and 9 categories. The overall spatial pattern is what analysts would expect: VMT are lower in grid cells near urban centers, but higher in suburban areas. It is also interesting to note that there is: (a) a large area in suburbs without vehicles or households; (b) a significant variability within suburbs depending on whether the grid cell is near the town center; and (c) the difference in patterns between VMT per vehicle and VMT per household.



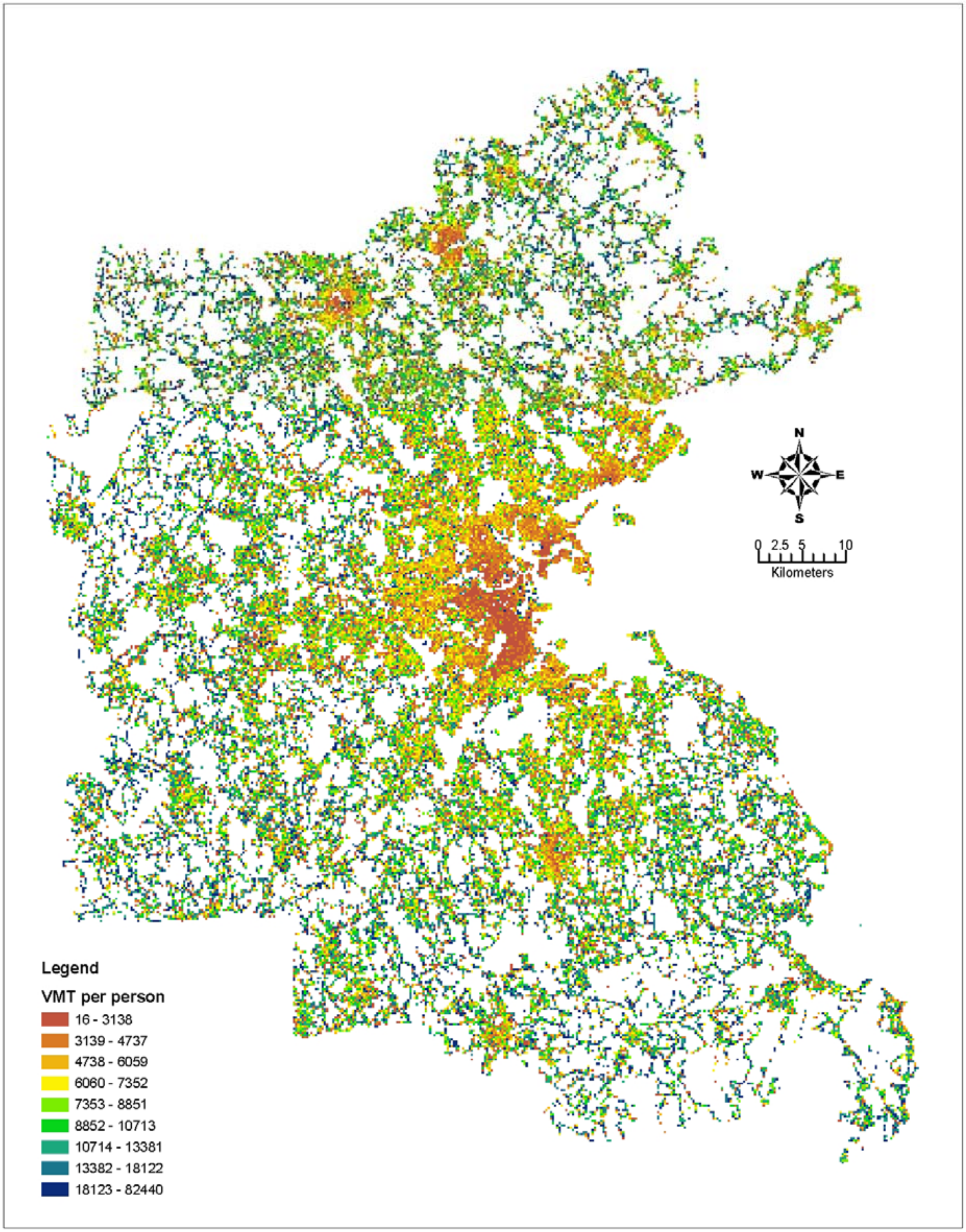
Source: The author

Figure 7: VMT per Vehicle across Grid Cells in Metro Boston



Source: The author

Figure 8: VMT per Household across Grid Cells in Metro Boston

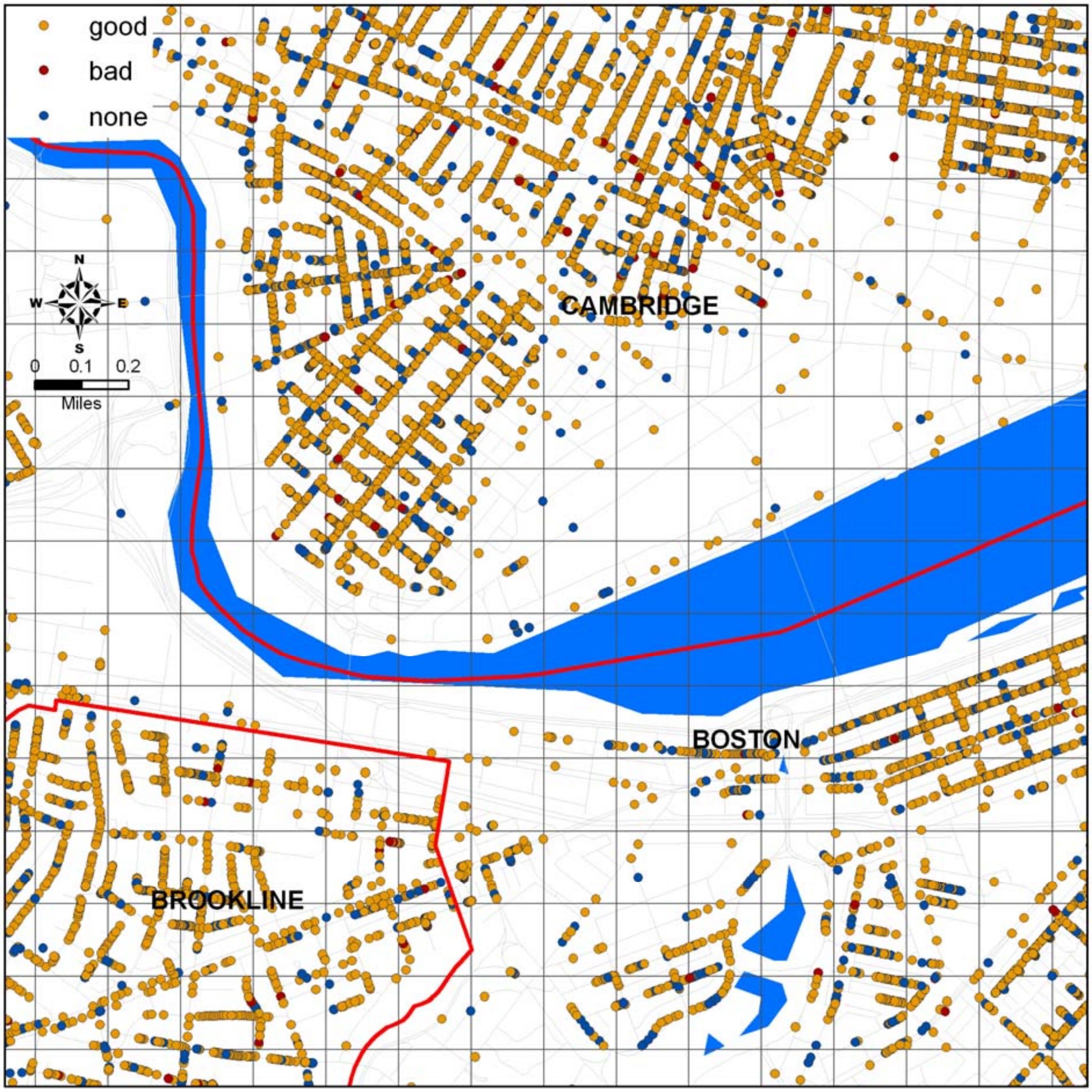


Source: The author

Figure 9: VMT per Capita across Grid Cells in Metro Boston

The dependent variables of the regression models are VMT per vehicle, VMT per household, and VMT per capita, computed for the 9-grid-cell catchment area of each grid cell, respectively. Figure 10 plots part of the study area. The vehicles are geocoded to a point layer based on the owners' street addresses. "Good" vehicles refer to vehicles with at least two credible odometer readings; "bad" vehicles refer to vehicles with less than two credible odometer readings; and "none" means vehicles without odometer readings at all. Due to the nature of the geocoding function in GIS softwares, the points are not located at the centroids of corresponding homes, but line up along roads. Points that are close to the boundaries of grid cells are likely to be assigned to the wrong grid cells. The catchment area could help analysts smooth the surface and reduce the noise in the raw data.

The total number of grid cells with at least one vehicle is 60,895. I exclude grid cells with annual VMT per household less than 100 miles or greater than 100,000 miles as well as grid cells without complete information. The final dataset for empirical analysis includes 52,929 grid cells.



Source: The author.

Figure 10: Geocoded Vehicles and Grid Cells

3.3.3 Built-Environment Variables

For this study, I computed 27 built-environment variables at fine-grained 250x250m grid cell level as described in Chapter 2.

3.3.4 Demographic Variables

Based on literature, I select 12 demographic variables at the block group level to control for the zonal difference of population, including percent of population below the poverty level, percent of owner-occupied housing units, percent of population with at least 13 years of schooling, median household income, percent of population that is white, per capita income, unemployment rate, percent of households with fewer than 3 members, percent of population three years old and over who are enrolled in elementary/high school, percent of population under 5, percent of population 65 years old and over, and percent of population 16 years old and over in labor force. Ideally, I should compute demographic variables at the grid cell level, but because of data limitations, I assign each grid cell the value of the block group to which it belongs. For population and household counts, block group counts were distributed among only those grid cells in the residential area.

3.4 EMPIRICAL ANALYSIS

In Section 2.4, I present the results of the empirical analysis for the Boston Metropolitan Area.

3.4.1 Factors Analysis

To deal with the multicollinearity among variables, I use factor analysis to reduce a large number of built-environment and demographic variable to several built-environment and demographic factors respectively. The factors are included in the regression models as explanatory variables.

The factor analysis for built-environment variables is presented in Chapter 2. Similarly, I also apply factor analysis to the 12 demographic variables at the block group level and extract from them 3 demographic factors: wealth, children, and working status. Factor 1 can be seen as an indicator of wealthy level. Block groups with higher values in Factor 2 tend to have more children and bigger household size. Factor 3 is related to residents' working status. The three factors explain 71.6% of the variance in the original variables. Factor loadings for each demographic variable are shown in Table 3. Table 4 presents the descriptive statistics of variables in the regression models.

TABLE 3: Factor Loadings of Demographic Factors

		Factor 1	Factor 2	Factor 3
		Wealth	Children	Working Status
1	Percent of population below poverty level	-0.863		
2	Percent of owner-occupied housing units	0.818	0.386	
3	Percent of population with at least 13 years of schooling	0.817		
4	Median household income	0.812		
5	Percent of population that is white	0.796		
6	Per capita income	0.707		
7	Unemployment rate	-0.613		
8	Percent of households with less than 3 members		-0.909	
9	Percent of population that are enrolled in elementary/high school		0.869	
10	Percent of population under 5		0.728	
11	Percent of population 65 years old and over			-0.856
12	Percent of population 16 years old and over in labor force	0.427		0.793

* I suppress factor loadings with an absolute value less than 0.35 for interpretation convenience.
Source: Calculated by the author using SPSS.

Table 4: Descriptive Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
VMT per vehicle	52929	12056.9	1770.8	5219.7	23843.7
VMT per household	52929	27120.6	13315.4	625.3	98954.6
VMT per capita	52929	9372.2	4204.0	85.0	50158.2
BE factor. 1: distance to non-work destinations	52929	-0.245	0.865	-2.594	3.983
BE factor 2: connectivity	52929	0.425	1.172	-1.644	11.130
BE factor 3: inaccessibility to transit and jobs	52929	-0.108	0.973	-2.271	4.583
BE factor 4: auto dominance	52929	-0.082	0.610	-1.210	6.409
BE factor 5: walkability	52929	0.080	0.921	-2.664	4.007
DEM factor 1: wealth	52929	0.568	0.654	-4.153	2.588
DEM factor 2: children	52929	0.413	0.764	-3.323	3.793
DEM factor 3: working status	52929	0.097	0.862	-6.923	4.104

Source: Calculated by the author.

3.4.2 Regression Results

Depending upon the selection of dependent variable and model specification, I estimate the following nine models:

1. OLS model for VMT per vehicle;
2. OLS model for VMT per household;
3. OLS model for VMT per capita;
4. Spatial lag model for VMT per vehicle;
5. Spatial lag model for VMT per household;
6. Spatial lag model for VMT per capita;
7. Spatial error model for VMT per vehicle;
8. Spatial error model for VMT per household; and
9. Spatial error model for VMT per capita.

I estimate the spatial-lag and spatial-error models with GeoDa 0.9.5 software. Table 5 summarizes statistics for the regression models. The R-squared of the OLS models range from 34.2% to 52.7%. Test of residuals indicates that the error term of the OLS models exhibit significant spatial autocorrelation. The likely reasons are the omission of spatially-correlated explanatory variables, and the effects of travel behavior in surrounding areas. Moreover, both the simple Lagrange multiplier tests for omitted spatially-lagged dependent variables (LM-lag) and error dependence (LM-error) are statistically significant, indicating the existence of spatial autocorrelation.

To capture the spatial effects, I estimate both spatial-lag and spatial-error models. Anselin et al.'s (1996) Lagrange multiplier tests of spatial-lag and spatial-error specifications being mutually contaminated by each other are employed to compare the two models. Both the test for error dependence in the possible presence of a missing lagged dependent variable (robust LM-error), and the test for a missing lagged dependent variables in the possible presence of spatially-correlated error term (robust LM-lag) are statistically significant. But the robust LM-error test rejects the null at the higher level of significance, favoring the spatial-error model. The log-likelihood statistics also support this conclusion, indicating that the spatial-error model has a better fit to the data than the corresponding spatial-lag model and OLS model. The goodness-of-fit statistics for VMT per vehicle models are higher than those for VMT per household and VMT per capita.

Table 6 presents the estimation results of the three models using the spatial-error specification.

Table 5: Estimation Summary

	VMT per Vehicle			VMT per Household			VMT per Capita		
	OLS	Spatial Lag	Spatial Error	OLS	Spatial Lag	Spatial Error	OLS	Spatial Lag	Spatial Error
Observations	52929	52929	52929	52929	52929	52929	52929	52929	52929
R-squared	0.527	0.789	0.810	0.418	0.626	0.631	0.342	0.566	0.573
Log Likelihood	-451127	-432073	-429930	-563448	-553582	-553497	-505660	-496458	-496291
Test	Statistic	p-value		Statistic	p-value		Statistic	p-value	
LM--Lag	86355.0	0.00		43966.2	0.00		41094.4	0.00	
LM--Error	115402.4	0.00		46425.7	0.00		43147.3	0.00	
Robust LM--Lag	621.6	0.00		619.4	0.00		305.3	0.00	
Robust LM--Error	29669.0	0.00		3078.8	0.00		2358.1	0.00	

Source: Calculated by the author.

Table 6: Estimation Results of the Spatial-Error Models

	VMT per Vehicle			VMT per Household			VMT per Capita		
	Coef.	t-stat.		Coef.	t-stat.		Coef.	t-stat.	
Built-Environment Factors									
Distance to non-work destinations	444.7	21.2	**	3820.9	23.1	**	859.7	15.8	**
Connectivity	-250.7	-23.4	**	-2970.3	-34.6	**	-833.6	-29.3	**
Inaccessibility to transit & jobs	1004.1	32.2	**	5905.6	30.1	**	1954.1	30.9	**
Auto dominance	-9.7	-1.0		581.2	6.0	**	271.5	8.3	**
Walkability	14.6	1.7		-1560.9	-19.4	**	-589.4	-21.8	**
Demographic Factors									
Wealth	-26.9	-2.0	*	737.7	5.5	**	296.9	6.6	**
Children	-9.1	-1.0		545.5	5.9	**	-45.9	-1.5	
Working status	29.6	4.4	**	160.3	2.3	*	58.1	2.5	*
Lambda	0.91	397.1	**	0.84	231.8	**	0.83	218.9	**
Constant	12409.4	313.4	**	30825.1	128.5	**	10456.6	135.1	**

* and ** denote coefficient significant at the 0.05 and 0.01 level respectively.
Source: Calculated by the author.

As shown in Table 6, most coefficients for demographic factors are statistically significant. One interesting finding is that higher wealthy level is associated with lower VMT per vehicle, but higher VMT per household and VMT per capita, which suggests that wealthier households tend to own more cars but drive each car less compared to other households. Household structure also influences vehicle usage. The number of children in the household tends to increase VMT per household, presumably because of child-related non-work trips. But its effects on VMT per vehicle and VMT per capita are insignificant. One possible explanation is that households tend to buy more vehicles as household size grows, but the usage of each vehicle does not change significantly. Factor 3 can be seen as a proxy for percentage of population that is working. This factor is positively associated with all three VMT variables, presumably due to the commuting trips.

After controlling for the influence of demographic factors, I find that built-environment factors are indeed important predicators of vehicle usage at grid cell level, with smart-growth-type neighborhoods associated with less vehicle usage than sprawl-type neighborhoods. The coefficients for the "distance to non-work destination" factor in the three models are positive and significant at the 0.01 level, suggesting that the spatial distribution of non-work activities is significantly associated with vehicle usage. As the distance to non-work destinations increase, VMT per vehicle, VMT per household, and VMT per capita all increase. The negative sign of the "connectivity" factor in all three models suggests that connectivity –an indicator of high-density, grid-type neighborhood tends to reduce household vehicle usage. The coefficients of the "auto dominance" factor are positive and significant in the VMT per household and VMT per capita models, while its coefficient in the VMT per vehicle model is insignificant. This suggests that an auto-friendly environment influences VMT by increasing the number of cars owned by households rather than by increasing the usage of each vehicle. As revealed by the estimated coefficients of the "walkability" factor, a good pedestrian environment is associated with lower VMT per household and VMT per capita, while its effect on VMT per vehicle is insignificant. The "walkability" factor tends to influence VMT by reducing the number of vehicles purchased.

By comparing the coefficients of the demographic and built-environment factors, I find that built-environment factors have a higher prediction power on VMT than demographic factors. Table 7 and Figure 11 present the change in annual VMT per vehicle, per household, and per capita due to one standard deviation increase in the individual factor. As is shown in Figure 11, accessibility to work and non-work destinations, connectivity, and transit accessibility make a much higher contribution to the model than other factors. The contributions are large for the VMT per household measure, where the average VMT per household at grid cell level for the study area is about 27,121 miles[23].

---

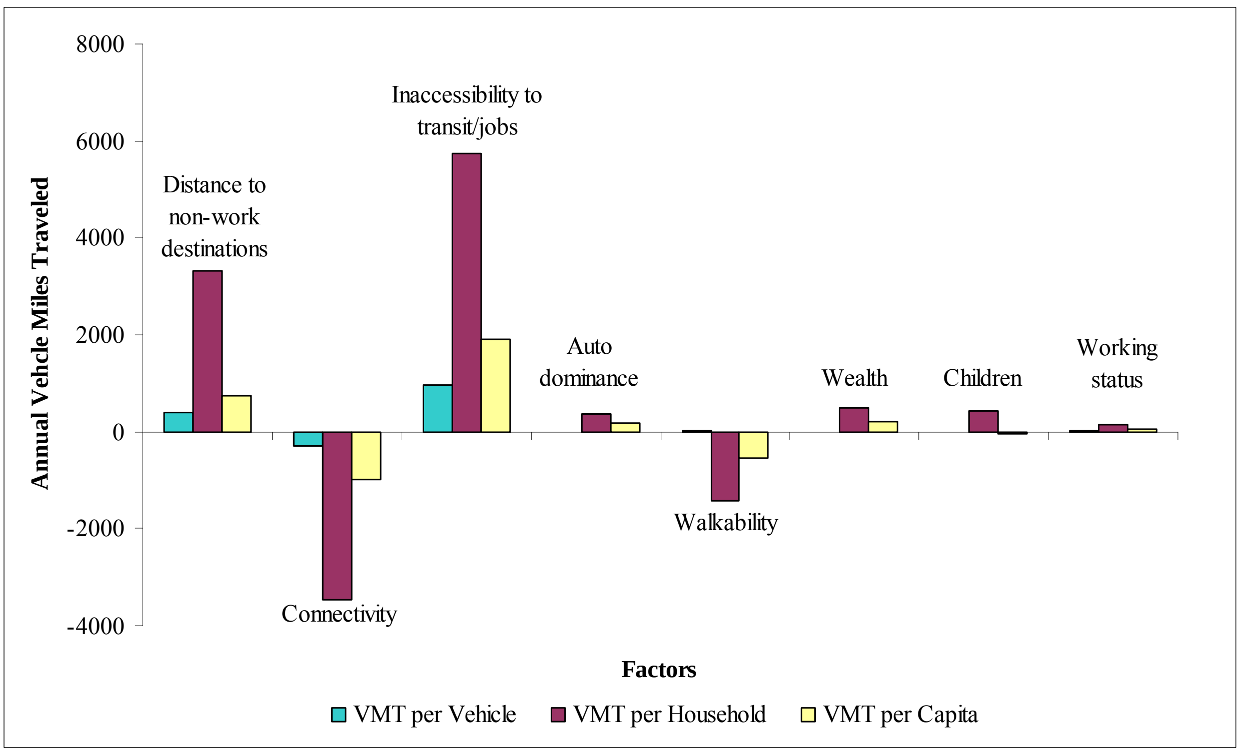
[2] For comparison purpose, I also calibrated the spatial error model with built-environment factors and 3 demographic variables, median household income, percent of households with less than 3 members, and percent of population 16 years old and over and in labor force. Each demographic variable represents one demographic factor. The estimation results and the change in VMT measures due to one standard deviation increase in the independent variables are presented in Appendices 1. The major conclusions of this essay still hold, except that the coefficient of the median household income variable has a positive and insignificant coefficient in the VMT per vehicle model.

[3] To account for the boundary effect in the "inaccessibility to transit and jobs" factor, I rerun the spatial error model after excluding the 10 percent grid cells with the highest scores in the "inaccessibility to transit and jobs" factor. The major conclusions still hold, which suggests that the impact of the boundary effect is not significant in this study.

Table 7: Change in VMT Measures Due to One Standard Deviation Increase in Factors

	VMT per Vehicle	VMT per Household	VMT per Capita
Built Environment Factors			
Distance to non-work destinations	384.8	3306.4	744.0
Connectivity	-293.8	-3480.5	-976.8
Inaccessibility to transit and jobs	976.7	5744.7	1900.8
Auto dominance	-5.9	354.6	165.6
Walkability	13.4	-1437.7	-542.9
Demographic Factors			
Wealth	-17.6	482.1	194.0
Children	-7.0	416.9	-35.1
Working Status	25.6	138.3	50.1

Source: Calculated by the author



Source: The author.

Figure 11: Contributions of Factors to the Model

3.5 CONCLUSIONS

In this study, I examine the relationship between the built environment and household vehicle miles traveled in the Boston Metropolitan Area. I derive the VMT measures using annual safety inspection records for all private passenger vehicles registered in Metro Boston. I compute a set of built-environment variables at 250x250m grid cell level using GIS techniques, apply factor analysis to mitigate multicollinearity, and integrate the built-environment and demographic factors into regression models to explain VMT variations. Spatial regression techniques are applied to correct spatial autocorrelation.

This study provides some clues to the relationships between the built environment and vehicle usage within the Boston Metro area. The spatial-error model outperforms the corresponding spatial lag and OLS models in goodness-of-fit statistics. The regression results of the spatial error model reveal that both the built-environment and demographic factors are significantly associated with VMT. On the demographic side, I find that wealth is negatively associated with VMT per vehicle, but positively associated with VMT per household, suggesting that households in wealthier neighborhood tend to own more cars than other households, but use each car less. Due to data limitation, I computed the demographic variables at the block group level, which is more aggregate than built-environment variables. Thus the results may be influenced by the Modifiable Areal Unit Problem. In this study, I show that the built-environment factors have higher impacts on VMT than demographic factors. In particular, accessibility to work and non-work destinations, connectivity, and transit accessibility are negatively associated with VMT, and their impacts are noticeably greater than other factors. In most studies using travel survey data, the bias is in the other direction – the individual characteristics are known, but the built-environment factors come from data aggregated at census tract or zip code zone scale. Many of these studies find that demographic or attitudinal factors explains most of the variations in VMT across households (e.g., Kitamura et al. 1997, Bagley and Moktarian 2002, and Frank et al. 2007), while the built-environment effect is minimal. The difference between my study and survey-based studies indicates the potential biases due to data aggregation, both on the demographic side and the built-environment side. The built-environment effect may be biased downwards in previous studies using aggregate built-environment measures, just like the demographic effect in this study.

Although finding a strong association between the built environment and travel patterns is not the same as showing that a change in the built environment will lead to a change in travel behavior (Handy 1996), these results still provide some support for those smart-growth policies that advocate increasing accessibility to destinations, creating traditional-type high-density, mixed-use neighborhoods, and improving transit accessibility. The research findings can facilitate the dialogue among regional-planning agencies, local government and the public regarding growth management and sustainable regional development strategies and scenarios.

This study also has implications for urban modeling by revealing the opportunities brought about by new spatial data infrastructure. With the development of information technology, the amount of administrative data with location information is rapidly increasing. For example, standardized GIS data layers are becoming more common for data about road networks, parcels, and building footprint, and for transaction information, such as housing transactions, vehicle safety inspections, transit fare cards, utility records, and cell phone use. These administrative datasets are collected regularly by various agencies. Calibrating urban models using administrative data can save the high expense of frequent surveys and enable improved monitoring and modeling of metropolitan areas at a spatially-detailed scale.

In the future, analysts can extend this study along multiple directions, for example

1. examine temporal trends in land use-transportation interconnection using time series safety inspection data;

2. construct profiles of fuel economy so that the built environment can be directly linked to energy consumptions and GHG emissions.

3. employ structural equations models to investigate the causal relationships among key variables, such as the built environment, automobile ownership, and travel behavior; and

4. extend the analysis to other North American metropolitan areas.

CHAPTER FOUR: RESIDENTIAL PROPERTY VALUES AND THE BUILT ENVIRONMENT: AN EMPIRICAL STUDY IN THE BOSTON METROPOLITAN AREA

4.1 INTRODUCTION

Over the last decade, planners have shown renewed interest in utilizing land-use-control policies to mitigate negative effects of sprawl-type development. Under the general name "smart growth", a group of planning strategies such as urban growth boundary, mixed-use planning, and transit-oriented development, is gaining popularity. Researchers have argued that built-environment features advocated by such strategies can curb travel demand, ease congestion, reduce emission, and contribute to improved quality of life (Tu and Eppli 1999).

From a policy perspective, it is important to understand how the built environment is valued in the market place. This information can help estimate the property-value effects of land-use change, and quantify the impacts of smart-growth policies on a neighborhood. Furthermore, it provides a potential financing mechanism via land value capture to fund infrastructure investment and help relieve the financial burdens of governments and agencies around the world.

Despite the policy motivations, a close look at the literature reveals that there have been few detailed and comprehensive analyses of the relationship between the built environment and residential property values. A number of analysts have empirically investigated the effects on housing price of certain built-environment features (e.g., Cao and Cory 1981; Song and Knaap 2004; Bowes and Ihlanfeldt 2001; Mattews and Turnbull 2007). However, they have been unable to draw a complete picture of the built environment, which is multi-dimensional in nature, due to data limitations and methodological challenges, such as measurement of the built environment, multicollinearity, and spatial autocorrelation.

Recent developments in information infrastructure and econometrics have led to a significant increase in the amount of available data with spatial attributes, spatial analysis tools, and modeling techniques dealing with spatial phenomena, which allow investigators to account for built-environment characteristics in their models (Case et al. 2004). Taking advantage of these new advances, I develop a comprehensive and spatially detailed analysis of the relationship between the built environment and residential property values.

The next section introduces related literature. Section 4.3 describes data and study area. Section 4.4 outlines the methodological framework of empirical analyses. Section 4.5 presents and discusses the modeling results. Section 4.6 concludes this second study of built-environment effects on travel demand, housing prices, and housing location.

4.2 LITERATURE REVIEW

This section summarizes the related literature, including the behavior framework of household location choice and hedonic value analysis of the built environment.

4.2.1 Behavioral Framework

Two strands of literature are closely related to household location choice. One line of research is the monocentric city model in urban economics. The concept of the monocentric city has its historical origin in the work of von Thunen (1966), and is further developed by Alonso (1964), Muth (1969), and Mills (1972). The Alonso-Muth-Mills model describes the equilibrium residential pattern in a monocentric city, whereby people commute to the central business district, where all jobs are located, with transportation cost depending on commuting distance. Each household maximizes utility by allocating household income to the consumption of a composite good, land (housing), and commuting. This model remains a powerful workhorse for the analysis of land values and location choices. However, analysts would like to consider more complex representations of dispersed destinations and the multi-modal transportation system in order to characterize modern polycentric metropolitan areas.

Another line of literature deals with the relationship between the built environment and travel behavior, which is widely researched in the transportation field. The built environment comprises land use, urban design, and transportation systems (Handy et al. 2002). Crane (1996) argues that the built environment can influence travel cost through speed and distance. He proposes individual choice of trip frequency and mode split as a constrained utility-maximization problem, with the built environment influencing travel behavior through the travel time of individual mode. Boarnet and Crane (2001) consider travel cost to be a generalized cost including time, out-of-pocket monetary expenditures, and psychological effects, and specify three alternative ways the built environment could affect travel cost. Fan and Khattak (2009) suggest two specific mechanisms through which the built environment may influence travel decisions: the built environment affects distance of trips, and the built environment affects time cost of driving. Cao et al. (2009) indicate that the extent to which travel costs are affected by the built environment is debatable. Built-environment characteristics may be good predictors for non-motorized travel costs, moderate predictors for auto travel costs, but inferior predictors for transit travel costs.

In summary, the built environment can influence travel costs either directly or indirectly, and thereby might influence household location choice and housing price.

4.2.2 Hedonic Price Analysis of the Built Environment

Hedonic-price models assume that goods are characterized as a bundle of inherent attributes, and the observed prices of goods reflect the implicit prices of these attributes (Rosen 1974).

Researchers have long sought to explain the variation in property values with hedonic-price models and location characteristics, such as public-service level, tax rate, and school quality (Edel and Sclar 1974; King 1974; Downes and Zabel 2002).

Analysts also apply hedonic-price models to investigate the built-environment effects on housing price. Existing studies indicate that certain built-environment features can be capitalized into property values, such as land-use mix (Cao and Cory 1981; Song and Knaap 2004), transit accessibility (Rowes and Ihlandeldt 2001; Rodriguez and Mojica 2009), and street network pattern (Matthews and Turnbull 2007). Cao and Cory (1981) show that increasing industrial, commercial, multi-family and public land uses tends to increase surrounding home values. Song and Knaap (2004) demonstrate that housing prices increase with proximity to public parks or commercial centers. Bowes and Ihlanfeldt (2001) look into both direct and indirect effects of transit stations, and they find that stations located away from downtown have positive impacts on property values, while stations in low-income neighborhoods or close to downtown generate negative externalities to nearby properties. Rodriguez and Mojica (2009) employ a before-and-after hedonic-price model to determine the effects of the Bus Rapid Transit (BRT) network expansion in Bogota. Compared with the control area, they identify asking price increases of 13-14% for the period after the BRT was extended. Matthews and Turnbull (2007) use measures of street connectivity and their interactions with other neighborhood attributes to evaluate how street layout affects property values, and they find a significant impact. Unlike the above studies that focus on one specific dimension of the built environment, Song and Knaap (2003) develop a comprehensive study on urban-form measures. They find that households pay a premium for some new-urbanism features, such as more connective street networks, shorter cul-de-sacs, smaller block size, better pedestrian accessibility to commercial uses, more evenly-distributed mixed land uses, and better proximity to light rail stations. Features such as higher density and containing more commercial, multifamily and public uses are not attractive to most buyers.

Compared with the large amount of literature on residential property values, studies focusing on the built environment are relatively few. Researchers face a number of dilemmas in probing the links between the built environment and residential property values. One significant barrier is the absence of spatially-detailed built-environment data. Data limitations have forced researchers to use built-environment measures that are more aggregate than is suggested by relevant theories (Song and Knaap 2003), or focus on narrow aspects of the built environment, taking a piecemeal approach to built-environment attributes (Matthews and Turnbull 2007). Moreover, some methodological challenges also contribute to the lack of substantive empirical results. To compute built-environment variables, such as density, land-use mix, street network layout, and pedestrian environment, many analysts have relied on a definition of neighborhood that is either dependent on census geography or on the delineation of a neighborhood. Thus, they are influenced by the Modifiable Areal Unit Problem (MAUP), one well-known problem in the analysis of spatial phenomenon. The MAUP often leads to the inconsistency of measurement results and statistical analyses. Due to the collinearity between built-environment attributes like density, mixed use, and walkability, it is questionable whether many built-environment variables will show up as statistically significant in the model (Cervero and Kockelman 1997). The spatial autocorrelation problem associated with the use of spatial data could lead to biased and inconsistent or inefficient estimation results in OLS models, depending upon the form of spatial autocorrelation (Anselin 1993). I aim to address some of these issues and develop a more comprehensive study of the built environment and residential property values.

4.3 DATA AND METHODOLOGY

In this section, I describe the methodology and datasets used in this study.

4.3.1 Built-Environment Measurement and Factor Analysis

Based on the behavioral framework discussed in Section 4.1 and related literature, I compute 27 built-environment variables that have the potential to influence travel costs. To deal with the potential multicollineairty among built-environment variables, I apply factor analysis to reduce a large set of built-environment variables to several factors and include the factors in regression models.

4.3.2 Hedonic-price models and Spatial Econometrics

A widely-used semi-log form hedonic-price model for housing properties is:

$$Ln(y_{it}) = \sum \alpha_j X_{ijt} + \sum \beta_t D_{it} + \varepsilon_{it} \tag{1}$$

For time period t, y_{it} is the transaction price of property i, X_{ijt} is a set of j housing attributes, D_{it} is a set of dummy variables which equal one for transactions taking place in time period t, and zero otherwise, and ε is a random error. Estimates of α can be used to compute the implicit marginal price for housing attributes. Estimates of β measure price movements associated with each time period, relative to a base period. Although there is no strong theoretical basis for choosing the functional form of a hedonic regression, Malpezzi (2002) argues that the semi-log specification has several advantages.

Literature has shown that if spatial autocorrelation is presented in an OLS model, the estimation results will be either biased and inconsistent or inefficient depending on the characteristics of the spatial autocorrelation (Anselin 1993). One reason for this phenomenon might be that houses in the same neighborhood share certain location characteristics. Following

Anselin (1993), we account for two types of spatial autocorrelation with two types of spatial econometric models. The spatial-lag model, which is analogous to the time-series lagged dependent variable model, is used to deal with autocorrelation related to a lagged term on the dependent variable. In this case, OLS will be biased and inconsistent. The second type of autocorrelation is analogous to time-series serially-correlated errors, which leads to unbiased and consistent, but inefficient, OLS estimation. Analysts use a spatial-error model to account for this type of spatial autocorrelation.

The spatial-lag model can be specified as:

$$Ln(y_{it}) = \rho W_{Ln(y_{it})} + \sum \alpha_j X_{ijt} + \sum \beta_t D_{it} + \varepsilon_{it} \qquad (2)$$

where ρ is the autoregressive coefficient, $W_{ln(yit)}$ is the NxN spatial weight matrix, $\varepsilon \sim N(0, \sigma^2 I)$.

The spatial-error model can be specified as:

$$\begin{aligned} Ln(y_{it}) &= \sum \alpha_j X_{ijt} + \sum \beta_t D_{it} + \varepsilon_{it} \\ \varepsilon_{it} &= \lambda W_{\varepsilon_{it}} + \mu_{it} \end{aligned} \qquad (3)$$

where λ is the spatial autoregressive coefficient, $W\varepsilon_{it}$ is the NxN spatial weight matrix, μ is a vector of i.i.d. standard normal error terms.

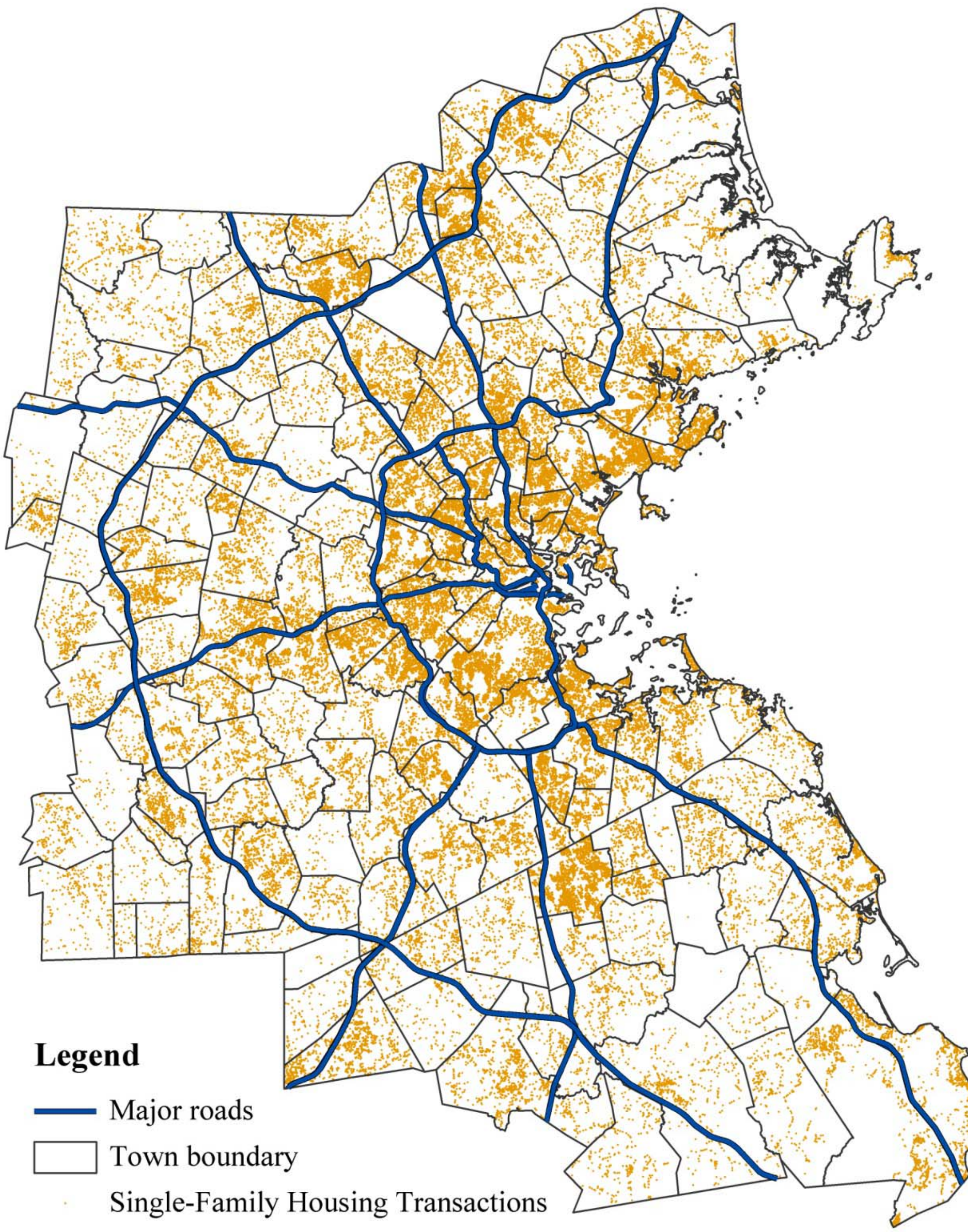
4.4 STUDY AREA AND DATA

I select the Boston Metropolitan Area as the study area. Boston exhibits a variety of built-environment characteristics, which makes it a compelling case for my study.

In this study, I use two recent datasets with exceptional spatial detail to measure housing price and the built environment. The primary housing dataset includes information on all single-family[4] housing transactions in the Boston Metropolitan Area from 2004 to 2006 provided by the

[4] In this study, single-family properties are defined as properties with state use code “101”.

Warren Group. This dataset contains date of sale, transaction price, location, and detailed structural characteristics of properties. I select 11 structural variables for the analysis: (1) lot size, (2) living area, (3) number of parking spaces, (4) number of fireplaces, (5) total rooms, (6) number of bedrooms, (7) number of full baths, (8) number of half baths, (9) a dummy variable indicating below average condition, (10) a dummy variable indicating good or above condition, and (11) a dummy variable showing the existence of air conditioning. After excluding transactions with unreliable data, I include 92,774 single-family housing transactions in the analysis. Transactions in the town of Tewksbury are missing. I plot the spatial distribution of housing transactions in Figure 12.



Source: The author

FIGURE 12: Single-family housing transactions in the Boston Metropolitan Area, 2004-2006.

Another dataset is the built-environment data from MassGIS, the State's Office of Geographic and Environmental Information, with unprecedented spatial detail. Detailed data description is provided in Chapter 2.

4.4.1 Dependent Variable

The dependent variable in the hedonic-price model is the natural logarithm of the nominal transaction price deflated to the first quarter of 2004.

4.4.2 Built-Environment Variables

I computed 27 built-environment variables in this study as described in Chapter 2.

4.4.3 Control Variables

To control for the influence of non-built-environment attributes, I include four additional sets of variables in the regression models: (1) structural characteristics (11 structural variables from the Warren Group data), (2) public service level (property crime rate, residential property tax rate, and school scores), (3) neighborhood socioeconomic characteristics (percent of white population and median household income of the block group), and (4) view amenity (distance to park).

Table 8 presents the descriptive statistics of variables in the model.

TABLE 8: Descriptive Statistics of Variables

Variable	Minimum	Maximum	Mean	Std. Dev.
Ln(transaction price)	11.00	15.00	12.91	0.45
Control Variables				
Lot size (k ft2)	0.400	2918.520	26.825	50.836
Number of parking space	0	8	0.357	0.712
Number of fireplaces	0	9	0.667	0.807
Living area (k ft^2)	0.32	15.43	1.957	0.947
Total number of rooms	1	23	7.001	1.685
Number of bedrooms	1	15	3.298	0.853
Number of full bathrooms	1	10	1.673	0.768
Number of half bathrooms	0	5	0.607	0.540
Dummy - below average building condition	0	1	0.024	0.152
Dummy - good or above building condition	0	1	0.351	0.477
Dummy - presence of air conditioning	0	1	0.382	0.486
Median household income (k$)	9.327	200.001	70.462	25.983
Percent of white population	0.000	1.000	0.905	0.131
Residential property tax per (k$)	7.270	15.110	10.276	1.511
Property crime rate (crime/population*1000)	0.000	48.079	18.637	11.873
School scores	49.000	194.000	148.122	27.490
Distance to park (km)	0.000	6.950	0.546	0.559
Built-Environment Variables				
Distance to church (km)	1.000	10.000	2.938	1.986
Distance to dentist (km)	1.000	15.000	3.538	2.531
Distance to grocery store (km)	0.000	8.381	1.408	1.155
Distance to gym (km)	1.000	15.000	3.882	2.172
Distance to hardware store (km)	0.000	7.537	1.567	1.079
Distance to shopping mall (km)	0.000	9.604	1.803	1.386
Distance to restaurant (km)	1.000	10.000	2.827	1.997
Distance to school (km)	0.000	6.800	1.057	0.881
Percent of roads with access control	0.000	0.977	0.026	0.088
Percent of roads with 30+ speed limit	0.000	1.000	0.037	0.095
Average road width (ft)	0.000	342.008	39.178	15.946
Distance to highway exit (km)	0.022	17.570	3.638	2.539
Distance to subway station (km)	0.006	58.829	19.748	13.977
Distance to commuter rail station (km)	0.021	24.639	4.444	3.779
Distance to bus stop (km)	0.002	51.914	11.297	11.498
Distance to MBTA parking lot (km)	0.005	24.453	4.680	3.975
Average sidewalk width (ft)	0.000	20.943	3.561	3.394
Percent of roads with curbs	0.000	1.000	0.366	0.294

Percent of roads with sidewalks	0.000	1.000	0.423	0.315
Population density (10k/km^2)	0.000	2.041	0.161	0.190
Land-use mix	0.000	0.994	0.247	0.229
Road density (km/km^2)	0.000	56.302	9.533	5.400
Intersection density (10/km^2)	0.000	28.444	5.466	3.937
Density of 3-way intersections (10/km^2)	0.000	18.489	3.580	2.581
Density of 4-way intersections (10/km^2)	0.000	9.067	0.823	1.042
Percent of 4-way intersections	0.000	1.000	0.114	0.104
Job accessibility (k)	5.869	690.722	174.308	136.451

Source: Calculated by the author.

4.5 EMPIRICAL RESULTS

4.5.1 Built-Environment Factors

To deal with the potential multicollinearity, I perform a principle component analysis with varimax rotation on the built-environment variables as detailed in Chapter 2. Table 9 reports the descriptive statistics of built-environment factors in the model.

Table 9: Descriptive Statistics of Built-Environment Factors

Built-Environment Factors	Minimum	Maximum	Mean	Std. Dev.
Distance to non-work destinations	-2.594	3.639	-0.367	0.767
Connectivity	-1.672	8.940	0.874	1.364
Inaccessibility to transit and jobs	-2.259	4.583	-0.166	1.014
Auto dominance	-1.245	7.508	-0.073	0.569
Walkability	-2.664	4.007	0.295	0.971

Source: Calculated by the author.

4.5.2 Regression Models

I estimate six models in this study, depending on the selection of model specification and the choice of factors:

Model 1: OLS model with built-environment variables

Model 2: OLS model with built-environment factors

Model 3: Spatial-lag model with built-environment variables

Model 4: Spatial-lag model with built-environment factors

Model 5: Spatial-error model with built-environment variables

Model 6: Spatial-error model with built-environment factors

Models 1 and 2 use OLS estimation, assuming absence of spatial autocorrelation. The value of Moran's I test for model 2 is 192.26, significant at the 0.01 level, suggesting a clear cluster pattern of residuals. The spatial-weight matrix for both spatial-lag and spatial-error models is developed assuming constant spatial dependence among properties up to a maximum distance. The maximum Euclidean distance used is 400m. Table 10 shows the summary statistics of the six models. Estimation results of the six models are given in Tables 11 and 12.

Table 10: Estimation Summary

	Model (1) OLS + BE Variables	Model (2) OLS + BE Factors	Model (3) Spatial Lag + BE Variables	Model (4) Spatial Lag + BE Factors	Model (5) Spatial Error + BE Variables	Model (6) Spatial Error + BE Factors
Observations	92774	92774	92774	92774	92774	92774
R-squared	0.750	0.733	0.751	0.735	0.794	0.797
Log Likelihood	5971.72	3008.82	6149.59	3238.25	13665.05	12797.12
AIC	-11831.4	-5949.64	-12185.20	-6406.50	-27218.10	-25526.20
SC	-11302.9	-5628.75	-11647.20	-6076.17	-26689.57	-25205.35

Source: Estimated by the author using GeoDa 0.9.5.

TABLE 11: Estimation Results of Models 1, 3, and 5

Variables	Model (1) OLS + BE Variables		Model (3) Spatial Lag + BE Variables		Model (5) Spatial Error + BE Variables	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Constant	11.476	787.71**	11.358	710.55**	11.576	502.10**
Control Variables						
Lot size (10k sq. ft)	0.003	19.55**	0.003	21.11**	0.004	26.25**
Number of parking space	0.022	18.96**	0.021	18.68**	0.020	17.39**
Number of fireplaces	0.039	34.95**	0.039	34.76**	0.033	29.23**
Living area (k sq. ft^2)	0.160	110.42**	0.160	110.16**	0.150	107.02**
Total number of rooms	0.018	22.07**	0.018	22.16**	0.014	18.95**
Number of bedrooms	0.009	6.76**	0.009	6.70**	0.014	11.61**
Number of full bathrooms	0.085	57.80**	0.084	57.61**	0.065	47.69**
Number of half bathrooms	0.073	45.55**	0.072	45.32**	0.062	42.14**
Below average building condition	-0.092	-18.45**	-0.092	-18.49**	-0.114	-24.29**
Good and above building condition	0.059	34.40**	0.058	34.12**	0.071	40.44**
Presence of A/C	0.009	6.05**	0.009	5.70**	0.010	6.60**
Median household income (k$)	0.003	59.38**	0.003	58.14**	0.002	34.73**
Percentage of white population	0.170	22.74**	0.167	22.33**	0.120	10.72**
Residential property tax rate	-0.016	-28.16**	-0.016	-27.90**	-0.017	-16.99**
Property crime rate	-0.001	-5.68**	-0.001	-5.96**	-0.001	-3.94**
School scores	0.003	52.44**	0.003	51.78**	0.003	36.93**
Distance to park (km)	-0.008	-4.90**	-0.008	-4.98**	-0.007	-2.76**
Built Environment Variables						
Distance to church (km)	0.001	1.08	0.001	1.05	0.001	1.09
Distance to dentist (km)	-0.004	-7.59**	-0.004	-7.35**	-0.004	-5.30**
Distance to grocery store (km)	-0.002	-1.99*	-0.002	-1.63	0.001	0.59
Distance to gym (km)	-0.002	-4.50**	-0.002	-4.45**	-0.002	-3.26**
Distance to hardware store (km)	0.009	9.02**	0.008	8.96**	0.009	5.84**
Distance to shopping mall (km)	0.004	5.48**	0.004	5.81**	0.006	4.75**
Distance to restaurant (km)	0.000	-0.12	0.000	-0.11	0.001	0.87
Distance to school (km)	0.009	7.48**	0.009	7.71**	0.006	2.94**
Percent of roads with access control	0.074	5.49**	0.076	5.75**	0.073	3.96**
Percent of roads with 30mph+ speed limit	0.014	1.30	0.013	1.23	-0.007	-0.46
Average road width (ft)	-0.001	-11.19**	-0.001	-11.13**	-0.001	-7.45**
Distance to highway exit (km)	-0.001	-2.84**	-0.001	-2.46*	-0.001	-1.66
Distance to subway station (km)	0.001	3.48**	0.001	3.51**	0.000	0.95
Distance to commuter rail station (km)	-0.015	-26.51**	-0.015	-26.34**	-0.016	-16.30**
Distance to bus stop (km)	0.000	-0.33	0.000	-0.53	0.000	0.59

Distance to MBTA parking lot (km)	0.013	24.17**	0.012	23.87**	0.013	14.63**
Average sidewalk width (ft)	-0.003	-3.18**	-0.002	-2.65**	-0.002	-1.71
Percent of roads with curbs	-0.044	-10.22**	-0.045	-10.45**	-0.032	-4.78**
Percent of roads with sidewalks	0.064	9.02**	0.059	8.35**	0.052	4.93**
Population density (10k/sq. km^2)	-0.002	-0.17	-0.002	-0.23	-0.029	-2.12*
Land-use mix	-0.018	-4.50**	-0.016	-4.06**	-0.035	-6.59**
Road density (km/sq. km^2)	-0.003	-9.05**	-0.003	-9.26**	-0.003	-7.93**
Intersection density (10/sq. km^2)	-0.003	-2.60**	-0.003	-2.82**	-0.005	-2.78**
Density of 3-way intersections (10/sq.km^2)	0.003	2.22*	0.003	2.35*	0.004	1.81
Density of 4-way intersections (10/sq.km^2)	0.007	2.92**	0.007	3.21**	0.007	2.42**
Percent of 4-way intersections	-0.055	-4.14**	-0.056	-4.25**	-0.051	-2.99**
Job accessibility (k)	0.009	66.48**	0.009	65.82**	0.010	42.90**

* and ** denote coefficient significant at the 0.05 level and 0.01 level respectively

Source: Estimated by the author using Geoda 0.9.5.

TABLE 12 Estimation Results of Models 2, 4, and 6

Variables	Model (2) OLS + BE Factors		Model (4) Spatial Lag + BE Factors		Model (6) Spatial Error + BE Factors	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Constant	11.493	897.01**	11.356	784.66**	11.692	475.35**
Control Variables						
Lot size (10k sq. ft^2)	0.003	18.79**	0.003	20.71**	0.004	27.48**
Number of parking space	0.032	27.90**	0.032	27.47**	0.022	18.98**
Number of fireplaces	0.043	38.64**	0.043	38.25**	0.033	29.11**
Living area (k sq. ft^2)	0.159	106.43**	0.158	106.25**	0.147	104.88**
Total number of rooms	0.020	24.12**	0.020	24.21**	0.014	18.73**
Number of bedrooms	0.008	5.80**	0.008	5.74**	0.015	12.44**
Number of full bathrooms	0.093	61.47**	0.092	61.21**	0.062	46.26**
Number of half bathrooms	0.079	48.32**	0.079	48.01**	0.061	41.91**
Below average building condition	-0.084	-16.42**	-0.084	-16.47**	-0.116	-24.90**
Good and above building condition	0.048	27.48**	0.047	27.26**	0.072	40.06**
Presence of A/C	0.008	4.77**	0.007	4.38**	0.009	6.25**
Median household income (k$)	0.003	66.01**	0.003	64.33**	0.002	28.18**
Percentage of white population	0.170	23.15**	0.166	22.61**	0.081	6.28**
Residential property tax rate	-0.021	-36.67**	-0.020	-36.13**	-0.020	-15.99**
Property crime rate	-0.001	-6.54**	-0.001	-6.83**	-0.001	-3.48**
School scores	0.003	57.91**	0.003	56.95**	0.003	32.72**
Distance to park (km)	-0.009	-5.66**	-0.009	-5.67**	-0.011	-3.50**
Built Environment Factors						
Distance to non-work destinations	-0.008	-6.96**	-0.007	-5.65**	0.001	0.54
Connectivity	0.036	47.85**	0.035	46.64**	0.016	12.21**
Inaccesibility to transit and jobs	-0.070	-77.10**	-0.069	-76.00**	-0.084	-42.92**
Auto dominance	-0.005	-3.72**	-0.005	-3.43**	-0.012	-5.47**
Walkability	0.015	17.52**	0.014	16.49**	0.014	9.15**
LAMBDA					0.637	177.43**
RHO			0.013	21.34**		

* and ** denote coefficient significant at 0.05 and 0.01 level respectively.

Source: Estimated by the author using GeoDa 0.9.5.

In terms of goodness-of-fit statistics, such as log likelihood, AIC, and SC, the spatial-error models outperforms spatial-lag models and OLS models. The existence of the spatial-error-type autocorrelation suggests that some variables not included in the OLS model are spatially-correlated. The impacts of these missing variables are captured by the spatially-lagged error term in the spatial error model. Models with built-environment variables generally have better fit statistics than corresponding models with built-environment factors, but the results are harder to interpret. I use the three pedestrian-environment related variables in Model 1 as an example. The variable “percent of roads with sidewalks” has a positive and significant coefficient, while “percent of roads with curbs” and “average sidewalk width” both have negative and significant coefficients. A model with such contradictory results cannot be used to inform policy making very well. A review of the correlation matrix shows that the three variables are highly correlated, which may contribute to the counter-intuitive results. Model 2 uses built-environment factors instead. The “walkability” factor captures the underling force of these individual road characteristics and gets a positive and significant coefficient, which is a more understandable and useful result.

In general, inclusion of built-environment variables/factors does not change signs of structural variables, but indeed affect magnitude of the coefficients. The structural variables have expected signs, and are statistically significant at the 0.01 level. The quarterly housing price index computed using the results of Model 6 has the same evolution pattern as the index by the Office of Federal Housing Enterprise Oversight. It increased gradually from Q1 2004, peaked at Q3 2005, and then began its decline to Q4 2006. This consistency shows that the model at least captures the fluctuation in the general housing market without significant mistakes. As for other control variables, high median household income, high percentage of white population, low residential tax rate, low crime rate, and good school scores tend to increase property values.

Built-environment factors appear to capture most of the explanatory power of built-environment variables, and are much easier to interpret. After controlling for these variables, we find that built-environment factors are indeed associated with property values. Next, we discuss the effects of built-environment factors based on results of Models 2, 4, and 6 (the OLS, spatial-lag, and spatial-error models with built-environment factors).

Distance to Non-Work Destinations

Both the OLS (Model 2) and spatial-lag (Model 4) models suggest households would like to pay a premium for proximity to non-work destinations. In both cases, the *t* values have significance at the 0.01 level. However, accounting for the spatial-error term (Model 6) renders the factor insignificant. It suggests that the error term may contain some unobserved variables that are correlated with this factor and relevant to housing price at a different level of spatial aggregation.

Connectivity

The positive sign of the connectivity factor in all three models suggests that other things being equal, households value good connectivity – an indicator of a higher-density, locally accessible, grid-type neighborhood. The magnitude of this effect based on the spatial-error model is about half that of the OLS and spatial-lag models. If the "connectivity" score increases by 1.364 units, which is one standard deviation of this factor, the property value will increase 2.2% (Model 6), or 8.39 thousand dollars for a house priced at 376.5 thousand dollars (the median value of all single-family housing transactions).

Inaccessibility to Transit and Jobs

The negative sign of the coefficients for the "inaccessibility to transit and jobs" factor indicates households demand a discount for inaccessibility to transit and jobs. A one standard deviation (1.014 units) decrease of this factor can increase the property value by 8.1%, or 30.65 thousand dollars for a house priced at 376.5 thousand dollars (Model 6).

Auto Dominance

The "auto dominance" factor has a negative coefficient, which means households prefer locations further away from high-speed roads. This result is somewhat contrary to our expectation based on its impact on travel costs. I speculate that the relationship between the auto dominance factor and housing price can be attributed to: (a) a positive impact of increasing auto speed and reducing travel costs; and (b) a negative impact of high-speed roads, due to noise, emissions, and safety. In this study, the negative effect outweighs the positive effect. The net effect is that property values are estimated to decrease 0.7% (about 2.56 thousand dollars for a house priced at 376.5 thousand dollars) for one standard deviation (0.569 units) increase in the "auto dominance" factor (Model 6).

Walkability

The three models have stable estimates on the "walkability" effect. Based on the coefficient of the spatial error model, the positive sign indicates that households pay a premium to live in neighborhoods with a good pedestrian environment, controlling for other variables. If the "walkability" score increases by one standard deviation (0.971 units), the property value will increase around 1.4%, or 5.34 thousand dollars for a house priced at 376.5 thousand dollars.

4.5.3 Built-Environment Effects in Sub-Markets

Analysts suggest that the built-environment effect may depend on the historical development of neighborhoods (Matthews and Turnbull 2007). Because transit-oriented development is an important smart-growth strategy, I investigate whether the built-environment effect varies between transit-oriented neighborhoods and other neighborhoods. To do so, I divide the data into two sub-samples, one for houses with good transit accessibility, defined as locating within 800m (walking distance) to a subway station or bus stop, and one for houses locating beyond walking distance to a subway station or bus stop. I estimate the spatial-error model for the two sub-samples separately. The estimation results are presented in Table 13. To simplify the presentation, only coefficients of the built-environment factors are shown.

TABLE 13: Estimation Results of Sub-Models

	Observations within 800m of subway station / bus stop		Observations beyond 800m of subway station / bus stop	
Variables	Coeff.	t-stat.	Coeff.	t-stat.
Distance to non-work destinations	-0.007	-0.73	0.000	0.15
Connectivity	0.017	4.53**	-0.008	-4.07**
Inaccessibility to transit and jobs	-0.155	-10.84**	-0.057	-28.48**
Auto dominance	-0.001	-0.09	-0.015	-6.70**
Walkability	0.013	3.38**	0.002	0.88
LAMBDA	0.824	152.36**	0.517	106.86**
No. of observations	28023		64751	
Pseudo R-squared	0.833		0.785	

* and ** denote coefficient significant at the 0.05 level and 0.01 level respectively.

Source: Estimated by the author using GeoDa 0.9.5.

The coefficient of the spatially-lagged error term is highly significant in both sub-models, which rejects the OLS model and confirms the existence of spatial-error-type autocorrelation. As shown in the table, signs remain the same for all significant built-environment factors except for connectivity, although magnitudes of coefficients vary between the sub-models.

For the "distance to non-work destinations" factor, all coefficients are once again insignificant, although the sub-sample of houses with good transit-accessibility has a coefficient of -0.007, suggesting households may demand a premium for proximity to non-work destinations.

In terms of the "connectivity" factor, households choosing to live close to transit stations pay a premium for traditional grid-type, high-density neighborhoods, as reflected by the positive coefficient of the connectivity factor. This premium is 2.4% of the housing value for one standard deviation of increase (1.364 units) in the factor score, or 8.93 thousand dollars for a house priced at 376.5 thousand dollars. However, households living beyond walking distance to transit stations value cul-de-sac-type street network more, and they want a 1.0% discount for one standard deviation increase in the connectivity score. Both effects are statistically significant.

The coefficients for the "inaccessibility to transit and jobs" factor are negative in both sub-models. Households choosing neighborhoods with good transit-accessibility pay a premium of 14.5% of the housing value for one standard deviation (1.014 units) of decrease in the factor score, while households in the other sub-sample would pay only 5.6% of the housing value.

There is no significant effect for the "auto dominance" factor in the good-transit-accessibility sub-model. However households in the other sub-model demand a 0.8% discount for one standard deviation (0.569 units) increase in the factor. Hence, the "auto dominance" factor shows little difference in the city, but it matters in suburban areas.

Households in good-transit-accessibility neighborhoods care more about the pedestrian environment than households in other neighborhoods. They pay a premium of 1.2% of housing value for one standard deviation (0.971 units) increase in the factor, while in the other sub-model, this effect is insignificant. The little difference of pedestrian environment in the suburban area may contribute to this insignificance.

The different premiums for the built environment between the two submarkets may be partly attributed to life style preference. Transit-oriented households may purposely choose to live in transit-friendly neighborhood, thus pay higher premium for built-environment features that favor transit. The coexistence of spatial-error-type autocorrelation and submarkets may suggest that some omitted variables, such as life style preference, are correlated at different spatial scales. These omitted variables may help explain the formation of submarkets.

4.6 CONCLUSIONS

In this paper, I examine the relationship between the built environment and residential property values. Taking advantage of two recent datasets with exceptional spatial detail, I compute a set of built-environment variables at 250x250m grid cell level, apply factor analysis to mitigate multicollinearity, and integrate the built-environment variables/factors into hedonic-price models. I apply spatial-regression techniques to correct spatial autocorrelation. Also, I divide the data into two sub-samples to investigate the built-environment effects in submarkets. By using a cross-sectional analysis, I cannot identify causal relationships between the built environment and property values, and the potential endogeneity could bias the estimates of the models. Solving these issues necessitates either before-and-after datasets, used by Rodriguez and Mojica (2009) or more complex econometric models, such as the instrumental-variable approach employed by Song and Knaap (2004). However, potential instruments, such as land-use regulations, applied at the municipal level will not enable differentiation at the 250x250m grid cell detail used in this study. Although I lack instrumental variables at the fine-grained spatial detail, my analysis reveals significant association between the built environment and property values at a very disaggregate scale – associations that will have to be explained if and when appropriate data become available from a before-and-after study.

Using goodness-of-fit statistics to rank the models, I find that the spatial-error model is the best model, followed by the spatial-lag model and the OLS model. Compared with the results of the OLS model, using spatial econometrics models changes the magnitude of the estimated coefficients of built-environment factors, but the direction of most built-environment factors does not change. Although models with built-environment variables have better fit statistics than corresponding models using built-environment factors, the multicollinearity between built-environment variables cause a number of insignificant and counter-intuitive coefficients, which impairs the power of the models in informing policy design. Factor analysis helps get more interpretable results.

The empirical results suggest that property values are positively associated with “connectivity” and “walkability”, and negatively related to “inaccessibility to transit and jobs” and “auto dominance”. The built-environment effects depend on neighborhood characteristics. Households living within walking distance to transit stations pay higher premiums for good accessibility to transit, jobs, and non-work destinations, good connectivity, and good walkability than other households.

The research findings have important policy implications. Generally, this study suggests smart-growth policies that focus on increasing transit accessibility, bringing jobs closer to residence, creating traditional type, well-connected, high-density neighborhoods, reducing auto speed with traffic management measures and improving pedestrian environment are positively associated with residential property values. Although finding association is different from constructing causality, the research findings still provide some support for the argument that smart growth can improve quality of life of neighborhoods, thus increase local property values (Nelson et al. 2002). Sorting out the impact of smart growth on local neighborhoods may help relieve the concerns about smart growth at the local level. The existence of submarkets for the built environment suggests that smart-growth-type built-environment characteristics do not have universal appeal to households, but they no doubt satisfy an important market segment.

CHAPTER FIVE: SELECTIVITY, SPATIAL AUTOCORRELATION, AND VALUATION OF THE BUILT ENVIRONMENT

5.1 INTRODUCTION

Houses are heterogeneous goods, and their prices depend on the level and quality of their characteristics. These characteristics include not only structural attributes of the house per se, but also characteristics of the location. As an important component of locational factors, the built environment could influence property values as indicated by various analysts (e.g., Cao and Cory 1981; Song and Knaap 2004; Bowes and Ihlanfeldt 2001; Matthews and Turnbull 2007; and Rodriguez and Mojica 2009).

From a policy perspective, analysts need to understand how the built environment is valued by households in the market place. To reduce transportation energy use and emissions and achieve sustainable metropolitan growth, various smart-growth policies are currently implemented by governments and planning agencies. These policies aim to reshape household travel behavior and curb travel demand by changing the built environment via such mechanisms as regional planning, zoning, and provisions of alternative transportation modes. On the one hand, gauging the built-environment effect on property values makes it feasible for analysts to discuss and quantify the implicit tradeoffs associated with smart-growth policies on a neighborhood. On the other hand, capturing the value-added effect of certain built-environment features such as transit accessibility provides policy makers a potential public-financing mechanism to relieve the heavy financial burdens facing governments and transit agencies worldwide.

The dominant technique to value housing attributes is hedonic-price analyses, pioneered by Griliches (1971) and formalized by Rosen (1974). This method is easily replicable, and is thus widely used in application. Many previous analysts have investigated the property-value effect of various built-environment attributes using hedonic-price models, such as land-use mix (e.g., Cao and Cory 1981; Song and Knaap 2004), transit accessibility (e.g., Bowes and Ihlanfeldt 2001; Rodriguez and Mojica 2009), and street network layout (e.g., Matthews and Turnbull 2007). However, the conventional hedonic-price approach may suffer from two major limitations in valuing housing attributes:

First, the OLS-based hedonic-price analysis can generate biased estimates of the willingness-to-pay (WTP) for housing attributes when the assumption that these attributes are exogenous to sample selection is violated. Heckman (1979) discusses the bias that results from using non-randomly selected samples to estimate behavioral relationships as an "omitted variables" bias. Analysts usually calibrate hedonic-price models with samples of sold properties. In the housing market, only a small fraction of properties sells in a single quarter or year. If the sample of sold properties is a non-random sample of the housing stock, the hedonic-price model may generate biased estimates (Gatzlaff and Haurin 1998). A number of analysts have explored the impact of sample selection in the housing market, such as Haurin and Hendershott (1991), Jud and Seaks (1994), Gatslaff and Haurin (1998), and Hwang and Quigley (2004).

The importance of the selection bias depends on the purpose of study. If it is intended to improve measures of the market prices of housing attributes of sold properties, then the selectivity issue is not relevant. If analysts intend to use the model to make an inference about the housing stock, however, they cannot ignore the sample selection bias. To assess the property-value effect of smart-growth policies on local neighborhood or design land value capture scheme to support infrastructure investment, analysts may find it relevant to understand the impact of the built environment on the entire housing stock.

Second, a hedonic valuation of housing attributes can be misleading when spatial autocorrelation exists. In spatial-data analyses, a spatial autocorrelation refers to the phenomenon that a value observed in one location depends on the values at neighboring locations. There is consistent evidence that property values exhibit a systematic pattern in their spatial distribution (see, e.g., Basu and Thibodeau 1998, among others). Analysts apply various approaches to deal with the spatial autocorrelation, for example, the spatial econometric techniques (Anselin 1993), the Cokriging approach (Chica-Olmo 2007) and the Geographically Weighted Regression approach (Fotheringham, Brunsdon, and Charlton 2002).

In this paper, I contribute to the literature by accounting for both selectivity and spatial autocorrelation in valuing the built environment. I apply the Heckman two-step procedure to correct for sample selection bias, and integrate spatial econometric techniques into the Heckman-selection model to overcome spatial autocorrelation. Based on the modeling results, I compute the willingness-to-pay for built-environment attributes and compare them with results of conventional OLS-based hedonic-price analysis to investigate the impact of selectivity and spatial autocorrelation in the valuation.

This paper is organized as follows. Section 5.2 describes the analysis techniques. Section 5.3 introduces an empirical study for the City of Boston, including datasets, variables, and modeling results. Section 5.4 summarizes research findings and discusses policy implications.

5.2 METHODOLOGY

Hedonic-price model is widely used in the valuation of housing attributes. A conventional hedonic-price model can be specifies as:

$$\ln P_{it} = \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + \varepsilon_{it} \tag{1}$$

where P_{it} is the transaction price of property i; X_{ijt} is a set of j structural characteristics; Z_{kjt} is a set of k locational characteristics, including built-environment attributes; D_{it} is a set of dummy variables such that they take the value 1 for transactions taken place in time period t, and 0 otherwise; and ε_{it} is normally distributed with a mean zero random error.

In this study, I employ a housing sales model used by Gatzlaff and Haurin (1998). This model represents a double-sided search market with heterogeneous participants and heterogeneous properties. Observable transaction prices are derived from the interaction between two populations of market participants: potential buyers on the demand side and potential sellers on the supply side. In the housing market, both the buyer and the seller have their own evaluations of the asset-specific characteristics, which lead to their prices for the properties. The hedonic-price equations for the buyer and the seller take the following forms, respectively:

$$\ln P_{it}^{b} = \sum \alpha_j^b X_{ijt} + \sum \beta_k^b Z_{ikt} + \sum \gamma_t D_{it} + \varepsilon_{it}^{b} \tag{2}$$

$$\ln P_{it}^{s} = \sum \alpha_j^s X_{ijt} + \sum \beta_k^s Z_{ikt} + \sum \gamma_t D_{it} + \varepsilon_{it}^{s} \tag{3}$$

where P_{it}^{b} is the offer price of the buyer and P_{it}^{s} is the reservation price of the seller for house i; $\alpha_j^b X_{ijt}$ and $\beta_k^b Z_{ikt}$ components reflect the systematic valuation of structural and locational characteristics common to all potential buyers; $\alpha_j^s X_{ijt}$ and $\beta_k^s Z_{ikt}$ reflect the systematic valuation of structural and locational characteristics common to all potential sellers; $\varepsilon_i^b (\varepsilon_i^s)$ is normally distributed with a mean zero random error.

I consider a transaction is completed when the buyer's offer price is higher than or equal to the seller's reservation price. Thus, properties sold in the market are not necessarily random draws from the population of houses. The possibility of sample selection bias arises when the unobserved housing characteristics affecting the transaction-sales propensity also influence the transaction-price level. The transaction price can be modeled as:

$$\ln P_{it} = \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + (\varepsilon_{it} \big| P_{it}^b \geq P_{it}^s) \tag{4}$$

It should be noted that the error term in Equation (4) may have a nonzero mean because the observed transaction sample contains only selected properties, i.e., houses with a buyer's offer price higher than or equal to the seller's reservation price. When $E[\varepsilon_{it} \mid P_{it}^b \geq P_{it}^s] \neq 0$, an OLS regression using the observed transactions produces biased estimates.

To correct for the potential sample selection bias, I apply the Heckman two-step procedure (Heckman 1979). In the first step, I model the probability that a property is sold with a binary-probit model. I use S_{it} to denote the outcome, and S_{it}^* to denote the difference between the offer and reservation prices. It should be noted that S_{it}^* is not observable, only the outcome S_{it} can be observed.

$$\begin{cases} S_{it} = 1, \text{if } S_{it}^* \geq 0 \\ S_{it} = 0, \text{otherwise} \end{cases} \tag{5}$$

Equation (5) is calibrated as a probit model using the entire housing stock:

$$\Pr[S_{it} = 1] = \Phi[\sum \varpi_j X_{ijt} + \sum \mu_k Z_{ikt}] \tag{6}$$

where *Φ* is the cumulative distribution function of standard normal distribution. Based on the estimation results of the probit model, I compute the inverse mills ratio as:

$$\lambda_{it} = \phi(\sum \varpi_j X_{ijt} + \sum \mu_k Z_{ikt}) / \Phi(\sum \varpi_j X_{ijt} + \sum \mu_k Z_{ikt}) \tag{7}$$

where *φ* and *Φ* denote the probability-density function and cumulative-distribution function of the standard normal distribution, respectively. In the second step of the Heckman procedure, the inverse mills ratio is included as an independent variable in the standard hedonic-price model, such that

$$\ln P_{it} = \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + \chi\lambda_{it} + \varepsilon_{it} \tag{8}$$

The inclusion of the inverse mills ratio corrects for the bias due to sample selection (Heckman 1979).

The classical Heckman procedure does not account for spatial autocorrelation. To solve the spatial autocorrelation problem, I integrate spatial econometric techniques into the Heckman-selection model. In the second step of the Heckman procedure, I expand the standard Heckman-selection model by adding in two spatial autoregressive terms to correct for two types of spatial autocorrelation respectively. For the first type of spatial autocorrelation, I assume that value of a property is influenced by the characteristics of neighboring properties. In this case, the OLS estimation will be biased and inefficient. This type of spatial autocorrelation can be solved by adding an additional regressor in the form of a spatially-lagged dependent variable to the regression, as is shown in Equation (9).

$$\ln P_{it} = \rho W_{\ln P} + \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + \chi\lambda_{it} + \varepsilon_{it} \tag{9}$$

where W_{lnP} is the spatial lag variable; ρ is a spatial lag correlation parameter, and ε is an Nx1 vector of i.i.d. standard normal errors. For the second type of spatial autocorrelation, I assume that housing attributes captured by the model have only local effects, but factors missing from the model specification are spatially correlated. In this case, the OLS estimation will be inefficient. This type of spatial autocorrelation can be corrected for by adding a spatially-lagged error term into the model, as is shown in Equation (10).

$$\begin{aligned} \ln P_{it} &= \sum \alpha_j X_{ijt} + \sum \beta_k Z_{ikt} + \sum \gamma_t D_{it} + \chi\lambda_{it} + \varepsilon_{it} \\ \varepsilon_{it} &= \tau W_{\varepsilon} + \mu_{it} \end{aligned} \tag{10}$$

where $W\varepsilon$ is the weighted average of error terms in neighboring areas; τ is a spatial-error correlation parameter, and μ is an Nx1 vector of i.i.d. standard normal errors.

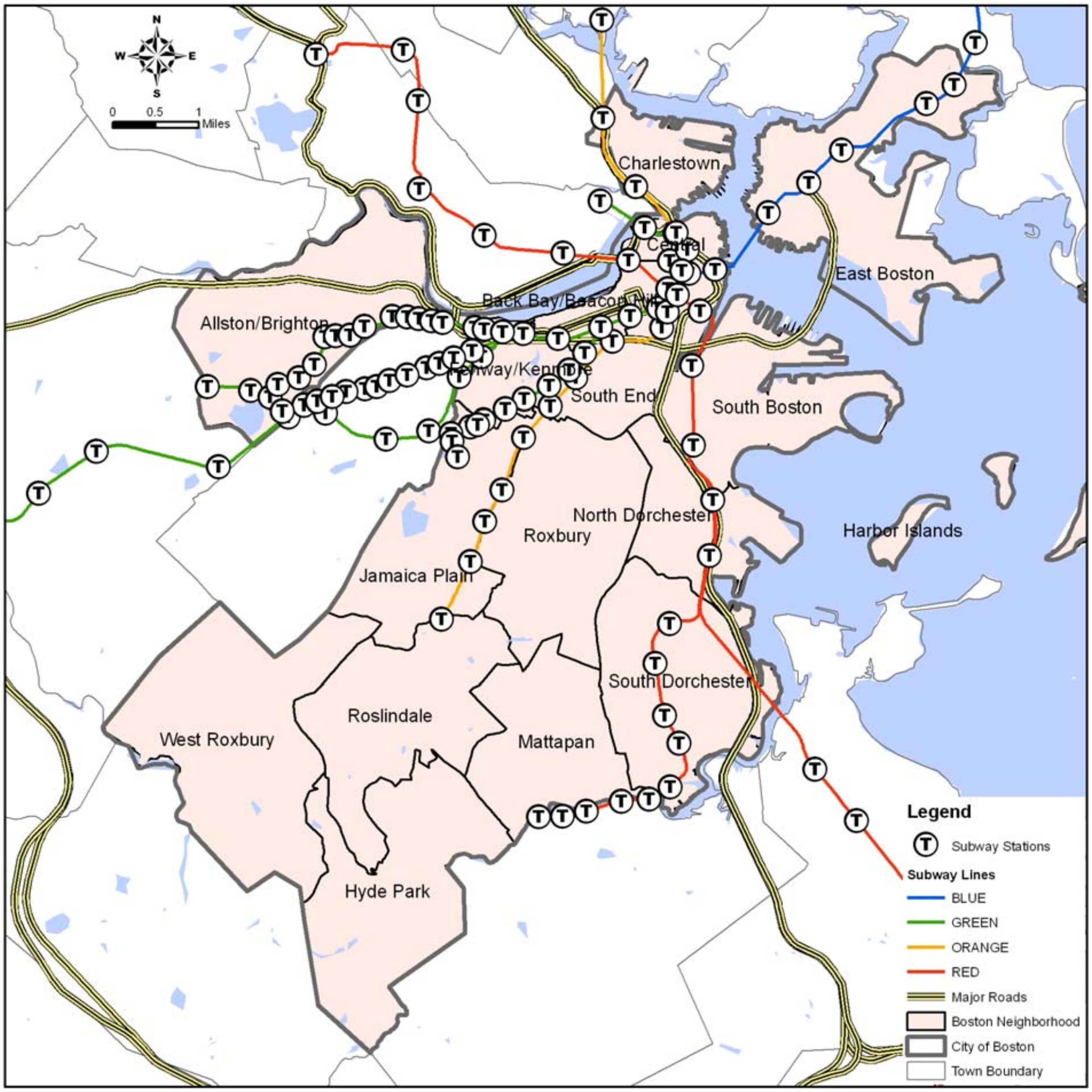
5.3 EMPIRICAL ANALYSIS

In this section, I present an empirical analysis based on the analytic framework discussed in Section 5.2.

5.3.1 Study Area and Data

The study area of the empirical analysis is City of Boston[5], the central part of the Boston Metropolitan Area. Figure 13 shows a map of Boston.

[5] I can only get housing stock data in the City of Boston. Therefore, in the third study, I use City of Boston as the study area. The first two essays use the Boston Metropolitan Area as the study area.



Source: The author.

Figure 13: City of Boston

On the housing side, I use both the transaction and stock data of single-family properties[6] in the City of Boston. The assessing records from the Assessing Department of Boston contain detailed information about all residential properties in the city, such as structural characteristics,

[6] In this study, single-family properties refer to properties with property type code "101" as defined by MassGIS.

tax information, and street address. The total number of single-family properties[7] in Boston is about 30,000, varying slightly over time. The housing transaction records from the Suffolk County Registry of Deeds provide data on all single-family housing transactions over the study period (1998-2007), including date of sale, transaction price, and street address. I link the transaction data to the assessing data based on street address using GIS tools. After excluding transactions that are unreliable or cannot be matched to the assessing data, there are 10,031 single-family housing transactions in the study period. The final datasets for analysis include a total of 1,198,031 observations, which is comprised of every parcel of single-family properties in Boston multiplied by the number of quarters the house was included in the assessing data. Because characteristics of unsold properties are included in the assessing data, I can apply the Heckman two-step procedure to correct for sample selection bias.

The built-environment data come from spatially-detailed datasets provided by MassGIS – the State's Office of Geographic and Environmental Information, including location of common-trip destinations, spatial distribution of households and jobs, land use, and transportation networks.

5.3.2 Variable Generation

For each single-family property in the City of Boston, I create four sets of variables: (1) built-environment variables; (2) structural attributes; (3) neighborhood socioeconomic characteristics; and (4) macroeconomic measures for each quarter during the study period.

One well-known challenge in spatial analysis is the Modifiable Areal Unit Problem (MAUP) – the inconsistency in measurement results and statistical analyses due to the choice of

[7] In this study, single-family properties are defined as properties with state use code "101".

neighborhood boundaries. To mitigate the MAUP, the basic spatial unit used in this study is a 250x250m grid cell layer developed by MassGIS as discussed in Chapter 2.

5.3.2.1 Built-Environment Variables

Based on literature, I compute 10 built-environment variables along four dimensions: density, land-use mix, street-network layout, and accessibility[8]. I use GIS and database management tools extensively in the computation.

Density: Density is an important indicator of the built environment. Population density (population divided by land area) is widely used in previous studies as a measure of density. However, the way density is measured can introduce significant bias when the proportion of residential use differs across neighborhoods. In this study, I compute residential density (population divided by the area of residential use) to capture a more realistic meaning of density. Population and household data are from the 2000 Census and constrained by MassGIS to those areas identified as residential by the 2000 land use dataset. MassGIS further allocated population and households to 250x250m grid cells. I assign the residential density in the 9-grid-cell catchment area to each grid cell.

Land-use mix: Land-use mix measures the degree to which land uses are mixed and balanced within the neighborhood. A greater mix of uses could facilitate walking and biking, reduce vehicle trips generated and vehicle miles traveled, and enhance urban aesthetics. This study uses a computational method based on the concept of entropy (Turner, Gardner, and

[8] The selection of 10 built-environment variables in Essay 3 is different from the previous two essays using 27 built-environment variables or 5 built-environment factors. The major reason is that the City of Boston has much smaller variation in the built environment than the Metro Boston. I tried applying factor analysis to the built-environment variables in City of Boston, but did not get meaningful factors. Therefore, in Essay 3, I select 10 built-environment variables that are identified as theoretically important and practically popular by literature.

O'Neill 2001). The idea is that a neighborhood containing each of the land-use types in the same proportions would obtain a maximum entropy value. It is computed as:

$$-\sum_j P_j * \ln(P_j) / \ln(J) \qquad (11)$$

where P_j is the proportion of land in the *j*th land-use category and J is the total number of land-use categories considered. In this study, J=5: single family, multi-family, commercial, industrial, and recreation and open space. This measure varies between 0 and 1. A value of 0 means the land is exclusively dedicated to a single use, while a value of 1 suggests perfect mixing of the five land-uses. I first compute the land-use-mix index for each 250x250m grid cell. Then, I assign each grid cell a value that equals the mean of the nine grid cells in the catchment area.

Street network layout: The layout of street networks is also an important factor of the built environment. To show the differences between the sprawl and traditional type of block patterns, I compute intersection density in each grid cell's catchment area as an indicator.

Accessibility: It is well-known that good accessibility can save the transportation cost of households, thus be capitalized into property values. In this study, I aim to capture the property-value effects of accessibility to activity centers such as jobs, non-work destinations, and the central business district (CBD), as well as the effects of accessibility to transportation networks, including subway station, commuter rail station, Massachusetts Bay Transit Authority (MBTA) park-and-ride lots, and highway exits,.

The job accessibility measure I use in this study is a gravity-type job accessibility indicator computed at the transportation analysis zone (TAZ) level, which takes the following form known as the Hansen accessibility model (Hansen 1959):

$$A_i = \sum_j O_j f(C_{ij}) \qquad (12)$$

where $f(C_{ij}) = \exp(-\beta * C_{ij})$, O_j is the number of jobs in TAZ j, $f(C_{ij})$ is an impedance function, C_{ij} is the network distance between TAZ i and j. β is set to 0.1, based on Zhang's calibration using an Activity–Travel Survey conducted by the Central Transportation Planning Staff for the Boston region (Zhang 2005).

MassGIS utilizes the Dun and Bradstreet business-location database to identify locations of 27 types of common non-work destinations in Metro Boston and computes a weighted average minimal Euclidian distance to major non-work destinations[9] at 250x250m grid cell level. They use the national average trip rate for each type of non-work destination from the 2001 National Household Transportation Survey as the weight in the computation.

The distance to CBD indicator measures the Euclidian distance to the Downtown Crossing subway station, which locates at the center of Boston's CBD area.

In this study, I compute four indicators to measure accessibility to transportation networks, including (1) presence of subway station within half mile, (2) presence of commuter rail station within half mile, (3) distance to MBTA parking lots, and (4) distance to highway exits. Presence of subway station (or commuter rail station) within half mile is a dummy variable, which takes the value 1 if a subway station (or commuter rail station) is within half miles of the property, and 0 otherwise. The distance to highway exits and distance to MBTA parking lots indicators are both measured as Euclidian distances to the corresponding transportation nodes. The MBTA provides parking space at some subway stations and commuter rail stations for travelers switching to transit.

---

[9] Common trip destination types covered include grocery stores, pharmacy, banks, daycare centers, auto repair stores, gas stations post offices, bars, clothing stores, convenience stores, dentist offices, drycleaners, fitness centers, beauty/nail salons and barber shops, hardware stores and home centers, motion picture theaters, museums, historical sites, performing arts centers/theaters, physician offices, non-physician, non-dentist, medical doctor offices, restaurants, sport facilities, veterinary service locations, religious institutions, and schools.

Unlike pervious studies that focus primarily on the built-environment effect on property values, this study also tries to investigate the built-environment effects on the probability of housing sales, which may lead to the sample selection bias in valuing the built environment.

5.3.2.2 Structural Variables

I select nine structural variables in the analysis: (1) lot size, (2) gross area, (3) year built, (4) number of floors, (5) total number of rooms, (6) number of full baths, (7) number of half baths, (8) a dummy variable showing the existence of air conditioning, and (9) number of fireplaces. Lot size and gross area are both measured in logarithms. The structural characteristics could influence both the probability of housing sale and transaction price, as suggested by previous analysts (e.g., Gatzlaff and Haurin 1998).

5.3.2.3 Socioeconomic Variables

Socioeconomic characteristics of the neighborhood could also influence property values. To control for this effect, I include percentage of white population as a measure of racial composition, and median household income as a measure of wealthy level. I measure both variables at the census-block-group level.

5.3.2.4 Macroeconomic Variables

National and local economic conditions may help explain variations in the probability of housing sales (Jud and Seaks 1994). To capture this impact, I include three variables, representing the gross national product (GNP), the national level mortgage rate, and the local unemployment rate. I expect that heightened economic activities increase the probability of housing sales.

Table 14 presents the descriptive statistics of the sold sample and the housing stock.

TABLE 14: Descriptive Statistics

Variable	Sold Properties Obs.	Sold Properties Mean	Sold Properties Std. Dev.	All Properties Obs.	All Properties Mean	All Properties Std. Dev.
ln(transaction price)	10031	12.644	0.598			
Structural Variables						
ln(lot size)	10031	8.267	0.690	1198031	8.366	0.640
ln(gross area)	10031	7.897	0.313	1198031	7.910	0.312
Year built	10031	1921	48.538	1198031	1924	43.973
Number of floors	10031	1.909	0.586	1198031	1.848	0.560
Total number of rooms	10031	7.142	1.780	1198031	7.148	1.795
Number of full bath	10031	1.362	0.613	1198031	1.291	0.560
Number of half bath	10031	0.526	0.549	1198031	0.518	0.547
Presence of A/C	10031	0.135	0.342	1198031	0.097	0.295
Number of fireplaces	10031	0.543	0.849	1198031	0.522	0.754
Built-Environment Variables						
Population density (k/km2)	10031	5.773	3.527	1198031	5.350	3.260
Land-use mix	10031	0.439	0.241	1198031	0.423	0.236
Intersection density (1/km2)	10031	116.771	40.629	1198031	113.288	37.416
Presence of subway sta. within half mile	10031	0.336	0.472	1198031	0.293	0.455
Presence of commuter rail sta. within half mile	10031	0.345	0.475	1198031	0.351	0.477
Distance to MBTA parking lots (km)	10031	1.717	1.173	1198031	1.643	1.139
Distance to highway exits (km)	10031	3.271	1.796	1198031	3.382	1.740
Distance to CBD (km)	10031	8.068	3.548	1198031	8.535	3.390
Job accessibility (k)	10031	461.346	94.496	1198031	448.635	89.111
Distance to non-work destinations (km)	10031	1.006	0.250	1198031	1.036	0.247
Macroeconomic Variables						
GNP (billion $)	10031	11406	1655	1198031	11224	1670
Mortgage rate	10031	6.531	0.716	1198031	6.617	0.739
Local unemployment rate	10031	4.204	0.999	1198031	4.121	1.043

Source: Calculated by the author.

Compared with the housing stock, the sold sample on average has a smaller lot size and gross area, more floors, bath rooms, and fireplaces, and is older in age and more likely to have air conditioning. Generally, the sold properties also tend to locate in smart-growth type neighborhoods with higher population density, land-use mix, and intersection density, better accessibility to transit stations and highway exits, a little further away from park-and-ride lots, but closer to jobs, non-work destinations and the CBD area than the housing stock. The differences between the sold properties and the housing stock suggest the potential existence of selection bias.

It may be helpful to look at the temporal change in the characteristics of the housing transactions. Table 15 presents the average structural and built-environment characteristics of the sold properties year-by-year. During the study period, the average transactions price started growing from 1998, peaked in 2005 and decreased slightly in 2006 and 2007. The last column in Table 16 shows the correlations of housing attributes and transaction price. When I focus on housing attributes with correlation coefficients significantly different from 0, I find that more transactions of relatively low-quality properties (small gross area, old in age, few floors and rooms) occurred as the housing price increases. One possible explanation is that households became "priced out" of the top tier of expensive properties. The average built-environment attributes of the sold properties also varies with time. In particular, population density, proximity to commuter rail stations, and job accessibility are negatively correlated with transaction price, while distance to CBD is positively associated with transaction price. Although a simple univariate analysis, it suggests that different pools of properties are transacted over time, which might be another indication of the sample selection problem.

TABLE 15: Annual Changes in Structural and Built-Environment Characteristics of the Sold Properties

Variables	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Corr. with ln(price)
ln (price)	12.04	12.15	12.36	12.49	12.65	12.77	12.88	12.94	12.89	12.89	
Structural Variables											
ln(lot size)	8.27	8.23	8.23	8.19	8.23	8.24	8.32	8.32	8.32	8.26	0.512
ln(gross area)	7.92	7.91	7.90	7.89	7.90	7.90	7.89	7.88	7.89	7.91	-0.685**
Year built	1917	1916	1917	1920	1917	1919	1926	1927	1926	1924	0.833***
Number of floors	1.95	1.93	1.96	1.96	1.94	1.91	1.88	1.84	1.85	1.92	-0.702**
Total number of rooms	7.27	7.28	7.20	7.12	7.22	7.20	7.12	7.00	6.94	7.17	-0.713**
Number of full bath	1.36	1.36	1.37	1.35	1.35	1.32	1.39	1.37	1.33	1.40	0.133
Number of half bath	0.51	0.52	0.50	0.53	0.49	0.48	0.55	0.55	0.55	0.56	0.452
Presence of A/C	0.12	0.12	0.14	0.13	0.12	0.09	0.15	0.15	0.14	0.17	0.452
Number of fireplaces	0.59	0.55	0.50	0.51	0.45	0.51	0.56	0.61	0.54	0.59	0.099
Built-Environment Variables											
Population density (k/km^2)	5.95	6.09	6.08	6.05	5.89	5.86	5.52	5.32	5.50	5.90	-0.738**
Land-use mix	0.44	0.45	0.44	0.44	0.45	0.46	0.42	0.42	0.43	0.44	-0.308
Intersection density (1/km^2)	118.21	118.29	120.48	118.28	116.91	118.16	114.37	113.39	113.98	119.00	-0.607*
Presence of subway station within half mile	0.34	0.36	0.36	0.38	0.36	0.35	0.30	0.30	0.30	0.36	-0.523
Presence of commuter rail station within half mile	0.36	0.38	0.35	0.35	0.35	0.33	0.31	0.35	0.34	0.34	-0.712**
Distance to MBTA parking lots (km)	1.71	1.77	1.78	1.75	1.77	1.76	1.69	1.60	1.69	1.73	-0.477
Distance to highway exits (km)	3.28	3.14	3.15	3.21	3.25	3.11	3.39	3.45	3.36	3.22	0.480
Distance to CBD (km)	7.93	7.80	7.73	7.63	7.82	7.86	8.38	8.60	8.48	7.95	0.630*
Job accessibility (k)	464.64	468.67	470.13	471.99	467.33	466.46	453.22	447.95	450.22	466.03	-0.614*
Distance to non-work destinations (km)	1.01	0.99	0.98	0.98	1.00	0.99	1.03	1.04	1.03	0.99	0.500

*, **, and *** denote significant at the 0.10, 0.05, and 0.01 level, respectively.

Source: Calculated by the author.

5.3.3 Estimation Results

In this study, I calibrate four models to value the built environment and assess the impact of selectivity and spatial autocorrelation by comparing the estimation results across models. The models are:

1. Conventional hedonic-price model (Equation 1);

2. Classical Heckman-selection model (Equations 6 and 8);

3. Heckman-selection model with spatially-lagged dependent variable (Equations 6 and 9, referred to as Heckman-selection model with spatial lag thereafter), and

4. Heckman-selection model with spatially-lagged error term (Equations 6 and 10, referred to as Heckman-selection model with spatial error thereafter).

Models 2-4 are based on the Heckman two-step procedure. In the first step of the Heckman procedure, I use structural variables, macroeconomic variables, built-environment variables, 15 neighborhood dummy variables, and 3 quarter dummy variables to predict the probability that a property is sold in the market using a probit model. I base the neighborhood dummy variables on planning districts defined by the Boston Redevelopment Authority, which is widely used in planning practice. The default is East Boston. Q2, Q3, and Q4 are three dummy variables that take the value of 1 when the transaction took place in quarter 2, quarter 3, and quarter 4 respectively, and 0 otherwise. Table 16 reports estimation results of the probit model.

TABLE 16: Estimation Result of the Probit Model

Variables	Coef.	t-Stat.
Constant	-0.9059	-3.06 ***
Structural Variables		
ln(lot size)	-0.0598	-5.75 ***
ln(gross area)	-0.1110	-5.96 ***
Year built	-0.0002	-3.54 ***
Number of floors	0.0254	2.69 ***
Total number of rooms	-0.0065	-2.17 **
Number of full bath	0.0866	11.01 ***
Number of half bath	0.0253	3.46 ***
Presence of A/C	0.1206	10.08 ***
Number of fireplaces	0.0157	2.86 ***
Macroeconomic Variables		
GNP	0.0074	2.77 ***
Mortgage rate	-0.0683	-7.04 ***
Unemployment rate	-0.0122	-1.90 *
Built-Environment Variables		
Population density (k/km^2)	0.0047	2.11 **
Land-use mix	0.0022	0.10
Intersection density (1/km^2)	-0.0003	-1.58
Presence of subway station within half mile	0.0163	1.43
Presence of commuter rail station within half mile	0.0082	0.86
Distance to MBTA parking lots (km)	-0.0025	-0.32
Distance to highway exits (km)	-0.0018	-0.30
Distance to CBD (km)	0.0201	2.23 **
Job accessibility (k)	0.0011	3.28 ***
Distance to non-work destinations (km)	-0.0734	-2.56 ***
Neighborhood Dummy Variables		
Charlestown	-0.1603	-3.58 ***
South Boston	-0.1411	-4.44 ***
Central	-0.1966	-2.62 ***
Back Bay	-0.3967	-7.17 ***
South End	-0.2751	-5.55 ***
Fenway	-0.3013	-2.86 ***
Allston/Brighton	-0.1375	-2.70 ***

Jamaica Plain	-0.0199	-0.45
Roxbury	-0.0843	-2.04 **
North Dorchester	-0.0336	-0.81
South Dorchester	-0.0031	-0.10
Mattapan	-0.0885	-2.39 **
Roslindale	0.0240	0.51
West Roxbury	-0.0523	-1.04
Hyde Park	0.0101	0.21
Quarter Dummy Variables		
Q2	0.1505	13.79 ***
Q3	0.1697	15.72 ***
Q4	0.0817	6.95 ***
Observations	1198031	
LR chi-square(40)	1174.76	(p=0.000)

*, **, and *** denote significant at the 0.10, 0.05, and 0.01 level respectively.

Source: Estimated by the author using Stata 10.

The probit model is highly significant as shown by the value of χ^2 for testing the null hypothesis that coefficients of independent variables are simultaneously 0. The probability of housing sale differs for properties with different structural characteristics. Generally, smaller properties with smaller lot size, smaller gross area, and fewer rooms are more likely to sell than are larger properties. Older properties have a higher sale propensity than newer ones. Meanwhile, the sale probability is positively associated with numbers of floors, bathrooms and fireplaces, and the existence of air conditioning. The estimated coefficients for macroeconomic variables suggest that increased economic activity raise the probability of sale. The GNP variable has a positive and significant coefficient, as expected. The local unemployment rate variable has the expected negative sign, but its impact is only marginally significant at the 0.10 level. The negative sign of the mortgage rate variable shows that lower rates increase housing sales. The significance of multiple built-environment variables confirms the impact of the built environment on the probability of sale. Single-family properties in dense area, with good job accessibility, close to non-work destinations, but far away from the CBD, are more likely to be sold in the market than those with the opposite characteristics. The impacts of other built-environment variables are insignificant. There is also evidence that the probability of sale varies across neighborhoods and quarters of year for identical properties.

Table 17 compares the estimation results of the hedonic-price model, Heckman-selection model, Heckman-selection model with spatial lag, and Heckman-selection model with spatial error. The spatial weight matrix for the last two models is developed assuming constant spatial dependence between properties until a maximum distance is reached. The maximum Euclidean distance I used is 400m.

TABLE 17: Estimation Results of the Price Model

	(1) Hedonic-price model			(2) Heckman-selection model			(3) Heckman-selection model with spatial lag			(4) Heckman-selection model with spatial error		
	Coef.	t-Stat.		Coef.	t-Stat.		Coef.	t-Stat.		Coef.	t-Stat.	
Constant	6.5999	34.43	***	3.6700	17.15	***	1.5741	7.35	***	6.5317	12.97	***
Structural Variables												
ln(lot size)	0.0677	8.50	***	-0.0062	-0.76		0.0248	3.16	***	0.0353	2.46	**
ln(gross area)	0.2590	17.24	***	0.0970	6.19	***	0.1495	9.95	***	0.1741	6.62	***
Year built[a]	-0.0281	-4.13	***	-0.0716	-10.58	***	-0.0500	-7.71	***	-0.0404	-5.21	***
Number of floors	0.1150	15.41	***	0.1373	18.95	***	0.1010	14.39	***	0.1012	11.68	***
Total number of rooms	0.0112	4.63	***	0.0008	0.35		0.0070	3.07	***	0.0057	2.27	***
Number of full bath	0.1154	18.49	***	0.2421	31.81	***	0.1702	22.21	***	0.1467	8.09	***
Number of half bath	0.0483	8.09	***	0.0835	14.15	***	0.0662	11.69	***	0.0647	8.90	***
Presence of A/C	0.1136	11.98	***	0.3164	26.81	***	0.2168	18.41	***	0.1938	7.58	***
Number of fireplaces	0.0922	22.30	***	0.1002	25.04	***	0.0719	18.36	***	0.0612	12.27	***
Socioeconomic Variables												
Percent of population that is white	0.3642	28.56	***	0.3546	28.81	***	0.2488	20.11	***	0.2405	6.63	***
Median household income (k$)	0.0048	20.67	***	0.0041	18.09	***	0.0022	9.77	***	0.0008	2.40	**
Built-Environment Variables												
Population density (k/km2)	0.0180	11.61	***	0.0237	15.68	***	0.0145	9.89	***	0.0049	1.60	
Land-use mix	0.0179	1.09		0.0063	0.40		0.0069	0.46		-0.0074	-0.27	
Intersection density (1/km2)[a]	-0.0158	-1.41		-0.0521	-4.77	***	-0.0179	-1.71	*	0.0046	0.21	
Presence of subway station within half mile	0.0570	6.48	***	0.0983	11.40	***	0.0539	6.47	***	0.0303	1.96	**
Presence of commuter rail station within half mile	0.0070	0.98		0.0136	1.98	**	0.0111	1.70	*	-0.0128	-1.01	
Distance to MBTA parking lots (km)	-0.0838	-19.97	***	-0.0707	-17.33	***	-0.0384	-9.39	***	-0.0639	-3.38	***

Distance to highway exits (km)	-0.0095	-3.47	***	0.0168	5.97	***	0.0115	4.28	***	0.0096	0.66	
Distance to CBD (km)	0.0650	19.97	***	0.0654	20.84	***	0.0303	9.42	***	0.0086	0.68	
Job accessibility (k)[a]	0.3570	27.19	***	0.3580	28.30	***	0.1872	13.78	***	0.1651	4.31	***
Distance to non-work destinations (km)	0.1276	6.08	***	-0.0387	-1.83	*	0.0039	0.20		-0.0862	-1.71	*
Inverse mills ratio				1.9482	27.23	***	1.1316	15.09	***	1.0801	4.83	***
Spatially-lagged error term										0.8792	78.76	***
Spatially-lagged dependent variable							0.3705	28.22	***			
R-square	0.7541			0.7711			0.7913			0.8091		
Log likelihood				-1671.86			-1222.22			-896.51		

*, **, and *** denote significant at the 0.10, 0.05, and 0.01 level respectively.

a Coefficient is x 10^{-2}.

Source: Estimated by the author using Stata and GeoDa 0.9.5.

In terms of goodness-of-fit statistics such as R-square and log likelihood, the Heckman-selection model with spatial error outperforms the other three models. The spatially-lagged dependent variable and the spatially-lagged error term are both significant in the corresponding model, which confirms the existence of spatial autocorrelation. The coefficients of the inverse mills ratio in the three models using the Heckman procedure have a negative sign and are statistically significant. It suggests that the sample of sold properties is a non-random sample of the housing stock. Exclusive reliance upon the sample of sold properties tends to underestimate the value of properties in the entire housing stock. This result is consistent with Gatzlaff and Haurin (1998), while Jud and Seaks (1994), Gatzlaff and Haurin (1997), and Hwang and Quigley (2004) find that the housing-price index is overestimated as a result of sample selection bias.

The inclusion of the spatially-lagged error term in the Heckman-selection model decreases the magnitude and significance level of the inverse mills ratio. My interpretation is that some omitted variables related to the choice of property are spatially correlated. Their effects on property values are partially captured by the spatially-lagged error term. Therefore, the inverse mills ratio, the independent variable used to correct sample selection, is correlated with the spatially-lagged error term, which explains the drop in the "importance" of the inverse mills ratio.

In general, coefficients of most structural variables have expected signs and are statistically significant. In all models, higher median household income and higher percentage of white population tend to increase property values. Both coefficients are statistically significant at the 0.05 level. After controlling for structural and neighborhood socioeconomic characteristics, many built-environment variables still show significant associations with the transaction price.

The estimation results confirm the important role of accessibility in the housing market. Households in Boston pay a premium for living within walking distance to a subway station, as reflected by the positive and significant coefficients in all models. Controlling for selectivity can significantly increase the magnitude of this effect, but when spatial effects are further controlled, this premium decreases to a lower level before correction. Distance to MBTA parking lots has negative and significant coefficient, suggesting that households demand a negative premium for living faraway to park-and-ride lots. Accessibility to activity centers can also be capitalized into property values. Job accessibility has positive and highly significant association with property values as expected. Households pay a premium for proximity to non-work destinations according to the Heckman-selection model with spatial error, but this effect is marginally significant at the 0.1 level. Other built-environment variables have insignificant coefficients in the Heckman-selection model with spatial error.

The estimation results of the four models can be used to derive a set of marginal implicit prices for each attribute that represents the household's willingness-to-pay (WTP) for marginal increase in the individual housing attributes. Following Halvorsen and Palmquist (1980) and Crane et al. (1997), the WTP for a particular housing attributes *i* can be computed by

$$WTP_i = (\exp(\hat{\beta}_i) - 1)P \tag{13}$$

where $\hat{\beta}_i$ is the estimated coefficient of housing attribute *i* in a semi-log form price model and *P* is the transaction price. In this study, the WTP for built-environment attributes is computed for a property priced at 325.0 thousand dollars (the mean sale price of the sold sample). The results are reported in Table 18.

TABLE 18: Willingness-to-Pay for Built-Environment Variables

	(1) Hedonic-price model		(2) Heckman-selection model		(3) Heckman-selection model with spatial lag		(4) Heckman-selection model with spatial error	
Variables	Coef.	WTP(k$)	Coef.	WTP(k$)	Coef.	WTP(k$)	Coef.	WTP(k$)
Population density (k/km^2)	**0.0180**	**5.903**	**0.0237**	**7.794**	**0.0145**	**4.738**	0.0049	1.605
Land-use mix	0.0179	5.861	0.0063	2.057	0.0069	2.259	-0.0074	-2.382
Intersection density (1/km^2)[a]	-0.0158	-0.051	**-0.0521**	**-0.169**	**-0.0179**	**-0.058**	0.0046	0.015
Presence of subway station within half mile	**0.0570**	**19.075**	**0.0983**	**33.562**	**0.0539**	**17.993**	**0.0303**	**9.987**
Presence of commuter rail station within half mile	0.0070	2.270	**0.0136**	**4.442**	**0.0111**	**3.623**	-0.0128	-4.137
Distance to MBTA parking lots (km)	**-0.0838**	**-26.119**	**-0.0707**	**-22.169**	**-0.0384**	**-12.234**	**-0.0639**	**-20.130**
Distance to highway exits (km)	**-0.0095**	**-3.079**	**0.0168**	**5.513**	**0.0115**	**3.766**	0.0096	3.127
Distance to CBD (km)	**0.0650**	**21.814**	**0.0654**	**21.968**	**0.0303**	**10.007**	0.0086	2.809
Job accessibility (k)[a]	**0.3570**	**1.162**	**0.3580**	**1.166**	**0.1872**	**0.609**	**0.1651**	**0.537**
Distance to non-work destinations (km)	**0.1276**	**44.232**	**-0.0387**	**-12.337**	0.0039	1.284	**-0.0862**	**-26.855**

a Coefficient is x 10^{-2}.

* Boldface denotes coefficients significant at the 0.1 level in the corresponding model.

Source: Calculated by the author.

Based on the estimation results of the Heckman-selection model with spatial error, households in the City of Boston would like to pay an additional 10.0 thousand dollars (or 3.1% of property values) for living within walking distance to subway stations, 20.1 thousand dollars (or 6.2% of property value) for every kilometer closer to MBTA park-and-ride lots, 26.8 thousand dollars (or 8.3% of property value) for every kilometer closer to non-work destinations, and 0.5 thousand dollars (or 0.2% of property value) for one thousand additional spatially-weighted job opportunities, for a property originally priced at 325.0 thousand dollars (the mean transaction price). The WTP estimates for the same built-environment attribute differ across the four models significantly, which suggests that selectivity and spatial autocorrelation have a significant impact in valuing the built environment. For example, based on the estimation results of the conventional hedonic-price model, the WTP for proximity to subway station is 19.1 thousand dollars for a property valued at the mean transaction price. However, the amount decreases to 10.0 thousand dollars, when I control for selectivity and spatial error type autocorrelation. The related bias is about 91.0%. A bias of such magnitude could misinform relevant policy designs, such as land value capture schemes to fund public transportation or transit-oriented development.

5.4 CONCLUSIONS

In this paper, I explore the role that selectivity and spatial autocorrelation could play in valuing the built environment. Using the transaction and stock data for single-family properties in the City of Boston from 1998 to 2007, I apply the Heckman two-step procedure and spatial econometrics techniques to account for sample selection and spatial autocorrelation respectively. I calibrate the following four models: (1) a conventional hedonic-price model, (2) a classical Heckman-selection model, (3) a Heckman-selection model with spatially lagged dependent variables, and (4) a Heckman-selection model with a spatially lagged error term. Based on the estimation results, I calculate the WTP for built-environment attributes.

The empirical analysis suggests that the sample of sold properties is a biased sample of the housing stock. Simply estimating a hedonic-price model using the sold sample generates biased estimates of the WTP for the housing stock. My results confirm the significant impacts of the built environment on both the probability of housing sale and transaction price. Higher density, better job accessibility, proximity to non-work destinations, and distance from CBD could increase the probability that a house is sold in the market. Spatial autocorrelation indeed exist in the empirical analysis. The Heckman-selection model with spatial error has the highest explanatory power among the four models. The estimation results of this model reveal that households in Boston pay a premium for living within walking distance to subway stations, closer to MBTA park-and-ride lots and non-work destinations, and proximity to job opportunities. Meanwhile, there are significant variations in the WTP estimates across the four models, which suggest that selectivity and spatial autocorrelation could lead to significant bias in valuing the built environment.

It should be noted that as the core part of the metro region, City of Boston exhibits a much smaller variation of built-environment characteristics compared to Metro Boston. This limitation may diminish the built-environment effects on both the probability of sale and transaction price and limit the generality of the results. Ideally, I hope to calibrate the same set of models for Metro Boston. However, I can only get all necessary data for the City of Boston, thus have to limit the study area to City of Boston.

Nonetheless, the findings of this study have important policy implications in metropolitan planning. Biased estimates of the WTP for the built environment due to sample selection and spatial autocorrelation might misguide policy recommendations for intervening urban-development patterns and distort estimations of the value-added effect of infrastructure investment for land value capture programs.

Smart-growth strategies often face the discrepancy between the regional and local interests in implementation. The region can benefit from smart-growth policies due to the reduction of transportation emissions, while local residents have to care about the impact of smart-growth policies on their own neighborhood. A fair estimate of the property-value effect of certain land-use-control policies could help assess the local effect of smart growth, reconcile regional and local benefits, and facilitate dialogues between regional planning agency, local government, and the public regarding alternative metropolitan growth scenarios. This study shows that in a dense urban area like Boston, properties values are positively associated with some smart-growth features such as transit accessibility, proximity to non-work and work destinations, after selectivity and spatial autocorrelation are accounted for. This may suggest that such smart-growth features can improve the quality of life and increase the property values in the local neighborhoods.

Smart growth encourages travelers to switch from auto to transit. However, transit agencies are facing significant financial challenges worldwide. Meanwhile, property owners and developers are benefiting from increased property values generated by transportation improvements as suggested by many previous studies including this essay. Such benefits create a rationale for the use of value capture policies such as land value taxes and tax increment financing to capture some of the value-added effect of transportation investment to relieve the financial burdens of transit agencies. One barrier in land use capture is the assessment of land value increment. This study demonstrates that conventional hedonic price analysis may bring significant bias in valuing the value-added effects of transit by omitting the selectivity and spatial autocorrelation issues. The methodology applied in this study could help governments and transit agencies to make informed decisions in designing land value capture programs.

CHAPTER SIX: CONCLUSIONS AND IMPLICATIONS

The world is undergoing a rapid urbanizing process. The UN (2001) projects that by 2030 an additional 2 billion people will be added to the world's urban areas. In the face of this urban growth, on the one hand, we need to accommodate the increasing travel and land needs for economic development and human welfare. On the other hand, we need to mitigate the associated negative effects, for example, congestions, emissions, and exhaustion of non-renewable resources, to make the metropolitan growth sustainable. The U.S. 2000 Census data and the vehicle safety inspection records from the Registry of Motor Vehicles used in this study draw a clear picture of the transportation emissions produced in the Boston Metropolitan Area. In 2000, 4.31 million individuals and 1.64 million households are living in the 164 municipalities of Metro Boston. They own 2.47 million private passenger vehicles[10]. On average, each vehicle drives 33.2 miles everyday, which adds up to 82.0 million miles per day, and 29.9 billion miles per year in the Metro. If we assume that the average fuel-efficiency of passenger vehicles is 22.1 miles per gallon[11] and a gallon of gasoline produces 8.8 kilograms of CO_2[12], then 1.35 billion gallons of gasoline are consumed and 11.9 million tons of CO_2 are generated annually. In Massachusetts, the transportation sector alone currently accounts for 36% of the overall carbon emissions[13], and this proportion is projected to continue increasing in the next decade[14].

The major focus of this study is a seemingly straightforward question: could the built environment play a role in reducing transportation emissions and achieving sustainable

---

[10] Based on vehicle safety-inspection records from 2005-2007.

[11] According to Research and Innovative Technology Administration, Bureau of Transportation Statistics, the Average U.S. passenger car fuel efficiency is 22.1 miles per gallon in 2005.

[12] Source: Greenhouse Gas Emission for a Typical Passenger Vehicle, U.S. Environmental Protection Agency (EPA) report EPA420-F-05-004.

[13] Source: Massachusetts Department of Transportation from U.S. Energy Information Administration.

[14] Source: Statewide Greenhouse Gas Emissions Levels: 1990 Baseline and 2020 Business as Usual Projections, MA DEP July 1st, 2009.

metropolitan growth? And if so, what role? To answer this research question, I structured the dissertation in three separate essays, focusing on two aspects of the land use-transportation interconnection, respectively: the impact of the built environment on travel behavior and the impact of the built environment on development patterns. This study benefits from several new-available administrative datasets with detailed location information and broad coverage: (1) the vehicle safety-inspection records for all the private vehicles registered in Metro Boston (about 2.47 million vehicles in total) from the Registry of Motor Vehicles; (2) the housing transaction records for all single-family housing transactions in Metro Boston during 2004-2006 (about 93 thousand transactions in total) from city and town assessors provided by the Warren Group; (3) the housing transaction records for all single-family housing transactions in the City of Boston during 1998-2007 provided by the Suffolk County Registry of Deeds; and (4) the assessing records for all single-family properties in the City of Boston from the Assessing Department of Boston. The study confirms the important role that the built environment can play in sustainable metropolitan growth. It demonstrates that a large portion of the variation in household vehicle miles traveled (VMT) can be explained by the variation in the built environment. Although the study is cross-sectional, the results suggest that smart growth could significantly reduce VMT by altering the built environment that requires people to drive. The variation in the built environment does appear to be capitalized into property values. Smart-growth-type built-environment features such as accessibility, connectivity, and walkability are positively associated with property values. The value-added effects of these smart-growth features provide a potential financing mechanism for governments and agencies to support environmental-friendly transportation modes and development patterns via land value capture. However, selectivity and spatial autocorrelation need to be accounted for when valuing land value increments.

6.1 SUMMARY OF EMPIRICAL FINDINGS

Boston is one of the few metropolises in the United States that offer a rich variety of built-environment characteristics and travel choices. The public transportation network and biking- and waking- friendly environment are supported by relatively dense and mixed land-use pattern in the urban center and sub-centers. The majority of the population and geography, however, is still auto-oriented. The diversity in the built environment and travel behavior make Boston a compelling case for the empirical analysis. The major findings are summarized below.

The first essay of my dissertation focuses on the relationship between the built environment and household vehicle usage. The empirical results reveal that both the built-environment and demographic factors are significantly associated with household vehicle miles traveled (VMT). On the demographic side, I find that wealthier neighborhood tend to have fewer VMT per vehicle, but considerably more VMT per household, suggesting that households in wealthier neighborhoods tends to own more cars and drive more total miles but use each car somewhat less. The built-environment factors have significantly higher impacts on VMT than do demographic factors. In particular, improving accessibility to work and non-work destinations, connectivity, and transit accessibility can significantly reduce VMT. In Metro Boston, one standard deviation increase in the "distance to non-work destinations" factor is associated with an increase in annual VMT per household of 3,306 miles; one standard deviation increase in the "connectivity" factor is associated with a decrease in annual VMT per household of 3,481 mile; and one standard deviation increase in the "inaccessibility to transit and jobs" factor is associated with an increase in annual VMT per household of 5,745 miles; However, one standard deviation increase in the "wealth" factor is associated with an increase in annual VMT per household of 482 miles[15].

The empirical results of the second essay suggest that built-environment characteristics can be capitalized into property values. The transaction price of single-family properties in Metro Boston is positively associated with the "connectivity" and "walkability" factors, and negatively related to the "inaccessibility to transit and jobs" and "auto dominance" factors. Based on the estimation results, for a single-family property originally priced at 376.5 thousand dollars (the median transaction price), one standard deviation increase in the "connectivity" factor and "walkability" factor could increase the transaction price by 8.39 thousand dollars (2.2% of property value) and 5.34 thousand dollars (1.4% of property value), respectively; one standard deviation increase in "inaccessibility to transit and jobs" and "auto dominance" could decrease the transaction price by 30.65 thousand dollars (8.1% of property value) and 2.56 thousand dollars (0.7% of property value), respectively[16]. These results represent the average built-environment effects across the region. The analysis also demonstrates the existence of submarkets for built-environment characteristics in Metro Boston. Households living close to transit stations pay higher premiums for smart-growth-type built-environment features than households living beyond walking distance to transit stations. The different premiums for the built environment between the two submarkets may be partly attributed to life style preference. Transit-oriented households may purposely choose to live in transit-friendly neighborhoods, thus would like to pay higher premium for built-environment features that favor transit. The

[15] One unit increase of the "distance to non-work destination", "connectivity", "inaccessibility to transit and jobs", and "wealth" factor is associated with 3,821, -2,970, 5,906, and 738 miles increase in annual VMT per household respectively.

[16] One unit increase of the "connectivity", "inaccessibility to transit and jobs", "auto dominance" and "walkability" factor is associated with 6.13, -30.25, -4.44, and 5.50 thousand dollars increase in property values for a single-family property valued at 376.5 thousand dollars.

coexistence of spatial-error-type autocorrelation and submarkets may suggest that some omitted variables, such as life style preference, are correlated at different spatial scales. These omitted variables may help explain the formation of submarkets and the variation in empirical measures reported in the literature.

In Essay 3, I investigate the impacts of selectivity and spatial autocorrelation in the valuation of the built environment. The empirical results suggest that the built environment has significant impacts on the probability of housing sales. Single-family properties in denser areas, with better job accessibility, closer to non-work destinations but farther away from the CBD, are more likely to be sold in the market. The modeling results indicate that the sample of sold properties is a biased sample of the housing stock and spatial autocorrelation indeed exists in the housing transactions. Therefore, when analysts apply conventional hedonic price analysis to the sample of sold properties to value built-environment features, they will get biased estimates. After correcting for sample selection bias and spatial-error-type autocorrelation, I find that households pay 10.0 thousand dollars (3.1% of property value) for living within walking distance to subway stations, 20.1 thousand dollars for every kilometer closer to MBTA transfer lots (6.2% of property value), 0.5 thousand dollars (0.2% of property value)for every one thousand additional spatially-weighted job opportunities, and 26.8 thousand dollars (8.3% of property value) for every kilometer closer to non-work destinations for a property valued at 325.0 thousand dollars (the median price of the sold sample). The magnitude of the biases due to selectivity and spatial autocorrelation could be big. For example, the WTP for proximity to subway stations computed based on the conventional hedonic-price model is about 91% higher than the one computed using Heckman-selection model with spatial error correction.

6.2 POLICY IMPLICATIONS

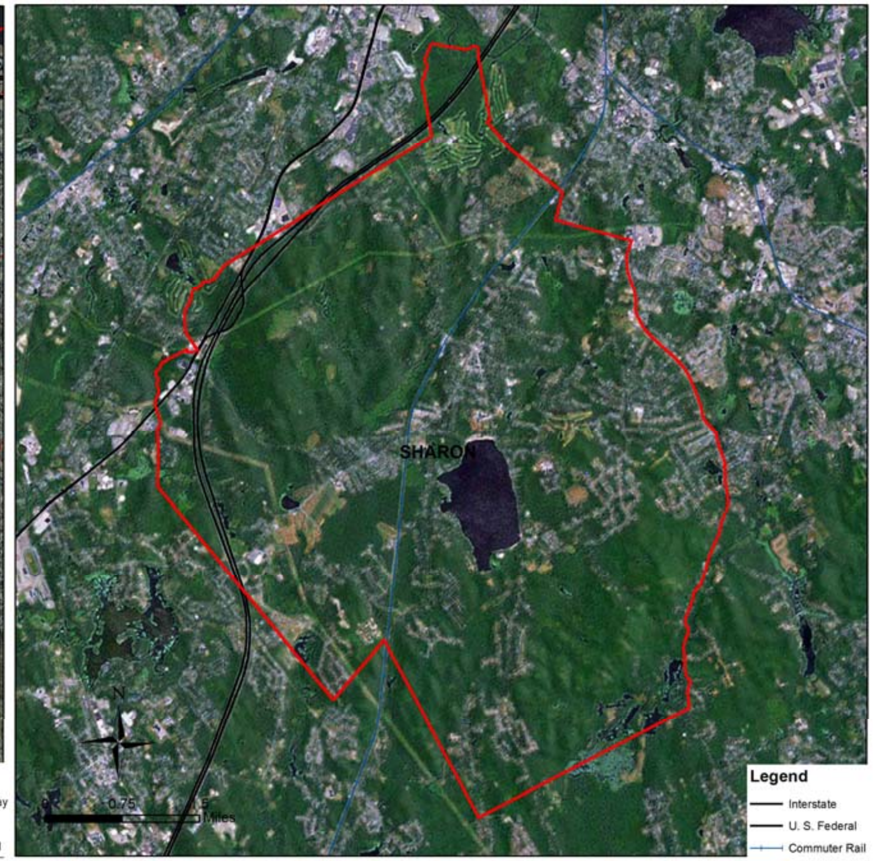
There has been a long-time debate about the policies to reduce auto-dependence and associated transportation GHG emissions.

In the short- to medium-term, technology alone will most likely not provide an easy answer. Heywood et al. (2003) conclude that based on the plausible vehicle technological improvements, both technology and demand management options will be required to reduce the U.S. private passenger vehicle annual fuel consumption over the next 20 years to levels below 500 billion liter per year in 2003. To reduce travel demand, economists often argue that proper pricing -- such as congestion tolls, fuel taxes, and parking surcharges -- would eliminate the need for smart growth and associated land-use-control policies. With substantially higher road price, people would move closer to jobs and switch to transit to economize on travel. However, road pricing remains something theoretically meaningful but practically difficult due to the enormous political barriers. By far only a few cities such as Singapore and London have implemented congestion pricing in practice. In the absence of true market-based pricing of transportation, smart growth and land use planning becomes a second-best response to transportation energy use and emissions.

This Boston-based study indicates that smart growth has the potential to significantly reduce VMT and associated transportation energy use and emissions, especially those policies that focus on increasing accessibility to destinations, creating traditional-type, high-density, well-connected neighborhoods, and improving transit accessibility. Figures 14 shows orthophotos of two towns in Metro Boston, Brookline and Sharon. Brookline is a town near urban core and Sharon is in the suburban area between the first and second ring roads. Figure 15 depict the different street network patterns of Brookline and Sharon at similar scales. Brookline (especially the dense northern half) has a traditional high-density, small-block, grid-type neighborhood design, while Sharon has relatively lower density and more cul-de-sacs and non-grid road network than Brookline. The average "connectivity" score is 2.17 for Brookline, and -0.23 for Sharon. The difference between them is about 2.04 standard deviations. Based on the modeling results, one standard deviation increase in the "connectivity" factor is equivalent to 3,481 miles decrease in annual VMT per household. Therefore, increasing the connectivity of Sharon to the level of Brookline could save about 7,098 miles in annual VMT per household, assuming other factors are the same. In fact, the actual annual VMT per household in Brookline and Sharon are 7,818 miles and 24,499 miles respectively as differences in other factors expand the difference in annual mileage between the two towns. The total saving in annual VMT amounts to 98.9 million miles if the 5,934 households living in Sharon had the VMT pattern of those in Brookline, which is equivalent to 4.48 million gallons of gasoline and 39.5 thousand tons of CO_2 emissions. It should be noted that this is only a very simplified computation -- a precise estimation of the CO_2 savings of curtain smart-growth project needs to deal with much more complex issues such as residential self-selection and necessitates a more complicated model structure, as suggested by Zegras et al. (2008). Nonetheless, this detailed analysis of actual VMT patterns provides some evidence of the potential effectiveness of smart growth in reducing vehicle usage and transportation emissions.

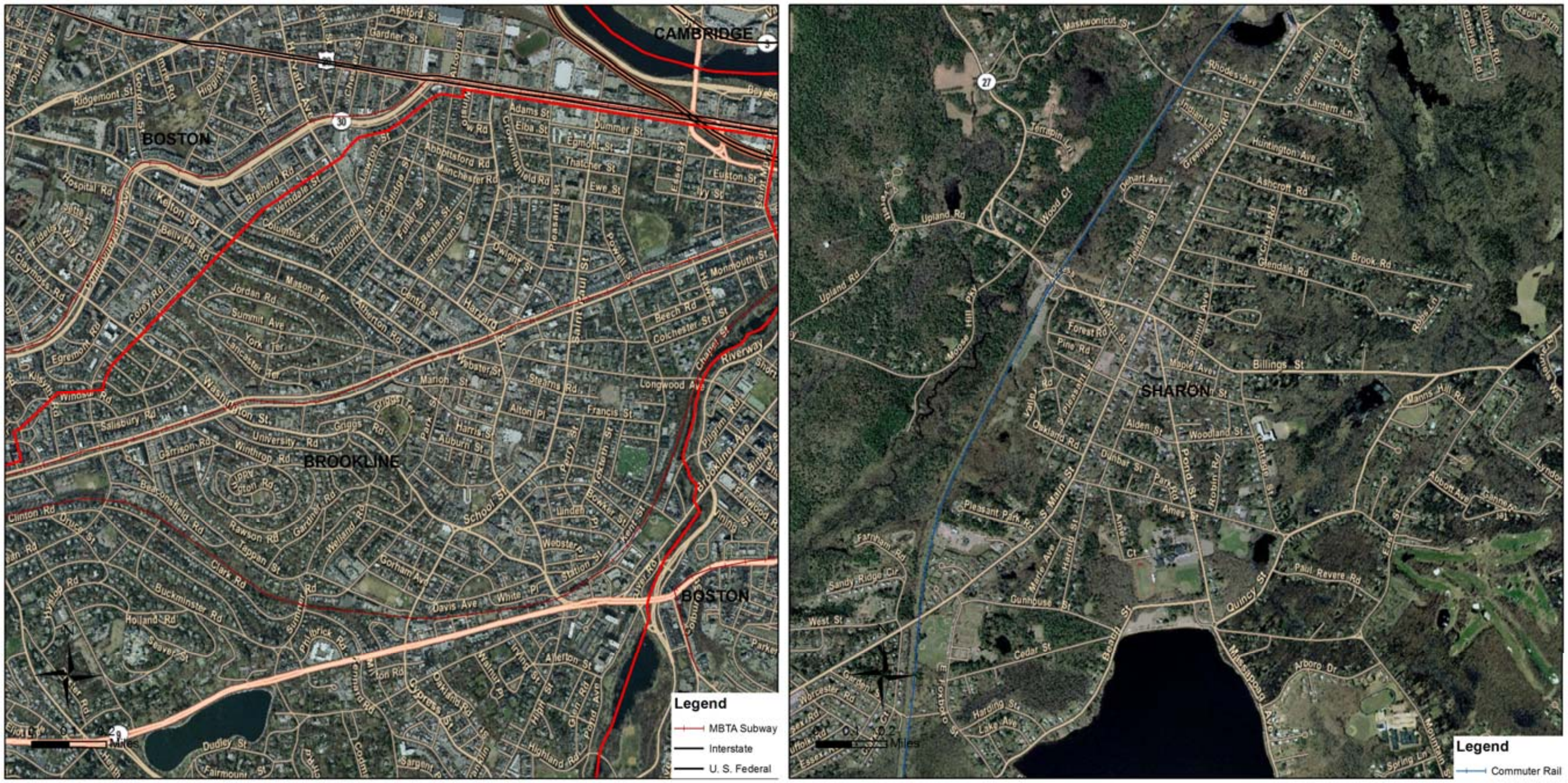
Town of Brookline

Town of Sharon



Source: The author.

Figure 14: Orthophotos of Brookline and Sharon



Source: The author.

Figure 15: Street Network Layout of Brookline and Sharon

The environmental benefit of smart growth is mostly felt at the regional levels. What about the impact of smart growth at the local level? What will local residents sacrifice for public gains? Until the benefits and costs of land-use-control policies on the neighborhoods are weighed fully, local residents may remain skeptical of smart growth. This study provides evidence that smart growth actually generate benefit to local neighborhoods. Properly-designed smart-growth programs plan for all development needs, such as access to public transportation and jobs, proximity to activity centers, and walkable neighborhoods. The empirical analysis indicates that smart-growth features such as connectivity, accessibility, and walkability are actually positively associated with residential property values. Although finding association is not equal to constructing causality, it still provides some support for the argument that by providing various amenities, smart growth could increase the desirability of the community, thus the property values (Nelson et al. 2002).

The built-environment effects on property values are not distributed evenly over space. Although households living in properties with good transit accessibility pay higher premiums for smart-growth-type built-environment characteristics than other households, most smart-growth features are still positively associated with properties values for both groups of households. The existence of submarkets for the built environment may suggest that the built-environment effect varies over space. In this case, calibrating a global model for the entire study area cannot capture the spatial variation of the relationships between the built environment and property values. Other modeling techniques such as geographically-weighted regression may help characterize this spatial effect.

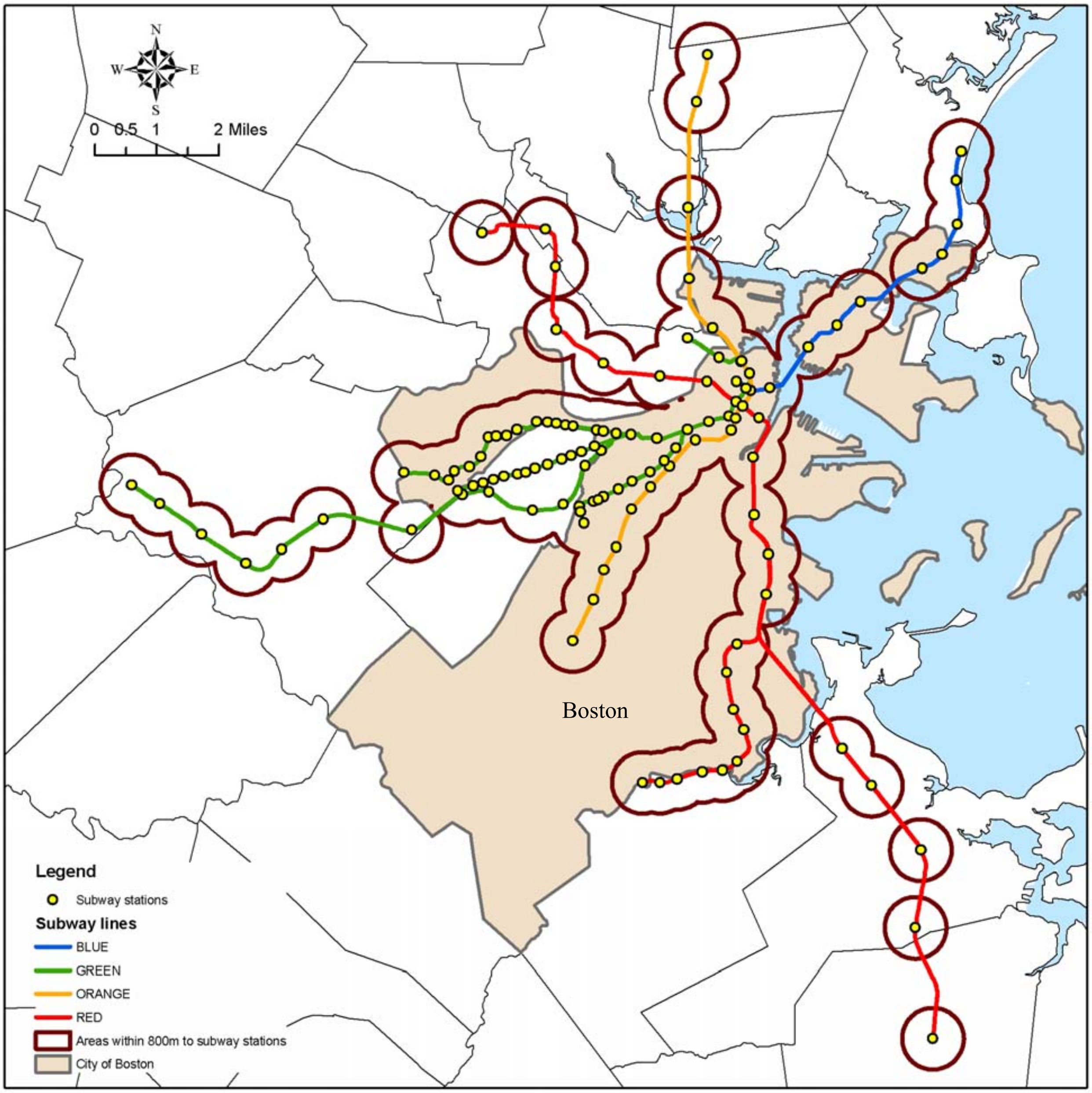
Smart growth needs coordinated land use and transportation planning. One impediment for the effective coordination of land use and transportation planning is the mismatch between where decisions on land development are made – locally – and the transportation impacts are felt – regionally. Local municipalities have their own concerns in making development decisions. For example, a more compact development pattern often means higher population density and more households, which in turn will bring more children to local schools and incur higher education spending. Smart-growth-type development will not necessary be implemented automatically at the local level just because it is valued positively by homeowners.

This study investigates the impact of the built environment at both the regional and local level. Regional planners could leverage such research findings to showcase the effectiveness of smart-growth strategies in reducing GHG emission, illustrate the potential improvement in the quality of life of the community, and facilitate the dialogue among regional planning agencies, local government and the public regarding alternative regional development scenarios. With a better understanding of the environmental benefits and the local amenities brought about by smart growth, local government might be more likely to give up some local interests for public gains, or at least agree to incentives or taxes to price the externality.

Smart growth encourages travelers to switch from cars to transit. However, a growing number of transit agencies around the world are facing increasing financial difficulties. For example, from 2004-2007, the MBTA (the transit authority in Metro Boston) has almost doubled the transit fares in order to cover a large part of its operating deficit. In the most recent proposal in 2009, the MBTA proposed to increase fares by 19.5 percent, which could raise 69 million dollars per year for the authority. The fare hikes could adversely influence the market share of transit. To ensure adequate and sustainable transportation investment for current and future needs, policy makers need to reassess the current mechanisms of transportation finance in the United States and explore alternative revenue sources. As a result, the feasibility of funding public transport systems through land value capture programs to recover part of the value-added effect of transit has become a keen concern of many researchers and policy makers.

Estimation of the land use increment is essential to effectively mobilizing land value capture programs in the public transit case. This study proposes a new estimation method to address two important methodological issues in the estimation: selectivity and spatial autocorrelation. Both issues could produce biased estimates in valuing the built environment. The study confirms the value-added effect of transit after correcting for sample selection and spatial autocorrelation, which provides a basis for value capture initiatives.

Although giving specific point estimates is not the major focus of this study, it is still of interest to do a “quick and dirty” computation to show the rough magnitude of the value-added effect of subway and the amount of value that could be captured. In a simplest hypothetical scenario, it is assumed that the value-added effect of subway is constrained to properties within walking distance (800m) to subway stations, and that property tax from these properties that is attributable to the proximity to subway stations will be earmarked to support the transit system. Figure 16 plots the locations of all MBTA subway stations in Metro Boston and their impact zone. Table 19 shows the computation results based on these admittedly strong assumptions.



Source: The author.

Figure 16: MBTA Subway Stations and Their Impact Zone

Table 19: Value-Added Effect of Subway Stations (Unit: Million Dollars)

					Hedonic-price model		Heckman-selection model		Heckman-selection model + Spatial Lag		Heckman-selection model + Spatial Error	
Property Type	Total Value in Boston	Property Tax in Boston	Total Value within Impact Zone	Property Tax within Impact Zone	Value added of Subway Station	Property Tax Attr. to Subway	Value added of Subway Station	Property Tax Attr. to Subway	Value added of Subway Station	Property Tax Attr. to Subway	Value added of Subway Station	Property Tax Attr. to Subway
1-Family	10472.4	112.4	3574.8	38.4	209.8	2.3	369.2	4.0	197.9	2.2	109.9	1.2
2-Family*	7092.3	76.1	2918.8	31.3	171.3	1.9	301.4	3.3	161.6	1.8	89.7	1.0
3-Family*	6440.4	69.1	3584.1	38.5	210.4	2.3	370.1	4.1	198.4	2.2	110.1	1.2
Condo.*	15113.3	162.2	12502.4	134.2	733.8	8.0	1291.1	14.2	692.2	7.6	384.2	4.2
Total	39118.4	419.7	22580.1	242.3	1325.3	14.5	2331.8	25.6	1250.1	13.7	693.9	7.6

* Numbers are computed using estimated coefficients of the single-family properties model.

Source: Calculated by the author.

The first row of Table 20 presents the computation results for single-family properties in the City of Boston using the modeling results of Essay 3. The total assessed value for all single-family properties in the City of Boston is 10.5 billion dollars, which generate annual property tax of 112.4 million dollars based on the tax rate of 2005. The aggregate assessed value for single-family properties within the impact zone is 3.6 billion dollars, or 34.1% of the total assessed value in the City. After the sample selection and spatial-error-type autocorrelation correction, the value-added effect of subway stations for single-family properties in the city is 109.9 million dollars. The corresponding annual property tax is 1.2 million. Single-family properties are only a proportion of the housing stock. Table 20 also presents the results for two-family, three-family and condo properties assuming that households living in these types of properties have the same WTP for subway accessibility as households living in single-family properties. The total value-added of subway stations is 693.9 million dollars, or 1.8 percent of the overall assessed value 39.1 billion. The corresponding annual tax revenue is 7.6 million dollars, which could be captured according to the hypothetical scenario to fund new transit facilities as well as transit-oriented development. The 7.6 million revenue is small compared to the $430 million revenue from transit fares in the 2008 budget of the MBTA, but similar in magnitude to the revenues from advertising (11.0 million) and Federal Government (8.0 million). It should also be noted other property types like multi-family apartments and commercial properties are not included in this calculation and subway stations outside Boston are also neglected.

As shown in Table 20, the estimates of the value-added of subway stations vary significantly across models. On the one hand, it shows the importance of correcting sample selection and spatial autocorrelation in the estimation. On the other hand, it also reminds policy makers to stay cautious when designing land value capture schemes.

In summary, the research findings of this dissertation suggest that: (1) the built-environment features advocated by smart growth could benefit the region as reflected by the significant reduction in vehicle usage and associated GHG emissions; (2) smart-growth-type built- environment features could improve the quality of life in local neighborhoods as reflected by the increase in property values; and (3) selectivity and spatial autocorrelation need to be corrected in valuing the built environment, if governments or agencies plan to apply value capture schemes to support environmental-friendly transport modes and resource-efficient land development patterns.

6.3 RESEARCH CONTRIBUTIONS

The study has made a number of contributions to the geography, transportation, and planning fields.

6.3.1 Spatial Unit of Analysis and the MAUP

This study enriches the built-environment literature by conducting a comprehensive and spatially-detailed analysis on the relationships among the built environment, place of residence and vehicle usage.

One significant challenge in spatial analyses is the well-known Modifiable Area Unit Problem (MAUP). The MAUP has two aspects, scale and zonal definition, which can lead to inconsistency in quantitative and statistical analyses. The scale effect refers to the inconsistency due to the change from one aggregation level to another (e.g., from block group to census tract). The zonal effect refers to the inconsistency due to the multiple ways in which areal units can be defined. Using disaggregate data and grid-cell type spatial unit are identified as one possible method to mitigate the MAUP.

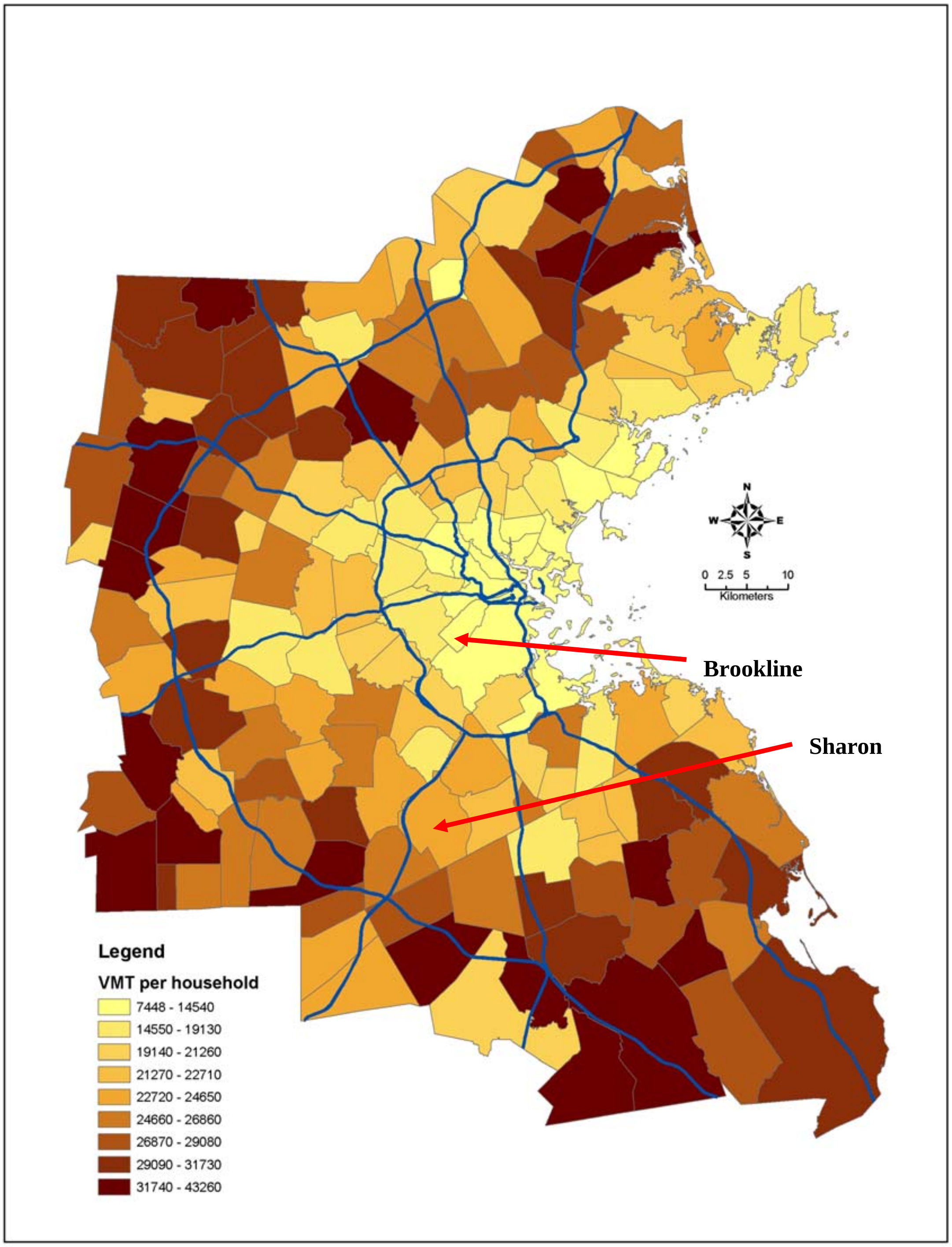
Table 20 summarizes the spatial units of analysis in several recent land use-transportation studies. Despite the MAUP effects, the TAZ or similar census geography remains a very common base unit for measuring the built environment in the relevant analyses. For example, Newman and Kenworthy (1999) use city-level data to analyze the relationship between density and energy use. Holzclaw et al. (2002) investigate the impact of neighborhood urban design and socioeconomic characteristics on car ownership and vehicle usage at the zip code level. At such aggregated levels, the intra-zone variations of built-environment, vehicle usage, and demographic measures could be too large to ignore. To deal with the MAUP, this study takes advantage of several spatially-detailed datasets and advanced GIS techniques and carries out the empirical analysis at fine-grained 250x250m grid cell level.

Table 20: Spatial Units of Analysis in Several Recent Studies

Study	Purpose	Spatial Unit of Analysis
Bhat and Guo (2007)	BE on car ownership	TAZ
Boarnet and Sarmiento (1998)	BE on VMT	Block group and zip code zone
Brownstone and Golob (2008)	BE on VMT and fuel use	Block group
Cervero (2002)	BE on mode choice	TAZ
Cervero and Kockelman (1997)	BE on travel demand	Census tract; 1 Hectare grid
Crane and Crepeau (1998)	BE on travel demand	1/2 mile buffer around HH for network; census tract for land uses
Greenwald (2003)	BE on non-work mode substitution	TAZ
Greenwald and Boarnet (2001)	BE on walk	TAZ, block group, HH buffers (1/4 -1mi)
Hess and Ong (2002)	Neighborhood on auto ownership	TAZ, census tract
Holzclaw et al. (2002)	BE on car ownership and vehicle usage	Zip code zone
IBI Group (2000)	Average HH Transport GHG emissions per TAZ	TAZ, in some cases TAZ centroid radii
Newman and Kenworthy (1999)	BE on energy use	Town
Rajamani et al. (2003)	BE on mode choice	Census block group boundary
Rodriguez and Joo (2004)	BE on mode choice	Block group for density; corridor measures for path, slope, sidewalk
Srinivasan (2001)	BE on travel demand	TAZ
Zhang (2004)	BE on mode choice	TAZ, 800m grid cell

Figure 17 shows VMT per household aggregated at the municipality level, using quantile classification method and nine categories. The spatial pattern is what analysts would expect, municipalities in the urban center have much lower VMT measures than municipalities in the suburban area. Although this municipality level map captures some interesting spatial patterns, it overlooks subtle phenomena exhibited at more disaggregate scale. In Figure 17, Brookline is a town close to the urban core and Sharon is a suburban town with higher VMT per household than Brookline. Figure 18 compares VMT per household at 250x250m grid cell level for Brookline and Sharon. Analysts can observe significant intra-town variations in both towns. The range of grid-cell level VMT per household is 2,986-37,154 miles for Brookline and 5,270-67,595 miles for Sharon. Although the town average of VMT per household in Sharon is much higher than that of Brookline, some communities in Brookline behave just like a suburban neighborhood, and households in part of Sharon drive even less than households in an average Brookline grid cell. These interesting spatial patterns diminish from Figure 17 due to data aggregation. The intra-zone variation is more severe in suburbs than in inner city because of the difference in the size of zones. In the inner city, a census tract may only contain a few city blocks, whereas in the suburb it is nor rare that an entire town is a single census tract.

What is the underlying factor on which VMT per household depend? The intra-town variation in VMT per household may stimulate further interest of analysts. The built-environment characteristics of Brookline and Sharon as shown in Figure 18, suggests that (1) proximity to subway stations and well-connected road network may have significant impacts on VMT in Brookline; (2) grid cells close to the commuter rail station may drive less than other grid cells in Sharon; and (3) detecting meaningful VMT difference requires disaggregate data at or near the 250x250m grid cell scale.

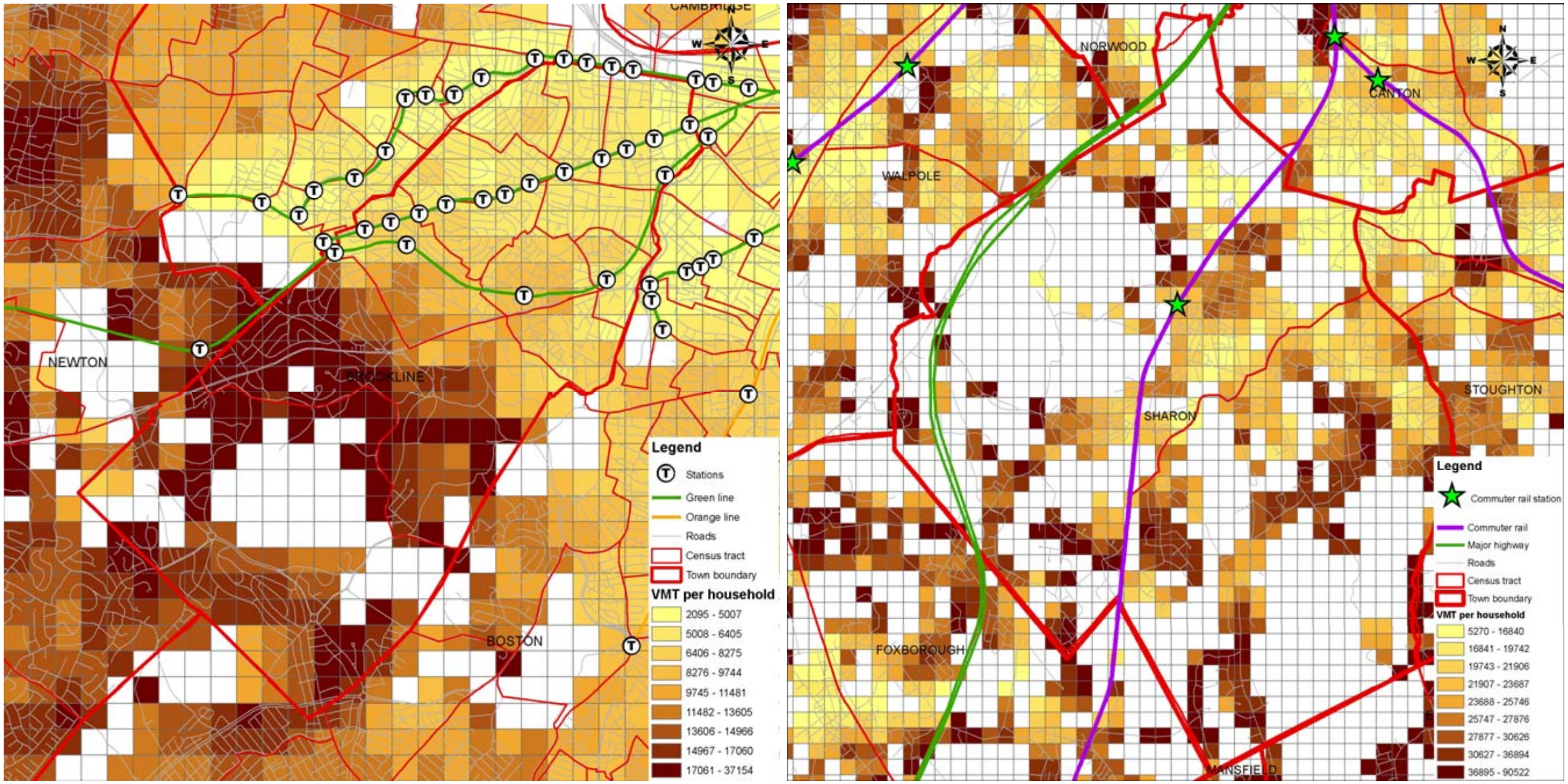


Source: The author.

Figure 17: VMT per Household at the Municipal Level

Town of Brookline

Town of Sharon



Source: The author.

Figure 18: Grid-Cell Level VMT per Household in Brookline and Sharon

The interaction between human behavior and the physical environment likely involves different processes at different spatial scales. Behavioral consideration and justification for a specific variable are important in selecting a specific scale and areal unit definition. For example, the impact of property tax rate is felt at the town level; the impact of school quality is constrained to the school district level; and Figure 18 suggests that assessing the built-environment effect on VMT needs to be carried out at much more fine-grained scale than analysts previously did, because average zonal travel and built-environment attributes may not necessarily reflect the characteristics of the specific locations where individual trip-making takes place.

6.3.2 Relative Effects of Built-Environment and Demographic Factors

The study provides new evidence of the relative effects of built-environment and demographic factors on vehicle usage. In this study, I find that the impact of the built-environment factors on VMT is significantly higher than that of demographic factors, contrary to the findings of many household-survey-based studies. These studies tend to find demographic characteristics and attitudinal factors explain a significant proportion of VMT variation, and the built-environment effects are minimal. To some extent, the different data aggregation schemes employed in these two types of studies might contribute to the different results. Data aggregation and associated MAUP could bring significant biases in statistical analyses. Previous studies usually use household-level demographic variables and aggregate built-environment variables at zip code or TAZ level, which is opposite to my study using aggregate (block-group-level) demographic factors and disaggregate (250x250m grid-cell-level) built-environment factors. For example, using travel diary data for 769 California households, Boarnet and Sarmiento (1998) found no stable link between density (computed at block group and zip code level) and VMT after using instrumental variables to control for the endogeneity of density. Using survey data for 2,954 households in San Francisco, Bhat and Guo (2007) find statistically significant but quantitatively small impact of built-environment measures (computed at the TAZ level) in vehicle ownership, while demographic and housing tenure variables have strong effects. Brownstone and Golob (2008) build a simultaneous equations model of residential density (computed at block group level), VMT and vehicle fuel use using the 2001 National Highway Transportation Survey, and find that the magnitude of the density effect is very small. My study suggests that the built-environment effects may be biased downward in previous studies because they use aggregate built-environment measures.

6.3.3 Transportation and Land Value Capture

This study also contributes to the existing literature of transportation financing by proposing a new analytical approach to evaluate the impact of transportation on property values. Assessing the property-value effect of transportation improvement is a prerequisite to design value-capture programs. The dominant method in valuing housing attributes is the hedonic price analysis. Table 21 lists some hedonic studies of the price effect of good transit access in North America. It shows that proximity to transit stations could increase property values by a wide range (4%-45%). The enormous variation in the magnitude of the impact could be attributed to either type of transit, other location characteristics, definition of proximity, model specification, or a combination of all these factors. However, none of these studies considers sample selection issue in the estimation.

Table 21: Property-Value Impacts of Transit Proximity in North American Cities

Case/Location	Impact on	Impact	Source
Boston Commuter Rail	Housing price	+6.7%	Armstrong 1994
Buffalo Light Rail	Housing price	+4-11%	Hess and Almeida 2007
Miami Metrorail	Housing price	+5%	Gatzlaff and Smith 1993
Portland Metro Express	Housing price	+10.5%	Al-Mosaind et al. 1993, Chen et al. 1998
San Francisco Bay Area BART	Residential rent	+10-15%	Cervero 1996
Santa Clara County Light Rail	Residential rent	+15%	Weinberger 2001
Santa Clara County Light Rail	Housing price	+45%	Cervero and Duncan 2002
St. Louis Metrolink	Housing price	+32%	Garrett 2004
Toronto Metro Subway	Housing price	+20%	Bajic 1983

Using data from the City of Boston, this study demonstrates that the widely-used hedonic price analysis calibrated with a sample of sold properties could lead to significant bias in valuing the built environment if sample selection issue is not corrected for. In this study, I apply Heckman 2-step procedure to correct for sample selection bias and integrate spatial econometric techniques into Heckman-selection models to resolve spatial autocorrelation. The proposed analytical approach, combining a Heckman procedure with spatial econometric techniques, could produce unbiased estimates of the WTP for built-environment characteristics. After the corrections, the value-added attributable to proximity to subway stations is 3.1% of property values in the City of Boston, compared with 5.9% of property values without the correction.

6.3.4 Administrative Data for Urban Modeling

Previous studies on land use and transportation primarily rely on household survey data. In this study, I demonstrate the benefits as well as difficulties in utilizing administrative data for urban modeling. With the rapid development of spatial database infrastructure in the last decade, the amount of available administrative data with spatial information has increased dramatically. For example, GIS data layers are often available on road networks, parcels, and building footprints, and transaction information like vehicle safety inspections records, assessing records, housing transaction records, and utility records.

This study shows at very low marginal cost, the administrative data can produce very rich information to support metropolitan planning. The administrative data, such as the vehicle safety inspection records and assessing records are routinely collected by corresponding agencies. The datasets are theoretically available to analysts at no cost, compared to the hundreds of dollars expense per observation in common surveys.

The administrative data have exceptionally broad temporal and spatial coverage. They usually cover the entire population of the subject of interest and are regularly updated. Both datasets are updated annually. Such pervasive administrative datasets enable analysts to compute reliable and comparable measures to better inform policy making. On the contrary, surveys usually have only a few thousand observations and are updated every 5-10 years.

These advantages together with other benefits such as accuracy, automatic collection and central storage make administrative data a compelling data source for urban modeling. However, inherent disadvantages of such administrative data also impose significant challenges in the exploitation of these datasets.

First, administrative data are usually not primarily designed for modeling purposes, so some critical information may be lacking, and the datasets are often not in an easy-to-use format, which restricts the usefulness of the raw data without intensive processing and careful interpretation. For example, both the vehicle safety inspection records and housing transaction records lack household-level demographic characteristics, which are indispensable to calibrate activity-based models to explore the underlying behavior mechanism of household choices of vehicle usage and residential locations.

Second, the administrative data are usually collected and maintained by different agencies in different formats with different spatial and temporal coverages, which makes cross-referencing among datasets a hard task and seriously limits the utilization of these datasets. In Essay 3, the housing transaction records from the Suffolk County Registry of Deeds use street address of properties as the only location identifier. The assessing records from the Assessing Department use parcel ID as the location identifier. Advanced GIS and DBMS tools are required to link these two datasets together. The data processing proved to be very time consuming and labor intensive.

Third, administrative data may introduce new sampling biases that need special attention. Some subgroups may be under-represented in the administrative datasets due to various reasons. For example, analysts need odometer readings from at least two successive safety inspections to compute the annual mileage of a vehicle. Therefore, VMT from new cars purchased within one year are missing from the analysis, which may bias the VMT measures downwards for zones with large numbers of new vehicles[17]. A well-designed survey can help sort out appropriate weights to remedy the bias.

In summary, both survey data and administrative data have their pros and cons. Although survey data still dominate current research efforts, administrative data indeed provide a meaningful alternative data source. The employment of administrative data in urban modeling is not to replace survey data, but to reduce the dependence on surveys and to complement their usage in metropolitan planning.

[17] Despite this limitation, the safety-inspection-based VMT dataset used in this study is still better than the California emission-inspection-based VMT dataset used in Holtzclaw et al. (2002). California exempts new vehicles from emission inspections for the first two years, while the safety-inspection-based VMT dataset only misses new vehicles bought within one year.

6.4 FUTURE RESEARCH DIRECTIONS

In this final section, I discuss methodological issues that need to be further clarified as well as directions that this study can be extended in the future.

6.4.1 Causality

Due to the cross-sectional nature of the empirical analysis, I cannot construct causal relationships between the built environment, vehicle miles traveled, and property values, and the potential endogeneity could bias the estimates of the models. For example in Essay 1, I found that VMT is negatively associated with smart-growth type built-environment features. However, the direction of the underlying causal link cannot be identified: whether the built environment influences household travel behavior or whether preferences for certain travel pattern affect the choice of the built environment. If the latter direction is the dominant one, the observed association between the built environment and vehicle usage may be attributable to residential self-selection. For example, those preferring transit may consciously choose to live in transit-friendly neighborhoods and thus use car less. If so, the ability to use land-use-control policies to change household travel behavior may be limited. There is similar mutual causality issue in the property-value study: built-environment attributes like accessibility, connectivity, and walkability may increase property values; in the meantime, good built-environment amenities could be more likely provided in neighborhood with higher property values. Solving the causality issues necessitates either before-and-after datasets, or more complex econometric models, such as structural equation models and instrumental variable approach.

6.4.2 Behavior Mechanism

Due to data limitations, I lack detailed household-level demographic information in the study. In the VMT study, I have to carry out the analysis at the grid cell level. Even though I use small grid cells (of 15.4 acres each) as the basic spatial unit, they measure behavior aggregated across multiple households in the grid cell. Hence, the underlying behavior mechanisms by which the built environment influences individual decisions cannot be revealed by the study. Household-level demographic information with broad coverage is usually unavailable for analysts due to confidentiality concerns. Future analyses on the same research questions using household or individual survey data for Metro Boston would be a good complement for this study, which enables more in-depth exploration of the underlying behavior mechanism.

6.4.3 Spatial Autocorrelation, Housing Submarkets and Sample Selection

This study provides some evidence on the existence of spatial autocorrelation, submarkets and sample selection in the housing market, but many issues remain to be further explored to reveal the nature of these issue and the underlying relationships among them. For example, calibrating Heckman selection models for each time period rather than a pooled model like I used in this study could provide more insights about the temporal change in the pool of properties transacted and the behavior of homebuyers in choosing a property. A geographically-weighted regression could do better in capturing the spatial variation in the relationships between the built environment and residential property values than global models such as OLS model, spatial lag model, and spatial error model.

6.4.4 Extension of Study Areas

Since the analytical framework developed in this study can be readily applied to further research, the empirical analysis conducted in this study can be extended to other metropolitan areas. There is considerable regional variation in urban structures in the U.S., and the nature of the land use - transportation interconnection varies from place to place. Boston is a metropolis with relatively high density and good transit provision among U.S. cities. Comparative studies of Boston with other metropolitan areas, especially sprawl-type cities like Los Angeles and Atlanta, would provide a more comprehensive picture of metropolitan variation in the land use-transportation interconnection.

6.4.5 Policy Evaluation

This study explores the interconnections between land use and transportation. Currently, various programs that leverage these interconnections to promote stainable metropolitan growth are being implemented, such as urban growth boundary, mix-use planning, and transit-oriented development. The efficacy of these programs in reducing GHG emissions, however, is not well-studied, possibly due to various methodological challenges, such as residential self-selection. More comprehensive program evaluation would help planners and policy makers formulate effective smart-growth strategies to achieve sustainable metropolitan growth.

To summarize, future research needs to generate more in-depth insights into the nature of the land use-transportation interconnection and should provide useful information for governments and agencies to make informed decisions regarding the sustainable development of metropolitan areas.

REFERENCES

1. Al-Mosaind, M. A., K. J. Dueker, and J. G. Strathman. (1993). Light-Rail Transit Stations and Property Values: A Hedonic Price Approach. *Transportation Research Record* 1400, pp. 90–94.

2. Alonso, W. (1964). *Location and Land Use: Towards a General Theory of Land Rent*. Harvard University Press, Cambridge, Massachusetts

3. Armstrong, R. J., Jr. (1994). Impacts of Commuter Rail Service as Reflected in Single-Family Residential Property Values. *Transportation Research Record* 1466, TRB, National Research Council, Washington, D.C., 1994, pp. 88–98.

4. Anselin, L. (1993). Discrete Space Autoregressive Models. In: Goodchild, M., B. Parks and L. Steyaert (Eds.) *Environmental Modeling with GIS*. Oxford University Press, New York, pp. 454-469.

5. Anselin, L. and A. Getis. (1992). Spatial Statistical Analysis and Geographic Information Systems. *Annals of Regional Science*, 26(1), pp.19-33.

6. Anselin, L., A. Bear, R. Florax, and M. Yoon. (1996). Simple Diagnostic Tests for Spatial Dependence. *Regional Science and Urban Economics*, 26, pp. 77-104.

7. Badoe, D.A. and E.J. Miller. (2000). Transportation-Land-Use Interaction: Empirical Findings in North America, and Their Implications for Modeling. *Transportation Research Part D*, 5, pp. 235-263.

8. Bagley, M., P. Mokhtarian. (2002). The Impact of Residential Neighborhood Type on Travel Behavior: a Structural Equations Modeling Approach. *Annals of Regional Science*, pp. 279-297.

9. Bajic, V. (1983). The Effects of a New Subway Line on Housing Prices in Metropolitan Toronto. *Urban Studies*, 20, 2, pp. 147–158.

10. Basu, S. and T. Thibodeau. (1998). Analysis of Spatial Autocorrelation in House Prices. *Journal of Real Estate Finance and Economics,* 17(1), pp. 61–85.

11. Bhat, C. and J. Guo. (2007). A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels. *Transportation Research Part B,* 41, pp.506–526.

12. Boarnet, M. and R. Crane. (2000). *Travel by Design: the Influence of Urban Form on Travel*. Oxford University Press, New York.

13. Boarnet, M. and R. Crane. (2001). The Influence of Land Use on Travel Behavior: Specification and Estimation Strategies. *Transportation Research Part A*, 35, pp.823-845.

14. Boarnet, M. and S. 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*, Vol. 35, No. 7, pp.1155-1169.

15. Bowes, D. and K. Ihlanfeldt. (2001). Identifying the Impacts of Rail Transit Stations on Residential Property Values. *Journal of Urban Economics*, 50(1), pp.1-25.

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

17. Brownstone D. and T. Golob. (2009). The Impact of Residential Density on Vehicle Usage and Energy Consumption. *Journal of Urban Economics*, 65, pp.91-98.

18. Cao, T. and D. Cory. (1981). Mixed Land Uses, Land Use Externalities, and Residential Property Values: A Re-evaluation. *Annals of Regional Science*, 16, pp.1-24.

19. Cao, X., P. Mokhtarian and S. Handy. (2009). The Relationship between the Built Environment and Nonwork Travel: A Case Study of Northern California. *Transportation Research Part A*, 43, pp.548-559.

20. Case, B., J. Clapp, R. Dubin and M. Rodriguez. (2004). Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models. *Journal of Real Estate Finance and Economics*, 29, pp.167-191.

21. Cevero, R. (1996). Transit Based Housing in the San Francisco Bay Area: Market Profiles and Rent Premiums. *Transportation Quarterly*, 50, 3, pp. 33–49.

22. Cervero, R. (2002). Built Environments and Mode Choice: toward a Normative Framework. *Transportation Research Part D*, 7, pp. 265-284.

23. Cervero, R., and M. Duncan. (2002). Benefits of Proximity to Rail on Housing Markets: Experiences in Santa Clara County. *Journal of Public Transportation*, 5, 1, pp. 1-18.

24. Cervero, R. and K. Kockelman. (1997). Travel Demand and the 3Ds: Density, Diversity, and Design. *Transportation Research Part D*, 2(3), pp.199-219.

25. Chen, H., A. Rufolo, and K. J. Dueker. (1998). Measuring the Impact of Light Rail Systems on Single-Family Home Values: A Hedonic Approach with Geographic Information System Application. *Transportation Research Record* 1617, pp.38–43.

26. Chica-Olmo, J. (2007). Prediction of Housing Location Price by a Multivariate Spatial Method: Cokriging. *Journal of Real Estate Research,* 29(1), pp. 91-114.

27. Crane, R. (1996). On Form versus Function: will the New Urbanism Reduce Traffic, or Increase it? *Journal of Planning Education and Research*, 15, pp.117-126.

28. Crane, R. (2000). The Influence of Urban Form on Travel: an Interpretive Review. *Journal of Planning Literature*, 15(1), pp. 3-23.

29. Crane, R. and R. Crepeau. (1998). Does Neighborhood Design Influence Travel? A Behavioral Analysis of Travel Diary and GIS Data, *Transportation Research D*, 3, 4, pp.225-238.

30. Crane, R., A. Daniere, and S. Harwood. (1997). The Contribution of Environmental Amenities to Low-Income Housing: a Comparative Study of Bangkok and Jakarta, *Urban Studies,* 34, pp. 1495-1512.

31. Downes, T. and J. Zabel. (2002). The Impact of School Characteristics on House Prices: Chicago 1987–1991, *Journal of Urban Economics,* 52, pp.1–25.

32. Edel, M. and E. Sclar. (1974). Taxes, Spending and Property Values: Supply Adjustment in a Tiebout-Oates Model. *Journal of Political Economy,* 82, pp.941-954.

33. Ewing, R. and R. Cervero. (2001). Travel and the Built Environment: a Synthesis. *Transportation Research Record*, 1780, pp. 87-113.

34. Fan, Y. and A. Khattak. (2009). Impact of the Built Environment on Travel Distance and Time Costs: Trip-Level Analysis. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.

35. Fotheringham, A.S., C. Brunsdon, and M. Charlton. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, John & Sons, Incorporated.

36. Frank, L.D. and P.O. Engelke. (2001). The Built Environment and Human Activity Patterns: Exploring the Impacts of Urban Form on Public Health. *Journal of Planning Literature*, 16(2), pp. 202-218.

37. Frank, L. et al. (2007). Stepping towards Causation: do Built Environment or Neighborhood and Travel preferences Explain Physical Activity, Driving, and Obesity? *Social Science and Medicine*, 65: pp. 1898-1914.

38. Garrett, T. A. (2004). Light-Rail Transit in America: Policy Issues and Prospects for Economic Development. Federal Reserve Bank of St. Louis, Mo.

39. Gatzlaff, D. and D. Haurin. (1998). Sample Selection and Biases in Local House Value Indices. *Journal of Urban Economics,* 43, pp.192-222.

40. Gatzlaff, D. and M. Smith. (1993). The Impact of the Miami Metrorail on the Value of Residences near Station Locations. *Land Economics*, 69, 1, pp. 54–66.

41. Greenwald, M.J. (2003). The Road Less Traveled: New Urbanist Inducements to Travel Mode Substitution for Nonwork Trips. *Journal of Planning Education and Research*, 23, pp.39-57.

42. Greenwald, M.J. and M.G. Boarnet. (2001). Built Environment as Determinant of Walking Behavior: Analyzing Nonwork Pedestrian Travel in Portland, Oregon. *Transportation Research Record*, 1780, pp. 33-42.

43. Griliches, Z. (ed.) (1971). *Price Indexes and Quality Change*. Cambridge, MA: Harvard University Press.

44. Halvorsen, R. and R. Palmquist. (1980). The Interpretation of Dummy Variables in Semilogarithmic Equations. *The American Economic Review*, 70: pp.474-475.

45. Handy, S. (1996). Methodologies for Exploring the Link between Urban Form and Travel Behavior. *Transportation Research D*, 1(2), pp. 151-165.

46. Handy, S., M. Boarnet, R. Ewing R, and R.E. Killingsworth. (2002). How the Built Environment Affects Physical Activity: Views from Urban Planning. *American Journal of Preventive Medicine*, 23, pp.64-73.

47. Hansen, W. (1959). How Accessibility Shapes Land Use. *J. Am. Inst. Plan*. 25, pp.73–76.

48. Haurin, D. and P. Hendershott. (1991). Housing Price Indexes: Issues and Results, *AREUEA Journal*, 19, pp.259-269.

49. Heckman, J. (1976). Sample Selection Bias as a Specification Error. *Econometrica*, 47, pp. 153-161.

50. Hess, D. B., and T. M. Almeida. (2007). Impact of Proximity to Light Rail Rapid Transit on Station-Area Property Values in Buffalo. *Urban Studies*, 44, No. 5 & 6, pp. 1041–1068.

51. Hess, D.B. and P.M. Ong. (2002). Traditional Neighborhoods and Automobile Ownership. *Transportation Research Record*, 1805, pp.35-44.

52. Heywood, J., M. Weiss, A. Schafer, S. Bassene, and V. Natarajan. (2003). The Performance of Future ICE and Fuel Cell Powered Vehicles and Their Potential Fleet Impact. Publication No. LFEE 2003-004 RP, Massachusetts Institute of Technology, Laboratory for Energy and the Environment, Cambridge, MA.

53. Holzclaw, J. (1994). Using Residential Patterns and Transit to Decrease Auto Dependence and Costs. Natural Resources Defense Council for California Home Energy Efficiency Rating Systems, June 1994.

54. Holtzclaw, J., R. Clear, H. Dittmar, D. Goldstein and P. Hass. (2002). Location Efficiency: Neighborhood and Socio-Economic Characteristics Determine Auto Ownership and Use – Studies in Chicago, Los Angeles and San Francisco. *Transportation Planning and Technology*, 25, pp. 1-27.

55. Hwang, M. and J.M. Quigley. (2004). Selectivity, Quality Adjustment and Mean Reversion in the Measurement of House Values. *Journal of Real Estate Finance and Economics*, 28(2/3): 191-214.

56. IBI Group. (2000). Greenhouse Gas Emissions from Urban Travel: Tool for Evaluating Neighborhood Sustainability. Healthy Housing and Communities Series Research Report, prepared for Canada Mortgage and Housing Corporation and Natural Resources Canada, February.

57. International Energy Agency (IEA). (2004). The IEA/SMP Transport Spreadsheet Model, developed for the World Business Council for Sustainable Development Sustainable Mobility Project.

58. Jud, G.D. and T.G. Seaks. (1994). Sample Selection Bias in Estimating Housing Sales Prices. *Journal of Real Estate Research*, 9(3), pp.289-298.

59. King, A. (1977). Estimating Property Tax Capitalization: A Critical Comment, *Journal of Political Economy,* 85(2), pp.425-431.

60. Kitamura, R., P. Mokhtarian, and L. Laidet. (1997). A microanalysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation*, 24, pp. 125-158.

61. Krizek, K. (2005). Perspectives on Accessibility and Travel. In K. Krizek and D. Levinson (Ed.), *Access to Destinations*. Elsevier Ltd, pp. 171-193.

62. Malpezzi, S. (2002). Hedonic Pricing Models: A Selective and Applied Review. Paper prepared for *Housing Economics*: Essays in Honor of Duncan Maclennan.

63. Matthews, J. and G. Turnbull. (2007). Neighborhood Street Layout and Property Value: the Interaction of Accessibility and Land-use mix. *Journal of Real Estate Finance and Economics*, 35, pp.111-141.

64. Miller, E.J. and A. Ibrahim. (1998). Urban Form and Vehicular Travel: some Empirical Findings. *Transportation Research Record: Journal of the Transportation Research Board*, 1617, pp. 18-27.

65. Mills, E. (1972). *Studies in the Structure of the Urban Economy*. The Johns Hopkins Press, Baltimore, 1972.

66. Muth, R. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. University of Chicago Press, Chicago, Illinois.

67. Nelson, A.C., R. Pendall, C.J. Dawkins, and G.J. Knaap. (2002). The Link between Growth Management and Housing Affordability: the Academic Evidence. A Discussion Paper Prepared for The Brookings Institution Center on Urban and Metropolitan Policy.

68. Nelson, J. (1982). Highway Noise and Property Values: A Survey of Recent Evidence. *Journal of Transport Economics and Policy*, 16(2), pp.117-38.

69. Newman, P. and J. Kenworthy. (1999). *Sustainability and Cities: Overcoming Automobile Dependence*. Washington, DC: Island Press.

70. Population Reference Bureau (2008). 2008 World Population Data Sheet. Washington, DC.

71. Price, L., S. de la Rue du Can, J. Sinton, E. Worrell, Z. Nan, J. Sathaye, and M. Levine. (2006). Sectoral Trends in Global Energy Use and Greenhouse Gas Emissions LBNL-56144. Ernest Orlando Berkeley National Laboratory, Environmental Energy Technologies Division, Berkeley, CA, July 2006.

72. Rajamani, J. C. Bhat, S. Handy, G. Knaap, Y. 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, pp. 158-165.

73. Robsen, B.T. (1969). *Urban Analysis: a Study of City Structure with Special Reference to Sunderland*. Cambridge University Press, Cambridge.

74. Rodriguez, D. and J. Joo. (2004). The Relationship between Non-Motorized Mode Choice and the Local Physical Environment. *Transportation Research Part D*, 9, pp. 151-173.

75. Rodriguez, D. and C. Mojica. (2009). Capitalization of BRT Network Expansions Effects into Prices of Non-Expansion Areas. *Transportation Research Part A*, 43(5), pp.560-571.

76. Rosen, S. (1974). Hedonic Price and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82, pp.34-45.

77. Schipper, M. and V. Moorhead. (2000). Odometer Versus Self-Reported Estimates of Vehicle Miles Traveled. http://www.eia.doe.gov/emeu/consumptionbriefs/transportation/vmt/vmt.html.

78. Song, Y. and G.-J. Knaap. (2003). New Urbanism and Housing Values: a Disaggregate Assessment. *Journal of Urban Economics,* 54, 2003, pp.218-238.

79. Song, Y. and G.-J. Knaap. (2004). Measuring the Effects of Mixed Land Uses on Housing Values. *Regional Science and Urban Economics* 34, pp.663-680.

80. Srinivasan, S. (2001). Quantifying Spatial Characteristics for Travel Behavior Models. *Transportation Research Record*, 1777, pp. 1- 15.

81. Thünen, J. H. von. (1966). Isolated State. An English edition of Der isolierte Staat. Translated by Carla M. Wartenberg. Edited with an introduction by Peter Hall, Oxford, New York, Pergamon Press, 1966.

82. Tu, C. and M. Eppli. (1999). Valuing New Urbanism: The Case of the Kentlands. *Real Estate Economics*, 27, pp.425–451.

83. Turner, M., R. Gardner and R. O'Neill. (2001). *Landscape Ecology in Theory and Practice: Pattern and Process*. Springer-Verlag, New York.

84. United Nations (UN). 2001. *World Urbanization Prospects.* United Nations Population Division.

85. Untermann, R. (1984). *Accommodating the Pedestrian: Adapting Towns and Neighborhoods for Walking and Bicycling*. Van Nostrand Reinhold, New York.

86. Weinberger, R. R. (2001). Commercial Rents and Transportation Improvements: The Case of Santa Clara County's Light Rail. Lincoln Institute of Land Policy, Cambridge, Mass.

87. Zegras, P.C., Y. Chen, and J.M. Grütter (2009). Behavior-Based Transportation Greenhouse Gas Mitigation under the Clean Development Mechanism: Transport-Efficient Development

in Nanchang, China. *Transportation Research Record: Journal of the Transportation Research Board,* 2114, pp.38-46.

88. Zhang, M. (2004). The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong. *Journal of the American Planning Association*, 70, 3, summer, pp.344-360.

89. Zhang, M. (2005). Exploring the Relationship between Urban Form and Nonwork Travel through Time Use Analysis. *Landscape and Urban Planning*, 73, pp.244–261.

90. Zhang, M. and N. Kukadia. (2005). Metrics of Urban Form and the Modifiable Areal Unit Problem. In *Transportation Research Record: Journal of the Transportation Research Board*, No.1992, Transportation Research Board of the National Academies, Washington, D.C., pp.71-79.

APPENDICES

APPENDIX 1: SPATIAL-ERROR MODELS USING BUILT-ENVIRONMENT FACTORS AND DEMOGRAPHIC VARIABLES

In the first study, for comparison purpose, I also calibrated the spatial error model with built-environment factors and 3 demographic variables, median household income, percent of households with less than 3 member, and percent of population 16 years old and over and in labor force. Each demographic variable represents one demographic factor. The estimation results and the change in VMT measures due to one standard deviation increase in the independent variables are presented in Tables A-1 and A-2. The major conclusions of Essay 1 still hold, except that the coefficient of the median household income variable has a positive and insignificant coefficient in the VMT per vehicle model.

.

Table A-1: Estimation Results of Spatial Error Model Using Built-Environment Factors and Demographic Variables

	VMT per Vehicle			VMT per Household			VMT per Capita		
	Coef.	t-stat.		Coef.	t-stat.		Coef.	t-stat.	
Built-Environment Factors									
Distance to non-work destinations	442.5	21.12	**	3842.7	23.39	**	878.5	16.21	**
Connectivity	-248.1	-23.14	**	-2990.9	-35.18	**	-849.2	-30.11	**
Inaccessibility to transit & jobs	1006.0	32.18	**	6017.5	30.51	**	1970.0	30.87	**
Auto dominance	-9.4	-0.97		571.3	5.92	**	267.6	8.21	**
Walkability	16.3	1.88		-1571.1	-19.60	**	-596.7	-22.15	**
Demographic Variables									
Median household income in thousand dollars	0.3	0.72		25.0	6.56	**	6.5	5.08	**
Percent of household with less than 3 members	104.7	1.70		-2515.8	-4.01	**	739.4	3.48	**
Percent of population 16+ years old and in labor force	177.7	2.50	*	152.3	0.21		818.3	3.32	**
Lambda	0.91	398.15	**	0.84	231.49	**	0.83	219.59	**
Constant	12194.3	148.82	**	30813.3	40.11	**	9188.8	35.48	**

* and ** denote coefficient significant at the 0.05 and 0.01 level respectively.

Source: Calculated by the author.

Table A-2: Change in VMT Measures Due to One Standard Deviation Increase in Built-Environment Factors and Demographic Variables

	VMT per Vehicle	VMT per Household	VMT per Capita
Built Environment Factors			
Distance to non-work destinations	383.0	3325.3	760.2
Connectivity	-290.7	-3504.7	-995.1
Inaccessibility to transit and jobs	978.6	5853.5	1916.4
Auto dominance	-5.7	348.5	163.3
Walkability	15.0	-1447.1	-549.6
Demographic Variables			
Median household income	7.6	683.7	178.5
Percent of households with less than 3 members	12.8	-306.9	90.2
Percent of population 16+ years old in labor force	15.3	13.1	70.2

Source: Calculated by the author.