Estimating Homebuyer Preferences Under Intensification:

Hedonic Modelling of Open Space and Multimodal Transit Amenities Preceding Light Rail in Kitchener-Waterloo

by

Robert Malcolm Babin

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Planning

Waterloo, Ontario, Canada

AUTHOR'S DECLARATION

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

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

X

Robert Malcolm Babin

ABSTRACT

Intensification is the preeminent growth management approach in Ontario, as well as across much of North America. Under this approach, the Region of Waterloo is currently constructing a light rail transit system and undertaking co-ordinated planning to support population and employment densification alongside other regional goals. The work presented in this thesis is centred on the development of spatially explicit hedonic models to estimate residential preferences for amenities associated with intensification. The results of this work are to be used as willingness-to-pay parameters in an agent-based land use-transport model.

The hedonic models presented in this thesis use 26,873 Kitchener-Waterloo residential property sales from 2005 to 2015 to estimate the joint effects of willingness-to-pay and willingness-to-accept for housing. Combined spatial lag and spatial error models are employed to test the effects of environmental home characteristics on assessed and appreciation-adjusted transaction values, for single-detached homes, semi-detached and duplex homes, and townhouses.

These models are specified to estimate effects related to the changing built form and housing market under intensification; specifically, this work estimates the price effects associated with access to public and semi-public open space, access to public transit (local bus stops), and walkability throughout the period preceding light rail. A large number of socioeconomic control variables were developed and analyzed to determine an appropriate model specification. Heterogeneous willingness-to-pay estimates are presented for intensification amenities within and outside of the Central Transit Corridor throughout three time periods. Models using assessed values and transaction values provided slightly different but comparable results. Results suggest a positive willingness-to pay effect of walkability for single-detached homes and semi-detached and duplex homes. However, walkability was found to only significantly relate to townhouse value through a synergistic interaction with open space and a negative interactive effect with transit. The regional Central Transit Corridor generally saw greater increases in property value due to walkability and in general throughout the implementation process of regional light rail, compared to homes outside of it.

The results of this thesis may be used to help inform policies and investments related to housing, public amenity distribution, and multimodal transit planning. While the precise estimates produced in this work are context-specific, broadly, these results can provide guidance in other municipalities and land-markets similarly undergoing intensification.

ACKNOWLEDGEMENTS

This research would not have been possible without the tremendous amount of support I have received over the past few years. I would like to acknowledge the varied contributions of the following individuals and offer them my sincere gratitude:

- Dr. Dawn Parker, for your mentorship, motivation, and bad jokes throughout the process. I cannot imagine a better person to have supported me in this work
- My research team members, for providing an outlet to brainstorm and collaborate
- The Social Sciences and Humanities Research Council, for providing the funding that made this research possible
    - SSHRC Partnership Development Grant (SSHRC # 890-2013-0034) entitled "*LIGHT RAIL TRANSIT AND CORE-AREA INTENSIFICATION: Unpacking Causal Relationships*
    - SSHRC Insight Grant (SSHRC # 435-2012-1697) entitled "*Urban intensification vs. suburban flight: An integrated residential land-use and transportation model to evaluate residential land-market form and function*
- The Region of Waterloo, and Margaret Parkin and David Stubbs in particular, for the incredible assistance with the data needs of this research, and for some great conversations too
- Scott MacFarlane & Dr. Xiongbing Jin, for assisting with the Python code to calculate the access metrics
- Attendees of the American Association of Geographers conference and Region of Waterloo Planning staff who attended presentations of my preliminary research results, for giving me the opportunity to share my work and for your valuable feedback that helped to strengthen it
- Dr. Jeff Casello, for letting me hang with the WPTI crew, pretending to be a transportation engineer around the world, and participating in my defence
- Dr. Christopher Higgins, for serving as the external reader for my defence, and some great insights on the research
- Robyn Roopchan, Becky Loi, and my sister, Rachelle Babin, for help with proofreading and formatting
- My friends, for consistently keeping my sanity in check
- My parents, for consistently chipping away at my sanity, as well as for your unfettered support over many years

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CT – Census Tract
CTC – Central Transit Corridor
DA – DA
GGGH – Growth Plan for the Greater Golden Horseshoe
KW – Kitchener-Waterloo
LRT – Light Rail Transit
MPAC – Municipal Property Assessment Company
NHS – National Household Survey
ROW – Region of Waterloo
TAZ – Traffic Analysis Zone
TTS – Transportation Tomorrow Survey
ROA – Rate of Appreciation
SEM – Spatial Error Model
SLM – Spatial Lag Model
OLS – Ordinary Least Squares
OS – Open Space
WARM – WAterloo Regional Model
WTA – Willingness-to-accept
WTP – Willingness-to-pay

LIST OF EQUATIONS

LIST OF MAPS

1 Introduction

Urban intensification, or the densification of the existing built area of cities, is a predominate force shaping regions across North America. Negative social and environmental repercussions of sprawl-type development, which include an automobile dependent built form and the degradation and loss of natural and agricultural lands, have motivated this regime of concentrated, recentralized development (Searle & Filion, 2011). The relationships between this intensification regime and the regional land-market are vast and complex. Naturally, these complex relationships present a number of avenues for academic inquiry. A broad and pertinent area of inquiry stemming from these complex relationships is in the study of residential location decisions and the price setting mechanisms of homebuyers in intensifying environments.

Conflicting values and preferences amongst homebuyers, planning officials, and the development community have produced divergent opinions on intensification's ability to fulfill housing market demand. A purported inability of intensification to fulfill housing demand was a central contestation from the development community of the Region of Waterloo (ROW)'s intensification-focused Official Plan (2012), where some members of the development community objected to the restriction of developable greenfield lands. This dispute reached a settlement of a modest increase in the amount of land open for development at the Ontario Municipal Board six years after the plan's regional approval (Desmond, 2015). A historical preference for the large homes, private yards, and ample parking that typify sprawl-type development stands oppositionally to the environment offered through intensification (DeFields, 2013). However, contemporary research in support of intensification suggests that preferences for an amenity-rich, multimodal urban environment exist, that they are growing, and that they can be leveraged to attract a knowledge-based labour-force (Florida, 2008). A topical question then arises: are homebuyers paying more to live in these intensified environments?

To answer this question, this work utilizes hedonic modelling, an econometric, revealed preference approach. Hedonic models, as will be discussed in the following chapter, are often employed to estimate homebuyer willingness-to-pay (WTP) for the environmental characteristics of homes. WTP is an economic term meant to capture the amount an individual or firm would pay to receive a certain good or service, which is constrained ultimately by one's ability to pay. Realistically, hedonic models capture the point of intersection between homebuyer WTP and home-seller willingness-to-accept (WTA), which is the amount a seller would accept for the transfer of their good or service to the buyer. However, the use of hedonic models in the literature as a means to identify WTP is widespread. Therefore, WTP will be generally be used to describe the hedonic model estimates in this work as well.

This work investigates the complex relationship between home values and three amenities commonly associated with the effective implementation of intensification. Specifically, for reasons that will be discussed throughout this thesis, this research estimates homebuyer WTP for access to public and semi-public Open Space (OS) (which will herein be referred to simply as public OS), proximity to local public transit, and walkability. As will be described in the following literature review, the contemporary land-market is spatially and temporally complex, and, as a result, estimating homebuyer preferences for these intensification amenities necessitates a thorough understanding of the concurrent spatial processes driving property price and the interactions between these processes.

The following section will outline the intensification context in Kitchener-Waterloo (KW), providing the case study that will be examined throughout this thesis.

1.1 Location and Policy Context

1.1.1 Provincial Intensification Context

The release of the Places to Grow Act in 2005 set the stage to move forward with a provincial growth management mandate in the province of Ontario (Ministry of Municipal Affairs, 2005). The Growth Plan for The Greater Golden Horseshoe (GGGH), first released in 2006 under the Places to Grow Act, is a policy document that outlines specifically where and how growth should occur in South Ontario – relying heavily on intensification as the means to accommodate that growth (Ministry of Municipal Affairs, 2014). The population and employment forecasts in the GGGH, which were updated in 2014, suggest a large amount of growth in the ROW in the forthcoming years; this plan estimates that from 2011 to 2041 the ROW will see an increase of 58% in population and a 50% increase in employment. The GGGH sets specific density targets for designated growth areas, which includes a target of 200 persons and/or jobs per hectare for Uptown Waterloo and Downtown Kitchener, or 1 person or job for every 50 square kilometres.

Ontario's Provincial Policy Statement, one of the primary pieces of provincial legislation guiding policy decisions, defines intensification as:

> "the development of a property, site or area at a higher density than currently exists through:
>
> a) redevelopment, including the reuse of brownfield sites;
>
> b) the development of vacant and/or underutilized lots within previously developed areas;
>
> c) infill development; and
>
> d) the expansion or conversion of existing buildings." (Ministry of Municipal Affairs and Housing, 2014, p.43)

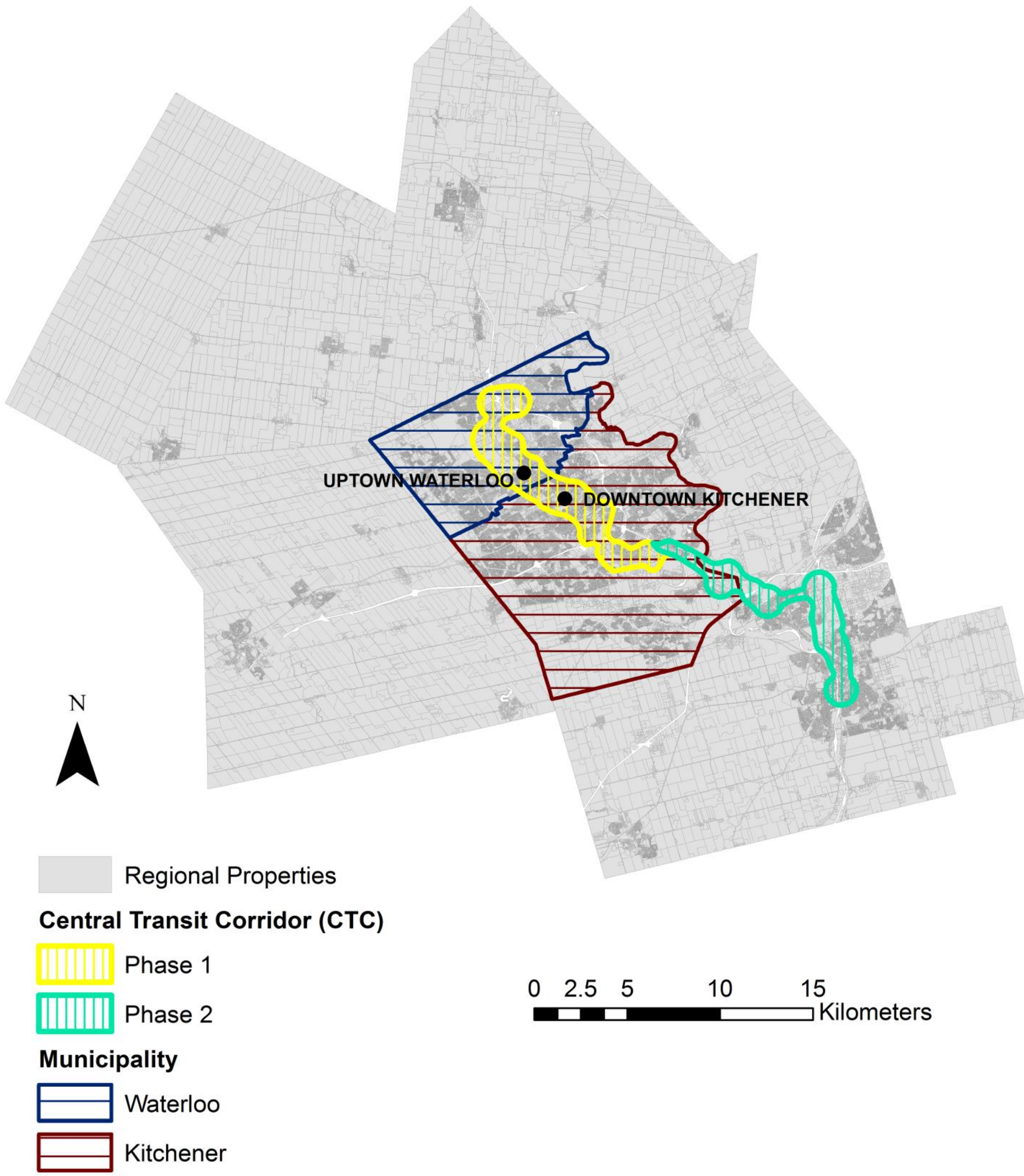
While this definition of intensification includes only consideration of building density, the policy documents that advocate for it stress the importance of a holistic approach to intensification. The GGGH supports growth through intensification while simultaneously advocating for the development of complete communities, where residents' needs are able to be met within the community, and where recreational amenities and opportunities for multimodal transit are provided (Ministry of Municipal Affairs, 2005, 2014).

1.1.2 Intensification in KW

The ROW is composed of seven separate municipalities, which include the cities of Kitchener, Waterloo, and Cambridge. Under the provincial mandate as stated above, the ROW is undergoing co-ordinated intensification efforts. Part of the regional intensification strategy is a light rail transit (LRT) system. The LRT implementation approach includes a strong focus on the development of communities alongside the promotion of intensification. As part of the LRT implementation process, the ROW has delineated a Central Transit Corridor (CTC). This CTC represents a roughly 800m buffer around stations where intensification and development is being promoted, modified to remove most stable residential neighbourhoods. The CTC and location of the study areas within the ROW is shown in Map 1. The CTC runs along the regional north-south spine, this is the boundary used in the ROW's monitoring of the CTC.

In this map, Waterloo is the northern municipality and Kitchener is the larger one in the south. Phase one of LRT is currently under construction and will run between Kitchener and Waterloo, while phase two will extend the line into Cambridge to the south, as shown in the map. Within the CTC, specific planning and monitoring is taking place to promote and ensure intensification is occurring in a way that is consistent with other regional goals. A Community Building Strategy is one of the regional policy documents meant to assist community development throughout the CTC (ROW, 2013a). Station areas were delineated by the ROW, which are areas surrounding future transit stations and are expected to be most impacted by the LRT. These station areas are undergoing specific planning efforts by the municipalities in which they are located, where the planning of these station areas has been the product of intensive community consultation in order to achieve intensification goals smoothly (ROW, 2013a). While the regional LRT is planned to extend to Cambridge, this is only after the first phase has been developed between Kitchener and Waterloo. For this reason, the modelling work conducted in this thesis does not investigate impacts in Cambridge.

Regional Context



Map 1: Context Map (Data from ROW)

1.2 Research Motivations

1.2.1 Applied

The primary motivation for this research is an applied need to estimate homebuyer WTP for an integrated land-use transport model of the ROW. This integrated model is called the WAterloo Regional Model (WARM), and it aims to simulate the processes of intensification in Waterloo Region in order to increase understanding of causal dynamics in the land-market. Specifically, WARM is an agent-based model (ABM), which operates by simulating the individual decisions of agents or actors and the interactions and feedbacks between these agents. ABMs are useful in understanding how large-scale patterns can be the result of very small-scale decisions (Buchmann, Grossman, & Schwarz, 2016). These models can be used for scenario testing, where they may be used to simulate the potential impacts of changes in regional policies or investments.

The model components currently in development consist of joined land-market and transportation models, which are shown in Figure 1. The land-market model consists of a residential location choice and land development model. The work presented in this thesis is a component of the residential location choice model.

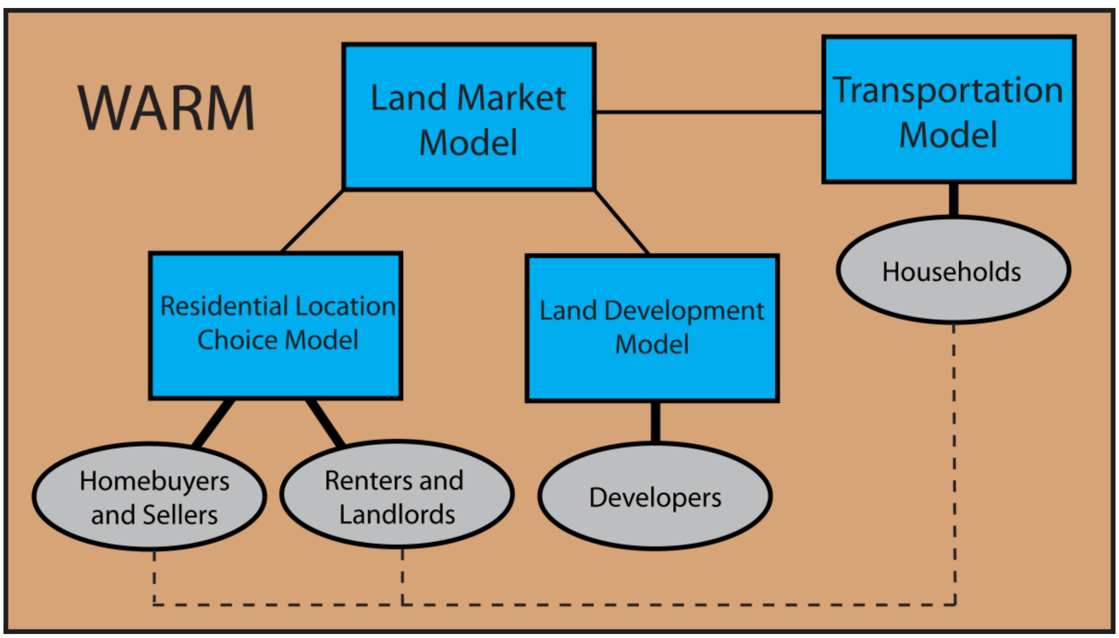


Figure 1: WAterloo Regional Model (figure created by author)

In the WARM model, households and developers will be the primary agents acting within the system. In the land-market model, households will be buying, selling, and renting homes, while the developer agents purchase and develop land into new residential building stock that may be inhabited by the households. The transportation model consists of an activity-based mode choice model, where households will allocate a daily set of household-specific travel resources to meet their respective travel demands (Yeung, 2015). When households in the transportation model are consistently unable to meet their travel demands, they may change their behaviour – where one potential option is to relocate to another residence that is better suited to their travel needs. This is when a household may decide to relocate, and the amount they bid on new homes in the land-market model will be based on the results from the research presented in this thesis.

1.2.2 Academic

With dissenting opinions on the potential land-market implications of intensification, there is a distinct need to explore the spatially complex relationship between this intensification and property values. The decision on how, where, and why to intensify is complicated, and beyond the scope of this work. However, this work can provide an important piece of the puzzle to support municipal and development decisions. As cities continue to work towards intensification while also attempting to maintain housing affordability and achieve transportation objectives, this work can be used as part of a holistic approach to decision-making.

Intensification is inherently connected to a host of land use and transportation changes. Particularly, this research is interested in investigating public OS access, access to transit, and walkability in relation to the land-market and homebuyer WTP. There is a need to develop a well-performing cross-sectional model, from which future causal, time-series models may be specified. This work supports these future causal models as it provides a basis from which to singularly identify the amount of land-value change caused by the LRT itself and by each of the built form characteristics that are evolving alongside the LRT's implementation.

Monitoring the impacts of LRT on the Region is also of relevance to regional decision-makers. The ROW is currently going through a process of monitoring the CTC throughout the development of the Regional LRT System (ROW, 2015c). The work presented in this thesis bears relevance on the ongoing monitoring of the regional CTC. Where the ROW has released a Baseline Monitoring Metrics report from which changes in the character of the CTC will be tracked over time, this work can be seen as a parallel; this thesis provides a baseline cross-sectional model from which future analysis can be used to understand land-market change throughout the corridor.

1.3 Research Goals and Objectives

The primary goal of this research is to understand how home prices in KW relate to various elements of intensification. Looking at this goal through a complex systems lens, it becomes apparent that many forces drive home values in non-constant ways. From this, a major sub-goal of this research is to understand these forces and include them in the analysis in order to make ceteris paribus (all else being held fixed) conclusions on intensification's relationship with home values.

Research Goals

- Characterize the housing market and spatial structure of home characteristics in the region

- Develop a well-performing hedonic model of KW, in terms of goodness of fit and the mitigation of model endogeneity, to understand homebuyer preferences for selected features of intensification
- Understand how preferences for intensification-related amenities may be heterogeneous across space and for different individuals, and incorporate this into the model

Research Objectives

- To estimate how homebuyer WTP is impacted by characteristics of intensified urban environments
- To understand the distributions of land values and their determinants across KW
- To identify and control for important spatial patterns in KW that relate to property value and intensification
- To quantify how homebuyer preferences for intensification amenities vary between the central area and the rest of the city, as well as between homebuyers of different property types
- To identify potential synergistic impacts of intensification amenities on property value
- To determine whether home value impacts of public OS are moderated by private OS access
- To understand how model results change between using observed transaction data and assessed home values as well as between ordinary non-spatial model and models that explicitly account for spatial processes

1.4 Research Questions

1.4.1 Primary:

Are homebuyers willing to pay for OS access, walkability, and proximity to public transit in KW, and if so, how much?

Hypothesis: Homebuyers express a positive WTP for these intensification amenities, with walkability providing the largest price effect.

Method: Hedonic Regression

1.4.2 Secondary:

1. How are the characteristics commonly associated with home price distributed across KW? Are intensification-related features correlated with each other, neighbourhood socioeconomic characteristics, or structural home characteristics in regional home sales? If so, is there a spatial pattern that explains this?

Hypothesis: Home characteristics are distributed in a spatially complex manner, which is conditioned largely by historical patterns of development.

Method: Visual analysis of maps

2. How does the characterization of OS access via a gravity-based model compare to those of spatial separation and cumulative opportunities models?

Hypothesis: Gravity-based models provide a more holistic measure of OS access that favours neither outlying areas with access to large OS nor central areas with access to many small OSs.

Method: 1. Calculating different access measures, 2. Visual comparison of maps, 3. Comparison of descriptive statistics and pairwise correlations

3. How do results of hedonic models using assessed or observed transaction values differ?

Hypothesis: Assessed value models will behave more predictably using common home price determinants than transaction data, while transaction value models will be able to find values related to a more spatially, temporally, and behaviourally complex set of home price determinants.

Method: Hedonic regression, error mapping

4. Are homebuyers willing to trade-off public for private OS?

Hypothesis: Private OS and public OS have a negative interaction effect, representing a trade-off in homebuyer decision making.

Methods: Interaction term in regression

5. Do transit access, public OS access, and walkability have synergistic impacts on homebuyer WTP?

Hypothesis: Homebuyers are willing to pay more for housing when multiple intensification amenities are present.

Method: Interaction term in regression

6. Are homebuyers' stated preferences consistent with their location choice decision?

Hypothesis: Homebuyers who state that OS and transit amenities were important in their decision to move have greater access to those amenities.

Method: Comparison of descriptive statistics

7. Have property prices increased throughout the planning of the regional LRT, and, if so, is this effect greater in the CTC?

Hypothesis: Homes in the CTC receive a premium that has increased at various points marking the approval of plans for LRT and CTC development.

Method: In CTC variable and related interaction terms

8. Do buyers of different types of homes express a different WTP for intensification-related variables?

Hypothesis: Homebuyers of single-detached homes will express a smaller, WTP for walkability and transit access, and OS access than buyers of semi-detached homes, duplexes, and townhouses.

Method: Comparing results from separate regressions by property type

9. How do estimates differ after accounting for spatial effects, is model fit improved?

Hypothesis: By controlling for spatial effects, the model will provide a better goodness of fit, and spatially autocorrelated variables and home prices will be able to be estimated unbiasedly.

Method: comparison of ordinary least squares (OLS) estimates and levels of significance, Pseudo r-squared, Log Likelihood

1.5 Thesis outline

Following this introduction, this thesis will contain six more chapters.

First, Chapter 2 presents a literature review, which will supply both the theoretical foundation from which this research was constructed as well as the applied framework, where specific concerns in hedonic modelling are outlined.

Next, Chapter 3 will explain the preliminary work done to collect, clean, manipulate, and compare the various candidate hedonic model variables for this work. This is followed by the results of this work, in Chapter 4, where descriptive statistics and maps explain the distributions of the variables across the region and property types.

Then, Chapter 5 will provide the theoretical model of land value guiding the selection of model variables. The results of the previous chapter will then be used to support the empirical regression model specification of this work. The results from these regressions are found in the following Chapter 6

Finally, Chapter 7 will summarize findings related to the research questions of this thesis and in general, and outline policy implications, limitations, and next steps of this research.

2 Review of the Literature

The first section of this literature review will outline the theoretical understanding of the contemporary residential land-market necessary to support this research; specifically it will focus on the evolution of the AMM model, intensification, and heterogeneity in homebuyer preferences and behaviour. Following that, the second half of this review will discuss hedonic modelling as is applied to estimate the relationship between environmental or locational amenities and home values. This second half will outline the hedonic modelling methodology commonly employed ini similar studies, which is used in this thesis to estimate homebuyer preferences related to regional intensification and LRT corridor planning.

2.1 Theoretical Framework: The land-market, intensification, and homebuyer preferences

The coevolution of urban form and the land-market is a rich and well-studied topic. Section 2.1.1 will explain the Alonso-Muth-Mills (AMM) model, a model from which many studies similar to this thesis are built upon. Further, Section 2.1.2 will theorize a revised AMM framework that considers a more contemporary and comprehensive set of land-market drivers. Section 2.1.3 will then discuss homebuyer preferences in the context of this revised theoretical model, and Section 2.1.4 will explain how some have simulated these preferences in ABMs. Finally, Section 2.1.5 will discuss the relative merits and disadvantages of revealed and stated preference methods to estimate these residential preferences in the land-market.

2.1.1 The evolution of theoretical models of urban structure relevant to the residential land-market

The AMM model of the monocentric city has been the preeminent theoretical framework underpinning urban microeconomic thought since the late 1960s. This model conceptualizes the regional economic structure as a monocentric city with a dense core of employment surrounded by housing and then agricultural lands (Alonso, 1964; Mills, 1972; Muth, 1969). The AMM model describes property value as a product of distance to the urban core, where locations closer to the centre are more valuable due to lower commuting or transport costs (Dziauddin, Powe, & Alvanides, 2014). This rather simple yet useful conceptualization of urban form has provided a basis for the development of modern theories that recognize and incorporate additional land-market drivers.

The identification of polycentricity within the urban spatial structure marked a significant improvement to the AMM model. Anas, Arnott, & Small characterize polycentricity as the presence of not simply a single employment centre, but differentiated functional nodes of activity within cities and regions (1998). This polycentric model has been applied in broad range of studies of the land-market and transportation, such as a 2007 study where this model was applied to analyze transit competitiveness in the Philadelphia, Pennsylvania region (Casello, 2007). Adding further complexity, more recently urban form has been described as the product of complex interactions between general processes and local conditions, including: costs and amenities, industrial mix, technological change, geography, and preferences (Keil & Young 2009; Shearmur, Coffey, Dubé, & Barbonne, 2007). This understanding of cities as polycentric entities characterized by complex interactions is well aligned with contemporary theories on spatial heterogeneity. Spatial heterogeneity arises as the result of the interaction between simultaneous, interacting spatial processes, leading to disparate outcomes over space that deviate from a strictly monocentric model of urban structure (Páez & Scott, 2005). Where the AMM framework would posit that land values decline in a monotonically linear fashion from the centre, spatially heterogeneous processes have complicated that relationship. This conceptualization of the contemporary land-market as a complex, spatially heterogeneous entity will help to inform the variable analysis and model specifications in this thesis.

In the context of the land-market and polycentricity, transportation plays a considerable role in producing spatially heterogeneous outcomes. Transportation infrastructure enables one to cross space in a shorter time, effectively replicating the effect of nearness or removing the friction of distance (Hooper, 2014) Continued investment in the North-American highway system in the late 20th century allowed cities to expand while still maintaining connectivity between residential development at the fringe and core employment areas (Mieszkowski & Mills, 1993). This highway investment was coupled with another opposing force to the central magnetism at the core of the AMM, which was a desire of the population to live away from the density and non-residential land-uses within core areas (Lambert, Clark, Wilcox, & Cho, 2011; DeFields, 2013). A residential preference for privacy and private space combined with the accessibility afforded by highway development allowed property value gradients to stretch across regions. This preference is best illustrated through moderate to high-value residential development with large amounts of private OS at the urban-rural fringe, or suburbanization more generally, which was common in the latter half of the 20th century (DeFields, 2013).

2.1.2 Intensification – Causes, Effects, and Links to Residential Preferences

Recently, a trend of renewed interest in the urban core as a residential environment has emerged (Riddell, 2004; Searle & Filion, 2011). Intensification is a prevalent force in cities and urban regions across Ontario, as well as across much of North America. Intensification presents a centripetal force in regional land-markets, pushing residential development back towards the existing built areas. This intensification is often supported through multimodal accessibility improvements, rather than improvements to the accessibility of the private car, which drives suburban development. In the context of the AMM model, municipalities are once again leveraging the access to jobs and amenities within the core to attract residents and development to the centre.

One primary rationale for the pursuit of intensification is Newman and Kenworthy's work linking density, and inherently urban form, to greenhouse gas emissions (1999, 2006). While this has provided a rationale to promote intensification, there is also perceived pressure for intensification on the side of homebuyer demand. Intensified core environments provide bountiful opportunities for socialization and consumption, which have been noted to appeal to certain homebuyers (Smith, 2007). In light of the work of Richard Florida, who sensationalized the notion of a "creative class", intensification can be seen as tool to provide the urban amenities and atmosphere he suggests will attract a labour force supportive of a knowledge-based economy (Florida, 2002; Florida, 2008). LRT is one transportation investment being employed by North American regions to support intensification efforts, attempting to encourage development and non-automobile transportation within central areas (Wolinsky, 1999).

Urban intensification can be explained through the AMM framework, polycentricity, and neighbourhood heterogeneity. According to the AMM model, levels of access drive property values in core areas. However, the rise of multimodalism and the heterogeneous levels of non-automobile accessibility within neighbourhoods have led to variations in the price gradient outward from the city centre. To understand the relevant drivers of the residential land-market under intensification, it is important to consider the potential motivations of homebuyers that drive the prices they pay for homes.

Before moving onto a discussion on homebuyer preferences, it is necessary to note that authors have cautioned that intensification, or renewed development of core areas generally, can play a large role in gentrification and the pricing-out of low-income populations (Smith, 2007; Zukin 1989). This is an important point to keep in mind for anyone studying the land-market and especially salient for those who aim to plan for inclusive intensified environments.

2.1.3 Homebuyer preferences

While the first sections have outlined the overarching theories of urban form and the land-market, this section aims to highlight how homebuyer preferences drive prices in the housing market. Historically, Tiebout proposed a model where the distribution of public services across locations had a direct effect on property values, as homebuyers were willing to pay more for housing with those services (1956). This model was extended by Oates to explain positive impacts on property values from increasing school expenditure to student ratios (Chiodo, Hernández-Murillo, & Owyangm, 2010; Oates, 1969). This Tiebout Model elucidates an understanding in the literature that homebuyers express a WTP for public investments and neighbourhood qualities. This understanding can be reasonably extended in the context of intensification to environmental and multimodal amenities, where these amenities are capitalized into nearby residential property values. While the Tiebout and AMM models may be viewed as competing models to describe the land-market, this thesis amalgamates the two – where the central price driver of the AMM model, access, may be seen as one public good within the Tiebout model.

While the impacts of intensification on property values are generally positive in the literature, as will be explained in the following section, this positive effect is neither definitive nor consistent across all cases. One confounding factor in the determination of property values that is often omitted in the literature is the heterogeneity in the preferences of those buying properties. Redfearn outlines that homebuyer preferences are not constant across space or time (2009). The idea that households exhibit heterogeneous preferences for amenities is a contemporary direction of academic inquiry. Households seek to maximize their individual utility in location choice decisions, by choosing where to live based on how well a location serves their own needs and desires. As such, WTP is dependent not only on the characteristics of residential properties, but also on characteristics of the homebuyers themselves.

Antoniou and Picard (2015) link household composition to WTP for urban amenities in relation to urban sustainability and household well-being. They state that households with children exhibit higher WTP for access to public OS and school quality, whereas young households without children prefer other amenities more related to adult recreation and leisure. Walker & Li (2007), through latent class choice models, found three types of lifestyles that are useful to inform location choice decisions, where they defined lifestyles as individuals' built environment preferences. The three categories they derived were urban dwellers, suburban dwellers, and transit-riders. These studies highlight the importance of considering various types of households and their heterogeneous preferences when trying to unpack household location choice decisions.

It is relevant to consider residential self-selection here, briefly. Residential self-selection is the process of people selecting themselves into certain residential environments because of their personal preferences. Because of this self-selection, a bias in estimating WTP can arise if resident characteristics are unaccounted for, as homebuyer WTP for locational characteristics is likely related to the locations these individuals ultimately decide to live in. This phenomenon is often explored in the mode choice and public health literature, where researchers often seek to identify built form impacts on rates of walking and physical activity exclusive of centrally located individuals' personal predisposition to walking (Boone-Heinonen et al., 2009; Cao, Mokhtarian, & Handy, 2007; Baar, Romppel, Igel, Brähler, & Grande, 2014). Since individuals who tend to walk more are more likely to locate in environments where walking is easy and enjoyable, it is difficult to distinguish how much walking behaviour is due to environmental factors and how much is due to the generally more active lifestyles of these residents. Self-selection can likewise be applied in considering homebuyers' valuation of locational amenities, where it is likely that homebuyers select themselves into neighbourhoods where the amenities they would be willing to pay more for than others are located.

2.1.4 Incorporating homebuyer heterogeneity into ABMs

There are two types of ABMs, which require very different data inputs, according to Buchmann, et al. (2016). 'Picasso' models are abstract and based in theory, usually used to explain one core process. What the authors call 'photograph' models are data-driven ABMs that model the unique traits of the particular case being simulated in detail. These photograph models often are highly accurate but have low overall generalizability to other contexts (Buchmann, et al., 2016). The WARM model, that this thesis is part of, would be described as a 'photograph' land-market-transport ABM. Buchmann, et al. state that a common way to include complexity in ABMS is through agent heterogeneity, where one major source of agent heterogeneity is through differentiated agent characteristics (2016). Filatova, van der Veen, & Parker provide one example of this, where they modelled agent heterogeneity in the context of the land-market transactions for the conversion of agricultural to urban land use (2009).

ABMs are a suitable platform for the incorporation of the heterogeneous agent preferences into complex systems modelling, due to the individual level simulations that define them. However, there is little consensus regarding the appropriate elements of that heterogeneity to represent and the effects of this heterogeneity on land-market outcomes at the aggregate level (Huang, Parker, Sun, & Filatova 2013). Generally, attempts to estimate heterogeneous agent-preferences for residential location decisions in ABMS have been driven by stated preference data (Brown & Robinson, 2006). This current research aims to estimate heterogeneous residential preferences through revealed preference methods, described in the following section, which are less prone to the biases outlined below.

2.1.5 Stated vs. Revealed Preferences

There are a number of methods to estimate the WTP for goods that are not traded directly or individually, such as individual housing characteristics. These methods consist of two broad categories: stated and revealed preference approaches. The predominate stated preference approach for assessing WTP for environmental features is the contingent valuation method, where survey respondents are asked to make hypothetical choices given a range of options with various criteria (Whitehead, Pattanayak, Van Houtven, & Gelso, 2008). One benefit of stated preference methods is that they are able to estimate the values of goods and services in markets where these goods and services do not yet exist (Bartholomew & Ewing, 2011; Kong, Yin, & Nakagoshi, 2007; Mitchell & Carson, 1989). Another benefit of stated preference approaches is that they are able to collect and identify the motivating factors that contribute to homebuyer preferences, whereas revealed preference approaches can simply estimate the preferences themselves (Bristow & Wardman, 2006).

While stated preference methods are more appropriate when assessing WTP for things that do not yet exist and in instances where understanding individual's motivations for determining WTP are necessary, a revealed preference method can more credibly capture individuals' WTP. Through statistical analysis of observed transactions, revealed preference approaches can determine how much an individual actually did pay for existing environmental amenities (Geoghegan, 2002). This avoids a potential bias present in stated preference methods, wherein respondents tend to overestimate the value they place in socially moral options (Whitehead et. al., 2008; Kong et al., 2007, Bartholomew & Ewing, 2011). Therefore, estimates from revealed preference approaches can provide greater validity and are preferred in situations where the amenity whose value is being estimated already exists.

2.2 Applied Framework: The hedonic method for the valuation of intensification amenities

The following sections operationalize the theoretical understanding of homebuyer's relationship with contemporary built form outlined in the previous section, placing this work squarely in the realm of the contemporary hedonic model. The following will explain the development process of hedonic models. First, a brief exposition on the historical antecedents to the modern hedonic model is presented. Next, the general hedonic model, which serves as the foundation for modern hedonic studies is provided. Following this, the process of model specification, consisting of variable selection and functional form decision, are outlined. As well, this section explains the use and interpretation of interaction terms in regression models. Finally, this section will briefly discuss the problem of spatial autocorrelation and how it is managed through the use of spatially explicit models.

2.2.1 Historical antecedents

Haas was the first to utilize hedonic methods in 1922, estimating prices for agricultural lands as a function of city size and distance to centre; however, it was not until roughly 15 years later that the term ‘hedonic’ itself came to existence (Colwell & Dilmore, 1999). The bulk of contemporary hedonic modelling literature relies on the work Lancaster, who first formally described the utility of good as being the product of the utility derived from that good’s constituent elements – rather than as a singular utility for the whole (1966). Rosen built on this theoretical framework set out by Lancaster and pioneered modern hedonic methods, with consideration as well to seminal theories of Becker (1965) and Muth (1969). Rosen (1974) first described the formulation of implicit, or *hedonic*, prices for differentiated goods as an estimable vector of attributes of those goods. Generally, early hedonic models relied heavily on the AMM framework, estimating distance to city centre as the primary source of variation in land values. Current hedonic methods are moving away from a strict adherence AMM model in favour of the view of land prices as complexly determined and polycentric rather than definable through distance to a singular centre (Redfearn, 2009).

The application of hedonic models in contemporary literature is quite broad, being used to estimate, among other things, the impacts of transit improvements, public amenities, crime, undesirable land uses, noise, natural disasters, and environmental rehabilitation and degradation on home values (Billings, 2011; Chen, 2015; Cho, Roberts, & Kim, 2011; Cohen & Coughlin, 2008; Dubé, Des Rosiers, Thériault, & Dib , 2011; Neupane & Gustavson 2008). Many hedonic models, predominately those interested in estimating transportation-related effects, have called upon the AMM model to provide theoretical support (Bartholomew & Ewing, 2011; Dziauddin et al., 2014; Geoghegan, 2002; Hess & Almeida, 2007; Higgins & Kanaroglou, 2016; Ibeas, Cordera, dell’Olio, Coppola, & Dominguez, 2012; Ottensman, Payton, & Man, 2008). Considering the revised AMM model presented above, the potential to estimate the effects of a number of complex property price determinants simultaneously makes hedonic modelling an appealing instrument for the deconstruction of home values.

While many studies have indeed estimated residential preferences for characteristics similar to those investigated in this thesis, it is understood that the WTP for access to public infrastructures and amenities differs based on geographic and socioeconomic context (Dubé et al., 2011). Therefore, this study is necessary to estimate the specific WTP for these amenities and infrastructures in KW. It is necessary to understand that the WTP estimates provided through hedonic model estimates are essentially a lower bound of actual homebuyer WTP. Since models estimated using transaction data are inherently the point of interaction between WTP and WTA, it can be assumed that homebuyers may actually be willing to pay more than what was actually paid, but not less. The difference between the WTP estimates produced in this work and true homebuyer WTP can be described as consumer surplus. For a fuller description of consumer surplus, see Mankiw (1998).

2.2.2 The General Hedonic Model

The broad types of hedonic model components are fairly consistent across studies. While often authors will choose different names for the vectors of attributes in hedonic models, they generally represent: homes’ structural or property characteristics, such as homes’ living area and age; neighbourhood characteristics, such as school quality and population density; and locational or environmental characteristics, such as access to amenities or proximity to disamenities (Dekkers & van Der Straaten, 2009; Neupane & Gustavson, 2008; Ottensman et al., 2008). In this work, the following structure of the hedonic model will be applied, consistent with that used by Neupane & Gustavson (2008):

$$Y_i = \beta_0 + \beta_1 \times S_i + \beta_2 \times N_i + \beta_3 \times E_i + \varepsilon$$

Equation 1: The general hedonic model

where Y_i is a vector the properties' values; β_0 is an estimated intercept; S, N and E are matrices of the structural, neighbourhood, and environmental characteristics, respectively; $\beta_{1,}$ $\beta_{2,}$ and β_3 are vectors of the estimated parameters for each independent variable; and ε is an error term.

Higgins and Kanaroglou (2016) provide an overview of forty years' worth of research quantifying the land-value uplift (LVU) associated with LRT. In this work, they identify three areas of potential improvement in these studies, two of which are relevant here. The first relevant suggestion is the explicit consideration and estimation of the LVU impacts associated with the components of TOD that accompany new transit implementation, such as walkability, density, and amenity provision. The second relevant suggestion is a more widespread usage of spatial methods and controls, including spatial econometric methods to account for spatial dependence and heterogeneity as well as the inclusion of cumulative opportunity or gravity based accessibility measure and variables representing local socio-economic conditions and trends. Their third suggestion is that there is a need to model the access provided by new LRT service explicitly; however, since the KW LRT is not yet operational these accessibility benefits do not yet exist (2016). These considerations are aptly addressed in this thesis through the selection of variables outlined a following chapters.

2.2.3 Time-dimension in hedonic models

Hedonic studies can be broadly separated into cross-sectional, panel, or time-series methods (Wooldridge, 2012). Many studies use panel and time-series econometric methods in hedonic modelling to estimate the causal impacts of public amenity provision (Dubé et al., 2011); however, other studies find meaningful inferences from simple cross-sectional analysis (Duncan, 2010). The difference-in-differences model, a panel data method, allows identification of causal relationships while controlling for unknown variables through a fixed effect and thereby reduces the negative impacts of autocorrelation and heteroskedasticity (Dubé et al, 2011; Gibbons & Machin, 2005). Likewise, Wang (2010) estimated property demand and amenity-specific value impacts through a time-series hedonic price model. While the literature points to panel and time-series methods as optimal to understand the causal impacts of intensification investments and infrastructures, many valuable insights have been gained through the use of cross-sectional hedonic models (Cotteleer & van Kooten, 2012; Lutzenhiser & Netusil, 2001; Kong et al., 2007). The work presented in this current thesis employs a cross-sectional model, and as such is able to identify only associative effects of property characteristics on property values, rather than causal effects.

2.2.4 Model Specification

Model specification includes two general components – the selection of variables for the model and the decision of an appropriate functional form to represent the relationships between them.

2.2.4.1 Functional Form

Various functional forms can be applied to estimate non-linear relationships between independent and dependent variables. In a review of hedonic studies related specifically to LRT, Higgins & Kanaroglou found that most hedonic studies had employed the level-level form(untransformed independent and dependent variables) (2016). The log-level form (logged dependent and untransformed independent variables) is also often used, where the log of home prices is regressed on the characteristics to obtain the semi-elasticity of price with respect to its characteristics— that is, the percent change in price attributable to a unit increase or decrease in the characteristics (Dubé et al., 2011). Geoghegan (2002) suggests that the relationship between lot size, year built, and square footage and a property's value is nonlinear and instead includes the log of these variables in her model. The functional form used in this research will be outlined in Chapter 5.

2.2.4.2 Variable selection

This section will first compare the two most common dependent variables in hedonic models, assessed values and transacted values. Following this, a discussion on the choice of independent variables will be presented, with a subsection dedicated to OS access specifically. Finally, this section will close with a short discussion of omitted variable bias (OVB) and multicollinearity.

Dependent

There are two common variables used as dependent variables in hedonic models, assessed values and observed transaction values. It has been argued that both assessed and sales values of homes are inherently proxies for the true value of a home, which is unknown (Doss & Taff, 1996; Cotteleer & van kooten, 2012).Many authors have investigated transaction value, while others solely assessed value, and some authors have simultaneously estimated both (Bowman, Thompson, & Colletti, 2009; Jaeger & Plantinga, 2007; Lee, Taylor, & Hong, 2008).

Assessed values are often created from hedonic models themselves, and so some authors have cautioned of the potential for hedonic models specified with assessed values to only recapture estimates from the assessors' original models (Cotteleer & van Kooten, 2012). Because of a time lag in property assessment, where neighbourhood appreciation is not immediately captured, there is also a potential for bias in assessed value models (Goolsby, 1997). Others have noted a source of potential bias from hedonic models using assessed values wherein assessed values are often derived from incomplete information about externalities affecting properties (Kitchen & Hendon, 1967; Bowman et al., 2009).

Transaction values, on the other hand, have the potential to contain values that do not represent true market sales. Various criteria exist to limit transaction data to only contain market rate, or arms-length, sales, which include removing sales less than $10,000, removing sales that are substantially different than their respective assessed values, and removing sales for less than a predetermined price to building area ratio (Chatman, Tulach, & Kim, 2011; Cervero, 2003; Henderson & Song, 2008).

From these studies, it would appear that transaction values, if adequately pre-screened to remove non-market rate sales, have the potential to more explicitly represent the true market value of a home considering its unique context. However, because of conflicting opinions on which measure best represents true market value, both will be investigated in this work. For a more in depth comparison of using expert opinion (assessed values) versus transaction values of homes, see Cotteleer and van Kooten (2012).

Independent

The choice of independent variables in a hedonic regression is generally informed by two factors: 1) that the model's variables should accurately represent the home price determinants of interest for the specific study being developed, and 2) that the appropriate intervening or moderating factors are appropriately accounted for. While the broad categories of independent variables included are fairly consistent among hedonic studies, the individual variables used can differ dramatically. Because of this high variation in potential model variables, misspecification is a serious concern in hedonic modelling – where coefficient estimates from hedonic models are highly sensitive to changes in model specification (Kuminoff, Parmeter, & Pope, 2010).

As many hedonic studies have their basis in the AMM framework, which was described in the first half of this chapter, a variable representing employment access is generally found in hedonic models. Following from the AMM conceptualization of urban structure, most studies control for this employment access through the use of a variable representing distance to city centre; however, other models have used distance to secondary centres as well as primary employment centres (Ottensmann, et al., 2008). It has been argued that travel times to destinations may control better for urban location than simple distance measures (Adair, McGreal, Smyth, Cooper, & Ryley, 2000; Des Rosiers, Thériault, & Villeneuve, 2000; Franklin & Waddell, 2003).

Sirmans, Macpherson, & Zietz conducted a "meta-regression" to understand the range of hedonic models, including 125 separate hedonic models in their analysis (2005). The most common structural variables they found were square footage, lot size, age, and garage spaces. Neighbourhood and environmental variables were far less consistent in the studies they reviewed, with many pages of candidate variables listed. One of the most difficult aspects of modelling the land-market is accounting for the enormous amount of potential neighbourhood controls (Can, 1990; Munroe, 2007). Neighbourhood controls aim to account for the socioeconomic character of areas, some often applied neighbourhood variables include school quality, population density, planning controls, crime levels, fixed effects for administrative neighbourhoods, employment access, racial composition, and many more (Hayes & Taylor, 1996; Ottensmann, et al., 2008; Yoo, Im, & Wagner, 2012)

In their review of studies specifically estimating the land value uplift associated with LRT, Higgins and Kanaroglou recommend that "variables should be specified to control for relevant price impacts in a particular study area, including gravity or cumulative measures of regional and local accessibility and indicators of neighbourhood quality and economic growth trends " (p.17, 2016). This arguably holds true for any hedonic model interested in estimating home price relationships to environmental features in intensifying environments. Redfearn found that studies estimating average values related to public amenities inadequately captured the non-static value-added by these amenities over space and time (2009), and so this research will employ methods to uncover heterogeneous preferences for the primary amenities investigated in this thesis over space.

Intensification related variables

Many Studies have employed WalkScore as a measure of walkability in hedonic models. Walkscore defines walkability as the ability to access amenities or points of interest over the pedestrian street network (Washington, 2013, Rauterkus & Miller, 2011; Pivo & Fisher, 2011). Raterkaus and Miller found in a hedonic analysis of Jefferson County, Alabama that the real estate premium associated with walkability, as indicated through WalkScore declined with distance city centre (2011). This shows that the effects of walkability on home values are not constant across space.

To capture the effects of access to transit in hedonic models, one typical method is to generate a dummy variable for homes within 800m from transit stations, representing approximately a 10 minute walk; however this 800m buffer method is inconsistent in the literature (Guerra, Cervero, & Tischler, 2011). For this current thesis, access to transit will be specified using a spatial separation model using pedestrian network distances, which is described in the following subsection on OS access. However, a dummy variable representing whether a home is located within the CTC, which is a roughly a modified 800m buffer from future LRT stations, will be included in the models of this thesis.

Relatedly, Atkinson-Palombo stresses the importance of considering the role of land-use planning that accompanies transit (2010). She found that overlay zoning, a planning tool used to promote certain types of development, had a significant impact on the state of TOD and on residential prices near transit stations. This work will use the boundary of the CTC in KW to identify how relationships between intensification amenities and home values differ between the area being planned, often up-zoned, specifically to support the regional LRT and regional development goals.

Access to OS

Through rapid urbanization and increases in urban density, the rate of OS being lost to development in North America has been increasing in the 2000s (McConnell & Walls, 2005). Intensification has the potential to reduce the amount of natural space in urban cores while preserving it at the fringe. OS presents many non-market benefits such as the opportunity for relaxation, socialization, and physical activity (Kaczynski, Potwarka, Smale & Havitz, 2009; Konijnendijk, Annerstedt, Busse Nielsen, & Maruthaveeran, 2013; Zhou & Rana, 2012). These non-market benefits represent themselves in the housing market through homebuyers' WTP for housing with access to OS. A great many studies have attempted to estimate the price effects associated with natural amenities on home values (Anderson & West, 2006; Crompton, 2001; Dehring & Dunse, 2006; Irwin, 2002; Luttik, 2000; Morancho, 2003; Smith, Poulos, & Kim, 2002; Troy & Grove, 2008). Urban OS is a common variable in hedonic models (Geoghegan, 2002). The estimation of homebuyer WTP for OS is pertinent in a context of urban intensification.

To estimate OS effects on home values, most studies employ landscape metrics. Landscape metrics are produced using specialized spatial analysis software such as FRAGSTATS with raster land-cover data (McGarigal, Cushman, & Ene, 2012). These landscape metrics commonly include patch richness, patch contiguity, measures of the radius and sizes of patches, the shape of patches, and edge density, to quantify landscape patterns (Geoghegan, 2002; Kong et al., 2007; Luttik, 2000). However, the research presented in this thesis uses *access to OS*, a more human-centric approach to estimating the amenity impact of OS. Where landscape metrics measure the spatial composition of OSs near homes, access represents the ease at which homeowners are able to reach these OSs. Zhang, Lu, Holt (2011) provide a detailed description of the common methods employed in measuring access to parks; they summarize that the

> "spatial accessibility of neighborhood parks in the literature can be categorized into three general approaches: 1) spatial proximity to parks, which measures travel costs in overcoming spatial separation between the locations of population and parks; 2) the container approach, which measures the existence or density of parks in a defined geographic area; and 3) the spatial interaction modelling approach, commonly known as gravity model-based approach, which measures the potential spatial accessibility of parks." (p.2)

These methods are consistent across the literature, although their naming conventions differ. In this work, the first measure will be called 'spatial separation', the second, 'cumulative opportunities' and the third, 'gravity-based' models. While in the case of Zhang, Lu, and Holt's research, where the interest was only in measuring access to parks (2011), this thesis uses these methods to measure access to a variety of publicly accessible OSs including forests and woodlots, parks and greenspaces, golf courses, and cemeteries.

Spatial separation models are the simplest form of access measure, which use only the distance between properties and the nearest opportunity, in this case OS. Cumulative opportunities models are slightly more sophisticated, as they consider not only the distance to access amenities, but also the abundance of opportunities within that distance. One drawback of cumulative opportunity models is that they inherently require the setting of an often-arbitrary distance threshold within which to consider accessible OSs. Neudorf, 2014 does a good job describing this problem, where cumulative opportunities measures using two different thresholds would produce drastically different characterizations of accessibility. Finally, gravity-based models are the most complex measure of access investigated in this thesis. The gravity-based model employed in this work is given in the following chapter, where access is the product of both the attractiveness of as well as the distance to amenities, which are both parametrized with decays. The attractiveness of OS to be used for this model will be determined by OSs' size, following from Kong et al. (2007). Of note, in a study investigating 20 different impedance functions in pedestrian measures of gravity-based OS access, the researchers found that medium distances, approximately 200-400m away were the most sensitive to changes in the distance decay (Vale & Pereira, 2016). Because of the sensitivity of gravity-based measures to different decay parameterizations, various parametrizations were tested for use in this thesis.

Additionally, as recommended by Koohsari, Kaczynski, Giles-Corti, & Karakiewicz (2013), in the calculation of access, this research will utilize network distances rather than Euclidean distance. Figure 2 gives a graphical representation of the difference between using Euclidean and network distances:

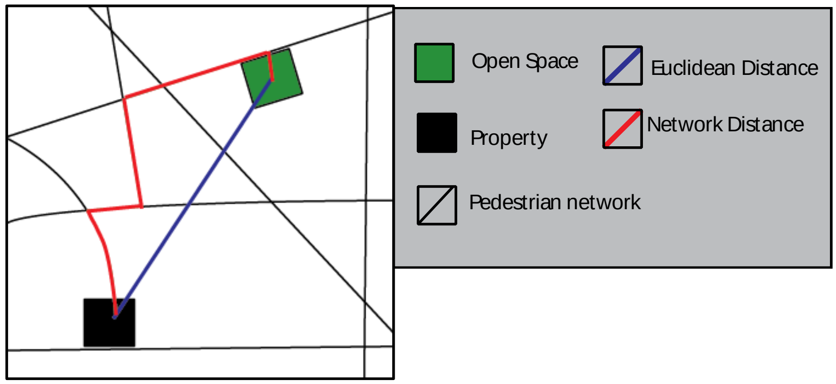


Figure 2 – Euclidean Distance versus Network Distance (created by author)

As can be seen from this diagram, using OS as the example opportunity or activity site, Euclidean distance measures will in most cases underestimate the real distance required to travel to a destination. Historically, Manhattan distance, a rectilinear measure of distance, has been suggested as more appropriate than Euclidean distance in urban contexts with regular grid pattern street network (Krarup & Pruzan, 1980); however the irregular, curvilinear street pattern associated with the suburban development style of the study region would not be appropriately represented using this method, and the availability of network data allows the use of actual network values which are most preferred.

In addition to the value of access to OS, home values have been shown to be affected by OS adjacency as well (Kitchen & Hendon, 1967). Homes that share borders with OS receive an aesthetic amenity value separate from the use value of OS. This adjacency will be controlled for in the modelling work of this thesis.

2.2.4.3 Omitted Variable Bias (OVB) and Multicollinearity

Considering the overall inconsistency in independent variables used in hedonic models, it is necessary to consider the potential problems of OVB and multicollinearity. OVB occurs when an important explanatory variable that is correlated with both the dependent and at least one of the independent variables of a regression is left out. This leads to biased results, as the model is unable to estimate the effect of the included independent variable(s) exclusive of the impact of the omitted one (Wooldridge, 2012). Alternatively, multicollinearity occurs when there are multiple predictor variables that are related to the dependent variable in the same way. When multicollinear variables are present, a regression model is unable to reliably estimate the collinear variables' singular impacts (Wooldridge, 2012).

OVB is the result of under-specification, while multicollinearity results from an over-specified model. While it would at first seem beneficial to estimate all factors known to affect home prices in hedonic models, multicollinearity prevents this (Anderson & West, 2006; Yoo et al., 2012). Clark & Hosking (1986) set a threshold of concern for collinear variables at a correlation of ± 0.7. Other authors have investigated concerns of multicollinearity amongst variables with much smaller correlations (Atkinson-Palombo, 2010). The effects of OVB produce a biased model, whereas the effects of multicollinearity only interfere with the estimation of parameters for the multicollinear variables themselves. Multicollinearity causes more immediately apparent problems than those of OVB, as estimates of multicollinear variables generally have large variances and often change signs unpredictably with unrelated changes in model specification (Yoo et al., 2012). The effects of OVB are sometimes less overt, where an affected model may or may not produce coefficient estimates of the correct sign, and these estimates may or may not appear robust under various model specifications provided that the relevant omitted variable is absent. Avoiding omitted variable bias in this research relies more so on a strong theoretical understanding of the forces shaping home prices, while multicollinearity is avoided through comparisons of candidate variables and iterative model specification.

In developing a hedonic model, a careful balance of variables is required to ensure that important determinants of home value are neither omitted nor included repetitively. The following chapters of this thesis will provide an extensive examination of the distributions of the built form and socio-economic variables in KW in effort to ensure that no important and relevant predictors are omitted and that there is enough variation between predictors to avoid multicollinearity.

2.2.4.4 Interactions to model synergistic or moderating effects

Interaction terms are used in regression to identify the synergistic or moderating impacts of independent variables on each other in the estimation of the dependent variable (Wooldridge, 2012). Interactions take the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

Equation 2: Simple Interaction Model

The specific interpretation of the equation changes when using either categorical or continuous data within the interaction, so both will be described separately below.

For the continuous case, the effect of the interaction between X_1 and X_2 is estimated through the β_3 parameter. A positive β_3 would indicate that there is a synergistic impact between X_1 and X_2. In the case of a hedonic regression, a positive β_3 would indicate that home values increase when two variables increase together, more so than should they increase on their own. Alternatively, a negative β_3 would indicate a moderating relationship between variables, where estimated WTP decreases when the variable values increase simultaneously, to a greater extent than should one rise individually.

For interactions with categorical variables, these interactions can be used to estimate different relationships by categorical groups. Indicator variables take either a 0 or a 1, depending on whether an observation meets the criteria for the category (1) or not (0). In an instance where X_1 in Equation 2 is an indicator variable and X_2 is continuous: β_1 would represent a fixed effect for group membership; β_2

would represent the effect of the continuous variable for non-members of the group X_1; and, β_3 would represent difference of the effect of X_2 for members of the group X_1 as compared to the effect for non-members.

Using interactions, Duncan found a synergistic relationship between public transit and pedestrian friendliness in his analysis of the San Diego housing market (2010).Henderson and Song investigated substitutional effects between public OS and private OS using interactions (2008). Both interactions between various intensification amenities and between private and public OS are investigated in this work.

2.2.4.5 Spatial Regression Methods

2.2.4.5.1 Spatial Autocorrelation

The spatial nature of the housing market presents a problem in appropriately estimating hedonic models. Spatial autocorrelation is often described using Tobler's first law of Geography – that everything is related, but closer things are more related than things farther away (Tobler, 1970), and where spatial dependence is the product of home prices being codetermined to some extent along with the home prices of nearby properties (Bowen, Mikelbank, & Prestegaard, 2001).

Traditionally, hedonic models have been estimated through the use of an OLS model. However, one assumption necessary for OLS to be the best linear unbiased estimator is that the model error is independent and identically distributed (Wooldridge, 2012). Spatially autocorrelated home transaction data violate this, as the values of homes to some extent do depend on the values of nearby homes included as observations. This non-independent data introduces endogeneity to the model, where the errors of OLS regressions on spatial data are heteroscedastic. Spatial models can be used when variables are autocorrelated over space, which correct for the bias that appears in the variance of OLS estimates (Cohen & Coughlin, 2008; Henderson & Song, 2008; Mueller & Loomis, 2008; Neupane & Gustavson, 2008). Páez and Scott explain that models that do not account for spatial effects can give misleadingly significant results, which disappear once spatial effects are controlled for (2005).

2.2.4.5.2 Spatial Models

While there are various models that aim to correct for these spatial effects, two of the most common are the spatial lag and spatial error models, (SEMs) which fall under the category of maximum likelihood estimation methods (Anselin & Bera, 1998; Krause & Bitter, 2012). Spatial Lag Models (SLMs) account for spatial dependence by weighting nearby property values and including them as independent variables in the regression, as shown in Equation 3:

$$Y = \rho W y + X\beta + \varepsilon$$

Equation 3: SLM (Anselin & Bera, 1998)

where the lagged dependent variable is represented through the ρWy term. SEMs account for the effects of spatial autocorrelation through the error terms of the regression, as in Equation 4:

$$Y = X\beta + u$$

$$u = \lambda W u + \varepsilon$$

Equation 4: SEM (Anselin & Bera, 1998)

where the total error, *u*, is estimated as an individual error, $\mathcal{E}$, plus the weighted errors of neighbouring observations, λWu.

A necessary consideration in spatial regression is of the choice of spatial weight matrix. A spatial weight matrix is a tool used to specify spatial relationships between observations (neighbours) within the data. Specification of the weight matrix necessary to perform spatial regression has been noted as *ad hoc* and *a priori* (Anselin, 1988; Henderson & Song, 2008). This means that the application of weight matrices across hedonic models inconsistent and generally developed through deduction of context-specific spatial effects. These neighbour specifications include: using observations within distance thresholds, using measures of adjacency between observations on a contiguous surface, and using a specific number of the nearest neighbours to each observation (called k-nearest neighbours or knn). Krause & Bitter caution that beta coefficients are often misinterpreted from SLMs, as estimates are largely impacted by the choice of spatial weight matrix, so it is necessary to carefully understand the weighting methods used when making inferences (2012).

2.2.4.6 Model Performance

Consideration must also be given to the indicators of model performance. Independent variables can be tested for significance with a t-test or alternatively a z-test if the sample size is large or population variance known, which identify whether a variable's impact on property values is significantly different from 0. Heteroskedasticity can be tested through a Breusch-Pagan test (Wooldridge, 2012). LeSage and Pace recommend comparing tables of coefficient estimates and standard deviations as well as associated t-statistics of spatial models to their non-spatial OLS counterparts, in order to assess potential misspecification (2009). Spatial effects can be tested for through a variety of statistics including Lagrange multiplier, and Moran's I (Anselin, 1988). In spatial models, R^2 is an unreliable measure of goodness-of-fit. Instead, authors suggest comparing OLS and spatial models using Akaike Information Criterion (AIC), Schwarz Criterion, Log-Likelihood, and alternative or pseudo R^2s (Anselin & Bera, 1998; LeSage & Pace, 2009). Likelihood ratio tests of spatial models can determine if spatial effects are significant in the spatially explicit models (Anselin & Bera, 1998).

2.3 Gaps in Literature

This research attempts to fill, to some extent, the following gaps in the literature:

1) There are many studies that use stated or revealed preference data to estimate homebuyer WTP, however there is little research that has yet attempted to incorporate stated preference and individual-level demographic data specifically in hedonic models to identify heterogeneous WTP values

2) There are no existing hedonic model results for KW, and since hedonic models are context-specific there is a need to explore this case
3) While some studies have found synergistic impacts between intensification amenities through hedonic modelling, there is still scant literature on the topic and a need for more applications
4) While there is a common understanding of the spatial layout of the housing market, there is little published research that demonstrates such an in depth mapping and comparison of the spatial patterns of candidate hedonic model variables along with the impacts of these patterns on model results
5) Heterogeneous homebuyer preferences have usually been incorporated into ABMs through stated preference methods and are not found in hedonic models at the individual level. This research aims to identify heterogeneous homebuyer preferences by combining stated and revealed preference approaches, and to use these to parameterize an ABM
6) While most hedonic studies have employed landscape metrics to capture the effect of OS, this work adds to the small amount of literature on the use of access to OS as a variable in hedonic model

3 Pre-Regression Methods: Variable collection, creation, manipulation, and comparison

Before running regression models, careful consideration must be given to the components of the models being run. This section will explain the methods used to collect, create, modify, and compare candidate model variables. The results of this pre-regression work, which are found in the following chapter, support the final empirical model specifications outlined in Chapter 5.

This section includes four parts. First, it puts forward a theoretical model of homebuyer WTP in KW from which candidate variables were selected. Second, a section outlines the four dependent variables that were investigated for the models. Third, it explains how the candidate independent variables for the models were developed. Lastly, it highlights the exploratory data analyses that were used to evaluate the potential model components and understand their relationships. All data used in the following methods are more fully described in Appendix 1 - Data.

3.1 Dependent Variables

This research examines both recorded sales price of homes as well as the assessed value of homes as dependent variables for hedonic regression. These two dependent variables were provided through a collaborative, project-specific license agreement with the ROW, dated September 9, 2015.

3.1.1 Assessed Value

As discussed in Chapter 2, studies have often employed assessed values as the dependent variable in hedonic models. However, the consequences of using assessed values as opposed to transaction values in terms of model performance are under-evaluated. MPAC property assessment values were obtained through the ROW. These values are estimates of property value determined by an arm's-length organization, which are used in the calculation of property taxes paid by property owners.

The MPAC estimates also contained a property code that classified the properties into specific categories. The property codes used for this research are located in Table 1. These property codes were manually selected from the 300 Series, residential, property codes. This removed certain types of properties that were expected to have a market value determined either by different property characteristics or by different mechanisms than the usual residential buyer-seller negotiation process.

The selected property types are grouped into single-detached, semi-detached and duplexes, and townhouses.

TABLE OF MPAC PROPERTY CODES USED

	Single-detached
	Semi-detached & duplexes
	Tow

Property Code	Description
301	Single family detached (not on water)
305	Link home – are homes linked together at the footing or foundation by a wall above or below grade.
309	Freehold Townhouse/Row house – more than two units in a row with separate ownership
311	Semi-detached residential – two residential homes sharing a common center wall with separate ownership.
313	Single family detached on water – year round residence
332	Typically a Duplex – residential structure with two self-contained units.

Table 1: Property Codes Included in Analysis (MPAC, 2016a)

3.1.2 Adjusted Value

Transacted home values were obtained through a project-specific license agreement with the ROW. In order to use a cross-section spanning multiple years in this analysis, transacted values were adjusted to January 2014 dollars using a regional home price index from Statistics Canada (2015a). This adjustment ensured that the data within the cross-section could be reasonably compared. This regional home price index controls for regional-scale temporal trends in real estate values, specifically price inflation and regional appreciation. The transacted home prices were adjusted in the following manner:

$$Adjusted\ Value_i = \frac{Transacted\ Value_{it}}{Home\ Price\ Index_t} \times 110.9$$

Equation 5: Adjusted value

Where the *Adjusted Value* of observation i is equal to the transacted value at time t divided by the HPI value at time t, multiplied by the 110.9, which is the index value in January 2014.

A cross-section of home sales from 1998-2015 was adjusted using the above Equation 5, which was used to calculate a rate of appreciation (ROA) variable described later in this chapter. This cross-section was further reduced to only those homes sold after 2005 in the final regression models, where the year 2005 was used as a minimum for two reasons. The first reason to select this year was that most of the independent variables in this research were calculated using data representative of 2012 and nearby years, so removing earlier years from the set increased the comparability of property sales, considering the time-invariant measures of those properties' characteristics. The second reason to use this year was that 2005 is the year that the Places to Grow Act was released, which was followed by a push for intensification across the province of Ontario.

Figure 4 shows the average values of the home price index for the years of the sample. The home price index follows a nearly linear pattern, with sthe highest variation in the growth rate found betwee 2005 and 2012

Figure 4: Home Price Index over Time

One challenge in using recorded sales price data in statistical home price modelling is the prevalence of observations that do not represent the homes' true market value. The following section will explain the methods used to eradicate these non-market sales from the dataset in this research.

3.1.2.1 *Removing non-market rate sales from transaction dataset*

All work performed to remove non-market sales was done in collaboration with ROW Planning staff member David Stubbs. A structured approach was taken to remove non-market sales from the dataset. This approach involved investigating variable relationships, identifying outliers and unexpected observations, and ultimately developing criteria to systematically exclude non-market sales. Exploratory data analysis and visualization were employed to find sales with values that were substantially different from market-rate sales and to identify criteria for eliminating similar cases. This work also helped to isolate sales to owner-occupiers, as opposed to landlords, developers, or investors – as this research is focused on estimating homebuyer preferences rather than unpacking the WTP exhibited by investors (see Antanaitis, 2014 for more information on the decision-making of developers and investors engaged in intensification in KW). Multifamily homes presented a complex range of unit and building sales, and it was determined that identifying individual unit sales of multifamily dwellings was beyond the scope of this research.

First, scatterplots were inspected between *adjusted value* and *living area* and between *assessed value* and *adjusted value* to identify outliers. These observations were then inspected tabularly in Stata and Access and visually in ArcMap to understand the source of these unexpected values. From this analysis of transaction values, home size, and assessment values, the following methods were devised to limit the database to only those sales representing a home's genuine market value paid by an intended owner-occupant:

1. Using only transactions of instrument type 'T' (Transfer (Grant)), as these were sales that occurred through regular deed transfer mechanisms.
2. Using only sales over $10,000 in value, as was done in Chatman, et al. (2011).
3. Removing observations with duplicated instrument numbers. Instrument numbers identify the document of a particular transfer, so cases where duplicated instrument numbers were found to represent cases where multiple amounts of money were transferred in the same transaction, which is an unusual circumstance.
4. Removing observations where an identical name was found in both a buyer name field as well as a seller name field. These were sales of partial ownership, where one person remained in ownership while another name was either added or removed to the ownership, usually due to the formation or dissolution of partnerships.
5. Removing sales of less than $20 per estimated square foot of the building's size. These sales were found through individual inspection to contain either:
    a. the same surnames in the to and from fields, indicating a sale within a family or the addition of a family member onto an existing ownership
    b. the same given name in the to and from field, indicating a change in an owner's surname name because of marriage or divorce, and a sale of partial ownership as in 4, above
6. Through manual inspection, all observations that did not contain commas, which are used to separate last and first names in the data, in their names , were identified to be companies, builders, developers, trusts, or other organizations, rather than owner-occupants. These buyers, who had no commas in their names, were omitted from the analysis. After removing sales to owners with no commas in their names, the sales dataset was further inspected manually for sales to other businesses and organizations whose names did contain commas, and these were also removed.
7. Removing sales that contained "(Estate)" in the seller name, meaning that the sale was not negotiated as would occur between a home seller and homebuyer in a regular market transaction.
8. Removing observations that contained a *Year Built* date that was later than the transaction date, indicating that the sale was for land rather than a home.

9. Removing sales over $2.5 million. This was intended to remove both sales of large multi-unit properties, which were likely to investors rather than owner-occupants, as well as sales of super luxury homes, whose values are largely determined by forces other than those that dictate the value of average properties.
10. Land sales and sales to family members were also further identified and removed manually by checking outliers. Specifically, outliers were identified in scatterplots of transacted value against living area and transacted values against assessed values. These observations were investigated using satellite imagery over time and ownership history information from the sales data, and removed if found to be a land sale or sale to family member.
11. Sales from companies/builders/investment trusts, those with no commas in their names or identified manually, that were less than $50 per square foot of a building's area were found to be part of a home sale structure wherein homebuyers pay separately for the land and the property structure, and were therefore removed.

3.2 Independent Variables

The following sections outline the candidate independent variables for the hedonic models. First, the environmental variables used in this research are described , including the methods used to calculate access to public OS, transit, and employment. Next, a section outlines the collection and calculation of various structural property characteristics. Finally, the last section will outline the various neighbourhood variables used in this work.

3.2.1 Environmental Variables

The environmental variables investigated in this research include access to OS, walkability, access to transit, as well as *OS adjacency.*

3.2.1.1 Access Variables

Following on the discussion on measuring access in the literature review, access variables were calculated using spatial separation, cumulative opportunities, and gravity-based access models. Access to OS was calculated with all three of the above models, while transit access was calculated as a spatial separation measure and employment access was calculated using a cumulative opportunities model. The following sections describe how these models of access were operationalized in this research.

Python Script for AcrGIS Origin-Destination Cost Matrix

A Python script, Appendix 2, was created to automate network analysis using the pedestrian network in KW. This script was used to calculate both access to OS and transit separately, as will be described in the following sections. The Python script uses the *arcpy* library to access and execute GIS functions in ArcGIS. ArcGIS's network analyst extension, in particular its *O-D Cost Matrix* tool, was essential to calculate distances along the pedestrian network between residences and public amenities. A loop was implemented in the script to overcome memory limitations in ArcGIS, where the *O-D Cost Matrix* tool was run on subsets of residential properties, and the results were merged into a single table.

Using the *O-D Cost Matrix* Tool, households within KW were set as origins, and amenity locations (OSs and transit stops) were set as destinations. Households were snapped to the road network using ArcGIS's *Near* tool to verify that they were located on an appropriate segment of the network. The *O-D Cost Matrix* tool uses Djikstra's Algorithm to find the shortest path from each of these origins to each destination within a specified threshold distance (ESRI, 2010). The settable parameters for this tool include a distance threshold between origins and destinations, where distances to destinations beyond the threshold are not calculated, as well as a threshold to snap the locations to the network, where observations that are a located farther from the network than the threshold are assumed to be inaccessible over the network.

This script, after calculating the distance from all origins to all destinations within the prespecified distance, converted all distances lower than 10m to 10m to mitigate the impacts of extremely small distances on the final access metrics. Extremely small distances were leading to extremely high measurements of access; however, it was decided that within a distance of 0m to 10m the impacts of this distance on access are practically unnoticeable. Additionally, as locations were represented only as points, a difference of 0-10m could result from the placement of the point that represented the actual polygon feature, rather than any actual difference in access.

3.2.1.2 OS Access

This OS access layer includes all available public and semi-public OS locations within KW. Specifically, the OS access calculations used OSs layers from The City of Kitchener: Parks (which also contained forests); the City of Waterloo: Parks, Environmental Lands, and Forests; and the ROW: Regional Forests, Cemeteries, and Golf courses (See Appendix 1for a full description of these data sources). These layers were merged in ArcGIS into a single layer. The *Dissolve* tool was used to remove overlapping OSs, so as to not double count their impact on OS access. With the ArcGIS tool *Feature Vertices to Points*, each OS polygon was transformed into a set of points. This conversion was decided upon after attempting to run the analysis using OS centroids and finding the results did a poor job calculating distance to both linear and large OSs. When a single point was used, the point representing the OS was assigned to only the nearest road segment, so not all access points for OSs were initially being included and the actual levels of OS access were being underestimated.

The *Near* tool was used to identify and remove the OS points that were not within 30m of the pedestrian network; this was done to ensure that vertices separated from the pedestrian network by untraversable features were not included in the analysis, as well as to reduce computing time. These points were then moved to be located along the nearest segment of the pedestrian network to verify that OSs would be accessed at the correct locations. A maximum distance of 1,000m was employed, as this provides a reasonable threshold for one to walk to an amenity within one's neighbourhood.

Spatial separation, cumulative opportunities, and gravity-based measures of OS access were calculated using the following methods:

Spatial Separation Model

For the spatial separation measure, the lowest distance to the nearest park was taken from the results of the origin-destination cost matrix for each property, and modified as per the following:

$$A_i = \min(d_{ij}) * -1$$

Equation 6: Spatial separation model

Where,

access at property i, A_i, is equal the minimum distance from i to any opportunity, j, multiplied by negative one. Multiplying by negative one is simply meant to reverse the order of the variable values, so that higher values represent greater levels of access as is the case with the other access measures. For this model of OS access, homes that were OS within the 1,000m threshold were set as being 1,200m from OS, so that they would neither contain null values, which would exclude them from the analysis, nor contain zero as a value, as this would represent the highest level of accessibility.

Cumulative Opportunities Model

For this measure, two different cumulative opportunities measures were calculated for a number of different binary thresholds. The following equation shows how the cumulative opportunities measure is calculated:

$$A_i = \sum_{i=1}^{n} (O_j \mid d_{ij} < t)$$

Equation 7: Cumulative opportunities model

Where access at the ith property, A_i, is equal to the sum of opportunities, O_j ,within a distance, d_{ij}, that is less than the binary threshold distance value, t.

For this work, two cumulative opportunity measures were tested. One measure calculated the number of OSs within the distance threshold, while the other calculated the sum of the area of OSs accessible within the distance threshold. Both of these measures were calculated for 250, 500, 750, and 1,000 metre (pedestrian network distance) values of the threshold value, t.

Gravity-Based Model

The most complex measure of OS access used in this research was a gravity-based model. The model is as follows:

$$A_i = \sum_{i=1}^{n} ({W_j}^{\alpha} \times {d_{ij}}^{-\delta})$$

Equation 8: Gravity-based model

Where,

A_i = access at property i

W_j = a measure of a destination's attractiveness

d_{ij} = the distance from origin i to destination j

and, α and δ are decay parameters for the attractiveness and distance values, respectively.

This model uses OS size as a measure of attractiveness. The decays on attractiveness and distance scale the effect of OS size and distance on the level of OS access attributed to that property. For instance, values of α lower than one reduce the marginal effect of increases in size on access as size increases, while values greater than one do the opposite. A value of 0.4 was used for α and a value of 0.15 was used for δ to ensure that even small OSs would have an impact on access if they were relatively close and that even OSs 1,000m away impacted the access value if they were large enough. These decay parameters were ultimately decided upon through visual examination of the resulting spatial patterns and data distributions of various potential specifications.

3.2.1.2.1 Public Transit Access

Public transit access was calculated using a spatial separation model, as in Equation 6. This model used the *O-D Cost Matrix* script, with transit stops set as the destinations. Distance along the pedestrian transportation network was calculated from each property centroid in KW to transit stops within 1,500m. From this table, the minimum distance from each property to a transit stop was taken. As with the spatial separation model of OS, zero values were updated to a value just beyond the search threshold for the *O-D Cost Matrix*, 1,700m.

3.2.1.2.2 Employment Access

Employment access for this model was computed as a binary threshold cumulative opportunities measure at the Traffic Analysis Zone (TAZ) scale. The binary threshold here was set to the generalized cost of the average commute made in the ROW – meaning that this variable represents the number of jobs that can be reached by the regional average commute.

The underlying data for this variable was supplied by Jason Neudorf (2014). In his work, he created an origin-destination cost matrix between each TAZ in KW. He also supplied employment data, in the form of employment counts in each zone, which he obtained from the Transportation Tomorrow Survey (TTS) (Data Management Group, 2006).

$$\mathrm{GC_{auto}} = (\mathrm{VOT} \cdot \mathrm{t_{ij\,(auto)}}) + (\mathrm{C_{var}} \cdot \mathrm{d_{ij\,(auto)}}) + \mathrm{C_{fixed}} = (\$12/\mathrm{h} \cdot \mathrm{t}) + (\$0.125 \cdot \mathrm{d}) + \$7.97$$

<u>Equation 9: Generalized Cost of Transportation</u> (Neudorf, 2014)

In this equation, the generalized cost (*GC*) of automobile transportation is calculated as: the value of a person's time (*VOT*) multiplied by time spent driving (t_{ij}), the variable costs of driving (C_{var}) multiplied by the distance driven (d_{ij}), plus a fixed cost associated with car ownership (C_{fixed}). The data and methods used for the value of time, variable costs, and fixed costs are taken from and explained more fully in Neudorf, 2014.

To calculate the average GC of commuting in the region, the average time and distance spent commuting in the ROW were taken from the TTS and National Household Survey (NHS), respectively. The average trip length for drivers in Waterloo Region between 6:00-9:00AM was 5.8km (Data Management Group, 2011). The average time spent commuting was 21.7 minutes (Statistics Canada, 2015b). These values were plugged into the generalized cost formula found above to produce an average generalized cost of commuting in Waterloo Region in a private automobile as $13.035. This average cost of commuting was applied as a binary threshold to a table of employment access costs from zone to zone, and for each zone, the number of employment opportunities within this binary threshold were summed using a *SUMIFs* function in Excel.

3.2.1.3 OS Adjacency

OS adjacency was specified as per the following:

$$OS\ adjacency_i = \begin{cases} 1, & shared\ edge\ with\ open\ space \\ 0, & no\ shared\ edges\ with\ open\ space \end{cases}$$

Equation 10: OS Adjacency

In addition to the benefits of *access to OS*, properties often see a discrete price effect attributable to the aesthetic value of OS adjacency. OS adjacency was determined using ArcMap, with the same OS and property layer as the OS access calculations described previously.

The *Neighbor Polygons* tool in ArcGIS was used to identify adjacencies between OS and property polygons. To use this tool, the property polygons and OSs were merged into a single layer, with binary identifiers to demarcate which layer each feature originated from. Adjacency between all polygons was identified using the *Neighbor Polygons* tool. This tool produced a table of the adjacencies between each OS and property polygon in the layer, with identifiers for both the source polygon of the adjacency and the neighbouring feature. Then, all observations containing OSs as the source and all containing property polygons as the neighbour were dropped from the table, which left only neighbour information for the property polygons that were neighbours to OSs. The *Delete Duplicates* tool was then used to leave only one instance for each property polygon neighbouring OS. This table was finally joined back to the original property parcel layer, with an identifier added, so that all observations from the adjacency table contained a 1 and all those found only in the original table contained a 0 in an OS adjacency field.

3.2.1.4 Walkability

Walkability index values were used from the ROW's NEWPATH model. This model was co-developed with researchers at academic institutions under the guidance of Principal Investigator Dr. Larry Frank (2009). The structure of this walkability index can be seen in the following, Equation 11, where it is the sum of the z-scores of separate measures of Intersection Density, Residential Density, Land Use Mix, and Retail Floor Area.

$$W = Z(Residential\ Density) + Z(Intersection\ Density) + Z(Land\ Use\ Mix) + Z(Retail\ Floor\ Area)$$

Equation 11: Walkability Index (ROW, 2009)

More information on the details of this model's development can be found in (ROW, 2009).

3.2.2 Structural Variables

Structural variables are the most consistently applied of the three categories of variables in hedonic models. The structural property variables investigated for this research include the type, estimated living area, street frontage, estimated private yard size, and age of the property structures. The formulations of these variables are outlined in the following section.

3.2.2.1 Property Type

Property type was obtained from the MPAC assessment data, as explained in Appendix 1, where Table 1 presents the property codes that were ultimately included in this analysis and how they were grouped for inclusion in the model. The three property types included in this work were used to calculate different descriptive statistics and to run different regressions for each property type.

3.2.2.2 Living Area

The formula for estimating living area in this research is as follows:

$$Living\ Area_i = Building\ Footprint_i \times Number\ of\ Storeys_i$$

Equation 12: Living area estimation

Both the building footprint and number of storeys came from the ROW building footprint layer. This estimate does not include basements. As well, this estimate does not account for space taken up by walls nor inconsistent dimensions throughout the various floors of the property structure.

3.2.2.3 Street Frontage

Street frontage was incorporated as the length of the edge of parcels along the street network, from the MPAC assessment data.

3.2.2.4 Private Yard Size

Yard size was estimated using the property parcel and building footprint layers. The calculation was completed as follows:

$$Yard\ Size_i = Lot\ Size_i - Building\ Footprint_i$$

Equation 13: Yard size estimation

The building footprint term in this equation uses the combined building footprint of residential and accessory structures on the lot. By design, this measurement of *Yard Size* includes paved ground, such as driveways, patios, and walkways as yard space. The yard size variable was used to identify a number of cases where incorrect building footprints led to negative or very small values of *Yard Size*, and these were manually corrected in the ROW's building footprint dataset.

3.2.2.5 Age of Residence

The age of the residence at the time of sale was calculated using data on the year the home was built and the year of the transaction, as follows:

$$Age_i = Sale\ Year_i - Year\ Built_i$$

Equation 14: Age of residence

As the Year Built data was only available at the Year scale, estimates of building age at the time of sale may actually be a year older or newer depending on when within the year the home was built and when it was purchased.

3.2.3 Neighbourhood Variables

This section will discuss variables that capture distinct land-market conditions and socioeconomic qualities of neighbourhoods or zones within the region. These include a rate of property appreciation, an estimate of neighbourhood perceptions of safety, nearby school quality, a CTC boundary dummy, and a municipality dummy.

3.2.3.1 Rate of Appreciation (ROA)

Appreciation rates are an important market characteristic related to homebuyer WTP. A neighbourhood scale ROA measure can be seen to represent some of the forward-thinking and economically rational ways that homebuyers behave within the residential land-market. It is assumed that individuals generally seek to maximize their potential for profit. This work seeks, then, to explicitly model the tendency of individuals to be willing to pay more for something should they perceive that this investment will provide greater returns in the future.

While a homebuyer's decision to invest in a certain property is also based on personal tastes and preferences, as accounted for through other independent variables in this work, this neighbourhood level appreciation rate accounts for the market response of homebuyers to invest wisely. A neighbourhood increasing in value relative to others can be a signal that encourages prospective homeowners to invest and be willing to pay more, following from an expectation that values within that neighbourhood are likely to continue to increase. Likewise, a low home appreciation rate may be a signal to homebuyers that an area may not be a particularly wise investment decision, so long as they expect the market conditions of that area to continue at the same rate.

This appreciation rate uses the adjusted value of homes. Using adjusted value for this calculation means that temporal effects at the regional scale, namely appreciation and inflation, have already been accounted for. Therefore, this appreciation rate represents a measure of home price appreciation of each neighbourhood relative to the region as a whole. A negative neighbourhood level appreciation rate calculated in this way may not mean that values were declining nominally in a certain neighbourhood, but that values were not increasing as quickly as the regional average. Similarly, neighbourhoods that were appreciating at high rates are doing so even after considering the appreciation affecting the regional residential market as a whole.

To calculate the ROA for various neighbourhoods, the dataset was first limited to arms-length sale of single-detached homes. The decision to use only single-detached homes was made to mitigate the impacts of price differences between different property types on the final appreciation measure. As well, observations were dropped where there were fewer than two home sales in a time period. In these ways, only homes that are roughly comparable have been used to calculate neighbourhood level price change.

After limiting the observations in the dataset, a lagged moving average of neighbourhood property value was created for each neighbourhood in each time period as follows:

$$LAV_{it} = \frac{1}{3} \times \left(\frac{\sum V_{it}}{n_{it}} + \frac{\sum V_{it-1}}{n_{it-1}} + \frac{\sum V_{it-2}}{n_{it-2}} \right)$$

Equation 15: Lagged average values

Where,

LAV_{it} is the lagged average value in neighbourhood i at time period t,

$\sum V_{it}$, $\sum V_{it-1}$, *and* $\sum V_{it-2}$ are the sums of transacted values in neighbourhood i at times t, t-1, and t-2, respectfully,

and, n_{it}, n_{it-1}, n_{it-2} are the number of sales in neighbourhood i at times t, t-1, and t-2, respectfully.

These lagged average values were then used to calculate a ROA for each neighbourhood at each time period as per the following equation:

$$RoA_{it} = \frac{LAV_{it} - LAV_{it-1}}{LAV_{it-1}} \times 100$$

Equation 16: ROA

Where,

RoA_{it} is the rate of change in lagged average values between period t-1 to period t in neighbourhood i

The script to calculate this variable is included as Appendix 3, which shows how the time-series *tssmooth* function in Stata was used alongside other functions to create the lagged moving average and ROA for each neighbourhood.

For t, the data was aggregated into spans of two years at a time, which ensured a sufficient number of observations within each time period to calculate a sensible metric. When testing this calculation with individual years as the time period, the appreciation rates were unrealistically volatile from year to year as they were heavily influenced by the sales of few homes with largely dissimilar prices.

Three neighbourhoods using administrative boundaries were tested for this variable: Census Tracts (CTs), TAZs, and the CTC boundary. TAZs were the smallest spatial unit, followed by CTs, and the CTC was the largest. Because neighbourhoods with only one sale in each time period were dropped from the model, the smaller size of TAZs produced the most null values when attached to observations in the dataset. The CTC produced the least null values; however, its large scale led to little variation between observations, which is an undesirable quality of statistical model variables. Ultimately, the CT was decided to be the neighbourhood definition that provided the most reasonable compromise between ensuring high variation and mitigating null values for certain neighbourhoods and time periods.

3.2.3.2 Perception of Safety

Police phone call data were used as a proxy to represent neighbourhood-level perceptions of safety. This aligns with the use of the data in the ROW's CTC monitoring program (2015c). This measure uses only public-facing calls—those regarding incidents that could inform neighbourhood perceptions of safety—which included calls regarding potential violence, theft, and public disorder in 2012. These were calls about potential criminal activities that resulted in police notification, visible to or directed toward the public.

Originally, a measure of calls within a certain radius around each property was considered. However, after visually inspecting the spatial distributions of the data, calls were found to be distributed more heavily in denser areas. Instead, a per capita measure of police call incidences is able to account better for the variation in call volumes that is a product of population size. Population data at the DA scale was available from the NHS, and used to create the per capita measure (Statistics Canada, 2011b). The variable was ultimately calculated as follows:

$$Perception\ of\ Safety_i = \frac{Police\ Calls_i}{(Population_i\ /\ 100)}$$

Equation 17: Perception of Safety

Where population was divided by 100 to mitigate very small, fractional values in the final variable.

There was a technical challenge in allocating the police calls to each DA. The location of the police phone call data was recorded as the nearest intersection to where the call was made. Many of these intersections happen to be located on the borders of contiguous population data polygons, between DAs. Using the *Near* tool in ArcGIS, as suggested by Honeycutt (2013), proved the most appropriate solution for assigning the police calls to zones, neither omitting or double counting the cases where calls were located on the border of zones.

This tool randomly assigns points that are equidistant from polygons, which is the case for points located on polygon borders, to a single polygon. While this randomness could be seen as potentially problematic, the measurement error associated with this independent variable would be randomly distributed and should not substantially affect the validity or interpretation of regression estimates.

3.2.3.3 School Quality

The *school quality* variable was created using two data sources: a layer of school catchment areas (Waterloo Region District Schoolboard, 2015), and school quality rankings from the Fraser Institute (Fraser Institute, 2014). The Fraser Institute rankings for all available public elementary schools in KW were copied into an Excel spreadsheet. Identifiers were then manually added to the school ratings corresponding to the identifiers in the Waterloo District School Board polygons, and the two were joined in ArcGIS. This research limited the rankings used to public elementary schools, to avoid the overlap between elementary and high schools. Elementary school catchments were more spatially disaggregate than secondary school catchments. In instances where elementary school catchments overlapped, because of divided junior and senior elementary schools, the junior elementary school districts were used. The junior elementary school catchments were generally smaller than the senior school catchments, and thus provided greater variation in the rankings.

3.2.3.4 Education Rate

An education rate was calculated for each neighbourhood. Education is a commonly used neighbourhood characteristic in hedonic models, as an indicator of neighbourhood level socioeconomic status. This neighbourhood characteristic is intended to represent the impact of socioeconomic status using CT level data from Statistics Canada (2011a). The education rate was calculated as the following:

$$Education\ Rate_i = \frac{Population\ with\ postsecondary_i}{Population_i} \times 100$$

Equation 18: Education Rate

Where education rate in neighbourhood i is equal to the proportion of the population that has undergone post-secondary education in that neighbourhood. The data used for this variable were obtained from the 2011 NHS (Statistics Canada, 2011a).

3.2.3.5 Population Density

Population density was calculated to account for the potential variation in property values associated with more or fewer people living in close proximity to one another, a natural by-product of residential intensification. Much literature supports the notion that individuals prefer to live in less densely populated environments; however, there is also support of positive social interaction amenity effects of more densely populated areas.

This variable was calculated from NHS data at the DA (DA) scale, as follows:

$$Population\ Density_i = \frac{Population_i}{Area_i}$$

Equation 19: Population Density

Where the population density in each neighbourhood i is equal to the population in each neighbourhood divided by their respective areas. In this work, the unit of measurement used for the area of each DA was square kilometres.

3.2.3.6 In CTC

A binary variable representing whether a property is located in the regional CTC was specified as follows:

$$InCTC_i = \begin{cases} 1, & contained\ in\ CTC\ boundary \\ 0, & outside\ of\ CTC \end{cases}$$

Equation 20: In CTC

This variable was used to identify differences in residential preferences between properties located within the CTC and those outside of it. As has been noted in the literature and outlined in Chapter 2, transit stations areas that have received specific, targeted plans for development to accompany the transit implementation see the highest potential to realize the value uplift associated with increased transit accessibility. The use of this variable is intended to control for the designated corridor that is slated to receive increases to accessibility from LRT, but also, and perhaps more importantly, the area for which plans are in place to coordinate the LRT with intensification and economic growth objectives.

3.2.3.7 Time Period

Time periods marking significant points in the approval of the LRT were selected to identify whether the decisions to develop LRT are related to increases in home values. These variables were specified as follows:

$$Time\ Period\ 2_i = \begin{cases} 1, & Sale\ occured\ between\ July\ 2009 - July\ 2011 \\ 0, & Otherwise \end{cases}$$
$$Time\ Period\ 3_i = \begin{cases} 1, & Sale\ occured\ after\ August\ \ 2011 \\ 0, & Otherwise \end{cases}$$

Equation 21: Time Periods

Where the base case is homes sold before any approvals of LRT occurred, before 2009. In July 2009, The ROW first approved regional LRT. In July 2011, council cemented their approval of LRT as the preferred mode alongside an implementation strategy, which has been followed by the carrying out of this implementation strategy (ROW, 2012).

3.2.4 Homebuyer Characteristics

As explained in Chapter 2, it is well understood that homebuyers exhibit heterogeneous residential location preferences. This section describes the various homebuyer characteristics that were expected to affect homebuyer WTP for intensification-related property characteristics. The subsections that follow discuss the demographic variables investigated in this research and how variables representing homebuyers' stated preferences for various locational characteristics were applied in this work.

Data on homebuyer characteristics was only available for a small sample of residential transactions. This data comes from a survey conducted by Emma DeFields on property size and outdoor space preferences (2013). For this reason, separate models incorporating homebuyer data were estimated using a limited set of independent variables, as will be described in Chapter 5.

3.2.4.1 Demographics

The intent of individual level demographic variables is to analyze the moderating impacts of demographic heterogeneity on homebuyers' WTP for residential characteristics. The demographic variables investigated in this work are homebuyer age, income, education level, and household composition.

These demographic characteristics were specified in the data as follows:

Household composition

This work includes only one dummy variable for household composition, whether a household includes children or not:

$$Children_i = \begin{cases} 1, & household\ contains\ children \\ 0, & otherwise \end{cases}$$

Equation 22: Household composition

Other household types were not investigated due to the small sample size. However, whether a household includes children has been noted as one of the largest demographic determinants of heterogeneity in residential preferences (Antoniou & Picard, 2015).

Income

Income data from the survey was collected in ranges. To create a continuous measure, the average of each range was attributed to that observation, as in the following:

$$Income_i = \frac{upper\ bound + lower\ bound}{2}$$

Equation 23: Income

Education

Education was input into the models as a continuous variable, which was the total number of years of education. The total number of years of education was calculated as the sum of years completed from grade one to thirteen plus years spent in university, college, vocational, or technical school.

3.2.4.2 Stated Preferences

In addition to demographics, the survey from Emma DeFields (2013) also obtained information about individuals' preferences for various property and neighbourhood characteristics [See Chapter 2 for a discussion of stated preference methods for environmental valuation]. These were included in this modelling work to identify whether homebuyers' stated preferences were consistent with their revealed preferences for various property characteristics.

Preferences for different neighbourhood amenities were obtained on a Likert-type scale. Specifically, the survey asked individuals to rank how important various factors were in their decision to move to their current neighbourhood. This scale ranges from one to five, where one, three, and five represent 'not at all important', 'somewhat important', and 'very important', respectively. The ratings used for this work include the importance of homes' 'access to transit' and 'close[ness] to parks or recreational opportunities' in the homebuyers' decision to move to their current neighbourhood.

These values were input into the model as a set of dummy variables, as follows:

$$Important_i = \begin{cases} 1, & \text{"somewhat important" and higher} \\ 0, & otherwise \end{cases}$$

Equation 24: Stated preferences

In this way, those who reported that a characteristics was less than 'somewhat important' are the base case. The *Important* variable represents those who stated that access to those amenities was at least somewhat important in their decision to live in their current neighbourhood.

3.3 Comparing Variables: univariate and bivariate analyses

Ensuring a well-performing model requires careful consideration of model specification. Omitted variable bias, multicollinearity, and misspecification are large concerns in the development of econometric models of any sort. As described in Chapter 2, these problems result from under-specification, over-specification, or otherwise erroneous variable choice in models.

Correlation matrices, scatter plots, descriptive statistics, and other univariate and bivariate analyses were used to compare different variables and combinations of variables to ensure an appropriate model specification. The following chapter will provide descriptive statistics necessary to understand the distributions of the variables presented in this current chapter. Preliminary models were tested to identify unexpected estimates that might be indicative of multicollinearity or omitted variable bias. The results of this analysis are used to justify the final model specifications presented in Chapter 5.

4 Pre-Regression Results

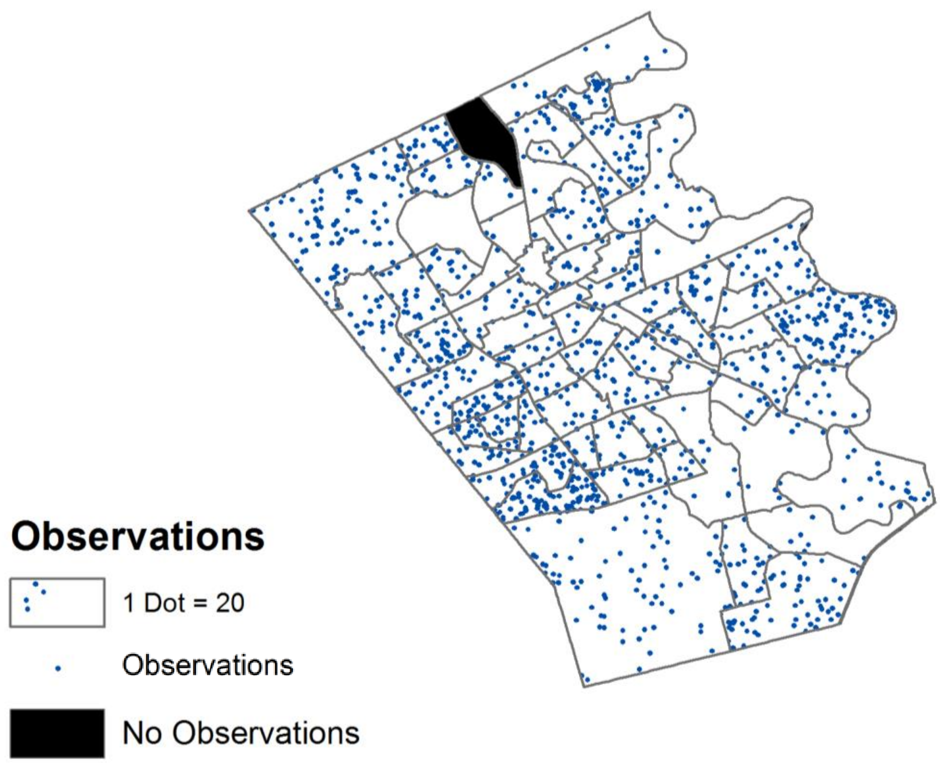
This chapter contains the results of the work described in the previous chapter, Pre-regression Methods. The research findings based on these results that are necessary to inform the empirical model specification in Chapter 5, while overall research findings are discussed in Chapter 7.

The intent of this section is to outline the distributions of and relationships between the candidate dependent and independent model variables of this research. First, the sample of home sales analyzed will be briefly described. Then, the descriptive statistics for the dependent variables will be presented. Following that, a long section will outline the results for each independent variable in the research, which includes maps showing the variables' spatial distributions and tables of descriptive statistics for the observations. The last section in this chapter will outline bivariate relationships of the data through correlation tables. Finally, this chapter will conclude with a brief comparison of home sales by homebuyer demographics and stated preferences using the available survey data.

4.1 Analysis of model variables

After limiting the data, the dataset was reduced to 26,873 arms-length home sales between January 2005 and February 2015.

Map 2 shows the spatial distribution of the final sample:



Map 2: Dot Density of Sales by CT

This map displays the number of observations in each CT, represented by the number of dots inside the tract. Because of a low number of observations in some tracts, with one as low as only including 5 observations, the exact locations of observations are not shown to protect data confidentiality. As can be seen in the map, sales were more densely distributed in CTs along the centre of the western and eastern borders. Sales were most sparse in the southernmost CTs as well as the CT containing the University of Waterloo in the northwest. The remainder of CTs show a fairly consistent dispersion of home sales.

Figure 5 shows the number of sales for each year. This number is gradually increasing, however it must be reiterated that homes sold multiple times in this period are only represented in the year of their latest sale. The transactions dataset includes sales up until the beginning of February 2015.

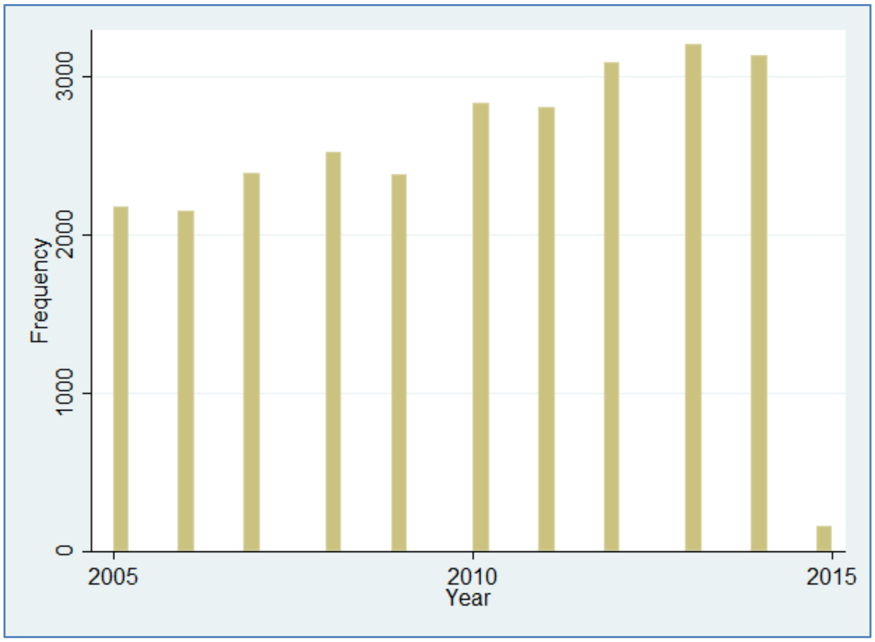


Figure 5: Observations by year

The following, Table 2, shows the distribution of observations by property type and whether they are located within the CTC:

K-W Home Sales, 2005-2016	Single-detached	Semi-detached & Duplexes	Townhouses	Total
In CTC	1,639	438	147	2,224
	73.7%	19.69%	6.61%	8.28%
Outside of CTC	20,162	2,264	2,223	24,649
	81.8%	9.19%	9.01%	91.72%
Total	21,801	2,702	2,370	26,873
	81.13%	10.06%	8.82%	100%

Table 2: Contingency table of observations by In CTC and Property Type

In this table, the percentages of each property type as a proportion of the total number of properties within and outside of the CTC are shown below their respective frequency in the sample. Overall there are 2,224 observations within the CTC, 8.28% of the total sample of 26,872. In KW, the CTC and the rest of the city shared the same hierarchy of sales by housing type, where detached dwellings were the most common, followed by semi-detached and duplex dwellings, then townhouses. However, there are notable differences between the proportions of each property type sold in the CTC and outside of it. Proportionally, there were over twice as many semi-detached and duplex sales within the CTC. Alternatively, outside of the CTC saw 8% more sales of single-detached homes and 2.4% more sales of townhouses by proportion.

4.1.1 Dependent Variables

This section will describe the distributions of assessed value and adjusted value. The following, Figure 6, shows the relationship between the two.

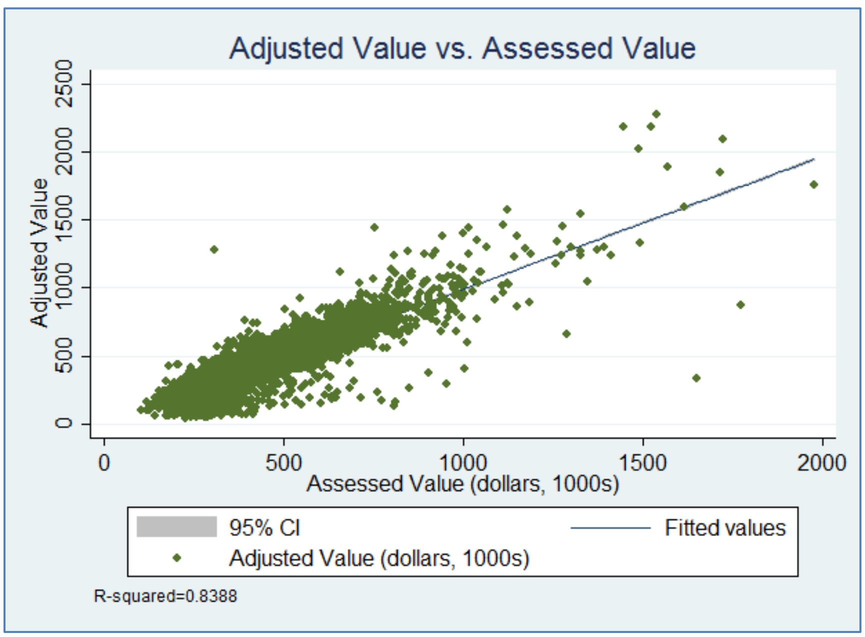
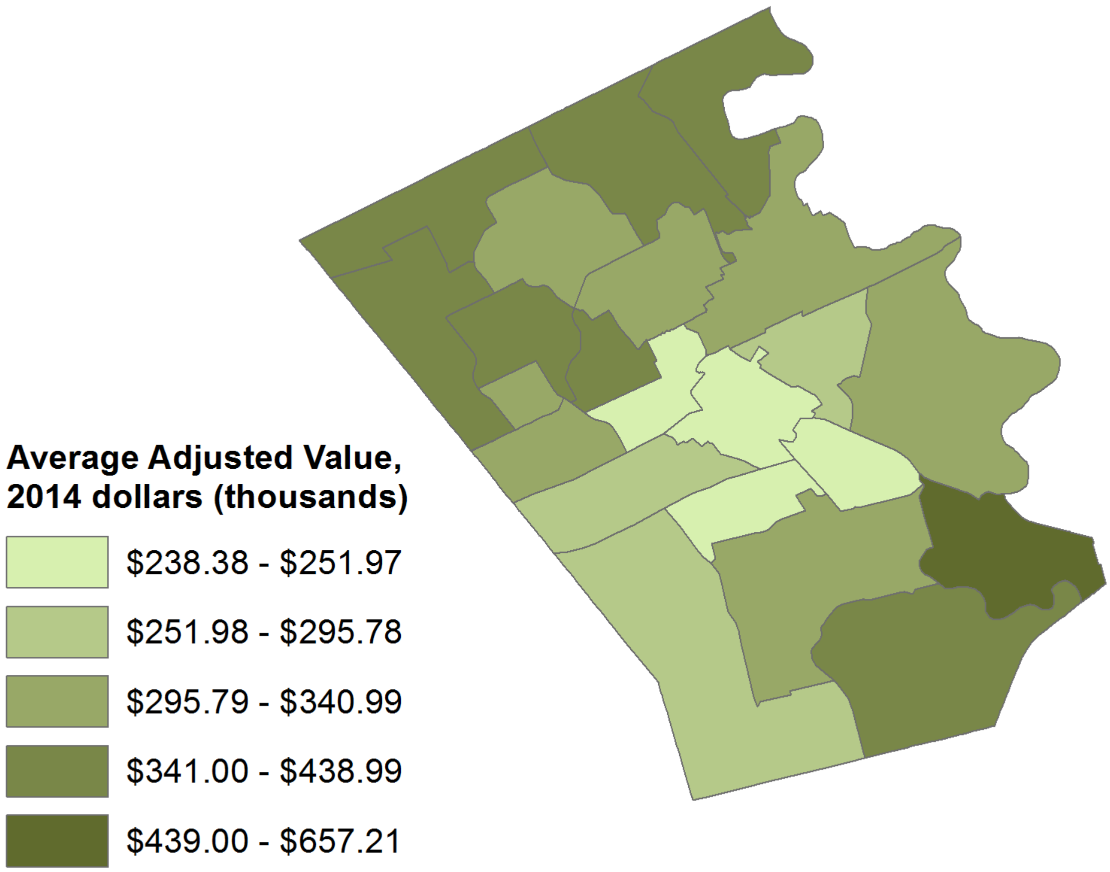


Figure 6: Adjusted and Assessed Value Scatterplot

Assessed Value is very closely related to *Adjusted Value*, describing 83% of the variance according to the OLS R^2 presented in the bottom left of the figure. There are a few outliers in the dataset in terms of assessed versus adjusted value, however these did not meet any of the criteria for excluding non-market sales in Chapter 4 and were included in the sample. An R^2 of 0.839 is a good indicator that the transacted data was adequately preprocessed and that assessments are reflective of observed sales values.

4.1.1.1 *Adjusted Value*

The following shows the average *Adjusted Value* of homes sold between 2005 and 2014, by 'planning neighbourhood' in KW. These 'planning neighbourhoods' are a spatial unit created by the ROW for analytical purposes, meant to represent relatively homogeneous neighbourhoods within the region.



Map 3: Average Adjusted Value

On average, properties in the centre of KW sold for less than those farther from the centre. The most expensive property purchases were located in the southeast of Kitchener and in the northern border of waterloo, up to nearly three times as expensive as homes in the centre, on average.

Table 3 shows the distribution of this variable in the regression dataset:

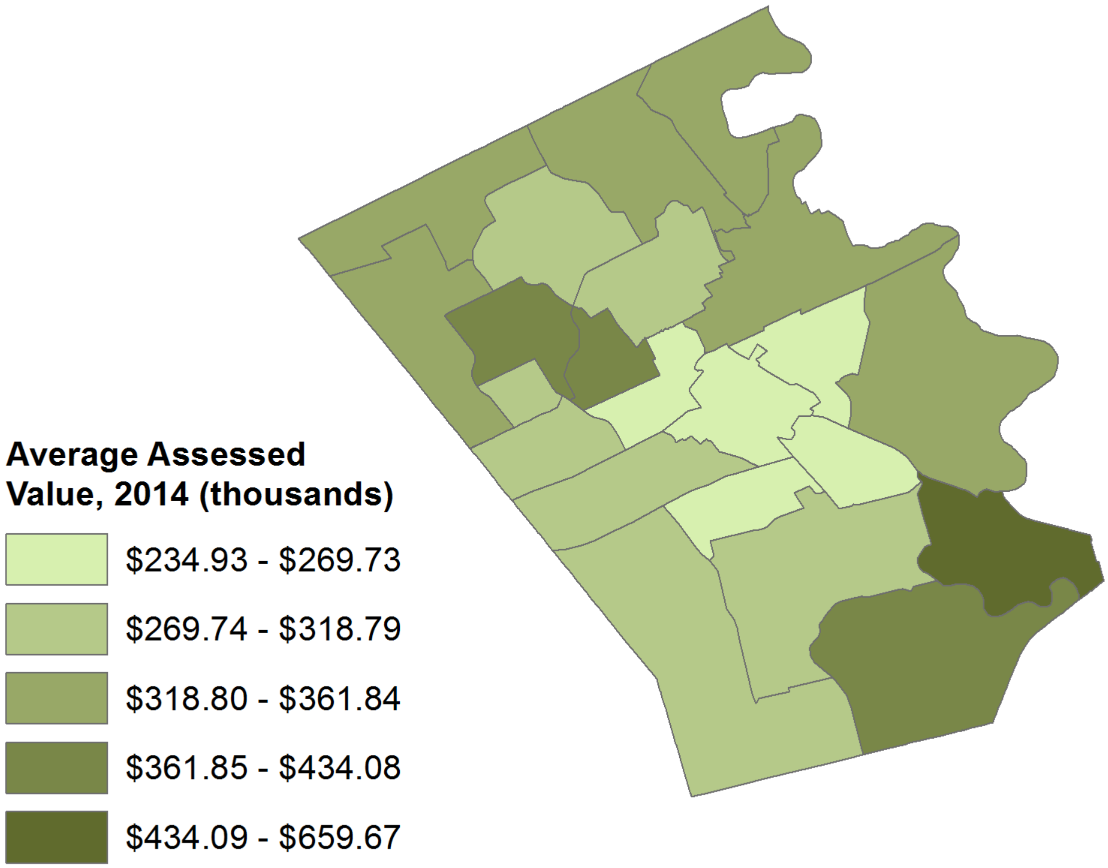
Adjusted Value (2014 Dollars)		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	$332,706	$281,816	$337,297	$351,199	$244,634	$262,968
Std. deviation	$120,743	$99,879	$121,408	$124,465	$51,380	$54,239
Minimum	$42,461	$47,800	$42,461	$42,461	$52,961	$70,659
Maximum	$2,274,316	$1,024,835	$2,274,316	$2,274,316	$638,273	$636,323

Table 3: Adjusted Value: descriptive statistics

Properties sold in KW between 2005-2007 had an average adjusted value of $332,706, in January 2014 dollars. Homes sold inside the CTC were just under $50,000 less than homes outside of it, on average, with more consistent prices. The semi-detached and duplex and the townhouse categories had similar average values and similarly low standard deviations. Single-detached homes were roughly $100,000 higher than semis, duplexes, and townhouses on average and were much less consistent.

4.1.1.2 Assessed Value

The following shows the average *Assessed Value* of homes sold between 2005 and 2014, by 'planning neighbourhood' in KW.



Map 4: Average Assessed Value

This map shows that assessed values share a similar pattern as the sales values; however, the highest assessed values are only in the southeast and two neighbourhoods in the Northwest, rather than along the entire Northern border of Waterloo as was seen in the Adjusted Value map.

Table 4 shows the distribution of this variable in the regression dataset

Assessed Value (2014)		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	$330,026	$275,180	$334,975	$349,184	$240,577	$255,742
Std. deviation	$112,790	$82,908	$113,814	$115,100	$45,453	$44,863
Minimum	$104,250	$104,250	$140,000	$129,750	$119,500	$104,250
Maximum	$1,720,250	$1,005,000	$1,720,250	$1,720,250	$616,250	$615,250

Table 4: Assessed Value: descriptive statistics

The *Assessed Values* of homes are very similar to the *Adjusted Values*. Like adjusted value, assessed values are higher outside of the CTC, and follow the same distribution among property types, where duplexes and townhouses were similar and the lowest value, and single-detached homes were roughly $100,000 higher.

4.1.2 Independent variables

This section contains the results produced by the pre-regression methods outlined in Chapter 4. Summary statistics and data visualizations will be presented and described for each model variable individually in the following sub-sections, while results comparing multiple variables will be presented at the end of this section.

All tables of summary statistics presented, for consistency, are of the sample of home transactions that was limited to only market-rate housing sold between 2000-2015. Most maps are presented for the entirety of KW, to highlight their overall spatial distributions; however some maps were created from the transaction data itself, and present only the average value of each variable in the transaction data by CT. Maps that were created from the transaction data will be identified as such when they appear.

The independent variables for this work are categorized into Environmental, Neighbourhood, and Structural characteristics, consistent with the general regression model specification, Equation 1. The following sections will describe these variables in detail.

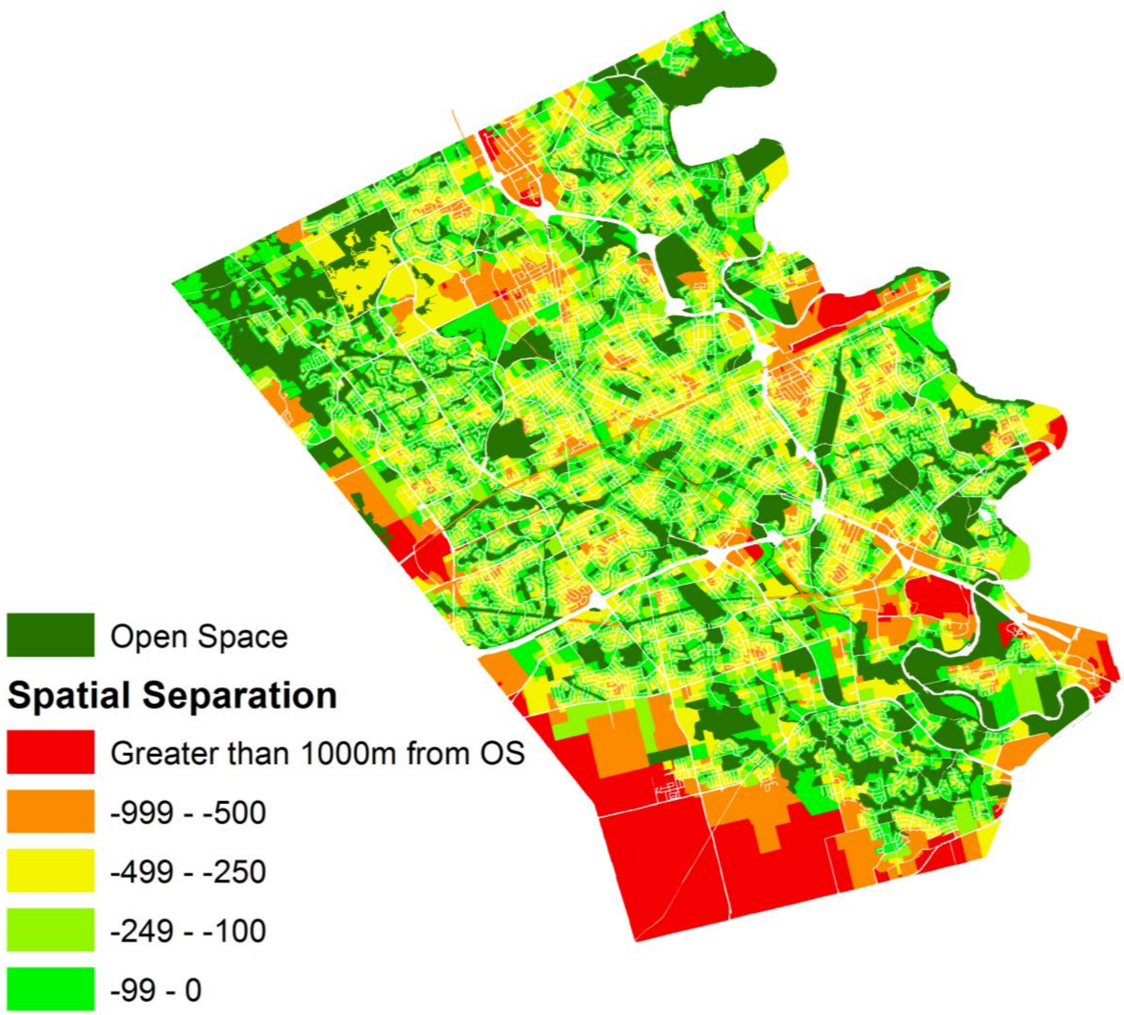
4.1.2.1 Environmental Variables

4.1.2.1.1 OS Access

Here, the results of the different OS access models (spatial separation, cumulative opportunities, and gravity based) are described. Comparisons of the candidate *access to OS* variables are left for later in this chapter.

Spatial Separation

The following, MAP 5, shows the result of the spatial separation *access to OS* access measure across KW.



Map 5: OS, Spatial Separation

As can be seen from Map 5, the spatial separation model produces equally distributed gradients of access surrounding each OS, irrespective of the OSs' sizes or the coincidence of multiple proximate OSs. There are a few locations within the two cities that are farther than 1000m from OS, and will by design contain a value of 0 in all of the following OS access measures; however for the descriptive statistics given below these properties were labelled as 1,200m from transit. These locations farther than 1000m along the pedestrian transport network to a public OS are indicated in red.

Table 5 shows the distribution of this variable in the regression dataset:

OS: Spatial Separation (m)		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	193.1	216.3	191.0	195.2	206.2	158.2
Std. deviation	148.9	155.2	148.2	151.3	143.4	126.9
Minimum	10	10	10	10	10	10
Maximum	1,200.00	1,200.00	1,200.00	1,200.00	850.8	598.9

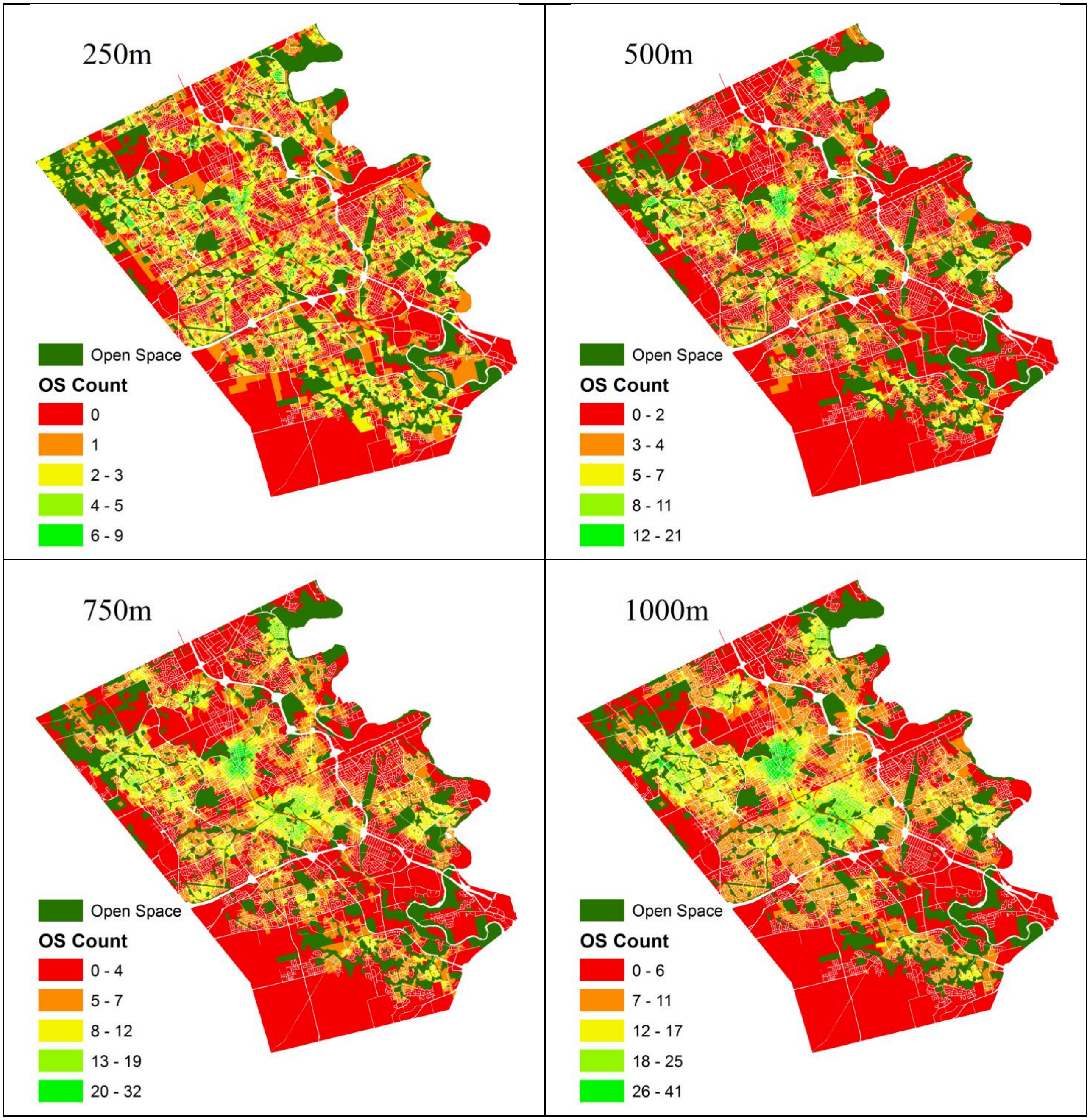
Table 5: OS, Spatial Separation Model: descriptive statistics

On average, homes sold outside of the CTC were roughly 25 metres closer to an OS than those within it. Townhouses were generally closest to OS, at an average of 158m away. Single-detached homes and semi-detached and duplexes were both under 50 metres farther from an OS than townhouses, at 195m and 206m respectively.

Cumulative Opportunities Model

Two specifications of cumulative opportunities models were run to assess *OS Access*. The first, using a count of nearby opportunities is presented in Map 6 and Table 6. The second, using the total area of nearby opportunities, is presented in Map 7 and Table 7.

Cumulative Opportunities: Count of opportunities



Map 6: OS, Cumulative Opportunities Model using count of accessible OS

As can be seen in Map 6, the result of the cumulative opportunity model using a count of nearby OSs generates a decreasingly fragmented spatial pattern as the binary threshold is increased. The first map of the series, at 250m shows that access at this level is highly decentralized. At a 250m threshold, access to many OSs is scattered in a salt-and-pepper-like pattern across the landscape, although a few area are able to be identified with contiguous high levels of access like in Uptown Waterloo. At higher levels of the distance threshold, two main distinct areas emerge with the highest concentration of OSs within a 1000m threshold. The two central hot spots of OS access that result from this model are Uptown Waterloo and Downtown Kitchener, in the central area of the map. Two smaller hotspots are located in the northwestern and northeastern areas.

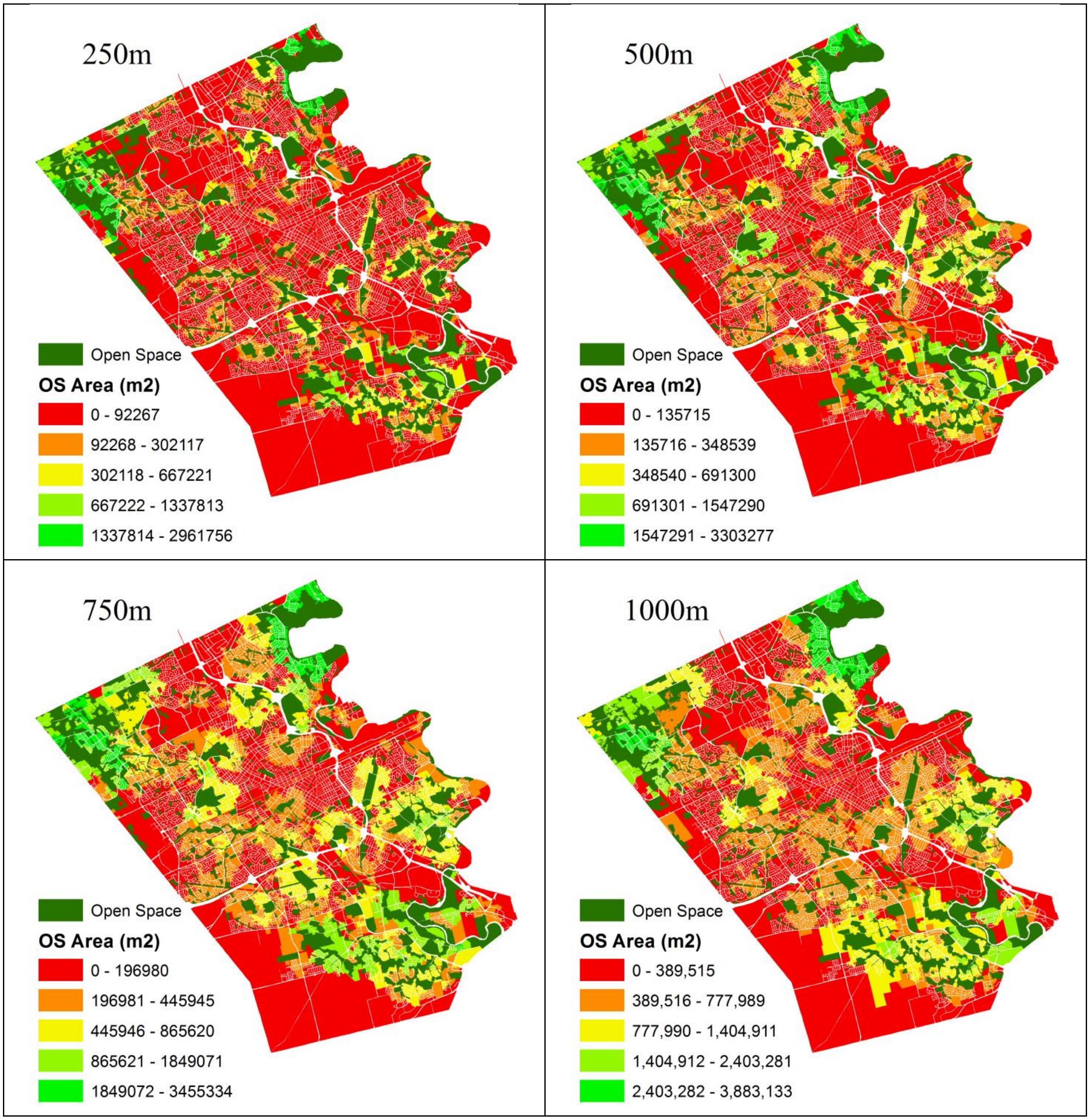
Table 6 shows the distribution of this variable using a 1000m network distance buffer in the regression dataset:

OS: Count in 1000m		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	7.8	9.3	7.7	7.7	8.4	7.8
Std. deviation	3.9	5.6	3.6	3.8	4.0	4.0
Minimum	0.0	0.0	0.0	0.0	1.0	1.0
Maximum	29.0	29.0	29.0	29.0	27.0	29.0

Table 6: OS, Cumulative Opportunities Count: descriptive statistics

On average, homes sold in the CTC have access to 1.6 more OSs than those outside of it; however, there was more variation in the number of accessible OSs in CTC than within any other category presented. Single-detached homes and townhouses had the lowest number of OSs accessible, at 7.7 and 7.8 respectively. Semi-detached and duplexes were on average able to access just over 0.5 more OSs within 1000m

Cumulative Opportunities: Sum of Area



Map 7: OS, Cumulative Opportunities Model using area of accessible OS

In the second cumulative opportunities model, different binary thresholds were tested on the sum of area of OS accessible from each property. This shows where residents are able to access the largest amount of OS, in terms of the OSs' areas in m^2. At very small levels of the binary threshold, only those few properties very near to large OSs receive relatively high values. As the binary threshold is increased, three main hotspots of OS access emerge. The most concentrated hotspots are located in the northwest in the northeast in Waterloo, while the southern and southeastern portion of Kitchener sees more deconcentrated high levels of access.

Table 7 shows the distribution of this variable in the regression dataset:

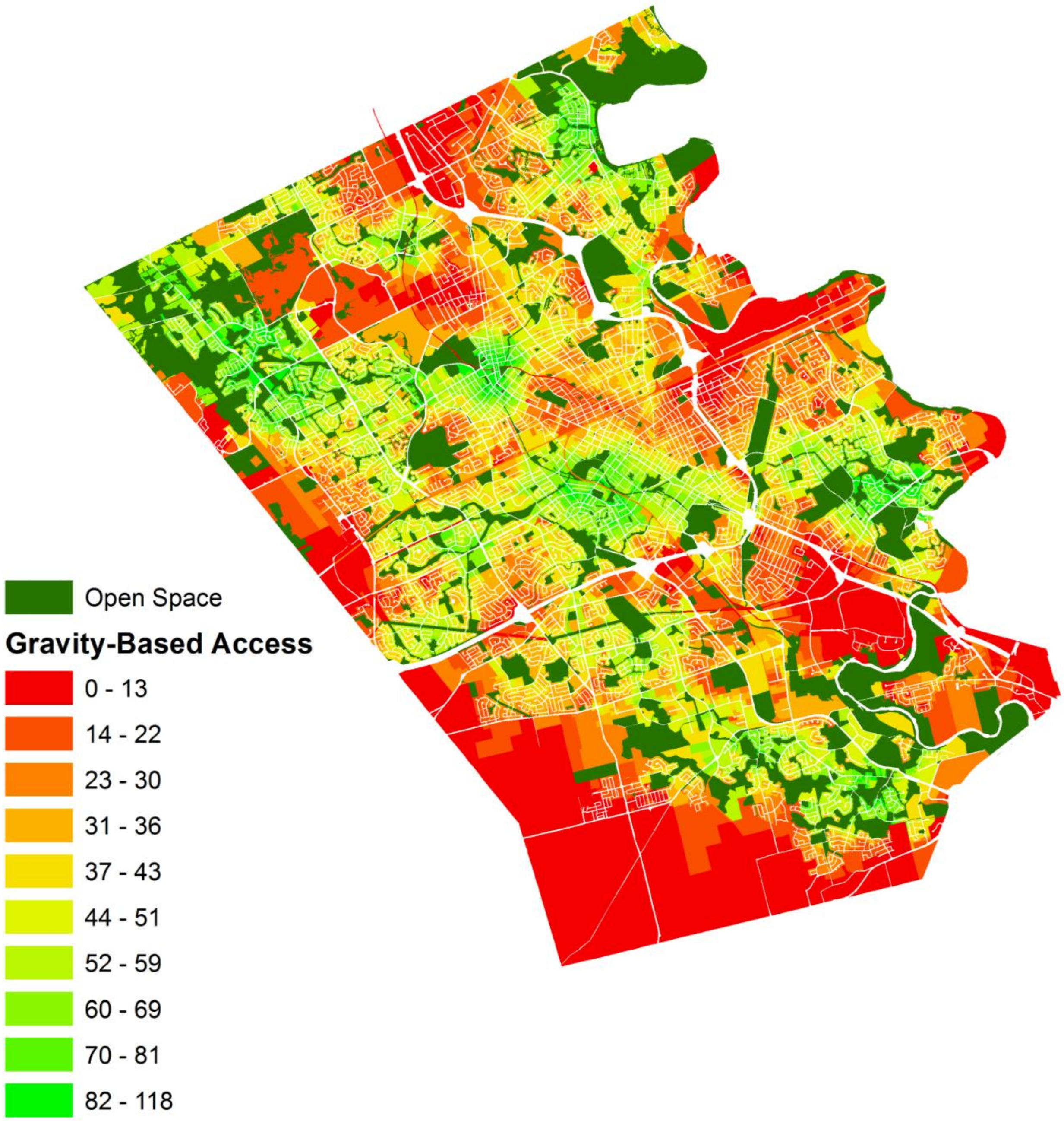
OS: Area in 1000m (m^2)		**CTC**		**Property Type**		
	Full Sample	**Inside**	**Outside**	**Single Detached**	**Semi & Duplex**	**Townhouse**
Mean	760,834	342,519	798,579	756,684	659,216	914,923
Std. deviation	757,615	225,779	777,150	745,474	709,139	887,626
Minimum	0	0	0	0	1,660	80,046
Maximum	3,883,133	1,667,638	3,883,133	3,883,133	3,562,971	3,641,020

Table 7 – OS, Cumulative Opportunities Area: descriptive statistics

Homes sold in the CTC were able to access less than half of the amount of OS in 1000m than the rest of the properties sold in the cities on average; however, the rest of KW had more variation in the amount of OS accessible. Townhouses had access to the greatest amount of OS overall, followed by single-detached homes, then semi-detached and duplexes.

Gravity-Based Model

The following, Map 8, shows the levels of access produced by the gravity-based OS access model, as specified in the previous chapter:



Map 8 – OS, Gravity-Based Model

The gravity-based model produces accessibility levels considering both the area of OSs in 1000m along the pedestrian network as well as the distance needed to travel to reach them. The highest levels of access from this measure are found in the northwest and northeast neighbourhoods, Uptown Waterloo and Downtown Kitchener in the centre, in the southern area of Kitchener, and just south of the centre of the eastern border of KW.

Table 8 shows the distribution of this variable in the regression dataset:

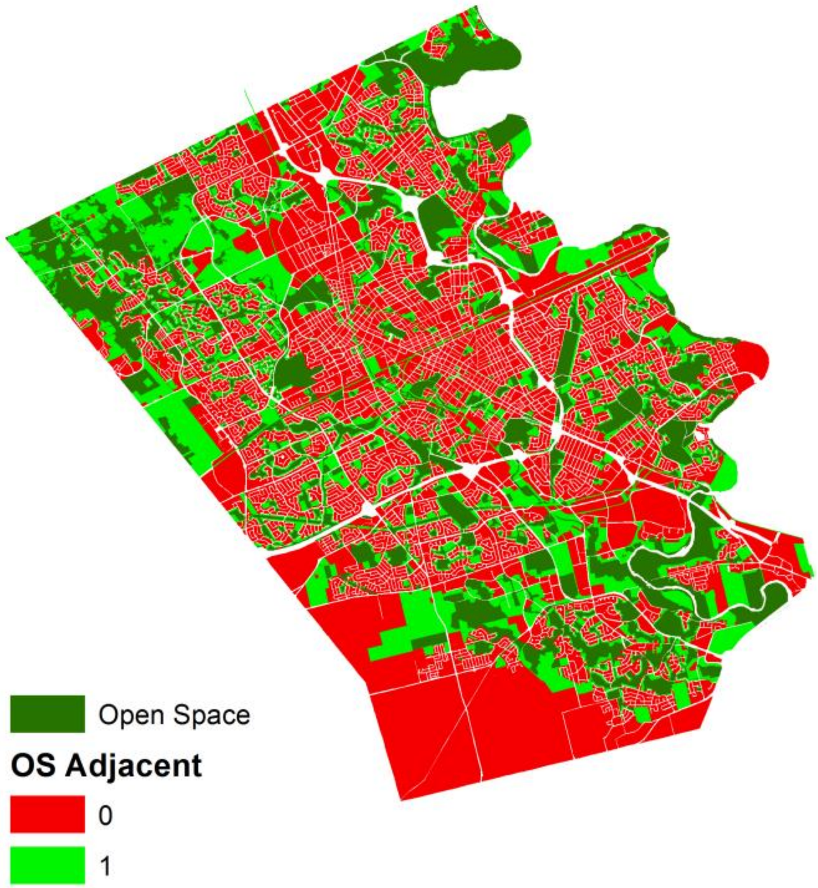
OS: Gravity Based		CTC		Property Type		
	Full Sample	**Inside**	**Outside**	**Single Detached**	**Semi & Duplex**	**Townhouse**
Mean	41.1	37.6	41.4	40.9	40.5	43.6
Std. deviation	17.7	18.4	17.6	17.7	17.2	18.1
Minimum	0.0	0.0	0.0	0.0	2.3	10.1
Maximum	117.9	95.4	117.9	117.9	107.7	99.2

Table 8: OS, Gravity-Based Model: descriptive statistics

While it is evident from Map 8 that OS access via the gravity based model is not constant across KW, the descriptive statistics indicate that the variation in OSs is fairly uniform both between the CTC and the rest of the city as well as across various property types. Homes sold within the CTC had an average *access to OS* value nearly four points lower than outside of it by this measure. Townhouses sold in KW hold the highest average level of access from the gravity-based model.

4.1.2.1.2 OS Adjacency

Map 9 shows the parcels within the region that share a border with OS:

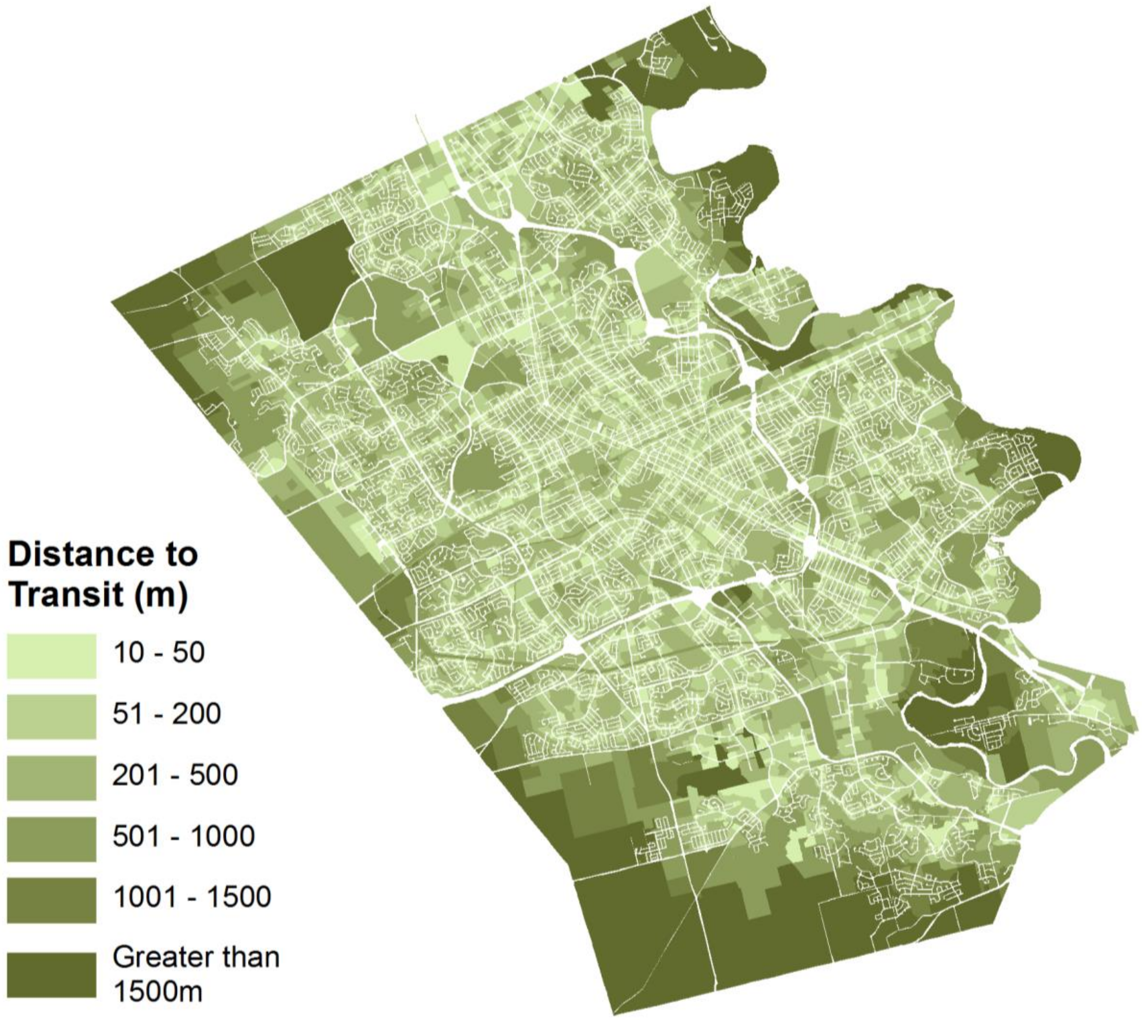


Map 9: OS Adjacency

In the northwest, southeast and northeast, large patches of land are shown to be adjacent to OS. This is the product of both the generally larger parcel sizes as well as the more expansive OSs, thereby increasing the opportunity for OS and property edges to connect.

4.1.2.1.3 Transit Access

Transit Access as computed by a spatial separation model and network distances is shown in Map 10:



Map 10: Transit Access, Spatial-Separation Model

The lowest distances to transit are found along major roads, mirroring the GRT bus routes within KW. The central area of the city contains the highest proportion of stops, particularly in downtown Kitchener. From Charles Street Terminal, in the centre of the map, levels of access according to this measure initially follow a loosely radial pattern outward that breaks down with distance from the hub. Outside of the core, transit stops are more dispersed as they follow the curvilinear road network. Since this model uses network distances between transit stops and properties, the disconnected nature of suburban the street pattern in some areas of the suburbs has led to lower values where intersection density is lower. Large distances between major roads lead to gaps where homes can be found more than 750m from a GRT stop. Very few areas of the cities are more than 1500m from a bus stop, which are the darkest shade on the map. These areas farther than 1500m from transit fall generally at the outer edges of KW as well as along the Grand River that runs within and along the eastern edge of the cities.

Table 9 shows the distribution of this variable in the regression dataset:

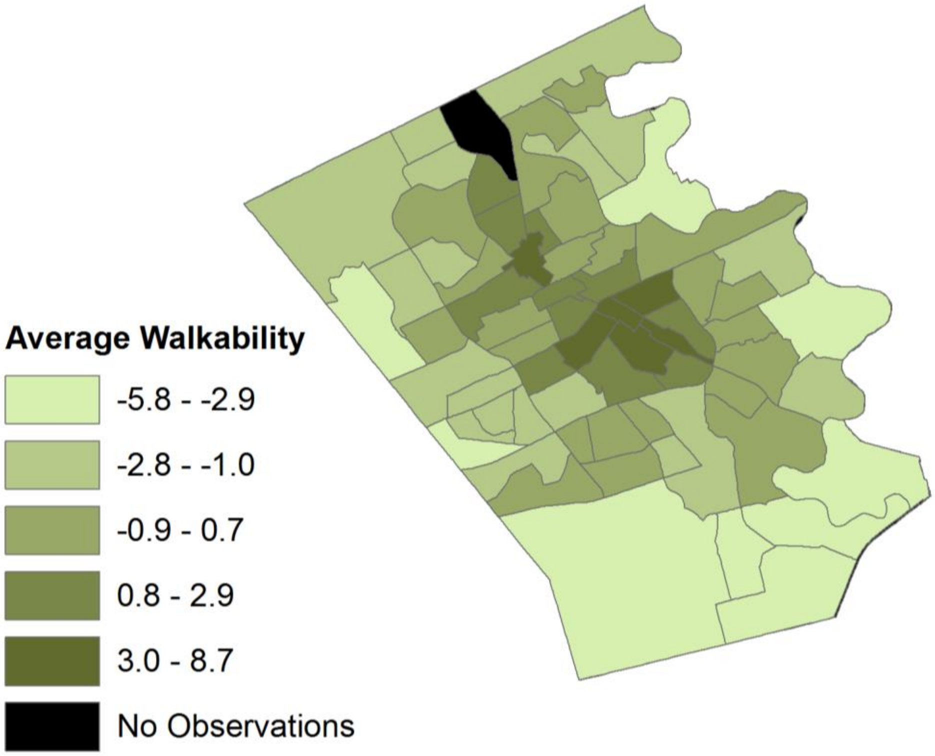
Transit Access × - 1		**CTC**		**Property Type**		
	Full Sample	**Inside**	**Outside**	**Single Detached**	**Semi & Duplex**	**Townhouse**
Mean	-364	-226	-377	-386	-213	-342
Std. deviation	338	192	346	353	175	294
Minimum	-1600	-1151	-1600	-1600	-1196	-1600
Maximum	-10	-10	-10	-10	-10	-10

Table 9: Transit Access, descriptive statistics

When interpreting this table, it should be noted that properties at greater distances than 1500m from transit stops were truncated to 1600m, and as such the estimates for average distances are lower than their true values in categories that contain minimums of 1600. Home sales within the CTC were approximately 150 metres closer to transit stops than those homes sold outside of it, and no home in the CTC was farther than 1151m from a transit stop. There were no semi-detached or duplex homes farther than 1196m from transit. Single-detached homes were alternatively the farthest, at 386m from transit stops on average, but with high variation.

4.1.2.1.4 Walkability

Map 11 shows the average walkability of homes sold between 2005-2015 in KW as an average in each CT:



Map 11: Walkability

As is clear from this map, average walkability of homes sold was highest in the core of KW, with the highest levels in Uptown Waterloo and Downtown Kitchener. Walkability tapers off with distance from the two urban cores, with the lowest levels found along the entire stretch of the southern border of Kitchener as well in four individual CTs on the eastern and western borders of the cities.

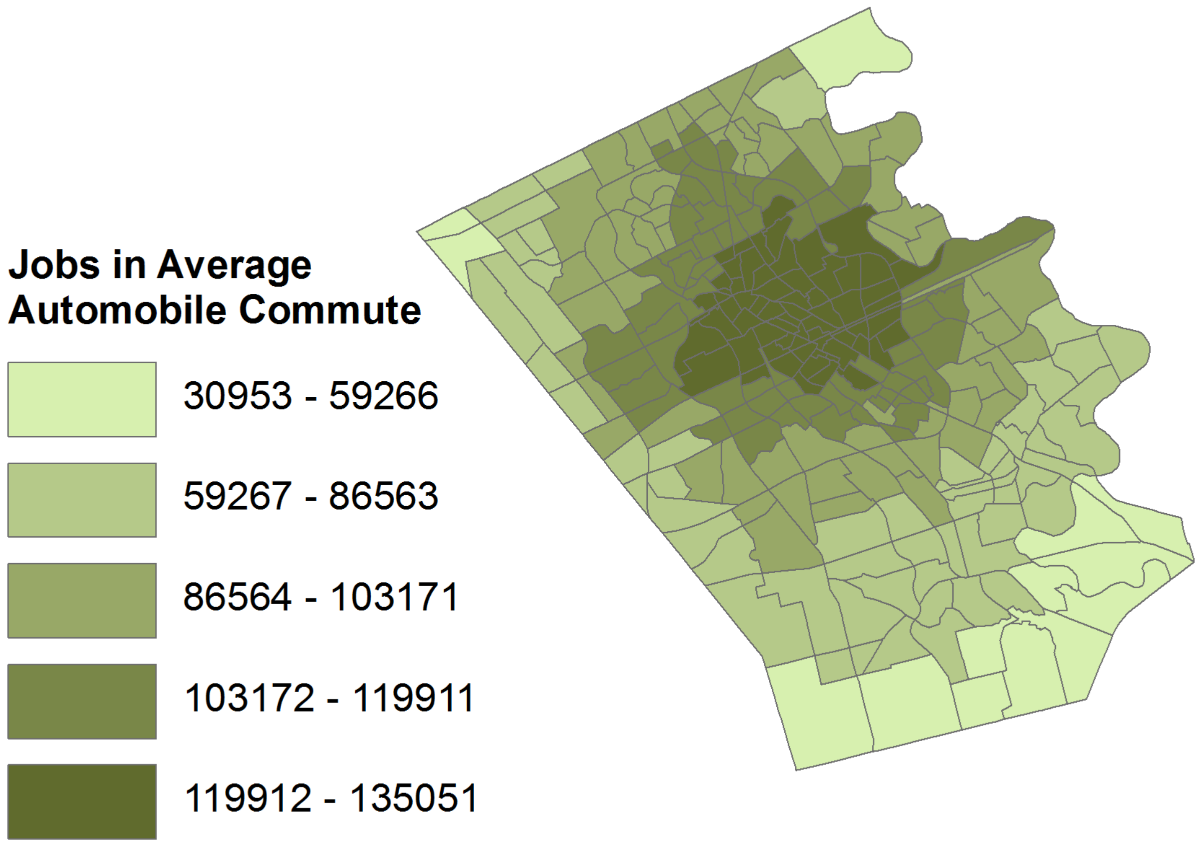
Walkability		**CTC**		**Property Type**		
	Full Sample	**Inside**	**Outside**	**Single Detached**	**Semi & Duplex**	**Townhouse**
Mean	-1.6	2.5	-2.0	-1.7	-0.5	-1.9
Std. deviation	2.6	3.0	2.2	2.6	2.4	2.6
Minimum	-8.9	-3.6	-8.9	-8.9	-7.3	-8.9
Maximum	10.6	10.6	4.0	10.6	10.6	10.4

Table 10: Walkability: descriptive statistics

The average walkability among sales was negative. However, average walkability was much higher, and positive, in the CTC than outside of it. With a total range of walkability from -8.88 to 10.89 in the data, no homes with a walkability value greater than 4 were located outside of the CTC. Townhouses sold were located in the least walkable locations, on average. Semi-detached and duplex sales were in the most walkable areas of the included property types, with an average walkability of -0.5.

4.1.2.1.5 Employment Access

Map 12 shows the number of jobs that can be reached within KW the generalized cost of the average automobile commute from each TAZ in KW:



Map 12: Employment Access (computed using data from Data Source: (Neudorf, 2014).

As can be seen in the employment access map, the highest access to observations via the average automobile commute time and distance within KW is in the centre of the two cities. Job access declines gradually moving away from the centre, producing a concentric gradient. The access gradient becomes slightly pointed in the north, south, east, and west as it follows the major highways that run through the cities; however this highway effect becomes unnoticeable at large distances from the centre. The zones at the edges of KW suffer from the omission of job information in neighbouring municipalities, where values would likely be higher had access to these jobs been considered.

Table 11 shows the distribution of this variable in the regression dataset:

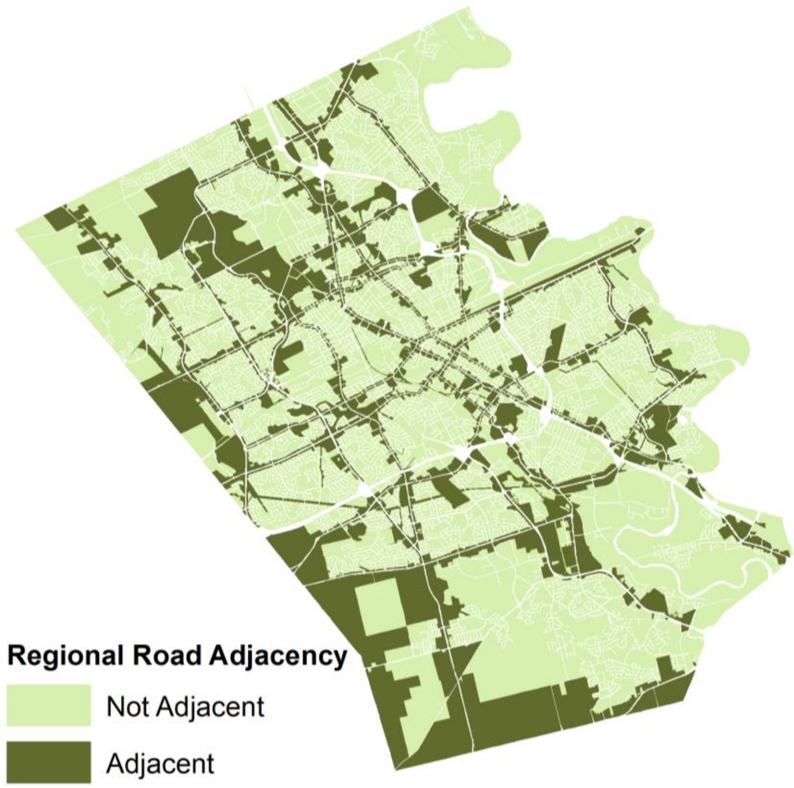
Employment Access		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	89,923	111,665	87,961	89,429	96,540	86,915
Std. deviation	21,789	19,930	20,863	22,727	16,389	16,269
Minimum	30,953	30,953	30,953	30,953	34,590	34,590
Maximum	135,051	135,051	135,051	135,051	135,051	134,308

Table 11: Employment Access: descriptive statistics

As is clear from the descriptive statistics, within the average commute time and distance, homeowners of homes bought within the CTC are able to reach on average nearly 25,000 more jobs. Townhouses sold in KW are able to reach slightly fewer jobs on average than single-detached homes. Semi-detached and duplex homes sold were able to reach around 10,000 more jobs than townhouses and 7,000 more than single-detached homes

4.1.2.1.6 Regional Road Adjacency

Map 13 shows the location of properties located within 25m of the centre of a regional road.



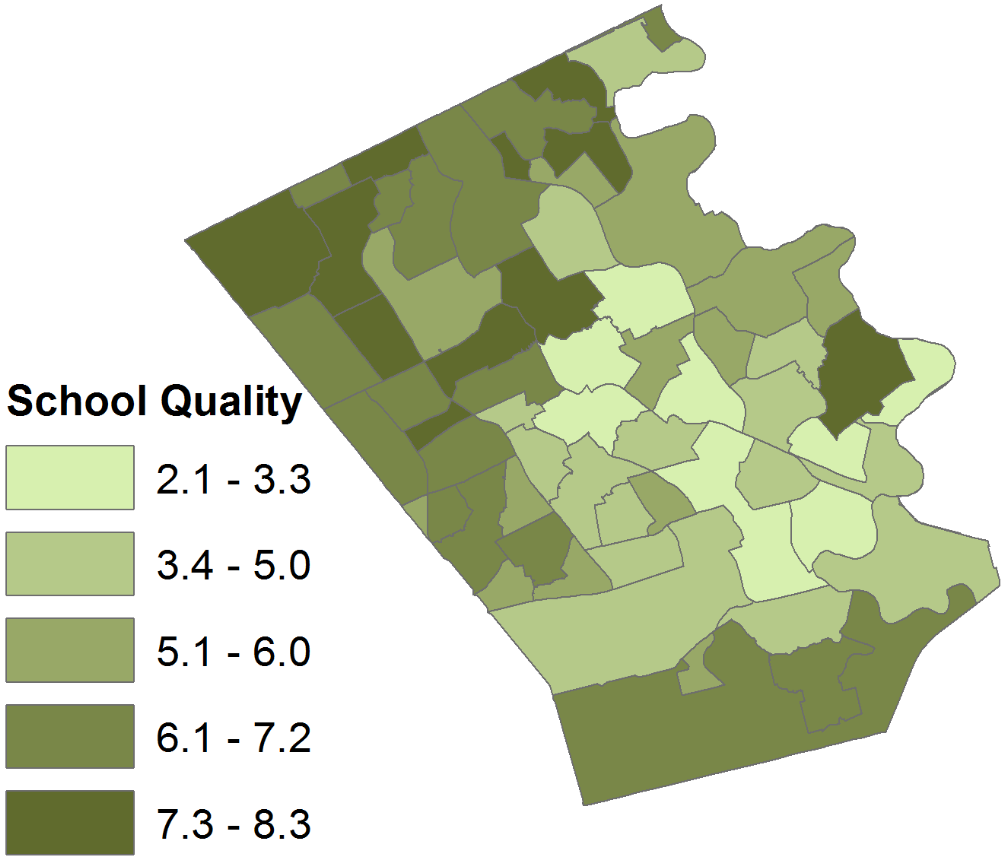
Map 13: Regional Road Adjacency

As can be seen from the map, there are many regional roads forming a non-uniform grid across KW.

4.1.2.2 Neighbourhood Variables:

4.1.2.2.1 School Quality

Map 14 shows the Fraser Institute school quality ratings of elementary school catchment areas in KW.



Map 14: School Quality

The quality of schools in catchments in the centre and southeast of the centre are the lowest in the region, while those in the northwest are the highest. Kitchener schools received lower ratings than Waterloo Schools as indicated by the generally higher values in the north and lower values in the south of the map.

Table 12 shows the distribution of this variable in the regression dataset

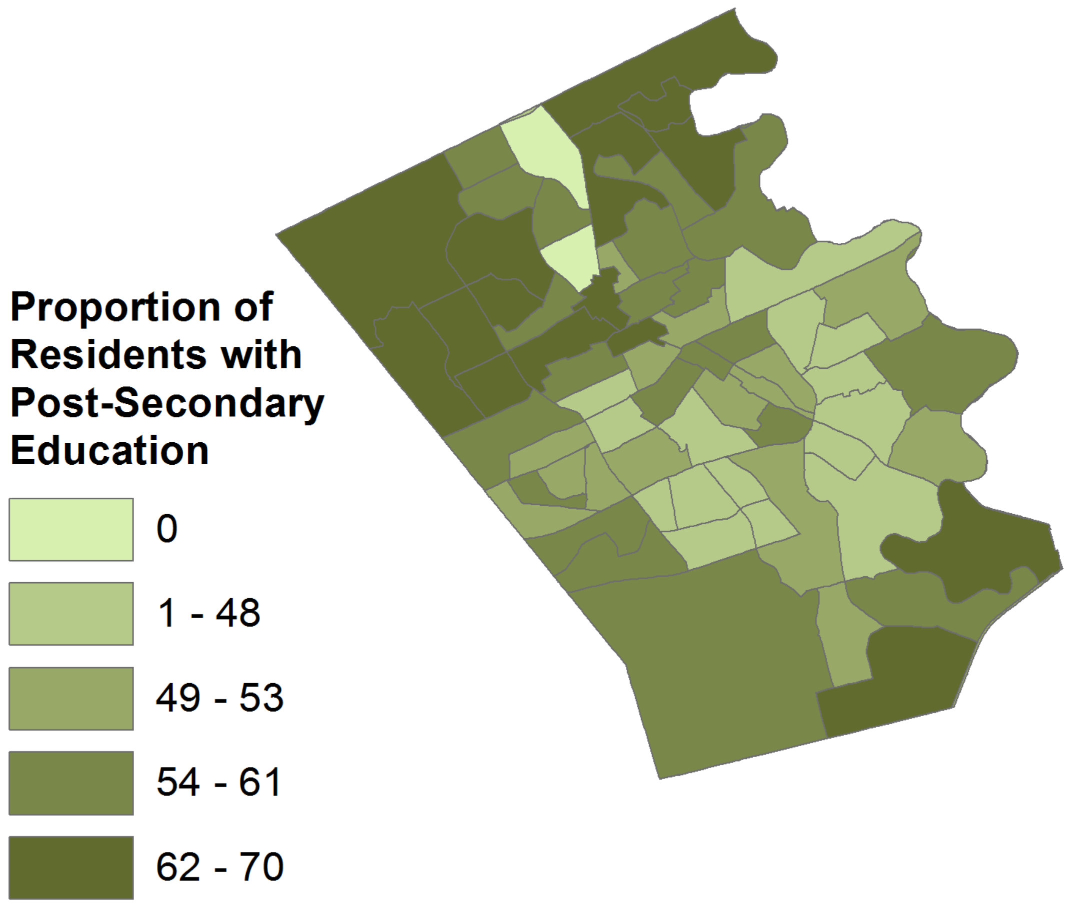
School Quality		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	5.9	4.9	6.0	5.9	5.6	5.9
Std. deviation	1.4	1.6	1.4	1.4	1.5	1.4
Minimum	2.1	2.1	2.1	2.1	2.1	2.6
Maximum	8.3	8.0	8.3	8.3	8.3	8.3

Table 12: School Quality: descriptive statistics

Generally, residential sales in the CTC were in catchment areas that scored approximately one point less in the school ratings on average. The average school quality of single-detached homes and townhouses were the highest, at 5.9, while the average score of semi-detached and duplex homes was 5.6. The standard deviation of school quality is fairly consistent, differing by a maximum of only 0.1 across property types and 0.2 between the CTC and the rest of the city.

4.1.2.2.2 Education Rate

Map 15 shows the proportion of educated residents in each CT in KW:



Map 15: Education Rate

The *Education Rate* map shows the proportion of residents who have undergone postsecondary education in each CT. Mid-level values are found in a linear pattern running north-south in the centre of the map. This linear patch of average education is surrounded on the east and west by areas of relatively low education rates. Two CTs in education data contained populations of zero, which are shown as darker in the map.

Table 13 shows the distribution of this variable in the regression dataset:

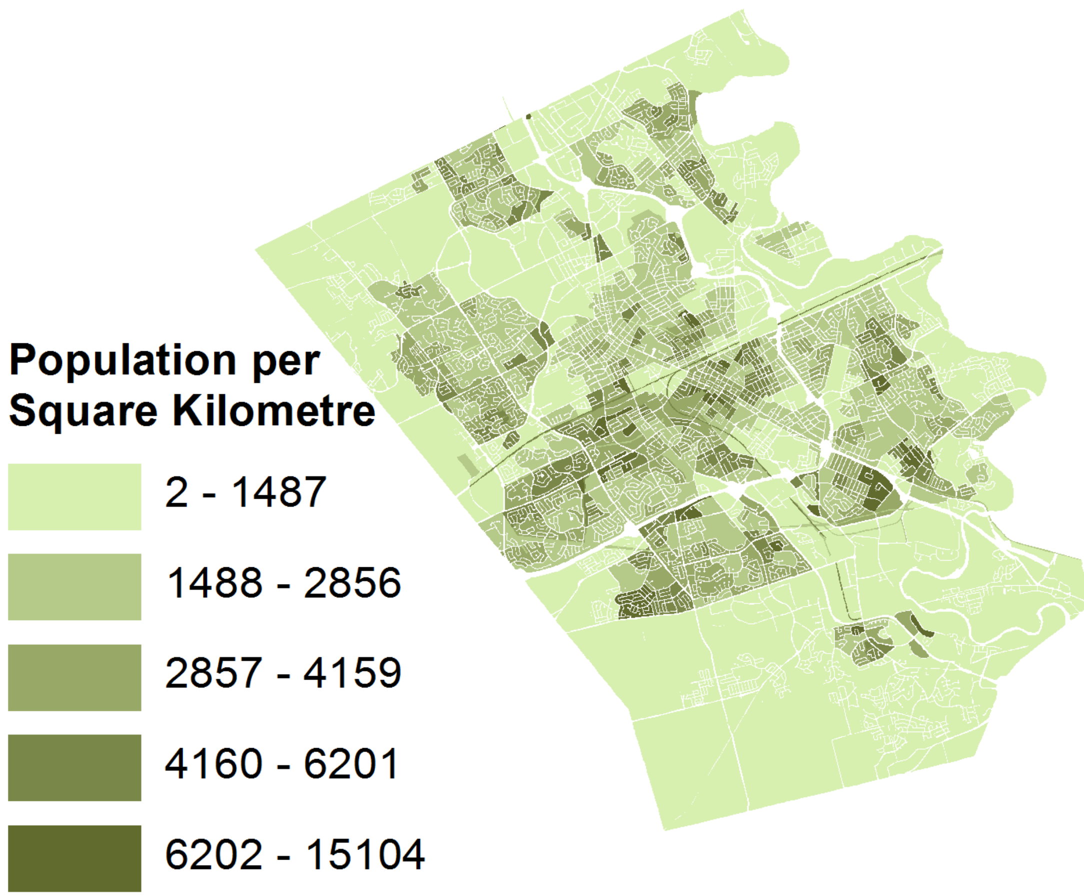
Education Rate (proportion with post-secondary)		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	57.2	52.1	57.7	57.3	54.1	59.9
Std. deviation	8.2	11.1	7.8	8.2	8.4	7.2
Minimum	0.0	0.0	0.0	0.0	0.0	38.9
Maximum	70.3	69.7	70.3	70.3	69.9	69.9

Table 13: Education Rate: descriptive statistics

Of properties sold, townhouses were on average located in the most educated neighbourhoods, followed by single-detached dwellings then semi-detached and duplex dwellings. Home sold within the CTC were on average in less educated CTs, however with a larger variation.

4.1.2.2.3 Population density

Map 16 shows the distribution of population density for each DA in KW.



Map 16: Population Density

Population is most densely populated in the centre of KW; however there are many DAs with high population density outside of the centre, including at a noticeable distance one cluster in the southern portion and another along the northern border. The lowest population densities are found along the borders of KW, with a very noticeable patch of low density in the south, which contains a cluster of higher-density DAs within it.

Table 14 shows the distribution of this variable in the regression dataset:

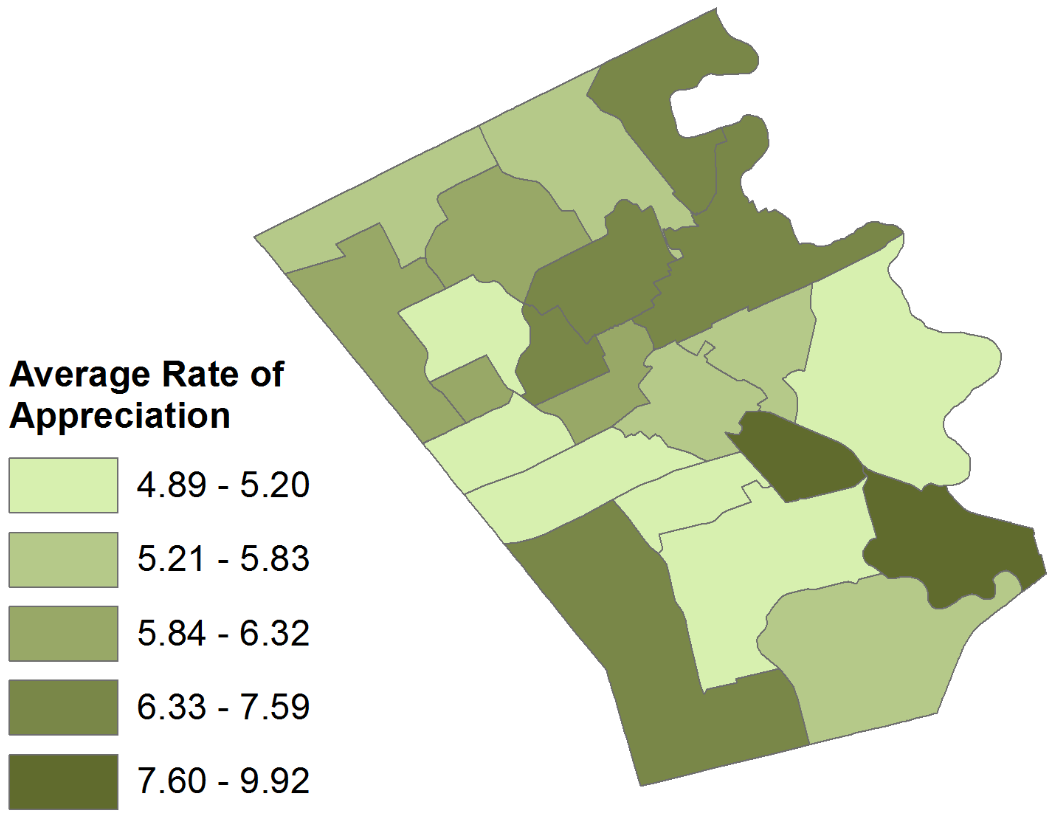
Population per square kilometer		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	2,539.5	2,969.2	2,500.7	2,420.9	3,503.9	2,530.9
Std. deviation	1,782.2	1,803.2	1,775.3	1,681.5	1,865.1	2,203.7
Minimum	74.5	121.0	74.5	74.5	110.1	201.3
Maximum	15,103.9	15,103.9	13,595.5	15,103.9	13,595.5	13,898.6

Table 14: Population Density: descriptive statistics

Homes sold inside the CTC were located in more dense DAs, averaging nearly 5000 residents per square kilometre more. Compared to the other property types, semi-detached and duplex home sales were located in the densest DAs, at an average of 3,500 people per square kilometre. Townhouses and single-detached home sales were in similarly dense neighbourhoods on average, but townhouses were surprising not found in any DAs with a population density less than 201 people per square kilometre.

4.1.2.2.4 Rate of Appreciation (ROA)

The following shows the average *ROA* homes sold between 2005 and 2014, by 'planning neighbourhoods' in KW. It should be remembered, the *ROA* variable represents the rate of change between lagged average home sales values in CTs between sets of two years (ex, the rate of change in the lagged average adjusted value of homes sold in 2000 and 2001 to the lagged average adjusted value of homes sold from 2002-2003).



Map 17: Average ROA

Map 17 shows, there is a high amount of variation in average appreciation rates between neighbourhoods, which follows no immediately apparent pattern. The fastest appreciating neighbourhoods, by this measure, are located in the southwest, southeast, and northeast. The slowest appreciating neighbourhoods include one neighbourhood in the east and a linearly contiguous set of neighbourhoods in the west.

The following, Table 15, shows the distribution of the sample's appreciation rate across property types and by whether they were located in the CTC.

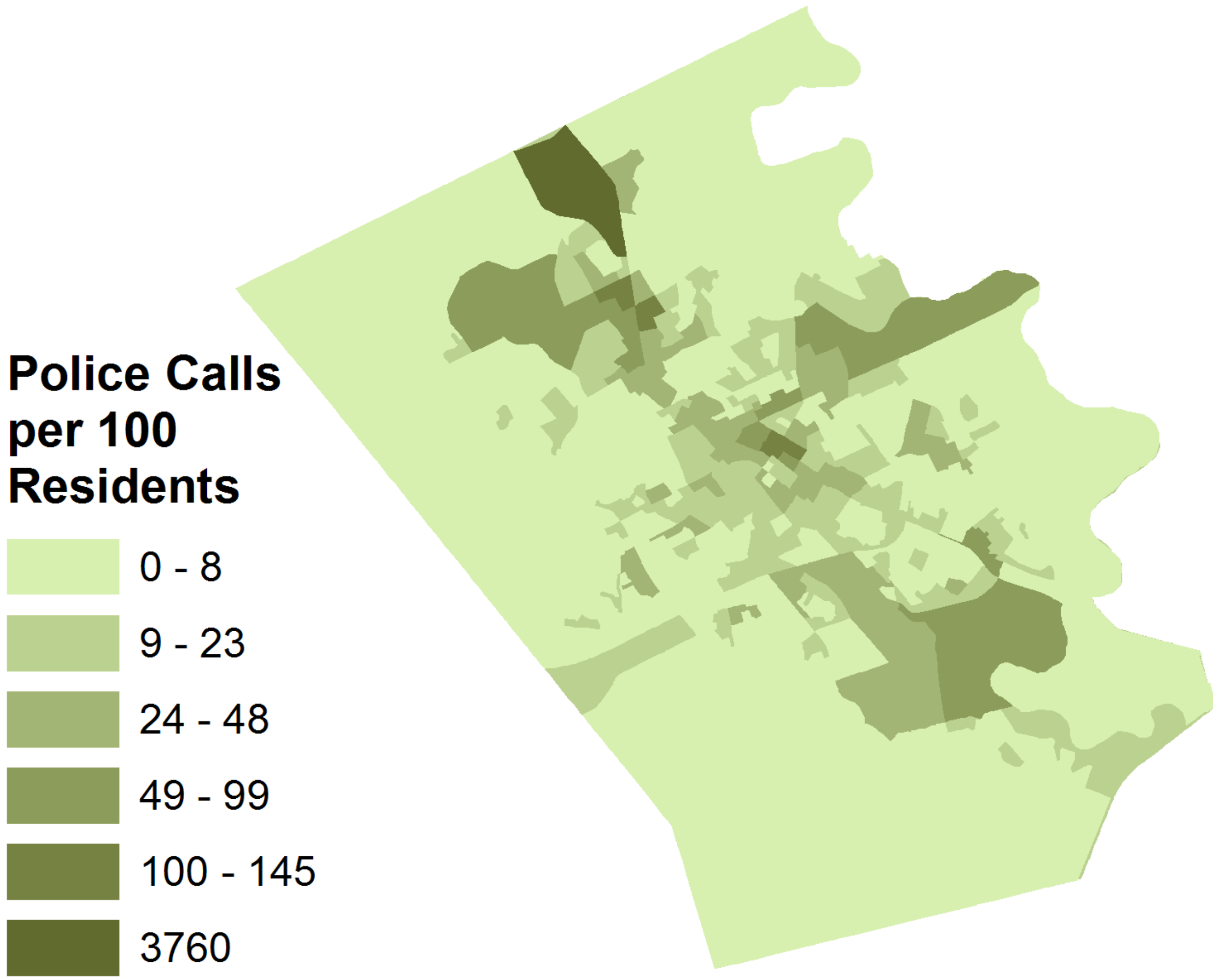
ROA		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	5.9	7.4	5.7	5.9	5.9	5.8
Std. deviation	3.0	4.1	2.9	3.1	2.8	2.6
Minimum	-10.8	-8.9	-10.8	-10.8	-10.8	-10.8
Maximum	25.0	25.0	25.0	25.0	25.0	16.4

Table 15: ROA: descriptive statistics

The descriptive statistics for the *ROA* show that homes sold in the CTC were located in neighbourhoods that were appreciating faster than those sold outside of it. Homes sold in the CTC were in CTs appreciating at an average of 7.4% per year, while the average for non-CTC sales was 5.7%. The various property types included had similar average appreciation rates, although no townhouses were sold in the fastest appreciation CT.

4.1.2.2.5 Perception of safety

Map 18 shows the distribution of the police calls per 100 residents variable calculated at the DA scale.



Map 18 – Police calls per 100 residents

High values of calls per 100 residents are located within the central north-south spine of KW. This variable declines with distance from the central spine of high values moving toward the east and west borders, although it does so in a non-continuous, fragmented pattern

While DAs are intended to contain relatively stable population counts, one DA in the northern edge in border of Waterloo in Waterloo contained a population of only 5 residents. This produced an extreme outlier in terms of calls per capita, with a value of 3260 calls per 100 residents. This zone was ultimately removed from the analysis, as it did not contain enough residential property sales needed to calculate the ROA variable.

Table 16 shows the distribution of this variable in the regression dataset:

Police phone calls per 100 residents		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	6.1	19.1	4.9	5.9	8.0	5.8
Std. deviation	10.5	21.7	7.8	10.1	14.6	7.1
Minimum	0.0	0.0	0.0	0.0	0.2	0.2
Maximum	144.6	144.6	135.8	144.6	144.6	41.6

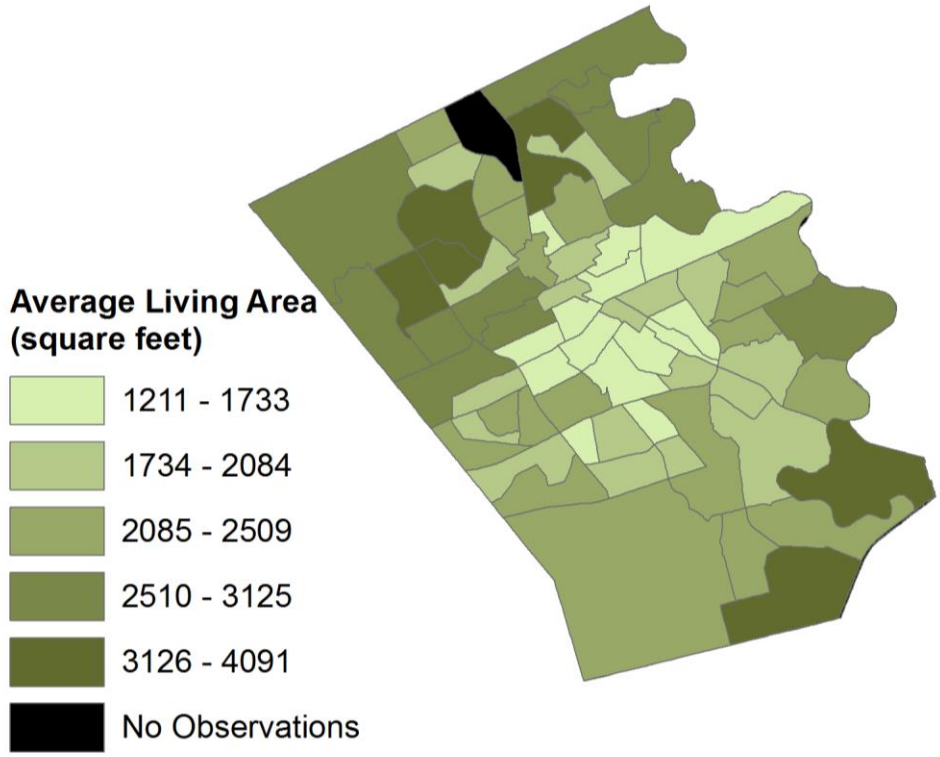
Table 16: Police Phone Calls per 100 residents: descriptive statistics

Inside the CTC has a much higher amount of calls per 100 residents. This may be due to the high employment density in the CTC, which attracts a greater amount of human activity and potential need for police. In addition, there is far greater variation in the CTC of police calls, with some areas of very high police calls and some areas with low calls. Single-detached homes and townhouses sold had similar average calls per 100 residents, with more variation in single-detached homes. Compared to single-detached and townhouses, semi-detached and duplex homes were sold in neighbourhoods that had higher and less consistent numbers of police calls per 100 residents. No townhouses were sold in areas with higher that 41.6 calls per 100 residents.

4.1.2.3 Structural Characteristics

4.1.2.3.1 Living Area

Map 19 shows the spatial distribution of living area in homes sold between 2005 and 2014 as an average in each CT.



Map 19: Living Area

Centrally located homes sold had less living area than those in outlying area, especially compared to the northwest, northeastern, and two isolated southeastern CTs. Homes sold in the southwestern, eastern, and some northern CTs had average living areas.

Table 17 shows the distribution of this variable in the regression dataset:

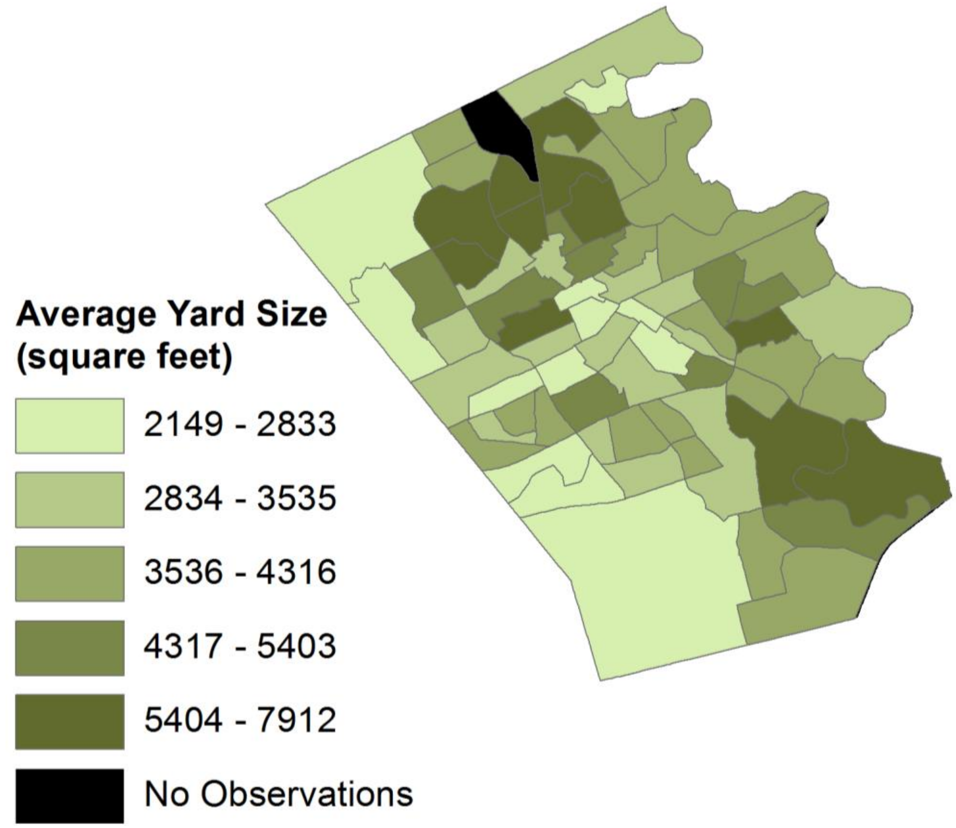
Living Area (square feet)		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	2,468	1,900	2,519	2,623	1,755	1,857
Std. deviation	1,047	918	1,043	1,077	562	481
Minimum	497	517	497	497	611	824
Maximum	14,134	7,602	14,134	14,134	6,230	4,538

Table 17: Living Area: descriptive statistics

Homes sold outside of the CTC were on average over 500 square feet larger than those inside. Semi-detached and duplexes and townhouses had similar sizes, although semi-detached and duplex homes varied slightly more than townhouses. Single-detached homes were the largest on average, at 2,623 square feet, but with both the lowest minimum and highest maximum sizes.

4.1.2.3.2 Yard Size

Map 20 shows the spatial distribution of Yard Sizes of homes sold between 2005 and 2014 as an average in each CT.



Map 20: Yard Size

Table 18 shows the distribution of this variable in the regression dataset.

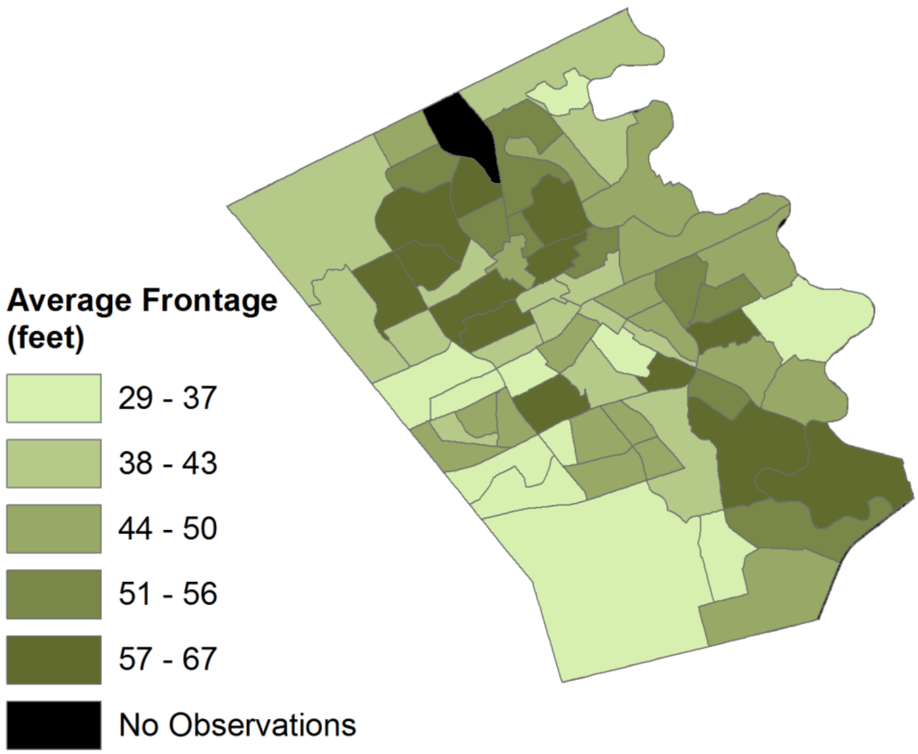
Yard Size (square feet)		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	3,606	3,415	3,623	3,938	2,659	1,629
Std. deviation	2,254	1,860	2,286	2,326	1,082	751
Minimum	51	51	587	394	51	293
Maximum	127,872	23,016	127,872	127,872	19,304	7,048

Table 18: Yard Size: descriptive statistics

Homes sold outside of the CTC had yards that were around 150 square feet larger. Townhouses sold had the smallest yards on average, less than half the size of the average single-detached home. Single-detached homes have the largest range of yard sizes. Semi-detached homes and duplexes had yard sizes between those of single-detached and townhouses on average, with the lowest minimum yard size of any of the property types included.

4.1.2.3.3 Frontage

Map 21 shows the spatial distribution of Frontages of homes sold between 2005 and 2014 as an average in each CT.



Map 21: Frontage (Feet)

Sales in centrally located CTs had relatively low frontages, while the highest frontages were found in the CTs to the northwest and southeast of the centre along a diagonal axis. Homes sold in the northwest and northeast tracts had similarly low frontages. The CTs where sales of properties with the lowest relative frontages took place are located in the southwest and west of KW.

Table 19 shows the distribution of this variable in the regression dataset

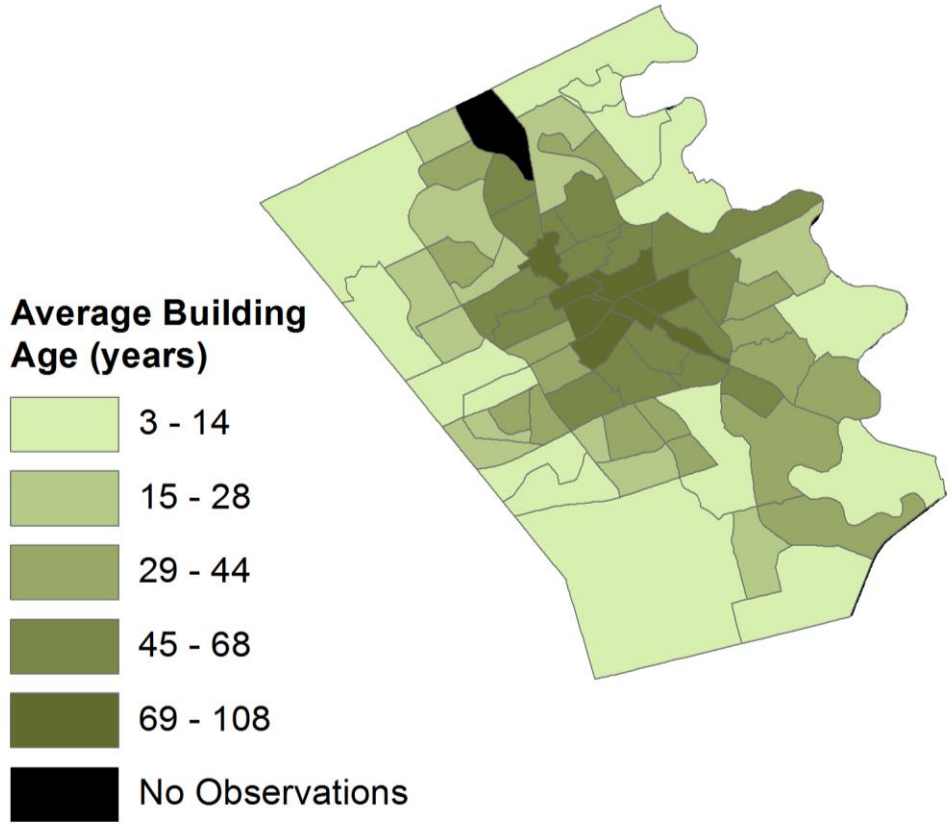
Frontage (feet)		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	42.6	44.7	42.4	45.8	33.6	23.4
Std. deviation	29.5	15.2	30.5	19.1	50.9	54.9
Minimum	0.1	2.9	0.1	0.1	5.8	2.9
Maximum	2,167.8	151.0	2,167.8	299.8	2,124.0	2,167.8

Table 19: Frontage: descriptive statistics

As seen in this table, the average frontages of properties sold inside and outside of the CTC were quite similar, although the frontages of those outside varied by a wider margin. There is a drastic difference between the frontages of property types that sold in KW. Of the sample of sales, townhouses had the smallest frontages, followed by semis and duplexes, then single-detached homes. On average, the frontages of single-detached homes were twice as large as those of townhouses.

4.1.2.3.4 Building Age

Map 22 shows the spatial distribution of building ages of homes sold between 2005 and 2014 as an average in each CT.



Map 22: Building Age

The average age of homes in KW declines moving away from the centre of the cities. This is consistent with the AMM model, where homes were initially built closest to the CBD, to allow for access to jobs. However, over time and with the introduction of the private automobile, homes were built at greater and greater distances. The CTs with the lowest average age of buildings are located at the rural-urban fringe, which reflects the persistence of sprawl-type development over the past few decades

Table 20 shows the distribution of this variable in the regression dataset

Building Age (years)		CTC		Property Type		
	Full Sample	Inside	Outside	Single Detached	Semi & Duplex	Townhouse
Mean	22.9	63.0	19.3	23.5	30.8	7.8
Std. deviation	24.4	35.4	19.4	24.4	27.1	10.3
Minimum	0	0.0	0	0	0	0
Maximum	201	164.0	201	201	164	112

Table 20: Building Age: descriptive statistics

From this table, KW homes sold between 2005-2015 were approximately 23 years old on average. Homes sold within the CTC were over three times as old as homes sold outside of it, although with more variation. The youngest property type sold is evidently the townhouse, with an average age of only 7.8 years and a relatively low standard deviation of 10.3 years. The townhouse is around one third the average age of single-detached homes sold and approximately one quarter the age of the average semi-detached or duplex home sold. Interestingly though, the oldest home in the dataset was a single-detached home located outside of the CTC, at 201 years old.

4.1.3 Bivariate results

This section provides a comparison of important correlations between candidate model variables. While a full matrix of the correlations between the independent variables can be found in Appendix 4, particular correlations will be described here as necessary to inform the model specifications found in the following chapter.

4.1.3.1 Dependent variables

The following tables show the correlations between the candidate dependent variables and the candidate independent variables. The first, table 21, shows correlations between the structural variables and dependent variables, Table 22 shows the correlations between the neighbourhood variables and dependent variables, and Table 23 shows the correlations between the environmental variables and dependent variables.

Correlations between candidate dependent and independent variables:

Structural Variables:	Living Area	Yard Size	Frontage	Building Age
Adjusted Value	0.75	0.41	0.2	-0.27
Assessed Value	0.81	0.43	0.21	-0.32
Log Adjusted Value	0.82	0.4	0.2	-0.37
Log Assessed Value	0.75	0.36	0.18	-0.31

Table 21: Dependent and Structural Variable Correlations

Neighbourhood Variables	Education Rate	ROA	Population Density	School Quality	Police phone calls
Adjusted Value	0.36	0.12	-0.32	0.23	-0.1
Assessed Value	0.38	0.11	-0.35	0.24	-0.11
Log Adjusted Value	0.42	0.09	-0.37	0.3	-0.14
Log Assessed Value	0.39	0.1	-0.33	0.28	-0.13

Table 22: Dependent and Neighbourhood Variable Correlations

Environment Variables	OS, Gravity-based	OS, Count in 1000m	OS, Area in 1000m	OS, Spatial Separation	OS, Adjacency	Transit	Employment Access	Walkability	Regional Road Adjacent
Adjusted Value	0.05	-0.06	0.21	0.08	0.17	-0.43	-0.29	-0.34	-0.08
Assessed Value	0.03	-0.09	0.2	0.07	0.17	-0.48	-0.34	-0.4	-0.08
Log Adjusted Value	0.04	-0.09	0.22	0.08	0.17	-0.47	-0.36	-0.43	-0.09
Log Assessed Value	0.05	-0.07	0.21	0.09	0.16	-0.41	-0.31	-0.37	-0.09

Table 23: Dependent and Environment Variable Correlations

For the most part, the correlations between the candidate variables and home prices match expectations. However, some environment variables show unexpectedly negative signs, which are *OS*

Count in 1000m, *Transit Access*, *Employment Access*, and *Walkability*, while Regional Road Adjacency has an unexpected positive relationship with property values.

Assessed Value and *Adjusted Value* show similar correlations with all of the independent variables; although overall the bivariate relationship between *Assessed Value* and the independent variables is slightly stronger than that of *Adjusted Value*. Correlations of the logged forms of *Adjusted Value* and *Assessed Value* show similar, although marginally stronger, correlations with the independent variables.

Of note from these tables, *Living Area* is the most strongly positively related variable to both adjusted and assessed value, followed by Yard Size. Walkability, employment access, transit access, population density, and building age all have a moderately negative correlation with home values. School Quality, Education Rate, and frontage all have moderate positive correlations with property values.

4.1.3.2 Comparing OS Variables

OS Access:	Gravity-based	Count in 1000m	Area in 1000m	Spatial Separation
Gravity-based	1			
Count in 1000m	0.7	1		
Area in 1000m	0.59	0.25	1	
Spatial Separation	0.42	0.22	0.19	1

Table 24: OS Access Correlations

The OS correlation matrix shows how the OS metrics relate to one another. OS count and area, from the cumulative opportunities model, are weakly, positively related. The gravity-based model is the most strongly correlated with the other metrics. The gravity-based model is moderately to strongly correlated with all of the other measures of OS access, while the others are all only weakly related to each other. Through a visual comparison of the gravity-based access, Map 8, to the cumulative opportunities maps, Maps 6 and 7, it can be seen that the gravity-based model has produced hotspots that coincide with both of the specifications of the cumulative opportunity model.

4.1.3.3 Collinear Variable Correlations

Collinear Variables 1:	In CTC	Transit Access	Walkability	Employment Access	Building Age	Population Density	Police Calls
In CTC	1						
Transit Access	0.12	1					
Walkability	0.48	0.49	1				
Employment Access	0.3	0.44	0.66	1			
Building Age	0.49	0.36	0.68	0.6	1		
Population Density	0.07	0.4	0.49	0.33	0.25	1	
Police Calls	0.37	0.09	0.34	0.26	0.38	-0.09	1

Table 25: Collinear Variables 1

Table 25 shows a subset of candidate independent variables that were found to have correlations stronger than ± 0.3. The most evident reason for this strong correlation is centrality. From the maps and descriptive statistics for each variable, it is clear that these most of these variable are highest in the CTC and the centre of the cities in general, while they are lowest at the periphery of KW.

Walkability and building age had the highest correlation, followed by walkability and employment access. Building age also had a very high correlation with employment access. Population density and police calls per 100 residents were related similarly to all other variables but not to each other. Transit access was simply moderately related with most of the other variables.

Collinear Variables 2:	Waterloo	Education Rate	School Quality	OS Access, area in 1000m
Waterloo	1			
Education Rate	0.53	1		
School Quality	0.5	0.49	1	
OS Access, area in 1000m	0.35	0.41	0.34	1

Table 26: Collinear Variables 2

Table 26 shows a second subset of collinear independent variables, with correlations stronger than ± 0.3. From a comparison of the maps of *education rate* and *school quality*, it is clear that high and low values of these two variables are found in similar locations, and that for the most part higher values of these variables are found in Waterloo than in Kitchener. As well, there was a strong correlation between the amount of OS accessible and the other three variables. Areas with the best schools and highest proportion of educated residents are also areas with the most amount of OS accessible within a kilometre.

Collinear Variables 3:	Living Area	Yard Size	Frontage	Building Age
Living Area	1			
Yard Size	0.33	1		
Frontage	0.19	0.43	1	
Building Age	-0.29	0.23	0.21	1

Table 27: Collinear Variables 3

Table 27 shows a subset of correlations for the different property characteristics assessed. The strongest correlations are between *Living Area* and *Yard Size* and between *Frontage* and *Yard Size*.

4.2 Analysis of Survey Sample

This section will outline the results related to the set of home sales data that corresponds to the survey conducted by Emma DeFields (2013). After limiting the survey dataset to contain only arms-length sales that occurred between 2005-1012, only 39 observations remained. Of these, 5 were property types other than single-detached and were removed so as to not influence estimates in models that do not control for property type. The remaining sample could be characterized as generally non-centrally located, with only 4 of the 34 observations located in the CTC.

This section is broken down into two subsections where survey variables are presented. The first subsection provides descriptive information on the demographic variables and how they relate to home values and intensification related variables. Then, the second section will outline the stated preference variables and their relationship with the preferences they concern.

4.2.1 Demographic Variables

4.2.1.1 Education and Income

The following table summarizes the continuous demographic survey variables from the survey sample, income and education, and how they relate to property values and intensification-related variables:

	Income (n=27)	Education (n=34)
Mean	$99, 000	12.2 Years
Range	$40,000-$200,000	0-20 Years
Standard Deviation	$42830	6.66 Years
Correlation with Adjusted Value	0.5449	0.087
Correlation with OS Access (gravity-based)	0.322	0.083
Correlation with Transit Access	-0.279	0.088
Correlation with Employment Access	-0.040	0.198
Correlation with Walkability	-0.113	0.1142

Table 28: Continuous Survey Variables, descriptive statistics

The average income of the 27 survey respondents who reported their income was almost $100,000, with a large range. Income of the survey respondent was moderately, positively correlated with the price they paid for their home. To a lesser extent, income was also associated with higher levels of OS access, and lower levels of transit access in the sample.

The average education of the sample was just over 12 years, with a range from 0-20 years. Correlations between years of education of homebuyers and the intensification-related elements of their homes were weak, with the strongest being a positive correlation of 0.198 with employment access.

4.2.1.2 With and Without Children

The following, Table 29, shows the average adjusted values and intensification-related metrics for households with and without children:

	With Kids (n=13)	Without Kids (n=21)
Average Adjusted Value	$239,815	$264,137
Average Access to OS (gravity-based)	40.8	43.2
Average Access to Transit	-400m	-274m
Average Access to employment	90,406 jobs	89,306
Average Walkability	-1.65	-1.46

Table 29: With Kids, descriptive statistics

Household with children purchased less expensive homes, on average. These homes had less access to OS, were farther from transit, and were located in less walkable areas than homes purchased by households without children. There was very little difference between the amount of accessible jobs from properties bought between households with and without children.

4.2.2 Stated Preference Variables

This section will summarize the difference in levels of access to amenities by those who stated preferences for them and those who did not.

4.2.2.1 Transit Preference

The following table shows the average *adjusted values* and *transit access* values for households, categorized by whether they rated transit access somewhat important or higher in their decision to move to their current neighbourhood:

	Transit preference (n=17)	No Transit Preference (n=17)
Average Adjusted Value	$248,192	$261,482
Average Access to transit	-281m	-364m

Table 30: Transit Preferences, descriptive statistics

Transit preference was split evenly across the sample. Households that expressed a preference for transit were located 83m closer to a transit stop than those who did. The average *adjusted value* price of homes purchased by those who stated a preference for transit were roughly $13,000 less than homes bought by those who did state a preference for transit access.

4.2.2.2 OS Preference

The following table shows the average *adjusted values* and various *OS access* values for households, categorized by whether they rated transit access somewhat important or higher in their decision to move to their current neighbourhood:

	OS preference (n=29)	No OS Preference (n=5)
Average Adjusted Value	$263,508	$204,548
Average Access to OS (Spatial Separation)	-230m	-251m
Average Access to OS (Cumulative Opportunities: Count in 1000m)	8.4	7.6
Average Access to OS (Cumulative Opportunities: Area in 1000m)	944,902m^2	778,480 m^2
Average Access to OS (gravity-based)	42.7	39.9

Table 31: OS Preferences, descriptive statistics

Only five households of the 34 in the sample did not express a preference for OS. Those who stated an OS preference bought homes located in areas with greater access to OS by all the OS measures presented, as well as purchased more expensive homes, on average.

5 Regression Methods

This chapter will outline the regression methods used for this research. First, this chapter will provide the theoretical model that guided the specification and calibration process. The chapter will operationalize the theoretical model in light of the capabilities and limitations of multiple regression modelling. This chapter uses findings from the results presented in the previous chapter alongside theory to justify decisions. Finally, this chapter will conclude with the final model specifications that were run to answer the primary research questions of this work. The results of these methods can be found in the following chapter. Attendees of two presentations outlining the preliminary results of this thesis provided guidance in the determination of an appropriate model specification: attendees of the Association of American Geographers conference and planning staff members at the ROW.

5.1 Theoretical Model

Initially, a broad theoretical model of homebuyer WTP was conceptualized based on the Review of the Literature and the KW intensification context. The following provides this overarching model that guided the initial selection of model variables:

Property Value = f (*Structural Attributes, Neighbourhood Socioeconomics, Spatiotemporal Market Trends, Intensification Amenities and Interactions, LRT & CTC Development Over Time)*

Equation 25: Theoretical model of property value in KW under intensification

This theoretical model posits that home prices, after conditioning on a range of structural, neighbourhood, and market controls, are affected complexly by concomitant intensification and LRT development. In this model, the home price impacts of intensification are thought to vary based on both synergistic effects of intensification amenities and on whether a home is in the CTC throughout the approval process of regional LRT. The interactions between intensification amenities are based on the idea that a home's price is not affected unitarily through individual elements of its intensification context, but rather that the spatial coincidence of the various amenities adds a distinct and relevant value in addition to these variables separately. This model aims to estimate discrete property value effects for different time periods corresponding to milestones in the LRT approval and development process and to test whether this impact is greater within the CTC.

5.2 Choice of Model

Initially, a model was run on the full dataset, with property types included as dummy variables interacted with the intensification amenities. As is clear from the results in the preceding chapter, housing characteristics and values are quite different depending on the type of property being considered. As well, housing types are distributed differently within and outside of the CTC, due to historical and contemporary development patterns and the marketability of different housing styles in each context. Considering this, and in light of literature on self-selection that would suggest different types of people may buy different types of homes, the sample was split into different models. Splitting the model aimed to identify heterogeneous homebuyer WTP for intensification amenities across buyers of different property types.

5.3 Functional Form

Generally, the hedonic modelling literature supports the use of the log-level functional form. While most of the hedonic models studied in Higgins & Kanaroglou (2016) employed the level-level specification, it has been stressed that the dollar-value of property characteristics is not constant across properties. The level-level specification was initially tested for this research, but is believed to have resulted in heteroskedasticity in the models. A log-level model assumes a non-linear relationship of variables to home values, where the coefficient estimates from the regression represent proportional effects on the dependent variable. This specification relies on the assumption that the effects of various home characteristics scale with total property value rather than present a fixed value to all properties.

The interpretation of coefficients from a log-level model is different from a level-level model. In a log-level model, coefficients are generally interpreted that a one unit increase in X will result in a $100\times\beta\%$ increase in Y. For example, level-level models are interpreted in such a way that with an additional 1,000 square feet, and a coefficient of 10 on the living area variable, a home's estimated value would increase by $10,000. In a log-level model, a coefficient of 0.000095 would indicate that an additional 1000 square feet, would mean a 9.5% increase in overall property value (0.0095*1000). The coefficients represent the effect of increasing the independent variables one unit, and these effects are described in relative terms rather than in concrete dollar values, essentially an elasticity. However, this is essentially an approximation that only holds true for values of coefficients between -0.1 and 0.1 (Kephart, 2013). The more precise estimate requires exponentiation of the absolute values of the coefficients, where $\exp(\beta)-1$ provides the estimated the percent change in price of a one unit increase in X. This can be achieved through the formula "=(exp(abs(*coefficient*))-1)*100"× ±1 (depending on the original sign of β) (Wooldridge, 2012). Because of a logged dependent variable, the total estimated sale price can only be returned to after all of the regression estimates are added together. This is because converting logs to their original form requires exponentiation, and exp(A) + exp(B) does not equal exp(A+B).

5.4 Variable Selection

The following sections operationalize the theoretical model's assumptions into a testable set of variables. Due to limitations of regression models in dealing with multicollinearity, operationalizing the theoretical model involved significant simplification from a model specified with all candidate variables.

The variable selection process involved consulting the available literature on hedonic modelling to identify the most important and commonly included determinants of property values, as discussed in Chapter 2. Structural variables were the most consistently applied in the literature, which included the *Living area*, *Yard Size*, *Frontage,* and *Building Age* variables in this work. Other common structural determinants, including number of bathrooms and measures of building or structural quality were unavailable to the researcher. The remaining candidate variables were selected that 1) related to the specific spatial and temporal setting of the study area (the In CTC and Time Period variables), 2.

represented aspects of the home's locational socioeconomic context (the neighbourhood controls), and 3. were related to intensification and thought to affect homebuyer WTP.

The appropriateness of candidate independent variables in the models in this research were evaluated based on their bivariate data distributions and on theoretical understandings of the residential land-market. Test models were then run to produce preliminary regressions using many different theoretically valid variable combinations. The results of these test regressions were compared to each other and to theoretical expectations, which helped to inform the analysis presented in the following subsections.

5.4.1 OS variables

Selecting an appropriate OS access variable involved consideration and comparisons of the distributions of OS access across KW for each of the various specifications of OS access presented in Chapter 4. From these results, it is clear that gravity-based access models provide values that are to some extent consistent with all other OS access variables presented in this work. Spatial Separation and various specifications of Cumulative Opportunities models provide important but dissimilar characterizations of access.

The correlations between OS access variables and home prices is given in 24. From these correlations, it would appear that gravity-based access is the least related to home values. However, preliminary OLS regression results indicated similar or stronger levels of significance for gravity-based than other OS access variables when structural and neighbourhood controls were present. Furthermore, estimates using OS access calculated with the cumulative opportunities models were highly sensitive to the removal or addition of various neighbourhood variables in preliminary models, with inconsistent effect directions between test models due presumably to multicollinearity.

Because the count and area cumulative opportunities models produce values that vary in relation to the city centres and outlying areas, respectively, it is believed that these variables represent a level of access that does not control for the heterogeneity between urban and suburban types of OS access. The gravity-based access measure, as described in the univariate results, shows similar standard deviations across *CTC* location as well as *Property Type*. This provides substantial justification to include gravity-based OS access over the others, as it produces identically distributed results across urban contexts. The variation presented, which is distributed uniquely from the other variables in this research, allows for the discrete and uninhibited estimation of OS effects on residential values. As well, interactions between In CTC and Property Type allow for the heterogeneous estimation of the different types of OS access that occur in the central and outlying areas.

OS adjacency was added to control for the scenic amenity value of OS, allowing for estimates of *OS access* to be estimated exclusive of OS's aesthetic impact on neighbouring properties. Regional road adjacency was likewise added to the list of variables to account for a disamenity affect that was believed to be interfering with the estimation of transit access, which will be described in the following.

5.4.2 Addressing collinear variables due to spatially distributed neighbourhood heterogeneity

While multicollinearity does not affect the overall predictive power of a model, it has the potential to produce individually meaningless coefficients for those variables in which multicollinearity is present (Yoo et al., 2012)

A large set of neighbourhood controls was developed to ensure that observations' neighbourhood level socio-economic situation did not bias estimates of WTP for intensification-related environmental features. These neighbourhood characteristics represent processes that occur at a larger scale than are thought to be controlled for through spatial regression methods (as these spatial regression methods use only a limited set of nearby observations to control for spatial autocorrelation and heterogeneity). This set of neighbourhood controls developed were found to be highly collinear,, as well as with some environmental and structural variables. Therefore, regressions run using this full set of neighbourhood variables produced estimates showing unstable signs that did not match theoretically understood relationships.

Specifically, three sets of moderately collinear variables were found, Tables 25-27 in Chapter 4. The spatial distribution and theoretical links between variables in these 3 tables will be discussed in the following subsections. This section explains how spatially dependent effects of these variables, which are distributed over space in a similar way, were managed for the final model specifications.

5.4.2.1 Collinear Variables 1: Centrality

Seven variables were found to relate to each other in a loose gradient outward from the central core of the two cities, as seen from a comparison of maps for variables found in Table 25. Because all of these variables share a similar spatial pattern, the model results of preliminary regressions were erratic and inconsistent with the theoretical relationships of variables to home value, indicative of multicollinearity as outlined in Chapter 2. These impacts of multicollinearity were prevalent in the model results using all of these variables, which included the production of coefficient estimates with unexpected signs and that changed erratically upon the addition and removal of other unrelated variables.

Ultimately, *population density* and the *In CTC* variable were deemed adequate to control for the effects of centrality necessary to estimate the relationships of the primary intensification-variables of this work. The following subsections outline in detail why variables were either removed based on their strong correlations or kept despite them.

Employment Access

Employment access was ultimately removed from the set of candidate variables. The decision to remove employment access was due to two factors:

1. That the employment access variable, as calculated, does not represent access at a small enough scale to identify significant variation in employment access within neighbourhoods. Moreover because of this variable's calculation, in relation to the geographic extent of Waterloo where most of

the area can be accessed within the average commute time and distance, the resulting variable is significantly affected by the shape of the study area. That is to say, central areas are disproportionately advantaged, where border zones omit employment accessible in other municipalities.

2. That, because of the problems highlighted in the last point, employment access did not represent true employment access as much as it represented distance from the centre at an aggregated scale. While distance to centre has an effect on property values that is explained by the AMM model and often estimated in hedonic models (see Chapter 2) the true effects related to centrality are the product of a more complex set of variables than solely distance to centre. In the context of developing parameters for a context-specific *`photograph'* ABM as outlined in Chapter 2, where it is known in reality the spatial outcomes of centrality are not equally distributed with distance to centre, it is better to apply variables that represent the constituent, disaggregate drivers of price rather than the larger process under which they are distributed. Since other variables in this research were calculated at smaller geographic scales that more directly capture effects that vary with distance to centre, like walkability, transit access, and population density, it was determined that to include the employment access variable would be redundant and lead to confused estimates.

Police Calls

Police Calls was ultimately removed from the set of variables in the final regressions.

In preliminary regressions *Police Calls* showed a positive relationship with property values, which was robust to the removal and addition of various neighbourhood controls including population density. This relationship is inconsistent with theory and literature regarding the relationship between neighbourhood safety and property prices. Ultimately, it was determined that police calls per 100 residents is either 1) an inadequate indicator of the perceptions of neighbourhood safety or 2. too correlated with other included neighbourhood quality variables for its impacts to be estimated explicitly.

In CTC

The In CTC variable was kept in this research.

The In CTC variable was deemed necessary in this research, because it 1) it represents a defined area that has implications for land-use values due to preferential zoning and the pedestrian access threshold for LRT (Atkinson-Palombo, 2010), 2) the question of the effects of CTC proximity on land use is central to the research question of the study, and 3) Estimates for this coefficient will serve as a comparison baseline in future research.

In CTC's relationship between the centrally distributed variables in this section was leveraged by using it as an interaction term with intensification-related amenities in the final regressions. These interactions are discussed in a later section in this chapter.

Building Age

Building Age was kept in the set of variables in the final regression.

While building age has strong correlations with other variables and is highly characterized by the distance to centre, the effects of building age on property price are independent of this centrality. Building age in hedonic models is generally used to capture building quality, where older homes are often assumed to be less desirable because of the potential costs of upkeep and repair. It is believed that after controlling for the effects of centrality in the coordinated, disaggregate way described throughout this section, the effects of building age are able to be uniquely identified within the model and do not interfere with other model estimates.

Transit Access and Walkability

Transit access and walkability were kept.

Estimating the effects of transit access and walkability is part of the core intent of this research. While walkable environments and transit access are most prevalent in the centre and where population densities are higher, their relationship with distance to centre is not completely linear. These two variables are calculated at a small enough scale to allow the values to vary within neighbourhoods. Once population density is controlled for and the effects of these variables are estimated heterogeneously between the CTC and different property types, these variables can be accurately estimated.

Regional Road Adjacency was added to the list of variables to account for initial negative impacts of transit access on property value found in preliminary models. As transit stops are frequently located on large regional roads with higher traffic volumes than most residential streets, it was important to ensure that the disamenity effect of this traffic was accounted for.

5.4.2.2 Collinear Variables 2: North-South Variation

School quality was removed and education rate was kept.

The second set of collinear variables included school quality, education rate, and the KW dummy variable. These variables generally showed higher values in the north and lower values in the south. Preliminary regressions of different combinations of these variables produced inconsistent results. As various controls were added and removed, the results of the school quality and KW variables changed erratically. The estimated relationship between education rate and home values remained generally stable under various specifications of control variables. For these reasons, the school quality and KW dummy variables were removed from the final model specification, allowing education rate to independently account for this north-south neighbourhood home price variation.

5.4.2.3 Collinear Variables 3: Structural Variables

Yard size and living area were kept, frontage was removed.

Yard Size and Frontage were also found to be moderately correlated. This is understandable, as Ultimately, frontage was dropped from the final regression because 1. Yard size inherently incorporates much of the effects of frontage by itself, and 2. It is the more important of the two in relation to the OS related research questions of this work.

As well, Yard Size and Living Area were found to be fairly correlated, however, their discrete effects on home value are large and different enough that relatively accurate estimates can be produced. As well, the structural variables of this work are not the primary focus for the estimates, so some degree of collinearity within them will not affect model estimates for the intensification-related variables included.

5.4.3 Interactions

Interaction terms provide the opportunity to add complexity to regression models. Interactions in regression can be used to estimate heterogeneous functional relationships between property characteristics and prices within the data. In this work, interactions are used to estimate the heterogeneous impacts of intensification related amenities for different types of homes and for homes located in the CTC. Interactions will also be used to identify synergistic or moderating impacts between intensification-related variables. Potential interactions were limited to only those between candidate intensification variables, to ensure the effects related to the research questions of this work would be easily interpretable from the regression results. Ultimately, the interactions designed for this research fall into three categories: In CTC, Property Type, and between variables.

The first two interactions discussed help to control for the effects of self-selection in homebuyers' decision-making. Self-selection – as outlined in Chapter 2, would suggest that homebuyers who have a preference for intensification related amenities would naturally 1. Want to be centrally located, where these infrastructures are most prevalent, and 2. Be more likely to live in a higher density property types naturally associated with denser environments. In these cases, interaction is used as a tool to organize the observations into groups who are expected to have significantly different preferences from each other. The last interactions detailed in this section attempt to account for theorized synergistic impacts between intensification related amenities, as well as for potential mediatory effects of private and public OS access. Interactions will also be used to identify heterogeneous WTP by homebuyer characteristics in the models.

5.4.4 In CTC:

Two types of interactions were employed with the In CTC variable: those with intensification amenities, and those with the timing of approvals for regional LRT.

Interactions of intensification-related variables with the CTC variable will provide information on how preferences for these amenities differ from those within and outside of the CTC. Naturally, the CTC contains a much different distribution of many of these amenities than the rest of the city. The CTC has much higher *access to transit* and *walkability* than the rest of the KW. The CTC does have a similar distribution of *OS access* according to the gravity-based measure; however, the gravity-based OS access measure contains high values inside and outside of the CTC for different reasons. Inside the CTC high levels of OS access result from a large number of OSs being nearby, while outside the CTC high values of access are due more so to the large area of nearby OSs. Because of these differences, the In CTC variable interactions will be able to estimate heterogeneous effects of these variables with potentially different relationships between these amenities and home values depending on urban situation.

Admittedly, the CTC is not itself a homogeneous entity, where land use distributions between station areas can vary dramatically. The walkability measure is thought to account for the different land use context within the CTC – where more urban locations within the CTC would have greater walkability, and more industrial or suburban locations will see a low value of walkability. The use of the CTC variable is not meant to represent an area with homogeneous character, but rather to specify the area which is planned to receive increases to non-automobile accessibility and increased development.

Interactions with the time period dummies will be used to test whether home values in the CTC increased alongside the LRT's approval and implementation. As was noted in Chapter 2, where new transit infrastructures are being developed in cities, higher land values were often found in transit station areas that received specific land-use plans to integrate the new transit into the urban realm, as opposed to those where no specific plans preceded the transit's implementation. Because of the land use planning process for the transit station areas in the CTC as well as the central area's density targets set out in the GGGH, the land within the CTC is expected to develop and appreciate faster than the rest of the region. This is supported by the ROW's CTC Baseline Monitoring Report (ROW, 2015c). The results of the *ROA* variable created for this work indicate that sales with the CTC have appreciated more than the KW average between 2005 and 2015, which may be due to the TOD and intensification amenities it provides. This difference in planning context and opportunity for disparate effects from the implementation of LRT justify the interactions of intensification-amenities with the In CTC variable.

5.4.4.1 Between Intensification Amenity Variables

Interactions between intensification related variables will be tested in the final model specification to identify potential synergies between amenities associated with intensified environments. As mentioned in the literature review, elements of TOD have been shown to have a synergistic relationship – where the land value impacts of singular elements of TOD are amplified through the co-presence of other elements of TOD. In this work, interactions between public OS, access to transit, and walkability are being investigated to identify whether these features have synergistic relationships with residential WTP. The theoretical basis for this is in the expected tastes of homebuyers and self-selection. It is believed that specific homebuyers have a preference for a variety of intensification amenities, rather than for these amenities singularly. It is expected that urban homebuyers desire not only to be able to walk, enjoy OS, and take transit, but to be able to do all of these within their neighbourhood.

5.4.4.2 Between Private and Public OS

An interaction between Public OS Access and Private OS Access is being incorporated in the final model. This interaction will be used to test the hypothesis that homebuyers will express a different WTP when both private and public OS are present. There are two competing theories as to the direction that is expected of this interaction effect. 1. It is theorized that public OS provides an alternative to private OS, and so a negative interaction would suggest that so long as a home had access to either public or private OS, the other would be less relevant to informing buyers' WTP . 2. The second theory of a potential positive value from this interaction would represent the preferences of homebuyers for a diversity of OS access, including both public and private OS access. In this case, it would be expected that homebuyers would express a higher WTP when purchasing a home where both private and public OS are present.

5.4.5 Spatial Models

Under the fulfilment of the Gauss-Markov assumptions outlined in Chapter 2, OLS regression is considered the best linear unbiased estimator. However, regression models fitted with spatially explicit data often suffer from spatial effects that violate of the gauss-markov assumptions. The primary violation is the production of endogenous errors due to spatially autocorrelated observations. For this work, OLS and spatial models will both be presented in the following chapter, accompanied by relevant statistics to compare their performance.

Technically, running a spatial regression on such a large set of data was difficult. Initially, Stata, which was used to perform the aspatial data analysis and run preliminary test OLS regressions, was attempted to run the spatial regressions. However, due to matrix size limitations, Stata was unable to generate the spatial weight matrix necessary to define relationships between observations – a necessary component of spatial regression methods. Ultimately, GeodaSpace was decided upon to run the final spatial models. GeodaSpace offered the capacity to run models that account for spatial effects through both a spatial lag and spatial error term. Initial spatial models using SLM and SEMs separately exhibited significant heteroskedasticity according to Breusch-Pagan test statistics. Therefore, a combined SLM and SEM with explicit corrections for heteroskedasticity was employed in this work.

The spatial regression methods required that no two observations are able to be located in an identical location. To overcome this limitation, when multiple sales of the same property were found in the dataset the most recent was kept and the rest dropped from the dataset.

In spatial regression, R^2 is not the best indicator of model goodness-of-fit, as discussed in Chapter 2. Instead, AIC, Schwarz Criteria, and the Log Likelihood test statistics are more appropriate indicators in comparing spatial regression results. The better model is the one that minimizes AIC and Schwarz Criteria, and maximizes Log Likelihood (Anselin, 1988). Additionally, GeodaSpace provides Pseudo R^2 and Spatial Pseudo R^2, which can be used to understand model goodness-of-fit in spatial models in a similar way to traditional OLS models through standard or adjusted R^2.

5.4.5.1 Spatial Weighting

Inherent in spatial regression is a weighting scheme to identify the spatial relationships between observations in the dataset. As presented in Chapter 2, there are a number of ways to specify spatial relationships in the data. This current research utilizes a k-nearest neighbours approach with ten neighbours specified for each observation. While a distance-based neighbourhood definition may have been able to better incorporate the effects of distance between observations in the spatial weighting, the inconsistent distance between observations made this unfeasible. With some properties very far from the next nearest observation, and some extremely close to many other included properties, no distance-based spatial weight matrix could be computed that allowed all observations to have neighbours without including an extremely large number of neighbours for some observations.

5.5 Final Model Specification:

The following outlines the model specification that was ultimately estimated. GeodaSpace was used to run the spatial regressions as explained in the previous section. All models were run as both OLS and combined SLM and SEMs.

Recall the general hedonic model:

$$Y_i = \beta_0 + \beta_1 \times S_i + \beta_2 \times N_i + \beta_3 \times E_i + \varepsilon$$

Equation 1, repeated

For this work, the dependent variable and vectors of structural, neighbourhood, and environmental characteristics were specified as follows:

$$Y_i = \begin{bmatrix} Logged\ Assessed\ Value \text{ or} \\ Logged\ Adjusted\ Value \end{bmatrix} of \begin{bmatrix} Single\ detached \text{ or} \\ Semi\ detached\ \ and\ duplex \\ \text{or } Townhouses \end{bmatrix}$$

$$S_i = \begin{bmatrix} Living\ Area_i \\ Yard\ Size_i \\ Building\ Age_i \end{bmatrix}$$

$$N_i = \begin{bmatrix} In\ CTC_i \\ Rate\ of\ Appreciation_i \\ Education\ Rate_i \\ Population\ Density_i \\ Time\ Period_i \\ In\ CTC\ \times\ Time\ Period_i \end{bmatrix}$$

$$E_i = \begin{bmatrix} Open\ Space\ Access_i \\ Transit\ Access_i \\ Walkability_i \\ Open\ Space\ Adjacent_i \\ Regional\ Road\ Adjacent_i \\ Open\ Space\ Access\ \times\ Yard\ Size_i \\ In\ CTC \times Open\ Space\ Access_i \\ In\ CTC\ \times\ Transit\ Access_i \\ In\ CTC\ \times\ Walkability_i \\ Open\ Space\ Access_i\ \times\ Transit\ Access_i \\ Open\ Space\ Access_i\ \times\ Walkability_i \\ Walkability_i\ \times Transit\ Access_i \end{bmatrix}$$

Equation 26: Final model variables

6 Regression Results

The following sections detail the hedonic model results. This chapter contains four sections, one for each included property type and a final section comparing effects between property types. Each property type section includes four components:

1. The results of OLS regressions without spatial corrections
2. Spatial diagnostic tests of the OLS models
3. The results of combined SLM and SEMs
4. A table showing the estimated average impact that an increase of one standard deviation in variables would have on home value, from the spatial lag and error model

The regression results presented in this chapter include, the coefficient estimates; the significances of estimates determined through t-tests for the OLS and z-tests for the combined spatial lag and error models; and the standard errors of these estimates. The Log Likelihood, AIC, Schwarz Criterion, and Adjusted R^2 are given to evaluate the OLS models. The model diagnostics provided by GeodaSpace are limited for the combined SLM and SEMs, which consist of a Pseudo R^2 and a Spatial Pseudo R^2 to assess model goodness-of-fit.

The summaries following the regression estimates use the approximation of a $100 \times \beta\%$ increase in *Y* due to an increase in *X.* The tables showing the relative contributions of effects as well as the comparative tables in the final section use the more precise formula for interpreting the model log-level coefficients. Model results of level-level models are located in Appendix 5, but not discussed explicitly.

6.1 Single-Detached Models

6.1.1 OLS

* p<0.05, ** p<0.01, *** p<0.001

Single-Detached Models: OLS						*n=21,801*
Variable	**Logged Adjusted Value Model**			*Logged Assessed Value Model*		
	Signif-icance	*Coeffic-ient*	*Std. Error*	*Signif-icance*	*Coeffic-ient*	*Std. Error*
CONSTANT	***	11.9424213	0.0132075	***	12.0697696	0.0095817
Living Area	***	0.0001468	0.0000014	***	0.0001467	0.000001
Yard Size	***	0.0000161	0.0000011	***	0.0000163	0.0000008
Building Age	***	-0.0024567	0.0000887	***	-0.002748	0.0000643
In CTC		-0.0109477	0.0163553	*	0.0264371	0.0118654
ROA	***	0.0070237	0.0004252	***	0.0047906	0.0003085
Education Rate	***	0.0043735	0.0001849	***	0.003518	0.0001341
Population Density	***	-0.0000086	0.0000009	***	-0.0000099	0.0000006
Time Period 2	***	0.087734	0.0035849	*	0.0065514	0.0026007
Time Period 3	***	0.1493403	0.0028793	***	0.0238777	0.0020889
In CTC * Time Period 2		0.0150111	0.012499		-0.0075268	0.0090678
In CTC * Time Period 3	***	0.0601204	0.0103603		0.0144128	0.0075161
OS Access	***	-0.0012119	0.0001551	***	-0.0010781	0.0001125
Transit	***	-0.0000601	0.0000142	***	-0.0000899	0.0000103
Walkability	***	0.009073	0.0018097	***	0.0054726	0.0013129
OS Adjacent	***	0.0469114	0.0037173	***	0.0366586	0.0026968
Regional Road Adjacent	***	-0.0635558	0.0071792	***	-0.0372219	0.0052084
Yard Size * OS Access	***	0.0000003	0	***	0.0000003	0
In CTC * OS Access		0.0000187	0.0003573		-0.0002994	0.0002592
In CTC * Transit		-0.0000005	0.0000269		-0.0000127	0.0000195
In CTC * Walkability	***	0.0209586	0.0020808	***	0.0238081	0.0015096
Transit * OS		0.0000001	0.0000003	***	0.0000007	0.0000002
Walkability * OS Access	***	-0.0002139	0.0000346	***	-0.0001504	0.0000251
Walkability * Transit	***	0.0000097	0.0000018	***	0.0000116	0.0000013
Log Likelihood		6703.817			13700.29	
AIC		-13359.6			-27352.6	
Schwarz Criterion		-13167.9			-27160.8	
Adjusted R^2		0.6424			0.7925	

Table 32: OLS Estimates, Logged Adjusted and Logged Assessed Value, Single-Detached

	Adjusted Value		Assessed Value	
TEST	VALUE	PROB	VALUE	PROB
Moran's I (error)	85.773	0	152.214	0
Lagrange Multiplier (lag)	8296.851	0	17466.9	0
Robust LM (lag)	2242.235	0	3685.792	0
Lagrange Multiplier (error)	7261.854	0	22925.93	0
Robust LM (error)	1207.238	0	9144.823	0
Lagrange Multiplier (SARMA)	9504.089	0	26611.72	0

Table 33: Spatial Dependence Tests, Single-Detached

All tests for spatial dependence show significant results in both the lag and error, indicating that the null of spatial independence can be rejected and that spatial models are appropriate.

6.1.2 Spatial Lag + Error

* p<0.05, ** p<0.01, *** p<0.001

Single-Detached Models ***Spatial Lag + Spatial Error***						***n=21,801***
	Logged Adjusted Value Model			***Logged Assessed Value Model***		
Variable	***Signif-icance***	***Coeffic-ient***	***Std. Error***	***Signif-icance***	***Coeffic-ient***	***Std. Error***
CONSTANT	***	5.4568606	0.2035113	***	6.0347905	0.2268263
lambda		0.0135647	0.0300799	***	0.4805893	0.0214296
W_Dependent Variable	***	0.5386543	0.0170336	***	0.5024415	0.0190217
Living Area	***	0.000095	0.000002	***	0.0000984	0.0000018
Yard Size	***	0.0000098	0.0000026	***	0.0000104	0.0000027
Building Age	***	-0.001268	0.0001649	***	-0.001845	0.0001793
In CTC		-0.0156967	0.0183749		0.025496	0.01629
ROA	***	0.003907	0.0003982	***	0.0014103	0.0002736
Education Rate	***	0.0010407	0.0002397	*	0.0005359	0.0002697
Population Density	***	-0.0000029	0.0000007	***	-0.0000048	0.0000008
Time Period 2	***	0.0859551	0.0031171	*	0.0045867	0.0020261
Time Period 3	***	0.1401735	0.0026694	***	0.0174524	0.0018088
In CTC * Time Period 2		0.0181127	0.0143331		0.0004764	0.0077002
In CTC * Time Period 3	***	0.0451095	0.0116707		0.0032477	0.0064838
OS Access	***	-0.0007493	0.0001916	***	-0.0008352	0.0001784
Transit		-0.0000143	0.0000122	*	-0.0000266	0.0000131
Walkability	*	0.0036828	0.0016904	*	0.0040024	0.0017681
OS Adjacent	***	0.0342894	0.0034006	***	0.0320289	0.0026097
Regional Road Adjacent	***	-0.0408812	0.0075459	***	-0.0280558	0.0049695
Yard Size * OS Access	**	0.0000001	0	***	0.0000001	0
In CTC * OS Access		0.0001902	0.0004126		-0.0001501	0.0003903
In CTC * Transit		-0.0000026	0.0000237		0.0000071	0.0000251
In CTC * Walkability	**	0.0088544	0.0027802	***	0.0111798	0.0027729
Transit * OS		-0.0000004	0.0000003		-0.0000001	0.0000003
Walkability * OS Access		-0.0000613	0.0000333		-0.0000139	0.0000346
Walkability * Transit		0.000001	0.0000016	*	0.000004	0.0000018
Pseudo R^2		0.7569			0.8850	
Spatial Pseudo R^2		0.7035			0.8242	

Table 34: Spatial Regression Estimates, Logged Adjusted and Logged Assessed Value, Single-Detached

Significant Results from Adjusted Value Model

- Walkability has a positive effect that is even greater within the CTC
- Yard Size and OS access have synergistic effect in both models
- Homes sold for an increasing amount over time, and even more so within the CTC after 2011
  - 8.6% and 14% outside of the CTC, and 18.5% inside the CTC after 2011
- OS access had negative impact on property values (-.0075% for every unit of access)

- OS adjacency had a positive impact on property values (+~3.5%)
- Regional Road Adjacency is related to a 4% lower property value
- Walkability and transit only significant in Assessed Value model
- Living Area = +0.0095% per square foot, and Yard Size = +0.00098%
- Every year a building ages is associated with a 0.1% decrease in total property value

<u>Differences between Adjusted Value and Assessed Value</u>

- Lambda (spatially error term) significant only in assessed
- In CTC * Time period 2 only significant in adjusted value model
- Negative effect of distance to transit only significant in assessed value model
- Walkability and transit, positive interaction significant in only assessed model
- Assessed value model estimated a 1.1% increase for each point of walkability for homes within the CTC
- Regional road adjacency only associated with a decrease of 2.8% in assessed values
- Slightly higher Pseudo and Spatial Pseudo R^2 values in assessed value model (roughly 10% more of variance explained)

6.1.3 Relative contributions of effects

Single-Detached Model		Adjusted Value Model		Assessed Model	
Variable	**Stnd. Dev. (or 1 for dummy)**	**Effect per Unit Increase**	**Estimated Effect of 1 Stnd. Dev. (or 1 for dummy)**	**Effect per Unit Increase**	**Estimated Effect of 1 Stnd. Dev. (or 1 for dummy)**
Living Area	1077.0	0.0095%	10.23232%	0.00984%	10.59855%
Yard Size	2325.9	0.00098%	2.27944%	0.00104%	2.419%
Building Age	24.4	-0.12688%	-3.09907%	-0.18467%	-4.5106%
In CTC	1				
ROA	3.1	0.39146%	1.21269%	0.14113%	0.4372%
Education Rate	8.2	0.10412%	0.85238%	0.0536%	0.43882%
Population Density	1681.5	-0.00029%	-0.48763%	-0.00048%	-0.80712%
Time Period 2	1.0	8.97574%	8.97574%	0.45972%	0.45972%
Time Period 3	1.0	15.04734%	15.04734%	1.76056%	1.76056%
In CTC × Time Period 2	1.0				
In CTC × Time Period 3	1.0	4.61424%	4.61424%		
OS Access	17.7	-0.07496%	-1.32414%	-0.08355%	-1.476%
Transit Access	353.1			-0.00266%	-0.93922%
Walkability	2.6	0.36896%	0.94589%	0.40104%	1.02814%
OS Adjacent	1.0	3.48841%	3.48841%	3.25473%	3.25473%
Regional Road Adjacent	1.0	-4.17283%	-4.17283%	-2.84531%	-2.84531%
OS Access × Yard Size	41087.9	0.00001%	0.41088%	0.00001%	0.41088%
In CTC × OS Access	17.7				
In CTC × Transit Access	353.1				
In CTC × Walkability	2.6	0.88937%	2.28005%	1.12425%	2.8822%
OS Access × Transit Access	6237.3				
OS Access × Walkability	45.3				
Walkability × Transit Access	905.2			0.0004%	0.36208%

Table 35: Illustrating Effects of Spatial Regression Estimates, Single-Detached

Structural

A standard deviation of Living Area had nearly five times the effect as a distributionally equivalent increase in Yard Size, which was more than double the amount of a standard deviation of living area in square footage. This difference between Living area and yard size in consistent between the Adjusted and Assessed Models.

Neighbourhood

The CTC's only significant effect in the base and second time period was through a walkability and a KW wide increase in values. Home prices in the CTC In addition to the walkability premium, the third period saw an additional appreciation, more than 4.5%, on top of a KW-wide increase of 15% from the base time period. Adjacency to OS and regional roads had comparable but opposite effects, of 3%

and negative 4% to home value when they are present, respectively. An increase of one standard deviation of neighbourhood density leads to a .5% decrease in values, while a standard deviation increase in the education level of the neighbourhood produces an 8.5% increase. A 3.1% increase in the neighbourhood ROA is estimated to increase values by 1.2%.

Environmental

A 17 point increase in OS access was associated with a 1.3-1.4% decrease in overall home price. Increasing a standard deviation of both yard size and OS access had an additional impact of 0.41%, robust between models. Transit access alone had a negative relationship with value, decreasing value by 0.9% per standard deviation. This negative impact is lessened through the interaction of walkability and transit, which exhibited a modest synergistic impact of 0.36% with a standard deviation increase of both – in addition to the 1% increase in value attributable to the walkability increase alone.

6.2 Semi-Detached and Duplex Models

6.2.1 OLS

*** p<0.05, ** p<0.01, *** p<0.001**

Semi-Detached Models OLS						*n=2,702*
	Logged Adjusted Value Model			***Logged Assessed Value Model***		
Variable	***Signif-icance***	***Coeffic-ient***	***Std. Error***	***Signif-icance***	***Coeffic-ient***	***Std. Error***
CONSTANT	***	11.9224366	0.0385903	***	11.9419498	0.0281173
Living Area	***	0.0001354	0.0000062	***	0.0001326	0.0000045
Yard Size	***	0.0000355	0.0000074	***	0.0000498	0.0000054
Building Age	***	-0.0021239	0.0002125	***	-0.0022472	0.0001549
In CTC	**	-0.1098411	0.0333652	*	-0.0498843	0.0243102
ROA	**	0.0034846	0.0012249	***	0.0031565	0.0008925
Education Rate	***	0.0031086	0.0004606	***	0.0029157	0.0003356
Population Density	***	-0.0000118	0.000002	***	-0.0000115	0.0000015
Time Period 2	***	0.1049256	0.0097962	**	0.0186602	0.0071376
Time Period 3	***	0.1493084	0.0080181	**	0.0166031	0.0058421
In CTC * Time Period 2		0.0079056	0.0234487		-0.0100735	0.017085
In CTC * Time Period 3		0.0352253	0.019833		0.0115748	0.0144505
OS Access	*	-0.00135	0.0006322		-0.0008023	0.0004606
Transit		-0.0000148	0.000065		-0.0000424	0.0000474
Walkability	***	0.0264654	0.0051004	***	0.0212187	0.0037162
OS Adjacent		-0.0055826	0.0120991		-0.0035128	0.0088155
Regional Road Adjacent		-0.0094818	0.0125516		-0.0063189	0.0091452
Yard Size * OS Access		0.0000001	0.0000002		0	0.0000001
In CTC * OS Access	***	0.0025407	0.0007225		0.0006988	0.0005265
In CTC * Transit		-0.0000701	0.0000721		-0.0000686	0.0000526
In CTC * Walkability		-0.0039906	0.0042708	***	0.0185592	0.0031118
Transit * OS		-0.0000014	0.0000014		-0.0000011	0.000001
Walkability * OS Access		-0.0002223	0.000092	*	-0.0001481	0.000067
Walkability * Transit		0.0000123	0.0000113	**	0.0000213	0.0000082
Log Likelihood		1034.922			1890.42	
AIC		-2021.85			-3732.84	
Schwarz Criterion		-1880.2			-3591.2	
Adjusted R^2		0.3691			0.4887	

Table 36: OLS Estimates, Logged Adjusted and Logged Assessed Value, Semi-Detached and Duplex

	Adjusted Value		Assessed Value	
TEST	VALUE	PROB	VALUE	PROB
Moran's I (error)	23.199	0	59.079	0
Lagrange Multiplier (lag)	533.937	0	2924.967	0
Robust LM (lag)	79.752	0	305.872	0
Lagrange Multiplier (error)	480.994	0	3223.538	0
Robust LM (error)	26.808	0	604.442	0
Lagrange Multiplier (SARMA)	560.746	0	3529.41	0

Table 37: Spatial Dependence tests, Semi-Detached and Duplex

All tests for spatial dependence show significant results in both the lag and error, indicating that the null of spatial independence can be rejected and that spatial models are appropriate.

6.2.2 Spatial Lag + Error

* p<0.05, ** p<0.01, *** p<0.001

Semi-Detached Models Spatial Lag + Error						*n=2,702*
Variable	***Logged Adjusted Value Model***			***Logged Assessed Value Model***		
	Signif-icance	*Coeffic-ient*	*Std. Error*	*Signif-icance*	*Coeffic-ient*	*Std. Error*
CONSTANT	***	5.5415364	0.610423	***	4.144149	0.4798546
lambda		-0.168004	0.1002069		0.1519734	0.0818609
W_Dependent Variable	***	0.5261355	0.0506208	***	0.6429866	0.0399456
Living Area	***	0.0001064	0.0000087	***	0.0000904	0.0000058
Yard Size	*	0.0000256	0.0000112	***	0.0000405	0.0000073
Building Age	***	-0.0013109	0.0002554	***	-0.0013495	0.0001858
In CTC		-0.0661635	0.0375496		-0.0186518	0.0264264
ROA		0.0016943	0.0011656		0.0002929	0.0008298
Education Rate	*	0.0011261	0.0004975		0.0009647	0.0005025
Population Density	*	-0.0000042	0.0000018		-0.0000028	0.0000015
Time Period 2	***	0.098911	0.0081876	*	0.0113621	0.0050064
Time Period 3	***	0.1414806	0.0065903	*	0.0095407	0.0042416
In CTC * Time Period 2		0.0089293	0.0281032		-0.0059539	0.0148852
In CTC * Time Period 3		0.031602	0.0255634		0.0138964	0.0140403
OS Access	*	-0.0013831	0.0006984		-0.000651	0.0005133
Transit		0.0000086	0.0000625		-0.0000259	0.0000463
Walkability	*	0.0115648	0.0052742		0.0030113	0.0043415
OS Adjacent		-0.0052872	0.0112421		0.0057208	0.0071997
Regional Road Adjacent		-0.0157127	0.011158	**	-0.0227528	0.0076234
Yard Size * OS Access		0.0000003	0.0000002		0.0000002	0.0000001
In CTC * OS Access		0.0012093	0.0007724		0.0003811	0.0006142
In CTC * Transit		-0.0000758	0.0000635		-0.0000231	0.0000615
In CTC * Walkability		-0.0028776	0.0048492		0.0066459	0.0041554
Transit * OS		-0.0000009	0.0000012		-0.0000001	0.0000009
Walkability * OS Access		-0.000079	0.0001002		0.0000272	0.0000876
Walkability * Transit		0.0000059	0.0000123		0.0000032	0.00001
Pseudo R^2		0.4577			0.7164	
Spatial Pseudo R^2		0.3898			0.523	

Table 38: Spatial Model Estimates, Logged Adjusted and Logged Assessed Value, Semi-Detached and Duplex

Notable Results from Adjusted Value Model

- Living area adds +0.01% per square foot
- Yard size add +0.0026%
- Semi-detached and duplex homes increased in value 9% from 2009-2011, and 14% from 2011 onward, with no significant difference for homes in the CTC
- Every walkability point increases property value of semi-detached and duplex homes by 1%

- Regional road adjacency only had a significant effect on assessed values, reducing property value by 2%

<u>Differences between Adjusted Value and Assessed Value:</u>

- Significant decrease from regional road adjacency only found in Assessed Value Model, at 2.28%
- Significant decrease for OS Access and increase for walkability not found in assessed values
- Assessed Value model had more predictive power (much higher Pseudo R^2 and slightly higher Spatial Pseudo R^2 values)

6.2.3 Relative contributions of effects

Semi-Detached & Duplex Model

		Adjusted Value Model		Assessed Model	
Variable	Stnd. Dev. (or 1 for dummy)	Effect per Unit Increase	Estimated Effect of 1 Stnd. Dev. (or 1 for dummy)	Effect per Unit Increase	Estimated Effect of 1 Stnd. Dev. (or 1 for dummy)
Living Area	562.5	0.01064%	5.98499%	0.00904%	5.08495%
Yard Size	1081.9	0.00256%	2.7698%	0.00405%	4.38195%
Building Age	27.1	-0.13118%	-3.54912%	-0.13504%	-3.65369%
In CTC	1				
ROA	2.8				
Education Rate	8.4	0.11267%	0.94821%		
Population Density	1865.1	-0.00042%	-0.78333%		
Time Period 2	1.0	10.3968%	10.3968%	1.14269%	1.14269%
Time Period 3	1.0	15.19782%	15.19782%	0.95864%	0.95864%
In CTC × Time Period 2	1.0				
In CTC × Time Period 3	1.0				
OS Access	17.2	-0.13841%	-2.38045%		
Transit Access	174.8				
Walkability	2.4	1.16319%	2.79337%		
OS Adjacent	1.0				
Regional Road Adjacent	1.0			-2.30136%	-2.30136%
OS Access × Yard Size	18608.4				
In CTC× OS Access	17.2				
In CTC × Transit Access	174.8				
In CTC × Walkability	2.4				
OS Access × Transit Access	3007.0				
OS Access × Walkability	41.3				
Walkability × Transit Access	419.9				

Table 39: Illustrating Effects of Spatial Regression Estimates, Semi-detached and Duplex

Structural

Living area was overall the strongest determinant in homebuyer WTP. Increasing Yard Size one standard deviation, or twice the area of one standard deviation of Living area, increases WTP by just over half the amount of a standard deviation of Living Area from the adjusted value model. The Assessed Value model found only a 0.7% difference between Living area and Yard Size. Building age estimates were consistent between the two models – with 27 years of age depreciating value by just over 3.5%.

Environmental

Increasing a property's walkability by a standard deviation increase property value by almost 3% from the adjusted value model. A standard deviation increase of OS access was found to decrease adjusted value by just under 2.5%.

Neighbourhood

The significant time period dummies but insignificant ROA variable in the adjusted value model show that the value of semi-detached and duplex homes has been increasing fairly consistently throughout KW.

6.3 Townhouse Models

6.3.1 OLS

* p<0.05, ** p<0.01, *** p<0.001

Townhouse Models OLS						*n=2,370*
Variable	**Logged Adjusted Value Model**			***Logged Assessed Value Model***		
	Signif-icance	*Coeffic-ient*	*Std. Error*	*Signif-icance*	*Coeffic-ient*	*Std. Error*
CONSTANT	***	11.7579403	0.0344831	***	11.8962018	0.0239846
Living Area	***	0.000179	0.0000061	***	0.0001709	0.0000043
Yard Size	***	0.000044	0.0000087	***	0.0000676	0.000006
Building Age	***	-0.0044361	0.0002942	***	-0.0063694	0.0002046
In CTC	***	-0.2742463	0.0622748		-0.0823936	0.043315
ROA	***	0.0050086	0.0010375	**	0.0020429	0.0007216
Education Rate	***	0.0024022	0.0004658		0.0003581	0.000324
Population Density	***	-0.000007	0.0000017	**	-0.000003	0.0000012
Time Period 2	***	0.1054031	0.0070062	***	0.0250743	0.0048731
Time Period 3	***	0.1845599	0.0058958	***	0.0569646	0.0041008
In CTC * Time Period 2	***	0.2098331	0.048034	*	0.0855336	0.0334099
In CTC * Time Period 3	*	0.0840447	0.0381874		-0.0383044	0.0265611
OS Access	***	0.0014432	0.0004123	***	0.0023002	0.0002868
Transit	*	-0.0000706	0.0000316	***	-0.0001159	0.000022
Walkability		0.0032119	0.0035811		-0.0012933	0.0024908
OS Adjacent		-0.0074563	0.0083854		0.0027914	0.0058325
Regional Road Adjacent		-0.0082023	0.0113383		-0.0090952	0.0078863
Yard Size * OS Access		-0.0000001	0.0000002	**	-0.0000004	0.0000001
In CTC * OS Access	***	0.0039969	0.0009164	**	0.002097	0.0006374
In CTC * Transit	**	-0.000138	0.0000427	***	-0.0001304	0.0000297
In CTC * Walkability		0.0129568	0.0070575	***	0.0207931	0.0049088
Transit * OS		-0.0000004	0.0000007	*	0.0000012	0.0000005
Walkability * OS Access	*	0.0001613	0.0000671		0.0000442	0.0000466
Walkability * Transit		-0.000003	0.0000039		0.0000018	0.0000027
Log Likelihood		1793.618			2654.065	
AIC		-3539.24			-5260.13	
Schwarz Criterion		-3400.74			-5121.63	
Adjusted R^2		0.6424			0.7572	

Table 40: OLS Estimates, Logged Adjusted and Logged Assessed Value, Townhouse

	Adjusted Value		Assessed Value	
TEST	VALUE	PROB	VALUE	PROB
Moran's I (error)	31.312	0	55.993	0
Lagrange Multiplier (lag)	672.931	0	1433.253	0
Robust LM (lag)	76.536	0	105.075	0
Lagrange Multiplier (error)	868.914	0	2840.209	0
Robust LM (error)	272.519	0	1512.031	0
Lagrange Multiplier (SARMA)	31.312	0	2945.284	0

Table 41 - Spatial Dependence Test, Townhouse

All tests for spatial dependence show significant results in both the lag and error, indicating that the null of spatial independence can be rejected and that spatial models are appropriate.

6.3.2 Spatial Lag + Error

* p<0.05, ** p<0.01, *** p<0.001

Townhouse Models Spatial Lag + Error						*n=2,370*
Variable	**Logged Adjusted Value Model**			**Logged Assessed Value Model**		
	Signif-icance	*Coeffic-ient*	*Std. Error*	*Signif-icance*	*Coeffic-ient*	*Std. Error*
CONSTANT	***	6.6386987	0.5974548	***	7.9680331	0.6917784
lambda	***	0.3498933	0.0839169	***	0.7189491	0.048845
W_Dependent Variable	***	0.4316912	0.0499763	***	0.3386007	0.0569221
Living Area	***	0.0001171	0.0000119	***	0.0001252	0.0000103
Yard Size	***	0.0000284	0.0000082	***	0.0000456	0.0000061
Building Age	***	-0.0026132	0.0006259	***	-0.0056717	0.0007148
In CTC	*	-0.2362981	0.1058904		-0.1222465	0.100804
ROA		0.0013941	0.0013771		-0.0013697	0.0007935
Education Rate		0.0009733	0.0006343		-0.0005133	0.0007731
Population Density		-0.0000034	0.0000026		-0.0000051	0.0000037
Time Period 2	***	0.1077929	0.0063303	***	0.0224376	0.0038137
Time Period 3	***	0.1743264	0.0062539	***	0.0420147	0.0050688
In CTC * Time Period 2	**	0.1694451	0.056967		0.0594832	0.0446404
In CTC * Time Period 3		0.0981846	0.0598786		-0.009348	0.0345832
OS Access		0.0005395	0.0004164		0.0003705	0.0003746
Transit		-0.0000361	0.0000328		-0.000046	0.0000332
Walkability		-0.0008717	0.0046976		0.0051223	0.0061773
OS Adjacent		-0.0119303	0.0127498		0.0066897	0.0059965
Regional Road Adjacent		-0.011694	0.0120728		-0.0065325	0.0114404
Yard Size * OS Access		0.0000001	0.0000002		-0.0000004	0.0000001
In CTC * OS Access		0.0028117	0.0019103		0.0021795	0.0019442
In CTC * Transit		-0.0000563	0.000052		-0.0000512	0.0000753
In CTC * Walkability		-0.0027232	0.0171991		0.00207	0.0147908
Transit * OS		-0.000001	0.0000007		-0.0000009	0.0000007
Walkability * OS Access	*	0.0001829	0.0000915		0.0000583	0.000106
Walkability * Transit	***	-0.0000075	0.0000042		-0.0000023	0.0000052
Pseudo R^2		0.7147			0.8204	
Spatial Pseudo R^2		0.6423			0.7346	

Table 42: Spatial Model Estimates, Logged Adjusted and Logged Assessed Value, Townhouse

Notable Results from Adjusted Value Model

- Townhouses increased in value over time, increasing by 10.8% between 2009 and 2011, and by 17.4% after 2011. Townhouses in the CTC increased by an additional 16% between 2009 and 2011, for a total of 26.8% from pre LRT.
- Walkability and OS access together increased townhouses' value, but transit access and walkability decreased it
- Townhouses in the CTC sold for less than outside of it in all time periods *ceteris paribus*

Differences between Adjusted Value and Assessed Value

- Negative walkability and transit and positive walkability and OS interactions only significant in adjusted value models
- Significant negative effect for In CTC in time period 1 and positive effect in CTC in time period 2 only found in adjusted value model
- Slightly higher Pseudo and Spatial Pseudo R^2 values in assessed value model (roughly 10% more of variance explained)

6.3.3 Relative contributions of effects

Townhouse Model

		Adjusted Value Model		Assessed Model	
Variable	**Stnd. Dev. (or 1 for dummy)**	**Effect per Unit Increase**	**Estimated Effect of 1 Stnd. Dev. (or 1 for dummy)**	**Effect per Unit Increase**	**Estimate d Effect of 1 Stnd. Dev. (or 1 for dummy)**
Living Area	480.5	0.01171%	5.62727%	0.01252%	6.01654%
Yard Size	750.6	0.00284%	2.1316%	0.00456%	3.4226%
Building Age	10.3	-0.26166%	-2.68504%	-0.56878%	-5.83655%
In CTC	1	-26.6552%	-26.6552%		
ROA	2.6				
Education Rate	7.2				
Population Density	2203.7				
Time Period 2	1.0	11.3817%	11.3817%	2.26912%	2.26912%
Time Period 3	1.0	19.04441%	19.04441%	4.29098%	4.29098%
In CTC × Time Period 2	1.0	18.46473%	18.46473%		
In CTC × Time Period 3	1.0				
OS Access	18.1				
Transit Access	293.8				
Walkability	2.6				
OS Adjacent	1.0				
Regional Road Adjacent	1.0				
OS Access × Yard Size	13614.0				
In CTC × OS Access	18.1				
In CTC × Transit Access	293.8				
In CTC × Walkability	2.6				
OS Access × Transit Access	5329.9				
OS Access × Walkability	46.9	0.01829%	0.85872%		
Walkability × Transit Access	760.5	-0.00075%	-0.57039%		

Table 43: Illustrating Effects of Spatial Regression Estimates, Townhouse

Structural

A standard deviation increase of Living area had aver twice the value impact of a proportionally similar increase in yard size for townhouses – with 270 square feet of yard size more than living area. A building age of just over 10 years decreased adjusted values by around 2.5% and assessed values by nearly 6%.

Neighbourhood

Education rate, population density, and ROA were all insignificant in the townhouse model. Compared to pre-2009 levels, homes increase by 11%, and 18% outside of the CTC. Within the CTC, homes sold for 26.65% less than outside of it pre-2009 and post 2011, but only 8.2% between 2009 and 2011.

Environmental

In the adjusted value model, the effect of increasing all intensification variables one standard deviation would be a net increase of ~0.3%, with a positive effect from OS access's and a negative from transit's interactions with walkability.

6.4 Comparison of Results between Property Types

This sub-section summarizes significant results from the models run on logged adjusted value, which will be used to answer research questions of this thesis in the following chapter. The following tables uses exponentiation to get the true value of the estimates. This table shows the estimated percent change in value from a base of homes outside of the CTC before regional LRT was initially approved in 2009.

6.4.1 In CTC over time comparison

Significant CTC and Time Variables (Logged Adjusted Value)	In CTC	Time Period 2	In CTC * Time Period 2	Time Period 3	In CTC * Time Period 3
Single Detached		8.97574%		15.0473%	4.61424%
Semis & Duplexes		10.3968%		15.1978%	
Townhouses	-26.6552%	11.3817%	18.4647%	19.0444%	

Table 44: Comparing CTC and Time Estimates by Property Type

This table shows the sub-regional effects of time on home values in KW, after accounting for regional appreciation effects through a home price index adjustment and accounting for neighbourhood level effects through the ROA. All property types across KW sold for more after the LRT was announced in 2009. Townhouses in the CTC were found to be less valuable than those outside before 2009 and after 2011; however, townhouses sold in the CTC between 2009 and 2011 were found to increase in value modestly. Overall for townhouses, the greatest positive effects over the course of the LRT's development were outside of the CTC. Single-detached homes, alternatively, saw the largest increase within the CTC after LRT was accepted as the preferred mode and an implementation strategy was established in 2011. Semi-detcahed and duplex homes saw a similar increase in both the CTC and the rest of KW throughout the approval process of LRT.

6.4.2 Intensification variables comparison

Significant Intensification Variables (Logged Adjusted Value)	Single Detached	Semis & Duplexes	Townhouses
OS Access	-0.074928%	-0.138095%	
OS Access* In CTC			
Transit Access			
Transit Access* In CTC			
Walkability	0.368979%	1.163213%	
Walkability * In CTC	0.889331%		
OS Access* Transit			
OS Access* Walkability			0.018302%
Transit* Walkability			-0.00075%

Table 45: Comparing Intensification Estimates by Property Type

The above table shows the percent change in adjusted value attributable to a one unit increase in the intensification variables, for each property type. Walkability and OS access were found to have significant effects on the adjusted value of all included property types. Walkability significantly increased single-detached, semi-detached, and duplex homes values across KW, with greater effects in the CTC for single-detached homes. In townhouses, walkability and OS were only found to only have interactional effects – where walkability it increased values coupled with OS access, but decreased them with transit access. OS access was associated with a significant decrease in values of single-detached, semi-detached, and duplex homes.

6.4.3 Model fit comparison

Comparison of Model Fit	*Adjusted R^2 – OLS*		*Pseudo R^2 – Spatial Lag + Error*		*Spatial Pseudo R^2 – Spatial Lag + Error*	
Model:	*Adjusted*	*Assessed*	*Adjusted*	*Assessed*	*Adjusted*	*Assessed*
Single-detached	0.6424	0.7925	0.7569	0.8850	0.7035	0.8242
Semi + duplex	0.3691	0.4887	0.4577	0.7164	0.3898	0.523
Townhouse	0.6424	0.7572	0.7147	0.8204	0.6423	0.7346

Table 46: Comparing Model Fit

Models run on assessed value produced a stronger fit than those run on adjusted values by all measures, for all property types. The spatial models of single-detached and semi-detached and duplex models had a greater fit than OLS according to both Pseudo and Spatial Pseudo R^2 values. The Spatial Pseudo R^2 of the townhouse models was lower than its Adjusted R^2, but the non-spatially corrected Pseudo R^2 of the townhouse model was higher than its Adjusted R^2, which is also uncorrected for spatial endogeneity effects.

7 Conclusions

As is evident from the results found throughout this thesis, the residential land-market is influenced by a number of complex and simultaneous processes. The section uses the findings from this thesis to answer specific questions regarding the residential land-market in KW and the models used to quantify it. A section then provides implications of this work for planners and actors in the development community. Finally, this thesis concludes with suggested next steps and areas for improvement in the modelling work presented here.

7.1 Research Questions

The following section summarizes findings of the research questions posed at the beginning of this work. Broader findings related to these questions are found throughout this thesis.

Answering Research Questions:

7.1.1 Primary:

<u>Are homebuyers willing to pay for OS access, walkability, and proximity to public transit in KW, and if so, how much?</u>

Hypothesis: Homebuyers express a positive WTP for these intensification amenities, with walkability providing the greatest effect.

Findings: There is little consistency between homebuyer WTP and these three variables amongst property types. Walkability was estimated to effect a 0.37% increase in homebuyer WTP per point for single-detached homes, and an increase of 1.16% increase for every point in semi-detached and duplex homes, in adjusted value. This walkability premium increases in the CTC for Single-Detached homes, where they see and additional increase of .9% for every walkability point compared to those outside of the CTC.

OS access by itself was found to be negatively related to property value in single-detached, semi-detached, and duplex homes – an estimated decrease in value of 0.075% per point from the gravity-based OS access measure for single-detached homes, and decreases of 0.138% per OS access point in semi-detached and duplex homes. However, this negative OS access value is partially offset by a positive interaction effect between public and private OS access, discussed under secondary Research Question 4.

Townhouses only saw significant price effects of these intensification variables through interactions between them, which will be discussed with secondary Research Question 5.

Secondary:

1. How are the characteristics commonly associated with home price distributed across KW? Are intensification-related features correlated with each other, neighbourhood

socioeconomic characteristics, or structural home characteristics in regional home sales? If so, is there a spatial pattern that explains this?

Hypothesis: Home characteristics are distributed in a spatially complex manner. Intensification related amenities are co-located in central, denser, older areas and will be highly correlated with structural property characteristics, and to a lesser degree to neighbourhood socioeconomic make-up.

Findings: See chapter 4 for a fuller discussion of the spatial characteristics of the housing market, which is summarized here. Structural home price determinants tend to follow a pattern of centrality, where living area, frontages, yard sizes, and building age decrease moving away from the centre. Socioeconomic neighbourhood factors tend to follow a loosely north-south pattern, with some degree of centrality also having an effect. In terms of environmental characteristics, walkability, transit access, and employment access generally decrease moving away from the centre representing the effects of suburbanization on urban form. The pattern of OS access is complex, with greater access to a variety of open spaces in central areas and access to a more area of open space in the suburbs.

2. How does the characterization of OS access via a gravity-based model compare to those of spatial separation and cumulative opportunities models?

Hypothesis: Gravity-based models provide a more holistic measure of OS access that favours neither outlying areas with access to large OS nor central areas with access to many small OSs.

Findings: See chapter 4 for a discussion of the various access measures investigated. Overall, cumulative opportunities models using the area of accessible OS provide higher results at the rural-urban fringe. Cumulative opportunities models using the count of accessible OSs favour central locations. Spatial separation model results are highest in the outlying areas and adjacent to OSs. Gravity-based models incorporate open space nearness, variety, and abundance, and therefore produce a pattern that to some extent reflects the measures of the cumulative opportunity and spatial separation models.

3. How do results of hedonic models using assessed or observed transaction values differ?

Hypothesis: Assessed value models will behave more predictably using common home price determinants than transaction data, while transaction value models will be able to estimate values related to a more spatially, temporally, and behaviourally complex set of home price determinants.

Findings: Models run on assessed values are easier to predict using commonly employed property characteristics, as indicated by generally higher spatial and pseudo R^2 than the adjusted R^2 values of the OLS models. Although, townhouses had similar levels of fit between OLS and Spatial Regressions. OLS models run on assessed values were found to exhibit much higher levels of spatial dependence than those on adjusted values.

4. Are homebuyers willing to trade-off public for private OS?

Hypothesis: Private OS and public OS have a negative interaction effect, representing a trade-off in homebuyer decision making.

Findings: Contrary to the hypothesis, yard size and OS access were found to have a significant positive interaction effect in the single-detached model. This suggests that rather than a trade-off between public and private OS, homebuyers place additional value on having access to both private and public OS.

5. Do transit access, public OS access, and walkability have synergistic impacts on homebuyer WTP?

Hypothesis: Homebuyers are willing to pay more for housing when multiple intensification amenities are present.

Findings: Significant synergistic impacts on adjusted value between intensification variables were found in only the model of townhouses. In the townhouse model, walkable areas were found to increase value in locations with higher OS access, but to decrease value with nearness to bus stops.

6. Are homebuyers' stated preferences consistent with their location choice decisions?

Hypothesis: Homebuyers who state that OS and transit amenities were important in their decision to move have greater access to those amenities.

Findings: Homebuyers' stated preferences appear to be consistent with their revealed preferences. Homebuyers who stated that transit or OS access were at least somewhat important in their decision to move to their current neighbourhood were closer to transit or had greater access to OS, respectively.

7. Have property prices increased throughout the planning of the regional LRT, and, if so, is this effect greater in the CTC?

Hypothesis: Homes in the CTC receive a premium that has increased at throughout the approvals of plans for LRT development.

Findings: It is evident that home prices have generally increased throughout the planning for the region LRT. Between 2009-2011, single-detached homes saw and increase of approximately 9%. After 2011, Single-detached homes saw an increase of 15% compared to pre-2009 values, with an additional 4.6% in the CTC. Semi-detached and duplex homes saw an increase of almost 10.5% between 2009 and 2011, and an increase of just over 15% after 2011, with no significant difference for homes within the CTC. Overall, townhouses sold for 11.4% more between 2009 and 2011, and 19% more after 2011, with an additional increase in value within the CTC of over 18% in the CTC between 2009 and 2011.Townhouses within the CTC before 2009 sold for 26% less than those outside of it, which was mitigated in the following period through a positive interaction.

8. *Do buyers of different types of homes express a different WTP for intensification-related variables?*

Hypothesis: Homebuyers of single-detached homes will express a smaller WTP for walkability, transit access, and OS access than buyers of semi-detached homes, duplexes, and townhouses.

Findings: Walkability was found to increase and OS to decrease values of single-detached, semi-detached, and duplex homes to varying degrees. Townhouses were found to have a negative interactive effects of transit and walkability but a positive interaction of walkability and OS access. No property type was significantly affected by solely transit access, although the transit access measure used was potentially too simplistic, as is discussed in the final section of this chapter.

9. *How do estimates differ after accounting for spatial effects, is model fit improved?*

Hypothesis: Spatial models will be able to give stronger, unbiased estimates, and outperform non-spatial models in terms of goodness-of-fit

Findings: Model fit was generally improved through the spatial models, determined by comparing the adjusted R^2 of the OLS models to the Pseudo and Spatial Pseudo R^2 pof the spatial models. However, OLS models generally showed stronger estimates of the spatial variables and interactions than the spatial models. This is discussed in more detail in the following section.

7.1.2 General Discussion:

7.1.2.1 Home characteristics in light of urban form and residential preferences

There are two clearly identifiable patterns in the spatial distribution of housing characteristics in the region that are consistent with the theoretical understanding of urban form presented in the literature review. Generally, in an imperfect and spatially heterogeneous way, patterns of centrality and a north-south variation were present in KW. The centrally oriented pattern is consistent with the AMM model, where historical demand for central land contributed to the development of smaller houses. It would be negligent to disregard the role of building age in this pattern, where the CTC contains older homes. Generally, older homes are smaller, as preferences for larger and larger homes grew over time (DeFields, 2013). While land values may be higher within the centre, due to its accessibility benefits, the value of the homes themselves is generally less than in outlying areas. This can be seen as predominantly the effect of building age, which can indicate the need for repairs or renovations, alongside home sizes. Once these two factors are controlled for, it becomes evident that many qualities of the central area, including intensification amenities, actually contribute more to home values than outside of the centre.

The north-south variation in home sales prices could potentially have begun as a municipal difference between KW, or a difference in public amenities at the local level according to the Tiebout Model; however, this research does not investigate the cause of these effects. Therefore, it is unclear whether the north-south variation in home values emerged due to a municipal-scale socio-economic effect, or whether these municipal-scale effects emerged over time as the product of accumulative heterogeneous location choices.

7.1.2.2 Spatial Models

The spatial models run on the larger dataset generally fared better than their OLS counterparts, in terms of model goodness-of-fit. As was expected from the review of the literature, the OLS models produced more significant estimates for the spatial variables than did the explicitly spatial models. In the OLS models, most of the spatial variables and interactions are significant, but in the spatial models they are rarely significant. When these spatial effects are left out of the regression, even while accounting for local spatial effects via environmental and neighbourhoods variables, a significant degree of spatial endogeneity remains.

A related, unresolved question for the researcher though, is whether spatial models might actually obscure the portion of property values' variance that could reasonably be attributable to the actual spatial variables and the interactions of these variables. It may be that while spatial models do provide sounder overall estimates of total property value, that these models might make it more difficult to discern discrete impacts that are due to specific neighbourhood scale spatial processes. In spatial models, could the impacts of a property's environmental characteristics be absorbed by the spatial lag, rather than be found within through the spatial variables themselves? Perhaps spatial models are most appropriate when the intent of a model is to control for spatial effects, rather than to specifically test for effects of inherently spatial variables.

7.1.2.3 Assessed vs. Adjusted Value

As stated in the literature review, regressions run on assessed value often provide a dataset more consistent with theory than those run on transaction values. This is often because assessed values are not influenced by the unobserved heterogeneous preferences of homebuyers. This was confirmed in this work, where *Assessed Values* were easier to predict than *Adjusted Values,* as indicated by the consistently stronger model fit. Interestingly, *Assessed Values* exhibited a higher degree of spatial dependence according to the provided tests. This likely stems from the use of nearby sales in assessors' valuation models, where spatial effects are included in the A*ssessed Values* by design.

7.1.2.4 Correlations vs. Regression Estimates

This work brings to light the importance of considering the model specification holistically rather than solely through its constituent elements. While some individual variables like *Regional Road Adjacency* were found to have unexpected bivariate relationships with home values on their own, when the model was specified with adequate controls the estimated relationships matched expectations. This highlights the potential for correlatory studies to provide misleading results. After accounting for neighbourhood quality and structural variables, it was found that *Regional Road Adjacency* did in fact represent the expected relationship with property values, but that this relationship was confounded by another factor.

7.1.2.5 *Heterogeneous Preferences*

There is clear heterogeneity in the impacts of various intensification amenities on homebuyer WTP. First, there is a some difference between the preferences for these amenities inside and outside of the CTC . Second, homebuyers' WTP for intensification-related features was also found to change depending on the type of property being purchased. The heterogeneity here highlights the importance of considering self-selection in hedonic models. It is well-understood that residential location choice is dependent on resident preferences. This work highlights the complex nature of residential preferences, which are not stationary across space or property types. Because of the greater number of homes outside of the CTC that are predominately single-detached, results obtained from a model that did not estimate these WTP values separately would have been biased toward the base case (single-detached homes outside of the CTC).

7.2 Planning and Development Implications

The first sub-section here provides the potential implications of this research on policy decisions from the perspective of urban and regional planning. The next outlines the potential for this work to be applied in residential development decisions made within the private sector.

7.2.1 Planners

Planners often aim to create vibrant, walkable, amenity-rich communities that support everyday transportation options besides the private automobile. The results of this thesis indicate that there is a generally positive relationship between walkable neighbourhoods and property values. As well, the models here showed positive impacts of time for homes in the CTC – indicating a potentially growing demand for central housing alongside the planning of the regional LRT. On one hand, the potential municipal value-capture from the public expenditures necessary to create these central communities, in the form of property taxes, are often framed in a positive light. However, there is potential for this positive relationship to play a role in gentrification and housing affordability issues.

Gentrification is high on the agenda of management issues associated with intensification. Providing affordable housing in these transit-served, walkable neighbourhoods is doubly important if considering the role of public transit as a public service, able to connect low- and middle-income residents with spatially mismatched jobs, as opposed to a city-building tool to attract a relatively wealthier creative class. In reality, the role of transit is not so dichotomous, where it has the potential to both attract this creative class while serving the needs of vulnerable populations.

However, this current research does not assist in identifying causal patterns. That is to say, this work cannot discern whether amenity-rich neighbourhoods become more valuable due to these amenities or whether the neighbourhoods that are more valuable receive a disproportionate amount of these amenities. These concerns should be at the forefront of decisions made about the development of intensified environments.

7.2.2 Development Community

Understanding homebuyer preferences and the spatial distribution of neighbourhood and environmental characteristics is essential in the pursuit of successful land development. Land development models rely on an assumption of market demand. However, little work has been done to identify the relative effects of environmental characteristics on homebuyer demand in KW. This work can be used to compare locations for development of different housing types, to decide whether investments in certain locations can be recouped in the selling price of homes; however this work is descriptive and not predictive, in that the values for these amenities in the time-context of the data do not necessarily provide relevant information in determining the current market value of developments.

Walkability was the most consistent amenity studied, providing broad positive effects to single-detached, semi-detached, and duplex homes. Homebuyers of townhomes will pay more for housing that is in walkable neighbourhoods with OS access. Developers are able to implement on-site measures to improve walkability and public amenity access, including the provision of semi-public OS, which would be attractive to prospective buyers. Walkability was found to be even more valuable within the CTC than outside of it. Townhouses seem to appeal more to suburban buyers than central buyers, but this may be changing with the development of LRT.

7.3 Limitations and Future Work

This hedonic modelling work is part of a larger integrated land use transport ABM. The hedonic models here provide a cross-sectional glimpse at homebuyer preferences in KW pre-LRT. However, as the LRT is constructed and becomes operational, there are a few specific points of consideration in updating these models.

7.3.1 Causality

A key component of the larger ABM project is identifying and modelling causality in the urban system. As has been iterated in this thesis, the cross-sectional models used here are able to identify only associations between property values and intensification characteristics. In the future, it is advised that time series or panel data methods be used to estimate the causal impacts of LRT and intensification in station areas. Difference-in-differences estimation provides the potential to identify causal impacts, and it is recommended in future work.

The current model specification does include an estimable effect for time in the time period dummies; however, these time variables are still not able to identify causal effects. To determine causality demands the collection of home characteristic data over time. While models testing impacts of LRT development are possible using time-invariant control variables, these models are unable to distinguish individual impacts when multiple processes inducing change occur simultaneously. If the interest of future work is to tease out distinct land-market impacts of the LRT and of the other elements of intensification that accompany it, new variables must be created that account for changing accessibility gradients and amenity locations over time.

7.3.2 Rate of Appreciation (ROA)

While the ROA variable in this work did generally perform as expected in the regressions, it is recognized that other methods of calculating home appreciation exist and may be more appropriate. Specifically, the repeat-sales approach provides the potential to estimate property appreciation while controlling for variation in home characteristics. In this method, only homes sold multiple times would be considered. With the assumption that no major renovations or deterioration have occurred, this would ensure that the estimates of the appreciation variable are not biased by variation in the homes that are sold over time.

7.3.3 Employment Access

Employment access is generally a core component of hedonic models developed under the AMM framework. However, the employment access variable attempted for this work captured neither neighbourhood level variation nor access to employment outside of KW. Instead, this variable was found to be highly correlated with other intensification-related factors, which made the estimation of individual WTP values infeasible. While an aggregate employment access measure generally suffices in hedonic models, the intent of this specific work, where many variables correlate strongly with distance to centre, requires a variable with more disaggregate variation. A cumulative opportunity model, as was evident in those developed for OS in this research, provides very different results with changes in the distance threshold applied. With a dataset of employment access across a farther extent, a smaller-scale variable – perhaps a gravity-based measure from each property – may better incorporate these elements.

7.3.4 Statistical spatial and categorical data analysis

While this work conducted an in-depth, qualitative overview of the categorical and spatial distributions of home characteristics and values in Chapter 5, future work would be better supported through the use of more advanced statistical methods. While the qualitative analysis and descriptive statistics presented do provide sufficient information to reasonably parametrize the final hedonic models, supplemental inferential statistical analysis would add greater depth to the research. Cluster analysis and principal components could provide interesting insights on the spatial and categorical patterns of the data to better understand heterogeneity between groups.

7.3.5 Multifamily Homebuyers

The intent of this research is to identify the residential preferences of owner-occupiers, and it was likely that multifamily properties in the dataset often represented multifamily buildings inhabited by many different households. Therefore, multifamily properties were excluded from the analysis. Due to the wide range of multifamily dwellings, it was beyond the scope of this work to accurately estimate unit sizes rather than entire building sizes and to discern which multifamily sales observations to use. Future work should generate a set of criteria for the identification of individual unit sales within the multifamily sales data in order to effectively estimate homebuyer preferences.

Since this work is centred around estimating the land-market impacts of intensification, modelling multifamily homebuyers, like condo owners, would be a pertinent objective of future research. It is expected that denser styles of housing would see a stronger relationship with intensification amenities than less dense homes. It is presumed that buyers of multifamily units are less likely to own private automobiles and access to private yards, and would therefore have greater preferences for walkability, transit access, and open space access.

7.3.6 Expand the spatial extent of data

In this work, variables were only calculated using data on Kitchener and Waterloo. The access metrics specifically suffer from an edge effect, where those properties on the edges may seem to have less accessibility simply because opportunities outside of KW are omitted. In the future, these access models should be calculated using observations that extend beyond the study area.

Furthermore, the study area of the regression models should be expanded to include all of the ROW. The results of models using only KW home sales do not capture potentially different preferences in Cambridge or in the rural townships of the ROW. This is important to consider, as it would be expected that preferences for intensification amenities may be lower in these contexts. Cambridge is not expected to receive LRT until the second phase of development; therefore, individuals with an immediate desire for intensification may be more likely to choose Kitchener or Waterloo to live. In the same vein, the rural townships are likely inhabited by a population that prefers a less intensive environment.

7.3.7 Spatial Weighting

Different spatial weighting schemes can have a significant influence on the estimates of hedonic models. In the future, models may be developed that utilize distance based weights rather than a k-nearest neighbours approach. In this way, the spatial weights can take into account the nearness of neighboring observations, which is a better approximation of Tobler's first law of geography, and as such may be more apt to control for the effects of spatial autocorrelation.

7.3.8 Data and Data Quality

Further work should investigate the sources of model error and strategies to further increase model goodness-of-fit. Maps to inspect the distribution of error showed no apparent spatial pattern. It is likely that the remnant model error is attributable to data quality issues and potential omitted variables.

Data quality issues include an estimated rather than an accurately measured living area. As well, some structural attributes were unavailable to the researcher. It is likely that these attributes, namely number of bathrooms and building quality, are highly correlated with living area – where larger homes generally have more rooms and are of greater quality. As such, it is expected that the coefficient on living area is somewhat biased, and includes some of the effects attributable to these omitted structural characteristics. Relatedly, the building age variable used in this work did not account for home renovations or remodelling, which has potentially biased results where these renovations are correlated with the included structural characteristics.

The level of service provided by transit is not necessarily captured through a measure of distance to a public transit stop. In the future, a transit access variable could be computed in a way that accounts for the level of service, with measures including transit frequency and connectedness of the route to important or popular destinations. This could then change the transit access measure from representing 'access to transit' to instead represent 'access by transit', which might present different, interesting results. Additionally, transit stop locations are a much less permanent amenity than the other included amenities. Bus stops move frequently, and it is likely that throughout the years of the sample used in this thesis access to transit itself changed, which is unaccounted for in the results.

7.3.9 Translating Hedonic Estimates to WTP

While hedonic estimates themselves are often said to represent WTP, this is not exactly the case. Hedonic estimates of sales data essentially represent the result of buyer-seller negotiation, and the point of intersection between WTP and WTA. Consumer surplus is used to explain the difference between prices and WTP. In recognition of consumer surplus, future work should investigate methods to account for heterogeneous differences in the hedonic model estimates provided here and actual homebuyer WTP.

The estimates of this thesis could serve as a lower bound in WTP, while an upper bound could be derived from stated preference data. Additionally, confidence intervals of the estimates may also be used to help derive a distribution of homebuyer WTP, which could be employed in the WARM ABM to stochastically generate distributions of heterogeneous homebuyer WTP values.

7.3.10 Archetypical Examples

In future work, the interpretation of results of this thesis could be supplemented through the development of archetypical models of regional homes. These archetypical homes could be used to further understand heterogeneity between property types and buyers by investigating potential varied effects based on different types of homes and different buyers who buy them. Archetypes could either be derived from the data itself, using statistical methods such as cluster analysis, or could be based on theoretical typologies representing common properties and local housing styles.

7.3.11 Demographic heterogeneity

This research began as an attempt to understand the impacts of homebuyer heterogeneity on WTP. Ultimately, after combining the available survey data to the transactions and cleaning it, the decreased sample size and low variation in survey respondents demographics made regression infeasible. An interesting and important area of future work should be in understanding the effects of demographic heterogeneity on WTP. Life-cycle stage and household composition are thought to vastly affect WTP, which may have moderating effects of homebuyer WTP for intensification amenities.

REFERENCES

Adair, A., S. McGreal, A. Smyth, J. Cooper and T. Ryley (2000). House prices and accessibility: The testing of relationships within the Belfast urban area. Housing Studies 15(5): 699-716.

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

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

Anselin, L. (1988). Spatial econometrics: Methods and models. Dordrecht: Kluwer Academic

Anselin, L., & Bera, A.K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economic Statistics.

Antanaitis, A. (2014). Developing Up and not Out: Understanding the Barriers to and Opportunities for Reurbanization along Waterloo's CTC (Master's Thesis, University of Waterloo). Retrieved form https://uwspace.uwaterloo.ca

Antoniou, C. and Picard, N. (2015). Urban sustainability and individual/household well-being, in A. Michelangeli (ed.), Quality of Life in Cities. Equity, Sustainable Development and Happiness from a Policy Perspective. Routledge, pp. 110-142.

Anas, A., Arnott, R., & Small, K. (1998). Urban spatial structure. *Journal of Economic Literature* 36(3): 1426-1464.

Atkinson-Palombo, C. (2010). Comparing the capitalisation benefits of light-rail transit and overlay zoning for single-family houses and condos by neighbourhood type in metropolitan Phoenix, Arizona. *Urban Studies (Edinburgh, Scotland)*, *47*(11), 2409–2426. http://doi.org/10.1177/0042098009357963

Baar, J., Romppel, M., Igel, U., Brähler, E., & Grande, G. (2014). The independent relations of both residential self-selection and the environment to physical activity. *International Journal of Environmental Health Research*, *25*(September), 1–11. http://doi.org/10.1080/09603123.2014.938029

Bartholomew, K., & Ewing, R. (2011). Hedonic Price Effects of Pedestrian- and Transit-Oriented Development. *Journal of Planning Literature*, *26*(1), 18–34. http://doi.org/10.1177/0885412210386540

Becker, G. (1965). A Theory of the Allocation of Time. *The Economic Journal*. 75(299), 493-517

Billings, S. B. (2011). Estimating the value of a new transit option. *Regional Science and Urban Economics*, *41*(6), 525–536. http://doi.org/10.1016/j.regsciurbeco.2011.03.013

Boone-Heinonen, J., Jacobs, D. R., Sidney, S., Sternfeld, B., Lewis, C. E., & Gordon-Larsen, P. (2009). A Walk (or Cycle) to the Park. Active Transit to Neighborhood Amenities, the CARDIA Study. *American Journal of Preventive Medicine*, *37*(4), 285–292. http://doi.org/10.1016/j.amepre.2009.06.006

Bowen, W. M., Mikelbank, B. a., & Prestegaard, D. M. (2001). Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications. *Growth and Change*, *32*(4), 466–490. http://doi.org/10.1111/0017-4815.00171

Bowman, T., Thompson, J., & Colletti, J., 2009. Valuation of OS and conservation features in residential subdivisions. Journal of Environmental Management 90, 321–330.

Brown, D. G., & Robinson, D. T. (2006) Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl, Ecology and Society, 11(1)

Buchmann, C. M., Grossmann, K., & Schwarz, N. (2016). How agent heterogeneity, model structure and input data determine the performance of an empirical ABM - A real-world case study on residential mobility. *Environmental Modelling and Software*, *75*, 77–93. http://doi.org/10.1016/j.envsoft.2015.10.005

Can, A. (1990) The measurement of neighborhood dynamics in urban house prices. *Economic Geography* 66 (3), 254–272.

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2007). Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation*, *34*(5), 535–556. http://doi.org/10.1007/s11116-007-9132-x

Casello, J. M. (2007). Transit competitiveness in polycentric metropolitan regions. *Transportation Research Part A: Policy and Practice*, *41*(1), 19–40. http://doi.org/10.1016/j.tra.2006.05.002

Cervero, R. (2003). *Effects of Light and Commuter Rail Transit on Land Prices : Experiences in San Diego County*. University of California, Berkeley

Chatman, D. G., Tulach, N. K., & Kim, K. (2011). Evaluating the Economic Impacts of Light Rail by Measuring Home Appreciation: A First Look at New Jersey’s River Line. *Urban Studies*, *49*(3), 467–487. http://doi.org/10.1177/0042098011404933

Chen, C. (2015). *Exploring Housing Market in Toronto , Ontario : Spatial Hedonic Modeling of Crime Rates, Subway Ridership, Dwelling Density & House Prices*. University of Waterloo.

Chiodo, A. J., Hernanndez-Murillo, R., & Owyang, M. T. (2010). Nonlinear effects of school quality on house prices. *Federal Reserve Bank of St. Louis Review*, *92*(3), 185–204.

Cho, S.-H., Roberts, R. K., & Kim, S. G. (2011). Negative externalities on property values resulting from water impairment: The case of the Pigeon River Watershed. *Ecological Economics*, *70*(12), 2390–2399. http://doi.org/10.1016/j.ecolecon.2011.07.021

City of Kitchener (2012) Parks. [Shapefile]. Provided by the University of Waterloo Geospatial Centre.

City of Waterloo (2011). Parks, Environmental Lands, and Forests. [Shapefiles]. Provided by the University of Waterloo Geospatial Centre.

Clark, W. A. V., & Hosking, P. L. (1986) Statistical Methods for Geographers. New York: Wiley.

Cohen, J. P., & Coughlin, C. C. (2008). Spatial Hedonic Models of Airport Noise, Proximity, and Housing Prices*. *Journal of Regional Science*, *48*(5), 859–878. http://doi.org/10.1111/j.1467-9787.2008.00569.x

Colwell P.F., & Dilmore G. (1999) Who was first? An examination of an early hedonic study. Land Economics 75(4): 620-626

Cotteleer, G., & van Kooten, G. C. (2012). Expert opinion versus actual transaction evidence in the valuation of non-market amenities. *Economic Modelling*, *29*(1), 32–40. http://doi.org/10.1016/j.econmod.2011.02.041

Cowley, & Easton (2015). Report Card on Ontario's Elementary Schools. Fraser Institute. Retrieved April 2016 from http://ontario.compareschoolrankings.org/pdfs/Fraser_Institute_Report_Card_on_Ontario%E2%80%99s_Elementary_Schools_2015.pdf.

Crompton, J. L. (2001). The impact of parks on property values: A review of the empirical evidence. Journal of Leisure Research, 33(1), 1–31

Data Management Group. (2006). Transportation Tomorrow Survey Boundaries [Shapefile]. Provided by DMG.

Data Management Group (2011). Regional Travel Summaries: Regional Municipality of Waterloo. Retrieved Jan 10, 2016 from: http://www.dmg.utoronto.ca/pdf/tts/2011/regional_travel_summaries/Waterloo.pdf.

Defields, E. (2013). *Property size preferences and the value of private and public outdoor spaces amid a shift to high-density residential development : A case study of Kitchener- Waterloo, Ontario* (Master's Thesis, University of Waterloo). Retrieved form https://uwspace.uwaterloo.ca

Dehring, C., & Dunse, N. (2006). Housing density and the effect of proximity to public OS in Aberdeen, Scotland. Real Estate Economics, 34(4), 553–566.

Dekkers, J. E. C., & van der Straaten, J. W. (2009). Monetary valuation of aircraft noise: A hedonic analysis around Amsterdam airport. *Ecological Economics*, *68*(11), 2850–2858. doi:10.1016/j.ecolecon.2009.06.002

Desmond, P. (2015). OMB accepts settlement of regional land dispute. Waterloo Region Record, June 18

Des Rosiers, F., Thériault, M., & Villeneuve, P.-Y. (2000). Sorting out access and neighbourhood factors in hedonic price modelling. Journal of Property Valuation and Investment 18, 291–315.

Doss, C.R., & Taff, S.J. (1996). The influence of wetland type and wetland proximity on residential property values. *Journal of Agricultural and Resource Economics* 21, 120e129.

Dubé, J., Des Rosiers, F., Thériault, M., & Dib, P. (2011). Economic impact of a supply change in mass transit in urban areas: A Canadian example. *Transportation Research Part A: Policy and Practice*, *45*(1), 46–62. http://doi.org/10.1016/j.tra.2010.09.002

Duncan, M. (2010). The Impact of Transit-oriented Development on Housing Prices in San Diego, CA. *Urban Studies*, *48*(1), 101–127. http://doi.org/10.1177/0042098009359958

Dziauddin, M. F., Powe, N., & Alvanides, S. (2014). Estimating the Effects of Light Rail Transit (LRT) System on Residential Property Values Using Geographically Weighted Regression (GWR). *Applied Spatial Analysis and Policy*, *8*(1), 1–25. http://doi.org/10.1007/s12061-014-9117-z

ESRI (2010). Algorithms used by Network Analyst. ArcGIS Desktop Help 9.3. Retrieved January 2016, from http://webhelp.esri.com/arcgisdesktop/9.3

Filatova, T., van der Veen, a, & Parker, D. C. (2009). Land Market Interactions between Heterogeneous Agents in a Heterogeneous Landscape—Tracing the Macro-Scale Effects of Individual Trade-Offs between. *Canadian Journal of Agricultural Economics*, *57*, 431–57. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1744-7976.2009.01164.x/full

Florida, R. (2008) *Who's Your City?: How the Creative Economy is Making Where to Live the Most Important Decision of Your Life.* Basic Books, Random House, New York

Florida, R. (2002) The economic geography of talent. Annals of the Association of Amer- ican Geographers 92(4):743–755

Franklin, J. P., & Waddell, P. (2003). A hedonic regression of home prices in King County, Washington, using activity-specific accessibility measures. Paper presented at the Transportation Research Board Annual Meeting.

Fraser Institute. (2014). 2013-2014 School Ratings. Retrieved April 14, 2016 from http://ontario.compareschoolrankings.org/elementary/SchoolsByRankLocationName.aspx.

Geoghegan, J. (2002). The value of OSs in residential land use, *19*(January 2001), 91–98.

Gibbons, S., & Machin, S., (2005). Valuing rail access using transport innovations. Journal of Urban Economics 57, 148–159

Goolsby, W.C., (1997). Assessment error in the valuation of owner-occupied housing. Journal of Real Estate Research 13 (1), 33e45.

Guerra, E., Cervero, R., & Tischler, D. (2011). The Half-Mile Circle: Does It Best Represent Transit Station Catchments? *UC Berkeley Center for Future Urban Transport*, (July).

Hayes, K. J. , & L. L. Taylor (1996). Neighborhood school characteristics: What signals quality to homebuyers? Economic Review (Federal Reserve Bank of Dallas), Fourth Quarter, pp. 2-9.

Health & Community Design Lab (2016) NEWPATH. University of British Columbia. Retrieved from http://health-design.spph.ubc.ca/research/current-research/newpath/

Henderson, K. K., & Song, Y. (2008). Can nearby OSs substitute for the size of a property owner's private yard? *International Journal of Housing Markets and Analysis*, *1*(2), 147–165. http://doi.org/10.1108/17538270810877772

Hess, D. B., & Almeida, T. M. (2007). Impact of Proximity to Light Rail Rapid Transit on Station-area Property Values in Buffalo, New York. *Urban Studies*, *44*(5/6), 1041–1068.

Higgins, C. D., & Kanaroglou, P. S. (2016). Forty years of modelling rapid transit's land value uplift in North America: moving beyond the tip of the iceberg. *Transport Reviews*, *1647*(April). http://doi.org/10.1080/01441647.2016.1174748

Hooper, J. (2014). A destination too far? Modelling destination accessibility and distance decay in tourism. *GeoJournal*, *80*(1), 33–46. http://doi.org/10.1007/s10708-014-9536-z

Huang, Q., Parker, D. C., Sun, S., & Filatova, T. (2013). Effects of agent heterogeneity in the presence of a land-market: A systematic test in an agent-based laboratory. *Computers, Environment and Urban Systems*, *41*, 188–203. http://doi.org/10.1016/j.compenvurbsys.2013.06.004

Ibeas, Á., Cordera, R., dell'Olio, L., Coppola, P., & Dominguez, A. (2012). Modelling transport and real-estate values interactions in urban systems. *Journal of Transport Geography*, *24*, 370–382. http://doi.org/10.1016/j.jtrangeo.2012.04.012

Irwin, E. G. (2002). The effect of OS on residential property value. Land Economics, 78(4), 465–480.

Jaeger, W.K., & Plantinga, A.J. (2007) How have land-use regulations affected property values in Oregon?. Oregon State University.

Kaczynski, A. T., Potwarka, L. R., Smale, B. J. a., & Havitz, M. E. (2009). Association of Parkland Proximity with Neighborhood and Park-based Physical Activity: Variations by Gender and Age. *Leisure Sciences*, *31*(2), 174–191. http://doi.org/10.1080/01490400802686045

Keil, R., & Young, D. (2009). Fringe explosions: Risk and vulnerability in Canada's new in-between urban landscape. *Canadian Geographer*, *53*(4), 488–499. http://doi.org/10.1111/j.1541-0064.2009.00270.x

Kitchen, J.W., & Hendon, W.S. (1967). Land values adjacent to an urban neighborhood park. Land Economics 43 (3), 357–361.

Kong, F., Yin, H., & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. *Landscape and Urban Planning*, *79*(3-4), 240–252. http://doi.org/10.1016/j.landurbplan.2006.02.013

Koohsari, M. J., Kaczynski, A. T., Giles-Corti, B., & Karakiewicz, J. A. (2013). Effects of access to public OSs on walking: Is proximity enough? *Landscape and Urban Planning*, *117*, 92–99. http://doi.org/10.1016/j.landurbplan.2013.04.020

Konijnendijk, C. C., Annerstedt, M., Busse Nielsen, A., & Maruthaveeran, S. (2013). *Benefits of Urban Parks A systematic review*. Copenhagen & Alnarp.

Krarup, J. & Pruzan P.M. (1980). The impact of distance on location problems. *European Journal of Operational Research 4, 256-269*

Krause, A., & Bitter, C. (2012). Spatial econometrics, land values and sustainability: Trends in real estate valuation research. *Cities*, *29*, S19–S25. http://doi.org/10.1016/j.cities.2012.06.006

Kuminoff, N. V., Parmeter, C. F., & Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? Journal of Environmental Economics and Management, 60, 145–160.

Lambert, D. M., Clark, C. D., Wilcox, M. D., & Cho, S. H. (2011). Distance, density, local amenities, and suburban development preferences in a rapidly growing East Tennessee county. *Agriculture and Human Values*, *28*(4), 519–532. http://doi.org/10.1007/s10460-010-9306-6

Lancaster, K. (1966). A new approach to consumer theory. Journal of Political Economy 74(2):132–157

Lee, S.W., Taylor, P.D., Hong, S.K., (2008). Moderating effect of forest cover on the effect of proximity to chemical facilities on property values. Landscape and Urban Planning 86, 171–176.

LeSage, J., & Pace, RK. (2009). Introduction to spatial econometrics. CRC Press, Taylor and Francis Group, Boca Raton

Luttik, J. (2000). The value of trees, water and OS as reflected by house prices in the Netherlands. Landscape and Urban Planning, 48(3–4), 161–167.

Lutzenhiser, M., & Netusil, N. R. (2001). The Effect of OSs on a Home's Sale Price. *Contemporary Economic Policy*, *19*(3), 291–298. http://doi.org/10.1093/cep/19.3.291

Mankiw, N. G. (1998). Principles of Microeconomics. The Dryden Press, Harcourt Brace College Publishers, Orlando Florida

McConnell, V., & Walls, M. (2005). The Value of OS: Evidence from Studies of Nonmarket Benefits. *Resources for the Future*, (January), 82.

McGarigal, K., Cushman, S.A., & Ene, E. (2012). FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: http://www.umass.edu/landeco/research/fragstats/fragstats.html

Mieszkowski, P., & Mills, E. S. (1993). The Causes of Metropolitan Suburbanization Published by : American Economic Association Tn. *The Journal of Economic Perspectives*, *7*(3), 135–147.

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

Mitchell, R. C., & Carson, R. T. (1989). Using Surveys to Value Public Goods: The Contingent Valuation Method. *Resources for the Future*, 7–9. http://doi.org/10.2307/2072944

Morancho, A. (2003). A hedonic valuation of urban green areas. Landscape and Urban Planning, 66, 35–41.

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

Ministry of Municipal Affairs and Housing (2006). Places to Grow: Growth Plan for the Greater Golden Horseshoe. Ontario, Canada.

Ministry of Municipal Affairs and Housing (2014). Provincial Policy Statement. Government of Ontario

Ministry of Municipal Affairs and Housing (2005). Places to Grow Act. Ontario, Canada.

MPAC (2014). Property Codes. Retrieved July 3, 2014, from http://www.mpac.ca/property_owners/how/propertyCodeInventory.asp.

MPAC (2015). MPAC Assessment Parcels [Shapefile]. Provided by the Region of Waterloo.

MPAC (2016a). Residential Properties. Retrieved 10 May 2016 from https://www.mpac.ca/PropertyTypes/ResidentialProperties.

MPAC (2016b). Property Valuation Explained. Retrieved 10 May 2016 from https://www.mpac.ca/PropertyOwners/Diva.

Mueller, J., & Loomis, J. (2008). Spatial Dependence in Hedonic Property Models: Do Different Corrections for Spatial Dependence Result in Economically Significant Differences in Estimates Implicit Prices? *Journal of Agricultural and Resource Economics*, *33*(2), 212–231.

Munroe, D.K. (2007). Exploring the determinants of spatial pattern in residential land-markets: amenities and disamenities in Charlotte, NC, USA. Environment and Planning B: Planning and Design 34, 336e354.

Neudorf, J. (2014). *Understanding Accessibility, Analyzing Policy : New Approaches for a New Paradigm* (Master's Thesis, University of Waterloo). Retrieved form https://uwspace.uwaterloo.ca

Neupane, A., & Gustavson, K. (2008). Urban property values and contaminated sites: A hedonic analysis of Sydney, Nova Scotia. *Journal of Environmental Management*, *88*(4), 1212–1220. http://doi.org/10.1016/j.jenvman.2007.06.006

Newman, P., & Kenworthy, J. (2006). Urban Design to Reduce Automobile Dependence. *Opolis: An International Journal of Suburban and Metropolitan Studies*, *2*(1).

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

Oates, W. (1969) The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis. Journal of Political Economy, November/December 1969, 77(6), pp. 957-71

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

Páez, A., & Scott, D. (2005). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, *61*(1), 53. http://doi.org/10.1007/sGEJO-004-0877-x

Pivo, G., & Fisher, J. (2011). The walkability premium in commercial real estate investments. Real Estate Econ., 39(2), 185–219.

Rauterkus, S. Y., & Miller, N. G. (2011). Residential Land Values and Walkability. *Journal of Sustainable Real Estate*, *3*(1), 23–43.

Redfearn, C.L. (2009). How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets. Regional Science and Urban Economics 39 (3), 297–306.

Region of Waterloo (2009). Development of a Walkability Index for Waterloo Region. Ontario Agency for Health Protection, retrieved May 10, 2016 from www.publichealthontario.ca.

Region of Waterloo (2010). Regional Forests, Cemeteries, and Golf courses. [Shapefiles]. Provided by the University of Waterloo Geospatial Centre.

Region of Waterloo. (2011). Pedestrian Network [Network dataset]. Provided by the Region of Waterloo.

Region of Waterloo. (2012). Official Plan. Region of Waterloo, Ontario, Canada.

Region of Waterloo (2013a). *Community Building Strategy.* Kitchener, Ontario

Region of Waterloo (2013b). Municipal Boundaries. [Shapefile]. Provided by the University of Waterloo Geospatial Centre.

Region of Waterloo. (2013c). Transit Schedule Data [GTFS dataset]. Retrieved March 28, 2014, from http://www.regionofwaterloo.ca/en/regionalGovernment/OpenDataCatalogue.asp.

Region of Waterloo (2015a). Building Footprints [Access database]. Provided by the Region of Waterloo.

Region of Waterloo (2015b). CTC Conceptual and Analytical Boundary. [Shapefiles] Provided by the Region of Waterloo.

Region of Waterloo. (2015c). *CTC Monitoring Program: Baseline Monitoring Report*. Kitchener, Ontario.

Region of Waterloo (2015d). MPAC Property Assessment Data [Access database]. Provided by the Region of Waterloo.

Region of Waterloo (2015e). Property Transaction Data [Access database]. Provided by the Region of Waterloo.

Region of Waterloo (2015f). Planning Neighbourhoods. [Shapefile] Provided by the Region of Waterloo.

Region of Waterloo (2016). Walkability. Provided by the Region of Waterloo.

Riddell, R. (2004). Sustainable Urban Planning. Malden, MA: Blackwell.

Rosen, S. (1974). Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition. *The Journal of Political Economy*, *82*(1), 34–55.

Searle, G ., & Filion, P. (2011) Planning Context and Urban Intensification Outcomes: Sydney versus Toronto., Urban Studies 48: 1419-1438.

Shearmur, R., Coffey, W., Dubé, C., & Barbonne, R. (2007). Intrametropolitan employment structure: Polycentricity, scatteration, dispersal, & chaos in Toronto, Montreal, and Vacouver, 1996-2001. Urban Studies, 44(9), 1713-38. http://dx.doi.org/10.1080/00420980701426640.

Sirmans, G. S., Macpherson, D. a, & Zietz, E. N. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, *13*(1), 1–44. http://doi.org/Article

Smith, V. K., Poulos, C., & Kim, H. (2002). Treating OS as an urban amenity. Resource and Energy Economics, 24, 107–129.

Smith, D. (2007). Studentification and 'Apprentice' Gentrifiers within Britain's Provincial Towns and Cities: Extending the Meaning of Gentrification. *Environment and Planning A*, 39: 142-161. doi: 10.1068/a38476

Statistics Canada (2011a) National Household Survey, CT Scale: Education. [Shapefile]. Provided by the University of Waterloo Geospatial Centre.

Statistics Canada (2011b) National Household Survey, DA Scale: Age. [Shapefile]. Provided by the University of Waterloo Geospatial Centre.

Statistics Canada. (2015a). *Table 327-0046 - New housing price index, monthly (index,2007 = 100),* CANSIM (database). (accessed: April 2015).

Statistics Canada (2015b). NHS commute Table: Table 2 - Usual commuting time to work, census metropolitan areas, 2011. Retrieved February 2015 from https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-012-x/2011003/tbl/tbl02-eng.cfm.

Statistics Canada (2016). DA. Retrieved from http://www12.statcan.gc.ca/census-recensement/index-eng.cfm

Tiebout, C.M. (1956). A pure theory of local expenditure. Journal of Political Economy 64 (5), 416–424.

Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region, Economic Geography, 46(2), pp. 234–240.

Troy, A., & Grove, J. M. (2008). Property values, parks, and crime: A hedonic analysis in Baltimore, MD. Landscape and Urban Planning, 87(3), 233–245.

Vale & Pereira, M. (2016) The influence of the impedance function on gravity-based pedestrian accessibility measures: A comparative analysis. *Environment and Planning B: Planning and Design* 0265813516641685, first published on April 1, 2016

Walker, J.L., & Li, J. (2007). Latent lifestyle preferences and household location decisions. Journal of Geographical Systems 9 (1), 77–101

Wang, L. (2010). Impact of Urban Rapid Transit on Residential Property Values. *Chinese Economy*, *43*(2), 33–52. http://doi.org/10.2753/CES1097-1475430203

Washington, E. (2013). Role of Walkability in Driving Home Values. *Leadership and Management in Engineering*, *13*(3), 123–130.

Waterloo Region District School Board (2015) 2011-2012 Catchment Areas. [Shapefile]. Provided by the University of Waterloo Geospatial Centre.

Whitehead, J. C., Pattanayak S.K., Van Houtven G.L., & Gelso, B. R. (2008). Combing revealed and stated preference data to estimate the nonmarket value of ecological services: An assessment of the state of the science. *Journal of Economic Surveys* 22:872-908.

Wolinsky, J. (1999). Light Rail: One Route To Livable Cities. *Railway Age*, *200*, 49. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=tspt&NEWS=N&AN=00770774

Wooldridge, J. (2012). Introductory Econometrics – A Modern Approach 5th ed. Mason: South Western Cengage Learning.

Yeung, K. (2015). *The Development of a Household Travel Resource Allocation Model for Kitchener – Waterloo* (Master's Thesis, University of Waterloo). Retrieved form https://uwspace.uwaterloo.ca

Yoo, S., Im., J., & Wagner, J. E. (2012). Variable selection for hedonic model using machine learning approaches: A case study in Onondaga County, NY. *Landscape and Urban Planning*, *107*(3), 293–306. http://doi.org/10.1016/j.landurbplan.2012.06.009

Zhang, X., Lu, H., & Holt, J. B. (2011). Modeling spatial accessibility to parks: a national study. *International Journal of Health Geographics*, *10*(1), 31. http://doi.org/10.1186/1476-072X-10-31

Zhou, X., & Rana, M. M. P. (2012). Social benefits of urban green space: A conceptual framework of valuation and accessibility measurements. *Management of Environmental Quality: An International Journal*, *23*(2), 173–189. http://doi.org/10.1108/14777831211204921

Zukin, S. (1989). Loft Living: Culture and Capital in Urban Change. Rutgers, NJ: Rutgers University Press.

APPENDIX 1: Data Sources and Uses

This appendix organizes and elaborates on the data used in this research and outline their role in the modelling work done. Two sets of data were joined, where one was held by the researcher and one was held by the ROW. The data used for this study and how they were applied to create variables in this research are shown in the following figure:

Data Framework

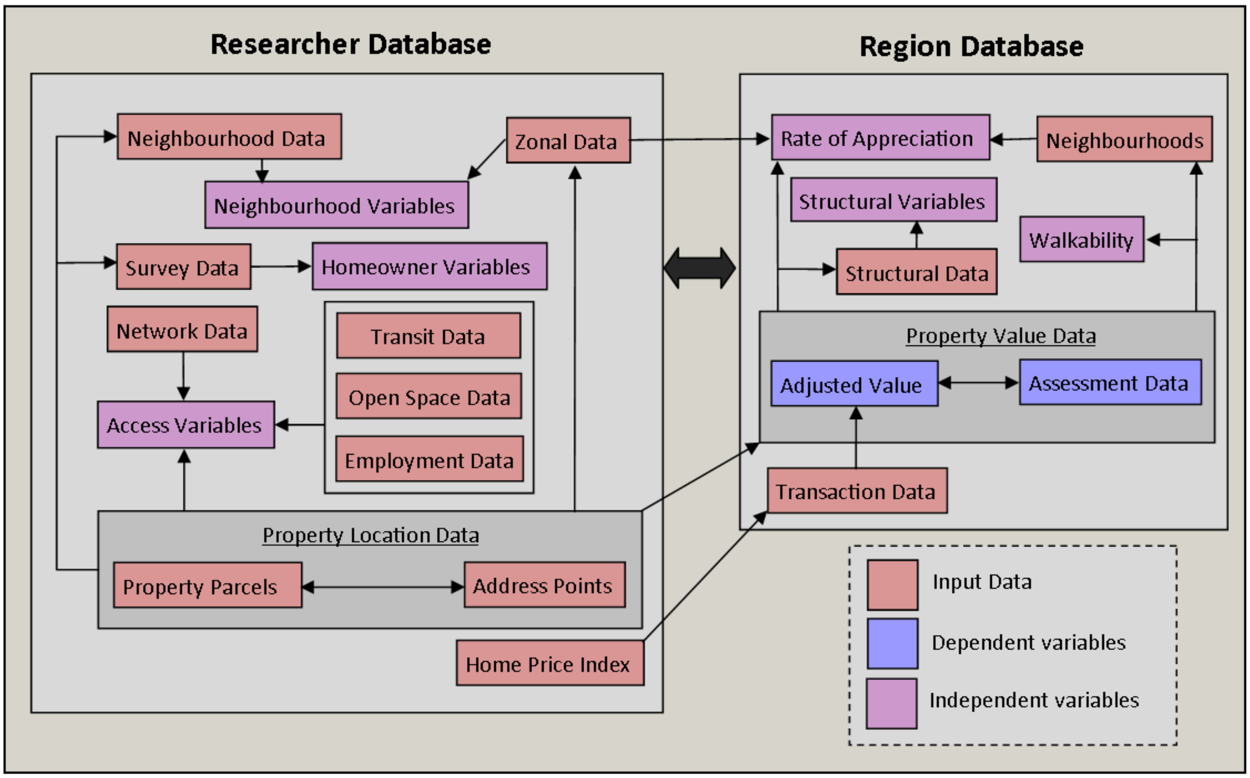


Figure 3: Data Framework

In this chart, data are classified as independent variables, dependent variables, or other input data elements used in the creation of model variables. The Researcher Database was composed of various data elements combined into a Stata format data file. The input data in the Region Database were combined in *Access* by region staff member David Stubbs, exported into a *csv* file, then converted to a *Stata* data file. The property *Roll Number* was the common identifier used to link the researcher and region databases using an m:1 merge. Following that, the raw from the Region database was used to create variables. The specific sources of each of the raw data elements are described in this chapter, while variables created from raw data are described in the following chapter.

Input Data

The following sections will describe the input data.

For reference, the order in which the input data will be presented is:

- Transaction Data
- Property Parcels
- Home Price Index
- Zonal Data / Neighbourhoods
- Neighbourhood Data
- Network Data
- Transit Data
- OS Data
- Employment Data
- Structural Data
- Demographic data

Transaction Data

Source: ROW, 2015e

This dataset contains records for property sales in the ROW since 1973; however, it contains a drastic increase in the number of sales from 1985 onwards, and a less drastic increase in 1997, which may indicate incomplete records prior to these years. This data contains the dollar amount for which each property was sold, the date of the sale, an indicator of the transfer mechanism used in the change of ownership, as well as multiple fields identifying the buyer(s) and seller(s) in the transaction.

Use:

- *Adjusted Value* variable

Property Parcels

Source: MPAC, 2015

This layer is a polygon shapefile representing all assessment parcels in the ROW in 2015, which are uniquely identified by Roll Number. This dataset was clipped in ArcGIS to contain only KW properties, leaving 93,371 properties.

Use:

- Joining researcher and region databases
- Mapping variables

Property Points

The property parcels layer was converted to points using the centroid of each polygon in ArcGIS.

- *OS* and *Transit Access* variables

Home Price Index

Source: Statistics Canada, 2015a

The home price index used for this work is the New Home Price Index from statistics Canada. The definition attached to the data states that "The New Housing Price Index (NHPI) is a monthly series that measures changes over time in the contractors' selling prices of new residential houses, where detailed specifications pertaining to each house remain the same between two consecutive periods. The survey covers the following dwelling types: single dwellings, semi-detached and row houses (town house and garden home). The survey also collects contractors' estimates of the current value (evaluated at market price) of the land. These estimates are independently indexed to provide the published series for land. The residual, (total selling price less land value), which mainly relates to the current cost of the structure is also independently indexed and is presented as the estimated house series." (Statistics Canada, 2015a). The home price index value used in this research are of the sum of both land and residual values (the total selling price).

Use:

- Adjusting Transaction Value

Zonal Data / Neighbourhoods

The zonal data in this work were used to create neighbourhood variables and to display aggregated confidential data in maps.

Traffic Analysis Zone (TAZ) Layer:

Source: Data Management Group, 2006

The TAZ is a spatial unit used for transportation planning; specifically, these TAZs correspond with the Transportation Tomorrow Survey. There are 270 unique TAZs within KW.

Use:

- Employment Access Variable
- Testing in ROA variable specification

CT (CT) Layer:

Source: Statistics Canada, 2011a

The Census

The CT layer used for this work comes from the 2011 NHS This layer includes population and employment information at the CT tract scale. This shapefile was used both as a neighbourhood or zone definition to create and display other variables, as well as to create variables using the NHS

Use:

- *ROA* variable
- *Education Rate* variable
- Averaging variables to display confidential data in maps

DA (DA) Layer:

Source: Statistics Canada, 2011b

The DA layer used for this work comes from the 2011 NHS. This layer includes population and employment information at the DA scale, which is smaller than a CT. Statistics Canada defines the DA as a "Small area composed of one or more neighbouring dissemination blocks, with a population of 400 to 700 persons. All of Canada is divided into DAs." (Statistics Canada 2016b, although they note that some DAs contain smaller or larger populations as they must conform to the boundaries of higher-order divisions (CTs and census subdivisions).

Use:

- *Perception of Safety (Police Calls per 100 residents)* variable
- *Population Density* variable

Planning Neighbourhoods

Source: ROW, 2015f

Planning neighbourhoods represent the different neighbourhoods used by the ROW for analysis. There are 22 of these planning neighbourhoods in KW.

Use:

- Aggregating confidential data for mapping (Adjusted Value, Assessed Value, and ROA)

CTC Analytical Boundary

Source: ROW, 2015b

The CTC analytical boundary is used by the ROW for planning efforts. This boundary aligns with the analysis found in the ROW CTC baseline monitoring described in the introduction of this thesis (ROW, 2015).

Use:

- *In CTC* dummy variable
- Comparing variable distributions between the CTC and rest of the KW

KW Boundary

Source: ROW 2013b

This shapefile contains the administrative boundaries of KW.

Use:

- *Waterloo* dummy variable

School Catchments

Source: Waterloo Region District School Board, 2015

This layer represents the school catchment in Waterloo Region in the 2011-2012 school year. After selecting only junior, senior, and composite elementary schools and removing redundant school zones, 81 school catchments in KW that were used for this work.

Use:

- *School Quality* variable

Neighbourhood Data

Most of the neighbourhood variables for this research were creating using the data within the zonal/neighbourhood shapefiles, except for the school quality ratings from the Fraser Institute and police phone calls data from the ROW, described below.

School Quality

Source: Fraser Institute, 2014

The Fraser Institute conducted a detailed analysis of school quality across Canada for the 2013-2014 school year. They provide a province-specific 'report card' for all schools, which contains objective ratings for each school on a 10-point scale. These ratings are based on the results of standardized provincial EQAO student assessments. Nine criteria are evaluated to establish the schools rating out of a potential score of ten. The first six are the average levels of each of reading, writing, and math for each of grades 3 and 6 students. The next two criteria are the difference between female and male test scores in grade 6 EQAO scores in each of the reading and the mathematics tests. The last criterion of the 10-point school rating is the proportion of EQAO scores that did not meet standards set out by the province. (Cowley & Easton, 2015).

Use:

- *School Quality* variable

Police Phone Calls

Source: Waterloo Regional Police Service, 2015 (modified by ROW)

The original police phone calls data was provided by the Waterloo Region Police Service, which is released annually. The police phone calls used in this research are represent the year 2012. From the raw police data, the ROW limited the dataset to contain only police phone calls for service for reasons believed to be linked to public perceptions. A detailed analysis of the police phone calls data is presented in the ROW Baseline Monitoring Report (2015).

Use:

- *Perception of Safety (Police Calls per 100 residents)* variable

Network Data

Source: ROW, 2011

This layer is an ESRI network dataset, including lines and junctions, for the pedestrian transportation network in 2012. This network data was updated after manual inspection by Jason Neudorf, as is explained in Neudorf (2014). This layer contains roadways (but not freeways), public trails, and pedestrian walkways in the ROW.

Use:

- *Transit* and *OS Access* variables
- *Regional Road Adjacency* variable

Transit Data

Source: ROW, 2013c

The transit data used in this research comes from the ROW, 2014. This is the GTFS dataset, which includes all Grand River Transit stops in 2013. Within KW, this file contains 2294 stop locations.

Use:

- *Transit Access* variable

OS Data

The OS data came from the City of Kitchener, City of Waterloo, and the ROW. Each is described individually in the following sub-sections. These OS shapefiles were merged in ArcGIS, and the *dissolve* tool was used to remove OSs that overlapped between the various datasets.

Use:

- *OS Access* variable
- *OS Adjacency* variable

City of Kitchener:

Source: City of Kitchener, 2012

The OS shapefile used from the City of Kitchener named Parks includes various categories of OS. These categories are 'city wide park', 'district park', 'green', 'green (commons)', 'green (parkette)', 'green (urban plaza)', 'greenway', 'natural area', 'neighbourhood park', and 'OS'. This layer contains 395 OSs in the year 2012.

City of Waterloo:

Source: City of Waterloo 2011

Three separate OS shapefiles were used from the city of Waterloo, Parks (2011), Env_Lands, and Forests. Parks contains 284 Waterloo parks, Env_Lands contains 228 environmental lands, which include a variety of types of OS including some lands owned by the Grand River Conservation Authority, and Forests contains 279 observations.

ROW:

Source: ROW, 2010

Three shapefiles representing OS were used from the ROW; these were Regional Forests, cemeteries, and golf courses. Regional Forests contains 4 forests located in KW, Golf courses contains 7 KW golf courses, and Cemeteries contains 24 cemeteries within KW.

Employment Data

Source: Neudorf, 2014

The employment data for this work was preprocessed by Jason Neudorf, who created an origin destination cost matrix between TAZs using TTS data. See the employment access section of the pre-regression methods of this work and his thesis (Neudorf, 2014), for more details.

Use:

- *Employment Access* variable

Average commuting time and distance

Source: Data Management Group, 2011 & Statistics Canada, 2015b

The average commuting time and distance in the ROW were taken from two sources. The TTS table provided the average commute distance, while the average commuting time was taken from Statistics Canada’s NHS.

Use:

- *Employment Access* variable (calculating generalized cost of the average commute in the ROW)

Structural Data

The structural data in this work come from the MPAC property lots layer and the regional building footprints. The property lots were described above in section X, while the building footprints layer will be described here.

Building footprints layer

Source: Region of Waterloo, 2015a

Data from the ROW building footprint layer was joined to the property value data in the ROW database. This building footprints layer included the building footprints of both the residential and accessory structures, separately, the number of storeys,

Use:

- *Yard Size* variable
- *Living Area* variable

MPAC Assessment layer

The ROW provided selected information for the sample from their MPAC assessment data, aside from the actual assessed values that are described later in this chapter, this data contained structural home characteristics as well. The fields that were used for structural home characteristics from this dataset include lot size, lot frontage, year the property structure was built, and property type. The specific uses of these data are described in the pre-regression methods chapter, under each variable that they were used to create.

Use:

- *Age of Residence* variable
- *Yard size* variable
- *Property* type variables
- *Frontage* variable

Demographic Data

Source: Defields, 2014

The demographic data comes from a survey conducted by Emma DeFields to understand KW residents' private yard and public OS preferences (2014). The sample of DeFields' survey was selected to omit homes with no private yards. Because of this, suburban households make up a proportionally larger share of the sample than the true distribution of households.

Use:

- *Stated Preference* variables
- *Homebuyer Demographic* variables

Dependent variables

Adjusted Value

Adjusted value was created from the transaction data, using a home price index and removing non-market sales as described in Pre-Regression Methods.

Use:

- Dependent variable in regression
- *ROA* variable

Assessed Value

Source: ROW, 2015d

Assessed value data for this research comes from the Municipal Property Assessment Corporation (MPAC). MPAC operates under the Ontario's Assessment Act., wherein it is tasked with estimating the current value of real property in Ontario. From the assessment act, "current value" means, in relation to land, the amount of money the fee simple, if unencumbered, would realize if sold at arm's length by a willing seller to a willing buyer" (ASESSMNENT ACT, CITE). Essentially, this means that assessments should represent the expected selling price of a home in a standard market transaction.

This assessment is necessary to calculate the property taxes owed by homeowners. The assessment methodology utilized by MPAC differs for various property types, and includes the direct comparison, cost, and income approaches. Using the direct comparison approach, which is similar to the approach taken in this thesis, MPAC assesses over 200 discrete variables. However, they found that there are five main factors that explain 85% of the variation in residential property values. These five factors are: the age of the residence, the living area, the location, the dimensions of the lot, and a rating of the building's construction quality – see MPAC, 2016 for more information on their valuation methodologies.

Use:

- Dependent variable in regression

Independent Variables

Except for walkability, all other independent variables in this thesis are the product of data inputs already described in this chapter. The methodologies of all the independent variables investigated in this thesis are outlined in the pre-regression methods chapter; however, this section will outline the source of the walkability data briefly.

Walkability

Source: ROW, 2016

The Walkability Index was created through a partnership between the ROW and academic institutions. This walkability index was created as part of are multi-institutional built form and public health study called NEWPATH. For more information on this study and its findings see Health & Community Design Lab, 2016. The components of the Walkability index are outlined in the thesis.

Use:

- Independent variable in regression

APPENDIX 2: O-D Cost Matrix Script

In Python:

```
# ---------------------------------------------------------------------------
# Description:
# Created by: Robert Babin with assistance from Xiongbing Jin and Scott
MacFarlane
# ---------------------------------------------------------------------------
from datetime import datetime

tstart = datetime.now()
print tstart

# Import arcpy module and checkout license
import arcpy
import os
import math
import numpy
arcpy.CheckOutExtension("Network")

# environment settings
arcpy.env.overwriteOutput = True
arcpy.env.workspace = "C:\\Users\\rmbabin\\Documents\\DATA\\KWres\\DATA"
#setting workspace
ws = "C:\\Users\\rmbabin\\Documents\\DATA\\KWres\\DATA"

# Local variables:

Network = "PednwUD_ND.nd" # Network Dataset
Lines = "OD Cost Matrix\\Lines" #OD Cost Matrix Output field: "Lines"
OD_Cost_Matrix = "OD Cost Matrix"
pnts_IJ = "pnts_IJ" #Address subsets with n="sample"
pnts_OS = "OSpoints.shp" #Point file of Destinations
OS_shp = "KW_OS.shp" #OS dataset
pnts_Property = "PropSnapped.shp"
ODLines_CopyFeatures_all = "ODLines_CopyFeatures_all.shp"
poly_Property = "April_KW_Properties.shp"

result = arcpy.GetCount_management(pnts_Property)
TotalAddCount = int(result.getOutput(0))
print TotalAddCount
sample = 1000
sample1 = 10000
outputfcs = ""

#Rounding up for looping
result = arcpy.GetCount_management(pnts_Property)
FullAddSet = int(result.getOutput(0))
def roundup(x):
    return int(math.ceil(x / 10000.0)) * 10000  #make sure this value is
equal to "sample1".0 and "sample1"
```

```
FullAddSetRnded = roundup(FullAddSet)

InitialSubsets = list(range(10000, FullAddSetRnded+1, sample1))

print str(InitialSubsets)

print "making initial origin point subsets"
for l in InitialSubsets:
    m = l - sample1
    print "Less than " + str(l) + " and greater than " + str(m)
    select_query = '\"FID\" < ' + str(l) + ' and \"FID\" >= ' + str(m) #
Selecting subsets
    arcpy.FeatureClassToFeatureClass_conversion (pnts_Property, ws,
"AddPnts"+ str(l) + ".shp", select_query)
    pnts_shp = "AddPnts" + str(l) + ".shp" #Point file of Origins
    result = arcpy.GetCount_management(pnts_shp)
    TotalAddCount = int(result.getOutput(0))
    print "starting " + str(l)

InitialSubsets = list(range(sample1, FullAddSetRnded+1, sample1))

print str(InitialSubsets)

arcpy.MakeODCostMatrixLayer_na(Network, OD_Cost_Matrix, "Length", "1000", "",
"", "ALLOW_UTURNS", "", "NO_HIERARCHY", "", "NO_LINES", "") #Generating OD
Cost Matrix with no lines displayed

print 'Loading Destinations'

arcpy.AddLocations_na(OD_Cost_Matrix, "Destinations", pnts_OS, "CurbApproach
CurbApproach #;Name ORIG_FID #", "1000 Meters", "ORIG_FID","Network
SHAPE;Network_Junctions NONE", "MATCH_TO_CLOSEST", "APPEND", "NO_SNAP", "1000
Meters", "INCLUDE", "Network #;Network_Junctions #")

for l in InitialSubsets:
    pnts_shp = "AddPnts" + str(l) + ".shp" #Point file of Origins
    result = arcpy.GetCount_management(pnts_shp)
    TotalAddCount = int(result.getOutput(0))
    print "starting " + str(l)

# FOR RUNNING THE OD COST MATRIX
    for i in range (0,TotalAddCount,sample): #Looping through address points
file to run the OD cost matrix on subsets
        j = i + sample

        select_query = '\"FID\" >= ' + str(i) + ' and \"FID\" < ' + str(j) #
Selecting subsets
        print select_query
        outputfc = "ODLines_CopyFeatures"+str(i)+"to"+str(j) # Defining
subset Feature Class
        print outputfc

        arcpy.MakeFeatureLayer_management(pnts_shp, pnts_IJ, select_query,
"", "FID FID HIDDEN NONE;Shape Shape HIDDEN NONE;FID_1_1 FID_1_1 HIDDEN
```

```
NONE;OBJECTID OBJECTID HIDDEN NONE;RollNumber RollNumber VISIBLE
NONE;ORIG_FID ORIG_FID VISIBLE NONE")

        arcpy.AddLocations_na(OD_Cost_Matrix, "Origins", pnts_IJ, "Name
RollNumber #;TargetDestinationCount TargetDestinationCount #;CurbApproach
CurbApproach #;Cutoff_Length Cutoff_Length #", "1000 Meters", "RollNumber",
"Network SHAPE;Network_Junctions NONE", "MATCH_TO_CLOSEST", "CLEAR",
"NO_SNAP", "1000 Meters", "INCLUDE", "Network #; Network_ND_Junctions #")

        print "solving "+ outputfc

        arcpy.Solve_na(OD_Cost_Matrix, "SKIP", "TERMINATE", "")
        arcpy.CopyFeatures_management(Lines,
"ODLines_CopyFeatures"+str(i)+"to"+str(j), "", "0", "0", "0") # Adding layer
to feature class
        outputfcs = outputfcs + outputfc + '.shp;'

    outputfcs = outputfcs[0:len(outputfcs)-1]
    print outputfcs

    print 'merging outputs...'
    arcpy.Merge_management(outputfcs,"ODLines_CopyFeatures_all")

    #clean up outputfcs
    fclist = outputfcs.split(';')
    for fc in fclist:
        print 'deleting ' + fc
        arcpy.Delete_management(fc)
    fclist = []
    outputfcs = ''
    print 'DONE solving and joining subsets'

    print 'Reidentifying observations'
    # CODE FOR SPLITTING IDENTIFIERS FROM NAME FIELD IN LINES FILE
    # Process: Adding "Address" Field
    arcpy.AddField_management(ODLines_CopyFeatures_all, "Address", "TEXT",
"", "", "", "", "NULLABLE", "NON_REQUIRED", "")
    # Process: Extracting "Address" from "Name" Field
    arcpy.CalculateField_management(ODLines_CopyFeatures_all, "Address",
"slicing(!Name!)", "PYTHON_9.3", "def slicing(field):\\n     head, sep, tail
= field.partition('-')\\n     return head.rstrip()")
    # Process: Adding ParkID field
    arcpy.AddField_management(ODLines_CopyFeatures_all, "ParkID", "DOUBLE",
"", "", "", "", "NULLABLE", "NON_REQUIRED", "")
    # Process: Extracting ParkID from "Name" field
    arcpy.CalculateField_management(ODLines_CopyFeatures_all, "ParkID",
"slicing(!Name!)", "PYTHON_9.3", "def slicing(field):\\n     head, sep, tail
= field.partition(' - ')\\n     return tail.rstrip()")

    print 'Getting minimum distance between each OS and property'

    #Code for getting smallest distance to park
    output_stats_table = "output_stats_table" + str(l) + ".dbf"
    stats_fields = [["Total_Leng", "MIN"], ["Address", "FIRST"], ["ParkID",
"FIRST"]]
```

```
    group_by_fields = "Name"

    # Statistical analysis to get the min length grouped by "Name"
    arcpy.Statistics_analysis(ODLines_CopyFeatures_all, output_stats_table,
stats_fields, group_by_fields)

    # The following lines of code convert the ParkID field from String to
Long so you can join with your park size table
    arcpy.AddField_management(output_stats_table, "ParkID", "DOUBLE")
    arcpy.CalculateField_management(output_stats_table, "ParkID",
'!FIRST_Park!', "PYTHON")
    arcpy.DeleteField_management(output_stats_table, [ "FIRST_Park",
"FREQUENCY"])
    arcpy.JoinField_management(output_stats_table, "ParkID", OS_shp,
"ORIG_FID", "AREA_M2")
    print 'done ' + str(l)


print "Appending address subsets"
for l in InitialSubsets[:-1]:
    arcpy.Append_management(["output_stats_table" + str(l) + ".dbf"],
"output_stats_table" + str(InitialSubsets[-1]) + ".dbf", "TEST","","")

arcpy.TableToTable_conversion ("output_stats_table" + str(InitialSubsets[-1])
+ ".dbf", ws, "output_stats_tableOSAll.dbf")

print "Appended"

print "deleting original files"

for l in InitialSubsets:
    arcpy.Delete_management("output_stats_table" + str(l) + ".dbf")
    arcpy.Delete_management("AddPnts" + str(l) + ".shp")

output_stats_table = "output_stats_tableOSAll.dbf"
AccessSum3 = "AccessSum3.dbf"

  #Process: Add Field
arcpy.AddField_management(output_stats_table, "Distance", "DOUBLE", "", "",
"", "", "NULLABLE", "NON_REQUIRED", "")

print 'Making minimum distance 10m'

 # Process: Calculate Field
arcpy.CalculateField_management(output_stats_table, "Distance", "!MIN_Total_!
if !MIN_Total_! > 10 else 10", "PYTHON_9.3", "")

#Process: Add Field
arcpy.AddField_management(output_stats_table, "Access3", "DOUBLE", "", "",
"", "", "NULLABLE", "NON_REQUIRED", "")
#
print 'Calculating Access Value'

#Process: Calculate Field
arcpy.CalculateField_management(output_stats_table, "Access3",
```

```
"math.pow(!AREA_M2!, 0.4) * math.pow( !Distance! , -0.15)", "PYTHON_9.3", "")

#Process: Summary Statistics
arcpy.Statistics_analysis(output_stats_table, AccessSum3, "Access3 SUM",
"FIRST_Addr")

print 'Joining back to original property points dataset'
# Process: Join Field
arcpy.JoinField_management(poly_Property, "RollNumber", AccessSum3,
"FIRST_ADDR", "OID;FREQUENCY; SUM_Access")


#Spatial separation measure
#Code for getting smallest distance to park
lowest_dist = "low_dist_table.dbf"
stats_fields = [["MIN_TOTAL_", "MIN"], ["FIRST_Addr", "FIRST"]]
group_by_fields = "FIRST_Addr"
# Statistical analysis to get the min length grouped by roll number
arcpy.Statistics_analysis(output_stats_table, lowest_dist, stats_fields,
group_by_fields)

##Cumulative Opportunities

print "starting buffer analysis"
bufferlist = [250, 500, 750, 1000]
for buffer in bufferlist:
    bufferstats = str (buffer) + ".dbf"
    currentbuffer = "buffer" + str(buffer) + ".dbf"
    select_query = '\"distance\" < ' + str(buffer)# Selecting buffer
    ##cumulative opportunities
    print "selecting buffer " + str(buffer)
    arcpy.TableSelect_analysis (output_stats_table, currentbuffer,
select_query)
    print "getting count of OS in buffer " +str(buffer)
    arcpy.Statistics_analysis(currentbuffer, bufferstats, "Distance COUNT",
"FIRST_Addr")
    print "renaming fields"
    arcpy.AddField_management(bufferstats, "OSin" + str(buffer), "DOUBLE")
    arcpy.CalculateField_management(bufferstats, "OSin" + str(buffer),
'!COUNT_Dist!', "PYTHON")
    bufferstats1 = str (buffer) + "_1.dbf"
    print "getting sum of opportunities in buffer " +str(buffer)
    arcpy.Statistics_analysis(currentbuffer, bufferstats1, "AREA_M2 SUM",
"FIRST_Addr")
    print "renaming fields"
    arcpy.AddField_management(bufferstats1, "OSm2in" + str(buffer), "DOUBLE")
    arcpy.CalculateField_management(bufferstats1, "OSm2in" + str(buffer),
'!SUM_AREA_M!', "PYTHON")
    print "done buffer " + str(buffer)
arcpy.MakeFeatureLayer_management(poly_Property, "Property_OSlayer")
arcpy.MakeFeatureLayer_management(poly_Property, "Property_OSlayer1")
for buffer in bufferlist:
    print "adding " + str(buffer) + " to polygon set"
    bufferstats = str(buffer) + ".dbf"
    bufferstats1 = str(buffer) + "_1.dbf"
```

```
    arcpy.JoinField_management(poly_Property, "RollNumber", bufferstats,
"FIRST_Addr", "OSin" + str(buffer))
    print "done OSin " + str(buffer)
    arcpy.JoinField_management(poly_Property, "RollNumber", bufferstats1,
"FIRST_Addr", "OSm2in" + str(buffer))
    print "done OSm2in " + str(buffer)
print "Done cumulative cumulative opportunity measures"
print "Adding lowest distance to polygon set"
arcpy.JoinField_management(poly_Property, "RollNumber", lowest_dist,
"FIRST_Addr")
arcpy.AddField_management(poly_Property, "Min_Dist", "DOUBLE", "", "", "",
"", "NULLABLE", "NON_REQUIRED", "")
arcpy.CalculateField_management(poly_Property, "Min_Dist", '!MIN_MIN_TO!',
"PYTHON")
arcpy.DeleteField_management(poly_Property, "MIN_MIN_TO")
arcpy.AddField_management(poly_Property, "No0_Dist", "DOUBLE", "", "", "",
"", "NULLABLE", "NON_REQUIRED", "")
arcpy.CalculateField_management(poly_Property, "No0_Dist", "!Min_Dist! if
!Min_Dist! > 0 else 0.0009", "PYTHON")
arcpy.AddField_management(poly_Property, "Dist_decay", "DOUBLE", "", "", "",
"", "NULLABLE", "NON_REQUIRED", "")
arcpy.CalculateField_management(poly_Property, "Dist_Decay",
'math.pow(!No0_Dist!, -0.25)', "PYTHON")
print "done"

arcpy.AddJoin_management("Property_OSlayer", "RollNumber", lowest_dist,
"FIRST_ADDR")
new_poly = "2014_Lots_OS_access.shp"
print "resaving polygon set as " + str(new_poly)
arcpy.CopyFeatures_management("Property_OSlayer1", new_poly)
arcpy.DeleteField_management(new_poly,
"April_KW_P;April_KW_1;April_KW_3;April_KW_4;April_KW_5;April_KW_6;April_KW_7
;April_KW_8;250_OID;250_FIRST_;250_FREQUE;250_COUNT_;250_1_OID;250_1_FIRS;250
_1_FREQ;250_1_SUM_;500_OID;500_FIRST_;500_FREQUE;500_COUNT_;500_1_OID;500_1_F
IRS;500_1_FREQ;500_1_SUM_;750_OID;750_FIRST_;750_FREQUE;750_COUNT_;750_1_OID;
750_1_FIRS;750_1_FREQ;1000_OID;1000_FIRST;1000_FREQU;1000_COUNT;1000_1_OID;10
00_1_FIR;1000_1_FRE;1000_1_SUM;low_dist_t;low_dist_1;low_dist_2;low_dist_4")
arcpy.AddJoin_management("Property_OSlayer1", "RollNumber", new_poly,
"April_KW_2")
new_poly1 = "Final_Lots_OS_access.shp"
arcpy.CopyFeatures_management("Property_OSlayer1", new_poly1)

for buffer in bufferlist:
    bufferstats = str (buffer) + ".dbf"
    bufferstats1 = str (buffer) + "_1.dbf"
    currentbuffer = "buffer" + str(buffer) + ".dbf"
    arcpy.Delete_management(bufferstats)
    arcpy.Delete_management(bufferstats1)
    arcpy.Delete_management(currentbuffer)
arcpy.Delete_management(new_poly)
from datetime import datetime
tend = datetime.now()
print 'runtime:'
```

```
print tend - tstart
```

APPENDIX 3: ROA Script

In STATA

```
ROA_Calculation*
Created by Robert Babin, 2016
```

```
set more off

//Script to run ROA and get output graphs
loc homeDir "C:\Insert"
loc homeDir1 "C:\Insert\Outputs"
cd `homeDir1'
mkdir RoA
log using RoA\Running_RoA.log, replace

//INPUTS TO SET:
loc Value1 "AdjValue"
loc Value2 "ValBysqft"
loc NbhdDef1 "num"
loc NbhdDef2 "ctuid"
loc NbhdDef3 "InCTC"
loc NbhdLabel1 "TAZ"
loc NbhdLabel2 "CT"
loc NbhdLabel3 "CTC"
loc time_period "TimeGroup"
loc TransData "AllTrans.dta"

//making directories for each neighbourhood and dependent variable specified
cd `homeDir1'\RoA
mkdir `NbhdLabel1'
cd `homeDir1'\RoA
cd `NbhdLabel1'
mkdir 1
mkdir 2
cd `homeDir1'\RoA
mkdir `NbhdLabel2'
cd `NbhdLabel2'
mkdir 1
mkdir 2
cd `homeDir1'\RoA
mkdir `NbhdLabel3'
cd `NbhdLabel3'
mkdir 1
mkdir 2
cd `homeDir'

//Running RoA on different dependent variables and neighbourhoods
forval DifDeps = 1/2 {
loc Value "`Value`DifDeps''"
forval NbhdNum = 1/3 {
```

```
loc neighbourhood "`NbhdDef`NbhdNum''"
use `TransData', clear //selecting transaction dataset to use
drop if dnu == 1 | PropType1 ~= 1
sum `Value' //displaying summary statistics for full dataset
egen nbhd_numbered = group(`NbhdDef`NbhdNum'')
quietly sum nbhd_numbered
local Nh = `r(max)'
quietly sum `time_period'
forval i = `r(min)'/`r(max)' {
forval j = 1/`Nh' {
quietly sum `time_period' if nbhd_numbered == `j' & `time_period' == `i'
quietly drop if r(N) < 2 & nbhd_numbered == `j' & `time_period' == `i'
}
}
collapse AVGin`time_period'=`Value', by (`time_period' `neighbourhood')
//taking means of all transaction values by time period and neighbourhood
egen nbhd_numbered = group(`neighbourhood') //to group by neighbourhood
save Nbhd_Means, replace //creating dataset of means by time_period and neighbourhood
sum nbhd_numbered, meanonly //summarizing dataset means by time_period and neighbourhood to get r(max)
local z = `r(max)' //creating z = to the number of neighbourhoods
forval i = 1/`z' { //looping through neighbourhoods to create rolling averages of each
use Nbhd_Means.dta, clear // using the file with means by time_period and neighbourhood
quietly sum nbhd_numbered if nbhd_numbered == `i'
ROA_FINAL* - Printed on 7/6/2016 11:53:20 AM
Page 2
if `r(N)' > 3 {
quietly keep if nbhd_numbered == `i' //dropping all except ith neighbourhoods
quietly tsset `time_period' //setting time variable as time_period
quietly tssmooth ma RollingAVG = AVGin`time_period', window(2 1 0) // making rolling average
quietly gen LastAvg = L.RollingAVG //creating variable of lagged RollingAvg value
quietly gen RoA_`NbhdLabel`NbhdNum''_`DifDeps' = (RollingAVG-LastAvg)/LastAvg*100 //calculating Rate change from previous time period
quietly save ROA`i', replace //saving each neighbourhood as individual datasets
local Q = `i' //defining Q as the last neighbourhood calculated
else {
display "ROA`i' does not exist"
}
}
}
use ROA`Q' // using last
local z = `Q'-1
forval i = 1/`z' { //looping through neighbourhoods except last to append
capture confirm file ROA`i'.dta
if _rc==0 {
```

```
append using ROA`i'.dta //combining all individual neighbourhood datasets with last ROA dataser
erase ROA`i'.dta // erasing individual neighbourhood datasets
}
else {
display "ROA`i' does not exist"
}
}
save ROA_`NbhdNum'`DifDeps', replace
erase ROA`Q'.dta // erasing last neighbourhood dataset
erase Nbhd_Means.dta // erasing mean Value by time_period and neighbourhood dataset
sort (`neighbourhood' `time_period')
egen NbhName = group(`neighbourhood'), label //making a variable to store labels of neighbouhood (works for strings)
sum NbhName, meanonly //getting r(max)
forval i = 1/`Q'{ //for each neighbourhood
quietly sum RoA_`NbhdLabel`NbhdNum''_`DifDeps' if nbhd_numbered == `i'
if `r(N)' > 1 {
graph twoway scatter RoA_`NbhdLabel`NbhdNum''_`DifDeps' `time_period' if nbhd_numbered == `i' , c(l) title("`NbhdLabel`NbhdNum''-`i'")
quietly graph export Outputs\RoA\\`NbhdLabel`NbhdNum''\\`DifDeps'\\`i'.png, replace //Saving scatterplot as "nbhd name"
display "`NbhdLabel`NbhdNum'' # `i' :"
sum RoA_`NbhdLabel`NbhdNum''_`DifDeps' RollingAVG if nbhd_numbered == `i' //summarizing results by neighbourhood
}
}
quietly sum `time_period'
forval i = `r(min)'/`r(max)' {
display "`time_period' `i' :"
sum RoA_`NbhdLabel`NbhdNum''_`DifDeps' RollingAVG if `time_period' == `i'
}
//merging ROA values to transaction data
use `TransData', clear //loading the original transaction dataset
merge m:1 `time_period' `neighbourhood' using ROA_`NbhdNum'`DifDeps', keepusing(RoA_`NbhdLabel`NbhdNum''_`DifDeps') //Adding RoA to each observation
drop _merge
save `TransData', replace //Saving new dataset as TransData
sum RoA_`NbhdLabel`NbhdNum''_`DifDeps'
}
}
log cl
```

APPENDIX 4: Correlation Matrix

	OS Access, Gravity Based	OS Access, count in 1000 metres	OS Access,Sum of Area in 1000 metres	OS Access, Spatial Separation	OS Adjacent	Transit Access	Employment Access	Living Area, square feet	Yard Size, square feet	Frontage, feet	BuildAge	Property Type	In CTC	100 Residents in Dissemination	Public School Quality	Walkability	Population per square kilometre	Rate of Appreciation in Census Tract	EDU_RATE	Kitchener/Waterloo	Regional Road Adjacent	b4_LRT	LRT_Aprv1	LRT_Aprv2	Adjusted Value, dollars	Assessed Value, dollars (2014)	Logged Assessed Value	Logged Adjusted Value
OS Access, Gravity Based	1																											
OS Access, count in 1000 metres	0.7	1																										
OS Access,Sum of Area in 1000 metres	0.59	0.25	1																									
OS Access, Spatial Separation	0.42	0.22	0.19	1																								
OS Adjacent	0.16	0.04	0.07	0.29	1																							
Transit Access	0.2	0.28	-0.11	0.07	-0.04	1																						
Employment Access	-0.02	0.23	-0.27	-0.07	-0.06	0.44	1																					
Living Area, square feet	0.09	-0.02	0.22	0.11	0.16	-0.33	-0.27	1																				
Yard Size, square feet	-0.05	-0.06	-0.08	-0.07	0.09	-0.06	0.05	0.33	1																			
Frontage, feet	0.01	0.01	-0.03	-0.02	0.03	0.03	0.08	0.19	0.43	1																		
BuildAge	-0.02	0.21	-0.29	-0.11	-0.04	0.36	0.6	-0.29	0.23	0.21	1																	
Property Type	0.03	0.03	0.04	0.05	-0.05	0.09	0.01	-0.28	-0.32	-0.24	-0.12	1																
In CTC	-0.06	0.12	-0.17	-0.05	-0.05	0.12	0.3	-0.16	-0.03	0.02	0.49	0.03	1															

Police Calls per 100 Residents	-0.09	0.04	-0.17	-0.1	-0.04	0.09	0.26	-0.16	0.03	0.03	0.38	0.02	0.37	1														
Public School Quality	0.21	0.14	0.34	0.06	0.06	-0.1	-0.12	0.3	-0.02	-0.02	-0.39	-0.03	-0.21	-0.24	1													
Walkability	0.1	0.34	-0.21	-0.03	-0.06	0.49	0.66	-0.32	-0.01	0.06	0.68	0.03	0.48	0.34	-0.22	1												
Population per square kilometre	0.06	0.23	-0.2	-0.05	-0.1	0.4	0.33	-0.27	-0.09	-0.03	0.25	0.09	0.07	-0.09	-0.09	0.49	1											
Rate of Appreciation in Census Tract	-0.02	0	0.06	0.02	0.02	-0.09	0.03	0.05	0.04	0	0.02	0	0.15	0.11	-0.01	0.06	-0.01	1										
EDU_RATE	0.24	0.17	0.41	0.14	0.04	-0.21	-0.26	0.39	-0.06	-0.03	-0.41	0.03	-0.19	-0.37	0.49	-0.33	-0.23	0	1									
Kitchener/Waterloo	0.28	0.27	0.35	0.11	0.05	0.13	0.21	0.2	0.06	0.07	-0.07	0.03	-0.05	-0.08	0.5	0.07	-0.02	0.05	0.53	1								
Regional Road Adjacent	0.05	0.07	-0.02	-0.02	-0.05	0.11	0.1	-0.08	0.05	0.02	0.1	0.05	0.05	0.06	-0.03	0.09	0.03	0.01	-0.03	0	1							
b4_LRT	0.03	0.02	0.02	0	0	0.04	-0.01	0.03	0.02	0.04	-0.09	-0.05	0	-0.01	0.05	0.02	0.03	-0.08	0.02	0.02	0.01	1						
LRT_Aprv1	0	0.03	-0.02	0.01	0.01	0.01	0.03	-0.01	-0.01	-0.02	0	0.01	0	0	-0.01	0.02	0.01	0.24	-0.02	-0.02	0.01	-0.38	1					
LRT_Aprv2	-0.03	-0.04	-0.01	-0.01	-0.01	-0.04	-0.01	-0.02	-0.01	-0.02	0.09	0.04	0	0.01	-0.04	-0.04	-0.04	-0.11	-0.01	0	-0.01	-0.66	-0.44	1				
Adjusted Value, dollars	0.05	-0.06	0.21	0.08	0.17	-0.43	-0.29	0.75	0.41	0.2	-0.27	-0.29	-0.13	-0.1	0.23	-0.34	-0.32	0.12	0.36	0.16	-0.08	-0.14	0.01	0.14	1			
Assessed Value, dollars (2014)	0.03	-0.09	0.2	0.07	0.17	-0.48	-0.34	0.81	0.43	0.21	-0.32	-0.32	-0.15	-0.11	0.24	-0.4	-0.35	0.11	0.38	0.14	-0.08	0	-0.01	0	0.92	1		
Logged Assessed Value	0.04	-0.09	0.22	0.08	0.17	-0.47	-0.36	0.82	0.4	0.2	-0.37	-0.38	-0.18	-0.14	0.3	-0.43	-0.37	0.09	0.42	0.16	-0.09	0.01	-0.01	0	0.89	0.97	1	
Logged Adjusted Value	0.05	-0.07	0.21	0.09	0.16	-0.41	-0.31	0.75	0.36	0.18	-0.31	-0.33	-0.16	-0.13	0.28	-0.37	-0.33	0.1	0.39	0.18	-0.09	-0.17	0.01	0.16	0.95	0.86	0.89	1

APPENDIX 5: Level-Level Regression Results

SINGLE DETACHED MODELS

REGRESSION

SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)

```
Data set            :Single_Detached.dbf
Weights matrix      :Single_Detached1.gwt
Dependent Variable  :    AdjValue            Number of Observations:      21801
Mean dependent var  :    351.2168            Number of Variables   :         25
S.D. dependent var  :    124.4542            Degrees of Freedom    :      21776
Pseudo R-squared    :      0.7888
Spatial Pseudo R-squared:  0.7009
N. of iterations    :           1            Step1c computed       :         No
```

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	-6.8541659	10.5413234	-0.6502187	0.5155509
BuildAge	-0.4631989	0.0862432	-5.3708432	0.0000001
C_LRT1	1.6459873	4.0034148	0.4111458	0.6809656
C_LRT2	5.3163765	3.5284965	1.5066974	0.1318882
C_OS	-0.0688907	0.1834811	-0.3754647	0.7073149
C_Trns	0.0200097	0.0116015	1.7247541	0.0845718
C_Walk	2.2692495	1.2448358	1.8229309	0.0683139
InCTC	14.4805490	8.3186296	1.7407373	0.0817296
LRT_Aprv1	28.5968597	1.2104937	23.6241287	0.0000000
LRT_Aprv2	50.2725679	1.0349598	48.5744168	0.0000000
ResT	0.0375732	0.0011874	31.6434637	0.0000000
RoA_CT_1	1.5322378	0.1828928	8.3777900	0.0000000
Trans_OS	-0.0003564	0.0002242	-1.5898305	0.1118730
W_AdjValue	0.5554597	0.0260822	21.2965359	0.0000000
WalkIndex	1.7467572	0.9017511	1.9370725	0.0527365
Walk_OS	0.0142081	0.0188737	0.7527950	0.4515731
Walk_Trans	0.0031241	0.0012559	2.4876074	0.0128606
YardSize	0.0065244	0.0023744	2.7477556	0.0060005
Yard_OS	0.0000243	0.0000441	0.5506189	0.5818950
edu_rate	0.2692281	0.1376099	1.9564589	0.0504111
os_adj	17.4558765	1.7513548	9.9670702	0.0000000
popperkm2	-0.0009724	0.0003269	-2.9749340	0.0029305
regroad	-14.2197563	2.4349646	-5.8398206	0.0000000
sum_acce_2	-0.1868685	0.1930338	-0.9680612	0.3330138
transit	0.0036119	0.0090601	0.3986642	0.6901406
lambda	0.3269978	0.0366048	8.9331891	0.0000000

```
Instrumented: W_AdjValue
Instruments: W_BuildAge, W_C_LRT1, W_C_LRT2, W_C_OS, W_C_Trns, W_C_Walk,
             W_InCTC, W_LRT_Aprv1, W_LRT_Aprv2, W_ResT, W_RoA_CT_1,
             W_Trans_OS, W_WalkIndex, W_Walk_OS, W_Walk_Trans, W_YardSize,
             W_Yard_OS, W_edu_rate, W_os_adj, W_popperkm2, W_regroad,
             W_sum_acce_2, W_transit
```

REGRESSION

SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)

```
Data set            :Single_Detached.dbf
Weights matrix      :Single_Detached1.gwt
Dependent Variable  :     RealtyT            Number of Observations:      21801
Mean dependent var  :    349.1843            Number of Variables   :         25
S.D. dependent var  :    115.0995            Degrees of Freedom    :      21776
Pseudo R-squared    :      0.8664
Spatial Pseudo R-squared:  0.7834
N. of iterations    :           1            Step1c computed       :         No
```

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	47.8972496	7.9928745	5.9924936	0.0000000
BuildAge	-0.6711828	0.0805898	-8.3283880	0.0000000
C_LRT1	3.1542464	2.6174264	1.2050946	0.2281668
C_LRT2	2.1696996	2.1863148	0.9924004	0.3210022
C_OS	-0.0970979	0.1471707	-0.6597641	0.5094052
C_Trns	0.0184197	0.0112494	1.6373894	0.1015491
C_Walk	2.4671213	1.1302516	2.1828071	0.0290500
InCTC	19.6934090	6.7969562	2.8973865	0.0037629
LRT_Aprv1	0.8865395	0.8968871	0.9884628	0.3229260
LRT_Aprv2	6.3098057	0.7941493	7.9453648	0.0000000
ResT	0.0397959	0.0009179	43.3561184	0.0000000
RoA_CT_1	0.5257501	0.1300425	4.0429089	0.0000528
Trans_OS	-0.0001990	0.0002002	-0.9942221	0.3201148
W_RealtyT	0.5179778	0.0260252	19.9029698	0.0000000
WalkIndex	1.9951866	0.8760441	2.2774956	0.0227566
Walk_OS	0.0275926	0.0184898	1.4923171	0.1356160
Walk_Trans	0.0039503	0.0012413	3.1823523	0.0014608
YardSize	0.0061713	0.0015689	3.9334457	0.0000837
Yard_OS	0.0000374	0.0000259	1.4444568	0.1486105
edu_rate	0.0689040	0.1310785	0.5256697	0.5991177
os_adj	14.2293762	1.3311772	10.6893182	0.0000000
popperkm2	-0.0015743	0.0003008	-5.2329318	0.0000002
regroad	-10.1764687	1.8642254	-5.4588190	0.0000000
sum_acce_2	-0.2180417	0.1226087	-1.7783541	0.0753457
transit	-0.0036925	0.0088846	-0.4156096	0.6776957
lambda	0.5591087	0.0245738	22.7522191	0.0000000

```
Instrumented: W_RealtyT
Instruments: W_BuildAge, W_C_LRT1, W_C_LRT2, W_C_OS, W_C_Trns, W_C_Walk,
             W_InCTC, W_LRT_Aprv1, W_LRT_Aprv2, W_ResT, W_RoA_CT_1,
             W_Trans_OS, W_WalkIndex, W_Walk_OS, W_Walk_Trans, W_YardSize,
             W_Yard_OS, W_edu_rate, W_os_adj, W_popperkm2, W_regroad,
             W_sum_acce_2, W_transit
```

SEMI DETACHED AND DUPLEX MODELS

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
------------------------------------------------------------------
Data set            :Semi_Detached.dbf
Weights matrix      :Semi_Detached.gwt
Dependent Variable  :    AdjValue               Number of Observations:       2702
Mean dependent var  :    244.6261               Number of Variables   :         25
S.D. dependent var  :     51.4563               Degrees of Freedom    :       2677
Pseudo R-squared    :      0.5320
Spatial Pseudo R-squared:  0.4388
N. of iterations    :           1               Step1c computed       :         No
```

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	34.9419410	13.6595757	2.5580546	0.0105260
BuildAge	-0.3477533	0.0705085	-4.9320738	0.0000008
C_LRT1	0.2999593	6.1275733	0.0489524	0.9609572
C_LRT2	4.9101146	5.7012163	0.8612398	0.3891060
C_OS	0.3205293	0.2181849	1.4690717	0.1418133
C_Trns	-0.0183608	0.0190650	-0.9630661	0.3355143
C_Walk	-0.3230621	1.3843658	-0.2333647	0.8154782
InCTC	-15.5416353	9.4291709	-1.6482505	0.0993013
LRT_Aprv1	24.5060865	1.8280649	13.4054792	0.0000000
LRT_Aprv2	35.2464649	1.5798668	22.3097697	0.0000000
ResT	0.0307236	0.0028135	10.9200347	0.0000000
RoA_CT_1	0.4142702	0.3427965	1.2085018	0.2268543
Trans_OS	-0.0001610	0.0003560	-0.4521122	0.6511881

W_AdjValue	0.5037130	0.0603619	8.3448876	0.0000000
WalkIndex	2.7520205	1.4542619	1.8923830	0.0584400
Walk_OS	-0.0185611	0.0272179	-0.6819446	0.4952740
Walk_Trans	-0.0009320	0.0038538	-0.2418323	0.8089101
YardSize	0.0081145	0.0032114	2.5268027	0.0115106
Yard_OS	0.0000681	0.0000671	1.0157083	0.3097683
edu_rate	0.1752331	0.1782929	0.9828385	0.3256869
os_adj	-0.9652640	2.8220630	-0.3420420	0.7323193
popperkm2	-0.0015088	0.0005027	-3.0012491	0.0026887
regroad	-5.4147209	2.8869007	-1.8756173	0.0607079
sum_acce_2	-0.3154600	0.1861350	-1.6947917	0.0901150
transit	-0.0073630	0.0185323	-0.3973088	0.6911397
lambda	0.0477886	0.1006093	0.4749921	0.6347926

Instrumented: W_AdjValue
Instruments: W_BuildAge, W_C_LRT1, W_C_LRT2, W_C_OS, W_C_Trns, W_C_Walk, W_InCTC, W_LRT_Aprv1, W_LRT_Aprv2, W_ResT, W_RoA_CT_1, W_Trans_OS, W_WalkIndex, W_Walk_OS, W_Walk_Trans, W_YardSize, W_Yard_OS, W_edu_rate, W_os_adj, W_popperkm2, W_regroad, W_sum_acce_2, W_transit

REGRESSION

SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)

Data set	:Semi_Detached.dbf		
Weights matrix	:Semi_Detached.gwt		
Dependent Variable :	RealtyT	Number of Observations:	2702
Mean dependent var :	240.5766	Number of Variables :	25
S.D. dependent var :	45.4525	Degrees of Freedom :	2677
Pseudo R-squared :	0.7166		
Spatial Pseudo R-squared:	0.5159		
N. of iterations :	1	Step1c computed :	No

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	24.2755686	11.3939003	2.1305758	0.0331241
BuildAge	-0.3420386	0.0506646	-6.7510313	0.0000000
C_LRT1	-3.5712597	3.7897595	-0.9423447	0.3460162
C_LRT2	2.1183194	3.7678147	0.5622143	0.5739700
C_OS	0.0626403	0.1910265	0.3279141	0.7429766
C_Trns	-0.0038571	0.0215620	-0.1788825	0.8580300
C_Walk	1.8968599	1.1401384	1.6637102	0.0961704
InCTC	-2.4812927	8.5672876	-0.2896241	0.7721039
LRT_Aprv1	3.0210748	1.3879114	2.1767058	0.0295025
LRT_Aprv2	2.5781016	1.1333043	2.2748538	0.0229147
ResT	0.0264512	0.0020196	13.0971139	0.0000000
RoA_CT_1	0.0741002	0.2404924	0.3081185	0.7579921
Trans_OS	-0.0000269	0.0002568	-0.1045900	0.9167012
W_RealtyT	0.6168664	0.0504677	12.2229886	0.0000000
WalkIndex	0.9186790	1.2162352	0.7553465	0.4500411
Walk_OS	0.0091354	0.0247781	0.3686900	0.7123588
Walk_Trans	0.0015197	0.0027999	0.5427493	0.5873024
YardSize	0.0099727	0.0020491	4.8667353	0.0000011
Yard_OS	0.0000828	0.0000448	1.8492988	0.0644147
edu_rate	0.1359523	0.1905898	0.7133240	0.4756453
os_adj	1.5061921	2.0174798	0.7465711	0.4553225
popperkm2	-0.0010787	0.0003945	-2.7344371	0.0062487
regroad	-7.0298095	1.9785958	-3.5529285	0.0003810
sum_acce_2	-0.2309869	0.1452506	-1.5902649	0.1117751
transit	-0.0072372	0.0128353	-0.5638485	0.5728572
lambda	0.2067724	0.0953921	2.1676053	0.0301887

Instrumented: W_RealtyT
Instruments: W_BuildAge, W_C_LRT1, W_C_LRT2, W_C_OS, W_C_Trns, W_C_Walk, W_InCTC, W_LRT_Aprv1, W_LRT_Aprv2, W_ResT, W_RoA_CT_1, W_Trans_OS, W_WalkIndex, W_Walk_OS, W_Walk_Trans, W_YardSize, W_Yard_OS, W_edu_rate, W_os_adj, W_popperkm2, W_regroad, W_sum_acce_2, W_transit

TOWNHOUSE MODELS

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
------------------------------------------------------------------
Data set            :Townhouse.dbf
Weights matrix      :File: Townhouse.gwt
Dependent Variable  :    AdjValue                Number of Observations:        2370
Mean dependent var  :    262.9464                Number of Variables   :          25
S.D. dependent var  :     54.2523                Degrees of Freedom    :        2345
Pseudo R-squared    :      0.7577
Spatial Pseudo R-squared:  0.6660
N. of iterations    :           1                Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT      41.9159817      17.7577808       2.3604291       0.0182538
            BuildAge      -0.5062994       0.1954240      -2.5907741       0.0095760
              C_LRT1      65.2899827      20.8682254       3.1286792       0.0017559
              C_LRT2      47.9939961      17.9478608       2.6740789       0.0074935
                C_OS       0.6330740       0.6004000       1.0544204       0.2916905
              C_Trns      -0.0074146       0.0190584      -0.3890469       0.6972415
              C_Walk       1.9705214       4.5319413       0.4348074       0.6637023
               InCTC     -83.6061802      33.6553315      -2.4841883       0.0129847
           LRT_Aprv1      27.0168890       1.5753540      17.1497255       0.0000000
           LRT_Aprv2      44.5662418       1.6694724      26.6948054       0.0000000
                ResT       0.0357535       0.0039901       8.9604554       0.0000000
            RoA_CT_1       0.1507965       0.4736289       0.3183853       0.7501927
            Trans_OS      -0.0005139       0.0002915      -1.7631751       0.0778710
          W_AdjValue       0.4656532       0.0585611       7.9515732       0.0000000
           WalkIndex       0.4581073       1.5154655       0.3022882       0.7624324
             Walk_OS       0.0591301       0.0317573       1.8619343       0.0626124
          Walk_Trans      -0.0024164       0.0016122      -1.4987611       0.1339356
            YardSize       0.0057925       0.0025821       2.2432933       0.0248779
             Yard_OS       0.0000680       0.0000579       1.1734899       0.2405994
            edu_rate      -0.0193795       0.2075637      -0.0933665       0.9256124
              os_adj      -2.0302222       3.0040694      -0.6758240       0.4991524
           popperkm2      -0.0022972       0.0008555      -2.6852147       0.0072483
             regroad      -3.7910329       3.6583566      -1.0362666       0.3000778
          sum_acce_2      -0.0008378       0.1249674      -0.0067045       0.9946506
             transit      -0.0069332       0.0118663      -0.5842732       0.5590365
              lambda       0.4900199       0.0824556       5.9428309       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_AdjValue
Instruments: W_BuildAge, W_C_LRT1, W_C_LRT2, W_C_OS, W_C_Trns, W_C_Walk,
             W_InCTC, W_LRT_Aprv1, W_LRT_Aprv2, W_ResT, W_RoA_CT_1,
             W_Trans_OS, W_WalkIndex, W_Walk_OS, W_Walk_Trans, W_YardSize,
             W_Yard_OS, W_edu_rate, W_os_adj, W_popperkm2, W_regroad,
             W_sum_acce_2, W_transit

REGRESSION
----------
SUMMARY OF OUTPUT: SPATIALLY WEIGHTED TWO STAGE LEAST SQUARES (HET)
------------------------------------------------------------------
Data set            :Townhouse.dbf
Weights matrix      :File: Townhouse.gwt
Dependent Variable  :     RealtyT                Number of Observations:        2370
Mean dependent var  :    255.7524                Number of Variables   :          25
S.D. dependent var  :     44.8569                Degrees of Freedom    :        2345
Pseudo R-squared    :      0.7858
Spatial Pseudo R-squared:  0.6869
N. of iterations    :           1                Step1c computed       :          No

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT     117.3412908      23.0304025       5.0950604       0.0000003
            BuildAge      -1.1623867       0.2113148      -5.5007360       0.0000000
```

C_LRT1	25.8645856	11.6944020	2.2117066	0.0269869
C_LRT2	-2.2661994	9.3676288	-0.2419181	0.8088436
C_OS	0.5164709	0.5451751	0.9473487	0.3434611
C_Trns	-0.0059135	0.0210822	-0.2804988	0.7790948
C_Walk	1.9809520	3.8578611	0.5134845	0.6076124
InCTC	-33.0901279	27.5756813	-1.1999750	0.2301491
LRT_Aprv1	5.0956130	1.1625305	4.3832079	0.0000117
LRT_Aprv2	9.2662175	1.5467770	5.9906617	0.0000000
ResT	0.0385457	0.0035061	10.9939588	0.0000000
RoA_CT_1	-0.4725138	0.2766614	-1.7079134	0.0876524
Trans_OS	-0.0002678	0.0002514	-1.0650527	0.2868521
W_RealtyT	0.3264711	0.0732146	4.4590994	0.0000082
WalkIndex	3.4064238	2.0078626	1.6965423	0.0897832
Walk_OS	0.0027032	0.0338987	0.0797440	0.9364409
Walk_Trans	-0.0002494	0.0018725	-0.1331837	0.8940481
YardSize	0.0111579	0.0018908	5.9012125	0.0000000
Yard_OS	0.0000131	0.0000414	0.3162637	0.7518023
edu_rate	-0.3594540	0.2339762	-1.5362848	0.1244685
os_adj	0.8657108	1.7644221	0.4906483	0.6236752
popperkm2	-0.0034537	0.0012632	-2.7339946	0.0062571
regroad	-1.9633734	3.1660721	-0.6201291	0.5351728
sum_acce_2	-0.0069621	0.1144539	-0.0608291	0.9514953
transit	-0.0157734	0.0120241	-1.3118138	0.1895830
lambda	0.7431284	0.0562992	13.1996348	0.0000000

```
Instrumented: W_RealtyT
Instruments: W_BuildAge, W_C_LRT1, W_C_LRT2, W_C_OS, W_C_Trns, W_C_Walk,
             W_InCTC, W_LRT_Aprv1, W_LRT_Aprv2, W_ResT, W_RoA_CT_1,
             W_Trans_OS, W_WalkIndex, W_Walk_OS, W_Walk_Trans, W_YardSize,
             W_Yard_OS, W_edu_rate, W_os_adj, W_popperkm2, W_regroad,
             W_sum_acce_2, W_transit
```