Exploring Housing Market in Toronto, Ontario: Spatial Hedonic Modeling of Crime Rates, Subway Ridership, Dwelling Density & House Prices

by

Chi Chen

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Planning

Waterloo, Ontario, Canada, 2015

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis explores house prices and its relationship to neighborhood characteristics in Toronto, Ontario. In particular, by applying spatial hedonic models at the census tract level, this study examines the association between house prices and three neighborhood characteristics: crime rates, subway ridership and dwelling density. House prices were first explored by cluster analysis at multiple listing service (MLS) district level and results showed a significant spatial clustering of the prices in Toronto. Spatial hedonic models were then conducted on census tract level to examine the role of neighborhood characteristics in explaining the spatial patterns. The spatial mode was applied both across the entire city and separately in three neighborhoods divided by income level. Findings indicate that on the citywide scale, crime rates and dwelling density do not significantly influence house prices, but subway ridership was positively associated with house prices. In the middle-income neighborhood, six types of crime were found to significantly decrease house prices. Densities of all types of dwelling were found to be positively associated with house prices while apartment density decreases house prices in the middle-income neighborhood. Findings from this research can be applied to inform housing and transportation policies, regarding neighborhood improvement, housing affordability and smart growth.

Acknowledgements

First of all, I would like to thank my advisor Dr. Jane Law, for her patient guidance, encouragement and insight that contributed to all aspects of this research. I would also like to thank Dr. Mark Seasons and Dr. Pierre Filion for their helpful advice. I am gratitude to the staff members of Geospatial Center in University of Waterloo, who provided data access and GIS support on many occasions. I would also like to thank the School of Planning, for providing endless opportunities to develop and discover. I express my gratitude to my mother and father, who have always been supportive during my ups and downs in Waterloo. Finally, I would love to thank my dear friends who have made my journeys more cheerful and fulfilling.

Table of Contents

List of Figures

List of Tables

Chapter 1.
Introduction

1.1 Motivation

There is no lack of attention to the housing market from the media, the development industry, government policies, or the academic literature. House prices, as a key indicator of the housing market, reflect the regional economy, neighborhood stability and individual household wealth (Leung, 2004). Local variation of house prices within urban areas is not uncommon (Gibbons & Machin, 2008). To explain the price variation in the simplest urban economic model (in a monocentric city), land values, which constitutes a large proportion of house prices, increase toward the city centre, because land users compete for the most accessible land (Alonso, 1964). However, in the modern city that are multi-centric, employment subcentres and transit nodes are exerting greater influences than the central business district (CBD) on the urban structure and the residential landscape.

The desirability of neighborhood is often associated with the local house price variation, and a diverse range of neighborhood characteristics has gained empirical attention (Gibbons & Machin, 2008), such as transit accessibility, school quality, air quality or views. To narrow down the scope, we focus on the roles of crime rates, transit ridership and dwelling density in explaining the local house price variation. There are plausible reasons for our interest in these three neighborhood characteristics. First, although crime threatens people's quality of life and is commonly viewed as decreasing property values, empirical challenges exist for crime studies and their findings are often mixed. Second, transit-oriented development (TOD) has been gaining popularity in urban development (Filion, 2011) and transit nodes are exerting great influences on neighborhood desirability, especially due to the urban lifestyle generated by the mixed land uses (Cervero, 2006). Third, as emphasis on higher density residential development is a major shift in recent redevelopment policies in large cities, home buyers' willingness to pay for increased dwelling density is of interest to both policymakers and property developers.

This study is motivated by the governing research questions: Do variation of house prices in Toronto exhibit spatial patterns? And if so, what are the roles of crime rates, transit ridership and dwelling density in explaining the price patterns? How can these findings be applied to inform planning practices? To explore the questions, four research objectives (Figure 1.1) are to be achieved: 1) Conducting exploratory spatial data analysis (ESDA, e.g. cluster analysis) to examine spatial patterns of Toronto house prices; 2) Conducting confirmatory spatial data analysis (CSDA) namely spatial hedonic price analysis to explore the association between house prices and neighborhood characteristics including crime rates, transit ridership and dwelling density; 3) Using the same spatial hedonic model to examine if the association differ among various neighborhoods within the city; 4) Exploring how the analytical results can be applied to planning policymaking, in particular housing, transportation and smart growth policies.

For the *first* research objective, cluster analysis including Moran's I techniques and mapping were used to identify house price clusters in Toronto. High-value house clusters were identified along the subway lines and near the four employment centers. Exploratory spatial analysis proves as a valuable starting point for the confirmatory spatial analysis (regression modeling). For the *second* objective, regression modeling approach namely ordinary least square (OLS), spatial error and spatial lag models were explored to determine the best fitting model. Spatial error hedonic function was the fittest model for the dataset. To examine the impacts of crime rates on house prices, crime was disaggregated into their component crime types (e.g. we distinguish the impact of crimes such as theft of a vehicle from break and enter), instead of the overall crime rate. When the best fitting model was applied on the citywide level, theft of a vehicle was the only type of crime to significantly influence house prices; subway ridership was positively associated with house prices; and dwelling density did not exhibit significant impact at the citywide level.

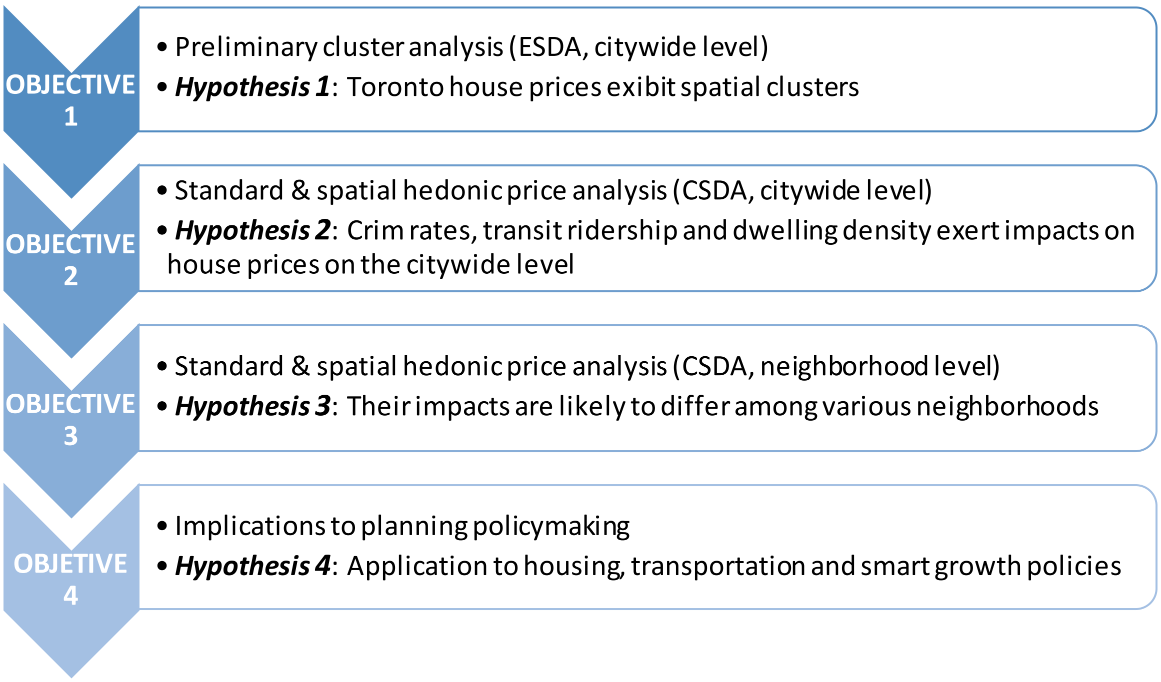

Figure 1.1 Objectives and hypotheses of this thesis

This naturally leads to the *third* objective, where the same spatial hedonic model was applied to three different neighborhoods divided by the income level of the census tracts. Findings indicate that six types of crimes had significant impacts on house prices in the middle-income neighborhoods, but not significant in high- and low-income neighborhoods. Also, higher house prices were associated with lower apartment density in the middle-income neighborhood. These various impacts of neighborhood characteristics on house prices further confirm that a disaggregated approach (dividing the market into submarkets) can be helpful for a detailed and more accurate analysis.

The *fourth* objective explores how the findings of this thesis can be applied to inform planning practices. First, allocating police resources to reduce neighborhood crime can have trickle-down effects on the welfare of households if houses are to be considered as their assets. Second, housing affordability can be considered as the cost of housing plus the cost of transportation. Transportation policies should synchronize with housing policies to acknowledge the trade-offs behind peoples' residential location decisions. Third, mixing house types can increase residential density as promoted by the smart growth movement and potentially improve housing affordability outcome, but such mixing projects should be carefully designed to attract target markets' by mitigating possible negative impacts such as traffic congestion due to increased activities. Incentive and Inclusionary zoning are possible planning implementation tools to promote housing affordability in transit-oriented development.

1.2 Outline

This thesis is organized into 7 chapters and proceeds as follows. *Chapter 2* briefly reviews past research regarding fundamental theories in urban economics and housing market, as well as empirical studies using spatial hedonic approaches. *Chapter 3* describes the study area: the city of Toronto, Ontario and provides descriptive statistics of house prices, crime, density, transit ridership and other socio-economic data involved in this thesis.

Chapter 4 focuses on a preliminary data analysis by examining spatial patterns of house prices in Toronto. Moran's I techniques and map-making were employed to uncover the spatial patterns, which are discussed in relation to Toronto's multi-centric urban structure. In *Chapter 5*, hedonic price models were developed to investigate the impacts of crime, transit ridership and dwelling density on house prices, both across the entire city and three income-based neighborhoods. Regression diagnostics were compared among standard hedonic models and spatial hedonic models (including spati al lag dependent model, spatial lag independent model, and spatial error model). Spatial error hedonic model was the best fitting model for our dataset.

Chapter 6 provides detailed interpretation of the findings as well as the limitations of this study, followed by implications to housing, transportation and smart growth policies. Finally, *Chapter 7* concludes this thesis by reviewing findings and contributions, discussing challenges encounter during the research and providing thoughts on possible future research areas. Figure 1.2 below outlines the structure of this thesis by chapters in aligns with the four objectives.

Chapter *1*: Introduction

Reserach motivation

Thesis outline

Chapter *2*: Theoretical & Empirical Overview

Urban housing market

Hedonic price function

Valuation of neighborhood amenities

Chapter *3*: Study Area & Data Description

Chapter *4*: Housing Market and Cluster Analysis

Preliminary analysis

Do Toronto house prices exibit spatial clusters? (*Objective **1***.)

Exploratory spatial data analysis (ESDA)

Chapter *5*: Regression Modelling Approach & Results

Data analysis & Results

Do crime rates, transit ridership and dewelling density influence house price? (*Objective **2***.)

Would their impacts differ among various neighborhoods? (*Objective **3***.)

Confirmatory spatial data analysis (CSDA)

Chapter *6*: Discussion and Interpretation

How can the findings be interpretated? (*Objective **2 & 3***.)

What are the implications to planning policymaking? (*Objective* **4.**)

Chapter *7*: Conclusions

Concluding thoughts

Future research areas

Figure 1.2 Thesis Structure and Chapters

Chapter 2.
Theoretical & Empirical Overview

Chapter Overview

Broadly, this research draws from theories within the field of urban economies. Specifically, the location theory and the residential sorting effects altogether form the theoretical basis of this study in explaining the urban housing price variation. Theories of hedonic price modeling are reviewed with a focus on the impacts of crime rates, transit and dwelling density on house prices. As the urban housing market is inherently spatial, the application of spatial hedonic models that acknowledge a fuller spatial effects than traditional hedonic models is expected to be valuable.

2.1. Urban Housing Markets and Residential Locations

Mainstream urban economics depicts real estate development as "relatively unproblematic [with] … transactions and investment seem to be activated by market signals as to land and property prices and rents" (Healey, 1991, p. 222). How diverse is a housing market and why house prices vary within the market are questions many studies attempt to address. We draw from two streams of basic urban economic theories: the location theory and residential sorting.

2.1.1 Location Theory

The fundamental characteristic of a urban housing (and land) market is that housing is more expensive at more advantageous sites (Dipasquale & Wheaton, 1996). In a simple model of a mono-centric city, for example, land prices increase towards the centre of a city, because land users (e.g. business owners, residents) compete for the most accessible land near the city centre (Alonso, 1964; Muth, 1969), and the amount they are willing to pay is called "bid rent". Similarly, the Ricardian Rent theory (Ricardo, 1817) states that the rent of a location equals to the economic advantage gained by utilizing the land site in its most effective use. As a result, a concentric pattern of land uses exists for a city's zoning model. According to this branch of theory, low-income neighborhoods would be found on the outskirt of a city because this is the location they can afford (Lerman & Kern, 1983; Duncan, 2010).

Doubts arise regarding the application of location theory and bid rent theory in modern cities that are no longer mono-centric, but poly-centric (or multi-centric), as employment subcentres are moving away from the central business district (CBD) and transit-oriented development (TOD, or nodal development) is gaining popularity (Heikkila, Gordon, Kim, Peiser & Richardson, 1989; Filion & Kramer, 2012). House prices may not necessarily decline with distance from the CBD and land values will have less variation within a city. Also, though typically density is high in CBD due to economies of scale and the scarce of land resources, density may not necessarily decline significantly with distance to CBD as in a monocentric model (Champion, 2001). Transit nodes and employment subcentres therefore are exerting increasing influences on the urban spatial structure and housing market.

2.1.2 Residential Sorting Effects

Apart from the above-mentioned location theories that relate land prices to urban structure (and land use), the residential sorting effects (also Tiebout sorting, 1956) associate the residential landscape with people's lifestyle preferences, attitudes and values. People face trade-offs when making decisions of residential locations, which allocates various types of people into different neighborhoods (Glaeser et al, 2006; Gibbons & Machin, 2008; Nhuyen-Hoang & Yinger, 2011). For example, low-income households trade off lower quality of public services for lower local taxes. Or sometimes they trade off greater living space for less commute time and cost. Therefore, low-income housing in many North American cities is found in the inner city, rather than the outskirt according to the bid-rent theory in a monocentric city. This trade-off process is further complicated by the multi-centric model, where different parts of a city have specialized functions and varied growth potential (e.g. as an employment centre or a residential area). Instead of having a single reference point (the CBD), people have to locate their homes with consideration to the importance they attach to various needs (Champion, 2001).

A matter similar to residential sorting can be found in environmental justice and urban gentrification literature, where economically disadvantaged households are sorted into unfavorable locations. In his influential paper "Just Garbage", Wenz (2001) argues that when it comes to locally undesirable land uses (LULU, e.g. underground toxic waste, waste management facilities, prison) that diminish property values, wealthy people has greater mobility: they can afford to move out and leave the less desirable areas to the economically disadvantaged households. Also, urban gentrification occurs when higher income neighborhoods, who have more power over lower income neighborhoods, bid for more desirable locations. When a once popular neighborhood falls out of favor with diminished property value, lower income households take over the location and replace the original residents (Hulchanski, 2010).

2.2. House Prices and the Valuation of Neighborhood Characteristics

As has been discussed above, it is the household willingness to pay for the most attractive locations that bid up land site prices and sort different people into various locations. Willingness to pay for neighborhood amenities, therefore, lies at the core of urban economic theories regarding city structure and residential landscape (Gibbons & Machin, 2008). Hedonic price function is widely used in both theory and empirical practices to evaluate consumer preferences (or willingness to pay). We briefly explain the model in this section, especially the application of the model in evaluating three neighborhood characteristics: crime rates, transit ridership and dwelling density.

2.2.1 Modeling house prices with hedonic regression

House buyers not only purchase the land and the house, but a whole package including surrounding amenities and neighborhood characteristics related to the house. When making a purchase, buyers are not made aware of the value of each individual component that make up the package. The most popular valuation model is the Rosen's hedonic model (1974), which has been regularly used for mortgage underwriting, property taxation, as well as property price generation (Fahrländer, 2006; Lehner, 2011). Briefly, hedonic models uses multiple regression analysis to disaggregate house prices into the values of the components of the house's characteristics.

Various attempts were made (Butler, 1982; Song & Knaap, 2004; Sirmans et al. 2009) to categorize the diverse housing and neighborhood characteristics associated with property values. Most commonly used categories are: structural attributes of houses (e.g. size, number of rooms and age of the dwelling unit), locational attributes (e.g. distance to central business districts) and neighborhood characteristics (e.g. crime rate, dwelling density). Additional attributes regarding internal features (e.g. bath, basement), external features (e.g. garage space, pool) and natural environment characte ristics (e.g. lake view, ocean view) of the dwelling were also included in some hedonic studies.

Despite the usefulness of hedonic price models, their estimation is often compromised by three problems: choice of functional form, omitted variable bias and spatial autocorrelation (Armstrong & Rodriguez, 2006). Simple functional forms of hedonic models include linear, log linear, semi-log and double-log. A study by Cropper, Deck and McConnell (1988) found that these simple functions perform better than complex ones. Omitted variable bias occurs when some factors are not incorporated in the model and the effects of other factors are over- or underestimated. One example of omitted variable problem is overlooking potential externalities associated with transit proximity (e.g. noise). Spatial autocorrelation as a result of spatial dependence and unobserved heterogeneity, if not accounted for, may lead to inaccurate estimation results. This particular problem will be addressed from a spatial perspective in *Section* 2.3.

A highly criticized problem of the hedonic model is that it does not consider heterogeneity within a general market (Islam & Asami, 2009): it assumes that all homebuyers are similar and a single demand equation is adequate for analysis. In other words, hedonic price function can be highly non-linear and the slope and the shape of the relationship is different depending on the market (Gibbons & & Machin, 2008; Tita, et al., 2006). Ample evidence suggest that who lives where is often determined by the demographic characteristics of residents such as income, race, education (Hulchanski, 2010; Cullen & Levitt, 1999; Morenoff & Sampson, 1997). Tita, et al. (2006) also found that by categorizing neighborhoods by income levels, impact of crime differ across the different neighborhoods. In this sense, linear function may not accurately reflect the varied willingness to pay for neighborhood characteristics among different homebuyers in one general housing market (e.g. housing market of Toronto).

2.2.2 Valuation of Neighborhood Characteristics

Characteristics of a place plays an important role in determining the market price of houses and the desirability of neighborhood is often associated with local price variation (Gibbons & Machin, 2008). Hedonic studies of house prices have analyzed the impacts of a variety of neighborhood characteristics such as racial composition, crime rates, education, air pollution, dwelling density and transit accessibility. To narrow down the scope, the focus of this study is on the impacts of crime rates, transportation and dwelling density on house price variations and in turn what information can reveal about homebuyers' willingness to pay for public safety, transit impacts and dwelling density in their neighborhoods. We discuss the reasons for our interest in these three factors.

First, as a public 'bad', crimes are generally expected to exert a downward impacts on house prices. Indeed, findings in the extant literature show that higher crime rates were associated with lower property values (e.g. Thaler, 1978; Dubin & Goodman, 1982; Haurin & Brasington, 1996). However, there is a body of studies showing that impacts of crime on house prices were very small or insignificant (e.g. Lynch & Rasmussen, 2001). Gibbons (2004) found that highly-visible crimes such as vandalism and graffiti have a greater negative impacts on house prices than break and enter in London, UK.

In their study, Tita et. al. (2006) pointed out that crime is often studied without further considering the impacts of different types of crime. They therefore examined the relationship between crime and house prices at a disaggregated level, which is both the disaggregation of total crime into their component ones and the disaggregation of a place into income-based neighborhoods. Their findings indicate that the average impacts of crime on house prices can be misleading and the degree that crime is capitalized differ for wealthy, middle-class and poor neighborhoods. This result corresponds to the previously noted theory that the housing market is more likely to respond to local rather than citywide variations and homebuyers' willingness to pay are likely to differ in one general market.

Second, transportation infrastructure is critical in shaping a city's urban structure and residential landscape. The impact of transportation on house prices has a long history in hedonic studies, but the results are often mixed and difficult to generalize or compare (Armstrong & Rodriguez, 2006; Dewees, 1976). Some studies have found the association between increased property values and better transit accessibility, while others found the link rather weak. For example, Dewees (1976) found that the new subway infrastructure alone does not substantially increase the rent along lines in Toronto, and further suggested examining the influences of multi-centric (rather than mono-centric) models and designing performance variables (instead of distance-based variables). Also, findings of these transit studies are often context-specific (e.g. regarding certain transit projects in a metropolitan area) and the approaches employed such as proximity measurement are far from standardized (Gibbons & Machin, 2008). It is also challenging in hedonic modelling design to separate out impacts of transportation externalities such as noise or congestion on house prices.

Third, density measurement, simply calculated and expressed in numbers, is commonly used in land use planning policies and regulations. Homebuyers' willingness to pay for development density (e.g. dwelling density) is also of interest to real estate developers, whose primary concern is the potential capacity and financial yield of a development project (Taylor & Nostrand, 2008). A great deal of research on property prices have designed measures of density such as dwelling unit density, population density, or combined employment and population density. In their study, Song & Knaap (2004) found that proximity to high density development depresses prices of nearby single detached houses in Portland. This result also corresponds to market surveys that consumer generally prefer low-density housing

(Neuman, 2005; Sloane, 2006). As high density is a key element of smart growth and urban intensification is on the agenda of provincial and municipal policies, we intend to further evaluate the impacts of dwelling density on house prices and in turn homebuyers' willingness to pay for density-related neighborhood characteristics, as well as the indication to planning policymaking.

2.3 Spatial Hedonic Models and the Housing Market

With improved geographic technology and better access to spatial data, researchers have become increasingly aware that neighborhood and locational attributes in standard hedonic modeling do not necessarily take into account the complete range of spatial effects (Anselin, 1992; Cho, Poudyal, & Roberts, 2008). Standard hedonic estimations are often flawed in terms of omitted spatial autocorrelation due to spatial dependence and unobserved spatial heterogeneity (Armstrong & Rodriguez, 2006). Also, in a property market where "location, location and location" is often said to be the primary determinant of house prices, spatial methods should be reasonably expected to be useful in explaining local house price variations.

The lack of spatial consideration in traditional models may cause biased estimation results, while spatial hedonic studies sometimes have remarkable results and profound implications. For instance, by incorporating into spatial hedonic models the negative externalities (e.g. noise) of commuter rail right-of-way, Armstrong and Rodriguez (2006) found that proximity to commuter rail station decreases property values in metropolitan Boston. Such willingness to pay revealed in their study validates concerns about the capitalization of transit infrastructure and service in transit-oriented housing. Another study employed spatial hedonic analysis to estimate the value of green open space in different neighborhoods (Cho, et. al., 2008). Their findings show that different features of open space vary according to the place's degree of urbanization: mixed forests are better valued in urban cores while diverse landscape and natural forest edges are better valued in rural areas. Therefore, it is our intention in this study to add spatial perspectives into the hedonic price analysis.

Chapter 3.
Study Region and Data

Chapter Overview

The first section of this chapter introduces the geography and demographics of the study area, as well as why it was chosen. The second section outlines relevant local and regional planning policies that help shape the housing market of the study area. The third section justifies the unit of analysis used, followed by descriptions of the outcome and explanatory variables upon which this research is based. The descriptions include data sources and variable creation methods. Descriptive statistics of each explanatory variable are also presented in tables.

3.1 Study Area: City of Toronto

This empirical study focuses on the City of Toronto, located on the North shore of Lake Ontario. It has a population of around 2.5 million, representing 44.8% of the total population in Toronto Census

Metropolitan Area (CMA). The City of Toronto is the main urban center in the Toronto CMA, which is the fasted growing CMA in Canada. Toronto CMA has a population growth rate of 9.2% from 2006 to 2011, while the national growth rate in the same period was 5.9%.

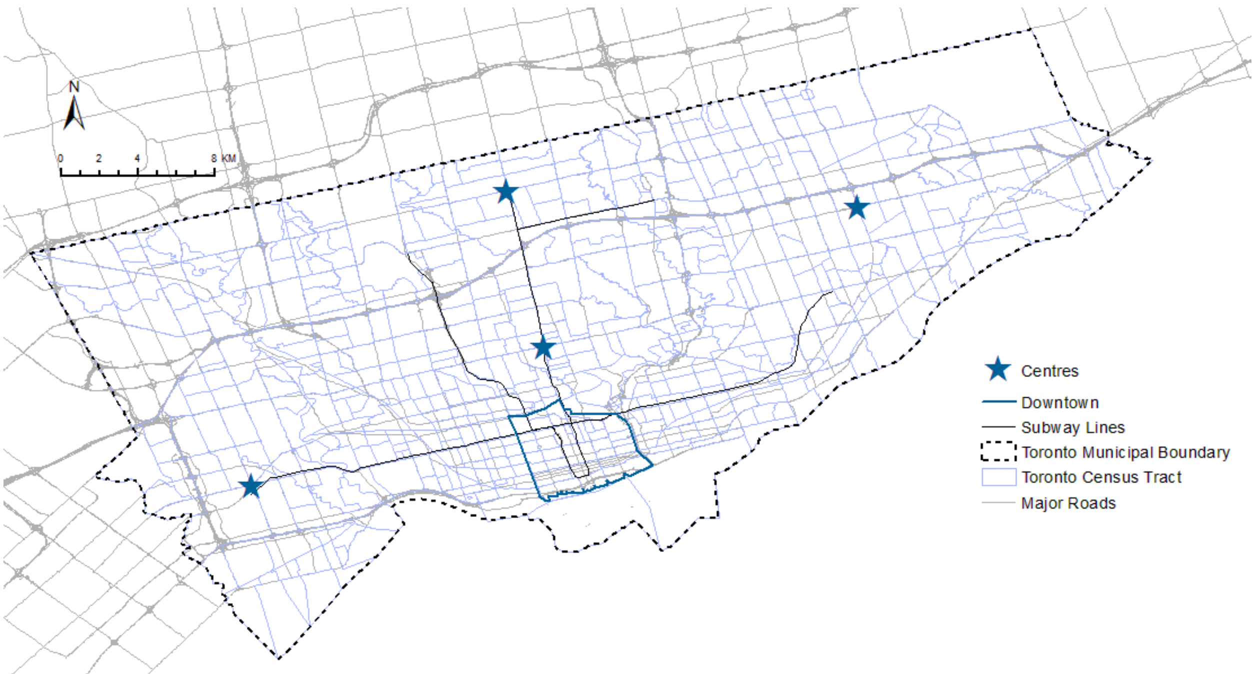

Figure 3.1.1. Map of the Study Area - City of Toronto.
Note: *For reference, downtown Toronto, the four Centres and subways lines are highlighted.*

The City of Toronto was chosen as the study area for pragmatic reasons. First, Toronto is a well-researched area, with large literature on its urban development, housing market and demographic trends. Second, for a large municipality like Toronto, house prices and socio-economic data are publicly available at small-area levels, which may not be the same case for other smaller municipalities. Third, in recent decades, Toronto has experienced relatively rapid population growth, due to international and domestic in-migration, which boosted job market and therefore increased local housing demand. Low interest rate has also contributed to the increasing housing activity in the study area.

Most importantly, Toronto has a strong commitment to public transit and high density development for more than 50 years (Filion & McSpurren, 2007), and is one of the most successful transit cities in the world (Robert, 1993). The city has remarkable concentrations of population and employment around rail (subway) stations. This clustered urban development around transit stations can be largely attributed to the local government's ability to acquire land along transit corridors and later lease or sell them to residential and commercial developers (Cervero, 1993). All those reasons make Toronto an interesting case for our study.

3.2 Overview of Local and Regional Planning Policies

Local and regional planning documents play an important role in shaping Toronto's urban structure and housing market. The *Growth Plan for the Greater Golden Horseshoe* (also called the "*Growth Plan*") promotes urban densification and restricts outward growth, which requires efficient use of land and existing infrastructure (Ontario, 2006). It has been considered as one of the most progressive planning document, reforming land use planning system, emphasizing smart growth and promoting nodal development along transit corridors (Figure 3.2.1). The *Greenbelt Plan* is another piece of prominent planning document having an impact on the region's residential structure. The plan protects agricultural and rural lands by supporting modest growth of towns and villages, during which their main characteristics and servicing capacities are remained. In this way, the plan has set restrictions on greenfield development, and consequently directed development projects towards brownfield development and urban intensification (Ontario, 2005).

The current *Official Plan* for the City of Toronto, first adopted in 2002, steered future growth within Toronto areas that are well served by transit and existing road network. It particularly defines strategic locations: the *Downtown* as the heart of Toronto, the four *Centres* as the vital mixed-use communities (including *Etobicoke Centre, Yonge-Eglinton Centre, North York Centre* and *Scarborough Centre,* which were also identified in the Growth Plan as urban growth centres in the GGH area, Figure 3.2.1)*,* the Avenues as corridors for new housing and job opportunities, and the Employment Districts as designated areas to support business and employment growth. Figure 3.2.2 highlight the locations of those strategic development areas, for which housing policies are specified as "to encourage a full range of housing opportunities (e.g. ownership, rental, emergency housing) and to reduce demand for in-bound commuting" (City of Toronto, 2010).

Figure 3.2.1. Urban Growth Centres in Greater Golden Horseshoe region.
Note: *This map is not to scale, does not accurately reflect approved land use and planning boundaries.*

(Source: Growth Plan for the Greater Golden Horseshoe, 2006)

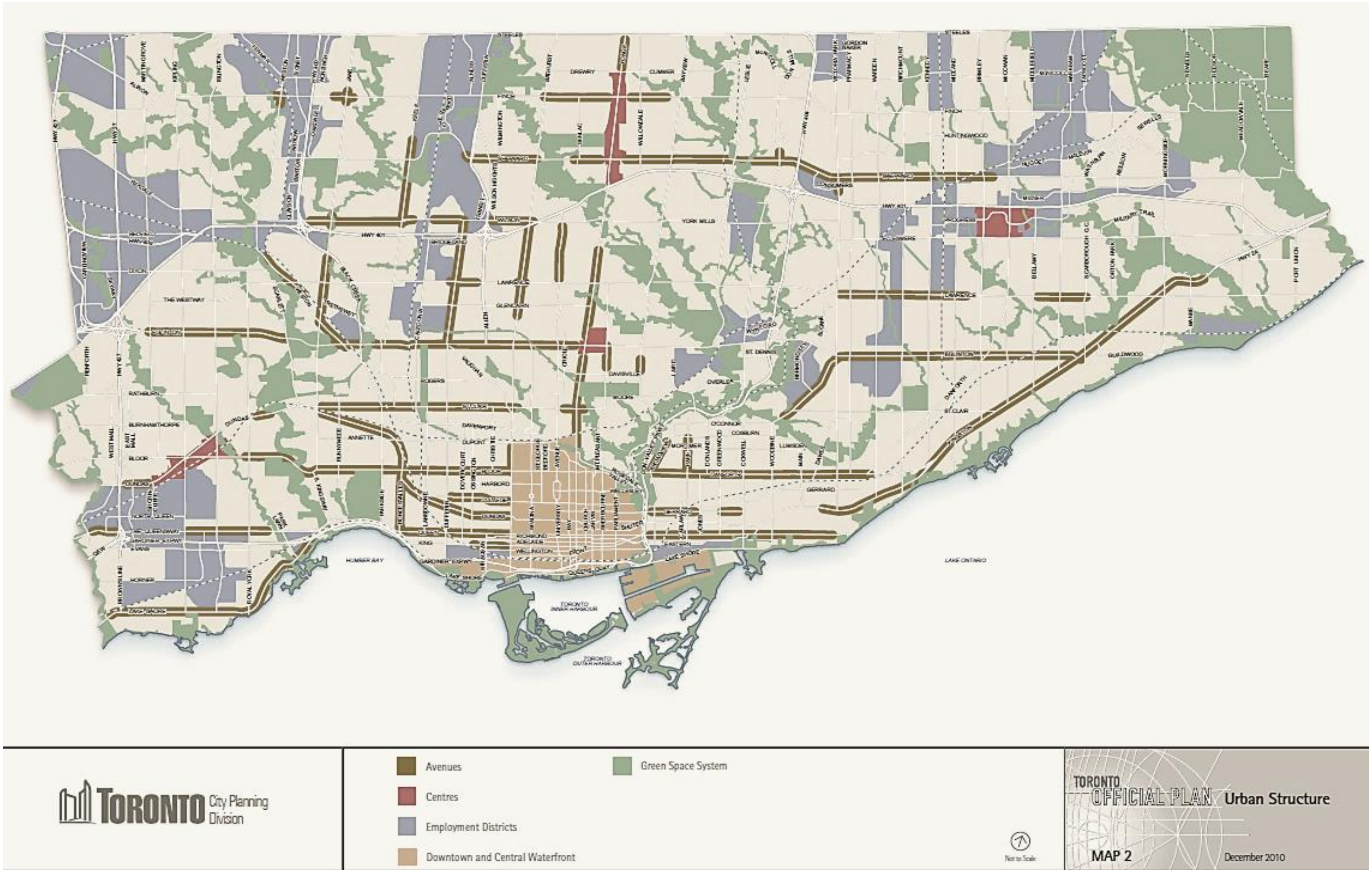

Figure 3.2.2. Urban Structure of Toronto (Strategic development areas: avenues, centres, employment districts highlighted)
Source: *Official Plan* for the City of Toronto, 2011

3.3 Data Description

In this section, different types of data used in this analysis are discussed including their data sources, how variables are generated by aggregation and calculation. The outcome (dependent) variable is 2006 house prices in Toronto and explanatory (independent) variables include household characteristics and neighborhood characteristics (crime rate, employment, income, ethnic, etc.). Descriptive statistics for each variable are provided including their minimum, maximum and mean values, as well as standard deviations.

3.3.1 Unit of Analysis – Census Tract

For analysis, the study area was disaggregate into smaller spatial units - census tracts. Census tracts are defined by Statistics Canada (2012) as small, relatively stable areas, with a population between 2500 to 8000 persons. The boundaries of census tracts follow permanent and identifiable physical features such as arterial roads (Statistics Canada, 2012).

The census tract was selected as the unit of analysis for the following three reasons. First, census tract as small-area unit captures spatial variations of socio-economic status of neighborhoods better than larger spatial scales (e.g. neighborhoods, census subdivision). Second, Toronto crime data was only available at census tract level, which is the spatial scale that other socio-economic data (e.g. income, education) are available at. Third, the size of census tracts are close to that of secondary plan areas as well as site and area-specific policies, identified in the *Official Plan* for the City of Toronto (2010). This means that the findings based on census tract analysis can be applied to local land use planning policies and practices.

3.3.2. House Price Data

The dependent variable is defined as house prices of Toronto, of which the data were extracted from 2006 monthly house sales and average price by Market Watch, Toronto Real Estate Board (TREB). TREB is a non-for-profit corporation and is the largest real estate board in Canada. It provides monthly and quarterly market report on average house sale price aggregated on Multiple Listing Service (MLS) district level. Figure 3.4.2.1 illustrates the 35 MLS districts within the city of Toronto. The monthly reports disaggregate houses types into single-family detached houses, semi-detached houses, condo apartment, and condo townhouse. In this study, the house price data only includes single family detached houses. This is because for other house types (e.g. semi-detached houses, apartments or condos), the structural characteristics (e.g. building units, number of rooms) among them vary drastically and TREB did not provide such information for us to include in the analysis. An alternative data source for house prices is Statistics Canada public-use micro data files (PUMFS), where the value of dwellings were self-assessed by homeowners. However, market data based on transactions are less biased from personal opinions and are preferred in almost all hedonic studies reviewed.

For analysis, house price data on MLS district level were assigned to each census tract. The MLS district boundaries align with major arterial roads, which also align with certain census tract boundaries. This means certain neighboring census tracts fall into the same MLS district, and therefore have the same average house price. (Drawbacks on this method is further discussed in *Chapter 6* Limitation part of the study; Figure 3.4.2.2 illustrates the boundary matches between MLS districts and census tracts). A natural log transformation for the house price was used. Previous studies found that a log transformation performs better than a linear function in that it corrects for heteroscedasticity, which is a major concern for regression analysis (Wooldridge, 2003).

Table 3.3.1. Descriptive Data for the Study Area (Toronto, 2006)

	Total Study Area	**Census Tracts**			
		Min.	Max.	Mean	S.D.
Geographic Area (km^2)	624.71	0.07	28.72	1.20	1.70
Log house price	13.14	12.54	14.23	13.07	0.38
MLS districts	35	-	-	-	-
Census Tracts	522	-	-	-	-

Figure 3.4.2.1 Multiple Listing District (MLS) districts in Toronto.
Note: *For reference, the city was divided into 35 MLS districts: 11 districts in the East (E01-E11), 14 districts in the Center (C01-C04, C06-C15), and 10 districts in the West (W01-W10) (Source: Toronto Real Estate Board, 2011).*

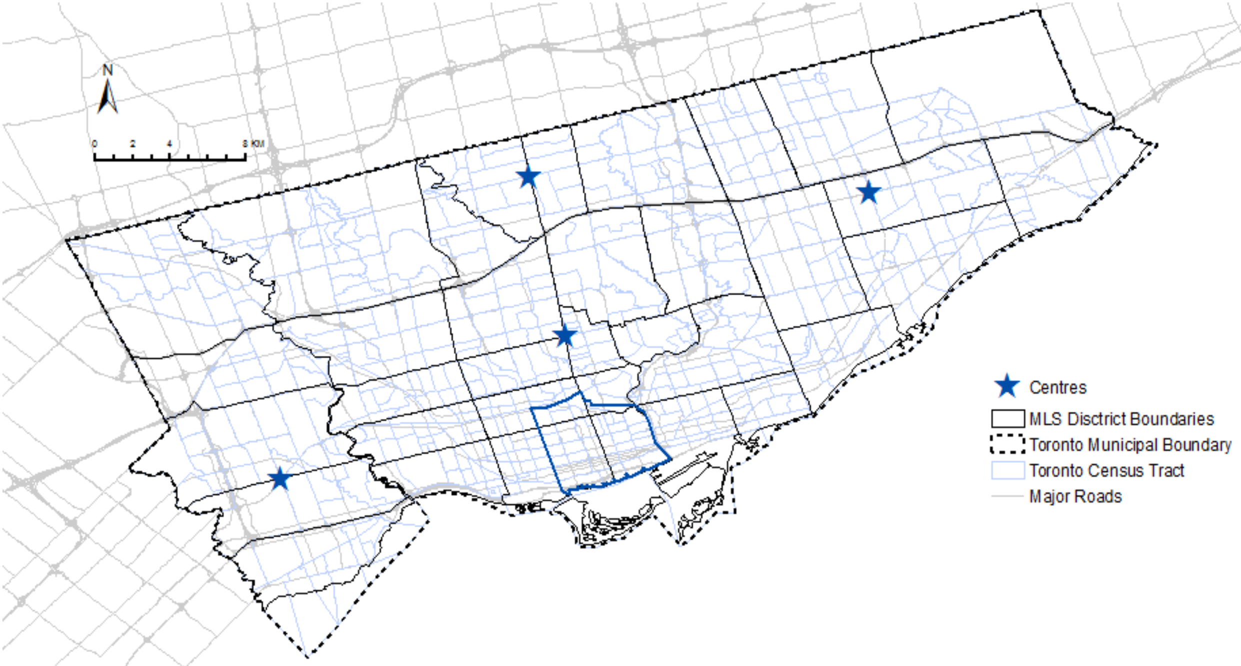

Figure 3.4.2.2 Illustration - boundaries of MLS districts and census tracts.

3.3.3. Household Variables

The family and households dimension of socio-economic characteristics of neighborhoods were hypothesized to have an impact on house prices. The first set of household variables measures the average physical housing attributes: average number of rooms or bedrooms; maintaining status of dwellings (e.g. needing minor or major repair). Since information about the age of dwellings was not provided, an alternative variable was generated by measuring percentage of dwellings built before 1946, based on data availability.

The second set of household variables measures the average family status and the proportion of household types (e.g. married with or without children, nonfamily, one family or multi-family households) of neighborhoods. Descriptive statistics of all household variables are provided in Table 3.4.3.1. All variables were extracted from Census *families, household and marital status data* (Statistics Canada, 2006) and were obtained from the University of Waterloo Geospatial Center.

Table 3.3.3. Descriptive statistics for household variables.

Household Variables	Description	Min.	Max.	Mean	Std. Dev.
Average number of rooms per dwelling*	--	3.2	10.3	5.75	1.20
Average number of bedrooms per dwelling*	--	1	4.3	2.4	0.60
Percentage of occupied private dwellings need major repair	Number of dwellings need major repair/total number of dwellings	0	0.28	0.08	0.04
Percentage of dwellings built before 1946	Number of dwellings built before 1946/ total number of dwellings	0	0.86	0.20	0.25
Percentage of nonfamily households	Number of nonfamily households/total number of households	0.04	0.77	0.32	0.14
Percentage of one family households	Number of one family households/ total number of households	0.24	0.86	0.64	0.12
Percentage of multi-family households	Number of multi-family households/ total number of household	0	0.18	0.04	0.03
Percentage of married couples with children at home	Number of married couples with children at home/ all census families in private households	0.11	0.66	0.43	0.10
Percentage of married couples without children at home	Number of married couple without children at home/ all census families in private households	0.07	0.61	0.26	0.07
Percentage of residents over 65	Number of residents over 65/total number of residents	0.02	0.32	0.14	0.05

[1.] Census tract area measured in km^2 * Data (Variables) extracted directly from Statistics Canada (2006), no calculation involved.

3.3.4 Neighborhood Characteristic Variables

3.3.4.1. Crime Variables.

Crime data were retrieved from 2006 incident-based Uniform Crime Reporting Survey (UCR) (Statistics Canada, 2006). The UCR Survey collects data on the number of incidents and their characteristics. Because only crimes reported to the police are included, the UCR is not a complete record of all crimes in Canada (Statistics Canada, 2006). For each census tract, crime counts between zero and five were round up to five for confidentiality reason. The incident counts of each crime type were converted into crime rates for analysis. The sum of residential and working population is used as the denominator. The crime data were obtained from the University of Waterloo Geospatial Center.

Instead of using a total crime index, crimes were disaggregated into their fifteen component crimes, in order to address the impact of different crime types on house values. Crime types were grouped into property crimes and non-property crimes. Property crimes include stealing and destroying property. In both cases, the property crime does not involve force or threat against a victim (Employment and Social Development Canada, 2015). Based on data provided by UCR, property crimes examined in this study include seven types: property crime (in total), theft of or from a vehicle, break and enter, mischief, shoplifting and other thefts. Table 3.4.2.1 presents descriptive statistics of each crime type. Mischief is classified as property crime according to Justice Laws (Government of Canada, 2014), where the property is destroyed or damaged in a dangerous, useless or inoperative status.

The rest of the UCR crimes are therefore classified as non-property crimes, including violent crime, robbery, sexual assault, uttering threats, minor assault, major assault, criminal harassment and drug offenses. Table 3.4.2.2 presents descriptive statistics for each of the non-property crime. Robbery is not classified as property crime, despite the fact that it involves taking someone's property. The main reason is that the crime involves force or threat against a victim, while all defined property crimes in this study does not.

For a better understanding of the crime categorization adopted in this study, it is worth noting that crimes can also be classified into three main categories base on offenders' bias motivation: expressive crime, acquisitive crime and other types of crimes. Expressive crimes are often motivated by expression of emotions and some of them are motivated by religious bias, racial bias or sexual-orientation bias. Acquisitive crimes, on the other hand, are motivated to reach tangible goals, such as obtaining physical goods (Cohn & Rotton, 2003).

All of the defined property crimes in this thesis are acquisitive crimes – primarily to obtain physical goods rather than to express emotions. However, the defined non-property crimes can be either expressive or acquisitive in nature. Robbery, for example, is an acquisitive crime since the motivation is to obtain tangible goods, but it involves violence against a victim, as noted previously. Criminal harassment and drug offenses are two crime types that do not exhibit characters of acquisitive or expressive crimes.

Table 3.4.4.1. Descriptive statistics for crime variables

	Crime Incident (count)		Crime rate (per 1,000 population at risk)			
	Total	**Mean**	**Min.**	**Max**	**Mean**	**Std. Dev.**
Property Crimes						
Property crime	74, 852	142.85	4.35	59.73	18.82	9.57
Theft from a motor vehicle	15, 663	29.89	0	18.99	4.00	2.61
Theft of a motor vehicle	5, 806	11.08	0	11.77	1.47	1.10
Break and Enter	11, 557	22.06	0	9.45	3.22	1.73
Mischief	14, 389	27.46	0	17.60	4.03	2.24
Shoplifting	9, 053	17.28	0	36.48	1.79	4.41
Other theft	19, 950	38.07	0	23.65	4.74	3.16
Non-property Crimes						
Violent crime	25, 985	49.59	0	28.26	7.23	4.25
Robbery	4, 204	8.02	0	5.95	1.21	1.02
Sexual assault	1, 1116	22.13	0	287.7	1.00	12.56
Uttering threats	5, 497	10.50	0	690.45	3.68	30.15
Minor assault	11, 648	22.23	0	89.42	5.03	5.99
Major assault	3, 976	7.59	0	26.27	1.73	2.23
Criminal harassment	1, 688	3.22	0	3.48	0.49	0.52
Drug offences	2, 942	5.62	0	13.87	0.81	1.17

3.3.4.2 Employment, Income, Ethnicity Variables

The occupation and income dimension of socio-economic characteristics of neighborhoods were hypothesized to influence house prices, as justified by past research in Chapter 2 of this study. The education variable is operationalized as percentage of residents aged 25 to 64 who hold a Bachelor's degree. Occupation variables include: location quotient of professional jobs (in natural, social, educational science), location quotient of business and administrative jobs, location quotient of manufacturing jobs, and unemployment rate of each census tract.

Location quotient is a ratio that quantifies the concentration of one factor in comparison to the concentration of the same factor in a larger reference context. For instance, location quotient of professional jobs (as "professionals" in the equation below) equals to the proportion of professional jobs in each census tract divided by the proportion of professional jobs in the city of Toronto. Detailed discussion of location quotient can be seen in *Section 5.5* of this thesis. Income variables include medium income, average income[1], composition of family income from government transfer payment, and percentage of low income families.

Location quotient of professionals = $\frac{(\text{number of professionals per CT})/(total\ number\ of\ jobs\ per\ CT)}{(\text{number of professionals in Toronto})/(Total\ number\ of\ jobs\ in\ Toronto)}$

Transfer payment in Canada refers to a redistribution of income to equalize social welfare. The low income threshold is defined by Statistics Canada (2006) as more than 20 percentage more of family income spent on food, shelter and clothing than the average family. Table 3.4.4 presents the descriptive statistics for occupation, education and income variables. All variables were extracted from Canadian Census *labour, education, occupation and income data* (Statistics Canada, 2006) and were obtained from the University of Waterloo Geospatial Center.

Ethnic origin and immigrant characteristics of neighborhoods are justified by previous studies to impact house prices in gateway cities like Toronto and Vancouver. Those variables include index of ethnic heterogeneity, percentage of aboriginal people, percentage of Caucasians, and percentage of visible minorities. The denominator for calculating the percentage is the population of each census tract. The index of ethnic heterogeneity is measured as '1' subtracted by the sum of squared ethnic proportions (Hirschfield & Bowers, 1997; Quick, 2013). All variables were extracted from Canadian Census ethnic origin and immigration status (Statistics Canada, 2006) and were obtained from University of Waterloo Geospatial Center

Index of ethnic heterogeneity = $1 - \sum Wi^2$ (where Wi is the proportion of residents of in ethnic group i for each census tract)

[1] Univariate and bivariate regression will be conducted to determine which income variable to retain, and will be discussed in Chapter 5. This further variable selection also applies to other neighborhood variables measuring similar aspects.

Table 3.4.4.2 Descriptive statistics for employment, income, ethnicity variables

	Min.	Max.	Mean	Std. Dev.
Employment variables				
Location quotient of professionals	0.28	2.11	1.00	0.33
Location quotient of business and administrative jobs	0.49	1.88	0.99	0.19
Location quotient of management jobs	0.11	3.04	1.01	0.53
Location quotient of manufacturing jobs	0	4.74	0.98	0.85
Unemployment rate	0	18.8	7.60	2.70
Percentage of residents aged 25-64 who hold Bachelor's degree	0.03	0.46	0.22	0.09
Income variables				
Average income	14,788	314,107	42,568	32591
Median income	12,078	65,269	26994	97777
Composition of family income from government transfer payment (%) for all economic families	0.5	34.5	11.07	5.78
Prevalence of low income (before tax) families (2005)	0	69	19.69	10.86
Ethnicity				
Index of ethnic heterogeneity	-0.82	0.92	0.14	0.36
Percentage of aboriginal people	0	0.07	0.01	0.01
Percentage of Caucasian	0.04	1.22	0.43	0.25
Percentage of visible minorities	0.02	0.96	0.44	0.25

3.4.4.3 Dwelling Density and Other Variables

Dwelling density characteristics of neighborhoods are hypothesized to be associated with house prices in many studies (Song & Knaap, 2004; Cho, Poudyal & Roberts, 2008). Since Toronto is densely populated and is under excessive development pressure, the city has a great number of apartment buildings. Thus, densities of apartments are included apart from density of detached houses. Total dwelling density of all types of dwellings above is also included.

Table 3.4.4.3 Descriptive statistics for density and other neighborhood characteristic variables

Dwelling density	Number of total dwellings/census tract area	25.86	29,695	2825	3080
Detached house density	Number of total detached houses/ census tract area	0	1956.90	485.48	360.48
Apartment (duplex) density	Number of apartment duplex units/census tract area	0	628.21	95.82	99.47

Apartment (building with 5 stories or more) density	Number of apartment units/census tract area	0	29173.91	1392.47	1392.47
Neighborhood stores (binary)		0	1	0.39	0.49
Ridership of subway stations	Average passengers of subway stations within each census tract	0	179,910	2848	12587

The presence of neighborhood (or convenient) stores is examined in some studies for its influence on house prices (Song & Knaap, 2004). Neighborhood stores are distinguished from local or regional commercial stores. A value of '1' is assigned to census tract with presence of neighborhood stores and a value of '0' is assigned to census tracts with no neighborhood stores. The variable was extracted from geographic and attribute information of Toronto address points (City of Toronto, 2014) and obtained from Toronto Open Data website.

Last but not least, ridership of subway stations measured by 'average passengers of subway stations' is included to examine the transit impact on house prices. The data originate from a 2006 yearly ridership for Toronto's four existing subway lines (including Yonge-University Line, Bloor-Danforth Line, Scarborough Line, Sheppard Line). It is argued that land use patterns, such as density and diversity, are closely related to transit ridership rather than transit adjacency (Sung & Oh, 2011; Cervero, 1993). This variable therefore also reflect the degree of centralization of each census tract. The greater the number of passengers, the more centralized or urbanized the census tract is. (Figure 3.4.6 is a map created based on ridership of each subway stations). A value of '0' is assigned to those census tracts with no subway stations inside of their boundaries. The variable was extracted from Streetfiles of major roads (DMTI Sptail Inc., 2006) and was obtained from University of Waterloo Geospatial Centre.

Figure 3.4.6 Average Passengers per Subway Station (yearly ridership) (TTC, 2006).

Chapter 4.
Preliminary Data Analysis

Chapter Overview

This chapter deals with a preliminary data analysis of the spatial dynamics of Toronto housing market. It centered on questions: Do Toronto house prices exhibit spatial patterns and what information can the patterns reveal? To answer these questions, exploratory spatial data analysis (ESDA) namely cluster analysis was conducted, in particular mapping and Moran's I techniques (e.g. LISA statistics). The results of the cluster analysis indicate a significant spatial clustering of high-value and low-value housing in Toronto. Implications of the clusters are analyzed with regard to the urban structure and development patterns of the study area.

4.1 Spatial Autocorrelation and Cluster Analysis

This section first discusses spatial autocorrelated phenomenon in social science and the differences between exploratory spatial data analysis (ESDA) and confirmatory spatial data analysis (CSDA). The second part deals with cluster analysis methods of ESDA, in particular specifications of global and local Moran's I.

4.2.1 Spatial Autocorrelation

Typically, when most social or economic phenomena are mapped, proximity in locations results in similarity in values. High values tend to be co-located with similarly high values, and low values with low values, which means exhibiting positive spatial autocorrelation (Voss, White & Hammer, 2006). This spatial phenomenon can be explained by the "first law of geography" that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p.236).

Studies of spatially autocorrelated phenomenon are widespread in social science. Regional voting clusters is one of the popular spatial analysis applications in North America, where voters' economic and ethnic background are considered to be associated with their political behaviors (West, 2005). Another example is studies of interdependent decision-making of central banks. It is widely accepted that policies made by central banks are constrained by their local contexts. Thus, studies using spatial analysis are examining how independent the decisions of central banks are from local authorities and among each other (Ward & Gleditsch, 2007). Other relevant studies include the spillover effects of pollution, as well as distribution of wealth and inequality.

Do house prices of Toronto exhibit spatial autocorrelation? If so, why do the patterns originate and what are the factors contribute to the patterns? To answer the questions, studies usually begin with exploratory spatial data analysis (ESDA). It visualizes spatial distribution patterns as well as trends, without examining why the patterns originate. Examining the "why" in the following question involves confirmatory spatial data analysis (CSDA), which tests relationships between outcome and explanatory variables. ESDA and CSDA are therefore two essential parts of spatial analysis. In this sense, ESDA is an invaluable starting point for the housing price analysis of this thesis.

4.2.2 Cluster Analysis and Moran's I

Cluster analysis, as a spatial technique of ESDA, is commonly used to "visualize spatial distributions, identify atypical locations or hotspots, and suggest spatial regimes or other forms of spatial heterogeneity" (Anselin et al., 2001). It can be applied to fields such as spatial epidemiology, spatial demographics, landscape ecology, and crime analysis to quantify geographic variation patterns. Common techniques involve creating and interpreting maps using geographic information systems (GIS).

Global and local clustering methods are two broad branches of spatial cluster analysis. Global diagnostics is the overall description or analysis across the data for the entire study area. A typical example of global clustering technique is global Moran's I. Local measurements, on the other hand, can be used for understanding clusters in a localized extent. Instead of generating a single set of global parameters, local analysis produces statistics corresponding with small-scale neighborhoods.

Global Moran's I measures the general extent of spatial distribution, but cannot by itself identify the exact spots of the clusters within the study area. Global Moran's I results range from a scale of negative one to one (-1, 1), with negative one indicating negative spatial autocorrelation (dispersion, dissimilar values among neighbors), zero indicating spatial randomness, and one indicating positive spatial autocorrelation (clustering, similar values among neighbors). When either negative one or one is present, the distribution of the variable is assumed to exhibit clustering (Anselin, 1995).

4.2 Global and Local Moran's I Results

In the spatial analysis of 2006 Toronto house prices, both global and local Moran's I were conducted by employing two useful ESDA techniques: Moran Scatterplot and LISA statistics (Local Indicators of Spatial Association, namely local Moran's I), as briefly introduced in the previous section. Results from both the global and local Moran's I analysis are shown below in tables and figures. Interpretation of the results as well as its indication to urban planning and Toronto's housing market are discussed. The cluster analysis in this section also prepares for the regression analysis in *Chapter 5*.

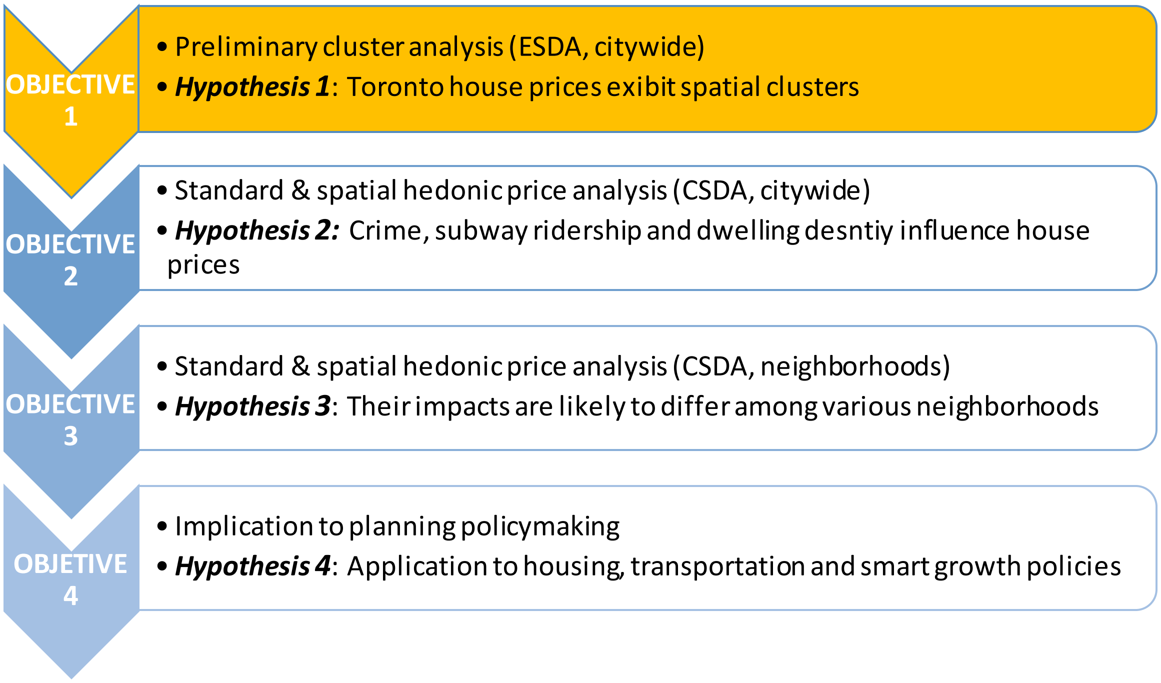

Figure 4.2.1 Objective 1 and Section 4.3 in Context

Global Moran's I was conducted using the original house price data, where MLS is the unit of aggregation. Results indicated that 2006 house price in Toronto exhibited positive global spatial autocorrelation. Table 4.3.1 below shows the parameters of the analysis. The Moran's I value is positive, falling within the interval 0-1, where the p-value indicates that the spatial autocorrelation is significant (p=0.001). This means, in plain language, the overall distribution of 2006 house price across Toronto is not randomly distributed. The Moran's I index over the entire study area were computed based on a spatial weight matrix (first-order queen). Both global and local Moran's I were conducted using GeoDa spatial analysis software.

Table 4.2.1 Global Moran's I results for house prices among 35 MLS districts in Toronto (2006)

Moran's I value	E[I]	p-value	z-value	St. Deviation
0.48	-0.029	0.001	4.53	0.11

Notes: *The result is based on permutations of 999.*

Local Moran's I was conducted by two techniques: Moran scatterplot and LISA map. Figure 4.3.1 presents the Moran scatterplot for the 2006 Toronto logged house prices (outcome variable). The data were standardized so that the distributional pattern of the scatterplot represents the standard deviation from the mean. The horizontal axis shows the standardized value of the logged house prices for each census tract, while the vertical axis shows the standardized value of the average logged house price for each district's neighboring districts. The neighbors were defined based on the "first order queen" convention, meaning that the neighbors for any district "A" are other districts that shares a common boundary.

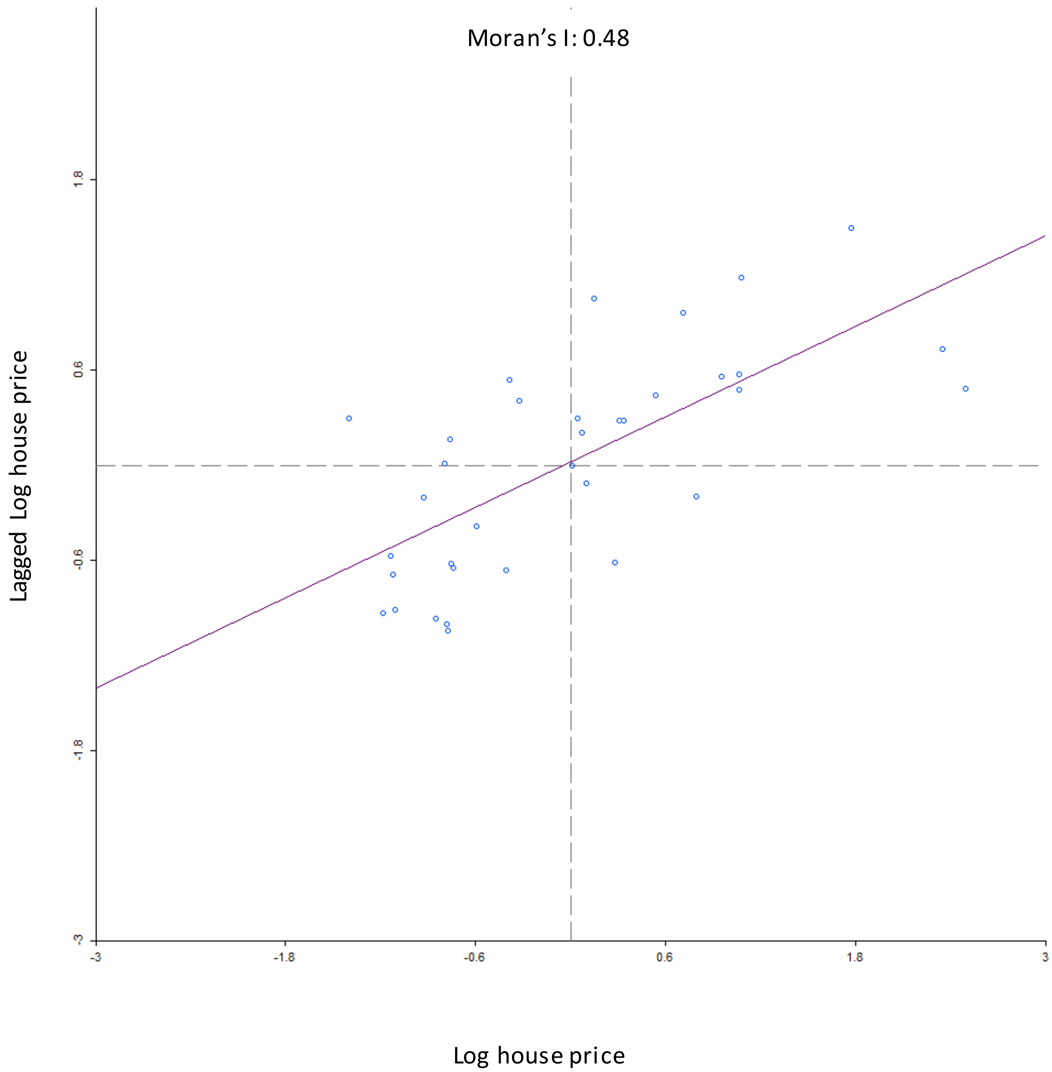

Figure 4.2.2 Moran's Scatterplot of house prices among 35 MLS districts in Toronto (2006)

In the Moran scatterplot (Figure 4.2.1), the upper right quadrant represents those MLS districts with above average house price, with adjacent MLS districts also having above average house price (high-high). The lower left quadrant represents the districts with below average house prices, which are also surrounded by MLS districts with below city average house prices (low-low). The upper left quadrat has

MLS districts with below average house prices, with neighbors that have above average house prices (low-high), and the lower right quadrant has the reverse (high-low). The slope of the line through all the points in Figure 4.2.1 expresses the global Moran's I value (Anselin, 1996). In this house price analysis example, the Moran's I value is 0.48 (also presented in Table 4.3.1). This statistic shows a strongly positive spatial autocorrelation (spatial cluster of similar values). As shown in the scatterplot, most MLS districts (points) can be found in the high-high and low-low quadrants.

Another technique of local Moran's I - LISA cluster map shows where in the city of Toronto the high-high and low-low neighborhoods were located. To denote the analysis and discussion in future chapters of this thesis, and to be consistent with the all the mappings, house prices were assigned to census tracts based on the MLS districts they fall within, since census tracts were the level where most data involved were available. Details of the aggregation can be found in *Section 3.4.1* of this thesis.

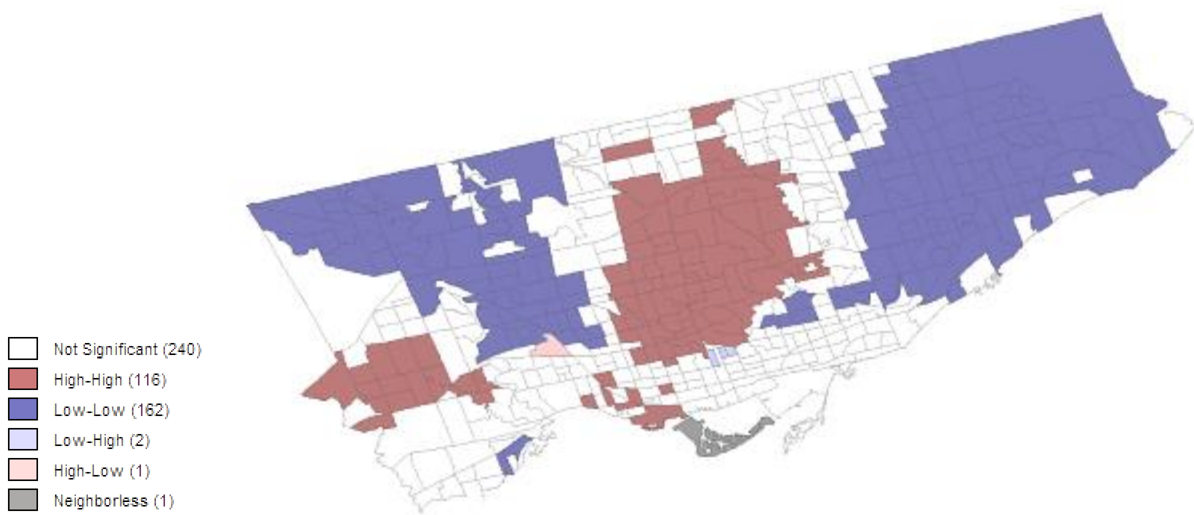

Figure 4.2.3 LISA cluster map of 2006 house prices by census tract level
Note: The results was based on 999 permutations; and meanings of "high-high, low-low, low-high and high-low" are explained in the text above.

Census tracts with insignificant local Moran statistic are not shaded in the map. Neighborless area is an island with no neighboring census tracts to be included in the analysis. Clusters of high-value house census tracts surrounded by high-value house tracts (high-high) are mostly located in the middle of the city (Yonge-Eglinton area), with some in the Southeast (Etobicoke area and East of downtown Toronto). Clusters of low-value house census tracts surrounded by low-value house census tracts (low-low) are located in the Northeast (North York) and Northwest (Scarborough) of the city.

4.3 Multiple Centers, Subway Lines and House Prices

Apart from the Moran's I analysis, an informative quantile map (Figure 4.3.1.3) was created to enhance the previous results regarding the spatial dynamics of Toronto's housing market. Cartographic displays, or visualization and mapping, useful in exploratory models, can reveal structure in the dataset that may not be readily available from tabulation. The map below divided the (logged) house price into six depicted quantile classes, with different shades of colors reflecting the spatial variation and clustering of the house prices in Toronto. Important geographic features with regard to Toronto's urban structure such as *Downtown Toronto, Centres* and subway lines were identified in the map. This visualization of various layers of spatial information is helpful for interpreting observed housing market dynamics. The map was generated by using Geographic Information System (GIS) software ArcMap 10.2.

Visually observed clustering of high-value houses in the study area (Figure 4.3.2.1) can be mostly spotted in the midtown, along the city's two subway lines and around three of the four *Centres* identified in the City of Toronto Official Plan. The three *Centres* are: *Etobicoke Centre* in the West, *North York Centre* in the North, and *Yonge-Eglinton Centre* in the middle of the city. The official plan defines the *Centres* as concentrated mixded-use development (e.g. jobs, housing and services mixed in a dynamical setting) and key locations on the rapid transit system. They are also four focal points of Toronto's urban structure, representing decades of planning policy and substantial i nvestment.

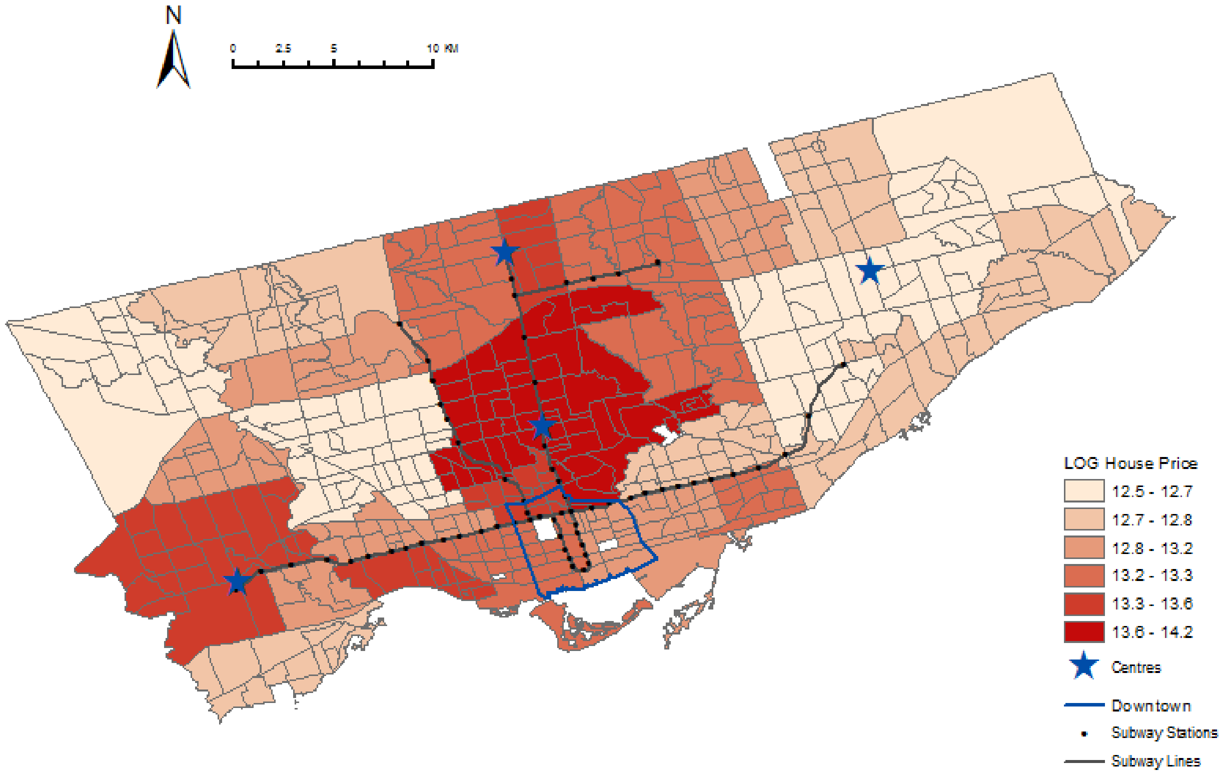

Figure 4.3.1 Quantile map of 2006 Toronto house prices by census tract.
Note: *'Centres' refers to four Centers across the city identified in the Toronto Official Plan 2006 & 2010.*

Each *Centre* is unique in terms of demographic composition and growth potential. *Scarborough Centre* is the only *Centre* not significantly clustered with high-value residential development. One possible reason is that Scarborough Centre functions as the major gateway within the city while *Centres* such as Yonge-Eglinton has greater development potential for urban residential infill projects. Also, growth potential (e.g. expected future rent increase) among other components for urban land price (e.g. the value of accessibility, cost of converting land uses) can account for more than half of the land price in rapidly growing cities like Toronto. The greater the growth premium, the higher the land price (Capozza & Helsley, 1989).

This visually observed pattern of house price distribution suggests an emerging urban development pattern – nodal development. Instead of having a single central business district (CBD), the City of Toronto has multiple centres as key nodes on the transit system (Figure 4.3.2.2) that also function as employment sub-centres. In traditional mono-centric urban theories, land values peak at the CBD and "better-off" population tend to live at the city periphery for lower land price and large r land parcels (Champion, 2001). In multi-centric urban structure, however, house prices may not necessarily decline with distance from the CBD and land values will have less variation, and the greater the number of centres, the flatter the land-value surface. As indicated in Figure 4.3.2.1, nodes (or sub-centres) are more influential than the CBD (downtown Toronto) itself in terms of house price distribution.

It is worth noting that although high value houses were found to be clustered around employment centres, residents may not necessarily work in the employment centre nearest to their neighborhoods. Further, not every employee can afford to live in the neighborhood where they are employed (Cervero, 1996), especially in cities like Toronto where living expenses are high. For example, residents who live above the ground floor of a mixed-use development are likely to be the consumers of the café, florists and dry cleaning businesses that located on the ground floor. Employees of those businesses, however, have to commute to a cheaper places that they can afford to live. Residents of those places possibly still have to commute to another office node or their high-tech firms in a business park at the urban periphery. Therefore, although location with respect to employment is argued to be a significant predictor of house prices (Ottensmann, payton & Man, 2008), the map does not necessarily imply the impact of employment accessibility on house prices.

The spatial patterns identified by this cluster analysis happen to be consistent with the findings by Hulchanski (2007, 2010) in his *Three Cities within Toronto*, where the city's neighborhoods fall into three distinct clusters based on their average income levels (Figure 4.3.2). The high-income neighborhoods were mostly located near the city centre and close to the two subway lines, especially near the waterfront, south of Bloor Street and Danforth Avenue, as well as central Etobicoke. Those areas are the high-value house clusters in our analysis. The low-income neighborhoods were mainly located in the northeast and northwest of the city, the area of which are the identified low-value house clusters. The middle income neighborhoods (as well as the medium-value house clusters) were located in between the neighborhoods. This consistency has implications for further confirmatory spatial data analysis (ESDA) in Chapter 5, where regression analysis were conducted in separate neighborhoods.

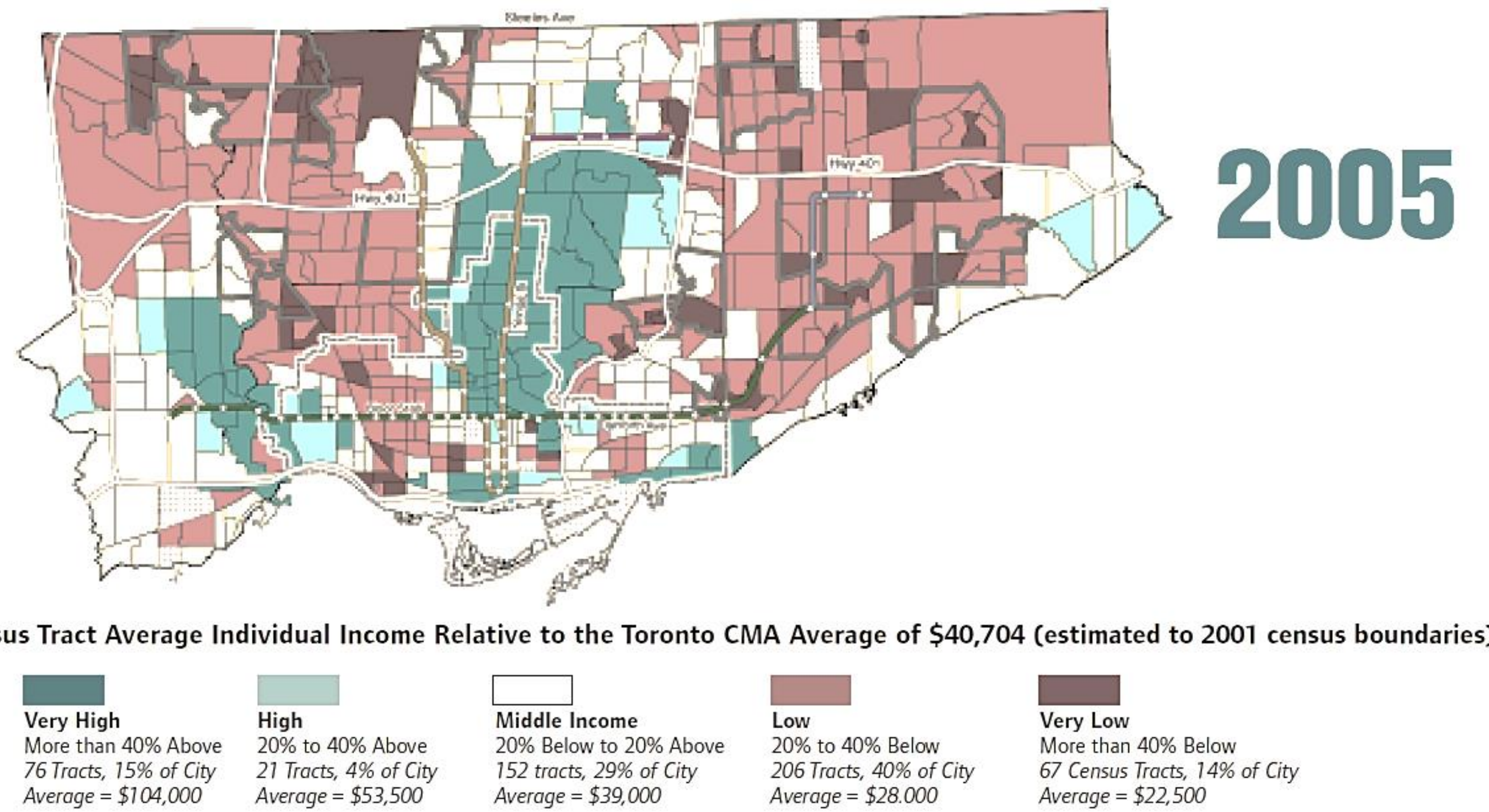

Figure 4.3.2 Average individual income by census tract (Toronto, 2005)
Source: Hulchanski, J.D. (2010). *The Three Cities Within Toronto.* pp.5. Map3.

Chapter 5.

Regression Modeling Approach & Results

Chapter Overview

This chapter focuses on the modeling strategies and regression results. In particular, it discusses the steps involved in the regression modeling and the process of selecting the best fitting model for the dataset. The best fitting model was first applied to all census tracts of the city. Then the same model was applied separately to three neighborhoods, defined by census tracts' average income levels. Results indicate that except for "theft of a vehicle", most crime types did not seem to be significantly associated with house prices across the entire city, or in the low- and high- income neighborhoods within the city. However, when tested in the middle-income neighborhood, six types of crime had significant impacts on house prices. Transit ridership is positively associated with house prices in the citywide estimations, but not significant in neighborhood estimations. In the middle income neighborhoods, dwelling density in general is positively associated with house prices while apartment density has a negative impact on house prices.

5.1 Spatial vs. Non-Spatial Regression Models

In the previous chapter of this thesis, results of preliminary data analysis (the ESDA) demonstrate that house prices at the MLS level in Toronto exhibits spatial clusters, where some areas are clustered by higher-value houses and others are clustered by lower-value ones. The "first law of geography" that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p.236) can explain this phenomenon as noted previously, but it does not account for why spatial clusters arise in practice. Or, in our case of housing prices, this short reminder cannot give a satisfying explanation to the local variation of house prices, such as why some neighborhoods are clustered with high-value houses.

The exploration of the local variation of house price leads to regression modelling, or the so-called CSDA (as opposed to ESDA) that examining the relationship between outcome variables and explanatory variables. Traditional property value research is often conducted in standard hedonic price models, based on the assumption that the relationship between house prices and crime is consistent across the city. However, when a large number of spatial units (e.g. census tracts) are involved in the dataset, especially in this case when house prices in Toronto exhibits non-random spatial distribution, standard regression approaches cannot properly incorporate aspects of space (e.g. proximity or spatial interaction) into their models, where location is part of the reason for a phenomenon.

Spatial models, recognizing the influence of geography, can improve the estimation by providing parameters less subject to statistical bias and inconsistency in a spatially structured dataset (Voss et al., 2006). For instance, if house prices in one neighborhood is similar to that of a nearby neighborhood, standard hedonic models can only capture the direct effects of physical features and neighborhood characteristics, but fail to capture indirect effects on house prices from nearby neighborhood or unobserved spatial variation (Gibbons & Machin, 2008; Cohen & Coughlin, 2008). This has tremendous implication to studies of property market.

5.2 Review of Regression Models

In this section, regression models including ordinary least squares (OLS), spatial lag dependent, spatial lag independent models, and spatial error models are briefly reviewed, with a focus on their basic functional forms. Standard and spatial hedonic price functions are also reviewed. Traditional hedonic price model is usually estimated with OLS. Spatial hedonic price models are divided into three subgroups of spatial models, based on how the spatial autocorrelation is expected to occur and by combining the basic forms of standard hedonic function and spatial regression models.

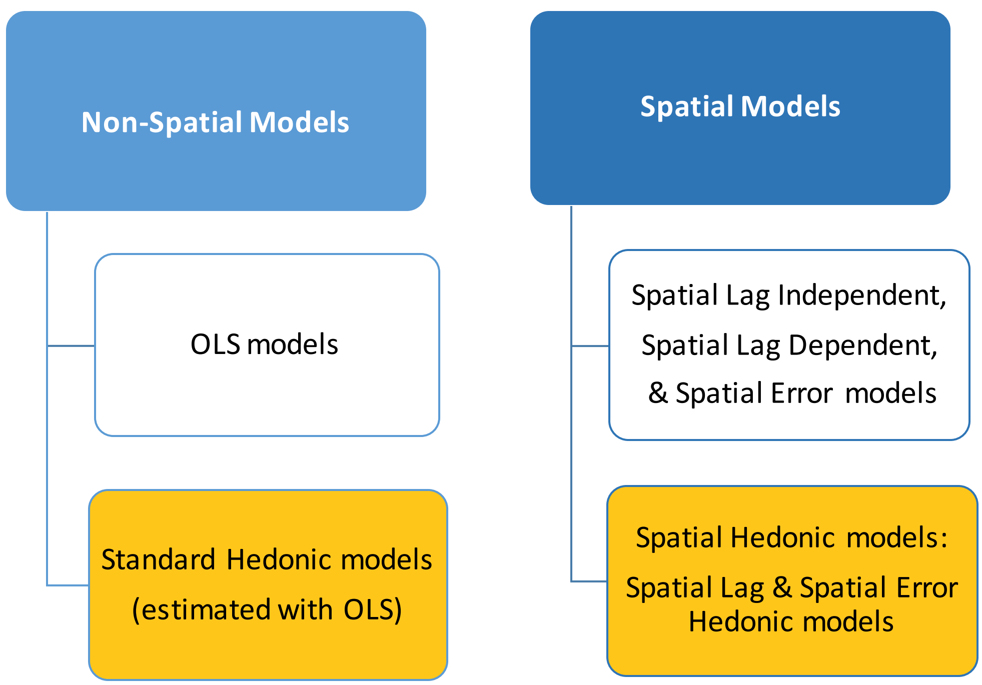

Figure 5.2 Regression Model Categories.

5.2.1 Ordinary Least Squares (OLS) Regression Models

Ordinary least squares (OLS) or linear least squares is a statistical method for estimating a linear relationship between explanatory and outcome variables. A univariate OLS means only one explanatory variable is involved, which takes the form of Equation 5.1.1 (a). A multivariate OLS regression expands univariate OLS by including more explanatory variables, which takes the form of Equation 5.1.1 (b).

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \qquad 5.2.1\ (a)$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \qquad 5.2.1\ (b)$$

In the case of housing price studies, y is the outcome variable - house price; β_0 is the regression constant; β_i is the regression coefficient for each explanatory variable x_i $(i = 1. 2, 3 \ldots)$ such as crime rate and household characteristics, and ε represents the regression error. A multivariate regression model can therefore be used to account for the relationship between house prices and crime rate, dwelling density, transit accessibility and other neighborhood characteristics such as ethnicity composition.

5.2.2 Standard Hedonic Price Models

Traditional hedonic price function is usually estimated using OLS. It allows estimation of the values of specific features by regressing property price with various attributes of the property as independent variables in a standard regression model. Common forms of hedonic housing price models can be seen in Equation 5.2.2 (a) or Equation 5.2.2 (b), where the dependent variable was specified as logged house prices (P), H is a matrix of housing characteristics, N is a matrix of neighborhood characteristics, and L is a matrix of locational characteristics. The β_0 is the constant, β_H, β_N, and β_L and corresponding parameters, and ε is error terms. Neighborhood and locational characteristics are defined together in some studies, since the two features are interdependent and neighborhood characteristics can be associated with a specific location (Mahan, Polasky & Adams, 2000; Cho et. al., 2008).

$$P = \beta_0 + \beta_H H + \beta_N N + \beta_L L + \varepsilon \qquad 5.2.2\ (a)$$

$$In\ (house\ price) = f\ (H, N, L) \qquad 5.2.2\ (b)$$

As noted in Chapter 2, choice of functional form is one problem associated with hedonic models. The semi-log model with the natural log of housing prices is a commonly used function and can reduce the heteroskedasticity problem involved in the hedonic price modelling (Cropper et. al., 1988; Song & knaap, 2004).

5.2.3 Spatial Regression Models

To capture the spatial variation, there are three commonly used spatial regression models: spatial lag dependent, spatial lag independent and spatial error models, which are briefly reviewed in the following three sections. Spatial effects were incorporated by creating a weight matrix, as noted previously in Moran's I analysis.

5.2.3.1 Spatial Lag Dependent Regression Models

Spatially lagged form of the outcome variable is included as an explanatory variable in a spatial lag dependent regression model, of which the structure can be seen in Equation 5.2.3.1, where y is a vector of observations on the dependent variable (e.g. house price) and the ρ is spatial autoregressive coefficient (the spatial lag term to be estimated), Wy is the spatially lagged dependent variable for weight matrix W, X is a matrix of observations on independent (explanatory) variables, β is the regression coefficient, and ε is random error terms (Anselin, 2005).

$$y = \rho Wy + X\beta + \varepsilon \qquad 5.2.3.1$$

In contrast to the spatial error regression model, which deals with spatial structure via an error term and considers spatial structure as nuisance, the spatial lag regression model adds explanatory terms to account for the spatial pattern (Zhukov, 2010). It is worth noting that the spatial lag dependent regression model is not suitable for this thesis, where the dependent variable (house prices) is aggregated from MLS level to census tracts. This means certain adjacent census tracts falling within the same MLS district would have the same house prices, which violates the underlying hypothesis in spatial lag dependent models that house prices are impacted by nearby house prices.

5.2.3.2 Spatial Lag Independent Regression Models

By adding a spatially lagged explanatory variable to a linear regression model, the spatial lag independent model can be seen in Equation 5.2.3.2, where y is a vector of observations on the dependent variable (e.g. house prices), X is a matrix of observations on independent (explanatory) variables, β is the regression coefficient, and the p is spatial autoregressive coefficient (the spatial lag term to be estimated). Wx is the spatially lagged independent variable for weight matrix W and ε is random error terms (Anselin, 2005).

$$y = X\beta + pWX + \varepsilon \qquad 5.2.3.2$$

In contrast to the spatial lag dependent model, spatial lag independent regression model takes into account of the spatial effects of independent variables from neighbors on the dependent variables. For instance, to estimate the house price of census tract A, the impacts of crime rate, income level or ethnic composition from census tracts adjacent to A are considered in the model. The definition of neighbors is determined by the spatial weight matrix.

5.2.3.3 Spatial Error Regression Models

Spatial error model adds a spatially lagged error term to a linear regression model. The error is modeled as a simultaneous spatial autoregressive model (Anselin et al., 2001). A spatial error regression model can be seen in Equation 5.2.3.2, where y is a vector of observations on the dependent (outcome) variable (e.g. house prices), X is a matrix of observations on the independent variable (explanatory) variables, β is the regression coefficient, ε is the spatially autocorrelated error terms, $W\varepsilon$ is the spatial weights matrix, λ is the autoregressive coefficient (the spatial error term to be estimated) and ξ is the normal distribution with mean and variables (Anselin, 2005).

$$y = X\beta + \varepsilon \quad (\textit{where } \varepsilon = \lambda W\varepsilon + \xi) \qquad 5.2.3.3$$

The spatial error model considers spatial effects as a nuisance and calculates the regression error from neighboring census tracts. It addresses the presence of unidentified explanatory variables and omitted variable bias (Zhukov, 2010). The spatial error model can also be applied when the neighboring observations are similar because of stimuli on a larger scale than the geographic unit of analysis (Fowler, 2011).

5.2.4 Spatial Hedonic Price Models

Spatial hedonic price model incorporates spatial effects by combining the basic forms of the standard hedonic model and the spatial regression models. Spatial hedonic models can therefore be divided into three types, depending on how the autocorrelation is expected occur. The spatial lag dependent hedonic model assumes inherent house price autocorrelation (Equation 5.2.4.1). The spatial lag independent hedonic model assumes the autocorrelation of the attributes of property and neighborhoods influence house prices (Equation 5.2.4.2). The spatial error hedonic model adds an error term to the standard hedonic function (Equation 5.2.4.3).

$$P = \beta_0 + \beta X + \rho WP + \varepsilon \qquad 5.2.4.1$$

$$P = \beta_0 + \beta X + \rho WX + \varepsilon \qquad 5.2.4.2$$

$$P = \beta_0 + \beta X + \varepsilon, \textit{ (where } \varepsilon = \lambda W\varepsilon + \xi) \qquad 5.2.4.3$$

Similar to Equations 5.2.2, P stands for a vector of logged house prices, X is a matrix of housing characteristics (H), neighborhood characteristics (N), and locational characteristics (L). The β_0 is the constant, β is a vector of corresponding parameters, and ε is the spatially lagged error terms. Spatial and standard (non-spatial) hedonic price models are employed to estimate the impacts of different crime, transit and dwelling density on house prices in the following sections.

5.3 Modelling Approach: Variable Selection and Model Specifications

This part discusses in details the steps of the regression modelling approach developed in this thesis. It is worth noting that, in an effort to include the impact of crime on house prices, spatial and non-spatial hedonic models were created for each component crime type (rather than a total crime index), along with other variables of neighborhood characteristics (e.g. accessibility or density).

Section 5.3.1 describes the univariate regression analysis and bivariate correlation test as the initial step of variable selection to determine what explanatory variables to include in the final model. In *Section 5.3.2*, model selecting decision process was discussed, with a focus on using regression diagnostics to justifying spatial lag error model as the best fitting model. *Section* 5.3.3 discusses how various neighborhood were defined to test the hypothesis that the impact of crime, transit and dwelling density are likely to differ in various neighborhoods.

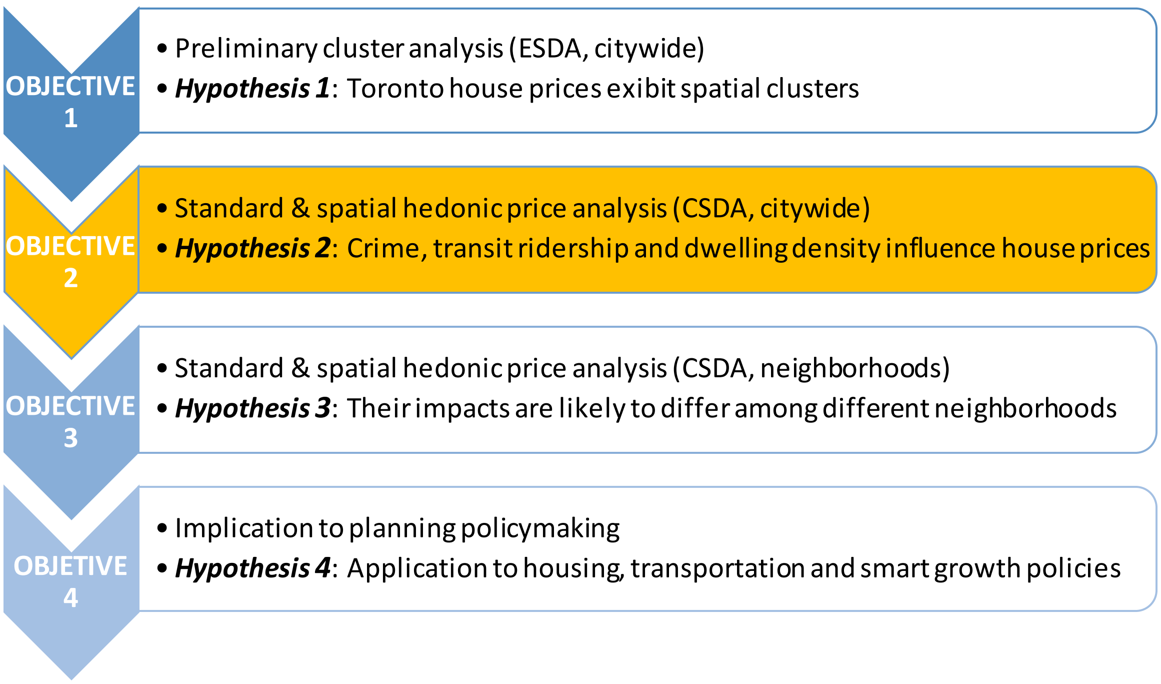

Figure 5.3.1 Objective 2 and Section 5.3 in Context

5.3.1 Univariate Regression Analysis and Bivariate Correlation Test

To begin with, univariate regression analysis was first conducted between logged house prices (dependent variable) and each of the explanatory variables in standard hedonic models (OLS). Most variables generated were significant and further examined for multicollinearity, which is a common problem of regression analysis where explanatory variables are highly correlated. This problem was addressed via bivariate correlation test and then OLS regression diagnostics. The bivariate correlation test was conducted in two steps and the first was to test between variables that measure the same dimension (e.g. government transfer payment and low family income are both indicators for households' economic status). For highly correlated explanatory variables (r>0.5), the variable with a larger residual sum of squares were withdrawn from our analysis, due to inferior model fit.

The second step of bivariate correlation analysis was conducted between variables measuring different dimensions. Highly correlated (r>0.5) explanatory variables (e.g. location quotient of management jobs and average income) were retained, but their sum of squares were recorded for use in multivariate regression analysis, during which the multicollinearity diagnostics will determine whether to include or exclude the variable (discussed in greater detail in section 5.2.3). Selected variables after the two-step bivariate correlation test were presented in Table 5.3.1. Bivariate correlation matrices can be seen in Appendix A.

Table 5.3.1. Selected explanatory variables for multivariate regression

Neighborhood and Locational Characteristics	
1. Dwelling Density	2. Single detached house density
3. Apartment (duplex) density	4. Apartment (with 5 stories +) density
5. Neighborhood stores	6. Average passengers of subway stations
7. Average income	8. Government transfer payment %
9. LQ management jobs	10. Percentage of bachelor degree
11. Crime rate (Property crimes)	
11.1 Property crime	11.2. Mischief
11.3. Theft of a motor vehicle	11.4. Break and enter
12. Crime rate (Non-property crimes)	
12.1. Drug offences	12.2 Robbery
12.3. Violent crime	
Household Characteristics	
1. Number of rooms per dwelling	2. Private dwellings need major repair
3. Percentage of visible minorities	4. Couples without children home
5. Percentage of nonfamily households	6. Percentage of houses before 1946

5.3.2 Multivariate Regression Approach

Variables selected from bivariate correlation analysis were tested in three regression models: standard hedonic model (estimated with OLS), spatial lag independent hedonic model and spatial error hedonic model (spatial lag dependent is not suitable for our dataset as noted in 5.2.3.1). OLS regression were first conducted and great attention were paid to the OLS regression diagnostics to address the problem of multicollinearity and to determine the best fitting model.

The first important indicator in the OLS regression diagnostics is the multicollinearity condition number, which signals the problem of highly correlated explanatory variables. The rule of thumb is that the number greater than 30 suggests multicollinearity, which may undermine the accuracy of regression results. In the case of multicollinearity, attention was paid to previously identified and retained highly correlated variables, and the ones with greater residual sum of squares (or the greatest probability in the model) are removed from the analysis until the multicollinearity number is below 30.

Another set of critical indicator is the OLS diagnostics is the Lagrange Multiplier test statistics, which were used to determine which spatial model (spatial lag or spatial error) is a better fit. If both Robust Lagrange Multiplier (LM) error and Robust LM lag are statistically significant ($p<0.05$), the one with greater significance (smaller probability value) suggests a better spatial model fit (Anselin, 2005). This model selection decision process can be seen in Figure 5.3.2 below. In addition, the largest log likelihood indicates the best model fit among the three models. All OLS and spatial regressions were conducted in the spatial analysis software Geoda. For spatial regression models, spatial effects were incorporated by creating a first-order row-standardized queen contiguity matrix.

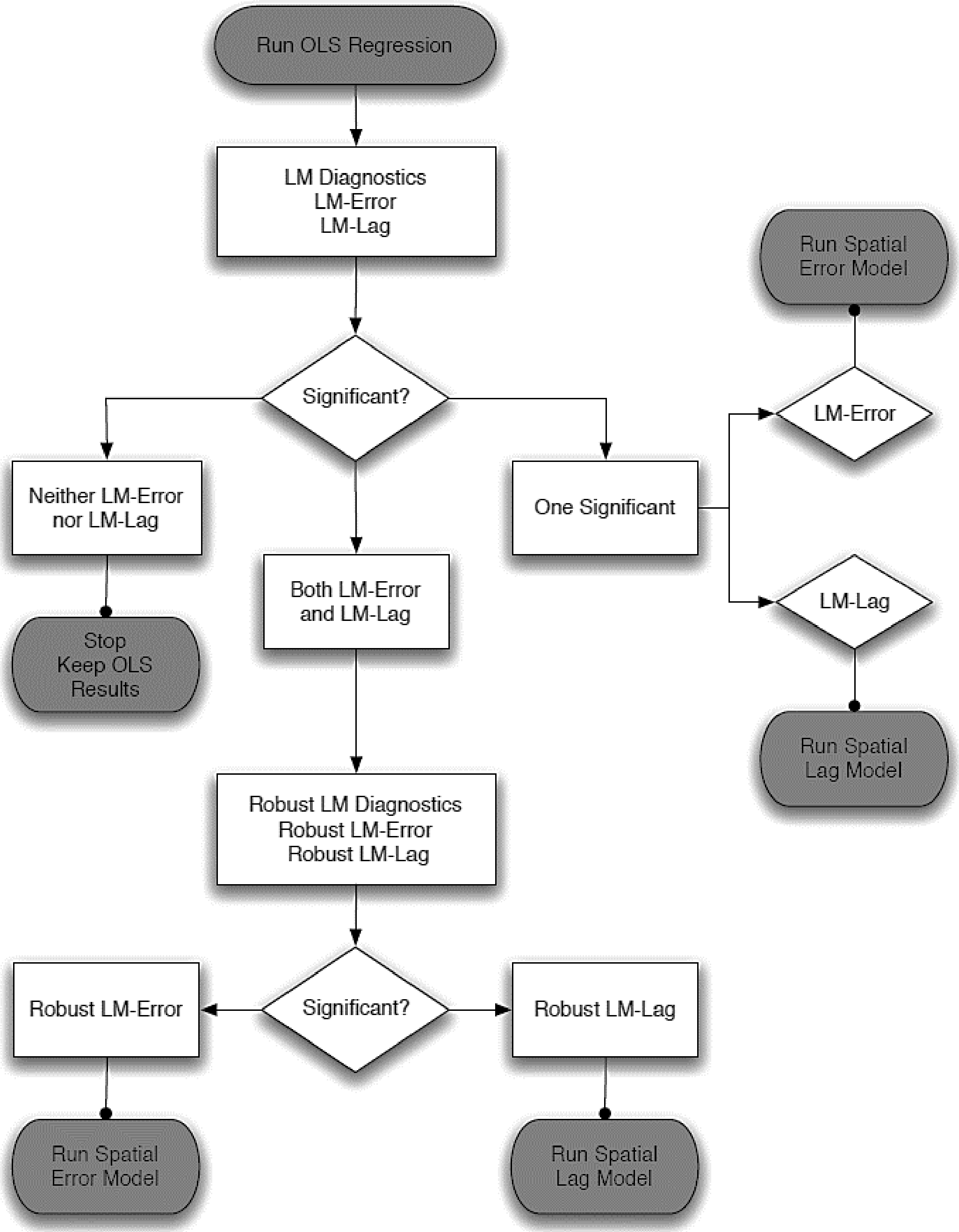

Figure 5.3.2 Spatial Regression Model Selection Decision Process
Source: Anselin (2005). *Exploring spatial data with Geoda™: a workbook*. pp. 199.

5.3.3 Categorizing Neighborhoods

Earlier, we hypothesize that the impacts of crime rates, transit ridership and dwelling density on house prices are likely to differ among various types of neighborhoods. This is, further disaggregating citywide housing market into localized ones. To operationalize this hypothesis, all census tracts were grouped into three neighborhoods based upon the average individual income level of each census tract. Because income level reflects other neighborhood characteristic such as unemployment rate or racial composition (Tita et al., 2006), average individual income level of each census tract was selected as the proxy for dividing the census tracts into three neighborhoods.

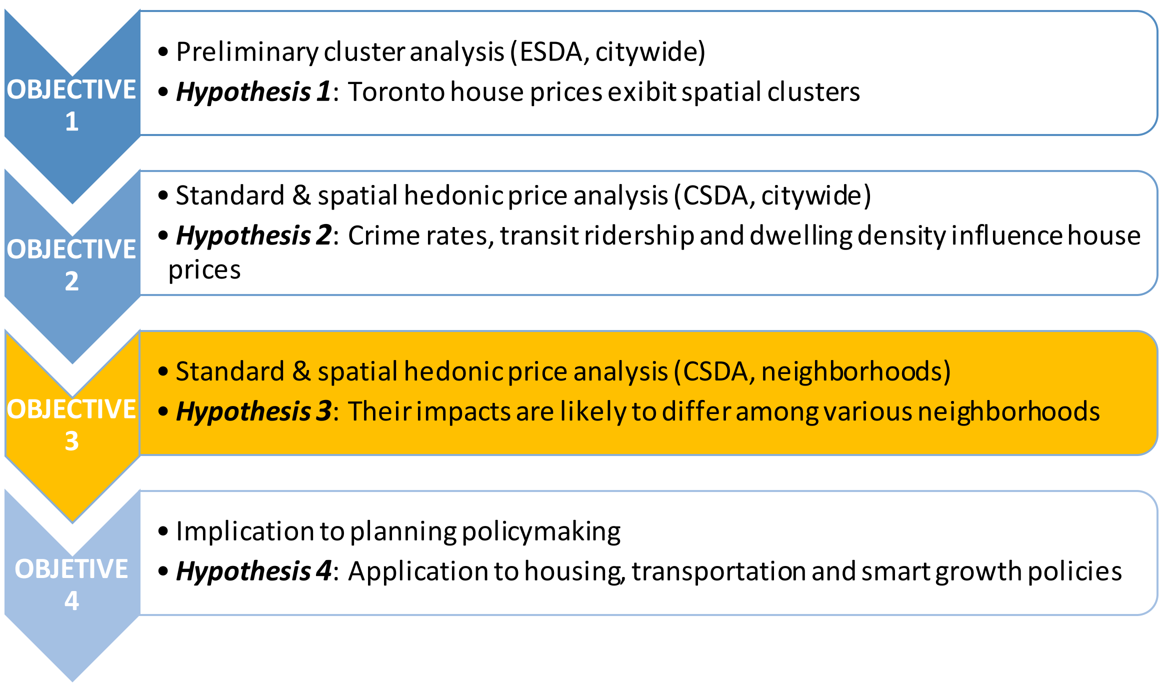

Figure 5.4.2 Objective 3 and Section 5.4.2 in Context

To define low-, middle-, and high-income neighborhoods, average individual income of each census tract were converted to location quotients (LQ). Instead of revealing merely absolute values of income, LQ compares average income of each census tract to that of the entire city. In generally terms, location quotient (LQ), ranging from 0 to infinity, is a ratio that quantifies the concentration of one variable in a smaller unit (e.g. census tract) in comparison with the concentration of the same variable in a larger reference context (e.g. the entire city). A LQ value of 1.00 indicates an equal income level between the tract and the entire city, while a LQ value less than 1.00 indicates income level lower than the city's average, and vice versa.

$$\text{LQ (Location quotient of income)} = \frac{\text{average income of each census tract}}{\text{average income of the City of } Toronto}$$

Specifically, high-income neighborhoods were defined as census tracts with LQ values greater 1.3, which indicates an income level 30% more than the city's average income. Middle-income neighborhoods were defined as census tracts with LQ values between 0.7 and 1.3, which indicates an income level 30% below to 30% above the city's average income. Low-income neighborhoods were defined as census tracts with LQ values less than 0.7, which indicates an income level more than 30% below city's average income. Map 5.3.3 below illustrates the results of the study area divided by 3 income categories, upon which the following analysis is based.

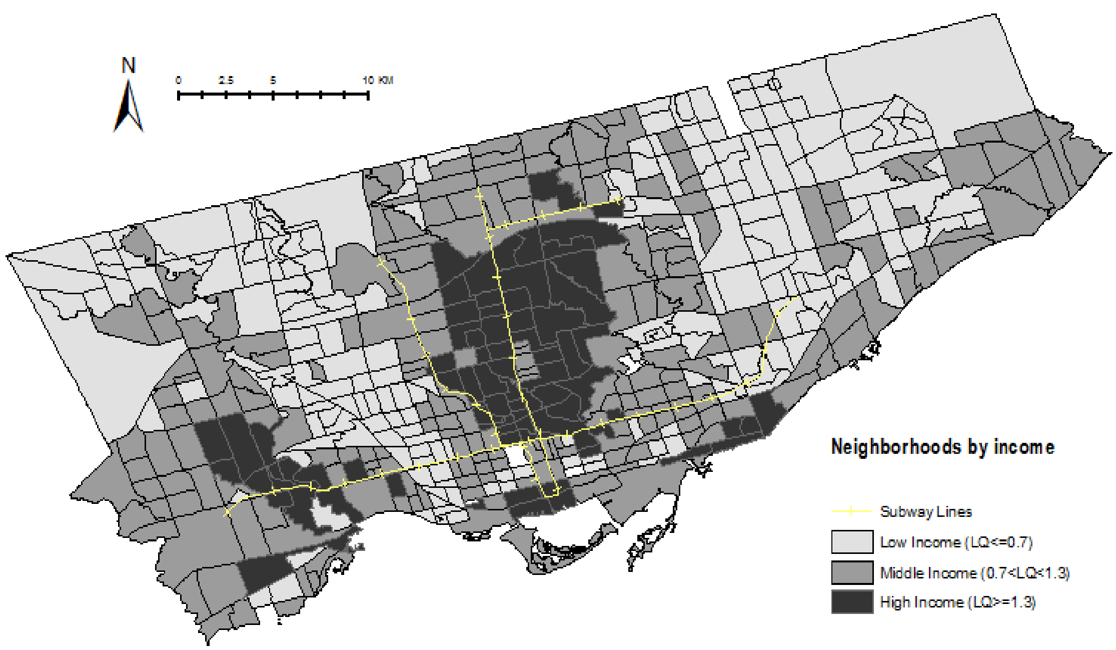

Figure 5.3.3 Neighborhoods by income category across the City of Toronto (census tract level, 2006)

5.4 Results

Section 5.4.1 examines the performance of the three models to justify our use of the spatial error hedonic model. *Section* 5.4.2 presents results where the regression models were first applied at the citywide scale, meaning the entire sample of 522 census tracts across the City of Toronto. *Section* 5.4.3 presents results of applying the model separately to three different neighborhoods defined previously. It is worth noting that, aside from dwelling density and transit ridership, in order to incorporate the impacts of crime rate of each component crime, the hedonic price model was created for each crime type. This means that when regressing house prices with explanatory variables, only one type of crime was included each time along with the same selected variables.

5.4.1 Model Performance: Spatial versus Standard Hedonic Models

Model specifications were compared among the three hedonic models to determine the best fitting model. Table 5.4.1.1 presents an example of regression diagnostics when modelling house prices with “theft of a vehicle” and other significant variables. The multicollinearity condition number was less than 30 in OLS regression model justifying that the correlation between explanatory variables has been controlled in the estimation. The value of robust LM probability for spatial error model (p=0.00) was smaller than that of spatial lag independent model, which suggests that the spatial error model outperforms the spatial lag independent model. Similarly, the value of log likelihood and R-square for spatial error model were the largest among the three models. The larger log likelihood and R-square further confirm that the spatial error model fits the data best for estimating impacts of neighborhood characteristics on house prices. Henceforth, spatial error hedonic model is called the “spatial hedonic model”.

Previously, our exploratory spatial data analysis (ESDA, *Chapter* 4) conducted at the MLS level demonstrated that spatial autocorrelation exists within Toronto house prices. To confirm that spatial dependence disappears after incorporating spatial effects into our models, residuals of regression analysis were examined in Moran’s I statistics. As regression analysis was conducted at the census tract level (due to data availability), spatial dependence test of residuals and house prices were also conducted at the census tract level for ease of comparison. Table 5.4.1.2 continues our example with the crime type "theft of a vehicle". House prices at the census tract level exhibit spatial dependence at one percentage level, as indicated by the positive Moran's I value and p-value (0.001). Spatial autocorrelation among residuals, on the other hand, had eliminated, as indicated by the p-value being greater than five percentage (0.151, not statistically significant). This justifies our use of spatial models.

Table 5.4.1.1 Comparison of Model Specifications

Model Indicator	OLS	Spatial Lag Independence	Spatial Error
Multicollinearity condition number	28.07	--	--
Robust LM probability	--	0.037	0.00
Log likelihood	-35.19	44.20	185.63
R-square	0.53	0.65	0.86
Observations	522	522	522

Notes: *Based on the regression modelling results of "theft of a vehicle".*

Table 5.4.1.2 Spatial Autocorrelation in House Prices (2006, census tract) and Residuals

	Moran's I value	E[I]	p-value	z-value	St. Deviation
House Prices	0.8135	-0.0019	0.001	34.20	0.0239
Residuals	-0.0268	-0.0019	0.151	-1.041	0.0251

Notes: *The result is based on permutations of 999, with "theft of a vehicle".*

5.4.2 Citywide Estimations

Table 5.4.2.1 summarizes relationships between property-crime rates, other significant neighborhood characteristics and house prices in the best fitting model (the spatial error model). Table 5.4.2.2 summarizes results of the same spatial hedonic regression model incorporating each non-property crime. Some variables that formed part of the initial hypothesis were dropped and omitted from result reporting because of insignificance, such as average number of rooms per dwelling, percentage of couples without children home, percentage of nonfamily household, and percentage of visible minorities. Theft of a vehicle is the only crime type (among both property crimes and non-property crimes) that is significant at the citywide scale.

In both Table 5.4.2.1 and Table 5.4.2.2, average income is positively associated with house prices at 0.1% significance level. Neighborhoods with greater percentage of older houses (built before 1946) have lower average house prices. Subway ridership is consistently significant at 0.1% level among all estimating models. Density variables are not significant with the citywide estimations. Model specifications and diagnostics are consistent across all spatial error hedonic models created for each type of crime. Detailed interpretation and discussion of the results can be seen in *Chapter 6*.

Table 5.4.2.3 presents the detailed regression result when crime rate of 'theft of a vehicle', along with dwelling density variables and subway ridership were included in the analysis. Regression diagnostics and specifications including coefficients of each explanatory variable, multicollinearity condition number, Robust LM error (lag), the value and significance level from log likelihood ratio test for spatial dependence, log likelihood and R square numbers were compared among all three models (OLS, spatial lag independent and spatial error models). Regression results created to incorporate each of other crime types can be found in Appendix B.

The column labeled OLS contains results for standard hedonic models, where most household, neighborhood and locational characteristics are significant. However, some of the coefficients are surprising and misleading: the indicated impacts on house prices are unreasonably positive or negative, which is supposed to be the opposite in practice. For example, "percentage of occupied private dwellings need major repair" is a variable measuring general housing maintenance. The expected results are negative coefficients indicating negative impacts on the house prices, but in the standard hedonic estimation, this variable increases house prices with coefficients ranging from 1 to 1.6. However, this problem is addressed in spatial models where the coefficient switched to negative (Table 5.4.2.3). Similar suspicious coefficients in standard hedonic models and their correction addressed in spatial models can be seen in Appendix B. This further confirms that the spatial hedonic model is a better fit than the standard hedonic model where house price data exhibit spatial dependence.

The column labeled spatial lag independent contains results for spatial lag independent hedonic models. Spatial lag dependent model is not suitable for the dataset as noted previously and are therefore omitted from the result reporting in this section. The column labeled spatial error contains results for spatial error hedonic model, which is the best fitting model as indicated by regression diagnostics and discussed previously in *Section* 5.4.1. As expected, many significant household and neighborhood characteristic variables (e.g. dwelling density, percentage of nonfamily households) in OLS had lost their significance in the spatial error hedonic model. Interestingly, the level of significance of yearly subway ridership ("average passengers of subway stations") had increased from 5 percent to 0.1 percentage.

5.4.3 Neighborhood Estimations

The same regression analysis including univariate regression and multivariate regression were repeated for each of the low-, middle-, and high-income neighborhoods (as defined in *Section 5.3.3*). Multicollinearity condition number is controlled under 30 for the problem of bivariate correlation. In the univariate spatial error regression, six types of crime were significant with the estimation of middle-income neighborhood, while none of the crimes were significant with the estimation of low- and high-income neighborhoods.

Table 5.4.3.1 summarizes the impacts of crime on houses prices by comparing regression results in the defined neighborhoods (citywide, low-, middle- and high-income neighborhoods). Middle income neighborhood was the most responsive to crime rates. Since most variables including dwelling density, crimes and subway ridership were insignificant in the low- and high-income neighborhoods, results for the these two neighborhoods were only included in Appendix for reference.

Table 5.4.3.2 presents spatial error hedonic results for the middle income neighborhood and only significant variables are included. Coefficients on other insignificant characteristics have been omitted for brevity. Six types of crimes: property crime, mischief, break and enter, theft of a vehicle, violent crime, and robbery exhibited significant impacts on house prices in the middle-income neighborhoods. Dwelling density had a positive association with house prices in this neighborhood, while apartment density significantly depressed house prices. Subway ridership lost significance in all models in this neighborhood. Regression specification diagnostics (e.g. R-square, log likelihood, coefficients) were consistent across the six models below, except that the magnitude of crime coefficients vary. Detailed interpretation of the results were presented in the following chapter.

Table 5.4.2.1. Results for spatial error hedonic models with property crime and significant variables (census tract level, citywide)

Dependent variable: Log house price				
Independent variables:				
Crime Types	**Property Crime**	**Mischief**	**Theft of a vehicle**	**Break and enter**
Coefficients on the crime type	-0.8825676	-4.391499	-16.77144**	-1.737558
Percentage of houses built before 1946	-0.1470271**	-0.141661**	-0.1478414**	-0.1497196**
Average income	1.888303e-006***	1.858552e-006***	1.834136e-006***	1.884721e-006***
Subway ridership	2.332117e-006***	2.291181e-006***	2.467415e-006***	2.290335e-006***
Model Indicators				
R-squared	0.86	0.86	0.86	0.86
Log likelihood	175.88	175.88	182.75	175.35
Number of observation	522	522	522	522

*P<0.05, **P<0.01, ***P<0.001. Each regression model includes only significant variables.

Table 5.4.2.2. Results for spatial error hedonic models with non-property crime and significant variables (census tract level, citywide)

Dependent variable: Log house price			
Independent variables:			
Crime types	**Violent Crime**	**Robbery**	**Drug offense**
Coefficients on the crime type	-1.217617	-7.086643	-3.031085
Percentage of houses built before 1946	-0.1513435**	-0.14794**	-0.1511856**
Average income	1.855168e-006***	1.866012e-006***	1.870133e-006***
Subway ridership	2.324756e-006***	2.369402e-006***	2.249897e-006***
Model Indicators			
R-squared	0.86	0.86	0.86
Log likelihood	175.44	175.75	175.37
Number of observation	522	522	522

*P<0.05, **P<0.01, ***P<0.001. Each regression model includes only significant variables.

Table 5.4.2.3 OLS and Spatial Models for 'Theft of a vehicle rate' (census tract level, citywide)

	Coefficient		
	OLS	**Spatial Lag Independent**	**Spatial Error**
Dependent variable*:* Log house price			
Independent variables*:*			
Household characteristics			
Private dwellings need major repair	1.085793**	0.7048654	-0.1018049
Percentage of nonfamily households	0.1909687	0.0996219	-0.04710702
Percentage of houses built before 1946	-0.17947*	-0.1010932	-0.1439551*
Neighborhood and Locational characteristics			
Theft of a vehicle (rate)	-40.64845***	-24.91037*	-18.3353**
Dwelling density	8.256083e-005***	2.723563e-005	2.196849e-005
Single detached house density	-6.641635e-005	-8.63482e-005	-5.787353e-005
Apartment (duplex) density	-0.0005175042**	-3.928192e-005	-0.0001274646
Apartment (with 5 stories +) density	-7.530272e-005***	-2.981785e-005	-2.784072e-005*
Average income	3.751118e-006***	1.959505e-006**	1.966823e-006***
Government transfer payment (%)	-0.009125519*	0.002891255	-0.0002484
Location quotient of management jobs	0.162793***	0.06743205	-0.03480377
Neighborhood stores (dummy)	-0.06455608*	-0.02920178	-0.004070613
Subway ridership	2.136707e-006*	1.52587e-006	2.382966e-006***
Model indicators			
Multicollinearity condition number	28.07	--	--
Robust LM probability	--	0.037	0.00
Log likelihood	-35.19	44.20	185.63
R-square	0.53	0.65	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001. Regression results of models incorporating every other crime type (seven types in total were examined) can be seen in Appendix B.

Table 5.4.3.1. Summary of crime impacts in defined neighborhoods at census tract level (spatial error hedonic models)

Dependent variable: Log house price						
Independent variables: Crime Types	**Property Crime**	**Mischief**	**Theft of a vehicle**	**Break and Enter**	**Violent Crime**	**Robbery**
Entire city of Toronto	-0.8825676	-4.391499	-19.36802**	-1.737558	-1.217617	-7.086643
Low income neighborhoods	-1.646405	-0.3308097	-11.86695	-7.813306	-1.268375	-8.49223
Middle income neighborhoods	-3.81728**	-20.49622**	-28.3025**	-18.08728*	-11.52201**	-41.06844**
High income neighborhoods	-5.300835	-33.75723	-45.72263	-0.081828	-8.814009	-3.136998

*P<0.05, **P<0.01, ***P<0.001.

Table 5.4.3.2 Regression results in the middle-income neighborhood at census tract level (spatial error hedonic models)

Dependent variable: Log house price						
Independent variables:						
Crime Types	**Property Crime**	**Mischief**	**Theft of a vehicle**	**Break and Enter**	**Robbery**	**Violent Crime**
Coefficients on the crime type	-3.81728**	-20.49622**	-28.3025**	-18.08728*	-41.06844**	-11.52201**
Dwelling density	3.982624e-005*	5.187332e-005**	3.011529e-005	5.291895e-005**	4.957121e-005**	4.237591e-005*
Apartment density	-4.201098e-005*	-5.369697e-005**	-3.233284e-005	-5.512254e-005**	-5.212618e-005**	-4.560197e-005*
Prevalence of low income households	0.005915358**	0.005380071**	0.005879841**	0.005652363**	0.0064103***	0.005948819**
Percentage of bachelor degree	1.102603***	1.003724***	0.9549201***	1.035035***	0.9555872***	1.017523***
R-squared	0.75	0.76	0.75	0.75	0.74	0.75
Log likelihood	75.5	72.76	72.11	70.76	72.00	72.58
Number of observation	233	233	233	233	233	233

*P<0.05, **P<0.01, ***P<0.001 Note: Spatial error modeling results of crime, dwelling density are insignificant in low- and high-income neighborhoods are included in the Appendix.

Chapter 6.
Interpretation and Discussion

Chapter Overview

In this chapter, insights drawn from the previous modeling analysis are discussed in three sections. Section 6.1 focuses on the roles of neighborhood characteristics in explaining the local house price variation, in particular how crime rates, subway ridership and dwelling density impact house prices in various defined neighborhoods. In section 6.2, limitations of this research are discussed, including methodological issues, data availability and quality. Section 6.3 addresses how the study findings can inform planning policy-making relevant to neighborhood improvement, transit-oriented development and housing affordability.

6.1 Valuation of Neighborhood Characteristics: Some Evidence

Based on the spatial hedonic model results in the previous chapter, we focus on the role of crime rate, transit and dwelling density in generating house price variations in hedonic models, and in turn what information they can reveal about homebuyers' willingness to pay for public safety (crime control), transit ridership impacts and dwelling density. Insights and evidence were drawn from recent empirical studies.

6.1.1 Crime Rate and House Prices

Crime threatens quality of life and disrupts neighborhood cohesion (Nasar & Jones, 1997). The cost of crime and fear of crime is one of the major issues in urban economies, and appear to be the forefront of people's concerns about urban life (Gibbons, 2004). The social, economic and psychological costs of crime can be capitalized into property values. Section 6.1.1.1 compares the two sets of results between citywide estimation and neighborhood estimation. Section 6.1.1.2 deals with in particular the estimation results in the middle-income neighborhoods, where six types of crime showed significant impacts on house prices.

6.1.1.1 Citywide vs. Neighborhood Estimations

In the citywide estimation, spatial hedonic model results suggest that theft of a vehicle is the only property crime among the five types that significantly influence house price. None of the non-property crime types showed significant impact on house values at the citywide scale. Property crime, as a type of crime in general (without disaggregation into its component crimes), does not seem to correlate with housing prices in Toronto.

However, by estimating the model separately in three different neighborhoods based on the average income level of census tracts, six types of crimes exhibited significant impacts on house prices in the middle-income neighborhood. The coefficients on the six types of crime were found to be negative in all three neighborhoods, but it is only in the middle-income neighborhoods that most coefficients were significant at one percent significance level. The only exception is the coefficient on 'break and enter' at five percent significance level (Table 5.4.3.1). Also, the magnitude of these coefficients in the middle-income model was larger than those in the citywide estimation models (Table 5.4.2.2).

A question that arises naturally is that: why the citywide housing market is unresponsive to crime rate variations, and why the middle-income neighborhood is? One plausible explanation is that: homebuyers and sellers in one general housing market (Toronto in this case) are heterogeneous in their economic status and individual preferences (Gibbons, Machin, 2008), and therefore their willingness to pay for a particular neighborhood characteristic such as marginal improvement in public safety or crime control would vary (Nhuyen-Hoang & Yinger, 2011). In econometrical terms, the hedonic price function can be highly non-linear (Gibbons & Machin, 2008; Tita, et al., 2006) and the slope of the relationship (e.g. between crime rate and house value) differ in various parts of one housing market. Although we attempt to address this problem by choosing a semi-log function, it is possible the log function cannot reflect the full range of the various willingness to pay of homebuyers.

For example, in housing market A where crime rate is generally high, the price associated with marginal improvement in public safety is often low. This type of neighborhood attract lower income population, who are relatively more adept at the exposure of violent or other disturbing events (Rountree & Land, 1996) and place little value on public safety. In this case, the slope of the relationship between house price and crime rate is shallow.

While in another type of housing market B where houses are generally upscale and residents generally have higher income, low crime rate of the neighborhood is already reasonably capitalized into the high house prices; the wealthy residents are also supposed to have more resource to address their crime concerns. Therefore, the room for improvement of public safety is narrow and the slope of the relationship is shallow. It is possible that market C exists, where buyers place greater value on lower crime rate and are willing to pay for marginal improvement in public safety of their neighborhood, and since there are potentials for the improvement, the slope is sharp.

To further explain the results, we consider two income-related factors: resource accessible and crime reporting behaviors across various neighborhoods. First, residents of the three income-based neighborhood categories does not have the same level of resource to address their concerns of crime. Although higher income neighborhoods are likely to be the targets of property crime because of the potential lucrative return, residents in wealthy communities often have greater resources for optimal precautionary measures (e.g. locked doors, security system). Residents of lower income communities are not necessarily concerned about property crimes such as 'break and enter' (Rountree & Land, 1996). This may explain the unresponsiveness of house prices to crime rates in lower or higher income neighborhoods.

Second, different neighborhoods differ in terms of crime reporting behaviors. As crime is widely accepted as underreported, a "dark figure" of unrecorded crime exists (MacDonald, 2001; Tita, et. al., 2006). The underreporting is associated with not only the nature of the offense, but also the socio-economic characteristics of the victim or witness (Skogan, 1999), such as gender, race, employment situation and education (MacDonald, 2000). Residents of wealthier neighborhoods are more likely to report crimes than residents in poor functioning neighborhoods (Lynch & Rasmussen, 2001). The possibly underreported crime in lower-income neighborhoods compromises the "official" local crime rate, and therefore leads to biased estimation. Both of the two discussed factors may counteract the impacts of crime on local house prices in the high- or low-income neighborhoods.

Interestingly, results of recent studies using standard (non-spatial) hedonic models seem to be disconnecting. For example, Tita et al. (2006) conducted standard hedonic analysis to examine impacts of crime on house prices across 189 census tracts in the city of Columbus, Ohio, where the model was also applied to different neighborhoods based on income levels. Their findings suggests that impacts of violent crime and property crime on house prices were only significant in low- and high-income neighborhoods. However, Tita et al. (2006) did not provide explanations as to why crime impacts were not significant in middle-income neighborhoods. Instead, they provided explanation on why the magnitude of crime impacts were smaller in higher-income neighborhoods than lower-income neighborhoods: that wealthier neighborhoods have more resources to address property or violent crimes

Also, in their estimations, some coefficients on the crime were surprisingly positive (Tita et al., 2006). The misleading and biased results from their standard hedonic model may indicate failure to control for spatial dynamics of the housing market, whose prices are often said to be dictated by "location, location, location". In particular, if spatial autocorrelation exists in house prices, standard hedonic models cannot capture indirect spatial effects from neighbors or unobserved spatial variation (Gibbons & Machin, 2008; Cohen & Coughlin, 2008).

6.1.1.2 Different Types of Crime and Their Various Impacts

In the middle-income neighborhood, house prices are responsive to six types of crime (Table 5.4.3.1). All the crimes reduce house prices as expected, but the degree of influence vary among crime types, as indicated by the coefficients ranging drastically from -3.8 to -41. Robbery has the greatest influence on house prices in this middle-income market (at 1% significance level), as suggested by the greatest magnitude of its coefficient (41). A likely reason is that robbery, as a type of violent crime, often receive greater media attention, which tend to increase the fear level among local residents.

Mischief (common examples include vandalism and graffiti) had the third greatest impacts on house prices in the middle-income neighborhood. This result is consistent with findings by Gibbons and Machin (2008): offences that are highly visible but rather trivial, such as criminal damage to property (vandalism), have significant influence on house prices. Possible reasons are that highly visible crimes can easily trigger fear of crime, which further leads to psychological costs of crime regarding residential choices, because they are perceived by potential home buyers as signals of community instability or neighborhood deterioration (Gibbons, 2004). This also possibly explains why hard-to-observe crimes (less visible by potential buyers) such as property crime and violent crime had weaker influences on house prices in our estimations.

Break and enter had a relatively slight influence on reducing house prices. Among the six types of crimes, it was the only type with influence on the 5% significance level, while the rest five types of crime were significant on the 1% level (Table 5.4.3.2). A possible explanation is that home buyers or residents can easily install effective yet inexpensive security measures to prevent break and enter (Rountree & Land, 1996; Gibbons, 2004).

Comparing between the impacts of property crime and non-property crime raises a question of whether crime against property have lesser or more influence on house price than crime against person. Our estimations suggest that property crime (e.g. break and enter, mischief, theft of a vehicle) seem to have more significant impacts on house prices than crime against persons (e.g. violent crime). It is worth noting that this result only reflects preferences of residential land users in the middle-income neighborhoods, while the influence of crime rates on house values may differ for other types of property such as commercial stores or offices.

6.1.2 Transit-Oriented Development and House Prices

Earlier, the cluster analysis and map-making (in Chapter 4) suggested clusters of higher value houses around transit nodes in Toronto. The regression analysis in our citywide model (Chapter 5, Table 5.4.2.1 & Table 5.4.2.2) further indicated an association between greater subway ridership and higher house prices at a 0.1% significance level. The results raise our interest in exploring the relationship between ridership impacts of transit-oriented development (TOD) and house prices in Toronto.

In our analysis, subway stations with greater ridership are inevitably transit nodes, with high-density, mixed-use development including office towers, high-rise residential apartments (or condos), retail shops, service commercial, and institutions. These structures are like pearls on a string, linked together along subway lines (Cervero, 1993). Transit nodes generate trips with efficient two-way flows (e.g. between workplaces and home) that support the operation of subways and other public transit that connect seamlessly with subway stations (Cervero, 2006). Ridership may also reflect the degree of centralization in each census tract. In fact, density and diversity of land uses have been argued to be closely related to transit ridership (Sung & Oh, 2011). In the Secondary Plans of the City of Toronto (2010), the level of commercial concentration is usually planned with the scale of subway stations in mixed use areas. The larger the subway station, the higher the ridership, and the greater the commercial concentration of the nearby land uses. Ridership is therefore a manifestation of TOD.

Although high-value houses were clustered around subway lines and higher house prices are associated with greater subway ridership, we cannot simply conclude that TOD on its own increases stationary house prices and generates issues of housing affordability. Our current model provides association, not necessarily causation. It is true that land prices around subway stations are high, but the house prices (the dependent variable) examined in our model is the price of single family houses, not rental houses that lower-income households often rely upon for affordable housing. In fact, public transit plays a role in explaining central city poverty as evidenced by a study of 16 cities in the United States (Glaeser, Kahn

& Rappaport, 2008). By estimating the costs of public transit and driving, the authors found that urbanization of poverty is not simply because of the centralization of old houses and apartments, not merely because that wealthy people want more space and want to live in the suburbs where land is cheaper, but to a greater extent because the poor needs better access to public transportation in the city centre for their daily life (Glaeser et. al., 2008).

This leads to an important theory behind the transit-oriented living: the residential self-selection or residential sorting. One of the possible reasons that people choose transit-oriented living is to save the time and money spent on commute. Savings in transportation costs can be critical for lower-income households that have to make every dollar count (Cervero, 2006). Also, young professionals who prefer a vibrant urban lifestyle or a greener lifestyle with less driving may choose to live in nodal areas with greater transit access and ample commercial services. This residential self-selection reveals the potential for mixed-income housing near transit nodes that can appeal to different home buyers' (or residents') preferences, either for the cost-saving or the urban lifestyle. It is worth noting that in conducting hedonic studies, the influences of those existing preferences on house prices are difficult to be separated from the impacts of the built environment (e.g. transportation infrastructure).

Except for the ridership impact, transit accessibility is another dimension of TOD often examined in hedonic studies that assume greater transit accessibility increases property prices[2]. However, a main problem with this assumption is that any measure of transit accessibility (usually measured in the form of distance) can also capture accessibility to many other local amenities, of which the locations are unlikely to be randomly determined. For example, "distance to major roads" of a stationary house also captures its proximity to employment, shopping centres or libraries (Gibbons & Machin, 2008), since commercial and public land uses (e.g. offices, parks) are generally located closer to major arterial roads for accessibility. Therefore, a statistical link between "distance to major roads" and residential house prices may not necessarily indicate accessibility benefits capitalized into property prices, but may simply imply that home buyers would like to pay a premium for proximity to employment or commercial services. Again, statistical correlation does not necessarily imply causation. Unobserved spatial

[2] The "transit accessibility" was not a variable in our model due to data availability. This part of the discussion is relevant to hedonic model design regarding the impact of TOD on house prices.

variations instead of pure transport accessibility, or put another way, intangible amenities and disamenities (e.g. noise), may account for house price differentials.

Following this logic, although transit brings to the stationary neighborhood better accessibility and increased commercial services, there are negative impacts associated with greater transit ridership that can be easily overlooked in hedonic modeling. For example, quality of living environment in stationary areas can be jeopardized due to increased activity intensity, causing congestion and chaos (e.g. crime rate increased due to greater outsider access). Armstrong and Rodriguez (2006) found that transit benefits were weakly reflected in property values when negative impacts of transit were included in the hedonic model. This requires a more comprehensive investigation of the costs (e.g. adverse impacts on the living environment) and benefits (e.g. boost in economic efficiency) of transit accessibility associated with house values. To include a “complete” set of data in an analysis, however, is empirically challenging often due to data restrictions. Even in a scenario where full data were available, considerable multicollinearity (So, Tse & Ganesan, 1996; Adair, Berry & McGreal, 1996) may be a priority concern as a result of the “complete” data.

6.1.3 Dwelling Density and House Prices

Increasing residential density is one of the top objectives of the smart growth movement and of many recent planning policies promoting intensification in Toronto. Greater residential density can be achieved by two ways: reducing the size of the land lot but maintaining the size of the house; and change the size and type of homes such as increasing the number of condominiums and apartments as opposed to single-family detached houses (Aurand, 2010; Song & Knaap, 2004). We are interested in how density characteristics of a neighborhood influence house prices and in turn homebuyers’ preferences for dwelling density features.

In our citywide models, dwelling density variables were not significantly associated with house prices in spatial models (Table 5.4.2.3). It was only when modeled with “theft of a vehicle” (the only significant crime variable in the citywide estimation), that the density of single detached houses has a significantly negative impacts on house prices. One possible explanation is that a city like Toronto has been under great development pressure due to increasing housing demand (mostly population growth by in-migration and immigration) and urban containment plans such as *the Greenbelt Plan* (2005). Growth is thus accommodated with high-density development in built-up areas and within the urban containment boundaries. In this case, residential density, which is expected to increase, have limited impact on house prices. Also, similar to previous discussion, in a general housing market, density preferences may vary among diverse homebuyers, which the citywide model failed to reflect.

In the neighborhood models, dwelling density variables (including overall dwelling density and apartment density) were significant only in the middle-income neighborhood. Overall dwelling density, which includes both single family detached houses and multi-unit structures such as apartments, was found to be positively associated with house prices in this neighborhood (Table 5.4.3.2). One explanation is that consumers are willing to sacrifice their demand for land and space with substitute of better home amenities (e.g. high-quality materials) and proximity to neighborhood amenities (e.g. corner grocery stores). Interestingly, the geographic distribution of middle-income neighborhoods are mostly surrounding the city's subway stations where urban development are denser and land prices are supposed to be higher (See Figure 5.3.3).

Greater density of apartment buildings in the middle-income neighborhood, on the other hand, had a negative impact on the prices of single-detached houses. This finding is consistent with existing property price research and market surveys, which reveal that houses in neighborhoods dominated by low-density single detached houses can be sold at relatively higher prices (Song & Knaapp, 2003, 2004; Grant & Bohdanow, 2008; Cebula, 2009). A likely explanation is that homebuyers of single family detached houses prefer low density neighborhood with exclusively single family houses and that homebuyers perceive the existence of apartment buildings as disamenities that generate traffic congestion and noise, which will diminish their property values (Song & Knaapp, 2004; Gibbons & Machin, 2008).

6.2 Limitations

There are at least three limitations that apply to this thesis. The first limitation is the modifiable area unit problem, which is also a drawback of most spatial research. The second is the ecological fallacy, which indicates that the association identified in this thesis are not necessarily representative of the analysis unit. The third one concerns the quality of the data involved in the analysis.

6.2.1 Modifiable Areal Unit Problem (MAUP)

The modifiable areal unit problem (MAUP) should be acknowledged for most research that involve spatial analysis. Spatial aggregation on varying scales will alter the results of spatial analysis. Specifically, cluster patterns of house prices can be different if the price data were available at a smaller (e.g. census dissemination area) or larger (e.g. census metropolitan area) areal scale. In this research, the house price data from Toronto Real Estate Board (TREB) has already been aggregated from point data of house sale to MLS district scale, which omitted the house price variations within each MLS district. In addition, crime data was only available at census tract level, which means the possible micro-spatial variation in crime rates within each tract is overlooked.

6.2.2 Ecological Fallacy

The ecological fallacy is a logical error in interpretation of statistics when inferences derived from aggregated data are applied to individuals (Schwartz, 1994). Specifically in this research, relationships derived from analysis at the census tract level are indications of these census tracts in general, not for individual single detached houses in the census tracts. In addition, clusters of house prices identified in this research are indications of the average house value in the area, not of individual house in the clustered area.

6.2.3 Data Availability and Quality

The first data limitation is that the house price data retrieved from market statistics published on TREB website were aggregated on the MLS level. Most hedonic studies managed to access transactional point of sale data of each property within their study area. Due to a lack of acce ss in this study, aggregated sales data on MLS level were assigned to census tracts, in order to match with the unit of analysis that other socio-economic data were available at. This limitation further restricted other measurement in this study such as proximity to neighborhood amenities (e.g. schools or transit stations).

Also, structural characteristics of the houses, critical in determining house prices, are usually available along with property point of sale data. Lacking access to the data, census data on household characteristics were used to make up for this limitation. For example, the variable "percentage of household need major repair" was used as an alternative to "the age of the house" commonly used in hedonic price models.

The second data limitation stems from using house prices exclusively of single detached dwellings. Structural characteristics for other types of house, such as semi-detached houses, townhouses or condo apartments vary across the city of Toronto and TREB did not provide relevant data, which makes it difficult to control in regression analysis. In addition, as discussed in previous chapters, it is possible that residents in townhouses or condos have different preferences than residents or buyers of single detached houses (Song & Knaap, 2004). Therefore, the picture revealed in this study is only partial.

The third data limitation is that the number of crimes in each census tract from UCR is based on reporting from victims and witnesses, which means the number of crimes in police records depends on their willingness to report incidents (Statistics Canada, 2013). Therefore, it is possible that certain crime were unreported to the police. Crime is likely to be under-reported in neighborhoods with higher composition of lower income individuals and less established immigrants. Past research has shown that younger, lower income and male victims are more prone to underreporting crimes than homeowners. This difference in reporting behavior would potentially undermine the accuracy of the association in our findings. In addition, as the UCR only records criminal offence that are punishable, incidents that involve both a property and violent offence may be only recorded as violent offenses, which underestimates the number of property offense and inevitably affects the accuracy of the regression results.

6.3 Implication for Planning Policy Development

This part discusses how the findings in this thesis can inform planning policy-making and practices. The first section deals with housing policies regarding crime and public safety. The second section discusses housing affordability issues in transit-oriented development (TOD), with a focus on integrating transportation and housing policies, as well as mixing house types and increasing residential density. The last one discusses the importance of recognizing the residential soring process in planning policies.

6.3.1. Crime and Neighborhood Improvement.

Since highly-visible crimes such as mischief (including graffiti and vandalism) often signal neighborhood disorder, encourage further property damage and induce fear of crime, policies of neighborhood improvement can target cleanup campaigns[3] and damage repair to prevent further vandalism and improve neighborhood status. Also, public policies tackling crime and anti-social behaviors have to base their decisions on the information about social cost of crime. Yet crime data available are often incomplete and inaccurate. Therefore, more resources should be allocated to improve the quality of crime statistics to inform policymaking (Hellman & Naroff, 1979).

Further, if we consider houses as assets of households, rather than consumer goods, then improving the desirability of house ownership can help build household wealth. Since the capacity of lower income families to accumulate financial assets is limited, residential property is typically their primary or only asset. After all, it is the accumulation of wealth, rather than wage, that account for intergenerational poverty (Flippen, 2004; Tita, et. al., 2006). Having said that, policies as a powerful mechanism should address factors associated with house values. Improving neighborhood safety and allocating policy resources to reduce crime rates in lower income neighborhoods can reduce the socio-economic costs of crime and improve financial status of the disadvantaged. This therefore can contribute to altering the distribution of household wealth and achieving the ultimate goal of social equality.

[3] Graffiti can be a controversial issue as some urbanists consider graffiti as art, rather than crime. For example, the Queen Street West Business Improvement Area (BIA) in Toronto has been organizing graffiti tours to showcase the street arts of the neighborhood. This is beyond the scope of discussion in our study.

6.3.2. Transit-based Housing and Affordability

Housing affordability has been an increasingly heated topic in housing policies across many large cities. The trade-off between housing and transit behind people's residential choices (e.g. trade off greater living space for less commute costs in a city centre) indicates that it is more reasonable to view housing affordability as a combination of housing costs and transportation costs, rather than the costs of housing on its own. To the extent that our regression analysis can be generalized, prices of single-family detached houses that near transit nodes are higher in Toronto. One of the main reason is that (expect for the obvious fact that land with better accessibility sells higher) demand for transit-based housing is growing, but supply failed to keep the pace due to obstacles regarding building affordable housing near transit nodes (CTOD, 2009). One of the many obstacles is that land acquisition and permitting process (e.g. rezoning) is lengthy in stationary areas, and government funding for building affordable housing is limited. These increases development costs on the developer's side and eventually passes on to homebuyers.

Condominium in stationary areas has been dominating the redevelopment of Toronto's urban core, especially around transit nodes. Although relatively more affordable than single-detached houses in stationary areas, this form of homeownership is often marketed to knowledge-intensive professionals and thus a higher end of the market (Hulchanski, 2004). This means that condo development does not necessarily help with improving affordability of transit-based housing in the broader market.

The rental and ownership housing market are different, as households in each tenure usually represent a different cohort in terms of both household size and income. Although the rental housing tenure is not the focus of our discussion, it is the most affordable tenure to most people and is expected to be included in the discussion of affordability in transit-based housing. However, the reality in cities like Toronto is that rental units in stationary areas are in most cases priced above the market level (Drummond, Burleton & Manning, 2004; Hulchanski, 2004) in order for developers to make a profit in such projects.

To promote affordable home ownership in transit-based housing, a consensus is growing on mixing house types (e.g. single-detached houses, townhouses, mid-rise condos and high-rise apartments) in stationary areas. As evidenced by a study of ten recent TOD examples across Canada, conducted by Canada Mortgage and Housing Corporation (CMHC, 2009), TOD projects are empirically successful with a broad spectrum of dwelling types[4]. Some of those projects are a mix of high-rise condo and townhouses, a mix of low-rise and mid-rise condos, and a mix of single-detached homes and low-rise apartments. The success of those housing mix projects in transit areas is partly due to close collaboration between the municipalities and the developers. The municipality provided flexibility on zoning and cost sharing on infrastructure, while developers in return provided required amenities.

It is true that mixing housing types for people with a range of income levels in the same stationary neighborhood imposes empirical challenges. A prominent one of them is the neighborhood opposition (NIMBYism) to intensification from residents of lower-density communities. Single-family homebuyers are usually willing to pay more for maintaining neighborhood homogeneity with low-density, single-detached dwellings (Duncan, 2010). Among all the TOD projects in the aforementioned study by CMHC, municipal planners conducted public consultation to address the residents' concerns. For example, as a result of high-rise buildings, increased density and activities can cause burdens of traffic congestion, sightlessness, a block of view and reduced neighborhood stability. In these cases, the developers worked on carefully design mixed housing projects in a way that the neighborhood would support. Further, due to promoting the benefits of such development (e.g. proximity to amenities, lively urban environment), it is possible to attract households who are previously foreign to high density living (CMHC, 2009).

There are several planning implementation tools that can be utilized to promote affordable housing in transit-oriented development. Inclusionary zoning in mixing house types is often enacted by a zoning ordinance over a large area (rather than for a project). It requires new construction to set 10 percent to 25 percent of the total units to be affordable. However, this policy bears the risk of affordable units being built far away from the transit station where land is cheaper (CTOD, 2009). Incentive zoning is

[4] The density of development and the mix of dwelling types depends on the nature of the stationary area. For example, subway stations with greater ridership examined in our study are located in bustling downtown areas or employment centres, which are mixed-use in nature and require high-rise (high-density) development. For avenues or arterial corridors, mid-rise buildings may be more appropriate (Pembina Institute, 2015). City of Toronto has proposed the SmartTrack transit lines which also calls for medium to high density development in its stationary areas to support the transit investment.

another planning tool to increase affordable housing units by rewarding developers with increased density. Increases in allowable residential density around transit nodes can lower construction costs per unit due to economies of scale and enhance ridership to support transit infrastructure maintenance.

Policies can also provide financial incentives to encourage adaptive reuse of parking lots, which is most useful for terminal stations that are soon becoming intermediate stations because of transit line extension (Cervero, 2006). In fact, the conversion of park-and-ride lots to housing reduces developers' risks of dooming a project due to negotiation with multiple property owners in the land purchase period (Cervero, 1993).

6.3.3. Residential Sorting and Planning Policies

For more integrated local planning, housing policies can be designed to synchronize with transportation policies by acknowledging the residential sorting process. People face trade-offs when choosing where to live. Their willingness to pay for neighborhood amenities are determined by interdependent factors such as costs of housing and transportation. These has tremendous implication to planning policymaking.

For example, although one of the goals of policies proposing new transit infrastructure is to increase labor supply, which is based on the assumption that reduced travel costs can move non-workers back to work, they overlook market forces that tend to sort low-income individuals to less accessible areas, where land values are lower and houses are more affordable. Reduced transportation costs are likely to raise local house prices, which makes the location more attractive to the employed (Gibbons & Machin, 2008). This possible impact is an alert to policymakers that in the long term gentrification may occur to victimize households who already live in the neighborhood by pushing them out (Hulchanski, 2010), due to rising rents and house prices - even though they can be the households that are most dependent on public transit.

Studies show that transit-oriented living are often associated with smaller households and families with fewer cars (e.g. under 2 cars). One obvious reason is that limited parking space and therefore expensive parking fees in the dense transit nodal areas often restrict household's ownership of cars. More importantly, this phenomenon reflects the residential self-selection (Cervero, 2006) that households with fewer family members and fewer cars prefer transit-oriented living. Thus, market-responsive policies around transit nodes should be in place such as flexible parking standards that allow reduced parking in exchange for discounted carpool parking and subsidized (or employer-paid) transit passes.

Chapter 7.
Conclusions

Chapter Overview

In this chapter, research findings are summarized for the study area, as well as contributions that this research has made, both to planning academia and practices. Areas for future research are suggested. Some final thoughts are also provided.

7.1. Summary of Findings

This thesis explores the housing market in the City of Toronto, Ontario and identifies the roles of neighborhood characteristics in explaining local variation of house prices. In particular, we employ spatial hedonic analysis to examine the impacts of crime rates, subway ridership and dwelling density upon single detached house prices in various defined neighborhoods. Overall, crime rates and dwelling density are not significantly associated with house prices when our spatial hedonic model is applied to the entire city, but they are significantly reflected in house prices in the middle-income neighborhood, which is mostly located in the south and central of the city, as well as along the subway lines. We attribute these varied impacts among neighborhoods to the heterogeneity of housing market, where people's willingness to pay for neighborhood amenities differ due to their various socio-economic status.

In the middle-income neighborhood, six types of crime significantly decrease house prices and their degree of impacts vary depending on the nature of the crime. In the same neighborhood, overall dwelling density is positively associated with house prices, which is not surprising for a city like Toronto where urban development emphasizes intensification. Apartment density is found to decrease single-family house prices, indicating that homebuyers of single detached houses are willing to pay more for neighborhood homogeneity. Our findings also reveal that higher-value houses are clustered along subway lines and subway ridership is positively reflected in house prices across the city, which raises our interest in the affordability issues in transit-oriented development.

7.2 Summary of Research Contributions

Briefly, this study has three humble contributions. First, the analysis extends the argument for a disaggregated approach to housing price analysis. Census tract data were used to reflect variations across small geographic unit and the impacts of location sensitive factors are captured. Also, instead of an overall "crime rate", crime was disaggregated into its seven component crime types, as different types of crime are likely to impose different impacts on house prices. For example, our findings indicate crimes of disorder (e.g. vandalism) were exerting a greater extent of impact on local house prices than property crimes, which corresponds to findings by Gibbons (2004). Crimes against properties (property crime) and crimes against people (non-property crime) were also distinguished from each other.

Second, heterogeneity within a housing market (e.g. Toronto housing market) is recognized in our analysis and addressed in two ways. On one hand, as spatial autocorrelation exists in house prices, spatial hedonic models, instead of standard (non-spatial) hedonic models, were used to capture the "complete" range of spatial effects across the study area. Traditional hedonic models capitalize locational effects by including a set of characteristics, which may not be adequate and can lead to errors of estimation. As the hedonic model is a popular analysis tool to inform policies regarding the estimation of costs and benefits, adding spatial perspectives is worth considering especially for property markets whose prices are said to be determined by location.

On the other hand, our analysis distinguishes among different types of neighborhoods by income level. Various neighborhoods are likely to have varied willingness to pay for public safety (crime rates), transit ridership impacts and dwelling density features of a neighborhood. Findings of this thesis suggest that house prices of middle income neighborhoods are more sensitive towards different types of crime and dwelling density, than high- and low-income neighborhoods.

Third, our research explores social and economic equality through the lens of housing market. Affordability issues in transit-oriented development should be considered as a combination of housing costs along with transportation costs. Housing policies can also synchronize with transportation policies to recognize the market forces of "sorting" and the trade-offs behind people's transit oriented living. Mixing house types in stationary areas can be a potential solution to housing affordability, but such projects should be financially incentivized to overcome development dilemmas and carefully designed to attract target homebuyers.

7.3 Areas for Future Research

In terms of factors that shape house prices, the influence from the demand side often outweigh the supply side (Hones, Leishman & Watkins, 2005). Toronto as a gateway city has been experiencing real estate booming partly due to the involvement of foreign capital (the demand side) that is less directly related to the local labour market (Moos, 2010). Therefore, future research can be enhanced with including more detailed demographic data on the socio-economic status of immigrants on the demand side. It may also be interesting for future studies to look into age group distribution of the housing market. Since young professionals constitute great percentage of employees in Toronto (e.g. financial and business districts), the inner city housing market can be closely associated with the segregation of neighborhoods by age and household type, which is worth being investigated.

In addition to ownership housing, a housing market as diverse as Toronto's also consists of rental housing as well as housing for special needs and emergence. Renters and owners each comprise approximately half of Toronto's households. Over the past decade, however, little increase in the supply of rental housing is observed in the city (City of Toronto, 2015). Future research addressing the rental housing market is therefore interesting and meaningful, since renters represent different cohorts with different preferences than home owners and government policies are striving to protect rental markets.

Housing if considered as an asset, rather than a consumer good, its price would reflect the present value of the potential growth in value in the near future. Therefore, cross-sectional analysis of how house prices has its weaknesses, since local house prices would react well in advance of time to neighborhood improvement such as accessibility upgrade in prospect (e.g. speculation when new transportation projects are planned). It is meaningful for future hedonic studies to conduct longitudinal analysis of house prices starting before the announcement of a certain project, for a more precise examination of its impacts on house values over the project lifecycle.

7.4 Concluding Thoughts

This is a fascinating time to be involved in the housing market research. Toronto housing market can be exceptional due to a variety of factors that are shaping and reshaping it: regulations such as urban intensification policies and green belt policies, macroeconomic restructuring, smart growth movement, and demographic shifts such as trends of smaller households. Understanding the nature and complexity of these factors is critical for analytical efforts. It is my hope that this research can contribute in a small, yet meaningful way to current planning literature and practices.

Appendix

Appendix A - Bivariate Correlation Matrix

Table A-1 All variables significant with univariate OLS regression

X10	Dwelling Density		
X11	Single detached house density	**X12**	Apartment (duplex) density
X13	Apartment (with 5 stories +) density	**X14**	private dwellings need major repair
X15	Average number of rooms/dwelling	**X16**	Percentage of nonfamily households
X17	Percentage of one family households	**X18**	Percentage of houses built before 1946
X19	couples without children home	**X20**	Average income
X21	(Medium income – Average income)^2	**X22**	Medium income – Average income
X23	Government transfer payment (%)	**X24**	Low income families (2005) (%)
X25	Median income	**X26**	Index of ethnic heterogeneity
X27	Percentage of visible minorities	**X28**	Percentage of Caucasian
X29	LQ (management jobs)	**X30**	Percentage of bachelor degrees
X31	Neighborhood Stores (dummy)	**X32**	Subway Stations (dummy)
X33	LQ (Professionals)	**X34**	Average passenger

Table A-2 Bivariate correlation between Household variables

	X10	X11	X12	X13	X14	X15	X16	X17	X18
X10	1.00								
X11	.067	1.00							
X12	.063	.137	1.00						
X13	.383	.157	.042	1.00					
X14	.027	.007	.062	.002	1.00				
X15	.030	.002	.001	.034	.083	1.00			
X16	.337	.063	.000	.224	.026	.033	1.00		
X17	.350	.083	.003	.230	.027	.016	.959	1.00	
X18	.005	.079	.178	.030	.119	.016	.108	.082	1.00
X19	.012	.000	.039	.026	.209	.075	.129	.108	.005

Note: The bivariate correlation test were conducted starting from variable X10 and excluding crime variables because multivariate regression will be created for each crime type, which means no two crime types will appear in one regression model.

Table A-3. Bivariate correlation between Economic Status variables

	X20	X21	X22	X23	X24	X25
X20	1.00					
X21	.693	1.00				
X22	.949	.812	1.00			
X23	.423	.095	.283	1.00		
X24	.270	.052	.144	.564	1.00	
X25	.660	.196	.435	.631	.562	1.00

Table A-4. Bivariate correlation between demographic variables

	X26	X27	X28	X29	X30
X26	1.00				
X27	0.58	1.00			
X28	0.44	0.58	1.00		
X29	0.56	0.43	0.45	1.00	
X30	0.00	0.00	0.00	.573	1.00

Table A-5 Selected Variables from within-sectional bivariate correlation

X10	Dwelling Density		
X11	Single detached house density	**X12**	Apartment (duplex) density
X13	Apartment (with 5 stories +) density	**X14**	private dwellings need major repair
X15	Number of rooms per dwelling	**X16**	Percentage of nonfamily households
X17	Percentage of houses before 1946	**X18**	Couples without children at home
X19	Average income	**X20**	Government transfer payment %
X21	Percentage of visible minorities	**X22**	LQ management jobs
X23	Neighborhood stores	**X24**	Average passenger
X25	Population density		

Table A-5 Bivariate correlation among all non-crime variables.

	X19	X20	X21	X22	X23	X24	X25
X10	.00	.00	.00	.00	.04	.02	.91
X11	.02	.05	.09	.05	.00	.02	.21
X12	.03	.00	.00	.02	.00	.00	.00
X13	.00	.00	.00	.00	.01	.03	.80
X14	.08	.17	.00	.12	.00	.01	.06
X15	.21	.12	.06	.16	.00	.03	.34
X16	.00	.07	.16	.05	.02	.05	.28
X17	.07	.11	.30	.12	.06	.00	.00
X18	.13	.23	.17	.26	.00	.04	.00
X19	1.00	.42	.25	.62	.02	.00	.03
X20		1.00	.37	.64	.02	.02	.02
X21			1.00	.43	.02	.00	.02
X22				1.00	.03	.01	.02
X23					1.00	.00	.04
X24						1.00	.01
X25							1.00

Note: variables from X10 to X18 are all household variables, between which the bivariate correlation are tested in Table 1.2.

Appendix B – Regression Results

Table B-1. OLS and Spatial Models for 'Property Crime'

	Coefficient		
	OLS	**Spatial Lag Independent**	**Spatial Error**
Dependent variable: Log house price			
Independent variables:			
Crime type			
Property crime (rate)	-7.799809***	-3.352216**	-0.8960124
Household characteristics			
Private dwellings need major repair	1.157856**	0.7363998*	-0.1215906
Percentage of nonfamily households	0.3135263**	0.1154755	-0.02511855
Percentage of houses built before 1946	-0.08525894	-0.0765965	-0.1530606*
Neighborhood status			
Dwelling density	9.941382e-005***	3.425639e-005	2.468674e-005
Single detached house density	-8.774673e-005*	-8.782036e-005	-4.598077e-005
Apartment (duplex) density	-0.0004367133***	-1.182252e-005	-0.000130851
Apartment (with 5 stories +) density	-8.538245e-005***	-3.916807e-005	-3.097394e-005*
Average income	3.829248e-006***	2.032962e-006***	2.061697e-006***
Government transfer payment (%)	-0.009134002*	0.00344476	-4.855582e-005
Location quotient of management jobs	0.1707377***	0.05503643	-0.03434629
Locational characteristics			
Neighborhood stores (dummy)	-0.05300978*	-0.02152354	-0.00459256
Average passenger of subway stations	2.100185e-006*	1.926596e-006*	2.256073e-006***
Model indicators			
Multicollinearity condition number	27.97	--	--
Robust LM probability	--	0.012	0.00
Spatial dependence test	--	11.71***	420.22***
Log likelihood	-27.81	75.33	182.30
R-square	0.54	0.69	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001.

Table B-2. OLS and Spatial Models for 'Theft of a Vehicle Rate'

	Coefficient		
	OLS	**Spatial Lag Independent**	**Spatial Error**
Dependent variable: Log house price			
Independent variables:			
Crime type			
theft of a vehicle (rate)	-40.64845***	-24.91037*	-18.3353**
Household characteristics			
Private dwellings need major repair	1.085793**	0.7048654	-0.1018049
Percentage of nonfamily households	0.1909687	0.0996219	-0.04710702
Percentage of houses built before 1946	-0.17947*	-0.1010932	-0.1439551*
Neighborhood Status			
Dwelling density	8.256083e-005***	2.723563e-005	2.196849e-005
Single detached house density	-6.641635e-005	-8.63482e-005	-5.787353e-005
Apartment (duplex) density	-0.0005175042**	-3.928192e-005	-0.0001274646
Apartment (with 5 stories +) density	-7.530272e-005***	-2.981785e-005	-2.784072e-005*
Average income	3.751118e-006***	1.959505e-006**	1.966823e-006***
Government transfer payment (%)	-0.009125519*	0.002891255	-0.0002484
Location quotient of management jobs	0.162793***	0.06743205	-0.03480377
Locational characteristics			
Neighborhood stores (dummy)	-0.06455608*	-0.02920178	-0.004070613
Average passenger of subway stations	2.136707e-006*	1.52587e-006	2.382966e-006***
Model indicators			
Multicollinearity condition number	28.07	--	--
Robust LM probability	--	0.037	0.00
Spatial dependence test	--	17.01***	441.65***
Log likelihood	-35.19	44.20	185.63
R-square	0.53	0.65	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001.

Table B-3. OLS and Spatial Models for 'Theft from a Vehicle'

Variables	Coefficient		
	OLS	**Spatial Lag Independent**	**Spatial Error**
Crime type			
Theft from a vehicle (rate)	-16.40837 **	-8.25115	-2.523593
Household characteristics			
private dwellings need major repair	1.081829 **	0.6937725	-0.1381411
Percentage of nonfamily households	0.2599583*	0.09657672	-0.0265164
Percentage of houses built before 1946	-0.152499	-0.09333276	-0.155368*
Neighborhood status			
Dwelling density	9.517113e-005 ***	3.022251e-005	2.559418e-005
Single detached house density	-7.367519e-005	-7.632147e-005	-4.502643e-005
Apartment (duplex) density	-0.0005466067 **	-3.817657e-005	-0.0001303425
Apartment (with 5 stories +) density	-8.545042e-005 ***	-3.207685e-005	-3.168086e-005*
Average income	3.814625e-006 ***	2.037979e-006 ***	2.072939e-006***
Government transfer payment (%)	-0.009236951**	0.003917408	0.0001544106
Location quotient of management jobs	0.1825799 **	0.07624469	-0.03498977
Locational characteristics			
Neighborhood stores (dummy)	-0.06348255 *	-0.02529494	-0.005658834
Average passenger of subway stations	1.627503e-006	1.180354e-006	2.198182e-006***
Model indicators			
Multicollinearity condition number	27.82	--	--
Robust LM probability	--	0.019	0.00
Spatial dependence test	--	14.03***	436.63***
Log likelihood	-36.30	53.03	182.01
R-square	0.52	0.66	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001.

Table B-4. OLS and Spatial Models for 'Mischief'

Variables	Coefficient		
	OLS	**Spatial Lag Independent**	**Spatial Error**
Crime type			
Mischief (rate)	-43.45432***	-11.09225	3.984502
Household characteristics			
private dwellings need major repair	1.449415***	0.8931271 **	-0.1130752
Percentage of nonfamily households	0.3171947**	0.2683749*	-0.02440755
Percentage of houses built before 1946	-0.05747911	0.03394683	-0.1521974*
Neighborhood status			
Dwelling density	9.829744e-005***	3.42754e-005	2.655689e-005*
Single detached house density	-4.395318e-005	-7.163054e-005	-4.351345e-005
Apartment (duplex) density	-0.0005792769***	-9.003776e-005	-0.0001284868
Apartment (with 5 stories +) density	-0.0005792769***	-3.80195e-005 *	-3.272003e-005*
Average income	3.439401e-006***	1.814894e-006 ***	2.051475e-006***
Government transfer payment (%)	-0.007856433**	0.002648636	0.00015072
Location quotient of management jobs	0.1916222***	0.05421102	0.00015072
Locational characteristics			
Neighborhood stores (dummy)	-0.06055636*	-0.03016777	-0.00547207
Average passenger of subway stations	1.609611e-006	1.383535e-006	2.233788e-006***
Model indicators			
Multicollinearity condition number	27.93	--	--
Robust LM probability	--	0.011	0.00
Log likelihood	-17.49	97.77	182.18
Spatial dependence test	--	6.18*	399.35***
R-square	0.56	0.71	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001.

Table B-5. OLS and Spatial Models for 'Break and Enter' (burglary)

Variables	Coefficient		
	OLS	**Spatial Lag Independent**	**Spatial Error**
Crime type			
Break and enter (rate)	-18.57258*	-11.49516	-1.553386
Household characteristics			
private dwellings need major repair	1.050329*	0.5741271	-0.1365102
Percentage of nonfamily households	0.2228852	0.1055372	-0.03448955
Percentage of houses built before 1946	-0.176060*	-0.07636448	-0.1601299**
Neighborhood Status			
Dwelling density	9.996801e-005***	3.927211e-005	2.623682e-005*
Single detached house density	-5.833911e-005	-7.958667e-005	-4.553314e-005
Apartment (duplex) density	-0.0005111384**	-0.0004235278	-0.0001210536
Apartment (with 5 stories +) density	-9.154226e-005***	-4.189218e-005*	-3.246933e-005*
Average income	3.963154e-006***	2.164087e-006 ***	2.062057e-006***
Government transfer payment (%)	-0.009667075**	0.003402244	9.172551e-005
Location quotient of management jobs	0.1784752***	0.06837922	-0.03406546
Locational characteristics			
Neighborhood stores (dummy)	-0.05885795*	-0.0246886	-0.005323279
Average passenger of subway stations	1.72141e-006	1.306111e-006	2.226769e-006***
Model indicators			
Multicollinearity condition number	28.18	--	--
Robust LM probability	--	0.03	0.00
Spatial dependence test	--	15.13***	440.45***
Log likelihood	-38.48	49.74	181.75
R-square	0.52	0.66	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001.

Table B-6. OLS and Spatial Models for 'Violent Crime'

Variables	Coefficient		
	OLS	Spatial Lag Independent	Spatial Error
Crime type			
Violent crime (rate)	-21.89547 ***	-7.217942*	-1.919727
Household characteristics			
private dwellings need major repair	1.56096***	0.9889986 **	-0.1047097
Percentage of nonfamily households	0.3241402**	0.118937	-0.0265164
Percentage of houses built before 1946	-0.1696655*	-0.1194146	-0.1587714**
Neighborhood status			
Dwelling density	9.8138e-005***	3.936188e-005	2.634361e-005*
Single detached house density	-7.409244e-005	3.936188e-005	-4.810197e-005
Apartment (duplex) density	-0.0004266822**	-6.400928e-005	-0.0001121787
Apartment (with 5 stories +) density	-9.195136e-005***	-4.299983e-005*	-3.261621e-005*
Average income	3.998312e-006***	2.21724e-006 ***	2.071815e-006***
Government transfer payment (%)	-0.003803292	0.004204954	0.0004489338
Location quotient of management jobs	0.1533551*	0.052695	-0.03515211
Locational characteristics			
Neighborhood stores (dummy)	-0.05874616*	-0.0273349	-0.005574381
Average passenger of subway stations	2.163038e-006*	1.594415e-006	2.274322e-006***
Model indicators			
Multicollinearity condition number	27.86	--	--
Robust LM probability	--	0.011	0.00
Spatial dependence test	--	7.54**	411.47***
Log likelihood	-23.71	70.53	182.03
R-square	0.55	0.68	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001.

Table B-7. OLS and Spatial Models for 'Robbery'

Variables	Coefficient		
	OLS	**Spatial Lag Independent**	**Spatial Error**
Crime type			
Robbery (rate)	-48.7806***	-28.82085*	-8.592288
Household characteristics			
private dwellings need major repair	1.240489**	0.8435526*	-0.1123185
Percentage of nonfamily households	0.1773387	0.07259391	-0.04182473
Percentage of houses built before 1946	-0.198760*	-0.1157945	-0.158711**
Neighborhood status			
Dwelling density	9.983663e-005***	4.285648e-005*	2.586464e-005*
Single detached house density	-6.699999e-005	-6.643003e-005	-4.872371e-005
Apartment (duplex) density	-0.0004849**	-7.969726e-005	-0.0001067489
Apartment (with 5 stories +) density	-9.138556e-005***	-4.35556e-005*	-3.227396e-005*
Average income	3.850272e-006***	2.190214e-006***	2.059486e-006***
Government transfer payment (%)	-0.007431515*	0.004224731	0.0004846046
Location quotient of management jobs	0.1766522***	0.06008555	-0.0337119
Locational characteristics			
Neighborhood stores (dummy)	-0.06232392*	-0.02208782	-0.005902328
Average passenger of subway stations	2.330494e-006*	1.470027e-006	2.31279e-006***
Model indicators			
Multicollinearity condition number	27.71	--	--
Robust LM probability	--	0.028	0.00
Spatial dependence test	--	12.92***	433.91***
Log likelihood	-34.62	57	181.75
R-square	0.53	0.66	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001.

Table B-8. OLS and Spatial Models for 'Drug Offense'

	Coefficient		
	OLS	**Spatial Lag Independent**	**Spatial Error**
Dependent variable: Log house price			
Independent variables:			
Crime type			
Drug offense (rate)	-33.69245**	-21.44542*	-3.549384
Household characteristics			
Private dwellings need major repair	1.166368**	0.7191872*	-0.1301123
Percentage of nonfamily households	0.2421652*	0.08416972	-0.03525012
Percentage of houses built before 1946	-0.2130768*	-0.1486998	-0.1616219**
Neighborhood status			
Dwelling density	9.816505e-005***	3.944942e-005	2.59976e-005*
Single detached house density	-6.674273e-005	-7.472322e-005	-4.66824e-005
Apartment (duplex) density	-0.0004715705**	-1.976726e-005	-0.0001157855
Apartment (with 5 stories +) density	-8.984344e-005***	-4.134034e-005*	-3.231128e-005*
Average income	3.838573e-006***	2.166176e-006 ***	2.05013e-006***
Government transfer payment (%)	-0.00857194*	0.005195612	0.0001642169
Location quotient of management jobs	0.182345***	0.06736253	-0.03411973
Locational characteristics			
Neighborhood stores (dummy)	-0.05300978*	-0.02338616	-0.005999187
Average passenger of subway stations	2.100185e-006*	9.461449e-007	2.239634e-006***
Model indicators			
Multicollinearity condition number	27.45	--	--
Robust LM probability	--	0.04	0.00
Spatial dependence test	--	8.59**	438.24***
Log likelihood	-37.29	57.05	181.83
R-square	0.52	0.67	0.86
Number of observation	522		

*P<0.05, **P<0.01, ***P<0.001.

Table B-9. Summary of Results for High-income Neighborhoods

Dependent variable: Log house price						
Independent variables:						
Crime Types	**Property Crime**	**Mischief**	**Theft of a vehicle**	**Break and Enter**	**Robbery**	**Violent Crime**
Coefficients on the crime type	-4.865042	-33.19457	-55.48835	-1.950341	-13.88511	-11.222
Dwelling density	3.054123e-006	6.961059e-006	1.399734e-006	3.482492e-006	3.067069e-006	5.247726e-006
Apartment density	-0.0001392922	-3.670435e-005	-5.549018e-005	-5.429682e-005	-3.534245e-005	-7.644652e-005
Prevalence of low income households	-0.0002614577	-0.001261677	-0.000199266	0.5644413	-0.001601254	-0.0003624464
Percentage of bachelor degree	-0.0002614577	0.100254	0.3119	1.035035***	0.5978339	1.017523***
R-squared	0.53	0.54	0.55	0.53	0.53	0.53
Log likelihood	-8.88	-7.48	-8.25	-9.20	-9.15	-8.95
Number of observation	80	80	80	80	80	80

Table B-10. Summary of Results for Low-income Neighborhoods

Dependent variable: Log house price						
Independent variables:						
Crime Types	**Property Crime**	**Mischief**	**Theft of a vehicle**	**Break and Enter**	**Robbery**	**Violent Crime**
Coefficients on the crime type	-1.693895	-0.111003	-12.6895	-7.04967	-5.886838	-0.727877
Dwelling density	3.778468e-006	3.730263e-006	3.572053e-006	3.833368e-006	3.584832e-006	3.802503e-006
Apartment density	-0.0001799469	-0.000169768	3.572053e-006	-0.0001544412	-0.0001541992	-0.0001648172
Prevalence of low income households	0.001124113	-0.000169768	0.001200287	-0.0001544412	0.001377211	0.00120172
Percentage of bachelor degree	0.5858225	0.6213249	0.6131794	0.563583	0.001377211	0.00120172
R-squared	0.73	0.73	0.73	0.73	0.73	0.73
Log likelihood	100.37	98.85	99.80	99.57	99.03	98.88
Number of observation	219	219	219	219	219	219

References

Adair, A.S., Berry, J.N. & McGreal, W. S. (1996). Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research*, 13, 67-83.

Anas, A., Arnott, R., Small, K. A. (1998). Urban Spatial Structure. *Journal of Economic Literature*, XXXVI, 1426-1464.

Anderson, S. T. & West, S. E. (2006). Open space, residential property values, and spatial context. *Regional Science and Urban Economics,* (36), 773-789.

Anselin, L. (1995). Local indicators for spatial association. *Geographical Analysis*, (27), 309-329.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In M. Fischer, H.J. Scholten, & D. Unwin (Eds.), *Spatial analytical perspectives on GIS* (pp. 111–125). London: Taylor & Francis.

Anselin, L., Cohen, J., Cook, D., Gorr, W., and Tita, G. (2001). Spatial analyses of crime. *Measurement and Analysis of Crime and Justice,* (4), 213-262. National Criminal Justice Reference Service.

Anselin, L. (2005). Exploring spatial data with Geoda™: a workbook. Center for Spatially Integrated Social Science. Retrieved October 2013 from http://geodacenter.asu.edu/system/files/geodaworkbook.pdf

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

Butler, R. V. (1982). The specification of hedonic indexes for urban housing. *Land Economics*, 56 (1), 96-108.

Bollinger, C. R. & Ihlanfeldt, K. R. (1997). The impact of rapid rail transit on economic development: the case of Atlanta's MARTA. *Journal of Urban Economics*, 42, 179-204.

Capozza, D. R. & Helsley, R.W. (1989). The fundamentals of land prices and urban growth. *Journal of Urban Economics*, (26), 295-306.

Cebula, R. J. (2009). The hedonic pricing model applied to the housing market of the city of

Savannah and its Savannah historic landmark district. *The Review of Regional Studies*, 39, 9-22.

Cervero, R. (1994). Transit-based housing in California: evidence on reidership impacts. *Transport Policy,* 3, 174-183.

Champion, A. G. (2001). A changing demographic regime and evolving polycentric urban regions: Consequences for the size, composition and distribution of the city populations. *Urban Studies*, 38, 657-677.

Charron, M. (2009). Neighborhood characteristics and the distribution of policy-reported crime in the City of Toronto. Statistics Canada. 85-561-M.

Cho, S-H., Poudyal, N. C., & Roberts, R. K. (2008). Spatial analysis of the amenity value of green open space. *Ecological Economics*, 66, 403-416.

City of Toronto. (2010). Chapter Six. Secondary Plans and Maps.

Cohen, P. J. & Coughlin, C. C. (2008). Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of Regional Science*, 48 (5), 859-878.

Cropper ML, Deck LB & McConnell KE (1988). On the choice of functional form for hedonic price functions. *Review of Economics and Statistics,* 70, 668-675.

Centre for Transit-Oriented Development (CTOD). (2009). Mixed-income housing: increasing affordability with transit. Centre for Transit-Oriented Development, Washington, DC (http://www.reconnectingamerica.org).

Cullen J. B., & Levitt, S. D. (1999). Crime, urban flight, and the consequences for cities. *The Review of Economics Statistics*, 81, 159–169.

Drummond, D., Burleton, D., & Manning, G. (2004). Affordable Housing in Canada: In Search of a New Paradigm. In J. D. Hulchanski & M. Shapcott (Eds.), *Finding room: options for a Canadian rental housing strategy* (pp. 15–68). Toronto: Centre for Urban and Community Studies, University of Toronto.

Dubin, R. A, Goodman, A. C. (1982) Valuation of education and crime neighborhood characteristics through hedonic housing prices. *Population and Environment*, 5, 166–181.

Duncan, M. (2010). The impact of transit-oriented development on housing prices in San Diego, CA. *Urban Studies*, 1-27.

Duffala, D. C. (1976). Convenience stores, armed robbery, and physical environmental features. *American Behavioral Scientist, 20,* 227-246

Employment and Social Development Canada. (2015). Security – victims of property crime. Retrieved from http://www4.hrsdc.gc.ca/.3ndic.1t.4r@-eng.jsp?iid=60

Fahrländer, S. S. (2006). Semiparametric construction of spatial generalized hedonic models for private properties, *Swiss Journal of Economics and Statistics*, 142 (4), 501–528.

Filion, P. & McSpurren, K. (2007). Smart growth and development reality: the difficult co-ordination of land use and transport objectives. *Urban Studies*, 44 (30), 501-523.

Filion, P. & Kramer, A. (2012). Transformative metropolitan development models in large Canadian urban areas: the predominance of nodes. *Urban Studies,* 49 (10), 2237-2264.

Flippen C (2004) Unequal returns to housing investments? A study of real housing Appreciation among black, white, and Hispanic households. *Social Forces,* 82, 1523–1552.

Fowler, C. S. (2011). Spatial Regression: Center for Studies in Demography and Ecology. University of Washington. Retrieved 9 April 2014 from http://csde.washington.edu/services/gis/workshops/Resources/SPREG_Presentation.pdf.

Gibbons, S. (2004). The costs of urban property crime. *The Economic Journal*, 144, F441-F463.

Gibbons, S. & Machin, S. (2008). Valuing school quality, better transport and lower crime: evidence from house prices. *Oxford review of Economic Policy*, 24, 99-119.

Grant, J. L. & Bohdanow, S. (2008). New urbanism developments in Canada: a survey. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 1(2), 109-127

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

Haurin DR, Brasington D (1996) School quality and real house prices: inter-and intra-Metropolitan effects. *Journal of Housing Economics*, 5, 351–368.

Healey, P. (1991) Models of the development process: a review. *Journal of Property*

Research, 8, 219– 238.

Heikkila, E., Gordon, P., Kim, J. I., Peiser, R. B., & Richardson, H. W. (1989). *Environment and Planning A*, 21, 221-232.

Helsley, R. W. & Strange, W. C. (1999). Gated communities and the economic geography of crime. *Journal of Urban Economics*, 46, 80-105.

Hellman, D. A., & Naroff, J.L. (1979). The impacts of crime on urban residential property values. *Urban Studies*, (16), 105-112.

Hirschfield, A. and Bowers, K.J. (1997). The effect of social cohesion on levels of recorded crime in disadvantaged areas. *Urban Studies,* 34 (8), 1275-1295.

Hulchanski, J. D. (2004). How Did We Get Here? The Evolution of Canada's "Exclusionary" Housing System. In J. D. Hulchanski & M. Shapcott (Eds.), *Finding room: options for a Canadian rental housing strategy* (pp. 179–193). Toronto: Centre for Urban and Community Studies, University of Toronto.

Hulchanski, J. D. (2010). The three cities within Toronto: income polarization among Toronto's neighborhoods, 1970-2005.

Islam, K. S. & Asami, Y. (2009). Housing market segmentation: a review. *Review of Urban and Regional Development Studies*, 21, 93 – 109.

Jacquez, G. M. (2008). "Spatial Cluster Analysis". In Fotheringham, S. and Wilson, J. (Eds.), *The Handbook of Geographic Information Science*. Blackwell Publishing, 395-416.

Jones, C. & Leishman, C. (2003). Structural change in a local urban housing market. *Environment and Planning*, 35, 1315-1326.

Jones, C., Leishman, C. & Watkins, C. (2005). Housing market process, urban housing submarkets and planning policy. *Town Planning Review*, 76 (2), 215-234.

Lehner, M. (2011). Modelling housing prices in Singapore applying spatial hedonic regression, *Master of Science Thesis*, Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich.

Lerman, S.R. and Kern, C.R. (1983). Hedonic Theory, Bid Rents, and Willingness-to-Pay: Some Extensions of Ellickson's Results. *Journal of Urban Economics,* 13 (3), 358–63.

Leung, C. (2004). Macroeconomics and housing: a review of the literature. *Journal of*

Housing Economcis, 13, 249-267.

Lin, J. J. & Gau, C. C. (2006). A TOD planning model to review the regulation of allowable development densities around subway stations. *Land Use Policy*, 23, 353-360.

Lynch, A. K. & Rasmussen D. W. (2001). Measuring the impact of crime on house prices. *Applied Economics*, 33, 1981 – 1989.

Mahan, B. L., Polasky, S. & Adams, R. M. (2000). Valuing urban wetlands: a property price approach. Land Economics. 76, (1), 100-113.

Matthews, S. A. & Parker, D. M. (2013). Progress in Spatial Demography. *Demographic Research*, 28 (10), 271-312.

McDonald, J. F. (1987). The identification of urban employment subcenters. *Journal of Urban Economics*, 21, 242-258.

Moos, M. & Skaburskis, A. (2010). The globalization of urban housing markets: immigration and changing housing demand in Vancouver. *Urban Geography*, 31, (6), 724-749.

Morenoff, J.D. & Sampson, R. J. (1997). Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. *Social Forces*, 76, 31–64.

Neuman, M. (2005). The Compact City Fallacy. *Journal of Planning Education and Research,* 25, 11–26.

Nguyen-Hoang, P. Yinger, J. (2011). The capitalization of school quality into house value: a review. *Journal of Housing Economics*, 20, 30-48.

Ottensmanna, J. R., Paytona, S., & Man, J. (2008). Urban Location and Housing Prices within a Hedonic Model. *The Journal of Regional Analysis and Policy*, 38(1), 19-35.

Quick, M. (2013). Exploring crime in Toronto, Ontario with applications for law enforcement planning: Geographic analysis of hot spots and risk factors for expressive and acquisitive crimes. A Thesis. University of Waterloo.

Rosen, S. (1974), 'Hedonic Prices and Implicit Markets: Product differentiation in pure competition. *Journal of Political Economy*, 82, 34–55.

Rountree, P. W. & Land, K. C. (1996). Burglary victimization, perceptions of crime risk, and routine activities: a multilevel analysis across Seattle neighborhoods and census tracts. *Journal of Research in Crime and Delinquency*, (33), 147-180.

The Pembina Institute (2015). Make way for mid-rise. How to build more homes in walkable, transit-connected neighborhoods. Retrieved 2015 June from http://www.pembina.org/reports/make-way-for-mid-rise.pdf

Sirmans, S. G., D. A. Macpherson & Zietz, E. N. (2009). The composition of hedonic pricing models, 13 (1) 1–44.

Schwanen, T. and Mokhtarian, P. (2004). The extent and determinants of dissonance between actual and preferred neighborhood type, *Environment and Planning B*, 31, 759–784.

Schwartz, S. (1994). The fallacy of the ecological fallacy: The potential misuse of a concept and the consequences. *American Journal of Public Health*, 84, 819-824.

Sloane, D.C. (2006). From congestion to sprawl. *Journal of the American Planning Association*, 72(1), 10–18.

So, H.M., Tse, R.Y.C. & Ganesan, S. (1996). Estimating the influence of transport on house prices: evidence from Hong Kong. *Journal of Property Valuation and Investment*, 15 (1), 40-47.

Song, Y & Knaap, G-J. (2003). New Urbanism and housing values: a disaggregate assessment. *Journal of Urban Economics*, 54, 218-238.

Song, Y. & Knaap, G-J. (2004). Measuring the effects of mixed land uses on housing value. *Regional Science and Urban Economics,* 34, 663-680.

Statistics Canada (2015). Retrieved from http://www.statcan.gc.ca/pub/85-002 x/2013001/article/11854-eng.htm

Sung, H. & Oh, J-T (2011). Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. *Cities*, 28, 70-82.

Taylor, Z. & Nostrand, J. V. (2008). Shaping the Toronto region: past, present and the future. An Exploration of the potential effectiveness of changes to planning policies governing greenfield development in the Greater Golden Horseshoe. Neptis The Architecture of Urban Regions.

Thaler, R. (1978). A note on the value of crime control: evidence from the property market. *Journal of Urban Economics*, 5, 137–145.

Tita, G. E., Petras, T. L. & Greenbaum, R. T. (2006). Crime and residential choice: a neighborhood level analysis of the impact of crime on housing prices. *Journal of Quantitative Criminology*, 22, 299-317.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 234–240.

Voss, P.R., White, K. & Hammer, R. (2006). Explorations in Spatial Demography. In Kandel, W.A. & Brown, D.L. (Ed*.), Population Change and Rural Society* (pp. 407-429). Netherlands, Springer.

Ward, M. D. & Gleditsch, K.S. (2007). An introduction to spatial regression models in the social science. Retrieved December 2013 from https://web.duke.edu/methods/pdfs/SRMbook.pdf

Wenz, P.S. (2001). Just Garbage. In Westra, L. & Lawson, B (Eds). Faces of Environmental Racism: Confronting Issues of Global Justice. 2nd Edition. Rowman & Littlefield Publishers.

Wilson, W. J. (1996). When Work Disappears. New York: knopf.

Zhukov, Y. M. (2010). Applied spatial statistics in R, Section 6. Retrieved 8 April 2014 from http://www.people.fas.harvard.edu/~zhukov/Spatial6.pdf.