TOWARDS A MICROSIMULATION RESIDENTIAL HOUSING MARKET MODEL: REAL ESTATE APPRAISAL AND NEW HOUSING DEVELOPMENT

TOWARDS A MICROSIMULATION RESIDENTIAL HOUSING MARKET MODEL: REAL ESTATE APPRAISAL AND NEW HOUSING DEVELOPMENT

By

XUDONG LIU, B.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2011) McMaster University

(School of Geography and Earth Sciences) Hamilton, Ontario

TITLE: Towards a Microsimulation Residential Housing Market Model: Real Estate Appraisal and New Housing Development

AUTHOR: Xudong Liu, B.Sc. (China University of Geosciences Beijing)

M.Sc. (Graduate University of Chinese Academy of Sciences)

SUPERVISOR: Dr. Pavlos S. Kanaroglou

NUMBER OF PAGES: ix, 74

ABSTRACT

As a mid-size industrial city in North America, the City of Hamilton has been increasingly experiencing urban sprawl in the past six decades coupled with population growth and economic development. The study of various interdependent processes driving the evolution of urban form requires the application of simulation models that offer urban planners and policy-makers an efficient means for evaluating urban development policies. This thesis focuses on the modeling efforts towards building a microsimulation residential housing market system for the City of Hamilton. To this end, two major tasks have been conducted in this research. First, a state-of-the-art agent-based microsimulation housing market framework has been designed. Second, two model components in the microsimulation framework, namely a real estate appraisal model and a new housing development model, have been estimated.

The objective of the real estate appraisal model is to assess the market values of existing dwellings based on the housing transactions in the previous period. Three model forms, including a traditional hedonic model, a spatial regression model, and a regression Kriging model, have been employed in estimations for comparison purposes. A series of independent variables that describe the characteristics of dwelling, location, and neighborhood are specified in the explanatory model. The comparisons among estimation results demonstrate that the spatial regression model has achieved a higher goodness-of-fit than the traditional hedonic model. In addition, we verified that spatial autocorrelation is present in the residuals of the traditional

hedonic model, which is explicitly captured by the spatial regression model. In terms of model prediction accuracy, spatial models (SAR and Kriging) both achieve a certain level of improvements over the traditional hedonic model. Overall, we end up recommending that the SAR model is more appropriate to be incorporated into the microsimulation framework, as it provides the best match between predicted and observed values.

The new housing development model enables the development of a dynamic housing supply module in the simulation framework by modeling the location and type decisions during the housing development process for each year. A parcel-level two-tier nested-logit model has been estimated. The model is able to deal with not only the decision to develop a specific vacant residential land parcel, but also the development type choice. In terms of the factors influencing the decision to develop, the picture revealed from the model estimation results is that land developers are more likely to start a development project in greenfields than in brownfields. As for the type choice decision during the development process, a variety of variables describing transportation accessibility, residential amenities, the characteristics of the land parcel and neighborhood are included in the model specifications.

ACKNOWLEDGEMENTS

I would like to express my most sincere gratitude to my supervisor, Dr. Pavlos Kanaroglou, for your guidance, support, patience in my research. Thank you for helping me get through those stressful times, and showing your respect on my career decision in the end.

Sincerely thank Dr. Antonio Paez and Dr. Darren Scott in my defense committee for their remarks on this research and all the knowledge I've learned from their courses.

Thanks to Dr. Jamie Spinney, Dr. Hanna Maoh, and Dr. Steven Farber who are very patient to answer my questions in research that gave me much confidence.

Special thanks to the department staff in main office, especially to Ann and LouAnne.

Also send my thanks and best wishes to my classmates, Ron, Nay, Tufayel, Monir, Jeff. Thank Pat and Laura for maintaining a very comfortable working environment.

I am very grateful to the support from my family, especially my Mom. I love you all forever.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Motivation and Significance of Research

In the City of Hamilton, Ontario, 37.4 percent of households moved residences over 2001-06, with 12.6 percent moving in the last year of this time period alone (2006 Canadian Census). Other Canadian cities experienced a similar trend. Such intensity of population redistribution can have a significant effect on urban form with consequences for the city's transportation, environment, and quality of life. In Canada, with the exception of a small percentage of the population living in public housing, approximately 95% of residents acquire their housing in the private market (Hulchanski, 2004, p223). This fact indicates the strong interdependence between urban residential mobility and real estate market, and further demonstrates the significant influence on regional economy. Dwellings in housing market are treated as commodities which usually occupy about two-thirds of one household's asset (Tracy et al., 1999). The enormous monetary allocation on housing of each household has an important effect on the expense of other goods, such as transport and living expenses, due to the financial constraint of each household. Meanwhile, new housing constructions and real estate transactions in market account for a large proportion in the economy, because of the huge investment and profit involved. The study of the housing market requires the development of tools, ideally simulation models, which are capable of capturing the diverse and dynamic nature of the urban real estate market. Such simulation models offer urban planners and policy-makers an efficient means for evaluating various urban development policies while enabling them to analyze the market under a specific policy scenario. In addition, as a fundamental component of an Integrated Urban Model (IUM), a housing market simulation model is capable of projecting the evolution of population and urban form as well as providing inputs for urban transportation and environmental analysis.

Using Hamilton, Ontario as the study area, researchers at the Centre for Spatial Analysis (CSpA) at McMaster University have been active in exploring and modeling the spatial processes underlying urban land use and transportation interactions. A variety of spatial models have been developed, including two prominent census tract based IUMs, namely IMULATE (Kanaroglou and Anderson, 1997) and IMPACT (Maoh et al., 2009), for the use of long-term simulation at a five-year interval. In recent years, with the advances in computer technologies and concerted efforts at micro-data collection, the application of microsimulation modeling in building urban models has attracted increasing attention from academic researchers and practitioners in planning institutions and industries. As for the microsimulation urban modeling efforts in the City of Hamilton, Maoh (2005) designed the framework for firmography and estimated the associated model components, and this model system is implemented by Yang (2011). On the other hand, a residential mobility microsimulation system was developed by Wang (2009), which mainly focuses on the household moves and associated location choices while the housing supply and housing prices are treated as exogenous.

Based on the above research context, this thesis estimates two models, namely, a real estate appraisal model and a new housing development model, which are perceived as endogenous components in the housing market microsimulation framework. The real estate appraisal model is developed to assess the market value of each specific private residential dwelling at a given year, so that the housing prices can be updated every year depending on the specific market. The price dynamics will affect the outcomes of behavioral decisions in the other model components, including residential mobility, new housing development, and also the location choices. The new housing development model is able to predict the outcomes of the housing development decisions in terms of whether to develop and which housing type to construct at each vacant residential parcel, which can affect the physical urban form by updating the spatial pattern of residential dwellings. With the inclusion of these two models, the dynamics of the whole simulation modeling system can be improved and strengthened.

1.2 Research Objectives

The aim of this thesis is the estimation of a real estate appraisal model and a new housing development model, which can predict the housing price dynamics and update the housing stock by incorporating new constructions respectively. To this end, the first stage is to design the general microsimulation framework for the residential housing market, based on the comparative review of current state-of-the-art microsimulation IUM systems. Then, as the two fundamental components in the simulation framework, the real estate appraisal model and the new housing development model are estimated. The key objectives of this research are outlined below:

1) Design a state-of-the-art microsimulation conceptual framework for the residential housing market of the City of Hamilton, Ontario.

2) Build the real estate appraisal model for the use of dwelling value assessment based on the comparisons of different modeling methodologies, including traditional hedonic modeling, spatial econometric modeling, and geostatistics modeling.

3) Build a nested logit new housing development model to simulate the housing stock dynamics, which is able to predict the outcomes of development decision and dwelling type decision for each vacant residential lot.

1.3 Thesis Structure

This thesis consists of five chapters. Chapter One describes the motivation and background of the research, and also the research objectives. Chapter Two first gives a brief introduction of the underlying processes in the housing market, and then reviews the current state-of-the-art housing market simulation frameworks that employ microsimulation modeling, followed by the review of the methodologies in housing price modeling and housing development modeling. Chapter Three starts with the description of the study area, the City of Hamilton, and then gives an overview of the data sources and derived variables that are utilized in this research; next, as a part of methodology, a microsimulation housing market model framework is designed and presented in this chapter, followed by the detailed description of modeling methods for the real estate appraisal model and new housing development model. Chapter Four presents the specifications and findings of the two above-mentioned models, including the interpretation of the results in detail; for the real estate appraisal model, the results of three models by using different modeling methods are presented and compared. Chapter Five summarizes the contributions of the thesis, and provides the direction for future research.

2. LITERATURE REVIEW

2.1 Introduction

The purpose of this chapter is to review the state-of-the-art housing market microsimulation models in the literature while focusing on two model components: real estate appraisal model and new housing development model. Besides, the explanatory modeling efforts for the above-mentioned two model components are reviewed in order to apply the up-to-date explanatory models in the simulation framework. This chapter starts with presenting the structure of housing market and the relevant underlying processes in section 2.2, and then moves to an overview of five microsimulation housing market frameworks in section 2.3. Followed by the detailed review of the real estate appraisal (or housing price) model in section 2.4 and new housing development model in section 2.5 in the five microsimulation frameworks, as well as the microscopic explanatory modeling efforts in these two components.

2.2 Housing Market Structure and Processes

There are two sides in the urban real estate market: housing demand side which refers to households wishing to be active in the market, and housing supply side which refers to vacant dwellings for sale in the market (Figure 2.1). The interaction of the two sides is driven by underlying housing market processes (i.e., household evolution, residential mobility and location choices, market transactions, and housing supply dynamics).

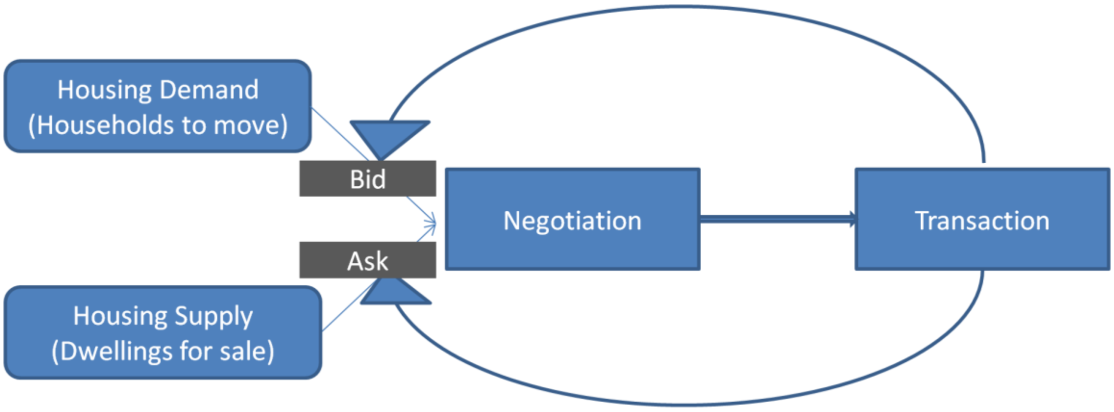


Figure 2.1 General scheme of urban residential housing market

(Simplified version from Filatova et al., 2009)

Two sources account for housing demand within a given housing market: new households (new migrants and new formed households) and existing households which have the desire to move into a more appropriate dwelling. The intra-urban residential mobility decision of a household would be influenced by many categories of factors, including demographic events (Rossi, 1955; Rabe and Taylor, 2010), neighborhood effect (e.g., Ioannides and Zabel, 2008; Rabe and Taylor, 2010), housing characteristics (e.g., Goodman, 1976; Diaz-Serrano, 2006), household income and economic status (e.g., Henley, 1998; Hassink and van Leuvensteijn, 2009), and moving cost effect (e.g., Nordvik, 2001; van Ommeren and van Leuvensteijn, 2005). One household would become active in the housing market if it has the intention to move, then a dwelling search process will be conducted. A choice set of potential appropriate dwellings will be generated in the end of searching. Housing search is a very complex process since it not only relies on household's preferences but is also influenced by the information sources (e.g., newspapers, friends, rental advertisements). In addition, past searching history and experience would also have an effect in this process (Habib and Miller, 2007). There are several methods to handle the spatial dwelling search process in modeling, such as imposing a nest searching structure on dwelling type and zonal choice (Hunt et al., 2010), generating an activity space where the moving household primarily conducts the searching (Blijie, 2005), or just employing a random selection (Habib and Miller, 2007; Miller et al., 2010). Location choices will be made within the choice set. Usually, utility maximization is assumed as the assumption in evaluating each dwelling. Zhou and Kockelman (2010) employed mixed logit modeling to rank the utilities of selecting each dwelling. Consumer surplus theory was introduced by UrbanSim (Waddell et al., 2003) and MUSSA (Martinez and Donoso, 2010) in modeling location choice decision. Habib and Miller (2009) utilized a reference-dependent approach with a focus on analyzing the gains and losses comparing to former residence by choosing a dwelling from the choice set. After the location choice decision was made, whether this household could acquire the dwelling or not depends on the market transaction process, which involves the negotiation of buyer and seller. In this process, the seller is assumed to choose the household with highest bidding price. Martinez and Henriquez (2007) modeled a random bidding process within an equilibrium framework. An auction-based bidding process was employed by Miller et al. (2010) and Zhou&Kockelman (2010). With the completeness of the transaction process, each dwelling sale price is finalized. Housing price, as an outcome of the market interaction process, will in turn have essential influences on the housing market activities. For dwelling seekers, the associated dwelling prices are relevant to the issue of affordability; for developers, the housing price level in one area determines the expected profit out of the potential new development projects. The literature on real estate appraisal modeling is reviewed in section 2.4.

On the other hand, the dynamics of housing supply also have an important impact on housing market activities, considering the direct influence on the supply-demand ratio. One source of supply dynamics is from the state transitions (e.g., demolition, renovation, split, combination) of existing dwellings, which can influence the dwelling property appraisal and the supply stock. The main source is from new construction, which happens either on vacant land or from the conversion of other land-use types (e.g., agricultural, commercial, industrial). In terms of new housing supply modeling, DiPasquale (1999) pointed out that most of the empirical literature has focused on aggregate data and very little effort has been expended at modeling the micro-level (e.g., parcel-level) land development process. On one hand, this is due to the fact that the process is very complicated and involves the multi-dimensional joint decisions of whether to develop or not, dwelling type, location choice, construction quantity, dwelling quality, and structural characteristics, such as number of bedrooms and bathrooms (Zhou and Kockelman, 2010; Farooq et al., 2010). In addition, the availability of appropriate micro-level data is another critical issue (DiPasquale, 1999; Haider and Miller, 2000), since the modeling not only requires the support of parcel-level land use data with detailed land use classifications, but also needs a variety of attribute data, such as site characteristics, location proximities, neighborhood characteristics, and market conditions. In the last decade, with the advancements in both data availability and modeling techniques, microscopic land development modeling has attracted increasing attention. The relevant modeling efforts are reviewed in section 2.5.

2.3 Urban Residential Housing Market Microsimulation Modeling Efforts

2.3.1. Microsimulation Modeling

Microsimulation modeling was firstly introduced by Orcutt (1957) who identified and represented individual actors in the economic system through the way in which their behavior changed over time. Microsimulation modeling has the capability to define the behaviors of individual agents and their interactions at disaggregated level. Macro-level pattern would be acquired as the outcome of micro-level interaction process. There are several advantages in using a microsimulation method over conventional aggregate modeling. Firstly, microsimulation modeling is more behaviorally sound than aggregate modeling. This is because the individual household is the actual decision-making unit playing role in the residential mobility process, and the aggregate pattern of whole population is generated from the activities made by all households and their interactions. Microsimulation modeling can explicitly distinguish the individual actors involved in the actual choice-making context by describing the characteristics of actors and the context; moreover, the context-specific behavior rules of actors are also defined and simulated (Miller, 2002). However, aggregate models are not capable of dealing with the above context, and usually only few categorized groups could be identified (Wegener and Spiekermann, 1996). In addition, there are heterogeneities and uncertainties existing in the decision-making process, based on specific characteristics and conditions of individuals. Microsimulation models can capture these heterogeneities because they model individual decisions with uncertainties, as well as their interactions through a bottom-up perspective. In contrast, aggregate models can only capture the macro-level patterns without distinguishing individual agents and disaggregated high resolution spatial units. Additionally, aggregate models are based on the spatial zonal system which assumes a homogeneous spatial distribution within the zone, incorporating heterogeneity bias into the models (Wegener and Spiekermann, 1996). In recent years microsimulation models have been increasingly applied to quantitative research in urban modeling.

2.3.2 Overview of Urban Residential Housing Market Microsimulation Models

Table 2.1 summarizes some basic information on the microsimulation-based housing market modules of five model frameworks. Among these , UrbanSim and Zhou&Kockelman are land use models, and the other three are integrated land use and transportation models, which explicitly have transport modules and establish the interdependence between land use and transportation. Some common modeling principles are employed in all five models, and other differences exist in each of them. Dynamic disequilibrium approach is taken by four models except Zhou&Kockelman, in order to represent the realistic housing market process; however, equilibrium is taken by Zhou&Kockelman to reduce the model computation complexity. Since agent-based microsimulation modeling is the simulation method, disaggregated unit

Table 2.1 Microsimulation Models of Urban Residential Housing Market

Model Name	Modeling Approach	Temporal & Spatial Unit	Treatment of Agents	Study Area
ILUTE (Salvini and Miller, 2005; Miller et al., 2010)	Random Utility Theory; Monte Carlo simulation; Dynamic disequilibrium bid-auction transaction; Hedonic analysis;	1 year; Dwelling	Individual; Household; Developer; Real estate agent	Greater Toronto Area, Canada
UrbanSim (Waddell, 2002; Waddell et al., 2003; Waddell and Ulfarsson, 2003)	Random Utility Theory; Monte Carlo simulation; Dynamic disequilibrium bid-rent theory; Hedonic analysis;	1 year; Land parcel and Dwelling or 150*150 meter grid	Individual; Household; Developer	Firstly fully implemented in Eugene-Springfield, Oregon; Now applied to around 20 urban areas in the world
ILUMASS (Moeckel et al., 2003; Moeckel et al., 2007)	Random Utility Theory; Monte Carlo simulation; Dynamic disequilibrium;	1 year; 100*100 meter grid and dwelling	Individual; Household; Developer	Dortmund, Germany
Zhou& Kockelman (Zhou and Kockelman ,2008; Zhou and Kockelman, 2010)	Random Utility Theory; Monte Carlo simulation; Dynamic equilibrium bid-auction market clearance;	1 year; Land parcel and Dwelling	Individual; Household; Developer	Austin, Texas, U.S.A
Oregon2 (Hunt et al., 2004; Hunt et al., 2010)	Random Utility Theory; Monte Carlo simulation; Disequilibrium market update by zone;	1 year; 30*30 meter grid and Dwelling	Individual; Household; Developer	Oregon, U.S.A

of time, space, and agents need to be explicitly defined. One-year time step is introduced by all five models, and at least three agents (individual person, household, developer) are distinguished, except that ILUTE defines one more agent (real estate agent) to monitor and gather real estate market generic information. About the treatment of spatial unit, land parcel is the analysis unit in modeling land development process in UrbanSim and Zhou&Kockelman. Grid unit is also introduced by UrbanSim, as well as ILUMASS and Oregon2. Moreover, individual dwelling is the transaction unit in the housing market for all five models to be allocated to one household which is looking for a new residence. Random Utility theory and Monte Carlo simulation are employed in each model to derive the probabilities of agents' choices (e.g., residential mobility decisions, residential location choices) under specific circumstances and calculate the choice outcomes respectively. Some economic rules are introduced in market transaction process; bid-rent theory is used in ILUTE, UrbanSim and Zhou&Kockelman; hedonic analysis is employed to simulate the housing price variations over time and across space in ILUTE and UrbanSim.

These five simulation models share a similar model structure, which consists of the following model components:

Demographic transition model. This submodel simulates the demographic transitions of individuals and households, and updates their attributes according to the transitions. Those demographic events include death, birth, education, marriage, and divorce. The changes in household structures brought by demographic transitions will impact the housing demand of associated households.

Residential mobility model. This model simulates household's decision about whether to move from its current residential location to a new location. A binomial logit model form is usually employed since this decision process involves two choice alternatives: stay or move.

New housing supply model. This model is developed to simulate the provision of new housing stock. Usually, the new development location and type are two main decisions that are considered in this model.

Residential location choice and market transaction module. This model is used to locate the suitable dwelling for each moving household based on relevant market rules. It starts with a location choice model which picks an appropriate candidate dwelling that the household would like to buy out of the available choice set, then employs a market transaction model to determine whether the dwelling seller would like to make this transaction happen.

Housing price model. The objective of this model is to estimate the market values of existing dwellings, based on the completed transactions in the last simulation period, in order to provide a basis for setting the asking prices before entering into the market.

2.4 Real Estate Appraisal Modeling

Housing price has a critical role in housing market, with the influence on both housing demand and supply activities. Hedonic price analysis is the most commonly used approach, which deconstructs the whole price into the implicit prices of each characteristic and usually uses different forms of Ordinary Least Square (OLS) estimation methods to determine the contribution of each component to overall price (Sheppard, 1999; Sopranzetti, 2010). However, traditional OLS hedonic price modeling cannot account for non-linearity and spatial effect which may have an influence on prices; therefore, several approaches were developed or applied into price modeling to overcome the shortcomings of the traditional hedonic approach (Brunauer et al., 2009; Paez, 2009; Osland, 2010). Haider and Miller (2000) applied spatial autoregressive (SAR) models to investigate the effects of transportation infrastructure and location on residential real estate values. Habib and Miller (2008) employed a multi-level model to address spatial-temporal heterogeneity in housing price modeling. Paez et al. (2008) applied a moving window approach into price estimation by comparing three methods (moving window regression, geographically weighted regression, and moving window kriging) to address market segmentation and spatial dependency issues. Similarly, Osland (2010) compared the performances of three modeling methods (geographically weighted regression, semi-parametric analysis, and mixed spatial Durbin model) against traditional hedonic approach in addressing issues of spatial dependency and spatial heterogeneity.

1) ILUTE

A multi-level hedonic price analysis is employed to estimate the asking price for each vacant dwelling for sale in the market (Habib and Miller, 2008). Spatial and temporal heterogeneity is incorporated into the price model by recognizing the spatio-temporal cluster which each dwelling exists in. At the top level, census tract is used as the unit in spatial cluster partition; the second level assumes that dwelling units are nested within spatio-temporal clusters that are obtained by dividing the spatial clusters through the identification of the quarter of the year. Multi-level modeling in this application clearly distinguishes between-cluster heterogeneity and heterogeneity among individual of the same cluster. Four groups of independent variables were introduced into the model specifications, including dwelling unit attributes, accessibility and location attributes, neighborhood attributes, and market variables. A set of non-spatial, spatial, and spatio-temporal models were built. Based on the model estimation results, the spatio-temporal model performs best in terms of model fit.

The above model can generate or adjust the asking prices for dwellings before entering into the market. In the market transaction process, the property prices will be adjusted based on the bidding intensity.

2) UrbanSim

In this model, the land price is influenced by two parts: the overall price level, and the relative price. The change in overall price level within the market for each real estate type is captured in response to shifts between aggregate supply and demand, which is executed in the end of market transactions. On the other hand, the traditional hedonic regression analysis is employed into relative price estimation to account for the price variations on attributes of the land and its environment. The independent variables influencing land prices are organized into four categories of site characteristics, regional accessibility, urban-design scale effects, and market conditions.

3) ILUMASS

The housing price is adjusted in the end of a simulation period in response to change in housing supply; however, details on the modeling methodology are not described in the literature.

4) Zhou&Kockelman

In this model, for the existing dwellings, the initial prices are exogenous and will be adjusted in the market; for the new developments, the improvement unit prices are determined by the developer model, which simulates the new housing supply outcomes classified by dwelling type, intensity (measured by floor-area-ratio) and quality (measured by improvement unit price per interior square feet). In the market clearing process, property prices will be adjusted in response to demand intensity at a fixed rate. A price range constraint is imposed on the adjusting process to ensure reasonable outcomes from bidders' competition. The maximum and minimum bid prices are determined as a function of land unit price, improvement unit price, and floor-area-ratio.

5) Oregon2

There are two modules within the Oregon2 simulation framework: household allocation which simulates the transitions and choices made by agents within one year, and land development module that microsimulates development transitions in 30m * 30m grids covering the model area. As a component in household allocation module, zonal residential space price is updated via a one-step adjustment based on the comparison between current vacancy rate in the zone and a reference vacancy rate value. The price will go up if current rate is lower than the reference value, and will decrease if higher, while the adjustment range is determined by a function of the current and reference vacancy rate values, the unit price, and the associated parameter.

2.5 New Housing Development Modeling

Profit maximization is taken as the fundamental assumption underlying the land development process, which refers to either the land developers or land owners pursue maximizing the net profit after subtracting the cost from the expected gross income. Therefore, in terms of modeling the outcomes of development decisions and the associated type choices at land parcels, the relevant data on prices and costs are necessary to be collected. However, due to the issue of data availability, in most cases, an extensive set of proxy indicators are introduced into the land development model instead of the price and costs of each parcel land. Generally speaking, five sets of variables, as seen in Table 2.2, have been employed in the following prominent models.

TABLE 2.2 Factors that have been employed in the microscopic Land Development models

Factors	References
Site Characteristics	
Spatial dependence influence	Mohammadian et al. (2008)
Distance to the nearest industry	Maoh et al. (2010a)
Distance to the lakeshore	Maoh et al. (2010a)
Recent built residential floorspace	Maoh et al. (2010a); Waddell and Ulfarsson (2003)
Zoning rules land use plan	Hunt et al. 2008; Waddell and Ulfarsson 2003;
Parcel area	Hunt et al. 2008; Zhou 2009
Existing space type and transition type	Waddell and Ulfarsson 2003; Hunt et al. 2008
Density of development	Hunt et al. (2008)
Age of development	Hunt et al. (2008)
Distance to developed parcel	Hunt et al. (2008); Waddell and Ulfarsson (2003);
Slope	Zhou (2009)
Transportation	
Road intersection density	Mohammadian et al. (2008); Haider and Miller (2004)

Distance to highway or major roads	Maoh et al. (2010a); Haider and Miller (2004) Waddell and Ulfarsson (2003); Zhou (2009)
Distance to highway ramps	Maoh et al. (2010a)
Proximity to subway	Haider and Miller (2004)
Residential Amenities	
School accessibility	Mohammadian et al. (2008); Maoh et al. (2010a)
Employment accessibility	Waddell et al. (2003); Mohammadian et al. (2008); Maoh et al. (2010a)
Distance to community centers	Maoh et al. (2010a)
Distance to the nearest woodlot	Maoh et al. (2010a)
Distance to open space	Maoh et al. (2010a); Haider and Miller (2004)
Distance to the shopping center	Maoh et al. (2010a); Haider and Miller (2004)
Proximity to CBD	Waddell et al. (2003); Haider and Miller (2004); Zhou (2009)
Retail density	Haider and Miller (2004)
Neighborhood Characteristics	
Residential inventory	Mohammadian et al. (2008); Waddell and Ulfarsson (2003);
Population density	Maoh et al. (2010a); Haider and Miller (2004); Zhou (2009)
Job density	Zhou (2009)
Changes of the average housing values	Maoh et al. (2010a)
Percentage of housing stock	Haider and Miller (2004)
Existing physical development	Haider and Miller (2004)
Units of development in the approval process	Haider and Miller (2004)
Percentage of land uses	Haider and Miller (2004); Waddell and Ulfarsson (2003);
Property values of housing type	Haider and Miller (2004)
Market Conditions	
Development charge	Mohammadian et al. 2008; Haider and Miller (2004)
Assessed market value of land and developed space on land	Hunt et al. (2008); Waddell and Ulfarsson (2003); Mohammadian et al. 2008; Zhou (2009)
Estimated demolition and construction cost	Hunt et al. (2008)
Vacancy rates	Waddell et al. (2003); Miller et al. (2010)

1) ILUTE

A New Housing Supply module is developed to simulate the provision of new housing stock by type and location. Two sub-models are used.

Housing Starts Model. This is a time-series econometric model to predict the total number of housing starts in the GTA in each year for four types of housing (detached, semi-detached, attached, and apartment). Monthly basis housing starts data are used in this regression model estimation.

New Housing Allocation Model. This model is developed to compute location choice probabilities in each zone by housing type and distribute new dwellings using Monte Carlo simulation based on location choice probabilities. Separate logit models for each type of housing are estimated (Haider and Miller, 2004).

Although the housing supply process in ILUTE is simulated in a zonal aggregate form, the outcomes of this module are delivered as the list of new individual dwellings in order to fulfill the needs of one-on-one based transactions in the market.

2) UrbanSim

The Real Estate Development Model of UrbanSim simulates several real estate construction events (new development, the intensification, conversion of existing development). Land parcels and 150*150 meter grid are used to construct a composite spatial representation. Twenty-five development types are classified by combining the number of housing units in each grid cell, the quantity of non-residential square footage in the cell, and the primary land use (residential, commercial, industrial, government, mixed residential/commercial, and vacant). After imposing the land use regulations in each cell, this multinomial logit (MNL) model predicts whether this cell will experience a development event or the specific type of event within the following year. Independent variables are organized into four categories, including site characteristics, urban design-scale, regional accessibility, and market conditions.

In the recent version of UrbanSim, a new mechanism in simulating land development process is designed as a top-down approach (UrbanSim, 2011). This model starts with a parcel evaluation process by considering the land use plan type and the remaining vacant land on the parcel, followed by an assessment process to determine whether this parcel fits one of the pre-defined development templates in UrbanSim. Next, the market values of proposed developments on the parcel will be calculated by using an Expected Sales Price Model which is the same model as of Real Estate Model. The Return on Investment (ROI) is computed as the result of expected sales price and the calculated development costs. In the end, a MNL model is estimated to sample the development projects by treating ROI as the profit. However, the details on model calibration process and model specifications have not been published yet.

3) ILUMASS

By considering the housing market supply-demand condition and profit expectation, the Residential Buildings component simulates three investment decisions of private developers: demolition, upgrading, and construction in a three-phase procedure. Supply-demand condition is evaluated in the exploration phase to estimate future profit expectation. If the developer believes that positive returns can be achieved from one investment type, the corresponding project will be planned in the decision phase. During this process, a zone and a cell are selected as the event occurrence site in consideration of land use regulations and location criteria. Then the project is executed in the implementation phase.

4) Zhou&Kocelman

In this land development component, the joint decisions (a combination of development type, intensity, and dwelling quality) are modeled in a multinomial logit form, following a similar methodology of the real estate development model in UrbanSim. Although the decisions about construction intensity and dwelling quality are continuous in nature, they are discretized into bins (low, medium, high) in order to operationally implement available estimation techniques (maximum simulated likelihood estimation). In addition, supply built space is coordinated to match expected demand based on likely growth rates, so that the new housing stock is determined. Different development behaviors for different constructions are revealed from the model results.

5) Oregon2

The Land Development (LD) module of Oregon2 updates the supply of residential space in response to the space price dynamics that are captured in the Household Allocation (HA) module. 30*30 meter grid is used as the spatial analysis unit to cover the study area. A nested logit model form is employed to represent the land development decisions: whether to implement a development and what type of development. Construction costs and rents, vacancy rates and associated uncertainties are the components of utility functions. It is to be noted that residential space prices are in the zonal form, while LD module is in a disaggregated form. To be able to include the space price in the LD model specification, prices as the outcome of HA module are disaggregated by considering the influences from factors including age, and development density in the grid (Hunt et al., 2008). The quantity of construction is determined by a sub-model, where the quantity value is selected from a uniform distribution with the maximum bound as of the maximum allowable floor area ratio in the zoning bylaws and the minimum bound as of zero for new construction and previous quantity value for additional construction.

6) Other microscopic explanatory models

A spatial multinomial logit (SMNL) model was developed by Mohammadian and Kanaroglou (2003) and applied into the dwelling type choice problem in the housing development process, in order to capture the spatial dependence effect among the observations. Multi-source data collected from January 1997 to April 2001 in the Greater Toronto Area (GTA) were used for model estimation. Four types of dwellings (Detached, Semi-detached, Apartment, and Others) were modeled against a series of independent variables including the price of the housing unit, municipal charge for the unit, street intersection density in the zonal area, school accessibility, employment accessibility, inventory of residential units, and the spatial dependence term. The spatial effect parameter is defined as a function of distances that evaluate the influence from the nearby housing projects with similar type choices. Throughout the comparison with a model in a MNL specification, the SMNL model achieves a better goodness-of-fit and the spatial dependence coefficient is demonstrated as significant statistically. By using the same dataset, Mohammadian et al. (2008) present a further model development in modeling type choice for new housing projects by incorporating the spatial dependence terms into a mixed logit (SML) model, which can allow each decision maker holds an unique set of utility coefficient and spatial dependency term parameters. The estimation results demonstrate that the SML model performs better than the SMNL model and the non-spatial MNL model.

Another prominent modeling effort conducted by Maoh et al. (2010a) was also contributed to the home builders' type choice issue by using the data covering from 1996 to 2001 in the City of Hamilton. Its purpose was to investigate the influence of the location factors on new housing projects. An extensive set of explanatory variables, mainly belonging to three categories including road infrastructure, residential amenities, and general site characteristics, were created and included in the model specifications. Similar to other studies, four alternative type-choices were modeled: detached, semi-detached, row-link, and condominium housing, and each type alternative has an associated model utility specification.

The above studies focus on one critical issue for new housing projects, i.e., which dwelling type to build at a given parcel of land, however, the question with respect to whether to build has not been captured. Based on the work of Maoh et al. (2010a), one purpose of this research is to integrate the two issues and model them simultaneously, so that fulfill the needs of capturing the housing supply dynamics in the simulation framework.

3 DATA AND METHODOLOGY

3.1 Study Area

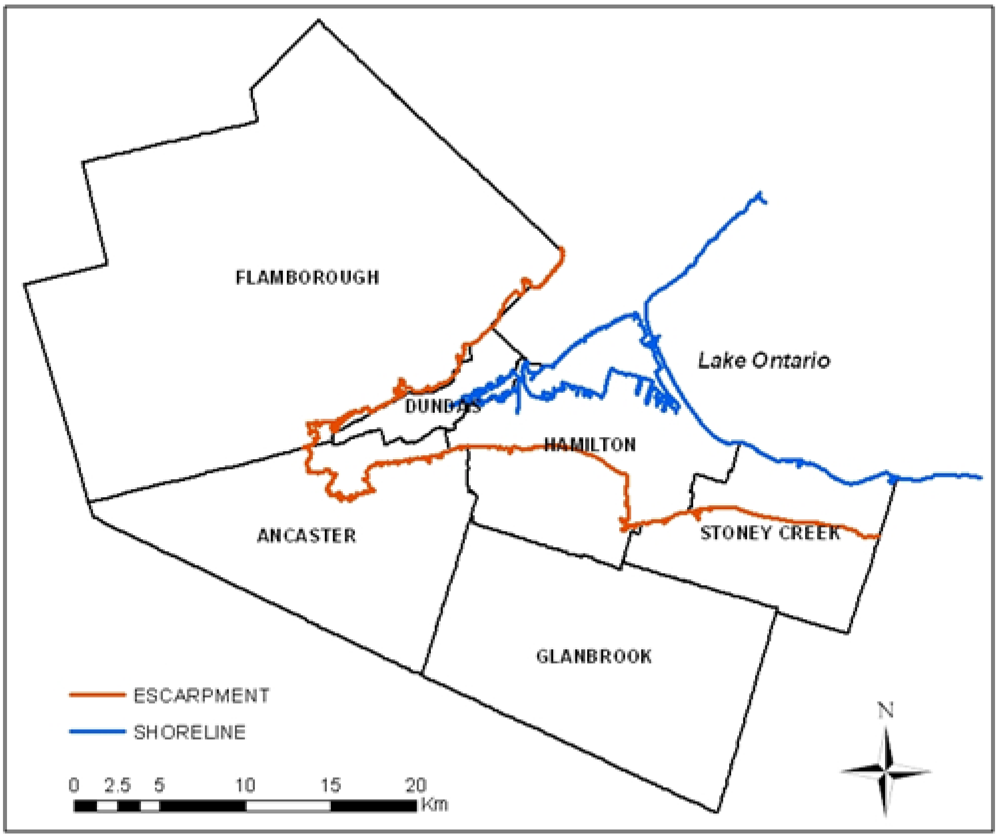


FIGURE 3.1 Location and Municipalities of the City of Hamilton

The area of study for this research, which is the City of Hamilton (see **Figure 3.1**), consists of six communities after the amalgamation in 2001: Hamilton, Ancaster, Dundas, Flamborough, Glanbrook, and Stoney Creek. As a middle-size industrial North American city, the City of Hamilton is increasingly experiencing urban sprawl. Over the period between 1986 and 2006, the population of Hamilton increased by 81,161 or 4,060 persons on average per year.Private dwellings increased by 38865, while the average value of dwellings in 2006 was almost three times as the price in 1986 (Statistics Canada). During the process, some historically agricultural centers surrounding Hamilton were experiencing urban development due to the outward growth of Hamilton (Spinney et al., 2009). In addition, the ongoing land development process has led progressively to a multinucleated city form with interdependence between land use activities at the nuclei (Maoh et al., 2010b). According to 2006 census, the approximate population, private households, and private dwelling units were 504,559, 140,810 and 194,455 respectively (2006 Canadian Census), making the City of Hamilton the ninth largest city in Canada.

3.2 Data

Data from multiple sources have been employed in the model estimation process other than designing the conceptual simulation framework.

1) ***Land-parcel GIS File***

Acquired from both the City of Hamilton and TERANET Inc., the land-parcel GIS dataset displays the land use state of each parcel in the City of Hamilton in 2003 and also records the details of the most recent land development in each parcel. 141,834 land-parcels which comprise the study area are described in the data with attributes including: location, current land use type of each parcel (general class, e.g. residential, commercial, industrial, etc., and sub-classification, e.g., single-detached house, office building, heavy manufacturing, etc.), the year of conversion to the current land use type, the floorspace created by this conversion, and the size of the parcel. Based on this rich dataset, residential developments and vacant residential land parcels in period from 1996 to 2001 were extracted as the base data for the new housing development model. On the other hand, data of dwellings which were sold in year 2001 are used for the real estate appraisal modeling. In addition, different variables were derived from this dataset, with the assistance of GIS software ArcGIS 10, such as the proximity measures to specific type of land uses, including industrial sites, schools, community centers, golf courses, university or college, shopping centers, woodlot, as well as the spatial relationship among the residential developed parcels.

2) ***Land Registry Transaction Data***

This dataset was obtained from TERANET Inc., which represents the population of real estate transactions in Hamilton between January 1995 and mid-September 2004. Details of each transaction are included: the registration date, sale price, and parcel details (e.g. area, perimeter, length, and ownership type). This dataset provides the sale price information for real estate appraisal model, while the land-parcel GIS file enables the use of spatial information in modeling.

3) ***DMTI CanMap***

The DMTI's CanMap Streetfiles dataset provides a comprehensive spatial representation of most populous areas across Canada, which contains the data for road infrastructures, points of interests, and topology. In this research, data on transportation networks and land use in year 2001 for the City of Hamilton are employed. The associated derived variables include proximities to highway or major roads, proximities to highway ramps, and percentages of specific land use types within a certain area.

4) ***Census of Canada***

Conducted every five years, the Census of Canada is an excellent tool for people to trace the changes and evolutions in almost all aspects of society, and also provides good quality data for research. The census tract profile data in year 1996 and the dissemination area profile data in year 2001 for the City of Hamilton were used for the new housing development model and the real estate appraisal model respectively. The population, the amount of dwellings, average value of dwellings, average household income, unemployment rate, education status, and migration information are extracted from the census data for the purpose of this research.

3.3 Designing the Urban Housing Market Microsimulation Conceptual Framework

3.3.1 Model overview

An agent-based microsimulation, disequilibrium, urban housing market model is designed in this research. Time within the model is accounted in one-year time steps and the parcel is used as the analysis unit for modeling the land development process. Three micro-level agents are identified: individual, household and land developer, and their behaviors are defined in the model. Details of model components are elaborated following the model conceptual structure in **Figure 3.2**.

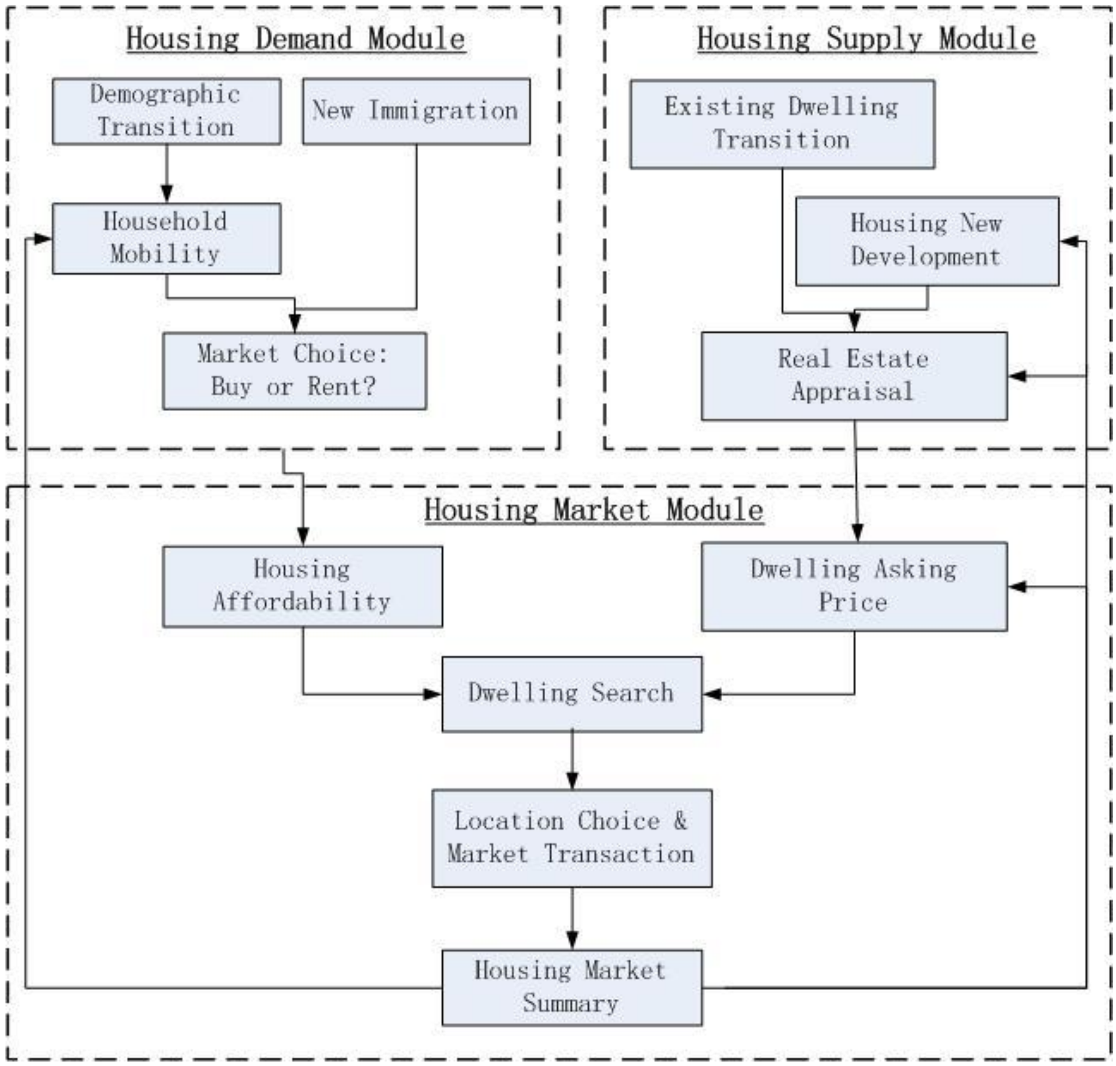


FIGURE 3.2 Housing Market Microsimulation Model Conceptual Structure

3.3.2 Housing demand module

This module is to pick out those households which have the intention to move into new residences, i.e., entities that are willing to enter the housing market.

3.3.2.1 Demographic Transition

Input: Each individual in existing households.

Process: The occurrence of life-cycle events (death, fertility, union formation, union dissolution, flat-mate formation, flat-mate leaving, nest leaving) will be simulated for each individual, as well as other events including job change and education status change.

Principle: The probability of each life-cycle event for a specific demographic group of people is derived from census data. Based on demographic characteristics of each individual in current life stage, also considering the sequence of events, Monte Carlo method will be utilized to simulate which event would happen to this individual. The probability of each life-cycle event for a specific demographic group of people can be derived from cross-tabulations of census data (Svinterikou, 2007; Wang, 2009). Based on demographic characteristics of each individual in current life stage, also considering the sequence of events, Monte Carlo method will be utilized to simulate which event would happen to this individual.

Output: Demographic characteristics of each individual would be updated in response to the simulated life-cycle event experienced in each simulation period; other information such as age and residence duration would also be updated.

3.3.2.2 Housing Mobility Decision

Input: A list of existing households along with their characteristics.

Process: Binomial logit modeling will be employed to analyze the mobility decision-making behavior of each household. Households would consider moving out if their living needs were not fulfilled by current residence due to relevant “push” and “pull” factors. Variables considered in this sub-model are demographic characteristics of household head (age, gender, race, education level, employment status) and household (size, size change, residence duration, and experienced life-cycle event),income condition (income class and change), current housing characteristics (maintenance cost, tenure, quality), neighborhood characteristics (average household income and change, employment level, accessibility, moving rate), housing market conditions (mortgage rate and change, transaction cost, average housing price and change, vacancy rate and change).

Output: In existing households, those which would like to move out will be picked out.

3.3.2.3 New Immigration (Exogenous)

Output: Immigrant households along with descriptions of household members.

3.3.2.4 Market Choice: Buy or Rent?

Input: Moving households.

Process: The market entry choice of each household will be simulated: do they want to buy or to rent a dwelling? The probabilities of home turnover (own-to-rent, rent-to-own) are given exogenously. Monte Carlo method will be then employed to simulate choice outcomes.

Output: The whole set of moving households will be classified into two groups: one for owner-based market, the other for rent-based market.

3.3.3 Housing supply module

This module is to identify all vacant dwellings for sale in the current period and new housing supply for sale in the next simulation period.

3.3.3.1 Existing Dwelling Transition

Input: Existing dwellings.

Process: Each dwelling is subjected to the probability of one of structural events (renovation, split, combination, conversion, demolition). Cross-tabulation probability of events for each category of dwellings will be calculated from collected data. Monte Carlo method will be used in simulation.

Output: Attributes of each dwelling (dwelling size, age, type, quality) would be updated based on experienced event. Existing vacant dwellings are also identified during the process. It is assumed that the dwellings of moving households would be available in the moment for this simulation period.

3.3.3.2 New Housing Development

Input: Vacant residential land parcels.

Process: In this sub-model, for each input parcel, the joint probability of residential development decision and associated type will be derived based on a two-level nested model structure. The top level concerns the construction decision that is whether the selected site would be considered for residential construction. The lower part of the nest concerns the decision of residential development type (detached, semi-detached, row-link, apartment) by using multinomial logit modeling, based on the previous work of Koronios (2009). Details are described in section 3.5.

Output: List of new developed dwellings will be generated for sale in next year with associated attributes of dwelling type, and location.

3.3.3.3 Real Estate Appraisal

Input: All existing dwellings in the current simulation.

Process: The market value of each dwelling will be assessed based on the sale prices in the previous simulation year. The dwelling value assessment is conducted by the spatial hedonic price analysis, which is not only able to explain the effect of dwelling characteristics on value variation, but also account for the effect of spatial dependence from previous sale prices in the neighborhood. Relevant details are followed in section 3.4.

Output: The assessed market value of each existing dwellings will be updated.

3.3.4 Housing market module

In this module, all housing transactions in the market are simulated, and the related housing market information emerging from market transactions are summarized in the end.

3.3.4.1 Household Affordability

Input: Households with mobility intention.

Process: The lower and upper bounds of the housing prices which the selected moving household could probably afford will be simulated based on current income category and current housing value. The limits of affordable housing price range are given exogenously as a percentage of current housing value taking account of the household's current income.

Output: Affordable housing price range for each moving household will be specified.

3.3.4.2 Dwelling Asking Price

Input: All vacant dwellings for sale in current simulation.

Process: The approximate asking price and acceptable bidding price range for each dwelling for sale will be specified. Based on assessed dwelling market value, the asking price for each dwelling will be estimated as a percentage of its value, while considering the influence of the ratio of housing demand and supply. If demand exceeds supply, which indicates an owner's market, the asking price will be specified as higher than value. In addition, the upper limit and the lower limit in acceptable bidding price range are given exogenously as percentages based on the asking price.

Output: Specified asking price and acceptable bidding price range for each vacant dwelling.

3.3.4.3 Dwelling Search

Object: Households with the intention to move out of their current residence.

Process: The potential dwelling choice set for each moving household will be located in this model. Two-stage process will be employed.

1) Stage 1: Rule-based dwelling filtering.

The first rule is based on area-constraint, which refers to the dwelling searching process conducted within the activity space for each household. This activity space is created by drawing a circle around the current residence, using a similar method as in Blijie (2005). The second rule is financial-constraint: only those dwellings within the housing affordability range by the household would be taken into consideration.

2) Stage 2: Logit-based process to identify appropriate dwellings.

In the list of vacant dwellings filtered by Stage 1, a randomly selected dwelling would be assumed as the residence of this household. Then the Housing Mobility Decision model will be employed to evaluate whether this household would like to continue living in this dwelling or move out. A number of dwellings (denoted as N) which this household could reside in will form the choice set for each household.

After experiencing the above two stages, if less than *N* dwellings could be found to be suitable for this household, the extent of the search would be broadened. The searching process returns to stage 1.

Output: The dwelling choice set for each moving household.

3.3.4.4 Location Choices and Market Transactions

Input: Each moving household and each for-sale dwelling.

Process: The moving household would offer its bidding price to the top priority dwelling in its choice set. Meanwhile, the owner of this selected dwelling would also evaluate offer(s) to decide whether to accept one offer. If the offer is accepted, the transaction will be executed, then this dwelling and moving household will leave the market. Meanwhile, associated attributes for these two agents will be also updated accordingly.

1) Stage 1: Rank the dwellings in the choice set.

The MNL model will be utilized here to calculate the probabilities of household's moving choices. Then dwellings in the choice set will be ranked by the probabilities. Factors influencing the rank include: dwelling characteristics (asking price, type, size, number of bedrooms, age, and quality), parcel characteristics (size, proximities), and neighborhood characteristics (commute times, employment level, percentages for land-use types).

2) Stage 2: Auction-based market transaction process.

The auction principle will be introduced in the transaction process. The dwelling with the top ranking in the choice set would be selected in the auction process. If more than one household makes an offer to the selected dwelling, this dwelling will be allocated to the household with the larger bidding capacity. Meanwhile, the transaction price will be specified as the upper bidding limit of the household with second largest bidding capacity. After the transaction process, those households which have not found matching dwellings would be subjected to the probability of out-migration. The other failing households would return to current (prior) residence.

Output: Real estate transactions and location choices outcomes for moving households.

3.3.4.5 Housing Market Summary

Input: Housing market transactions.

Process: Market characteristics and attributes can be concluded for each group of dwelling type and household type based on market transactions. This updated summary information will influence the behaviors of market agents in the next simulation period. Three categories of information can be summarized.

1) Housing price. The final transaction price of sold dwelling, the average price level for each housing category, and the associated housing price change by comparing to the price in the last period.

2) Housing demand. In each neighborhood and the whole urban area, the following values will be calculated: number of households which would like to move into new residence and number of successful moving households.

3) Housing supply. Similar to the summarized housing demand data in each neighborhood and the whole urban area, the following numbers will be calculated:

vacancy rate, number of each dwelling type for sale, and number of sold dwellings in each dwelling type.

Output: Housing market characteristics in each period.

3.4 Real Estate Appraisal: Incorporating the Spatial Variations

As described in section 3.3.3.3, the purpose of real estate appraisal model is to assess the market values of existing dwellings in order to propose the appropriate asking prices for the for-sale dwellings in the market. To this end, based on the review of relevant modeling approaches in the literature, a set of models with different specifications, i.e., traditional hedonic model, spatial regression model, and Kriging model, are estimated and compared.

3.4.1 Traditional hedonic modeling

In the hedonic model specification, the dwelling sale prices are reconstructed as the combinations of the implicit prices of each value-adding component, and the ordinary least squares (OLS) regression analysis is used to examine the exact contributions from each individual piece to the overall value (Sheppard, 1999; Sopranzetti, 2010). In this research, as a commonly used approach in the literature, the semi-log OLS regression model form is employed which gives marginal effects of independent variables, a convenient interpretation of percentage changes. Malpezzi (2008) summarized five five advantages in favour of using the semi-log functional form in hedonic modeling. First is that the marginal value of a specific independent variable is allowed to vary proportionately with the values of other variables. In addition, the heteroskedasticity problem is often mitigated by the semi-log form. Other advantages in using the semi-log form include easy computation, simple interpretation of model coefficients, and the flexibility in building model specification.

Three categories of explanatory variables are tested in the model specification, including dwelling attributes, location attributes, and the neighborhood characteristics. Equation 3.1 is the model specification for traditional hedonic regression.

$$\mathbf{LogP} = \boldsymbol{\beta_0} + \sum \boldsymbol{\beta_1 X_1} + \sum \boldsymbol{\beta_2 X_2} + \sum \boldsymbol{\beta_3 X_3} + \boldsymbol{\varepsilon} \qquad \mathbf{3.1}$$

Where the dependent variable is the natural logarithm of the dwelling sales price; β_0 refers to the constant parameter; β_1, β_2, β_3 and X_1, X_2, X_3 are the coefficients and independent variables for dwelling attributes, location attributes, and the neighborhood characteristics; ε denotes the error term.

3.4.2 Spatial effects

As seen from the model specification, traditional hedonic modeling is not capable of capturing the spatial effects in the process of housing price dynamics. Usually, there are two types of spatial effects, namely spatial heterogeneity and spatial dependence. Spatial heterogeneity refers to the variations in the properties (e.g., mean, variance, or covariance structure) associated with location specific sampling distributions of the outcome of a spatial process over space (Bailey and Gatrell, 1995; Buliung and Kanaroglou, 2007). The presence of spatial heterogeneity can lead to the non-stationary relationship between the outcome and explanatory predictors. The diagnostic of the existence of spatial heterogeneity in the context of housing price is usually associated with market segmentation. The specific structures of the process in the sub-markets need to be examined with respect to their differentiation and from the overall structure (Long, 2006). Spatial dependence refers to "the possible occurrence of interdependence among observations that are viewed in geographic space", and its presence violates the assumption of uncorrelated observations in statistical and econometrical modeling (Can, 1990). It occurs when individuals in population are related based on locations, and/or individuals residing in proximate locations make similar choice decisions (Anselin, 1988; Mohammadian et al., 2005). In this research, the sale price of a dwelling is usually affected by the prices of previous sales in the neighborhood, showing a cluster pattern which indicates the existence of the similarity. Moran's I is the common approach of the diagnostics of spatial autocorrelation (Anselin, 1988; Osland, 2010).

$$\mathbf{I} = \frac{\mathbf{N}}{\sum_i \sum_j \mathbf{W_{ij}}} * \frac{\sum_i \sum_j \mathbf{W_{ij} E_i E_j}}{\sum_i \mathbf{E_i E_i}} \qquad \mathbf{3.2}$$

In the expression 3.2, I denotes the Moran's I statistic; E refers to the residuals between the observed and estimated values; W is the spatial weight between two observations; N is the sample size.

The values of Moran's I range from -1 to +1. Negative (positive) values indicate negative (positive) spatial autocorrelation, and a zero value reveals a random spatial pattern.

3.4.3 Spatial regression modeling

One common approach to incorporate spatial dependence effect into the traditional OLS model is the spatial regression modeling (Anselin, 1988). Usually two model specifications, namely spatial lag model and spatial error model, are utilized based on different assumptions. For housing price modeling, spatial lag model (Equation 3.3) assumes that spatial autocorrelation exists in the housing prices, i.e., the price of a given dwelling is directly influenced by the prices of adjacent dwellings, while spatial error model (Equation 3.4) separates the spatial autocorrelation component (denoted as the $\lambda W\varepsilon$ in Equation 3.4) from the unobserved part which usually consists of omitted variables and measurement errors.

$$\mathbf{P} = \rho \mathbf{W}\mathbf{P} + \mathbf{X}\beta + \mathbf{u} \qquad \mathbf{3.3}$$

$$\mathbf{P} = \mathbf{X}\beta + \varepsilon; \ \varepsilon = \lambda \mathbf{W}\varepsilon + \mathbf{u} \qquad \mathbf{3.4}$$

Where P means the price; W denotes the spatial weight matrix over neighboring observations; ρ is the spatial lag coefficient; λ refers to the spatial error autoregressive parameter; u and ε are the error terms, while u is independently distributed.

Three types of spatial weight matrix are constructed based on specific spatial relationships: K-nearest neighbors, distance-based, and contiguity-based. Since the sample observations are point distributed, a Delaunay TIN is generated from the GIS layer of prediction sample so that the contiguity for the observations to be inferred. Furthermore, to be able to test the model sensitivity to the choice of spatial weight matrix, a group of weight matrix is built for each matrix type.

The Lagrange Multiplier (LM) Tests are frequently used to assist the choice of using spatial lag model or spatial error model if the existence of spatial effects in the data is demonstrated. Accordingly, the LM-lag statistic tests whether the spatial autocorrelation is statistically significant effective in the dependent variable, while the LM-error statistic tests the existence of significant spatial error autocorrelation (Osland, 2010).

3.4.4 Kriging combined with regression modeling

Kriging refers to a suite of spatial interpolation techniques including Universal Kriging, Ordinary Kriging, Co-Kriging, Kriging with External Drift and Regression Kriging among others. It can be used for spatial prediction which can generate the surface of the targeting process covering the whole space based on known observations. Kriging uses the covariance structure of the spatial observations to represent the spatial dependence. As the two most common methods in Kriging family, Universal Kriging and Ordinary Kriging implement the interpolations based on the coordinates of the observations. The difference between these two approaches is that Universal Kriging assumes there is a spatial trend in the data that can be modeled using a deterministic function, while Ordinary Kriging assumes the trend can be represented by a constant (global mean) over the study area (Spinney, 2010, p100). Regression Kriging divides the value of the target variable into a deterministic part and a stochastic part; for the deterministic component, regression analysis is used to fit, while the residuals from the regression model are interpolated by Ordinary Kriging (Hengl et al., 2003). In this research, the residuals extracted from the OLS hedonic model are used as the base dataset for interpolation by using Ordinary Kriging.

After the interpolation, the predicted dwelling value at a location with specific attributes is the sum of the outcome of the hedonic model and the interpolated residual generated by Kriging.

3.4.5 Model validation and comparison

Following the methodology in Paez et al. (2008), a group of indicators are employed for validation and comparison. An out-of-sample validation is conducted which randomly selects 10% of the whole population as the validation sample, while the rest of the observations enter into the model estimation pool. To ensure that the selected validation sample is representative of the estimation sample, the sample means difference test is employed. On the other hand, the estimation results of different models are compared from three aspects: the model goodness-of-fit, the explanatory power, and the prediction accuracy. Log Likelihood and AIC are employed in describing the model fit. The statistical significance and signs of independent variables are considered to assess the models' explanatory power. In terms of model prediction accuracy, the following indicators are used: the square root of the variance of the residuals (RMSE), the mean absolute error (MAE), the correlation parameter of the observation values and the predictions, and three percentages indicate whether the prediction value is within 5%, 10%, and 15% away from the observation value.

3.5 New Housing Development Model: A Nested-Logit Form

As the previous work, Koronios (2009) conducted a thorough exploratory spatio-temporal study of the land development process in the City of Hamilton and built a type choice model in housing development process by employing the multinomial logit (MNL) model. In order to operationalize the housing supply module in the microsimulation framework, the model in this section expands the model of Koronios (2009) and is capable of dealing with a two-tier decision-making process: whether to develop and which dwelling type to develop for a specific vacant residential land parcel. Nested-logit (NL) discrete choice modeling is employed which is consistent with the conceptual structure of the research issue. Another reason of using NL model form is that NL relaxes the limitation of Independence from Irrelevant Alternatives (IIA) assumption in MNL by grouping the correlated alternatives into an independent nest.

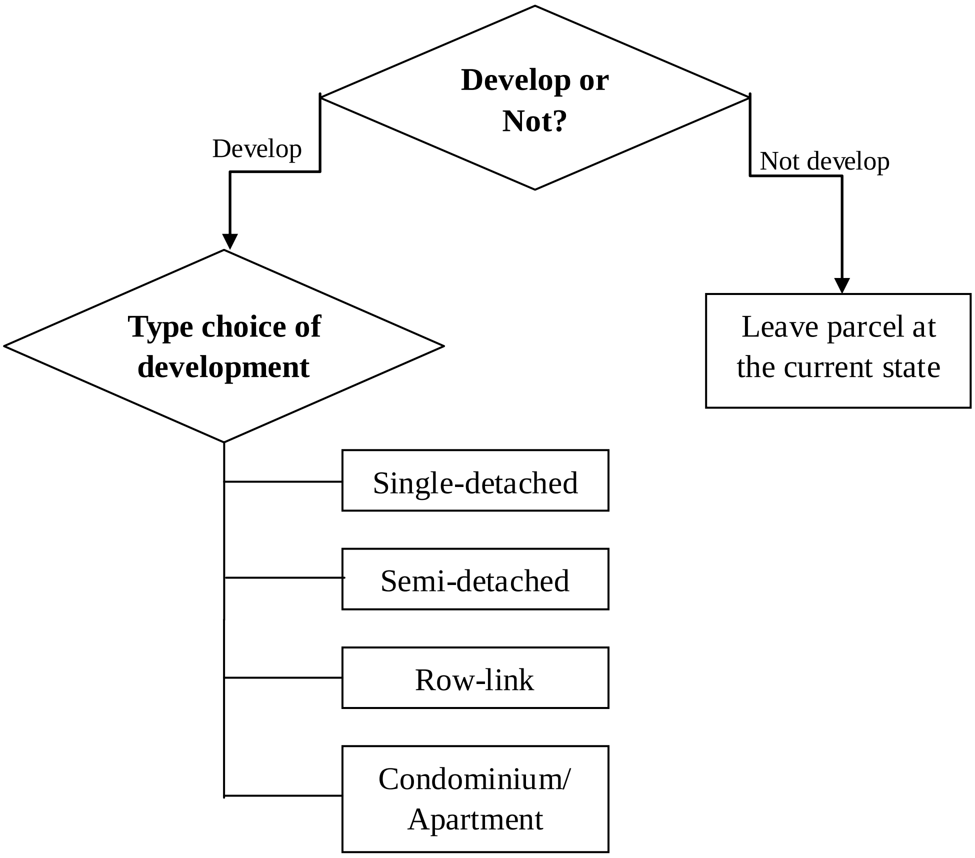


FIGURE 3.3 The Two-Tier Structure of the Nested-Logit New Housing Development Model

As seen in **Figure 3.3**, the top-tier of the model concerns with the developer's decision about whether to develop a specific vacant residential land parcel. The branch of non-development is a degenerated nest which has one leaf alternative, while the branch of development deals with the type choice issue which consists of four alternatives: single-detached housing, semi-detached housing, row-link housing, and condominium or apartment.

The strategy in selecting the explanatory factors follows the experience in the literature (see **Table 2.3**), more importantly from Koronios (2009), and also is based on the available data sources. Four categories of factors are investigated and included in the model specification, including site characteristics, proximity to transportation infrastructures, residential amenities, and neighborhood characteristics. The utility for an alternative ***i*** for the developer ***n*** can be expressed in Equation 3.5:

$$\mathbf{U}_{\boldsymbol{ni}} = \mathbf{V}_{\boldsymbol{ni}}(\mathbf{X},\ \boldsymbol{\beta}) + \boldsymbol{\varepsilon} \qquad \textbf{3.5}$$

Where ***V*** denotes the observed component; ***X*** refers to the vector of independent variables, $\boldsymbol{\beta}$ is the associated vector of parameters; $\boldsymbol{\varepsilon}$ is the unobserved component.

Formula 3.6 gives the choice probability for alternatives ***i*** in nest ***k*** (Train, 2009, p80).

$$\mathbf{P_{ni}} = \frac{\mathbf{e^{V_{ni}/\lambda_k}\,(\sum_{j \in B_k} e^{V_{nj}/\lambda_k})^{\lambda_k - 1}}}{\mathbf{\sum_{m=1}^{K}(\sum_{j \in B_m} e^{V_{mj}/\lambda_m})^{\lambda_m}}} \qquad \textbf{3.6}$$

Where $\boldsymbol{\lambda_k}$ is the measure of the degree of independence in unobserved utility among the alternatives in nest ***k***.

With regards to model estimation method, full information maximum likelihood estimation is used instead of sequential estimation, since the former approach can generate more efficient unbiased estimation results. The open source discrete choice modeling software Biogeme, which follows the Utility Maximization Nested Logit (UMNL) specification and uses the simultaneous estimation, is employed.

4. RESULTS

Following the data and methodology in Chapter three, the real estate appraisal model and the new housing development model are built and the associated estimation results are presented in Sections 4.1 and Section 4.2, respectively.

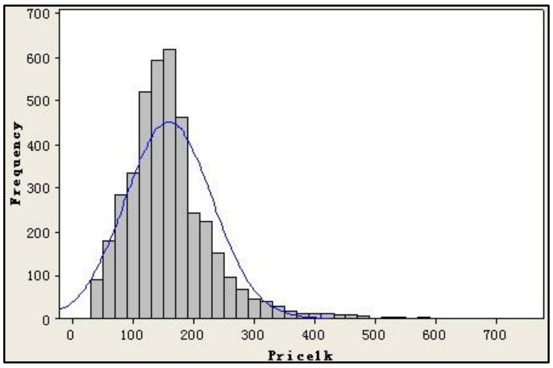
4.1 Estimation Results for Real Estate Appraisal Model

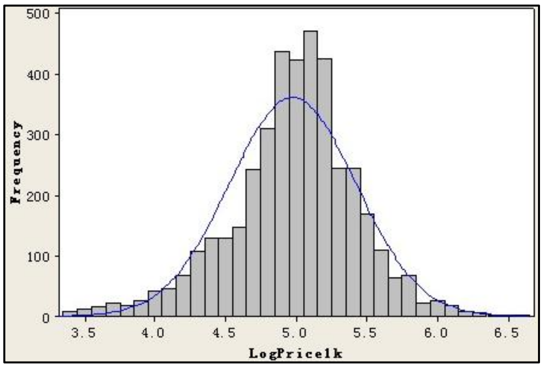
The definitions and associated descriptive statistics of variables that are included in the final model are presented in Table 4.1.

TABLE 4.1 Definitions of Variables and Descriptive Statistics

Variables	Description	Mean /Proportion	Standard Deviation
	Dependent Variable		
Price1k	Sales price (in 1,000 CAD$)	158.25	76.47
LogPrice1k	Natural log of sales price	4.94	0.57
	Dwelling Attributes		
LotSize100	Lot size (in 100 sqft)	66.99	74.51
FSize100	Unit size (in 100 sqft)	14.58	5.91
Age	Age of the dwelling (years)	45.38	31.6
DuDetach	Dummy: 1 if dwelling is single-detached	89.42%	
	Location attributes		
DuHA	Dummy: 1 if in the community of Hamilton	64.94%	
DuFL	Dummy: 1 if in the community of Flamborough	10.19%	
DuOpen500	Dummy: 1 if within 500m from nearest open space	11.34%	
DuIndt500	Dummy: 1 if within 500m from nearest heavy industrial site	6.25%	
DuUniv500	Dummy: 1 if within 500 from the nearest university/college	2.76%	
	Neighborhood attributes (Dissemination Area)		
EduBlUniv	Percentage of people with education not higher than high school	43.29	16.24
EduAbUniv	Percentage of people with education higher than a Bachelor's degree	1.88	2.02
UnEmp	Percentage of unemployment	5.81	4.56
RecLand	Percentage of recreational land use	2.89	7.96
	N (Estimation sample size)	4064	

From the Land Registry Transaction Data, the transactions of private-owned individual dwellings in year 2001 were extracted, followed by a series of data processing before entering in the modeling pool. In order to reduce the effects of outlier sales, the following filtering criteria were implemented: excluding the sales with price below 30,000 CAD or with the lot size over 1.5 acre (65,340 square feet). The distributions of native sales prices after filtering and the natural logarithm of the prices are presented in Figure 4.1.





(a) **(b)**

FIGURE 4.1 Distributions of Dependent Variables with the Normal Distribution Curve

(a- distribution of sales prices; b-distribution of natural logarithm of the dwelling sales price)

Next, a 10% validation sample is randomly extracted and held out for the use of testing the predictive accuracy. T-test is employed to assess whether the means of two samples are statistically different from each other with the null hypothesis that the variables in two samples do not have different means. The results in Table 4.2 do not return significant statistics, which indicates that the validation sample is representative of the estimation sample. A total of 4,596 observations are retained in the final basic dataset, with 4,064 records in the estimation sample and 532 in the validation sample. The distribution of the observations is displayed in Figure 4.2.

TABLE 4.2 Statistics of Estimation and Validation Sample

				T-test for Equality of Means	
Variables	**N**	**Mean**	**Std. Deviation**	**T-Stat**	**Prob. (two-tailed)**
Price1k	4064	159.33	71.63	0.738	0.460
	532	156.90	69.99		
LogPrice1k	4064	4.97	0.45	0.564	0.573
	532	4.96	0.44		
Age	4064	45.38	31.60	0.334	0.738
	532	44.89	30.93		
LotSize100	4064	66.99	74.52	0.911	0.362
	532	63.89	69.51		
FSize100	4064	14.58	5.91	0.843	0.399
	532	14.35	5.61		

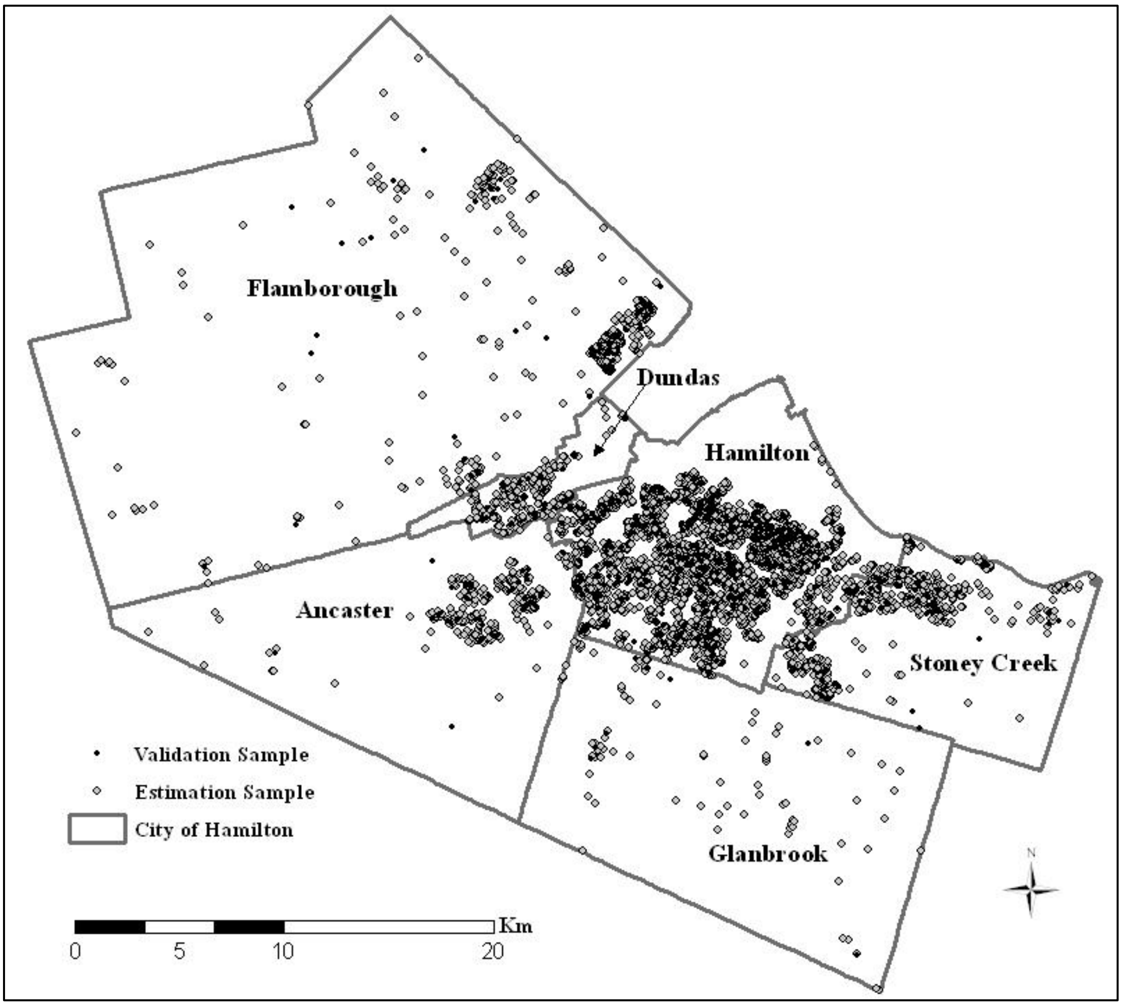


FIGURE 4.2 Housing Price Observations within the City of Hamilton, 2001

4.1.1 Traditional hedonic model results

The estimated coefficients for the independent variables and the overall model parameters for OLS model are presented in Table 4.3. In general, the model reaches a relatively high (0.66) goodness-of-fit, which means that about 66 percent of the variability in the dependent variable is accounted for by the model. In addition, the signs of the independent variable coefficients turn out to be consistent with intuition. The influence of each variable on the change of housing prices is described below in the context of all other factors being equal. It is noted that the coefficient represents the price elasticity brought by the specific independent variable, since a semi-log model form is employed in this research.

Specifically, the areas of land property and dwelling are positively correlated with the sale price; furthermore, one percentage point change in dwelling size has more influence on the percentage change in sale price than one unit change in land property size, as indicated by the coefficients of FSize100 (0.028) and LotSize100 (0.00055). This is consistent with common sense that the property of the land is valued more than the land itself because of the investment of housing construction. Age has a negative impact on price, which confirms the expectation that housing prices decrease with age. Single-detached dwellings have higher prices than other types on average. With respect to the location effects, dwellings in the Community of Hamilton pull down the price level, while houses in Flamborough are relatively more expensive, if all other factors are equal. Additionally, dwelling prices are higher if close to the university, colleges or open area, and lower for houses that are closer to heavy industries. As for the neighborhood influence, dwellings in a neighborhood with more recreational land use or having more advanced educated residents have a higher price potential; on the other hand, a neighborhood with a higher unemployment rate or high concentration of population with low education has negative impact on the housing price level.

TABLE 4.3 Estimation Results of Traditional Hedonic Model (OLS) and Spatial Autoregressive (SAR) Model

Variables	**OLS**	**SAR** (K=9[1])
Dependent Variable	Log(Price)	Log(Price)
Dwelling Attributes		
LotSize100	0.00055***	0.00033***
FSize100	0.028***	0.025***
Age	-0.0047***	-0.0035***
DuDetach	0.19***	0.17***
Location attributes		
DuHA	-0.044***	-0.0021
DuFL	0.073***	0.041**
DuOpen500	0.042**	0.038**
DuHIndt500	-0.15***	-0.084***
DuUniv500	0.082**	0.057*
Neighborhood attributes(Dissemination Area)		
EduBlUniv	-0.56***	-0.28***
EduAbUniv	0.82***	0.55**
UnEmp	-0.003**	-0.003**
RecLand	0.18***	0.14**
Constant	4.83***	3.01***
Rho		0.34***
Adj. R-square	0.66	
Log Likelihood	-320.5	-176.42
AIC	671	384.4
N(Total number of observations in modeling sample)	4064	4064

Note: 1- K is the number of nearest neighbors that are used in constructing the spatial weight matrix

***- Significant at 0.001 level

** - Significant at 0.01 level

* - Significant at 0.05 level

4.1.2 Spatial regression model results

Spatial regression models are specified and compared with the benchmark OLS model. Moran's I is employed as the diagnostic test of global spatial autocorrelation in the residuals of OLS model. Meanwhile, three types of spatial weigh matrices are generated in order to test the sensitivity of spatial autocorrelation to the choice of spatial market segmentation.

Figure 4.3 presents the relevant Moran's I statistics that are all statistically significantly different from zero. For K-nearest neighbor matrix, with values of K range from 1 to 10 corresponding to the dots from left to right in Figure 4.3. For the

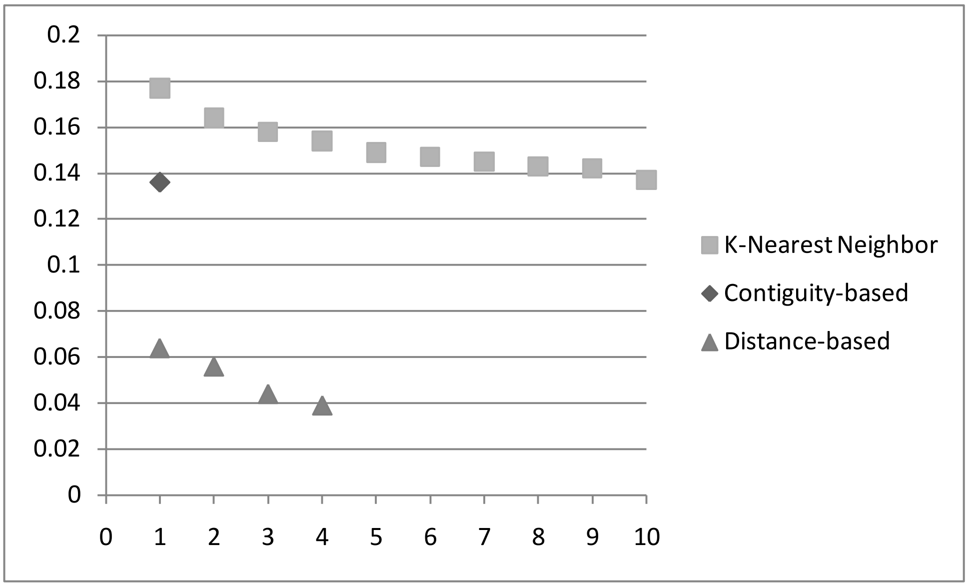


FIGURE 4.3 Moran's I Statistics Based on Different Types of Spatial Weight Matrix.

contiguity-based matrix, rook and queen contiguities are both utilized. For distance-based matrix, the threshold distance is 3,712m which ensures that each observation has at least one neighbor. Four distance bands are used to create the spatial weight matrix, including 3,712m, 4km, 4.5km, and 5km, and they are plotted as the four distance-based dots in the Figure 4.3 from left to right. As seen in Figure

4.3, k-nearest neighbor type of spatial weight matrix captures more spatial autocorrelation existing in the OLS residuals than other types; therefore, it is employed in building the spatial regression models.

In order to choose between spatial lag model and spatial error model, as described in Section 3.4.3, the Lagrange Multiplier (LM) tests are used as a rule of thumb. On the other hand, model goodness-of-fit measurement (AIC) and the model prediction capability are also taken into account. In terms of LM-tests, in most cases, both LM-lag test and LM-error test return significant results, i.e., these two model specifications are demonstrated to be able to produce reliable results. On the other hand, with respect to the implementation of spatial prediction, spatial lag model is much more straight-forward than spatial error model, particularly in the simulation context. This is because, for spatial error model, the error component is unobserved during the prediction so that the spatial autocorrelation component is hardly able to be incorporated into the prediction value, while for spatial lag model, the spatial autocorrelation effect is brought by the prices from adjacent sales observations (Yu et al., 2007). Considering the fact that both models produce comparable AICs based on the same spatial weight matrix, as seen in Figure 4.4, the spatial lag model specification is chosen. Furthermore, the number of nearest neighbors is set as nine because of the smallest AIC being achieved as presented in Figure 4.4.

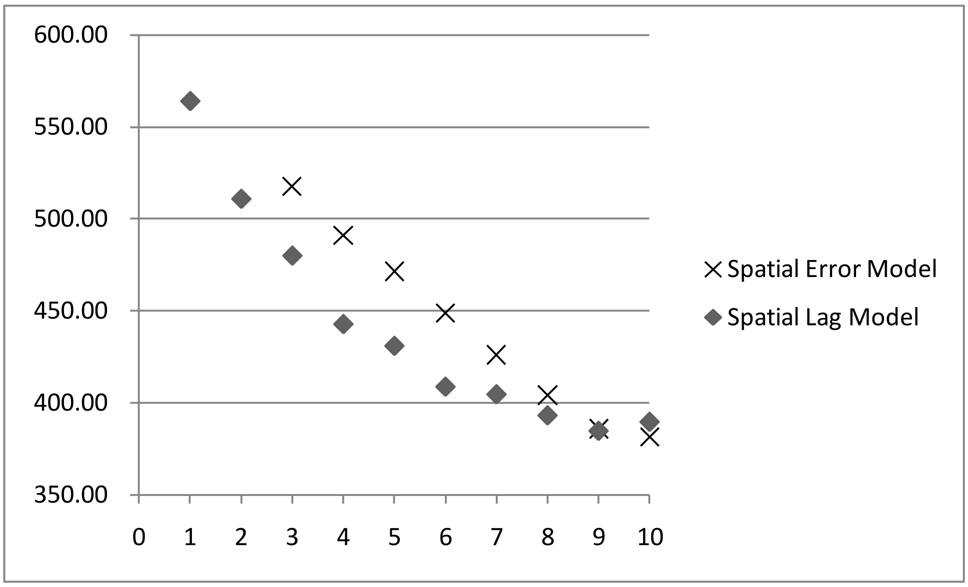


FIGURE 4.4 AIC Statistics of SAR Models Based on K-Nearest Neighbor Weight Matrix (X axis shows the number of nearest neighbors in the spatial weight matrix; Y axis displays the associated model AIC values corresponding to each SAR model specification based on specific weight matrix)

The estimation results of SAR model are presented and compared against OLS model results in Table 4.3. The comparison illustrates stable coefficients for the independent variables in explaining housing prices. On the other hand, a certain level of improvement was achieved over the OLS model, as demonstrated by the AIC value which is reduced to 384.4 for the SAR model from 671 for the OLS model. In addition, the reduction in the value of the constant also points out that the explanatory power of SAR model is higher than OLS hedonic model. Besides, the spatial autocorrelation coefficient (**ρ**) is significant, which further verifies the necessity of employing a spatial model.

4.1.3 Model validation and comparison

Table 4.4 provides the validation results for three models, namely OLS model, SAR model, and Kriging model with combination of regression. As described in Section

3.4.4, the Regression Kriging model is composed of two procedures: OLS regression, and Kriging prediction by taking the OLS residuals as the input data. Therefore, the coefficients for independent variables are same as the ones in OLS model, while the prediction values are different from OLS model.

TABLE 4.4 Out-Of-Sample Model Validations

	RMSE	MAE	Correlation	15p	10p	5p
OLS	43.92	26.29	0.79	0.61	0.41	0.23
SAR	40.84	24.42	0.82	0.62	0.42	0.26
Regression Kriging (RK)	42.81	22.96	0.81	0.67	0.46	0.24

Through the model comparisons, we can observe that spatial models (SAR, RK) perform better than non-spatial (OLS) model in terms of prediction accuracy. As indicated in Table 4.4, the values of RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) in spatial model validation results are both lower than the value in OLS model. On the other hand, the correlation parameters between predictions and observations in spatial model results are higher than the parameter in OLS model. In terms of the comparison between SAR and RK, these two models achieve comparable results. SAR model reach better numbers in RMSE, Correlation, and 15p (proportion of the predictions within 15 percent deviation from the validation prices), while RK has the best results in MAE and two correlation measurements, i.e., 15p and 10p. In the end, SAR model is proposed to be incorporated as the real estate appraisal component of the simulation framework, because in general the predictions out of SAR model are closer to the validation values, as indicated by the highest correlation parameter.

4.2 Estimation Results for New Housing Development Model

A two-tier nested-logit model, is successfully estimated that deals with two essential decisions during the housing development process: whether to develop and which dwelling type to build on a specific vacant land parcel. The type choice among four alternatives (single-detached house, semi-detached house, row-link house, and condominium/apartment) was modeled by Koronios (2009) using the multinomial logit model and the same parcel data. This nested-logit model brings new insights with respect to the development decision, and integrates these two decision-making components as a whole so that being feasible to be applied in the simulation framework.

In total, there are 10,643 observations in the dataset which describe the basic and derived attributes of residential developments (6,494 observations) and vacant residential land parcels (4,149 observations) during 1996-2001 in the City of Hamilton. A summary of the dataset is given in Table 4.5.

TABLE 4.5 Data Summary for New Housing Development Model

	Count of Land Parcels	**Aggregated Area (Acre)**
Residential Developments	6494	1,531.72
Single-detached	5441	1,297.98
Semi-detached	821	56.20
Row-link	163	11.23
Condominium/Apartment	69	166.31
Vacant Residential Parcels	4149	11,129.50
Total	10,643	12,661.22

Four groups of independent variables are retained in the final model: proximities to transportation infrastructures, residential amenities, site characteristics, and neighborhood characteristics. Distances to major roads and highway ramps are employed as the measurements on the travel convenience of the parcel. Residential amenity variables include the proximities to nearest recreational land uses and neighborhood shopping centers. In terms of site characteristics, the closeness to the nearest industrial site, the aggregated area of open space, residential land, recreational land, and the dwellings of the same development type within the local area (1.5km from the centroid of the parcel) are calculated. Neighborhood characteristics include the census tract data on migration, average dwelling value, and the proportion of each housing type.

TABLE 4.6 Estimation Results of New Housing Development Model

Covariate	**Definition**	**Alternative**	**Beta**	**T-stats**
ASC_D	Alternative specific constant for D	D	-4.25	-16.74
ASC_S	Alternative specific constant for S	S	-4.79	-14.68
ASC_R	Alternative specific constant for R	R	-4.49	-16.62
ASC_C	Alternative specific constant for C	C	-4.46	-14.46
Transportation				
B0_RoadG1k	Proximity to major roads; 1 if >=100m	N	0.780	7.30
B1_Road200	Proximity to major roads; 1 if <200m	D	-0.101	-3.69
B2_Road400	Proximity to major roads; 1 if <400m	S	0.367	3.23
B3_Road100	Proximity to major roads; 1 if <100m	R	0.0889	3.16
B4_Road600	Proximity to major roads; 1 if <600m	C	-0.132	-2.04
B1_Ramp200	Proximity to highway ramps; 1 if <200m	D	-0.212	-3.94
B3_Ramp800	Proximity to highway ramps; 1 if <800m	R	0.181	4.10

Residential Amenities				
B0_RecG1k	Proximity to parks, golf course and recreational land uses; 1 if >1000m	N	0.289	4.90
B3_Shop400	Proximity to neighborhood shopping centers; 1 if <400m	R	0.209	4.29
Site Characteristics				
B0_Indt200	Proximity to industrial sites; 1 if <200m	N	1.07	12.37
B1_Indt800	Proximity to industrial sites; 1 if <800m	D	-0.137	-3.53
B2_Indt400	Proximity to industrial sites; 1 if <400m	S	-0.213	-2.72
B0_OpenLU1k5	Area of open space within 1.5km from the parcel	N	-0.269	-9.57
B0_RecLU1k5	Area of recreational land within 1.5km from the parcel	N	-0.785	-7.51
B0_ResLU1k5	Area of residential land within 1.5km from the parcel	N	-0.444	-11.53
B4_CD9196	Aggregated area of recently built apartments in 1991-1996 within 1.5km from a specific land parcel	C	0.434	3.52
Neighborhood Characteristics				
B2_SemDpt	Percentage of semi-detached houses in census tracts, 1996	S	1.91	3.49
B0_Move5y	Percentage of population of a census tract in 1996 that moved in during 1991-1996	N	-2.52	-11.28
B1_IntraMove5y	Percentage of population of a census tract in 1996 that moved in from other census tracts within the city during 1991-1996	D	1.43	4.53
B1_AvHValue	Average dwelling value at census tract level in 1996	D	0.773	9.69

B2_AvHValue	Average dwelling value at census tract level in 1996	S	0.571	5.98
B3_ AvHValue	Average dwelling value at census tract level in 1996	R	0.708	8.88
B4_ AvHValue	Average dwelling value at census tract level in 1996	C	0.488	4.08
μ	Coefficient for the inclusive value of the development nest		0.20	4.28
n	Number of observations		10643	
L0	Init log-likelihood at zeros		-17129.248	
LC	Log-likelihood constants only		-10691.420	
L*	Final log-likelihood at convergence		-9413.117	
LR test	-2(L0 - L*)		15432.261	
ρ^2	Adjusted rho-square		0.449	
D:Detached, S:Semi-detached, R:Row-link, C:Condominium, N: No development				

Table 4.6 presents the estimation results of new housing development model. In general, the model has a fairly high goodness-of-fit, as measured by the adjusted rho-square (0.449). The coefficient for the inclusive value of the development nest is statistically significant, which confirms the nest structure of the model. In addition, the coefficients for all the covariates are significant at the 95% confidence level with reasonable signs, which confirms the influences on housing development process from the employed factors. The alternative specific constants for each development type are all negative, which means that it is more likely to leave the parcel as is, other things being equal.

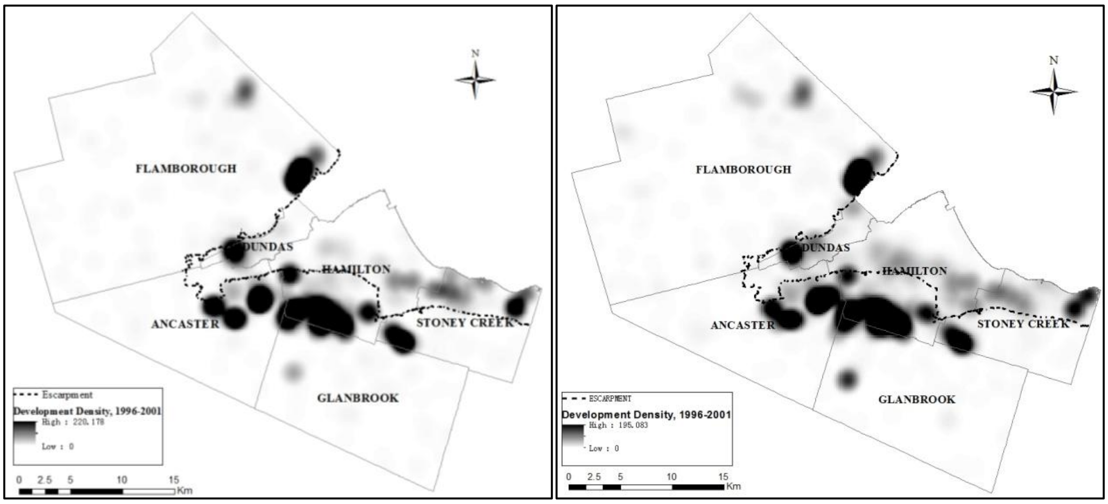
4.2.1 The Decision to Develop

Based on the results in Table 4.6, the utility function for non-development is expressed as follows:

$U_{Non\text{-}Dev}$ = 0.780 * ***B0_RoadG1k*** + 0.289 * ***B0_RecG1k*** + 1.07 * ***B0_Indt200*** - 0.269 * ***B0_OpenLU1k5*** – 0.785 * ***B0_RecLU1k5*** – 0.444 * ***B0_ResLU1k5*** – 2.52 * ***B0_Move5y***

(4.1)

If a vacant residential land parcel is far away (greater than one kilometer) from the major roads, then on average the likelihood of having development on this parcel is decreased. This is because the long distance from road network will reduce the site's transportation accessibility. A similar explanation applies to the influence of distance to recreational land uses. If this parcel is at least one kilometer away from parks, golf courses, or other recreational amenities, it is more unlikely to start a construction. In general, the available residential amenities are important attractive elements for both the dwelling buyers and developers. This is verified by another variable, which indicates the proportion of recreational land uses in the parcel's proximity increases the probability of developing this parcel. In addition, proximity to industrial land plays an important role. Developers tend not to build a dwelling in a location that is close (200 meters) to an industrial site. The living environment around the industrial area is mainly taken into account. Two other land use variables appear significant in the model. The areas of both open space and residential land uses around the specific land parcel add a certain amount of disutility to the alternative of non-development. The influence of existing residential land uses is a form of spatial dependence effect, which the new developments can share some residential amenities in neighborhood so that the cost involved in the development process is reduced. On the other hand, the negative sign attached to the area of open space around the parcel explains the attractiveness of greenfield for new developments. Figure 4.5(a) is a density map of residential developments from 1996 to 2000 in the City of Hamilton, while Figure 4.5(b) presents a density map of predicted new residential developments in the same period. The similarity in the patterns of two density maps verifies the model capability in prediction use.



(a) **(b)**

FIGURE 4.5 Density Maps of Residential Developments, City of Hamilton, 1996-2001 (a-observed pattern; b- predicted pattern by the model)

Most new development were conducted at the urban-fringe of the City of Hamilton, where more lands converted from open area (agricultural farms, etc.) are available for constructions, instead of in the Hamilton downtown area where most vacant parcels are brownfields. By definition, brownfield sites are “real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant.” (US Environmental Protection Agency). Obviously, brownfield sites are less attractive than greenfields because of less cost-effective and greater risks for land developments (De Sousa, 2000). Other than above mentioned factors, the proportion of movers in the census tract also has a certain amount of influence on development decision; if the census tract has a high percentage of residents who settled down from other places (other countries, other provinces, other cities, or other census tracts) in last five years, the probability of the non-development decision is lowered down. This can be accounted for by the higher housing demand in this census tract from this high proportion of movers.

4.2.2 The Decision for the Type of Development

The utility functions for each type of development are given in Equation 4.2 to 4.5. The variable selection followed Maoh et al. (2010a), since the same research question is dealt with, and the results produced are consistent. For single-detached developments, as seen in Equation 4.2, the closeness (200 meters) to major roads or highway ramps decreases the likelihood of starting a new development on the land parcel. It is also less likely to build a new single-detached house in a location around an industrial site (800 meters), other things being equal.

$$\mathbf{U_{Single\text{-}detached}} = -4.25 - 0.101 * \boldsymbol{B1\_Road200} - 0.212 * \boldsymbol{B1\_Ramp200} - 0.137 * \boldsymbol{B1\_Indt800} + 1.43 * \boldsymbol{B1\_IntraMove5y} + 0.773 * \boldsymbol{B1\_AvHValue} \quad \mathbf{(4.2)}$$

For semi-detached housing developments (Equation 4.3), the closeness to industry still has a negative impact on starting a new development; however, if the parcel is within 400 meters away from major roads, the probability of a new semi-detached development will go up. Besides, the proportion of semi-detached houses in the census tract will also positively influence the outcome of building a semi-detached house.

$$U_{Semi\text{-}detached} = -4.79 + 0.367 * \textbf{\textit{B2\_Road400}} - 0.213 * \textbf{\textit{B2\_Indt400}} + 1.91 * \text{B2}\textbf{\textit{\_SemDpt}} + 0.571 * \textbf{\textit{B2\_AvHValue}} \quad \textbf{(4.3)}$$

Row-link constructions are more likely to start at parcels that are close (100 meters) to roads; and if the site is within 800 meters from the nearest highway ramp, the probability will also go up. The closeness to the shopping center plays a positive role in favor of row-link projects.

$$U_{Row\_link} = -4.49 + 0.0889 * \textbf{\textit{B3\_Road100}} + 0.181 * \textbf{\textit{B3\_Ramp800}} + 0.209 * \textbf{\textit{B3\_Shop400}} + 0.708 * \textbf{\textit{B3\_AvHValue}} \quad \textbf{(4.4)}$$

Comparing to other alternatives, if the parcel locates within 600 meters from the major roads, it is less likely to build a condominium, as shown in Equation 4.5. On the other hand, a form of spatial inertia is demonstrated in condominium developments. In the local area (1.5 kilometers around the parcel center), the higher proportion of recently built condominium in last five years will bring up the probability of starting a new condominium development at the parcel.

$$U_{Condo} = -4.46 - 0.132 * \textbf{\textit{B4\_Road600}} + 0.434 * \textbf{\textit{B4\_CD9196}} + 0.488 * \textbf{\textit{B4\_AvHValue}} \quad \textbf{(4.5)}$$

Other than above discussed factors which were employed in Koronios (2009), two other variables are included in this model. The first variable is a finer representation of migration status at the census tract level, which describes the proportion of population in one census tract which were moved from other census tracts within the city in last five years. This variable holds a positive sign in the utility expression of single-detached developments, which indicates it is in favor of starting a new single-detached development where the area of the parcel attracts more intra-urban movers. Another variable concerns the average dwelling value of the census tract. It is expected that this variable holds a positive sign in the utility function of each development choice, since higher profits for developers is expected to be achieved in areas with higher rents. The findings meet the expectation for each development type.

5. CONCLUSION

5.1 Thesis Summary

Two major tasks have been conducted in this research. First, a microsimulation residential housing market framework for the City of Hamilton has been designed based on the comparative review of current state-of-the-art operational models. Second, two models in the microsimulation framework, namely a real estate appraisal model and a new housing development model, have been estimated. The objective of the first model is to assess the market values of existing dwellings based on the housing transactions in the previous period, while the latter model enables the development of a dynamic housing supply module by modeling the location and type decisions during the housing development process for each year.

For the real estate appraisal model, three model forms, including a traditional hedonic model, a spatial regression model, and a regression Kriging model, have been employed in estimations for comparison purpose using the housing transaction data for the City of Hamilton in 2001. A series of independent variables that describe the characteristics of dwelling, location, and neighborhood are specified in the explanatory model. A fairly high proportion of data variability is explained by the non-spatial hedonic model. The comparisons among model estimation results demonstrate that the spatial regression model has achieved a higher goodness-of-fit than the hedonic model. In addition, we verified that spatial autocorrelation is present in the residuals of the traditional hedonic model, which is explicitly captured by the spatial regression model. In terms of model prediction accuracy, spatial models (SAR and Kriging) both achieve a certain level of improvement over the traditional hedonic model. As for the comparison between SAR and Kriging, SAR model performs better in the correlation parameter between predicted and observed values, and also in the RMSE (Root Mean Square Error), while Kriging reaches a smaller MAE (Mean Absolute Error) value. Overall, we end up recommending that the SAR model is more appropriate to be incorporated into the microsimulation framework.

For the new housing development model, a parcel-level two-tier nested-logit model has been estimated based on the observations between 1996 and 2001 for the City of Hamilton. The model is able to deal with not only the decision to develop on a specific vacant residential land parcel, but also the development type choice. As for the factors influencing the decision to develop, the picture revealed from the model estimation results is that land developers tend to start the development project in greenfield. The City of Hamilton is a traditional industrialized city in North America, which has a large portion of its economy rooted in heavy industries, especially in the production of steel. Over the last several decades, there have been many brownfield sites in the City of Hamilton, mainly in the older industrial and downtown area, that were previously used for industrial or certain commercial purposes. Government encourages the developers to redevelop the brownfields in order to improve the urban living environment in those areas. However, a common perception among developers is that such development is less cost-effective and involves higher risks than greenfield development. This perception is evident in development location decisions, as confirmed by the model. De Sousa (2000) points out that proper policy changes can increase the attractiveness of residential brownfield projects. The municipal government for the City of Hamilton has introduced the Environmental Remediation and Site Enhancement (ERASE) – Community Improvement Plan (CIP) to encourage and promote development in brownfield sites. Initially approved in 2001 and then expanded in 2005, the plan is designed to clean up brownfields in the older industrial area of the City and replace them with productive economic land uses, thereby achieving economic, environmental, and social benefits throughout the City (Hamilton ERASE Community Improvement Plan, 2010).

With respect to the development type choice, the model results are consistent with Koronios (2009) that solely dealt with type choice modeling based on the same dataset. The developers tend to build a single-detached housing on sites which are away from industrial area as well as not very close to highway and major roads. In addition, parcels in census tracts with higher average housing value and large group of intra-urban movers in population appear to be more attractive for the development of detached housing. For semi-detached housing development, developers are likely to select sites with proximity to major roads and away from the industrial area. Also, the construction of semi-detached housing is more likely to occur in those census tracts with a large percentage of semi-detached houses and higher average housing value. As for the row-link development, the parcels with a high level of transportation accessibility, such as being close to highway ramps, major roads, and shopping centers are more attractive to developers. Similarly, the average housing value in a census tract has a positive influence on the row-link development, and also the condominium constructions. The spatial inertia effect is found in condominium development. Developers tend to start projects in locations where there are a large proportion of recently built condominiums. In addition, the model suggests that it is less likely to have a condominium construction started in parcels with proximity to major roads.

5.2 Directions for Future Research

Several aspects of work can be carried out in future research, based on the results of this research and previous studies.

1) Real estate appraisal modeling. The modeling in this research relies on global regression approach. It is worthwhile to investigate the relationship between housing prices and the variations in variables at local scale. Several modeling methods have been employed in explanatory studies and used as spatial predictors (e.g., Paez, 2008; Harris et al., 2010), such as geographically weighted regression (GWR), moving window regression (MWR), moving window Kriging (MWK), Kriging with a GWR trend component amomg others. Another future direction is to look into the influence of the spatio-temporal dependence effect on property values, while this research only focuses on spatial autocorrelation by using observations in one year for estimation.
2) New housing development modeling. Spatial dependence effect in this research is captured by employing derived variables, such as the aggregated area of the recently built same development type in local area around a specific parcel. This can be explicitly incorporated into the spatial discrete choice models, as was done

by Mohammadian et al. (2008) that developed a spatial mixed logit model to analyze the homebuilders' choice behavior. In addition, the influence of housing market indicators, particularly at the neighborhood scale, hasn't been investigated in this research due to lack of appropriate data. Factors such as vacancy rates and micro-scale housing/land prices are essential in defining the expected profit for developers. Thus, this is a potential fruitful avenue of research if relevant data are available in future.

3) Microsimulation housing market model implementation. Wang (2009) developed and validated a microsimulation residential mobility model for the City of Hamilton that focuses on the simulation of housing demand and residential relocation. However, the housing supply module is not implemented, neither is the housing market module. The new housing development model in this research enables the implementation of a dynamic housing supply module in the simulation framework, and the real estate appraisal model is an essential component to develop the housing market module that simulates the interactions between the housing demand and the housing supply. Therefore, based on current explanatory models and available data, it is possible to be able to implement the microsimulation residential housing market model for the City of Hamilton and carry out the model validation in order to test the applicability of the simulation model.

BIBLIOGRAPHY

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic Publishers.

2006 Canadian Census. http://www12.statcan.ca/census-recensement/2006/rt-td/index-eng.cfm.

Bailey, T. and A. Gatrell (1995). Interactive spatial data analysis. Addison-Wesley, Harlow.

Blijie, B. (2005). The Impact of Accessibility on Residential Choice: Empirical Results of a Discrete Choice Model. 45th Congress of the European Regional Science Association 23-27 August 2005, Vrije Universiteit Amsterdam.

Brunauer, W.A., Lang, S., Wechselberger, P. and S. Bienert (2009). Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna. *The Journal of Real Estate Finance and Economics*, 41(4):390-414.

Buliung, R.N. and P.S. Kanaroglou (2007). Activity–travel behaviour research: conceptual issues, state of the art, and emerging perspectives on behavioural analysis and simulation modelling. *Transport Reviews*, 27(2): 151-187.

Can, A. (1990). The Measurement of Neighborhood Dynamics in Urban House Prices. *Economic Geography*, 66:3, 254-272.

Hamilton ERASE Community Improvement Plan (2010). http://www.investinhamilton.ca/images/stories/publications/ERASE-CIP-Apr-2010a.pdf.

De Sousa, C. (2000). Brownfield Redevelopment versus Greenfield Development: A Private Sector Perspective on the Costs and Risks Associated with Brownfield Redevelopment in the Greater Toronto Area. J*ournal of Environmental Planning and Management*, 43:6, 831-853.

Diaz-Serrano, L. (2006). Housing Satisfaction, Homeownership and Housing Mobility: a Panel Data Analyses for twelve EU countries. Discussion paper, IZA. The Institute for the Study of Labour (IZA), Bonn.

Filatova, T., Parker, D. and A. van der Veen (2009). Agent-Based Urban Land Market: Agent's Pricing Behavior, Land Prices and Urban Land Use Change. *Journal of Artificial Societies and Social Simulation*, 12(1):3.

Farooq, B., Miller, E. J. and M. Haider (2010). Modelling The Evolution of Office

Space Supply. Presented at the 12th World Conference on Transport Research, July 11-15, 2010, Lisbon, Portugal.

Goodman, J.L. (1976). Housing consumption disequilibrium and local residential mobility. *Environment and Planning A*, 8:855-874.

Habib, M.A. and E.J. Miller (2007). Modeling Residential and Spatial Search Behaviour: Evidence from the Greater Toronto Area. Sixth Triennial Symposium on Transportation Analysis, Phuket Island, Thailand.

Habib, M.A. and E.J. Miller (2008). Influence of Transportation Access and Market Dynamics on Property Values: Multilevel Spatiotemporal Models of Housing Price. *Transportation Research Record*: Journal of the Transportation Research Board, No. 2076, pp. 183-191.

Habib, M.A. and E.J. Miller (2009). Reference-Dependent Residential Location Choice Model within a Relocation Context. *Transportation Research Record: Journal of the Transportation Research Board*, 2133, 92-99.

Haider, M. and E.J. Miller (2000). Effects of Transportation Infrastructure and Location on Residential Real Estate Values. *Transportation Research Record: Journal of the Transportation Research Board*, 1722:1-8.

Haider, M. and E.J. Miller (2004). Modeling Location Choices of Housing Builders in the Greater Toronto, Canada, Area. *Transportation Research Record*: Journal of the Transportation Research Board, No. 1898, TRB, National Research Council, Washington D.C., pp. 148-156.

Harris, P., Fotheringham, A.S., Crespo, R. and M. Charlton (2010). The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets. Mathematical Geosciences, 42:6, 657-680.

Hassink, W. and M. van Leuvensteijn (2009). The Importance of Income and Housing Wealth Constraints for Future Residential Mobility. Discussion Paper Series 09-05, Tjalling C. Koopmans Research Institute, Utrecht School of Economics.

Hengl, T., Geuvelink, G. and A. Stein (2003). Comparison of Kriging with External Drift and Regression-Kriging. Technical note, ITC, Available online at http://www.itc.nl/library/academic_output.

Henley, A. (1998). Residential Mobility, Housing Equity and the Labour Market. *The Economic Journal*, 108:414-427.

Hulchanski, J. D. (2004). What Factors Shape Canadian Housing Policy? The

Intergovernmental Role in Canada's Housing System. In Young, R. and C. Leuprecht (eds.), *Canada: The State of the Federation. Municipal-Federal-Provincial Relations in Canada*. Institute of Intergovernmental Relations. Montreal and Kingston. PP. 221-247.

Hunt, J.D., Abraham, J.E. and T.J. Weidner (2004). Household Allocation Module of OREGON2 Model. Transportation Research Record No. 1898.

Hunt, J.D., Abraham, J.E., De Silva, D., Zhong, M., Bridges, J. and J. Mysko (2008). Developing and Applying a Parcel-Level Simulation of Developer Actions in Baltimore. Presented at Transportation Research Board 87th Annual Meeting, Washington D.C.

Hunt, J.D., Abraham, J.E. and T.J. Weidner (2010). Household Behaviour in the Oregon2 Model. In F. Pagliara et al. (eds.), *Residential Location Choice, Advances in Spatial Science*, DOI 10.1007/978-3-642-12788-5_9, # Springer-Verlag Berlin Heidelberg 2010.

Ioannides, Y.M. and J.E. Zabel (2008). Interactions, Neighborhood Selection and Housing Demand. *Journal of Urban Economics*, 63:229-252.

Koronios, M. (2009). Exploring and Modeling the Urban Land Development Process in Hamilton, Ontario. M.A. Thesis, McMaster University, Hamilton, Ontario, Canada.

Long, F. (2006). Modeling Spatial Variations in Housing Prices - An Investigation of the Individual and Joint Effects of Spatial Autocorrelation and Spatial Heterogeneity. M.A. Thesis, McMaster University, Hamilton, Ontario, Canada.

Malpezzi, S. (2008). Hedonic Pricing Models: A Selective and Applied Review. In *Housing Economics and Public Policy* (eds T. O'Sullivan and K. Gibb), Blackwell Science Ltd, Oxford, UK. doi: 10.1002/9780470690680.ch5.

Maoh, H., Kanaroglou, P., Scott, D., Páez, A. and B. Newbold (2009). IMPACT: An Integrated GIS-based Model for Simulating the Consequences of Demographic Changes and Population Ageing on Transportation. *Computers, Environment and Urban Systems*, 33(3): 200-210.

Maoh, H., Koronios, M., and P. Kanaroglou (2010a). Modeling type choice outcomes of housing development projects in the City of Hamilton, Canada. Proceedings of the 89th Annual Meeting of the Transportation Research Board: Compendium of Papers DVD, Washington DC, January 10-14, 2010.

Maoh, H., Koronios, M. and P. Kanaroglou (2010b). Exploring the land development process and its impact on urban form in Hamilton, Ontario. *The Canadian Geographer*, 54(1): 68-86.

Martinez, F.J. and R. Henriquez (2007). A Random Bidding and Supply Land Use Equilibrium Model. *Transportation Research Part B*, 41:632-651.

Martinez, F. and P. Donoso (2010). The MUSSA II Land Use Auction Equilibrium Model. In F. Pagliara et al. (eds.), *Residential Location Choice, Advances in Spatial Science*, DOI 10.1007/978-3-642-12788-5_5, # Springer-Verlag Berlin Heidelberg 2010.

Miller, E. J. (2002). Microsimulation. In *Transportation Systems Planning: Methods and Applications*, Goulias, K. G. eds., CRC Press, Chapter 12.

Miller, E.J., B. Farooq, F. Chingcuanco, and D. Wang (2010). Microsimulating Urban Spatial Dynamics: Historical Validation Tests Using the ILUTE Model System. Presented at the Workshop on Urban Dynamics, Marbella, Chile, March 22-25, 2010.

Moeckel, R., Spiekermann, K., Schurmann, C. and M. Wegener (2003). Microsimulation of Land Use. *International Journal of Urban Sciences*, 7(1):14-31.

Moeckel, R., Schwarze, B., Spikermann, K. and M. Wegener (2007). Simulating Interactions Between Land Use, Transport and Environment. Presented at the 11th World Conference on Transport Research (WCTR), 24-28 June 2007, UC Berkeley.

Mohammadian, A. and P. Kanaroglou (2003). Applications of Spatial Multinomial Logit Model to Transportation Planning. Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

Mohammadian, A., Haider, M. and P. Kanaroglou (2005). Incorporating Spatial Dependencies in Random Parameter Discrete Choice Models. Paper presented at the 84th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2005.

Mohammadian, A., Haider, M. and P. Kanaroglou (2008). Homebuilders Choice Behaviour Analysis. *Canadian Journal of Regional Science*, 31:2, 225-240.

Nordvik, V. (2001). Moving Costs and the Dynamics of Housing Demand. *Urban Studies*, 38(3):519-533.

Orcutt, G.H. (1957). A New Type of Socio-economic System. *Review of Economics and Statistics*, 39:116-123.

Osland, L. (2010). An Application of Spatial Econometrics in Relation to Hedonic House Price Modelling. *Journal of Real Estate Research*, forthcoming.

Paez, A., Long, F. and S. Farber (2008). Moving Window Approaches for Hedonic Price Estimation: An Empirical Comparison of Modelling Techniques. *Urban Studies,* 45(8):1565-1581.

Páez, A. (2009). Recent research in spatial real estate analysis. *Journal of Geographical Systems*, 11(4):311-316.

Rabe, B. and M. Taylor (2010). Residential mobility, quality of neighbourhood and life course events. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(3):531-555.

Rossi, P. (1955). *Why Families Move: A Study of the Social Psychology of Urban Residential Mobility*. New York: The Free Press.

Salvini, P. and E.J. Miller (2005). ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. *Networks and Spatial Economics,* 5(2):217-234.

Sheppard, S. (1999). Hedonic analysis of housing markets. in *Handbook of Regional and Urban Economics* Eds E S Mills, P Cheshire (Elsevier Science, Oxford).

Sopranzetti, B.J. (2010). Hedonic Regression Analysis in Real Estate Markets: A Primer. In C.-F. Lee et al. (eds.), *Handbook of Quantitative Finance and Risk Management*, DOI 10.1007/978-0-387-77117-5_78, Springer Science+Business Media, LLC 2010.

Spinney, J. (2010). Geography, Land Values, and Municipal Taxation: A Spatial Paradigm for the Estimation and Reclamation of Rent. Ph.D. Thesis, McMaster University, Hamilton, Ontario, Canada.

Svinterikou, M. (2007). Microsimulation Models in Geography Using Object-Oriented Programming: An Application to Residential Mobility, Ph.D. Thesis, University of the Aegean, Lesvos, Greece.

Tracy, J., Schneider, H. and S. Chan (1999). Are Stocks Overtaking Real Estate in Household Portfolios? *Current Issues in Economics and Finance*, 5(5):1-6.

Train, K. (2009). Discrete Choice Methods with Simulation. Cambridge University Press, Second Edition.

UrbanSim (2011). Development Proposal Choice Model.

http://urbansim.org/Documentation/Parcel/DevelopmentProposalChoiceModel.

US Environmental Protection Agency. Brownfields and Land Revitalization. http://www.epa.gov/brownfields/overview/glossary.htm.

van Ommeren, J. and M. van Leuvensteijn (2005). New Evidence on the Effect of Transaction Costs on Residential Mobility. *Journal of Regional Science*, 45 (4):681-702.

Waddell, P. (2002). UrbanSim: Modelling Urban Development for Land Use, Transportation and Environmental Planning. *Journal of the American Planning Association*, 68 (3), 297–314.

Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke, and G. Ullfarsson (2003) Microsimulation of urban development and location choices: design and implementation of UrbanSim. *Networks and spatial economics*, 3, 43-67.

Waddell, P. and Ulfarsson, G.F. (2003). Dynamic Simulation of Real Estate Development and Land Prices within an Integrated Land Use and Transportation Model System. CD-ROM, 135 Annual Meeting of Transportation Research Board (TRB), Washington DC, January 12-16.

Wang, Y. (2009). A Microsimulation Model for Residential Mobility: An Application to the City of Hamilton, M.A. Thesis, McMaster University, Ontario, Canada.

Wegener, M. and K. Spiekermann (1996). The potential of microsimulation for Urban Models. In *Microsimulation for Urban and Regional Policy Analysis*, Clarke, G. ed, Pion, London, 149-163.

Yu, D., Wei, Y. and C. Wu (2007). Modeling Spatial Dimensions of Housing Prices in Milwaukee, WI. *Environmental and Planning B: Planning and Design*, 34:1085-1102.

Zhou, B. (2009). Land Use Change through Market Dynamics: A Microsimulation of Land Development, the Bidding Process, and Location Choices of Households and Firms. Dissertation, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

Zhou, B. and K.M. Kockelman (2008). Microsimulation of Residential Land Development and Household Location Choices. *Transportation Research Record*, No. 2077, PP. 106-112.

Zhou, B. and K.M. Kockelman (2010). Land Use Change through Microsimulation of Market Dynamics: An Agent-based Model of Land Development and Locator

Bidding in Austin, Texas. *Presented at the 89th Annual Meeting of the Transportation Research Board*, Washington D.C.