4-6-2021

Calibration and Validation of The Population Mobility and Housing Price Sub-Modules of The Smartplans Integrated Urban Model

Mohamed Abdo
University of Windsor

CALIBRATION AND VALIDATION OF THE POPULATION MOBILITY AND HOUSING PRICE SUB-MODULES OF THE SMARTPLANS INTEGRATED URBAN MODEL

By

Mohamed Abdo

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada

2021

CALIBRATION AND VALIDATION OF THE POPULATION MOBILITY AND HOUSING PRICE SUB-MODULES OF THE SMARTPLANS INTEGRATED URBAN MODEL

by

Mohamed Abdo

APPROVED BY:

W. Anderson
Department of Political Science

Y. H. Kim
Department of Civil and Environmental Engineering

J. Tofflemire
Department of Civil and Environmental Engineering

H. Maoh, Advisor
Department of Civil and Environmental Engineering

February 23, 2021

DECLARATION OF CO-AUTHORSHIP / PREVIOUS PUBLICATION

I. Co-Authorship

I hereby declare that this dissertation incorporates material that is the result of joint research, as follows:

Chapter 2 of the dissertation is based on a draft of a journal article that was co-authored by Dr. Hanna Maoh and Mr. Terence Dimatulac. The chapter is currently under consideration for publication in Environment and Planning B: Urban Analytics and City Science. Chapter 3 is based on a draft of a journal article that was co-authored by Dr. Hanna Maoh. The chapter is prepared to be submitted for publication in the Journal of Geographical Systems.

In Chapter 2, co-author Terence Dimatulac contributed to the specification of the population mobility model for the period 2001-2006. He also assisted with proof-reading the text. In Chapter 3, co-author Dr. Maoh contributed to the validation of the SAR model. He also assisted with proof-reading the text in all Chapters as well as providing directions with data exploration and brainstorming. In Chapters 2 and 3, data analysis, model estimation and validation, and the write-up was done by the author.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my dissertation and have obtained written permission from each of the co-author(s) to include the above material(s) in my dissertation. I certify that, with the above qualification, this dissertation, and the research to which it refers, is the product of my own work.

II. Previous Publication

This dissertation includes the following two original papers that have been submitted for publication in peer reviewed journals:

Thesis Chapter	Publication title/full citation	Publication status
Chapter [2]	Abdo, M., Dimatulac, T. & Maoh, H. (2021). Analysis of Population Mobility within the SMARTPLANS Integrated Urban Model: Application to Four Canadian Metropolitan Areas. Submitted to *Environment and Planning B: Urban Analytics and City Science.*	Under Review
Chapter [3]	Abdo, M. & Maoh, H. (2021). Modeling and validating the price of residential housing in the SMARTPLANS integrated urban model. To be Submitted to *Journal of Geographical Systems.*	In Progress

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

III. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Since the 1960s, Integrated Urban Models (IUMs) have consistently been applied to simulate the future of cities. Technological advancement in recent years has opened the doors for sophisticated IUMs to be developed, ones requiring extreme computing power. The SMARTPLANS IUM is one example. While the development and application of SMARTPLANS exists in the literature, exploring potential improvements in the model's predictive ability is lacking. This dissertation aims to fill the gap in the literature by focusing on two sub-modules of SMARTPLANS to test and ultimately advance their performance.

The research conducted in this thesis explores the population mobility and land price submodules within the Land Use Module of SMARTPLANS. The models were estimated using relevant parameters, compared over time, and validated with Canadian census data. The results show that the population aged 24-35 is the primary influencing factor to impact population mobility in all study areas. Additionally, the number of detached dwellings and household income were found to positively impact house prices in all models. Further, the number of row houses and the distance from the central business district (CBD) negatively influenced prices.

The estimated models for the two sub-modules suggest stable transferability over time in regions experiencing steady pace growth. Furthermore, the analysis confirms a strong spatial influence present in the data associated with both submodules. As such, the utilization of spatially oriented techniques, namely the Simultaneous Auto-Regressive (SAR) model, resulted in superior predictions when compared to the predictions obtained from Ordinary Least Squares (OLS) regression models. The implementation of SAR models within SMARTPLANS will therefore improve its predictive ability.

ACKNOWLEDGEMENT

First and most importantly, I would like to express my gratitude to Dr. Hanna Maoh for all his efforts and guidance throughout the duration of my research. His devotion to the field of transportation combined with his hard-working attitude inspired me to be the best version of myself. He directed me step by step during the two-year program and could not have accomplished this great milestone of my life without him.

I would like to also thank the members of my dissertation committee: Dr. Bill Anderson, Dr. Yong Hoon Kim, and Prof. John Tofflemire for their insightful comments and support. I would like to also thank Dr. Chris Lee for chairing my defence. I extend my thanks to everyone at the Cross Border Institute for offering a positive and friendly environment. I am also grateful for Mr. Terence Dimatulac and Dr. Shakil Khan for always being there when I needed assistance. I would like to further extend my gratitude to Health Canada for funding parts of the research conducted in this thesis.

Last but not least, I thank my parents, my sisters, and my wife for all their support and love throughout my journey. Thank you for always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Planning the future of land use patterns and associated urban travel demand activities is critical to achieve urban sustainability. This topic has gained the attention of policy-oriented decision makers in recent years. Many cities throughout Canada have focused on devising integrated land use and transportation strategies to cope with the steady increase in population and travel activities. While not extensively applied in Canada, Integrated Urban Models (IUMs) have emerged as one of the methods to assist decision makers with the development of their future planning strategies. IUMs are virtual laboratories that simulate the relationship between land use and transportation systems and approximately 200 models have been utilized around the globe since the early 1960s (Miller, 2018). In practice, most Canadian cities have focused on the transportation system while using exogenous land use inputs. Such approach assumes a one-way relationship between land use and transportation. That is, land use has an influence on travel demand but not the other way around. However, in reality a two-way relationship exists in which land use affects travel demand and also travel demand drive land use changes. IUMs are well suited to capture the two-way relationship between the land use and transportation systems.

IUMs have consistently been developing and improving their predictive ability by incorporating different principles to their modeling approaches. Initially, the idea of using gravitation and entropy-maximization to develop land use models emerged in the late 1960s. These models were based on Newton's universal law of gravitation where the attraction between two bodies increases as the distance between them decreases, like the Lowry model (Gross, 1982).

Further progression in the space led to the development of economic based models involving the principles of macro-economics. The Leontief Input-output (IO) model followed by spatial IO models (for example the MEPLAN model) predicted the affects of movement of goods and services on the national economy and determine flows between traffic analysis zones (TAZs) (Ebiefung and Kostreva, 1993). Additional improvement led to the introduction of discrete choice models, which emerged as a means to spatially predict people's behaviour and choices. Discrete choice models are based on the theory of utility maximization, whereby an individual selects an alternative from a set of well-defined alternatives such that the selected alternative is associated with the highest utility (Train, 1986). The most commonly used discrete choice model is the Multinomial Logit (McFadden, 1978). Martínez (1992) combined the utility maximization framework with the bid-rent economic theory and proposed the Bid-Choice Model to predict location decisions of households in urban areas. Such approach has been used since then as the foundation for developing contemporary land use models such as the MUSSA model (Martinez, 1996) and Urbansim model (Waddell et al., 2003). More recently, a full-fledged IUM, known as SMARTPLANS, that has the capability to simulate various land use and transportation processes and the interactions between them has been develop and applied for a number of Canadian cities.

SMARTPLANS, **S**imulation **M**odel for **A**ssessing the **R**amification of **T**ransportation **P**olicies and **Lan**d use **S**cenarios, is an IUM used to simulate the relationship between land use and transportation and to assess the impact of such interaction on the environment and health in Canada (Maoh et al. 2019). It has the capability to be applied to any urban area since the parameters are not hardcoded but rather configurable by the user through the Graphical User Interface of the software. To date, SMARTPLANS has been applied to five major Canadian cities: London, Halifax, Ottawa, Vancouver and Calgary. As shown in Figure 1-1, SMARTPLANS is primarily composed of six modules: regionwide aggregate controls module, land use module, transportation module, spatial disaggregation module, health benefits module, and sustainability indicators module. As will be highlighted later, this thesis will focus on testing and improving the performance of certain sub-modules of the land use module, more specifically the population mobility and the land price sub-modules.

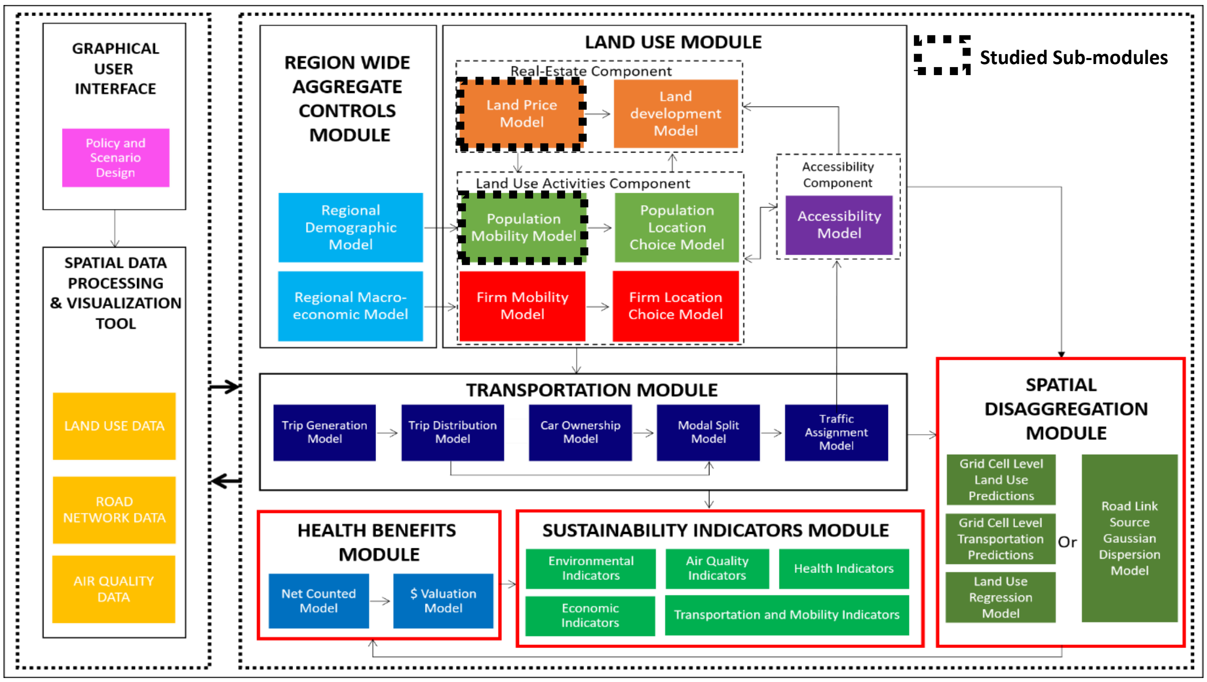


Figure 1-1 SMARTPLANS Modeling Framework

1.2 Research Objectives

The SMARTPLANS IUM is a fairly new model within the transportation field and has rarely been explored in the literature. Works such as ones presented in Maoh and Gingerich (2016) and Maoh et al. (2019) have mainly explored the model's development and its application, however studies to improve the model's predictive ability are not discussed in the literature. The research presented in this thesis strives to investigate the areas unexplored in the literature by mainly focusing on the following:

- Estimate the population mobility model within the Land Use Module of SMARTPLANS using population mobility data from different census periods (namely: 2001-2006 and 2011-2016) for the following Canadian census metropolitan areas: Halifax, London, Ottawa and Calgary
- Improve the specification of the land price model within the Land Use Module of SMARTPLANS by focusing on housing prices in the following Canadian cities: Ottawa and Calgary
- Compare the model parameters obtained from calibrating the population mobility and land price models using data from different census periods/years to explore the stability of the parameters over time
- Improve the population mobility and land price models by examining spatial models to account for the spatial nature of the modeled data
- Simulate population mobility and land price values for the years 2011 and 2016 using the base year 2006 model parameters to then validate the predicted results with official Canadian census data for the years 2011 and 2016

1.3 Thesis Outline

The remainder of this thesis is organized as follows:

Chapter 2 explores the outcomes obtained from calibrating, comparing, and validating a series of population mobility models in the Canadian CMAs of Halifax, Calgary, London, and Ottawa. Ordinary Least Squares (OLS) regression models are initially estimated to test the predictive performance of the mobility model. Simultaneous Auto-Regressive (SAR) models were then introduced to explore the potential improvement in results. Canadian census data for the periods 2001-2006 and 2011-2016 are utilized in the development and assessment of all models.

Chapter 3 investigates the findings achieved by specifying, estimating, comparing, and validating a series of land price models in the Canadian cities of Calgary and Ottawa. Four distinct models are explored: OLS models with and without region-specific parameters, as well as SAR models with and without region-specific parameters. The performance and accuracy of the models are then evaluated for their ability to recreate official price values provided by the Canadian census.

Finally, Chapter 4 provides conclusions of the research in this thesis by combining the findings obtained in Chapters 2 and 3. Significant factors influencing both population mobility and land price figures are described. The chapter further explores the enhancements provided by the SAR modeling technique when compared to its OLS counterpart. Directions for future research are also presented in this chapter.

1.4 Chapter 1 References

Ebiefung, A.A. & Kostreva, M.M. (1993). The generalized Leontief input-output model and its application to the choice of new technology. *Annals of Operations Research*. 44 (2). p.pp. 161–172.

Gross, M. (1982). *The Lowry model of land use simulation and its derivatives*. [Online]. Available from: https://reader.elsevier.com/reader/sd/pii/0198971582900072?token=1FE8EE5DC823488C37EDDD126C16E2605AA38D8838555E4E9FC270B8D94522F8784F7B26505F92E98A15A77D0B92DBAD. [Accessed: 12 March 2020].

Maoh, H., Dimatulac, T., Oiamo, T. & Johnson, M. (2019). Implementation of the SMARTPLANS Integrated Urban Model for Halifax, Nova Scotia and London, Ontario. In: *Book of the Canadian Transportation Research Forum (CTRF) 54th Annual Conference*. 2019, Vancouver, British Columbia, Canada.

Maoh, H. & Gingerich, K. (2016). SMARTPLANS: SIMULATION MODEL FOR ASSESSING THE RAMIFICATIONS OF TRANSPORTATION POLICIES AND LAND USE SCENARIOS. In: *Book of the Canadian Transportation Research Forum (CTRF) 51st Annual Conference*. 2016, Toronto, Ontario, Canada.

Martinez, F. (1996). MUSSA: Land Use Model for Santiago City. *Transportation Research Record*. 1552 (1). p.pp. 126–134.

Martínez, F.J. (1992). The Bid—Choice Land-Use Model: An Integrated Economic Framework. *Environment and Planning A: Economy and Space*. 24 (6). p.pp. 871–885.

McFadden, D. (1978). MODELING THE CHOICE OF RESIDENTIAL LOCATION. *Transportation Research Record*. [Online]. (673). Available from: https://trid.trb.org/view/87722. [Accessed: 19 March 2020].

Miller, E.J. (2018). Integrated urban modeling: Past, present, and future. *Journal of Transport and Land Use*. 11 (1). p.pp. 387–399.

Train, K. (1986). Google-Books-ID: bUUNKbTx590C. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. MIT Press.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M. & Ulfarsson, G. (2003). Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics*. 3 (1). p.pp. 43–67.

CHAPTER 2

ANALYSIS OF POPULATION MOBILITY WITHIN THE SMARTPLANS INTEGRATED URBAN MODEL: APPLICATION TO FOUR CANADIAN METROPOLITAN AREAS

2.1 Introduction

Different individuals have various reasons to change their residence to improve their quality of life, where they could either move to a different neighborhood (i.e., intra-urban migration) or even a different country (i.e., external immigration). According to Statistics Canada (2016), approximately 38% of the Canadian population relocated between 2011 and 2016. When dissecting population mobility, the 2016 Canadian census reported that around 54% of relocations were intra-urban, 28% were intra-provincial, 7% were inter-provincial, and 11% were attributed to external migrants. These figures suggest that the spatial distribution of population in urban areas over time is largely driven by their mobility. Therefore, it is imperative to account for mobility when modeling urban land use changes.

In the past few decades, Integrated Urban Models (IUMs) have been developed to study the relationship between land use and transportation. Several city-specific IUMs around the globe have been developed to examine urban sustainable solutions, such as ILUTE (Chingcuanco and Miller, 2018), UrbanSim (Waddell, 2002), and PECAS (Miller, 2018). These models have their own dedicated population mobility models to simulate relocation decisions and levels within the area of interest. More recently, a full-fledged IUM called SMARTPLANS (**S**imulation **M**odel for **A**ssessing the **R**amification of **T**ransportation **P**olicies and **Lan**d use **S**cenarios) has been developed to study land use and transportation problems in a number of Canadian Census Metropolitan Areas (CMA). Like other IUMs, SMARTPLANS has a dedicated population mobility model, which is incorporated within its land use module. SMARTPLANS can be applied to any urban area since that the model parameters are not hardcoded in the software, but rather configurable through the graphical user interface (Maoh et al. 2019).

Despite of being a fully operational IUM, SMARTPLANS has rarely been explored in the literature. Works such as ones presented in Maoh and Gingerich (2016) and Maoh et al. (2019) have mainly explored the model's development and application. However, studies to validate and improve the model's predictive ability are still lacking in the literature. The research presented in this chapter tries to fill this gap by studying the population mobility submodule of the land use module of SMARTPLANS. The conducted research is focused on (1) estimating logistic regression models using population mobility data for different census periods to evaluate the stability of the estimated parameters over time, (2) utilizing spatial regression models to examine if the introduction of spatial terms can improve the performance and predictive ability of the population mobility sub-module, and (3) validate the predictive ability of the estimated models over time. The conducted analysis is applied to the following Canadian CMAs: (1) Halifax, Nova Scotia; (2) Calgary, Alberta; (3) London, Ontario; and (4) Ottawa, Ontario, using data for the following census periods: 2001-2006 and 2011-2016.

The remainder of this chapter is organized as follows. The second section provides a background on push and pull factors regarding population relocation and discusses model validation as is performed within IUMs. Next, the third section describes the study areas. The fourth section then explores the development of the mobility model and the method used to compare the parameters. The fifth section discusses the obtained results, and the final section provides the conclusion of this research.

2.2 Background

2.2.1 Population Mobility Push and Pull Factors

There tends to be certain motives behind the decision of population movement, whether that be urban, provincial, or international. Parkins (2010) explores four factors that encourage this type of behavior: lack of safety, skill mismatch, scarce economic opportunity, and absence of social opportunities. On the other hand, a generalized classification is explained in Martin and Zürcher (2008), where the factors are grouped into three main categories: demand-pull, supply-push, and networks all with economic and non-economic migrant types. Demand-pull factors are often defined as the reasons a specific location is attractive to a mover, whereas supply-push factors are the reasons enticing stayers. Network factors deal with the conveniences or accessibility regarding the movement.

For many decades, achieving economic prosperity is considered a key determinant for individuals to relocate. Areas with promising economic opportunities are likely to attract skilled workers in deprived regions in hopes of earning higher wages (i.e., developing to developed country immigration). For instance, approximately 47% of the residents of Toronto, ON (Canada's largest city and economic hub) are foreign born (Picot, 2008). Many developing countries witnessed what is known as "brain drain", where many educated people leave their territories, which severely impact their country's human capital assets (Lowell and Findlay, 2001). The country of destination, although benefits in the process, also experiences some drawbacks. For example, skilled migration allows for an increase in both household income and population, which ultimately leads to rising rent and house prices. According to the Ontario Human Rights Commission (2007), immigration has increased the demands for rental properties, especially in the core of many Ontario's cities. Moreover, Saiz (2007) argues that any acceleration or deceleration in immigration has a direct impact on the rise and fall of both rent and house prices in many American cities.

Additional important factors influencing relocation decisions are with respect to migrant characteristics, more specifically their age and citizenship. Intuitively, younger age demographic is more likely to be footloose and tend to relocate more smoothly than older individuals. Searching for new experiences and/or foreign study opportunities often encourage many young people to move to different destinations (Martin and Zürcher, 2008). Hare (1999) explores several determinants of migration and shows that groups between 16 to 25 years and 26 to 35 years are 30% more likely to migrate than those who are 45 and above. The implementation of citizenship parameter, specifically those from developed countries, into a migratory model is scarcely available in the literature. Therefore, to better understand the effects of citizenship on population movement, we turn to mover information of a developed country (e.g., United States), where the majority of the population are citizens. The United States Census Bureau (USCB) reported that approximately 93% of the American population are U.S. citizens (United States Census Bureau, 2018). With that being said, the USCB estimates that approximately 9.8% of the total population moved in 2019, 60% of which came from the same county (2019). This pattern, however, has been slowly declining since the late 1940s. Within the Canadian context, approximately 35% of all Canadian households have moved within the past 5 years, where 61% of them moved within the same city (Statistics Canada, 2019). As such, most residential relocations in North America can be seen as intra-urban in nature.

The study by Li and Siu (2001) further supports the aforementioned occurrence, in which the authors suggest that the majority of residential mobility are short distance movers and generally relocate within the same city district or a neighbouring district. Moreover, they observe that most migratory patterns within the urban form are inner district to an outer adjacent district with very rare cases of movement within inner city districts. Such migratory pattern is often referred to as suburbanization, which is prominent in many metropolitan areas around the world (Mieszkowski and Mills, 1993). Typically, improvement in accessibility, such as the development of the U.S. Interstate highway system, is considered a major catalyst for this phenomenon. In a study by Baum-Snow (2007), empirical estimates show that approximately 18% of the city's population is lost due to a single highway passing through its core. In the same manner, intra-urban relocation has also been fuelled by the nature of land development, which preferred areas far away from the core. Such pattern is well observed in the Canadian context in the case of Hamilton, Ontario (Maoh et al., 2010). On the other hand, middle and high income earners enjoy larger single family homes rather than older, smaller residential units centrally located at the city's core (Mieszkowski and Mills, 1993).

2.2.2 Model Validation

For the past few decades, IUMs have been used consistently to plan the future of land use patterns and travel demands for different cities around the globe (Wegener, 1994). IUMs are typically defined as a type of modeling framework that integrates a transportation model with a land use model. As of 2013, there are approximately 200 state-of-the-practice models that have been developed in the last 40 years, in which around 40 models are still being used (Miller, 2018). A reliable model must incorporate a robust validation technique to examine its performance and accuracy. Model validation is a method used to assess the performance of a model based on its intended purpose (Vliet, 2013).

There are two primary validation techniques normally used in transportation research: independent and dependent validation. Independent validation is when 100% of the data is used to calibrate the model from time *t1* to time *t2* then validated from time *t2* to time *t3* (Vliet, 2013). This method is commonly used when sufficient data is available (Kok et al., 2001). On the other hand, dependent validation is used when data is scarce. This validation technique reserves a portion of the calibration data to be used in the validation process, basically splitting the data into separate samples (Sullivan et al., 2010). The quality of both validation techniques is observed through goodness of fit measures (e.g., correlations, root mean square errors, etc.) and its ability to regenerate the known state (Engelen and White, 2008).

Based on available research, not all IUMs have incorporated validation as part of their simulation processes. Recently, the Integrated Land Use, Transportation, Environment (ILUTE) model has taken the forefront in the IUM space with its consistent upgrades and improvements allowing it to be a credible transportation and land use model (Salvini and Miller, 2005). To achieve this, the ILUTE model has undergone several validation processes within its separate sub-models to evaluate accuracy, more specifically the demographic and housing market sub-models. Miller et al. (2011) applies independent validation techniques on both sub-models for a twenty-year period (1986 to 2006) using Canadian census and transportation survey data. The demographic model is then re-visited by Chingcuanco and Miller (2018) and validated once again using a similar approach.

2.3 Study Areas

This chapter focuses on the following Canadian CMAs: (1) Halifax, Nova Scotia; (2) Calgary, Alberta; (3) London, Ontario; and (4) Ottawa, Ontario. The Halifax CMA is located east of Canada along the Atlantic coast of Nova Scotia and has a land area of about 5,496km^2. The Calgary CMA is in southern Alberta, which is west of Canada, and has a comparable land area to Halifax (5,110km^2). The London and Ottawa CMAs are in southwestern and northeastern Ontario, respectively. London has the smallest land area among the four study areas, with about 2,662km^2, while Ottawa has the largest land area, with about 6,767km^2. Figure 2-1 shows the locations of the CMAs in the Canadian context.

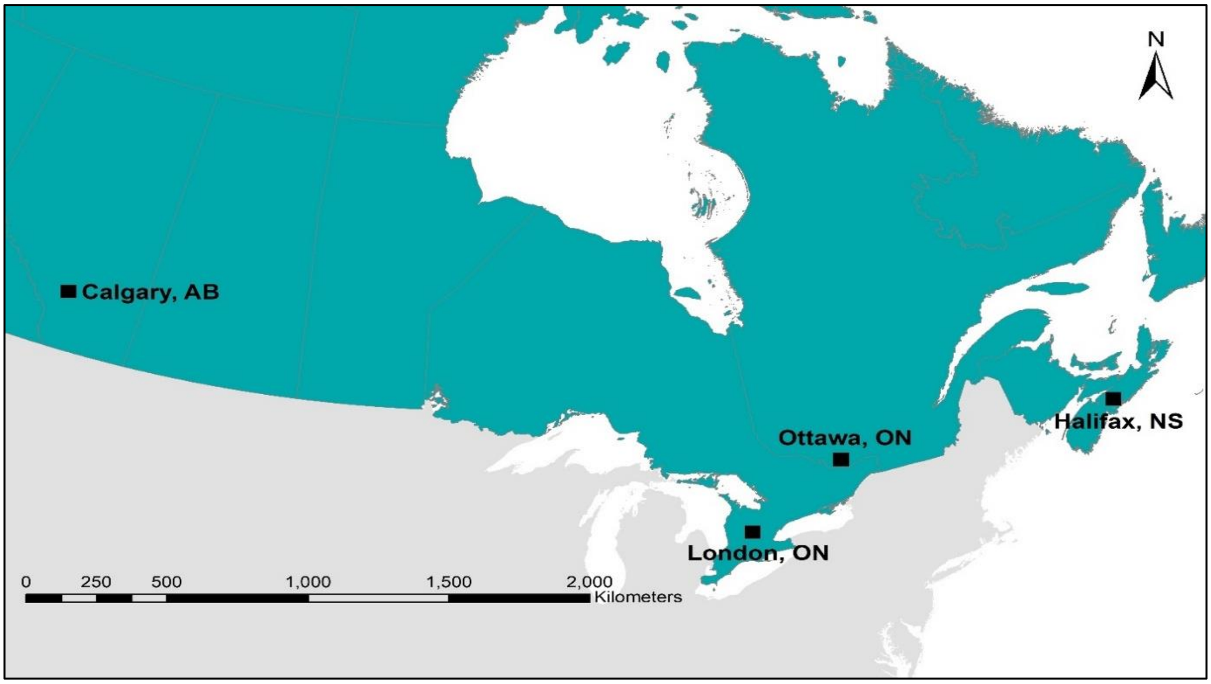


Figure 2 - 1 Map of Canada Highlighting the Four CMAs

Several factors have led to the growth of population in each region. For example, compared to other Canadian CMAs, housing prices in Halifax are among the least expensive. The Canadian real estate association (2020) reports that an average home in Halifax costs about $366,000, compare to the national average price of $539,000. This has attracted many interprovincial movers to the area, while still maintaining its medium city feel. More specifically, about 47.8% of all migrants in Halifax are inter-provincial, followed by external migrants with 28.7%, and finally intra-provincial migrants which comprise 23.5% (Statistics Canada, 2016). On the other hand, being at the centre of Canada’s oil industry, Calgary has witnessed a boost in its economy and population growth over the last two decades. In addition, its average residential property tax has been lower than the national average in the past decade. More specifically, Calgary's average property tax is 28% lower as of 2018 (Altus Group, 2018).

London is within close proximity to many significant locations across Ontario. For example, it is only a few hours from Toronto, Ontario's capital city and a major economic hub in Canada. Additionally, it has close access to the three major border crossings to the United States: Ambassador Bridge, Peace Bridge, and Blue Water Bridge. Meanwhile, Ottawa, as the capital of Canada, has experienced a great deal of economic prosperity and infrastructure growth in the last decade. That is, many large-scale transportation projects (e.g., light rail, major bus terminals etc.) are currently being planned and under construction to improve the city's commute (City of Ottawa, 2020)

According to Statistics Canada, the total population of each CMA has significantly increased in the last two decades, where Halifax experienced the least (about 12%) and Calgary experienced the most (more than 46%), as shown in Table 2-1 (Statistics Canada, 2001; 2006; 2011; 2016). The continuous development in the suburbs has aided in the occurrence of sprawl and population decentralization. Consequently, the CMAs have expanded horizontally and witnessed an increase in their spatial footprints. Therefore, determining population movement and understanding patterns associated with such mobility are of great importance.

Table 2 - 1 2001 - 2016 Population Counts in the Four Study Areas

CMA	**2001**	**2006**	**2011**	**2016**
Halifax	359,183	372,858	390,328	403,390
Calgary	951,395	1,079,310	1,214,839	1,392,609
London	432,451	457,720	474,786	494,069
Ottawa	1,063,664	1,151,141	1,254,919	1,323,783

2.4 Methods of Analysis

2.4.1 Specification of the Population Mobility Model

The objective of the mobility model in SMARTPLANS is to predict the probability of the population staying $P(S_i)$ in the same census tract i (i.e., zone) between two simulation periods t and t+1. That is, the zonal population in census tract i at time t will either stay or leave the census tract over the simulation period. As shown in Figure 1-1, this model is part of a larger land use activities module within the SMARTPLANS IUM (Maoh et al. 2019). The mobility model is integrated with a population location model to predict the spatial distribution of population in the census tracts comprising the study area over time. A logistic regression model composed of a series of socio-economic attributes (X_1, $X_2 \ldots X_q$) is used to model the stay probability, as shown in the following equation:

$$P(S_i) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 X1_i + \beta_2 X2_i + \cdots + \beta_q Xq_i\right)\right)} \qquad \text{(Eq. 1)}$$

where β's are parameters to be estimated. The reasoning behind selecting a logistic regression model is due to the categorical nature of the dependent variable (i.e., stay or move). This modeling technique uses a series of independent variables to predict the outcome of the dependent variable. Since our data is based on observed zonal totals, equation 1 can be expressed as the following linear function:

$$\ln\left(\frac{P(S_i)}{1 - P(S_i)}\right) = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \cdots + \beta_q Xq_i \qquad \text{(Eq. 2)}$$

The left-hand side of equation 2 is known as the log of odds. The parameters in equation 2 can now be estimated using the Ordinary Least Square (OLS) method. However, given the spatial nature of the analyzed data, the dependent variable (i.e., $y_i = \ln\left(\frac{P(S_i)}{1-P(S_i)}\right)$) might exhibit spatial autocorrelation. Spatial autocorrelation can be present in spatial data if the observed values of two or more neighboring census tracts (areas) are highly correlated. The Moran's *I* statistic can be used in such case to examine the presence of spatial autocorrelation. According to Bailey and Gatrell (1995), Moran's *I* statistic can be calculated as follows:

$$I = \frac{1}{s^2}\frac{\sum_i \sum_j (y_i - \bar{y})(y_j - \bar{y})}{\sum_i \sum_j w_{ij}} \qquad \text{(Eq. 3)}$$

where i and j are the observations, $\bar{y}$ is the mean, s^2 is the sample variance, and w_{ij} is the weight matrix of observations i and j. The range of values of Moran's *I* is between -1 and +1. A value in the vicinity of -1 represents perfectly dispersed data with clustering of unrelated values. A value of and close to 0 represents complete random values and correlation does not exist in the data. Finally, a value of and around +1 suggests that the data is perfectly clustered with similar values.

The application of the OLS method to estimate the linear model will lead to biased parameters if the dependent variable y_i (in this case, the log of odds) exhibits spatial autocorrelation. To remedy the problem, the Simultaneous Auto-Regressive (SAR) model can be used instead. The SAR model can be formulated as follows:

$$y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \cdots + \beta_q Xq_i + \rho \sum_{j=1}^{N} w_{ij} y_j \qquad \text{(Eq. 4)}$$

Notice that the SAR model includes the additional spatial lag term $\sum_{j=1}^{N} w_{ij} y_j$ which is associated with the spatial lag parameter ρ. The addition of the spatial lag term will account for any potential spatial autocorrelation, leading to un-biased parameter estimates in the model.

The model is based on Statistics Canada Census data (Statistics Canada, 2001; 2006; 2011; 2016). Several attributes at the census tract (CT) level are explored to understand the factors affecting population movement. In doing so, the factors listed in Table 2-2 are deemed important explanatory variables. We hypothesize that the probability of staying in the same CT decreases if the CT houses more people in the age class 25 to 34. This age cohort is likely to be more footloose compared to older cohorts, who are more established, and as such will have the tendency to relocate in the search for better opportunities. Additionally, CTs with higher average housing rent reduces their attractiveness; therefore, the probability of staying reduces. With regards to Canadian citizens, the probability of staying in their current census tracts increases since they tend to be more settled and hence, less likely to relocate compared to other population groups (e.g., immigrants and/or refugees). Similarly, CTs with high average family income suggest social stability and that increases the probability of staying in the census tract. Finally, as the distance from the CBD increases, the probability of staying increases due to the effects of sprawl and suburbanization.

Table 2 - 2 Covariates Used in the Specification of the Logistic Regression Model

Covariates	Definition
Citizen	Number of Canadian citizens located in census tract *i*
Pop2534	Number of persons aged between 25 and 34 residing in census tract *i*
DistCBD	Euclidian distance between the centroid of census tract *i* and the centroid of the CBD, in kilometers
Rent	Average rent price in census tract *i*, in $CAD
FamInc	Average family income in census tract *i*, in $CAD

It is important to note that the covariates listed in Table 2 – 2 represent data for the beginning of the time period. That is, if the dependent variable represents the share of stayers between 2001 and 2006, the covariates pertain to data for the year 2001.

2.4.2 Parameter Comparison Over Time

IUMs perform predictions using base year datasets to simulate future results. The accuracy of the predictions is usually determined by the span of the simulation period. Typically, as the gap between the base year and the target simulation year increases, errors accumulate and lead to poorer results. Here, we seek to explore if there is significant difference in predictions using parameters obtained by estimating the model with data from two different time periods. Therefore, the mobility model would be calibrated for the years 2006 and 2016, and their coefficients would be compared to determine significance. This will help answer the question: *Do the parameters remain fixed over time?* To perform comparisons, the Wald Chi-Square test will be used to determine whether the 2016 parameters are significantly different from the 2006 ones. The formula for this statistic is:

$$\chi^2 = \frac{(\beta_{t+1} - \beta_t)^2}{[se(\beta_{t+1})]^2 + [se(\beta_t)]^2} \qquad \text{(Eq. 5)}$$

where

β_{t+1} = coefficient for the 2016 parameter

β_t = coefficient for the 2006 parameter

$se(.)$ = standard error for each period

When calculating χ^2, we aim to examine the null hypothesis that β_{t+1} is no different from β_t. A significant χ^2 value suggests that we must reject the null hypothesis, leading us to conclude that the two parameters are different from each other. Here, the critical values for significance for 1-degree of freedom at 90% and 95% levels are 2.706 and 3.841, respectively.

2.5 Model Estimation Results

2.5.1 Model Calibration and Comparison

The estimated OLS regression model results for the years 2006 and 2016 in the case of the four CMAs are shown in Table 2-3. Similarly, the estimated SAR model results are shown in Table 2-4. High correlation values between observed and predicted results suggest good predictive ability of both OLS and SAR models. The SAR models yield slightly improved R-Square values compared to its OLS counterpart in all study areas, but both modeling techniques produced respectable R-square values indicating acceptable goodness of fit. Further, Moran's *I* values (which are estimated and are compared to 999 randomly generated permutations), suggest that the dependent variable used in the different OLS models are clustered to some degree (i.e., exhibit positive spatial autocorrelation). This finding is reinforced by the values of the spatial lag coefficient (ρ), except for London in the 2006 period. In general, the SAR model parameter values are relatively different from the OLS values since the spatial lag parameter is able to capture and account for the presence of spatial autocorrelation in the data.

The signs of all parameters are in line with our initial hypotheses, except for the *Rent* variable in the case of Calgary for both simulation periods. Such results imply that individuals in Calgary are more likely to stay at their current CT despite of increasing average rental cost on each zone. This unusual outcome could be attributed to Calgary's economical boost in the last two decades caused by rising oil prices. High demand in the housing market, combined with scarce rental properties, have impacted the housing rental prices in Calgary. The *Pop2534* variable has the highest coefficient values, while *DistCBD* has the lowest coefficient values in all study areas. The impact of the 25-34 age group on relocation is very significant and is clearly observed in the results. On the other hand, few parameters are not significant at the 90% confidence level; hence, they are not included in the validation process. However, they are displayed here to provide context to the results.

Figure 2-2 displays the number of movers per 100 persons in 2016 for the four study areas. All four CMAs experience high relocation from the core, whereas the suburbs witness more stayers. This result is in-line with the current state of sub-urbanization of North American cities, where a significant portion of the population has been gradually leaving the cores to settle in the suburbs. On the other hand, the relocation patterns observed in Calgary’s suburbs differ from the other study areas, where relocation is more prominent. This suggests that the entire CMA observed strong population movement, whether that be relocation between different areas of the suburbs or entirely leaving the CMA in 2016.

Table 2 - 3 OLS Parameter Estimates of the Mobility Model, 2006 and 2016

Covariates	**Halifax**		**Calgary**		**London**		**Ottawa**	
Year	**2006**	**2016**	**2006**	**2016**	**2006**	**2016**	**2006**	**2016**
Citizen **	0.3252	0.2282	0.0992	0.0672	0.1877	0.1538	0.1530	0.1068
	(6.778)	(6.311)	(5.654)	(5.948)	(5.706)	(5.870)	(8.431)	(7.004)
Pop2534 **	-1.5586	-1.2548	-0.6060	-0.3309	-1.0584	-0.9285	-0.7769	-0.6150
	(-6.565)	(-6.818)	(-9.318)	(-7.546)	(-5.406)	(-6.514)	(-10.010)	(-7.068)
DistCBD	0.0109	0.0121	0.0124	0.0152	0.0067	0.0035	0.0155	0.0153
	(3.816)	(2.899)	(4.206)	(4.614)	(2.176)	(1.079)	(7.988)	(7.091)
Rent *	-0.0728	-0.0525	0.0345	0.0610	-0.0372	-0.0377	-0.0199	-0.0045
	(-2.775)	(-2.574)	(3.766)	(5.310)	(-1.499)	(-2.788)	(-2.312)	(-0.6185)
FamInc ***	0.4700	0.5417	0.4284	0.0639	0.8956	0.5880	0.7216	0.5585
	(1.692)	(2.864)	(4.514)	(1.225)	(4.120)	(3.990)	(9.293)	(8.144)
Correlation (%)	97.2	97.9	95.9	95.7	97.5	98.7	96.5	98.6
No. of Obs.	85	80	191	191	100	99	233	235
R-Square	0.72	0.74	0.70	0.59	0.69	0.64	0.70	0.66
Adj R-Square	0.69	0.72	0.69	0.58	0.67	0.62	0.70	0.66

Number in parenthesis is the t-stats value

** parameters scaled by 100; ** parameters scaled by 1,000; *** parameters scaled by 100,000*

Table 2 - 4 SAR Parameter Estimates of the Mobility Model, 2006 and 2016

Covariates	**Halifax**		**Calgary**		**London**		**Ottawa**	
Year	**2006**	**2016**	**2006**	**2016**	**2006**	**2016**	**2006**	**2016**
Spatial Lag (ρ)	0.3706	0.2675	0.2000	0.4297	0.1049	0.2596	0.2002	0.2878
	(3.993)	(2.876)	(2.775)	(6.237)	(1.002)	(2.432)	(2.971)	(4.151)
*Citizen ***	0.2444	0.1846	0.0885	0.0471	0.1788	0.1243	0.1360	0.0895
	(5.449)	(5.315)	(5.121)	(4.545)	(5.598)	(4.590)	(7.551)	(6.144)
*Pop2534 ***	-1.1506	-1.0537	-0.5333	-0.2404	-1.0133	-0.7562	-0.7067	-0.5276
	(-5.310)	(-5.824)	(-8.115)	(-5.965)	(-5.315)	(-4.867)	(-9.139)	(-6.356)
DistCBD	0.0073	0.0078	0.0084	0.0076	0.0054	0.0017	0.0109	0.0092
	(2.546)	(1.743)	(2.649)	(2.349)	(1.664)	(0.459)	(4.489)	(3.629)
*Rent **	-0.0194	-0.0452	0.0296	0.0407	-0.0382	-0.0331	-0.0234	-0.0082
	(-0.654)	(-1.619)	(3.314)	(3.880)	(-1.597)	(-2.004)	(-2.789)	(-1.169)
*FamInc ****	0.6555	0.5555	0.4283	0.1181	0.8556	0.6361	0.6939	0.5192
	(2.672)	(3.210)	(4.679)	(2.538)	(3.911)	(3.961)	(9.137)	(7.918)
Correlation (%)	98.0	97.9	95.9	96.9	97.5	98.7	96.6	98.9
Moran's *I*	0.64	0.58	0.48	0.54	0.43	0.39	0.54	0.55
No. of Observations	85	80	191	191	100	99	233	235
R-Square	0.75	0.77	0.71	0.67	0.69	0.66	0.72	0.69

Number in parenthesis is the t-stats value

** parameters scaled by 100; ** parameters scaled by 1,000; *** parameters scaled by 100,000*

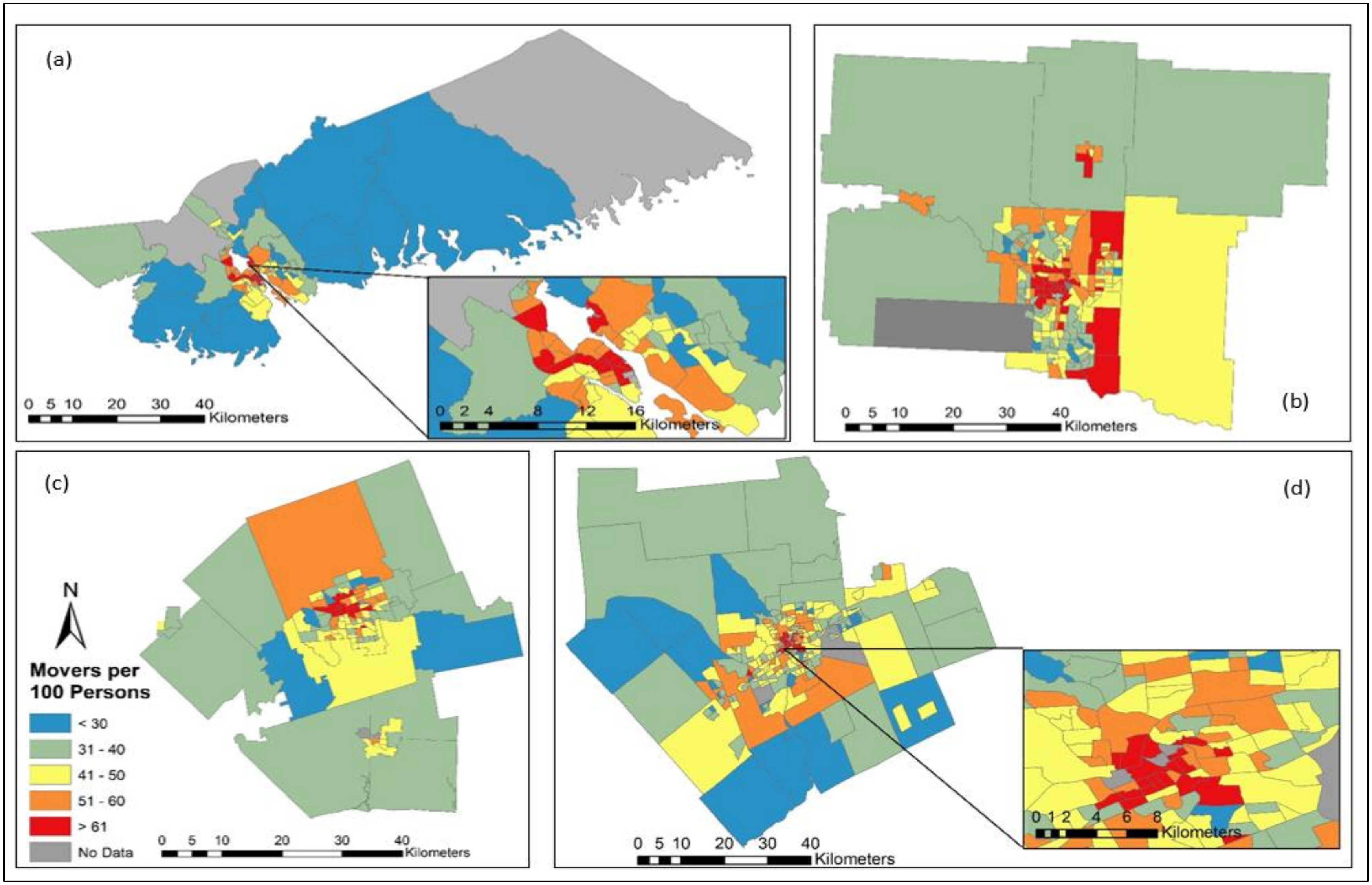


Figure 2 - 2 2016 Movers for (a) Halifax; (b) Calgary; (c) London; (d) Ottawa

2.5.2 Parameter Comparison Over Time

Table 2-5 displays the Wald Chi-Square statistic for the different parameters of both OLS and SAR models. Overall, the majority of the variables are not significant, suggesting that they do not change over time, albeit a few exceptions are noticed. Most of the significant parameters in Table 5 are related to the Calgary models. The differences between the 2006 and 2016 parameters in Calgary's case show that they change over time. The model comparison captures the clear relocation that occurred two decades ago in Calgary, with model variables such as *Citizen*, *Pop2534*, and *FamInc* differing significantly between 2006 and 2016 model years. It is possible that an overflow of population aged 25-34 into Calgary during its economic peak allowed for a sharp increase in household income. Such a scenario would throw off the model's predictive ability due to variation in historical trend of mobility in this Canadian CMA. With the exceptions found in the case of Calgary, the results suggest that the 2006 parameters would produce similar results compared to the 2016 parameters especially in those regions with stable and consistent changes in economic and demographic patterns over time.

Table 2 - 5 Wald Chi-Square for Difference in Parameters for 2006 and 2016 Models

	Halifax		**Calgary**		**London**		**Ottawa**	
Model Technique	**OLS**	**SAR**	**OLS**	**SAR**	**OLS**	**SAR**	**OLS**	**SAR**
Spatial Lag (ρ)	-	0.62	-	5.31*	-	1.07	-	0.82
Citizen	2.61	1.11	2.34	4.23*	0.65	1.70	3.80	4.04*
Pop2534	1.02	0.12	12.3*	14.4*	0.29	1.09	1.93	2.49
DistCBD	0.06	0.01	0.40	0.03	0.50	0.61	0.01	0.22
Rent	0.37	0.40	3.27	0.66	0.00	0.03	1.85	1.95
FamInc	0.05	0.11	11.4*	9.13*	1.37	0.65	2.48	3.03

** significantly different at the 0.05 level*

2.5.3 Model Validation

The calibrated 2006 parameters of both OLS and SAR models are used to predict 2011 and 2016 population of non-movers at the CT level in the four CMAs. As mentioned earlier, statistically insignificant parameters below the 90% confidence interval are excluded from the prediction process. Moreover, the predicted results are validated against the observed data obtained from the Canadian Census. The validation is based on calculating the correlation and root mean square error (RMSE) terms. Tables 2-6 and 2-7 provides a summary of the OLS and SAR validations, respectively. The results show that the RMSE significantly increases between the 2011 and 2016 values in three out of the four CMAs. Further, the SAR models reduce the RMSE and improves the correlation in most of the estimates. Moreover, scatter plots comparing the observed versus predicted population of non-movers of the OLS model for the years 2011 and 2016 are shown in Figures 2-3 and 2-4, respectively. The plots show that the models used to predict future population of non-movers are stable and perform well over time. Several outliers are observed within the Calgary CMA, which is expected due to the significant differences in some of the parameters of this CMA as reported earlier in Table 2-5, and because of the high RMSE values in Table 2-6. The outliers are determined to be the CTs located in the outskirts of Calgary. This further suggests the presence of urban sprawl in this region. Figure 2-5 presents the prediction error (i.e., residuals) in the 2016 population mobility SAR model in all four study areas. The error, which is based on the difference between the predicted and observed population movers, was normalized using the observed 2016 movers to obtain the percent values shown. Based on the results, the majority of the errors are within 20%, although Calgary and Ottawa have a larger number of zones with greater than 30% errors. This is in-line with the RMSE obtained in Table 2-7. Nonetheless, the general performance of the model is relatively strong.

Table 2 - 6 2011 and 2016 OLS Validation Results

	Halifax		Calgary		London		Ottawa	
Year	**2011**	**2016**	**2011**	**2016**	**2011**	**2016**	**2011**	**2016**
Correlation (%)	97.9	97.8	88.3	80.5	97.9	98.3	98.4	98.1
RMSE	367	336	1016	2235	286	352	334	791

Table 2 - 7 2011 and 2016 SAR Validation Results

	Halifax		Calgary		London		Ottawa	
Year	**2011**	**2016**	**2011**	**2016**	**2011**	**2016**	**2011**	**2016**
Correlation (%)	98.2	97.9	89.9	83.2	98.0	98.4	98.4	98.3
RMSE	401	401	937	2093	280	348	344	729

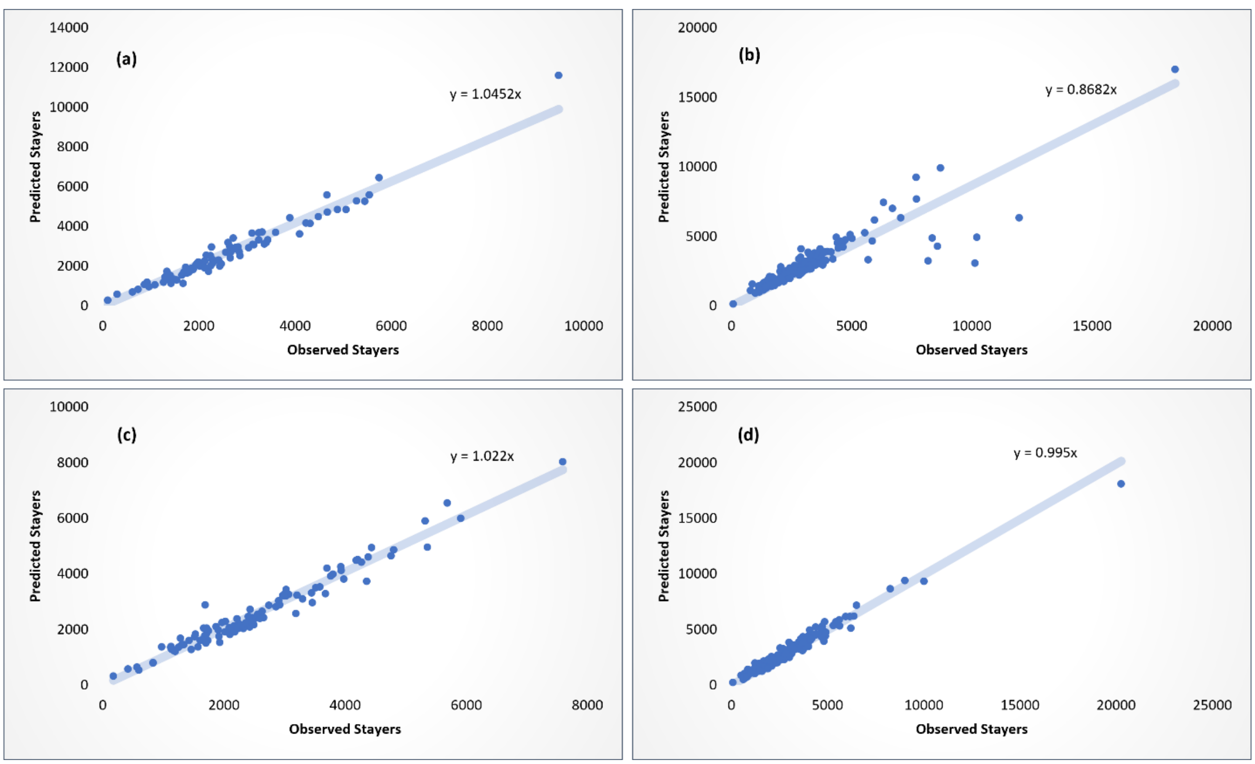


Figure 2 - 3 2011 Validation Scatters for (a) Halifax; (b) Calgary; (c) London; (d) Ottawa

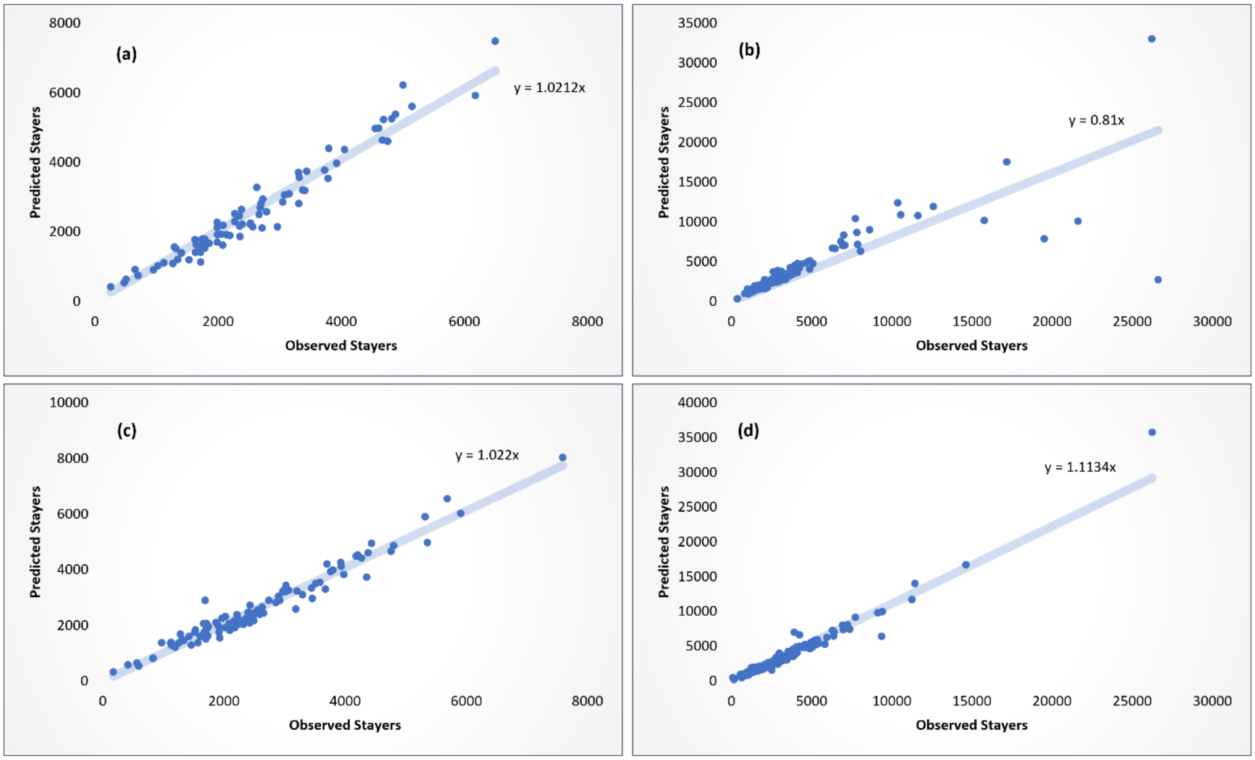


Figure 2 - 4 2016 Validation Scatters for (a) Halifax; (b) Calgary; (c) London; (d) Ottawa

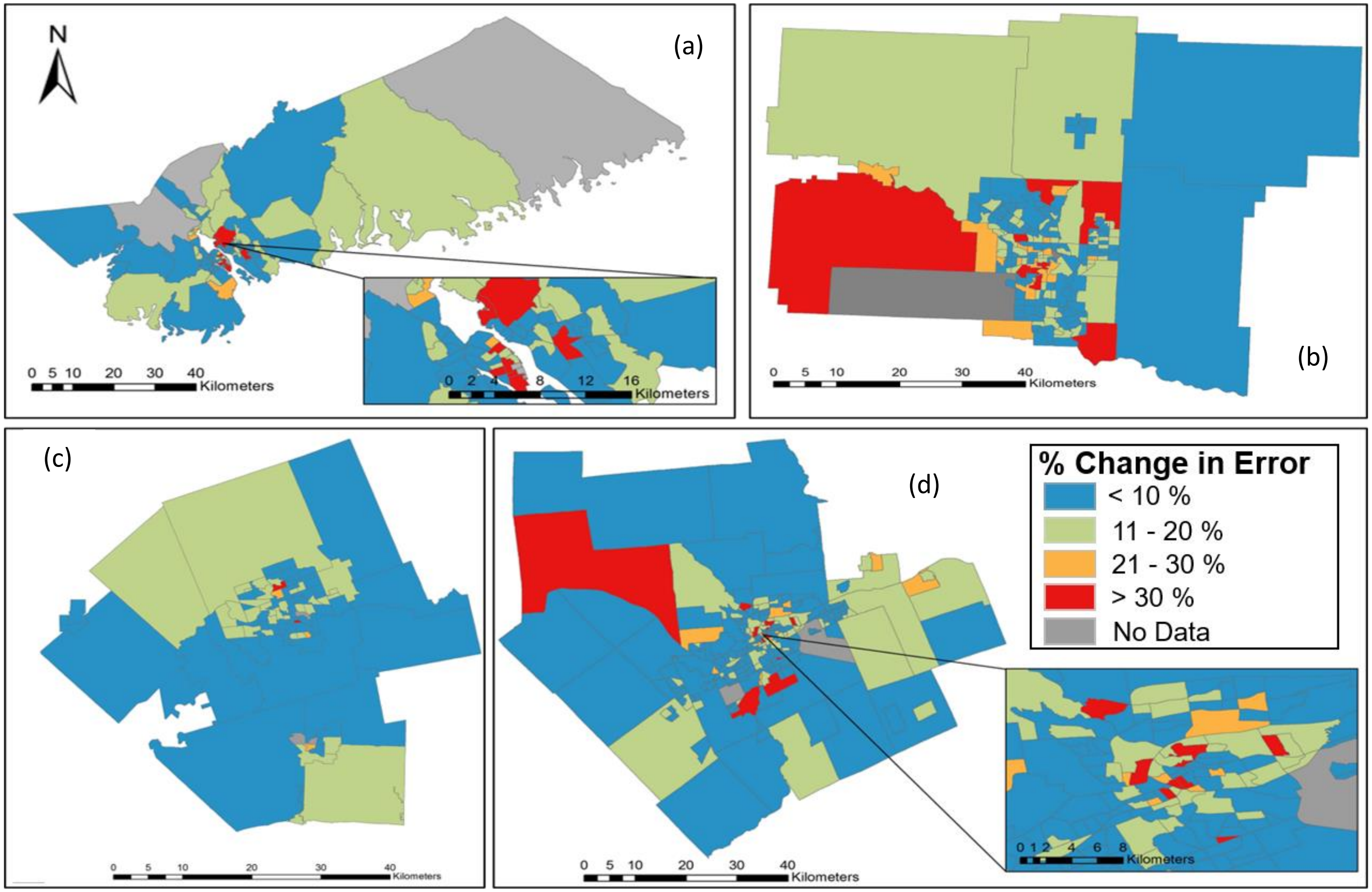


Figure 2 - 5 2016 Validation Residuals for (a) Halifax; (b) Calgary; (c) London; (d) Ottawa

2.6 Conclusions

This chapter presents the findings obtained through calibrating and validating mobility models for four Canadian metropolitan areas: Halifax, Calgary, London, and Ottawa. The 2006 and 2016 OLS logistic regression and SAR models are developed using a series of socio-economic attributes. The parameters are compared, and relocation figures are validated using official data from Statistics Canada. The results obtained from the models are in-line with the given hypotheses except for the *Rent* variable in the case of Calgary for both simulation periods. Individuals in Calgary are less likely to leave their current residence even as rent prices increase. This outcome could be due to Calgary's economic boost caused by rising oil prices. Among the attributes tested to determine population relocation, population aged 24-35 is the primary influencing factor to impact this decision. Both OLS and SAR results exhibit accurate estimates. The dependent variables of all study areas experience spatial autocorrelation as shown by the significant Moran's *I* coefficient.

The 2006 parameters are compared to the 2016 values to test for significance and investigate how the models compare over time. Most of the parameters are statistically insignificant, meaning the parameters do not change over time; however, a few exceptions are observed. Most of the significant parameters are located within the Calgary CMA. Results suggest that an influx of population aged 25-34 would likely increase income standards in the CMA in this time period, which negatively impacted the model's predictive ability. Relocation figures obtained from the models are then validated with Statistics Canada values using correlation analysis and RMSE values. Once again, the Calgary CMA produces sub-standard results with respect to other CMAs as shown with lower correlation and high RMSE. Moreover, observed versus predicted relocation scatter plots display a few abnormal relocation figures within the Calgary CMA, all located in the suburbs of the city.

In summary, the results obtained from calibrating and validating the mobility models shine light on several key aspects. First, the mobility process within the Canadian context is a stable one. Based on the findings, the Wald Chi-Square tests show that the parameters rarely vary overtime in all four CMAs. Second, the data associated with population movement in one zone is dependent on that of its neighbouring zone, suggesting the presence of spatial autocorrelation. Significant Moran's *I* coefficient as well as significant spatial lag parameters (ρ) justifies this case. We can further conclude that the SAR models used to estimate population mobility are stable and consistent in the regions they are applied to and are far more superior than their OLS counterparts. Improved R^2 values, along with slightly better correlation between the data, are seen throughout the models. The Calgary CMA has not performed as well as the other three CMAs; however, validation figures suggest satisfactory results.

While the present work can draw important conclusions regarding population mobility within the Canadian context, it has opened the doors for further research work. OLS regression models have consistently been applied within IUMs to predict population movement; however, SAR models can potentially enhance the predicted outcomes. Integrating the developed SAR models within the SMARTPLANS IUM and testing its overall performance is what we seek to explore. Further, we would like to perform comparisons and validations on the predicted results of the SMARTPLANS IUM before and after implementing these improvements and to test the significance of the differences.

2.7 Chapter 2 References

Altus Group (2018). *Canadian Property Tax-Rate Benchmark Report*. [Online]. 2018. Available from: https://www.altusgroup.com/wp-content/uploads/2019/08/Canadian-Property-Tax-Rate-Benchmark-Report-2018.pdf. [Accessed: 6 February 2021].

Bailey, T. & Gatrell, A. (1995). *Interactive Spatial Data Analysis*. [Online]. Available from: http://www.personal.psu.edu/faculty/f/k/fkw/rsoc597/Introduction.pdf. [Accessed: 6 February 2021].

Baum-Snow, N. (2007). Did Highways Cause Suburbanization?. *The Quarterly Journal of Economics*. 122 (2). p.pp. 775–805.

Chingcuanco, F. & Miller, E.J. (2018). The ILUTE Demographic Microsimulation Model for the Greater Toronto-Hamilton Area: Current Operational Status and Historical Validation. In: J.-C. Thill & S. Dragicevic (eds.). *GeoComputational Analysis and Modeling of Regional Systems*. Advances in Geographic Information Science. [Online]. Cham: Springer International Publishing, pp. 167–187. Available from: https://doi.org/10.1007/978-3-319-59511-5_10. [Accessed: 22 July 2020].

City of Ottawa (2020). *Construction and infrastructure projects*. [Online]. 1 April 2020. Construction and Infrastructure Projects. Available from: https://ottawa.ca/en/planning-development-and-construction/construction-and-infrastructure-projects. [Accessed: 22 January 2021].

Engelen, G. & White, R. (2008). Validating and Calibrating Integrated Cellular Automata Based Models of Land Use Change. In: S. Albeverio, D. Andrey, P. Giordano, & A. Vancheri (eds.). *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach*. [Online]. Heidelberg: Physica-Verlag HD, pp. 185–211. Available from: https://doi.org/10.1007/978-3-7908-1937-3_10. [Accessed: 19 July 2020].

Hare, D. (1999). 'Push' versus 'pull' factors in migration outflows and returns: Determinants of migration status and spell duration among China's rural population. *Journal of Development Studies*. 35 (3). p.pp. 45–72.

Kok, K., Farrow, A., Veldkamp, A. & Verburg, P.H. (2001). A method and application of multi-scale validation in spatial land use models. *Agriculture, Ecosystems & Environment*. 85 (1). p.pp. 223–238.

Li, S.-M. & Siu, Y.-M. (2001). Residential Mobility and Urban Restructuring under Market Transition: A Study of Guangzhou, China. *The Professional Geographer*. 53 (2). p.pp. 219–229.

Lowell, B.L. & Findlay, A.M. (2001). MIGRATION OF HIGHLY SKILLED PERSONS FROM DEVELOPING COUNTRIES: IMPACT AND POLICY RESPONSES. p.p. 38.

Maoh, H., Dimatulac, T., Oiamo, T. & Johnson, M. (2019). Implementation of the SMARTPLANS Integrated Urban Model for Halifax, Nova Scotia and London, Ontario. In: *Book of the Canadian Transportation Research Forum (CTRF) 54th Annual Conference*. 2019, Vancouver, British Columbia, Canada.

Maoh, H. & Gingerich, K. (2016). SMARTPLANS: SIMULATION MODEL FOR ASSESSING THE RAMIFICATIONS OF TRANSPORTATION POLICIES AND LAND USE SCENARIOS. In: *Book of the Canadian Transportation Research Forum (CTRF) 51st Annual Conference*. 2016, Toronto, Ontario, Canada.

Maoh, H.F., Koronios, M. & Kanaroglou, P.S. (2010). Exploring the land development process and its impact on urban form in Hamilton, Ontario. *The Canadian Geographer / Le Géographe canadien*. 54 (1). p.pp. 68–86.

Martin, P. & Zürcher, G. (2008). *Managing Migration: The Global Challenge*.

Mieszkowski, P. & Mills, E.S. (1993). The Causes of Metropolitan Suburbanization. *Journal of Economic Perspectives*. 7 (3). p.pp. 135–147.

Miller, E.J. (2018). Viewpoint: Integrated urban modeling: Past, present, and future. *Journal of Transport and Land Use*. [Online]. 11 (1). Available from: https://www.jtlu.org/index.php/jtlu/article/view/1273. [Accessed: 11 March 2020].

Miller, E.J., Farooq, B., Chingcuanco, F. & Wang, D. (2011). Historical Validation of Integrated Transport–Land Use Model System. *Transportation Research Record*. 2255 (1). p.pp. 91–99.

Ontario Human Rights Commission (2007). *Human Rights and rental housing in Ontario: Background paper*. [Online]. 2007. Available from: http://www3.ohrc.on.ca/en/human-rights-and-rental-housing-ontario-background-paper/rental-housing-landscape-ontario. [Accessed: 17 July 2020].

Parkins, N.C. (2010). *Push and Pull Factors of Migration*. p.p. 19.

Picot, G. (2008). Immigrant Economic and Social Outcomes in Canada: Research and Data Development at Statistics Canada. *Statistics Canada, Analytical Studies Branch, Analytical Studies Branch Research Paper Series*.

Saiz, A. (2007). Immigration and housing rents in American cities. *Journal of Urban Economics*. 61 (2). p.pp. 345–371.

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

Statistics Canada (2001; 2006; 2011; 2016). *2001-2016 Census Profile*. [Online]. 2016. Available from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMACA&Code1=205&Geo2=PR&Code2=12&SearchText=Halifax&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=205&TABID=1&type=0. [Accessed: 23 July 2020].

Statistics Canada (2019). *Results from the new Canadian Housing Survey, 2018*. [Online]. 22 November 2019. Available from: https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2019079-eng.htm. [Accessed: 24 January 2021].

Sullivan, J., Aultman-Hall, L., Troy, A., Azaria, D. & Lawe, S. (2010). *Integrated Land - Use ,Transportation and Environmental Modeling : Validation Case Studies*. [Online]. Vermont: University of Vermont. Available from: https://rosap.ntl.bts.gov/view/dot/20578. [Accessed: 20 July 2020].

The Canadian Real Estate Association (2020). *National Price Map*. [Online]. 2020. Available from: https://www.crea.ca/housing-market-stats/national-price-map/. [Accessed: 23 July 2020].

United States Census Bureau (2018). *NATIVITY AND CITIZENSHIP STATUS IN THE UNITED STATES*. [Online]. 2018. Available from: https://data.census.gov/cedsci/table?q=Citizenship&hidePreview=false&t=Citizenship&tid=ACSDT1Y2018.B05001&vintage=2018. [Accessed: 21 July 2020].

US Census Bureau (2019). *CPS Historical Migration/Geographic Mobility Tables*. [Online]. 2019. Available from: https://www.census.gov/data/tables/time-series/demo/geographic-mobility/historic.html. [Accessed: 18 July 2020].

Vliet, J. van (2013). *Calibration and validation of land-use models*. phd. [Online]. S.l.: s.n. Available from: https://library.wur.nl/WebQuery/wurpubs/432594. [Accessed: 18 July 2020].

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

Wegener, M. (1994). Operational Urban Models State of the Art. *Journal of the American Planning Association*. 60 (1). p.pp. 17–29.

CHAPTER 3

MODELING AND VALIDATING THE PRICE OF RESIDENTIAL HOUSING IN THE SMARTPLANS INTEGRATED URBAN MODEL

3.1 Introduction

Potential buyers of a property consider many different attributes in the assessment process, from size to location to nearby amenities, however price for most buyers represents a significant element in their decision. Future homeowners tend to follow the theory of utility maximization, trying to take full advantage of their hard-earned dollar to allocate the most value for their new property. Therefore, the presence of a price model within Integrated Urban Models (IUMs) has become a standard in the land use and transportation modeling space. Since the 1960s, IUMs have been frequently used as a modeling technique to formulate a relationship between land use and transportation. Recent IUM developments have incorporated some form of a price model, whether that be to estimate land price, average zonal house price, or individual dwelling prices such as UrbanSim (Waddell, 2002), MUSSA (Martinez, 1996), and more recently SMARTPLANS (Maoh et al. 2019).

SMARTPLANS is a full-fledge IUM that utilizes land use patterns and transportation systems to simulate the affects on sustainability (Maoh et al. 2019). To date, SMARTPLANS has been applied to five Canadian cities: Calgary, AB, London, ON, Ottawa, ON, Halifax, NS, and Vancouver, BC. It has the capability to be programed and implemented for limitless regions since the parameters are configurable instead of being hardcoded into the model (Maoh and Gingerich, 2016). A dedicated price model is built-in within the land use module, as shown in Figure 1-1.

Regression techniques are often implemented to estimate price models due to their simplicity and effectiveness (Gingerich et al., 2013). A great deal can be understood about the different determinant attributes comprising the regression model, however, developments in the field of spatial statistics and advancements in geographic information system (GIS) has opened the doors for using more sophisticated models such as simultaneous autoregressive (SAR) models (Wilhelmsson, 2002). Examples of implementing spatial techniques within price models include Martínez and Viegas (2009) and Bidanset and Lombard (2014).

This chapter reports on the outcomes achieved from calibrating and validating a series of price models for the Canadian cities of Calgary, AB and Ottawa, ON. The best performing and most efficient model will be incorporated into the land price model component of SMARTPLANS to improve the IUMs predictive ability. Historical data obtained from the Canadian census for the years 2006, 2011, and 2016 will be used to calibrate a series of Ordinary Least Squares (OLS) and Simultaneous Autoregressive (SAR) models using attributes that directly influence land prices. The model parameters will be compared together to test for significance and confirm whether they change over time. Finally, the 2006 model will be used to estimate 2011 and 2016 price figures and the results will be validated using official Canadian census values.

Succeeding information in this chapter is organized as follows. The second section will provide a background on factors impacting house prices and spatial nature of house prices. The third section will present a description of the study areas. Next, the model development and the methods used to compare the parameters will be discussed. The fifth section presents and discusses the attained results, and the final section provides conclusions and directions for future research.

3.2 Background

3.2.1 Factors Impacting House Prices

Numerous researchers have studied the influence of transit on land values. Many have discovered a positive influence on property values (McMillen and McDonald, 2004; Bartholomew and Ewing, 2011) whereas some have witnessed weaker impacts (Gatzlaff and Smith, 1993). Gallo (2018) estimated a hedonic model to explore the effects of transit on real estate property values and applying the model to the city of Naples. Several external parameters where incorporated and selected based on correlation to determine which transit system was significant. The three major transit systems explored were: high-frequency metro, low-frequency metro, and bus lines. Similarly, Hopkins (2018) implemented a hedonic regression model to study the possible relationship between housing values and their proximity to transit on 25 metro areas around the United States. With a proximity of a half mile from the transit stop, results show that six regions significantly impact the price of a property. Additionally, Kim and Lahr (2014) explain that the value of a property decreases as the distance from a transit stop to the property increases. Whether there exists a positive or negative impact of transit on property values, it remains an essential factor to consider in a price model.

Household income is another fundamental factor to consider in the estimation of house prices. One study conducted by Gallin (2006) indicates that house prices and household income are not cointegrated for 95 U.S metropolitan areas. Conversely, many studies have shown that there exists a long-term relationship between income and house prices (Abraham and Hendershott, 1996; Meen, 2002). The ability to own a home is directly influenced by its price which, as a consequence to rising home prices, a larger income is required to sustain the mortgage payments (Linneman and Megbolugbe, 1992; Chen et al., 2007). As such, many low-income families experience affordability difficulties in the housing industry. Moreover, Chen et al. (2007) found that there does exist an equilibrium trend between house prices and income in the long run and increases in income allow housing to become more attainable.

An additional component impacting house prices is derived from urban theory, such as suburbanization or distance from the central business district (CBD). The luxury of a suburban lifestyle has been the desire of many homeowners in North America. Larger lots, reduced noise levels, cheaper land, and easy accessibility to the core have increased the attractiveness of suburban living (Jansen, 2020). Many business sectors are relocating or opening branches in the suburbs due to lower prices and consequently causing a disruption in the urban form (Margulis, 2002). Such increase in demand triggered upward pressure on housing prices (Voith, 1999). Helbich (2015) found that the suburban properties indicate an independent housing market which significantly impact overall house prices.

Some studies have shown the significant impact of schools on house prices. Haurin and Brasington (1996) observed that there exists a direct relationship between quality schools and house prices where schools with significantly higher passing rates increase neighboring house prices. Moreover, Kiel and Zabel (2008) indicate that the most important determinant of house prices is its location and that certain school districts influence its value. On the other hand, one study conducted by Livy (2017) show evidence of periodic influence of quality schools on house prices. During times of market declines, proximity to quality schools significantly impact neighboring housing values but show no relationship during market inclines.

The expansion of urban areas and residential neighbourhood development must go through rigorous planning in order to maximize the value of the properties. The real estate market is said to be determined by location which adds to the fact that locational characteristics should be of great value in house price modeling (Cohen and Coughlin, 2008). Property planners and realtors consider many different neighbourhood qualities when evaluating their prices. Proximity to schools, transit, entertainment, as well as accessibility and noise levels are key factors that influence the price of houses (McMillen and McDonald, 2004; Kiel and Zabel, 2008; Cohen and Coughlin, 2008).

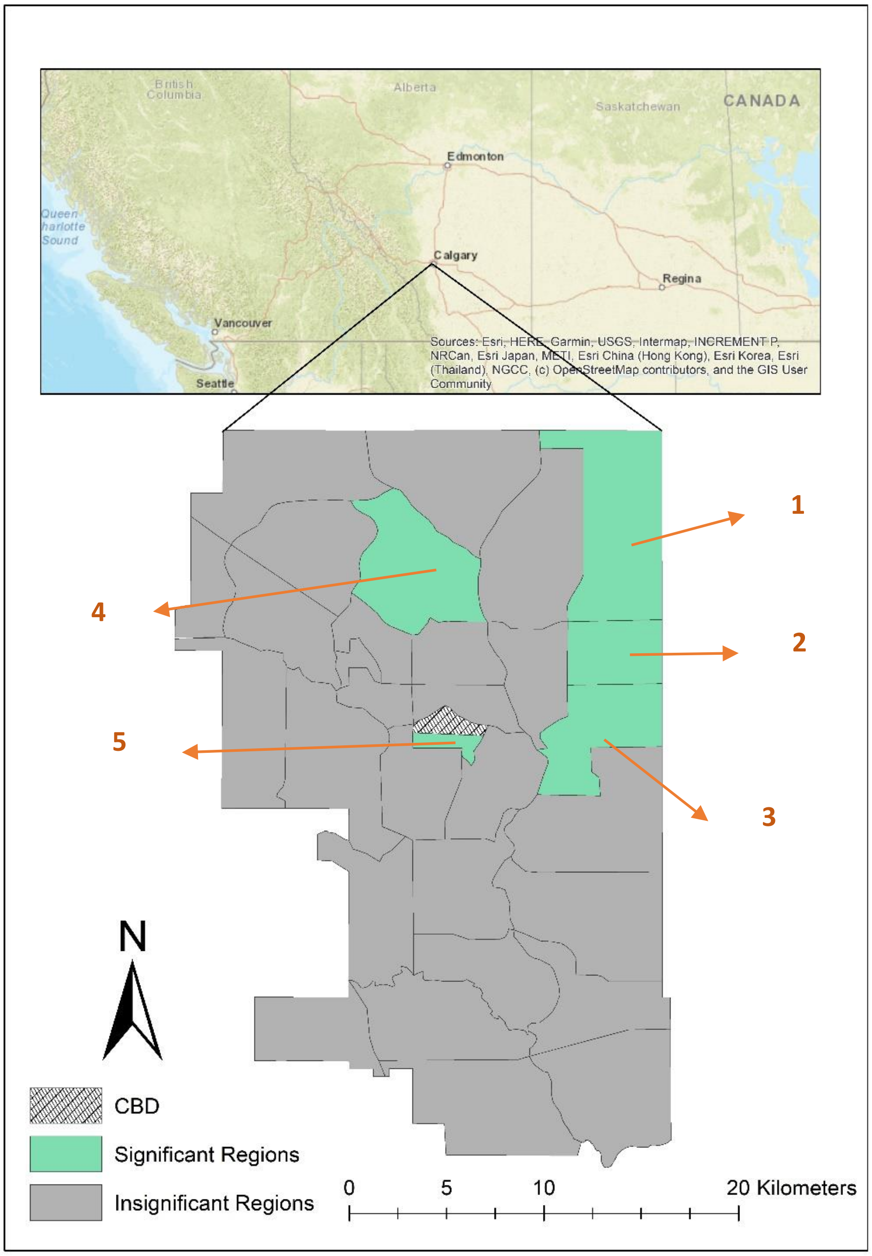
3.2.2 Spatial Nature of Housing Prices

A key determinant of property and residential neighbourhood prices is its location. As such, location is an important element to consider in any house price modeling since a strong relationship between them exists (Spinney et al., 2014). The presence of spatial autocorrelation (SA) in property prices has long been a challenging factor when determining the true value of a property (Potoglou et al., 2018). In fact, the general lack of success in incorporating the neighbourhood influence on house prices is mainly due to the complexity in evaluating them (Dubin, 1992). SA exists when the price in one region is directly influenced by that in a neighbouring location. Moran's I statistic is the most common test used to determine the existence of SA in the data (Kelejian and Prucha, 2001). This has opened the doors for an inflow of location-based modeling, such as simultaneous autoregressive (SAR) models. Such models have recently been utilized in modeling real estate markets to improve their performance when compared to the widely used traditional regression models, like the works presented in Osland (2010) and Bourassa et al. (2010). The work reported in Gingerich et al. (2013), Spinney et al. (2014) and more recently in Potoglou et al. (2018) encourages the application of spatial methods in land price models. Given the inherently spatial nature of land prices, the application of spatial regression models is expected to improve the conducted analysis.

3.3 Study Areas

This chapter focuses on two major Canadian cities: Calgary, Alberta and Ottawa, Ontario. Calgary is the most populous city in Alberta, located west of Canada, and has a land area of about 5,110km^2. Ottawa is the capital of Canada and located in northeast Ontario with an approximate land area of 6,767km^2. Figures 3-1 and 3-2 display maps of both Calgary and Ottawa, respectively, highlighting the CBD and significant regions when applied into the price models.

The city of Calgary is primarily known for its activity in the energy sector, more specifically the oil industry. The steady increase in oil prices in the last few decades triggered high inflow of population into the region, increasing the inhabitants by more than 46% between the years 2001 and 2016 (Statistics Canada, 2001; 2016). During the same period, average dwelling values in Calgary increased from just under \$202,000 to over \$527,000. The city of Ottawa has experienced relatively strong economic growth in the last decade with many large-scale transportation projects under construction. The Ottawa region is home to approximately 300,000 new residents between census years 2001 and 2016 (Statistics Canada, 2001; 2016). The housing market saw a healthy growth in dwelling prices, increasing from an average of about \$174,000 to almost \$400,000 during the same period.



Key:

1) Saddle Ridge

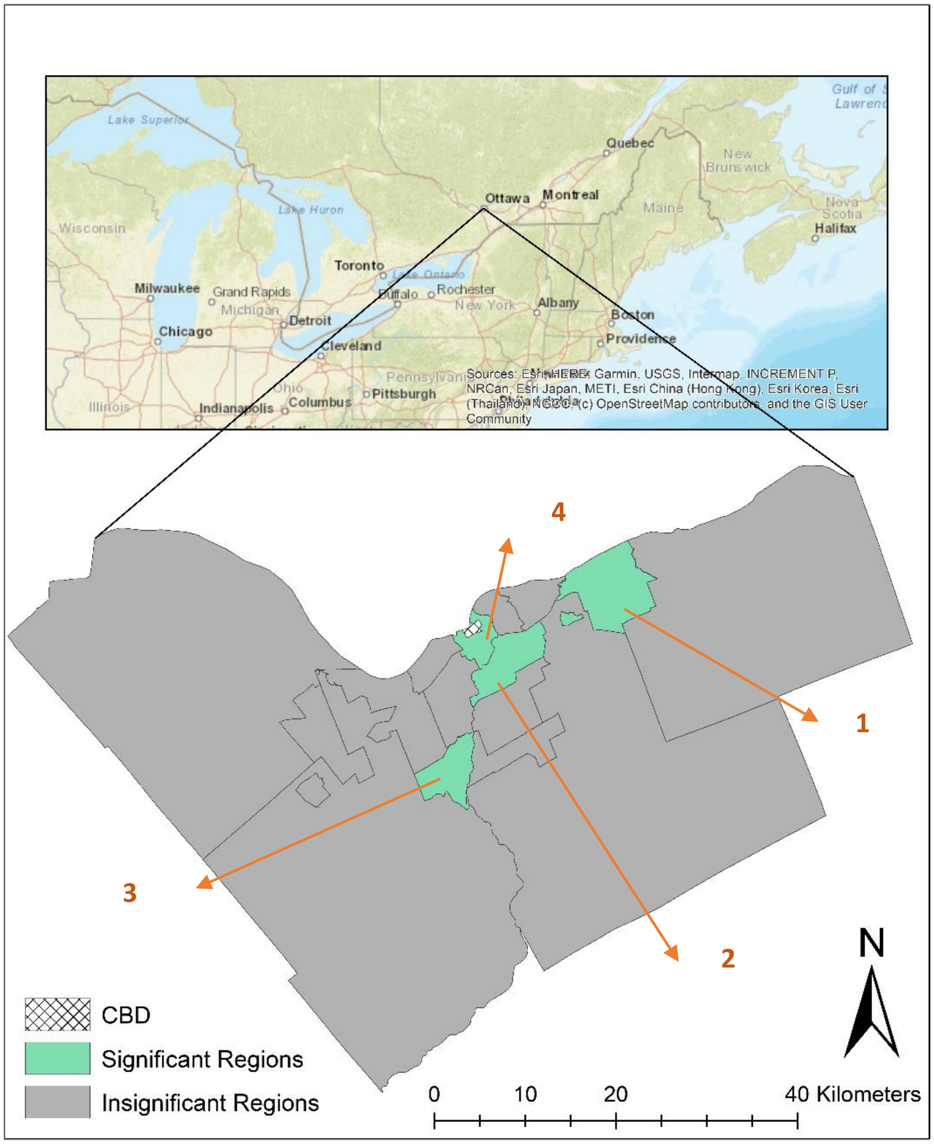
2) The Properties

3) Forest Lawn

4) Beddington

5) CBD Fringe

Figure 3 - 1 Map of Calgary, AB



Key:

1) Orleans

2) Alta Vista

3) South Nepean

4) Ottawa Inner Area

Figure 3 - 2 Map of Ottawa, ON

3.4 Methods of Analysis

3.4.1 Specification of the Land Price Model

The data used in the land price model involved average residential house prices at the Dissemination Area (DA) level obtained from the Canadian census for years 2006, 2011, and 2016. According to the 2016 census, approximately 89% of the Calgary CMA population (75% of the Ottawa CMA) reside in the city and therefore estimations were performed exclusively on the city as opposed to the CMA.

The covariates used in the analysis were primarily selected using the existing literature as a guide. Several categorical variables were implemented where a value of 1 indicates a factual record and 0 otherwise. Region-based parameters and buffer variables were also applied to the analysis. The list of covariates used in the land price model was shaved down to exclude insignificant parameters and ones observing high levels of multicollinearity. Additionally, DAs with no land price data or ones observing suppressed data (regions with 0 land price values) were excluded from the analysis to improve the model's predictive ability. Table 3-1 displays the non-regional covariates used in the land price models for both Calgary and Ottawa.

Table 3 - 1 Covariates used in the Specification of the Land Price Model

Covariate	Description
Detached	Number of detached houses in the DA
Apartment	Number of apartments in the DA
Semi-Detached	Number of semi-detached houses in the DA
Row-House	Number of row-houses in the DA
H.H Size	Average household size in the DA
Income	Average household income in the DA
School	DAs within 500m buffer of an education facility; 1 if true, 0 otherwise
DistCBD	Euclidian distance between the centroid of the DA and the centroid of the CBD, in kilometers
Bus Stop	DAs that contain at least one bus stop; 1 if true, 0 otherwise

Initially, an Ordinary Least Square (OLS) regression was used to develop the land price model. The formulation of such model is as follows:

$$\boldsymbol{P} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \text{(Eq. 1)}$$

where:

$\boldsymbol{P}$ is an $n \times 1$ vector representing average housing price in each DA i ($i = 1, 2, \ldots, n$).

$\boldsymbol{X}$ is an $n \times (k + 1)$ matrix representing the k independent covariates measured at DA level. First column is set to unity to account for the constant in the model.

$\boldsymbol{\beta}$ is an $(1 + k)$ coefficients associated with covariates k. First β coefficient represents the constant.

$\boldsymbol{\varepsilon}$ is an $n \times 1$ vector representing the error terms. These terms are assumed to be normally distributed with zero mean and constant variance.

Due to the spatial nature of house prices, a spatial lag model was also implemented to accommodate for the presence of spatial autocorrelation. Such a model is used when the dependent variable P_i for a given DA i is highly correlated with the price in neighboring DAs j. The Simultaneous Autoregressive (SAR) model was applied here and takes the following form:

$$\boldsymbol{P} = \boldsymbol{X}\boldsymbol{\beta} + \rho \boldsymbol{W}\boldsymbol{P} + \boldsymbol{\varepsilon} \qquad \text{(Eq. 2)}$$

where ρ is a spatial lag parameter, $\boldsymbol{W}$ is a weight matrix with elements w_{ij} capturing the relationship between neighboring DAs i and j. All other terms are previously defined.

3.4.2 Parameter Comparison Over Time

To explore if a parameter is significantly different between two time periods, the Wald Chi-Square test is regularly exercised. This technique will help determine if the 2006, 2011, and 2016 parameters are significantly different from each other. A major contributor of prediction accuracy is the span of the simulation period, where an increase between the base year and target year increases the probability of error accumulation. Here, we try to investigate if there exists a significant difference in parameter values by estimating the land price model with data from different time periods. The formula for this statistic is:

$$\chi^2 = \frac{(\beta_{t+1} - \beta_t)^2}{[se(\beta_{t+1})]^2 + [se(\beta_t)]^2} \qquad \text{(Eq. 3)}$$

where

β_{t+1} = coefficient for the target year parameter

β_t = coefficient for the base year parameter

$se(.)$ = standard error for each period

When estimating χ^2, the objective is to test the null hypothesis that β_{t+1} is no different from β_t. If χ^2 is greater than the critical values of significance (2.706 and 3.841 for 90% and 95% levels, respectively), we conclude that the two parameters are statistically different.

3.5 Model Estimation Results

3.5.1 Model Calibration and Comparison

Numerous testing was conducted on the land price model to verify which modeling approach provides respectable results combined with the most efficient analysis method. OLS and SAR modeling techniques (with and without region-specific parameters) were performed to test significance and predictive ability, summarized in Table 3-2 for Calgary and Table 3-3 for Ottawa. Initial analysis on the land price model were conducted using the OLS approach without region-specific parameters, labelled OLS-1. The R^2 coefficient for the analysis years 2006, 2011, and 2016 is 0.35, 0.61, and 0.72, respectively for Calgary and 0.52, 0.49, and 0.58, respectively for Ottawa.

To improve the results, a series of region-based parameters for both Canadian cities were incorporated into the models, identified as OLS-2. Both Calgary and Ottawa have predefined regions as outlined locally and a series of regions were implemented to the land price model based on their significance (refer to Figures 3-1 and 3-2). Five regions were incorporated into the Calgary model (namely: Forest Lawn, Properties, Saddle Ridge, Beddington, and CBD Fringe) and four regions in the Ottawa model (namely: Alta Vista, Orleans, Ottawa Inner Area, and South Nepean). These regions were treated as categorical variables, having a value of 1 if the DA falls in the specified region, 0 otherwise. The R^2 coefficient improved to 0.45, 0.65, and 0.77 for Calgary and 0.58, 0.57, and 0.65 for Ottawa for analysis years 2006, 2011, and 2016, respectively.

Next, two SAR models were tested to explore whether the results would improve when compared to the OLS model with region-specific parameters. Again, one model does not incorporate region-specific parameters (SAR-1) while the other does (SAR-2). Based on the obtained findings, both models perform very similarly and show improved results when compared to the OLS-2 model. The analysis suggests that the SAR modeling techniques does a superior job at accounting for spatial autocorrelation even without incorporating region-based parameters.

Table 3 - 2 Parameter Estimates of the Land Price Models for Calgary, AB

Covariates	OLS-1			OLS-2			SAR-1			SAR-2		
	2006	2011	2016	2006	2011	2016	2006	2011	2016	2006	2011	2016
Constant	2.5472	3.1055	3.4653	1.7088	2.6716	2.8812	0.1437	1.1018	1.0730	-0.0842	1.1095	1.1197
	(12.10)	(16.63)	(17.48)	(8.19)	(13.88)	(14.09)	(0.82)	(5.81)	(5.66)	(-0.45)	(5.50)	(5.44)
Rho (ρ)	--	--	--	--	--	--	0.6205	0.4187	0.4724	0.5404	0.3562	0.3467
							(27.33)	(18.79)	(23.55)	(21.71)	(14.81)	(17.52)
Detached	0.0011	0.0013	0.0009	0.0007	0.0008	0.0006	0.0007	0.0009	0.0007	0.0005	0.0006	0.0005
	(5.28)	(5.11)	(5.48)	(3.55)	(3.17)	(3.79)	(4.17)	(3.90)	(4.75)	(3.12)	(2.80)	(3.70)
Apartment	-0.0012	-0.0007	-0.0014	-0.0005	-0.0002	-0.0004	-0.0004	-0.0004	-0.0006	-0.0003	-0.0002	-0.0002
	(-2.30)	(-1.53)	(-3.79)	(-0.86)	(-0.32)	(-1.06)	(-1.06)	(-1.04)	(-1.78)	(-0.61)	(-0.43)	(-0.49)
Semi-Detached	-0.0031	-0.0016	-0.0001	-0.0018	-0.0006	0.0005	-0.0018	-0.0010	0.0001	-0.0012	-0.0005	0.0005
	(-2.35)	(-1.50)	(-0.15)	(-1.47)	(-0.64)	(0.66)	(-1.74)	(-1.04)	(0.08)	(-1.14)	(-0.51)	(0.62)
Row-House	-0.0033	-0.0034	-0.0034	-0.0021	-0.0028	-0.0027	-0.0025	-0.0030	-0.0031	-0.0020	-0.0027	-0.0027
	(-4.76)	(-5.31)	(-6.14)	(-3.31)	(-4.65)	(-5.18)	(-4.62)	(-5.38)	(-6.59)	(-3.66)	(-4.90)	(-5.81)
H.H Size	0.3114	-0.0192	0.1019	1.1091	0.5141	0.7483	0.4871	0.2007	0.3294	0.9250	0.4953	0.6878
	(3.9)	(-0.28)	(1.44)	(12.13)	(6.15)	(8.73)	(7.53)	(3.27)	(5.47)	(11.50)	(6.39)	(8.78)
Income*	1.4084	1.9324	1.8500	0.9022	1.6871	1.6369	0.7020	1.3872	1.3512	0.5154	1.3233	1.3193
	(19.42)	(42.70)	(56.08)	(12.28)	(35.57)	(49.15)	(11.90)	(30.39)	(39.60)	(8.17)	(28.22)	(38.65)
School	-0.2703	-0.0500	-0.1598	-0.2098	-0.0282	-0.1082	-0.1961	-0.0201	-0.1027	-0.1728	-0.0131	-0.0814
	(-3.64)	(-0.76)	(-2.45)	(-3.07)	(-0.45)	(-1.79)	(-3.32)	(-0.34)	(-1.85)	(-2.97)	(-0.22)	(-1.49)
DistCBD	-0.1308	-0.0961	-0.1191	-0.1798	-0.1368	-0.1735	-0.0873	-0.0736	-0.0841	-0.1189	-0.1009	-0.1227
	(-11.47)	(-10.15)	(-12.81)	(-16.55)	(-14.35)	(-18.83)	(-9.26)	(-8.63)	(-10.36)	(-12.09)	(-10.97)	(-13.72)
Bus Stop	0.2966	0.1021	-0.0538	0.2845	0.0959	-0.0868	0.1886	0.0611	-0.1159	0.2028	0.0660	-0.1192
	(3.16)	(1.24)	(-0.58)	(3.28)	(1.22)	(-1.01)	(2.53)	(0.84)	(-1.47)	(2.75)	(0.91)	(-1.54)
Forest Lawn	--	--	--	-1.6391	-1.1839	-1.5828	--	--	--	-0.8042	-0.6433	-0.8921
				(-13.76)	(-10.33)	(-14.36)				(-7.37)	(-5.73)	(-8.25)
Properties	--	--	--	-1.4428	-1.0962	-1.3294	--	--	--	-0.8582	-0.6669	-0.8099
				(-9.70)	(-7.76)	(-9.69)				(-6.56)	(-4.96)	(-6.28)
Saddle Ridge	--	--	--	-1.5402	-0.9885	-1.0852	--	--	--	-0.9791	-0.6540	-0.7283
				(-8.85)	(-5.98)	(-6.56)				(-6.48)	(-4.22)	(-4.80)
Beddington	--	--	--	-1.1122	-0.9411	-0.9426	--	--	--	-0.5736	-0.5750	-0.5118
				(-8.72)	(-7.66)	(-8.01)				(-5.16)	(-4.93)	(-4.68)
CBD Fringe	--	--	--	-0.7571	-0.8453	-1.3830	--	--	--	-0.1822	-0.4135	-0.7207
				(-2.90)	(-3.44)	(-5.60)				(-0.82)	(-1.81)	(-3.18)
No. of Obs.	1388	1560	1561	1388	1560	1561	1388	1560	1561	1388	1560	1561
R^2	0.345	0.606	0.724	0.455	0.646	0.768	0.587	0.690	0.799	0.604	0.698	0.808

**Parameter scaled by 100,000*

Table 3 - 3 Parameter Estimates of the Land Price Models for Ottawa, ON

Covariates	OLS-1			OLS-2			SAR-1			SAR-2		
	2006	2011	2016	2006	2011	2016	2006	2011	2016	2006	2011	2016
Constant	2.8102	3.5757	4.2793	1.9982	2.5668	2.9995	1.2685	0.9151	0.9934	1.0322	0.6772	0.7594
	(20.04)	(20.18)	(19.84)	(13.47)	(13.98)	(13.92)	(8.55)	(5.55)	(5.27)	(6.66)	(3.92)	(3.89)
Rho (ρ)	--	--	--	--	--	--	0.4347	0.5730	0.5891	0.3482	0.5091	0.5133
							(15.85)	(25.43)	(28.64)	(12.07)	(21.33)	(23.24)
Detached	0.0015	0.0024	0.0025	0.0015	0.0024	0.0022	0.0011	0.0018	0.0017	0.0012	0.0019	0.0017
	(6.16)	(7.61)	(8.17)	(6.84)	(8.10)	(8.13)	(5.20)	(7.07)	(7.13)	(5.88)	(7.45)	(7.21)
Apartment	-0.0002	0	-0.0002	0.0001	0.0002	-0.0001	-0.0001	0.0001	-0.0002	0	0.0002	-0.0001
	(-0.99)	(-0.07)	(-0.93)	(0.43)	(1.11)	(-0.33)	(-0.86)	(0.62)	(-1.00)	(0.17)	(1.34)	(-0.59)
Semi-Detached	-0.0015	-0.0012	0	-0.0018	-0.0015	-0.0004	-0.0013	-0.0010	0.0001	-0.0016	-0.0012	-0.0002
	(-1.92)	(-1.32)	(0.04)	(-2.53)	(-1.75)	(-0.50)	(-1.87)	(-1.32)	(0.15)	(-2.37)	(-1.67)	(-0.22)
Row-House	-0.0017	-0.0029	-0.0027	-0.0015	-0.0026	-0.0025	-0.0015	-0.0023	-0.0021	-0.0014	-0.022	-0.0021
	(-6.88)	(-9.43)	(-8.93)	(-6.42)	(-9.04)	(-8.95)	(-6.79)	(-9.34)	(-8.98)	(-6.48)	(-9.12)	(-8.99)
H.H Size	-0.3840	-0.1881	-0.3697	-0.1211	0.1365	0.0577	-0.1706	0.1107	0.0297	-0.329	0.2557	0.2022
	(-6.67)	(-2.69)	(-4.31)	(-2.08)	(1.95)	(0.69)	(-3.29)	(1.98)	(0.45)	(-0.61)	(4.33)	(2.92)
Income*	1.8586	1.7775	2.1621	1.7949	1.6919	2.0576	1.3775	1.1142	1.4036	1.4307	1.1417	1.4469
	(25.82)	(23.18)	(29.76)	(26.55)	(23.57)	(30.90)	(19.88)	(17.27)	(23.24)	(21.09)	(17.92)	(24.34)
School	-0.1615	-0.2203	-0.2085	-0.1458	-0.2078	-0.2054	-0.1151	-0.1422	-0.1164	-0.1111	-0.1423	-0.1257
	(-3.58)	(-3.94)	(-3.21)	(-3.41)	(-3.97)	(-3.45)	(-2.85)	(-3.17)	(-2.32)	(-2.79)	(-3.22)	(-2.54)
DistCBD	-0.0319	-0.0575	-0.0758	-0.0252	-0.0468	-0.0598	-0.0253	-0.0407	-0.0491	-0.0225	-0.0375	-0.0446
	(-9.80)	(-14.22)	(-16.48)	(-7.81)	(-11.63)	(-13.22)	(-8.55)	(-12.04)	(-13.00)	(-7.46)	(-10.86)	(-11.62)
Bus Stop	-0.1395	-0.3425	-0.4337	-0.0497	-0.2125	-0.2574	-0.0903	-0.2213	-0.2646	-0.0395	-0.1636	-0.1955
	(-2.50)	(-4.58)	(-5.02)	(-0.94)	(-3.03)	(-3.24)	(-1.81)	(-3.69)	(-3.96)	(-0.81)	(-2.76)	(-2.97)
Alta Vista	--	--	--	-0.2056	-0.3295	-0.3718	--	--	--	-0.1726	-0.2403	-0.2259
				(-2.85)	(-3.54)	(-3.52)				(-2.57)	(-3.06)	(-2.57)
Orleans	--	--	--	-0.4114	-0.4723	-0.6558	--	--	--	-0.3064	-0.2817	-0.3482
				(-6.54)	(-6.35)	(-7.87)				(-5.15)	(-4.44)	(-4.93)
Ottawa Inner Area	--	--	--	0.7710	1.0799	1.3764	--	--	--	0.4983	0.5593	0.6936
				(10.37)	(11.34)	(12.70)				(6.82)	(6.62)	(7.30)
South Nepean	--	--	--	-0.2371	-0.3775	-0.5375	--	--	--	-0.1688	-0.2135	-0.2956
				(-3.00)	(-3.96)	(-4.98)				(-2.28)	(-2.64)	(-3.27)
No. of Obs.	1190	1305	1300	1190	1305	1300	1190	1305	1300	1190	1305	1300
R^2	0.517	0.495	0.577	0.579	0.565	0.651	0.611	0.674	0.746	0.631	0.687	0.758

** Parameter scaled by 100,000*

At first glance, all models are relatively in-line with each other and are well behaved. The models provide important feedback on the performance of several parameters with regards to land prices in both Calgary and Ottawa. The number of detached dwellings and household income positively and significantly impact land prices in all models. Moreover, the number of row houses and distance from the CBD decreases the value of land prices in all models. Such a result is fairly intuitive and in parallel with the literature. Both apartment and semi-detached residence units showed almost no impact in the land price models. Household size and proximity to bus stops show a mix of positive and negative impacts on land price and therefore do not skew the results to a specific direction. Interestingly, the results show that DAs with proximity to schools suggest a general negative impact on prices. A possible reason behind such an outcome is twofold: 1) the general school quality in both regions is relatively poor or, 2) the residence units in the DAs within close proximity to schools are smaller, older properties. Region specific parameters in both Calgary and Ottawa mainly negatively impact land prices. This is in-line with the *DistCBD* parameter and further suggests that other things being equal, prices are lower in areas outside the city's core, except for the CBD Fringe parameter in Calgary's case. Lastly, all models experience highly significant ρ parameter suggesting that the observations are not independent and spatial autocorrelation is at play here.

3.5.2 Model Comparison over Time

Tables 3-4 and 3-5 display the Wald Chi-Square statistics comparing both 2006-2011 and 2011-2016 model parameters for Calgary and Ottawa, respectively. The results show a very interesting trend among both regions. Significant parameters (meaning they change over time) are mainly present in the 2006-2011 models whereas are seldom present in the 2011-2016 models. One possibility of such an observation occurring is the severe increase in Canada's population during the period 2006 to 2011. According to Statistics Canada, the population grew by almost 5.9% between census years 2006 and 2011 compared to only about 1% between census years 2011 and 2016 (2006; 2011; 2016). Such a large population inflow into Canada is expected to influence the demand for housing and consequently prices. Therefore, the models estimated using the 2011 data behave differently when compared to the models estimated using the 2006 data. By comparison, the small growth in population during the period 2011 – 2016 stabilized the prices and as such the models estimated using the 2011 and 2016 datasets did not change significantly. This allows us to conclude that if an urban area is not going to experience dramatic population growth over time, then the estimated models can be used to predict future land prices within an IUM.

Table 3 - 4 Wald Chi-Square Test for 2006-2011 and 2011-2016 Model Parameters, Calgary

Covariates	2006-2011				2011-2016			
Model Technique	OLS-1	OLS-2	SAR-1	SAR-2	OLS-1	OLS-2	SAR-1	SAR-2
Rho (ρ)	--	--	40.24*	28.34*	--	--	0.09	0.09
Detached	0.45	0.11	0.53	0.22	1.15	0.27	0.37	0.09
Apartment	0.56	0.18	0.00	0.02	1.59	0.18	0.10	0.00
Semi-Detached	0.84	0.54	0.37	0.27	1.10	0.84	0.75	0.62
Row-House	0.01	0.65	0.42	0.97	0.00	0.03	0.02	0.00
H.H Size	9.78*	23.06*	10.32*	14.79*	1.49	3.83	2.24	3.05
Income	37.58*	80.60*	84.41*	105.66*	2.17	0.75	0.40	0.01
School	4.91*	3.80	4.47*	3.78	1.40	0.84	1.04	0.73
DistCBD	5.49*	8.86*	1.17	1.79	3.02	7.66*	0.79	2.90
Bus Stop	2.42	2.60	1.49	1.76	1.57	2.46	2.71	3.04
Forest Lawn	--	7.59*	--	1.06	--	6.30*	--	2.55
Properties	--	2.86	--	1.04	--	1.40	--	0.59
Saddle Ridge	--	5.28*	--	2.26	--	0.17	--	0.12
Beddington	--	0.93	--	0.00	--	0.00	--	0.16
CBD Fringe	--	0.06	--	0.53	--	2.38	--	0.91

***** *significantly different at the 0.05 level*

Table 3 - 5 Wald Chi-Square Test for 2006-2011 and 2011-2016 Model Parameters, Ottawa

Covariates	2006-2011				2011-2016			
Model Technique	OLS-1	OLS-2	SAR-1	SAR-2	OLS-1	OLS-2	SAR-1	SAR-2
Rho (ρ)	--	--	15.17*	18.46*	--	--	0.28	0.02
Detached	5.24*	5.07*	4.03*	3.57	0.03	0.11	0.12	0.35
Apartment	0.31	0.37	1.06	0.80	0.43	0.98	1.33	1.79
Semi-Detached	0.05	0.09	0.11	0.17	0.89	0.75	1.07	1.03
Row-House	9.02*	8.73*	5.75*	6.09*	0.27	0.12	0.41	0.24
H.H Size	4.68*	8.01*	13.60*	12.94*	2.70	0.52	0.87	0.35
Income	0.59	1.09	7.74*	9.64*	13.26*	13.95*	10.73*	12.27*
School	0.67	0.84	0.20	0.27	0.02	0.00	0.15	0.06
DistCBD	24.3*	17.55*	11.76*	10.64*	8.94*	4.58*	2.70	1.91
Bus Stop	4.74*	3.44	2.82	2.60	0.64	0.18	0.23	0.13
Alta Vista	--	1.11	--	0.43	--	0.09	--	0.02
Orleans	--	0.39	--	0.08	--	2.70	--	0.49
Ottawa Inner Area	--	6.54*	--	0.30	--	4.23*	--	1.12
South Nepean	--	1.28	--	0.17	--	1.23	--	0.46

***** *significantly different at the 0.05 level*

3.5.3 Model Validation

Table 3-6 represents the results obtained from validating all the models. To validate the models, the calibrated 2006 parameters were used to predict 2011 and 2016 prices. These predictions were then compared to official data from the Canadian census for the years 2011 and 2016, respectively. When performing the predictions, only significant parameters at the 90% confidence interval in the 2006 model were utilized. Predictions for the OLS case were done by applying the linear model arithmetically using the 2006 significant parameters. On the other hand, the predictions for the SAR case were done with the help of matrix algebra, that is:

$$\widehat{\boldsymbol{P}} = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}(\boldsymbol{X\beta}) \qquad \text{(Eq. 4)}$$

Where $\widehat{\boldsymbol{P}}$ is a vector with predicted price values at the DA level, $\boldsymbol{I}$ is an identify matrix, ρ is the estimated spatial lag parameter from the 2006 model, $\boldsymbol{\beta}$ is a vector of significant parameters, and $\boldsymbol{X}$ is a matrix of the significant covariates for a given year. The accuracy values were obtained using the following expression:

$$A = 1 - \left(\frac{\sigma}{\bar{x}}\right) * 100 \qquad \text{(Eq. 5)}$$

where σ is the standard deviation of both observed and predicted prices, and $\bar{x}$ is the average of the observed prices. The results obtained from the validation show that all models have decent predictive abilities, ranging between 65-75% in accuracy. Additionally, the SAR models offer improved performance when compared to their OLS counterparts in 7 out of the 8 models with improvements ranging from around 1% to about 4% in accuracy.

Table 3 - 6 Model Accuracy (%) for Calgary and Ottawa

	Year	**Calgary**	**Ottawa**
OLS-1	2011	66.9	69.9
	2016	67.2	65
SAR-1	2011	65.4	72.2
	2016	68.6	67.8
Difference = (SAR-1) – (OLS-1)	2011	-1.5	2.3
	2016	1.4	2.8
OLS-2	2011	68.2	71.3
	2016	67.6	66.1
SAR-2	2011	69.5	74.7
	2016	70.1	69.9
Difference = (SAR-2) – (OLS-2)	2011	1.3	3.4
	2016	2.5	3.8

Figure 3-3 presents the scatter plots depicting the relation between the observed and predicted values obtained from the OLS-2 model specification for the two study areas. Also, Figure 3-4 presents the scatter plots associated with the predictions obtained from the SAR-1 model specification for the two study areas. The results further solidify the stability of the model predictions as the correlation between the observed and predicted results hovers around 0.72 – 0.80. Further, the SAR-1 model shows more accurate results when compared to the OLS-2 results in three out of four scatters, namely: (b); (c); and (d). Figures 3-5 and 3-6 present the prediction error (i.e., residuals) in the 2016 land prices in the city of Calgary and Ottawa, respectively. The error, which is based on the difference between the predicted and observed prices, was normalized using the observed 2016 price values to obtain the percent values shown. The patterns suggest that most of the errors are below the 50% level in both study areas. Further, the SAR-1 model results show an overall improvement in the accuracy when compared to the OLS-2. This is visible through the reduction in the number of red DAs throughout the study area, as well as DAs that were orange in the OLS model map but became green in the SAR model map (i.e., smaller residuals).

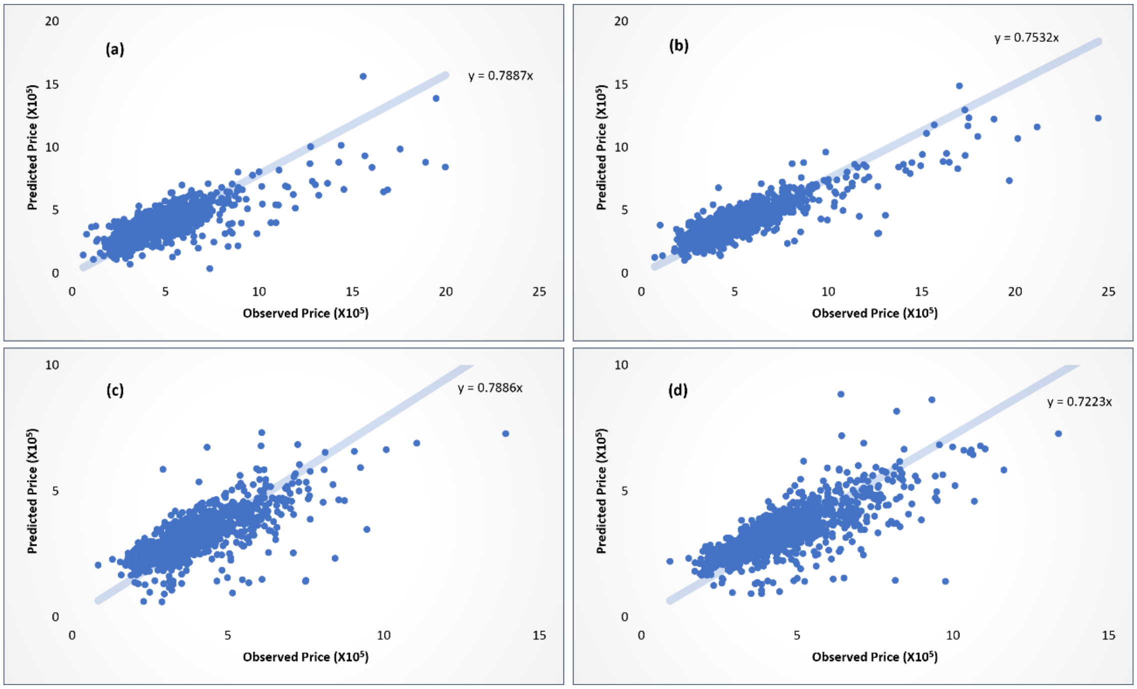


Figure 3 - 3 OLS-2 Validation Scatters for (a) 2011 Calgary; (b) 2016 Calgary; (c) 2011 Ottawa; (d) 2016 Ottawa

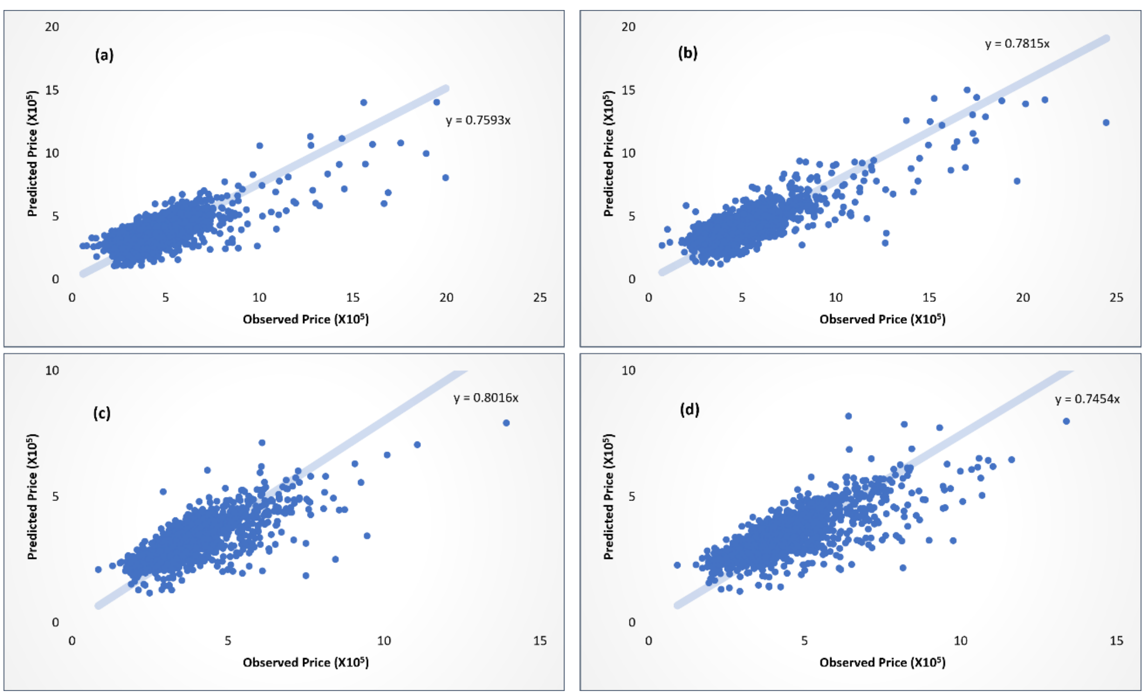
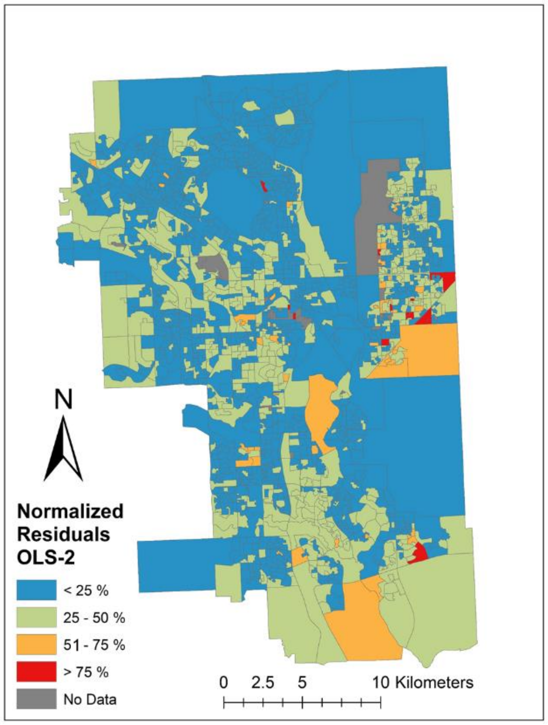


Figure 3 - 4 SAR-1 Validation Scatters for (a) 2011 Calgary; (b) 2016 Calgary; (c) 2011 Ottawa; (d) 2016 Ottawa



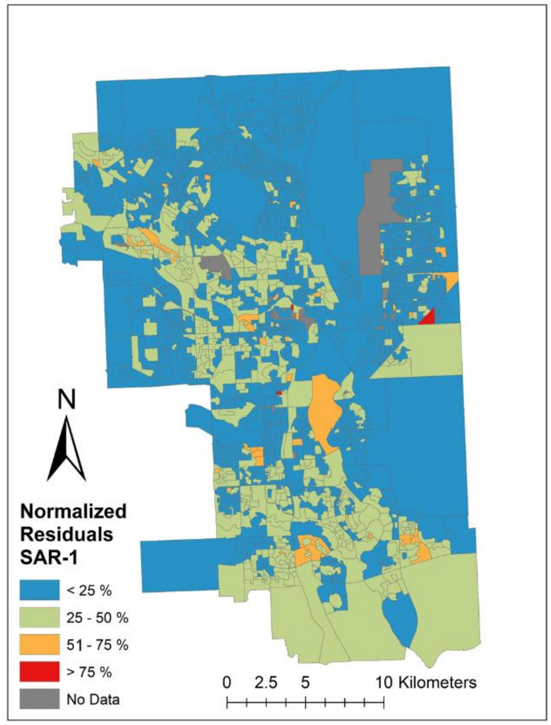
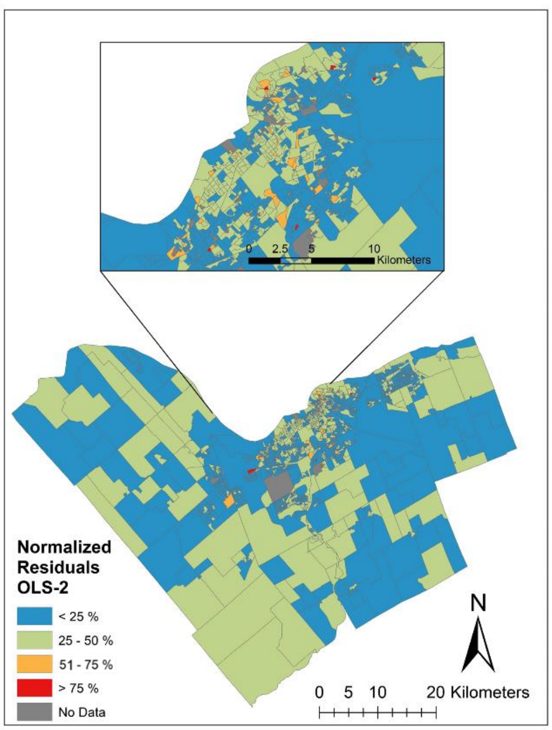


Figure 3 - 5 2016 Normalized Residuals of Calgary



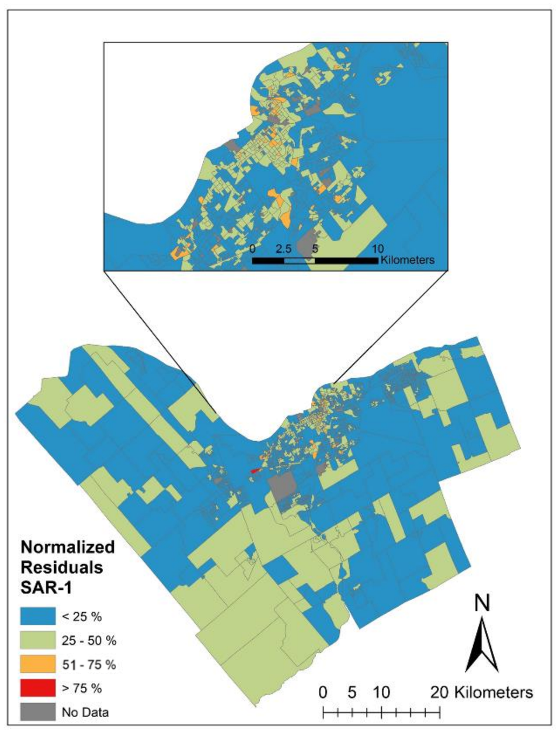


Figure 3 - 6 2016 Normalized Residuals of Ottawa

3.6 Conclusions

This chapter presents the findings obtained through specifying, estimating and validating a series of housing price models for the Canadian cities of Calgary, AB and Ottawa, ON. OLS regression and SAR models for the years 2006, 2011, and 2016 were developed using variables that directly impact zonal average housing prices. These variables were inspired by the literature and then finalized based on statistical significance. Region-specific parameters were also used to test if they impact the performance of the models. The parameters obtained were then compared to explore whether they change over time. The comparison was based on the Wald Chi-Square test. Further, the ability of a 2006 base year model to predict future housing values was examined by predicting the average housing prices for the years 2011 and 2016 and validating the results against observed values from the Canadian census data.

Judging by the achieved R-square values, the SAR models better fit the data when compared to their OLS counterparts. Additionally, region-specific parameters were found to significantly improve the performance of the OLS family of models, however, have less of an impact on the SAR models. The number of detached dwellings along with household income were found to positively impact zonal average prices in all models. Furthermore, the number of row houses and the distance from the CBD negatively influence prices.

The Wald Chi-Square tests suggested that the significantly different parameters were mostly present between the 2006 and 2011 models. We conclude that such an occurrence could be the result of the severe inflow of people into Canada in this time period. Finally, the validation of the 2011 and 2016 predictions using the 2006 models suggested that the trends are stable over time. That is, the 2006 models can produce stable predictions with an accuracy ranging between

65-75%. Also, the SAR models improved the prediction accuracy by approximately 1-4% when compared to the OLS ones.

Based on the findings presented in the conducted research, a few conclusions can be drawn with respect to price models within IUMs. First, model transferability over time within the real estate market is feasible given the relatively stable predictions. However, sudden and extreme changes in the market could hinder the predictive ability of the models. Second, house prices are spatial in nature and the modeling process needs to account for spatial autocorrelation (SA) in the data. Significant spatial lag parameters (ρ) in all the SAR models and the improved statistical outcomes in terms of goodness of fit measures (R^2) along with enhanced prediction accuracy values justify the need for using spatially oriented regression models.

As the present work was able to obtain valuable insights regarding price models within IUMs, it has opened the doors for further research work. Incorporating a SAR-based model within the SMARTPLANS IUM and conducting estimations and validations would be appealing. It would be interesting to examine how SAR models can be incorporated in the modeling framework of SMARTPLANS to examine the possible improvements in predictions when running simulations. Furthermore, testing the model transferability over space (i.e., from one study area to another) is another aspect worth investigating. Areas or regions that do not have the proper data to estimate the models require an alternate evaluation method. Model transferability can be beneficial in this case. Therefore, future research could focus on investigating the effectiveness of model transferability over space to examine how such transferability will impact the predictions of housing prices over the planning horizon when running simulations.

3.7 Chapter 3 References

Abraham, J.M. & Hendershott, P.H. (1996). *Bubbles in Metropolitan Housing Markets*. Working Paper Series. [Online]. National Bureau of Economic Research. Available from: http://www.nber.org/papers/w4774. [Accessed: 6 October 2020].

Bartholomew, K. & Ewing, R. (2011). Hedonic Price Effects of Pedestrian- and Transit-Oriented Development. *Journal of Planning Literature*. 26 (1). p.pp. 18–34.

Bidanset, P.E. & Lombard, J.R. (2014). Evaluating Spatial Model Accuracy in Mass Real Estate Appraisal: A Comparison of Geographically Weighted Regression and the Spatial Lag Model. *Cityscape*. 16 (3). p.pp. 169–182.

Bourassa, S., Cantoni, E. & Hoesli, M. (2010). Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods. p.p. 23.

Chen, M.-C., Tsai, I.-C. & Chang, C.-O. (2007). House prices and household income: Do they move apart? Evidence from Taiwan. *Habitat International*. 31 (2). p.pp. 243–256.

Cohen, J.P. & Coughlin, C.C. (2008). Spatial Hedonic Models of Airport Noise, Proximity, and Housing Prices*. *Journal of Regional Science*. 48 (5). p.pp. 859–878.

Dubin, R.A. (1992). Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics*. 22 (3). p.pp. 433–452.

Gallin, J. (2006). The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets. *Real Estate Economics*. 34 (3). p.pp. 417–438.

Gallo, M. (2018). *The Impact of Urban Transit Systems on Property Values: A Model and Some Evidences from the City of Naples*. [Online]. 4 May 2018. Available from: https://www.hindawi.com/journals/jat/2018/1767149/. [Accessed: 5 October 2020].

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

Gingerich, K., Maoh, H. & Anderson, W. (2013). Location and Transportation Effects on Nonresidential Real Estate Price Regressions in Windsor, Ontario, Canada. *Transportation Research Record*. 2397 (1). p.pp. 99–107.

Kelejian, H. & Prucha, I.R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*. 104 (2). p.pp. 219–257.

Haurin, D.R. & Brasington, D. (1996). School Quality and Real House Prices: Inter- and Intrametropolitan Effects. *Journal of Housing Economics*. 5 (4). p.pp. 351–368.

Helbich, M. (2015). Do Suburban Areas Impact House Prices?: *Environment and Planning B: Planning and Design*. [Online]. Available from: https://journals.sagepub.com/doi/10.1068/b120023p. [Accessed: 10 October 2020].

Hopkins, E.A. (2018). The influence of public transportation on housing values. *International Journal of Sustainable Development & World Ecology*. 25 (3). p.pp. 206–215.

Jansen, S.J.T. (2020). Urban, suburban or rural? Understanding preferences for the residential environment. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*. 13 (2). p.pp. 213–235.

Kiel, K.A. & Zabel, J.E. (2008). Location, location, location: The 3L Approach to house price determination. *Journal of Housing Economics*. 17 (2). p.pp. 175–190.

Kim, K. & Lahr, M.L. (2014). The impact of Hudson-Bergen Light Rail on residential property appreciation. *Papers in Regional Science*. 93 (S1). p.pp. S79–S97.

Linneman, P.D. & Megbolugbe, I.F. (1992). Housing Affordability: Myth or Reality? *Urban Studies*. 29 (3–4). p.pp. 369–392.

Livy, M.R. (2017). The effect of local amenities on house price appreciation amid market shocks: The case of school quality. *Journal of Housing Economics*. 36. p.pp. 62–72.

Maoh, H., Dimatulac, T., Oiamo, T. & Johnson, M. (2019). Implementation of the SMARTPLANS Integrated Urban Model for Halifax, Nova Scotia and London, Ontario. In: *Book of the Canadian Transportation Research Forum (CTRF) 54th Annual Conference*. 2019, Vancouver, British Columbia, Canada.

Maoh, H. & Gingerich, K. (2016). SMARTPLANS: SIMULATION MODEL FOR ASSESSING THE RAMIFICATIONS OF TRANSPORTATION POLICIES AND LAND USE SCENARIOS. In: *Book of the Canadian Transportation Research Forum (CTRF) 51st Annual Conference*. 2016, Toronto, Ontario, Canada.

Margulis, H.L. (2002). Suburban Housing Resale Prices and Housing Market Restructuring. *Journal of Urban Affairs*. 24 (4). p.pp. 461–477.

Martinez, F. (1996). MUSSA: Land Use Model for Santiago City. *Transportation Research Record*. 1552 (1). p.pp. 126–134.

Martínez, L.M. & Viegas, J.M. (2009). Effects of Transportation Accessibility on Residential Property Values: Hedonic Price Model in the Lisbon, Portugal, Metropolitan Area. *Transportation Research Record*. 2115 (1). p.pp. 127–137.

McMillen, D.P. & McDonald, J. (2004). Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983–1999. *Real Estate Economics*. 32 (3). p.pp. 463–486.

Meen, G. (2002). The Time-Series Behavior of House Prices: A Transatlantic Divide? *Journal of Housing Economics*. 11 (1). p.pp. 1–23.

Osland, L. (2010). *An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling*. p.p. 33.

Potoglou, D., Maoh, H., Wang, Y. & Orford, S. (2018). The impact of public transport infrastructure on land values: using spatial analysis to uncover policy-relevant processes. In: H. Briasoulis, N. Soulakellis, & D. Kavroudakis (eds.). *The practice of spatial analysis: Essays in memory of Professor Pavlos Kanaroglou.* [Online]. Springer, pp. 275–293. Available from: http://orca.cf.ac.uk/112204/. [Accessed: 5 October 2020].

Spinney, J., Maoh, H. & Kanaroglou, P. (2014). Location and land values: comparing the accuracy and fairness of mass appraisal models. *The Canadian journal of regional science.* 37(1/3) p.p 19-26

Statistics Canada (2001). *2001-2016 Census Profile*. [Online]. 2006;2011;2016 2001. Available from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMACA&Code1=205&Geo2=PR&Code2=12&SearchText=Halifax&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=205&TABID=1&type=0. [Accessed: 23 July 2020].

Statistics Canada (2016). *2016 Census Profile*. [Online]. 2016. Available from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMACA&Code1=205&Geo2=PR&Code2=12&SearchText=Halifax&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=205&TABID=1&type=0. [Accessed: 23 July 2020].

Voith, R. (1999). The Suburban Housing Market: Effects of City and Suburban Employment Growth. *Real Estate Economics*. 27 (4). p.pp. 621–648.

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

Wilhelmsson, M. (2002). Spatial Models in Real Estate Economics. *Housing, Theory and Society*. 19 (2). p.pp. 92–101.

CHAPTER 4

CONCLUSIONS

4.1 Overview

Integrated Urban Models (IUMs) are effective tools for planning the future of cities. They can be considered as virtual laboratories used to simulate complex processes that shape the structure of urban form and the type of travel activities observed on a daily basis (Maoh et al., 2019). As of 2013, there are approximately 200 state of the practice models that have been developed in the last 40 years; with around 40 models still being used in practice (Miller, 2018). In general terms, there is a consensus in the literature that these models are extremely data hungry, complicated to program and in many cases require high computing power. Also, the developers of these integrated models have very high chances to encounter technical challenges due to the feedback relationships between the different modules and sub-modules forming an IUM. Fortunately, ongoing technological advances in both software and hardware have made it more possible to develop sophisticated IUMs in recent years. One such model is the SMARTPLANS IUM.

The SMARTPLANS IUM has been developed as a full-fledge stand alone model with programable parameters. One of the key objectives of SMARTPLANS is to become a model that can essentially be applied to any urban area (Maoh et al., 2019). Among its six major modules (namely: regionwide aggregate controls, land use, transportation, spatial disaggregation, health benefits, and sustainability indicators), the land use module is of great importance to many decision makers and city planners. In this thesis, two submodules within the land use module are extensively explored to provide enhancements and ultimately improve the overall performance of the land use module. The population mobility submodule (Chapter 2) and the land price submodule (Chapter

3) were specified, estimated and validated to explore the potential improvements that could improve their predictive ability.

4.2 Population Mobility Submodule

The population mobility model to be incorporated within the SMARTPLANS IUM is calibrated and validated for four Canadian CMAs: Halifax, Calgary, London, Ottawa. Data from the Canadian Census was utilized to develop the model. Ordinary Least Square (OLS) logistic regression and Spatial Auto-Regressive (SAR) models were estimated for the periods 2001-2006 and 2011-2016, respectively. The 2006 parameters were compared to the 2016 ones to examine if they change over time, and the 2006 parameters were then used to predict 2016 population stayers which were then validated with observed census data. The results showed that the variable representing young people in the age group 25-34 years significantly impacted relocation decisions compared to other assessed attributes. The Calgary CMA behaved differently than the other studied regions. The results also showed that most of the parameters did not change between 2006 and 2016, and the predicted 2016 population stayers are in-line with census data.

In this chapter, it was found that the mobility process within the Canadian context is a stable one. This is justified by the results obtained using the Wald Chi-Square test where the estimated parameters rarely vary overtime in all four CMAs. Additionally, the modeled mobility data exhibited spatial autocorrelation based on the obtained results from the Moran's *I* statistic tests. This led us to test Simultaneous Auto-Regressive (SAR) models on the data. The results confirmed the presence of spatial autocorrelation based on the obtained significant spatial lag parameters (ρ) in all CMAs. Further, we found that the SAR models were far more superior at predicting population movement when compared to the OLS ones. The improvement in the obtained R^2 and correlations between the observed and predicted values support this conclusion.

Finally, the results attained from the Calgary CMA are not as well-behaved as the other three CMAs, however validation figures suggest adequate results.

4.3 Land Price Submodule

Modeling the price of residential real-estate is essential when developing comprehensive and robust IUMs. In this part of the research, we expanded the specification of the price model of SMARTPLANS IUM and evaluated its performance by applying both OLS and SAR modeling techniques. More specifically, we calibrated and validated a series of OLS and SAR price models for the Canadian cities of Calgary, AB and Ottawa, ON. We also incorporate region-specific parameters in the model of both cities to test how they will influence the results. In terms of model specification, we found that the number of detached dwellings and household income positively impact house prices whereas the number of row houses and distance from the central business district (CBD) reduce prices. The region-specific parameters were shown to be useful when applied to the OLS models, however this was not the case in the SAR models. Moreover, the SAR modeling technique was found to better fit the data in all modeling years. The validation of the predicted results also suggested that the SAR models outperforms the conventional OLS technique usually used in land price models. Based on the validation results, the SAR models improved the prediction accuracy by approximately 1% to 4% in 7 out of 8 estimated models.

In the research conducted in this part of the thesis, it was comprehended that the model transferability over time within the real estate market is relatively stable. Such a conclusion was deduced through the results obtained from the Wald Chi-Square test where stable variation in real estate influencing factors allows for consistent and predictable future assessments. Next, the house prices within the studied areas are spatial in nature and reveal a high degree of spatial autocorrelation. Region-based parameters within OLS based models as well as spatial oriented models (like the SAR) significantly improve the model performance. Improved R^2 figures combined with enhanced accuracy values is visible throughout the SAR models when compared to their OLS counterparts.

4.4 Contribution and Policy Implication

The analysis presented in this thesis offers influential efforts to address the current gap within the SMARTPLANS IUM in the transportation discipline. To the best of the author's knowledge, testing and validating the population mobility and land price sub-modules to offer possible spatial-related improvements has not been explored in the literature. The contributions of this thesis are as follows: (1) it explores the attributes that have a direct impact on population mobility and house prices within the Canadian context over time; (2) it examines and confirms the strong spatial influence present in the data for both sub-modules; and (3) it offers improvements in performance and accuracy by using the SAR modeling approach as a suitable alternative to the conventional OLS approach usually utilized in these models.

Integrated land use and transportation strategies have become a standard mechanism of planning the future of cities in most Canadian Transportation Master Plans. This is evident in the publicly available TMP documents published by the transportation planning departments of the analyzed cities (see for example: City of London (2013); City of Ottawa, (2019); City of Calgary (2020); City of Halifax, (2020)). While integrated strategies that could lead to sustainable futures are formulated in various TMPs, their long-term impacts are usually not examined due to the lack of operational IUMs. Policy oriented decision-makers can lean on the outcomes of these IUMs to make informed decisions regarding their integrated land use strategies. However, the accuracy of the predictions from such IUMs is of great importance.

As the findings from this research suggests, the estimated models provide more reliable results and thus potentially improving the predictive ability of SMARTPLANS. In doing that, reliable policies about population mobility and land prices can be tested. The conducted analysis found that Calgary was behaving differently than the other regions in terms of both population mobility and land prices. It is believed that these differences are related to the oil industry and the economic boom that Calgary experienced during the first 10 years of the new millennium. This indicates that the parameters estimated for the base year should not be used blindly to predict future outcomes in regions experiencing sharp growth or decline. In comparison, base year models estimated for Canadian cities with steady pace growth tended to do a good job in predicting future values.

4.5 Limitations and Recommendations

The SMARTPLANS IUM is a large-scale model with numerous modules and sub-modules. As such, the conducted research is limited only to the improvements offered by the two sub-modules tested and explored. To fully understand the potential capability of spatial analysis in IUMs, more sub-modules must be further investigated. The data used in this analysis was primarily composed of records obtained from Statistics Canada. While the data was ample and sufficient, it was not fully comprehensive of the study area as some records were suppressed. These records were removed from the analysis to avoid potentially disturbing the model's predictive ability, but a complete dataset would have been ideal.

Future research on this topic would be to first implement both SAR models into their respective sub-modules within SMARTPLANS software to test the improvements on a larger scale. Performing comparisons and validations before and after implementing the SAR model would be of interest. Additionally, testing the model transferability over space (i.e., from one study area to another) is another aspect worth investigating. Many regions around the world do not have sufficient data to perform the data hungry analysis needed by IUMs. Here, model transferability between two regions would be advantageous. Therefore, it would be interesting to explore the usefulness of transferability over space as it associates with population mobility and price values.

4.6 Chapter 4 References

City of Calgary (2020). *Calgary Transportation Plan (CTP)*. [Online]. 2020. Available from: /transportation/tp/planning/calgary-transportation-plan/calgary-transportation-plan-ctp.html. [Accessed: 6 February 2021].

City of Halifax (2020). *Integrated Mobility Plan*. [Online]. 2020. Available from: https://www.halifax.ca/about-halifax/regional-community-planning/transportation-planning. [Accessed: 6 February 2021].

City of London (2013). *Transportation Master Plan*. [Online]. 2013. Available from: https://london.ca/government/council-civic-administration/master-plans-strategies/plans-strategies. [Accessed: 6 February 2021].

City of Ottawa (2019). *Transportation Master Plan*. [Online]. 2019. Available from: https://ottawa.ca/en/planning-development-and-construction/official-plan-and-master-plans/transportation-master-plan. [Accessed: 6 February 2021].

Maoh, H., Dimatulac, T., Oiamo, T. & Johnson, M. (2019). Implementation of the SMARTPLANS Integrated Urban Model for Halifax, Nova Scotia and London, Ontario. In: *Book of the Canadian Transportation Research Forum (CTRF) 54th Annual Conference*. 2019, Vancouver, BC.

Miller, E.J. (2018). Viewpoint: Integrated urban modeling: Past, present, and future. *Journal of Transport and Land Use*. [Online]. 11 (1). Available from: https://www.jtlu.org/index.php/jtlu/article/view/1273. [Accessed: 11 March 2020]

VITA AUCTORIS

NAME:	Mohamed Abdo
PLACE OF BIRTH:	Sana’a, Yemen
YEAR OF BIRTH:	1995
EDUCATION:	University of Windsor, BASc., Windsor, ON, 2018
	University of Windsor, MASc., Windsor, ON, 2021