Analyzing housing market dynamics and residential location choices concurrent with light-rail transit investment in Kitchener-Waterloo, Canada

by

Yu Huang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Planning

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Christopher Higgins Assistant Professor, Dept. of Human Geography University of Toronto Scarborough
Supervisor(s):	Dawn Parker Professor, School of Planning, University of Waterloo
Internal Member:	Leia Minaker Assistant professor, School of Planning, University of Waterloo
Internal-External Member:	Derek Robinson Associate professor, Dept. of Geography, University of Waterloo
Other Member(s):	Paul Anglin Professor, Dept. of Marketing & Consumer Studies University of Guelph

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contributions

This thesis consists of three manuscripts written for publication. As the first author for all manuscripts, Yu Huang initiated the studies, conducted literature review, designed and conducted the housing survey, conducted data manipulation, performed data analyses, prepared figures and tables, and drafted and revised the manuscripts.

This thesis is based, in part, on data provided by the Municipal Property Assessment Corporation. Any opinions, findings, conclusions or recommendations expressed in this material are those solely of the author(s) and are not necessarily the views of the Municipal Property Assessment Corporation. Contributions to the three manuscripts in Chapter 2, Chapter 3, and Chapter 4 are described below.

Chapter 2: *A spatio-temporal multilevel housing price model: integrating the spatial and temporal dependence and neighbourhood effects*

This paper was co-authored with Dr. Dawn Parker and Robert Babin. Dr. Parker provided feedback on model specification, data analysis and model interpretation. Robert Babin developed many of the independent variables (including the concept, calculation scripts, and sensitivity analysis) and originally estimated the baseline model for this paper.

Chapter 3: *Identifying heterogeneous residential preferences during the construction of a new light-rail transit line*

This paper was co-authored with Dr. Dawn Parker and Dr. Paul Anglin. Dr. Parker and Dr. Anglin both provided constructive feedback on the paper structure, writing and revision.

Chapter 4: *Who Prefers to Live in Transit-Oriented Development Areas? Evidence from a Residential Location Choice Survey in Canada*

This paper was co-authored with Dr. Dawn Parker and Dr. Leia Minaker. Dr. Parker and Dr. Minaker both provided valuable feedback on the paper structure and revision.

Abstract

Transit investment and transit-oriented development (TOD) have become the predominant planning policies to manage growth and limit sprawl. Waterloo Region implemented a light-rail transit (LRT) system aiming to provide alternative transit options and shape urban communities. Meanwhile, as one of the most fast-growing urban areas, the region has experienced rapid growth in population and employment. The booming high-tech industries, the international immigrants and migrants from the Greater Toronto Area (GTA) have all contributed to the increasing attractiveness of the region and its changing demographics, which in turn have heavily shifted the housing markets in the region. The housing prices have risen dramatically since 2014 and reached a peak in 2017 when the average sales price increased by over twenty percent from 2016. These changes occurring in the region have motivated this thesis to investigate 1) How have different housing markets in the region reacted to the LRT investment? 2) How might the LRT investment have influenced the residential location choices of various households? 3) Who might hold strong preferences for living in the TOD area? This thesis addresses these questions through three empirical studies.

The first study presents a spatio-temporal autoregressive multilevel model to better examine the relationship between housing characteristics, transit investment and housing prices. The proposed model is expected to improve the purely spatial hedonic price modelling in three aspects: i) controlling for both the spatial and temporal relations on housing price determination, i.e., the dependence on "recent comparable sales"; ii) considering the nesting structure of housing in neighbourhoods; and iii) accounting for the neighbourhood-level spatial interactions. Using 68,258 housing transactions occurring in Kitchener-Waterloo (KW) during 2005-2018, this study finds better performance of the proposed models and provides strong evidence of the three distinct effects that underly the price generating process. According to the preferred model results, this study finds significant housing price increase in the central-transit corridor (CTC), compared to housing outside the CTC, while the impacts vary for different housing types at different stages of the LRT implementation process.

The second study seeks to delineate the housing demand structure in the region during the LRT construction. To this end, this research conducted a housing survey in KW through 2016-2017 and obtained 357 complete responses from homebuyers. Based on the survey data, this study performs a second-stage demand analysis and reports heterogeneous preference estimates of different demographics for dwelling and neighbourhood attributes. Household structure and age seem to be the major demand shifters. This study also finds that both couples without children and seniors aged 55 and over are more willing to pay for the CTC area.

The third study aims to identify household groups holding different preferences for TOD. Based on the survey responses regarding the importance of TOD features in residential location choices, this study conducts a latent-class analysis (LCA) and finds that 36.2 percent of households (primarily couples with children and with medium income) in our sample show a strong desire for TOD features, including LRT access, bus access, walkability, ease to cycle, access to urban centre and access to open space, although they purchased outside the CTC. This indicates a possible undersupply of housing in the CTC for these families with children. Through further examination of their preferences for other housing attributes, this study finds the adequate living space, garage and school quality are more important to these households.

This thesis provides updated knowledge on housing market dynamics, housing demand and TOD preferences, which may help inform housing policies in the region to provide home options for a wide range of households inside and outside the central transit corridor and thus create vibrant and complete communities.

Acknowledgements

I would like to acknowledge three funding sources for this project: 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; and China scholarship programme.

I would like to acknowledge the Municipal Property Assessment Corporation (MPAC) and the Teranet company for providing the housing transaction data. I would like to thank Shelley Graham and Robert Babin from MPAC for continuously assisting me on data-related questions and offering professional knowledge on housing markets. This thesis would not be possible without the data provided by MPAC and the insights and help offered by the staff from MPAC.

I would like to thank my supervisor Prof. Dawn Parker. Without your guidance and support, this thesis would not have been possible. I appreciate all the conversations with you and all your handwritten comments on my papers. Thank you for guiding me on academic writing, proposal preparation, ethics application, and research conduction in many aspects. Thank you for believing in me all the way. Your encouragement and support have strengthened my confidence and prompted me to move forward. I appreciate the time working with you. Your vision and inspiration will continue to impact my academic journey. I would like to thank my committee members and all the professors and staff from the Faculty of Environment who helped and inspired me in many ways. Thank you for my colleagues, officemates, and all of my friends in Waterloo.

I would like to thank my parents. Dad and Mom, I am truly grateful to have you as my parents. Thank you for your unconditional support and love for me. You are the best gifts of my life. Thank you for offering me the best opportunities to obtain knowledge and always encouraging me to pursue higher education. Without your strong backing, I would not be able to study abroad and complete a doctoral program. I would like to thank my brother. You are the most amazing brother. Thanks for loving me and always being proud

of me. You know what, I am also very proud of you. I would like to thank my husband. Thank you for your understanding, patience, love, and company. I appreciate all the efforts that you made to better our life.

Finally, I would like to thank my grandma. Thank you for taking care of me for almost ten years. Although I missed the chance to see you again, I will always remember you and have you in my heart.

Table of Contents

List of Tables xiii

List of Figures xv

1 Introduction 1

1.1 Scope of the thesis 1

1.2 Context 2

1.2.1 Location and Policy Context 2

1.2.2 Housing market introduction 3

1.3 Research objectives 5

1.4 Research questions 6

1.5 Thesis outline 7

2 A spatio-temporal multilevel housing price model: integrating the spatial and temporal dependence and neighbourhood effects 8

2.1 Introduction 9

2.2 Spatio-temporal hedonic modelling 11

2.2.1 Spatial hedonic modelling 11

2.2.2 Spatio-temporal hedonic modelling 13

2.3 Methodology 15

2.3.1 Model specifications 16

2.3.2 Estimation method 22
2.4 Data 23
2.4.1 Market description 23
2.4.2 Data preparation 23
2.5 Results 28
2.5.1 Condo housing models 28
2.5.2 Single-detached housing models 33
2.5.3 Semi-detached/duplex housing models 38
2.5.4 Townhouse models 41
2.6 Conclusion and discussion 43
2.6.1 Synthesis of key findings 43
2.6.2 Discussion 44

3 Identifying heterogeneous residential preferences during the construction of a new light-rail transit line 48

3.1 Introduction 49
3.2 Theoretical foundations 51
3.2.1 A model of residential location choice 51
3.2.2 Preference identification 52
3.2.3 Estimation method 54
3.3 Study area and data 60
3.3.1 Study area 60
3.3.2 Data 61
3.4 Results 65
3.4.1 First-stage hedonic results 65
3.4.2 Heterogeneous residential preferences 67
3.5 Discussion 73

3.5.1 LRT development, property value changes and residential location choices . . . 73
3.5.2 Caveats . . . 76

4 Who prefers to live in Transit-Oriented Development area? Evidence from a residential location choice survey in Canada 77

4.1 Introduction . . . 78
4.2 Literature Review . . . 80
4.2.1 TOD and residential location choices . . . 80
4.2.2 TOD and socio-demographics . . . 80
4.3 Data and Methods . . . 82
4.3.1 Research context . . . 82
4.3.2 Data collection . . . 82
4.3.3 Latent class modelling . . . 84
4.4 Findings . . . 87
4.4.1 Identifying preference classes . . . 87
4.4.2 Travel mode choice . . . 91
4.4.3 Demographic profiles . . . 92
4.4.4 Moving motivation . . . 92
4.4.5 Stated preferences for housing attributes . . . 94
4.4.6 Residential location choices . . . 96
4.5 Discussion . . . 98
4.5.1 Current TOD households . . . 98
4.5.2 Missing target in TOD housing . . . 98
4.5.3 Caveats and Future Research . . . 99

5 Conclusion 100

5.1 Key findings . . . 100

5.2 Contributions and planning implications 104
5.2.1 Shaping communities in the corridor by providing a variety of home options 104
5.2.2 Creating complete communities by satisfying various housing demand 106
5.3 Limitations and future work 107

References **108**

Appendices **125**

A1 Appendices in Chapter 2 125
A1.1 Semi-variograms 125
A1.2 Data source 127
A1.3 Descriptive statistics of variables 128
A1.4 Correlation matrix 129
A2 Appendices in Chapter 3 130
A2.1 Hedonic demand identification 130
A2.2 Survey sampling 132
A2.3 Data source 134
A2.4 Linear regression results 135
A2.5 Stated preference from the housing survey 136
A2.6 Correlation matrix 137
A3 Appendices in Chapter 4 138
A3.1 Estimated class probabilities from the LCA 138
A3.2 Density plots of class probabilities 148
A4 R code for Chapter 2 148
A5 R code for Chapter 3 154
A6 R code for Chapter 4 168
A7 Housing survey questionnaire 192

List of Tables

2.1 Description of variables 25

2.2 Summary of model statistics - condominium housing 29

2.3 Estimates of coefficients from the four models - condo housing 31

2.4 Estimates for the variables of interest from the four models - condo housing 33

2.5 Summary of model statistics - single-detached housing 34

2.6 Estimates for the variables of interest from the four models - single-detached houses 36

2.7 Estimates of coefficients from the four models - single-detached housing . . 37

2.8 Summary of model statistics - semi-detached/duplex housing 38

2.9 Estimates of coefficients from the four models - semi-detached/duplex housing 40

2.10 Summary of model statistics - townhouses 41

2.11 Estimates of coefficients from the four models - townhouses 42

3.1 Descriptive statistics for the variables used in the study 62

3.2 Summary statistics - homebuyers and comparison with the Census statistics 64

3.3 Estimates from the first-stage nonparametric hedonic regression 66

3.4 Estimates of the willingness to pay for private yard 67

3.5 Willingness to pay for the structural housing attributes 68

3.6 Probit estimates of the demand for single-detached housing 70

3.7 Probit estimates of the demand for the CTC 71

3.8 Willingness to pay for the locational and neighbourhood attributes 72

4.1 Statistics for latent class models 87
4.2 LCA results with estimated class-specific response probabilities and odds ratios 88
4.3 Demographic profiles of three groups 93
4.4 Residential location choices of three groups 97

A-1 Data source 127
A-2 Descriptive statistics of variables 128
A-3 Correlation Matrix for the single-detached housing observations 129
A-4 Variables and data source 134
A-5 Linear regression results 135
A-6 Estimated class probabilities of each survey respondent - LCA results . . . 138

List of Figures

1.1 Residential housing market trends in KW over the years from 2008 to 2018 4

2.1 Four model specifications by different weight matrices and neighbourhood effect control 16

2.2 A two-level geographically hierarchical housing data structure 17

2.3 Study area and the Central Transit Corridor (CTC) 24

2.4 The New Housing Price Index for Waterloo Region 26

2.5 The sales price trend for KW from 2005 to 2018 before and after price adjustment by the NHPI 27

2.6 Time fixed effects estimates within vs. outside the CTC - condo housing . 32

2.7 Time fixed effects estimates within vs. outside the CTC - single-detached housing 35

2.8 Time fixed effects estimates within vs. outside the CTC - semi-detached housing 39

2.9 Time fixed effects estimates within vs. outside the CTC - townhouses . . . 43

4.1 Survey questions of stated importance of physical neighbourhood and accessibility-related attributes 84

4.2 Latent class model structure 85

4.3 Three groups of the total survey sample 90

4.4 Spatial distributions of the three groups 91

4.5 Car ownership and travel mode choice of the three groups 92

4.6 Moving motivations 94

4.7 Stated preference for TOD features 95

4.8 Stated preferences for structural housing features 95

4.9 Stated preferences for socio-demographic characteristics in neighbourhoods 96

A-1 Empirical semi-variogram fit to transactions of different housing types . . . 126

A-2 Illustration of hedonic demand identification 130

A-3 The percentage of observations in each census tract from the survey sample and the sales dataset, respectively 132

A-4 The absolute difference of the percentage of observations in each census tract between the survey sample and the sales dataset 132

A-5 The spatial distribution of the absolute difference of the observation percentage in each census tract between the survey sample and the sales dataset 133

A-6 Survey sample vs. sales dataset with respect to housing types and housing prices 133

A-7 Stated importance of housing attributes in location choices 136

A-8 Partial correlation matrix of explanatory variables 137

A-9 Density plots of class probabilities for the three groups 148

Chapter 1

Introduction

1.1 Scope of the thesis

Light-rail transit (LRT) investment and transit-oriented development (TOD) have become a focus of urban planning in North American cities. Waterloo Region (the Region) implemented a light-rail transit (LRT) system aiming to provide alternative transit options and shape urban communities. Meanwhile, the Region has seen continuous population growth and booming housing markets in recent years. First, this thesis aims to examine the relationship between housing characteristics, transit-related characteristics and housing prices through a hedonic model, which simultaneously accounts for the spatial and temporal effects on price determination. The model estimates help evaluate how different housing markets (including condos, single-detached houses, semi-detached houses and townhouses) react to the transit development over the years and how the locational amenities contribute to the housing prices of different housing types.

Second, few studies have explored the impacts of LRT investment on individual households' residential location choices. A better understanding of different households' preferences for housing helps explain the residential patterns within a metropolitan area and offers valuable information for policy makers to evaluate and devise housing policies. This thesis aims to examine heterogeneous housing preferences underlying the individual households' residential location choice behaviours. Further more, this thesis attempts to identify the demographic groups with distinct preferences for TOD housing and seeks to guide TOD policies to create home options based on the needs of various households. Updated knowledge on housing demand is expected to help the region build complete and vibrant communities for a range and mix of residents.

1.2 Context

1.2.1 Location and Policy Context

Waterloo Region had a total population of 535,154 in 2016 (Statistics Canada, 2017), making it the fourth largest urban area in the Province of Ontario. The Region is internationally known for its leading-edge technology industries and innovative universities, and is one of the fast growing areas in the province. As an increasingly attractive place to live and work, the Region is projected to reach 742,000 people by 2031 (Growth Plan, 2017). In light of anticipated growth in population and employment, the Region has taken innovative steps in growth management. Back to 2003, the Regional Growth Management Strategy (RGMS, 2003) was approved by the Regional Council, and it identified six goals for managing growth, including "building vibrant urban places" through *reurbanization* and "providing greater transportation choice" through *a rapid transit system.* Ontario's *Places to Grow - Growth Plan for the Greater Golden Horseshoe* (2006) also identified rapid transit as a key catalyst to encourage intensification in existing urban area. Since then, the Region's rapid transit plan went through several milestones, which are summarized below.

- 2010 - The Provincial and Federal governments announced their funding commitments towards rapid transit in Waterloo Region.
- 2011 - Regional Council approved the LRT implementation option with a two-staged approach.
- 2014 - The Phase-One LRT construction began.
- 2018 - The Phase-One LRT construction ended.

- 2019 - The Phase-One LRT started services in June.

To guide urban growth along with the LRT investment, the Region released an updated Regional Official Plan (ROP, 2015) for managing growth in Waterloo Region to 2031. The ROP implements principles set out in the RGMS (2003)) and conforms to the provincial policies and legislations including the Growth Plan (2006) and the land-use planning policies in Provincial Policy Statement (2014). The key elements include directing a greater share of new development and investment towards the existing Built-Up Area (BUA) and improving integration of transit. These policies encourage the Region to build up instead of out, and thus strive to create balanced and sustainable growth. Apart from the general development goal of *reurbanization*, the ROP introduces specific policies to guide Transit Oriented Development in major transit station area, which include promoting medium and higher density development, creating a more compact urban form, providing a mix of land uses that allow people to walk or take transit to various destinations, creating pedestrian-friendly environments, facilitating multi-module transportation, and enhancing social integration. Under these policies, the Waterloo Region is working to create vibrant and complete urban communities.

1.2.2 Housing market introduction

The housing market in KW has experienced dramatic shifts over the years from 2008 to 2018. As shown in Figure 1.1-a, the average residential sales price increased with a relatively stable rate (3-5% annual increase rate) before 2014, while the price sees a sharp rise after that. The housing price has increased by 7.7% from 2015 to 2016, and then reached a peak in 2017 with an over 20% increase from 2016. When looking at Figure 1.1-b, the number of home sales peaked at the second quarter of both 2016 and 2017. The historic low level of listings since 2014 is illustrated as well.

The housing boom occurring in KW has been mainly attributed to the sudden demand increase from the Great Toronto and Hamilton Area (GTHA). The GTHA is Canada's largest urban region, and the housing market is one of the hottest markets in Canada.

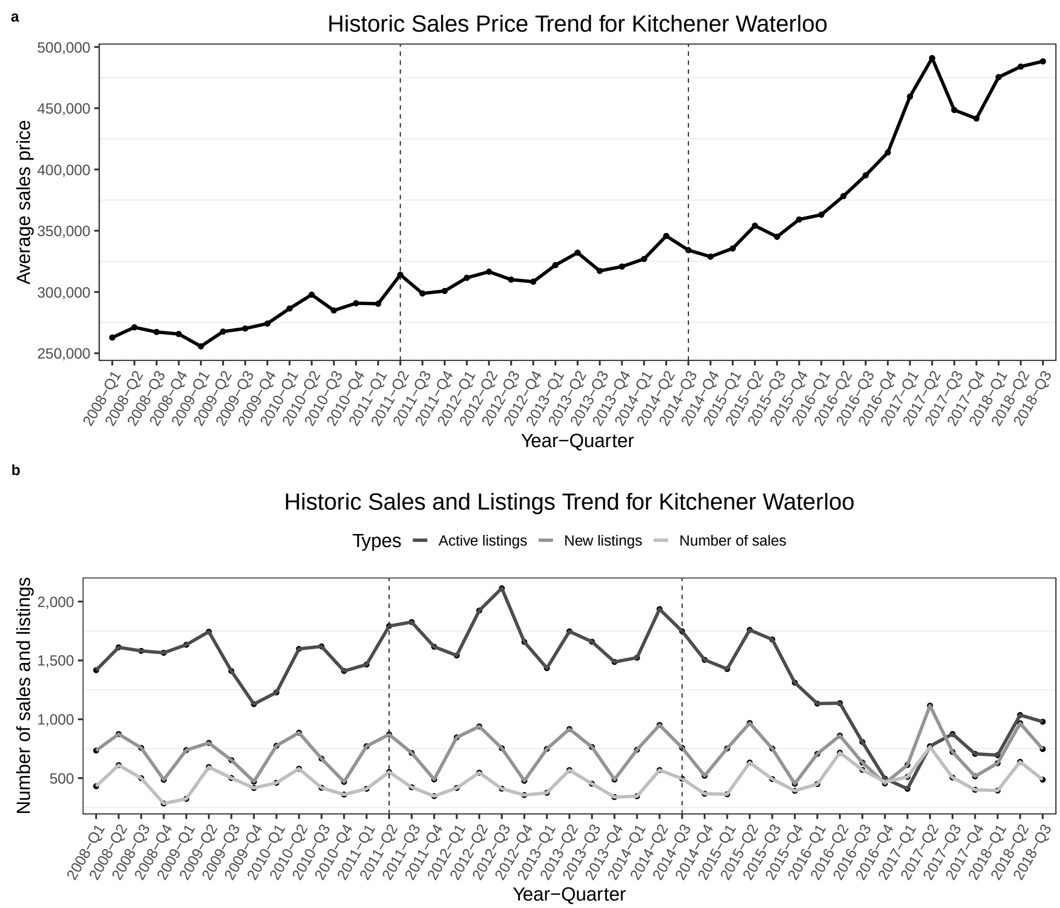


Figure 1.1: Residential housing market trends in KW over the years from 2008 to 2018

Note: (a) shows the average residential sales price of each quarter in KW. (b) shows the total residential sales and listings of each quarter in KW. Note that the LRT line got approved in 2011 and started construction since 2014. Source: Kitchener-Waterloo Association of REALTORS®.

However, escalating prices and the mortgage stress test in 2016[1] have prompted some GTHA buyers to seek homes in KW. The relatively less expensive housing, the fast growing economy and its regional accessibility to the GTHA have made the KW market more attractive to these GTHA buyers. As a result, the unrelenting demand in particular with GTHA buyers migrating to KW as well as the low inventory appear to have contributed to the housing boom in KW from 2016 to 2017.

[1] The federal mortgage stress test rules aimed to ensure that homebuyers can afford their mortgages even if interest rates rise much higher in the future.

1.3 Research objectives

This thesis is highly motivated by the region's growth policy changes, in particular with the new LRT investment and TOD polices. As many other medium-size cities in North America, the region expects the LRT to intensify urban land uses and concentrate more residents into urban cores, and major transit station area. Therefore, it is particularly important to examine how the residential housing markets reacted to the policy changes, and what housing and neighbourhoods people prefer to reside in. This thesis aims to analyze the complex relationship between LRT investment, housing market fluctuations and residential location choices through three empirical analyses. The two main objectives are summarized below.

Objective 1: To build better hedonic pricing models to investigate the relationship between housing characteristics, LRT-related characteristics and housing prices.

To achieve this objective, the first study presents an innovative spatio-temporal multilevel model to simultaneously control for the spatial and temporal relations underlying the housing price determination. This study also includes a range of intensification-related characteristics: bus transit access, open space access, and intersection density. With the preferred model specification, this thesis is able to better understand how different housing markets react to the LRT investment over the years and how the locational amenities contribute to the housing prices of different housing types.

Objective 2: To investigate the residential location choices and preferences of the individual households during the LRT construction stage.

The first study analyzes the housing market prices from the interaction between home buyers and sellers. However, it offers little information on housing demand or residential preferences. With a particular interest in examining the relationship between LRT and residential location choices, disaggregated information about homebuyers, their location choices and attitudes toward the LRT is needed to facilitate this analysis. This study starts with a detailed housing survey during the LRT construction stage. Taking advantage of the survey data, the second study conducts a second-stage demand analysis and recovers the heterogeneous housing preferences of different households groups.

The third study aims to further identify different household groups holding different preferences for TOD. Based on the survey responses regarding the importance of TOD features in residential location choices, this study conducts a latent-class analysis (LCA) to examine who holds a strong desire for TOD communities and who still prefers the car-oriented neighbourhoods. Results from the two studies are expected to inform housing policies with up-to-date knowledge on housing demand and TOD preferences.

1.4 Research questions

The key research questions to be addressed in this thesis are detailed below.

Q1: How do the "recent comparable sales" impact the housing prices? How do different neighbourhoods impact housing prices? What are the main advantages of specifying a spatio-temporal multilevel model for housing prices, compared to the purely spatial hedonic model?

Q2: What are the associations between housing prices and housing characteristics including the structural and neighbourhood attributes? What trends are seen in the time fixed-effects over the years 2005-2018? What is the relationship between the LRT development and housing prices of different housing types?

Q3: Based on the housing survey analysis, do households have heterogeneous preferences for dwelling and locational attributes of housing?

Q4: Are there significant differences in the survey sample in terms of stated preferences for TOD? Do the demographic profiles, housing preferences and home choices of these groups differ significantly?

1.5 Thesis outline

This thesis is organized into five chapters and proceeds as follows. Chapter 2, 3, 4 are presented based on the three manuscripts. Each chapter consists of introduction, literature review, data and estimation method, results and discussion. Chapter 2 proposes a spatio-temporal hedonic model for analyzing the housing price dynamics over the years. Chapter 3 employs a demand analysis to estimate the underlying preference heterogeneity across different households during the LRT construction stage. Chapter 4 further analyzes different households' preferences for TOD neighbourhoods. Chapter 5 summarizes thesis findings and contributions and introduces planning implications and future work.

Chapter 2

A spatio-temporal multilevel housing price model: integrating the spatial and temporal dependence and neighbourhood effects

Each housing transaction has a specific location and occurs at a specific moment. Housing prices are theoretically determined by the location of each property and the time when it is transacted. In recent years, the housing literature has exhibited a growing interest in the specification and estimation of space-time hedonic models for housing prices. This paper presents a spatio-temporal autoregressive multi-level model (STAR+MLM) to simultaneously account for spatial and temporal effects on housing prices. First, we introduce temporal restrictions to spatial interactions and define a spatio-temporal weight matrix to control for both spatial and temporal dependence at the property level. We further consider the nesting structure of housing, where houses are nested within aggregated clusters or neighbourhoods (such as census tracts), and thus control for spatial heterogeneity through a multi-level modelling (MLM) approach. This study uses 68,258 housing transactions between 2005 and 2018 in Kitchener-Waterloo, Canada, including condos, single-detached houses, semi-detached houses and townhouses. Results indicate that the STAR+MLM models produce better model performance and explicitly identify three effects on housing price determination: i) the impact of recent comparable sales; ii) neighbourhood/contextual effects; and iii) neighbourhood dependence.

2.1 Introduction

Housing research has made considerable progress in accounting for spatial effects, including spatial dependence (known as "spillover effects") and spatial heterogeneity (known as neighbourhood/contextual effects), especially along with the development of the spatial econometric techniques (Anselin et al., 2004). In contrast to the enormous spatial hedonic applications, few attention has been put on the time dimension in housing price determination (Füss and Koller, 2016). Theoretically, housing prices are determined by both the location of each property and the time when it is sold. Empirically, the real estate professionals rely on the "recent comparable sales" to determine the sales price of a particular property. Therefore, it is essential to control for the spatio-temporal dependence in housing price determination, not the solely spatial dependence. Following the seminal work of Pace et al. (1998), the spatio-temporal autoregressive (STAR) models have aroused increasing attention in housing price modelling (Dubé and Legros, 2014; Thanos et al., 2016; Liu, 2013; Hyun and Milcheva, 2018). These works demonstrate the need to consider both the spatial and temporal relations underlying the transaction data generating process (DGP).

Most of the STAR literature focuses on the examination of the spatio-temporal relations at the property level, but they fail to further account for the effects derived from the higher neighbourhood level. Houses are naturally nested within aggregated units or neighbourhoods, where the housing prices within the same neighbourhood are expected to be similar in part due to the same neighbourhood effect. The impact of neighbourhoods on housing prices can manifest itself through many channels. First, the housing prices can be affected by the observable neighbourhood characteristics, such as the education rate, population density as well as the public services such as school quality and security. Second, housing prices can also be influenced by the behaviour or interactions of people in the neighbourhood. For instance, desirable social interactions and beneficial social capital of local communities can increase the attractiveness of neighbourhoods (Ioannides and Zabel, 2003) and thus being capitalized into housing prices. However, the social aspects of neighbourhood effects are not easy to be measured. The multilevel modelling (MLM) provides an approach to account for the unobserved effects in the neighbourhoods. Housing research has increasingly applied the multilevel modelling (MLM) approach (Goldstein, 1987) in hedonic models (Glaesener and Caruso, 2015; Law, 2017; Orford, 2002), but few are found in the current STAR literature.

This paper presents the spatio-temporal multilevel model aiming to test three hypotheses: i) the past sales of neighbouring properties determine the sales price of a particular property - spatio-temporal dependence; ii) sales prices of residential properties are in part determined by different neighbourhoods - neighbourhood heterogeneity or neighbourhood/contextual effects; iii) neighbourhoods nearby are similar in price determination - neighbourhood-level dependence. We start with the classic spatial autoregressive (SAR) model and then combine with the MLM technique to control for the neighbourhood effects through building the SAR+MLM model; we then specify a spatio-temporal weight matrix in the STAR model to take the temporal causality into account; and finally we control for the neighbourhood effects in the STAR model and present the STAR+MLM model.

This study uses a large transaction data set from the Municipal Property Assessment Corporation (MPAC) and Teranet in 2005-2018 in Kitchener-Waterloo, Canada to test the model performances of the STAR+MLM models and the impacts on parameter estimates. The region has experienced rapid economic growth in high-tech industries and dramatic housing market dynamics, along with a new light-rail transit (LRT) investment in the region during the study period. To the best of our knowledge, this study is the first to couple the STAR model with the MLM technique to better control for the spatial and temporal effects in housing price determination. The proposed STAR+MLM model is expected to improve both model performance and estimation efficiency. Empirically, through a better control for the spatial and temporal effects, we are able to examine the neighbourhood effects on prices of different housing types and the time fixed effects over the years.

2.2 Spatio-temporal hedonic modelling

2.2.1 Spatial hedonic modelling

Space plays an important role in housing price determinations (Bockstael, 1996). The two key features of spatial effects, spatial dependence and spatial heterogeneity, have been long investigated in housing studies. According to Tobler (1970)'s first law of geography, nearby things are more related than distant things. Given the geographic nature of housing, the price obtained on a house tends to be similar to the prices of neighbouring houses, and such dependency may diminish as the distance between the houses increases (Osland, 2010). This is well known as spatial dependence in housing research. Along with the development of spatial econometric techniques (Anselin, 1988), hedonic studies have widely applied spatial autoregressive (SAR) models (including spatial lag models, spatial error models, and the general spatial models) to account for spatial dependence (Lesage and Pace, 2014; Small and Steimetz, 2012; Gibbons and Overman, 2012; Cohen and Coughlin, 2008; Koschinsky et al., 2012; Trojanek and Gluszak, 2018). These studies confirm that ignoring spatial dependence in hedonic models significantly impact the price effects of various variables and the predictive accuracy of housing prices (Krause et al., 2012).

Spatial heterogeneity generally refers to the spatially varying relationships between housing prices and attributes (Brunsdon et al., 1998), which might be due to the underlying heterogeneity in housing demand and supply across space. Studies commonly deal with the possible heterogeneous market structures with the use of local regression methods, primarily the geographically weighted regression (GWR) models proposed by Fotheringham et al. (1998). The GWR models assume that market structures vary continuously across space and allow for representing continuous variations of the relationships over space (Yu et al., 2007; Crespo and Gret-Regamey, 2013; Fotheringham and Oshan, 2016).

Some studies assume market heterogeneity to be discrete across space and apply the multilevel modelling (MLM) approach (Goldstein, 1987) to account for price variations in different geographic scales/levels. MLM recognizes the hierarchical nature of housing (Goodman and Thibodeau, 1998), where dwellings are generally nested within neighbourhoods, districts or cities, and decomposes the unexplained price variations into different spatial scales (Orford, 1999). It allows identificaiton of the extent to which price variations come from the lower-level differences and from the higher-level environmental/locational differences (Chasco and Gallo, 2012). Thus, MLM has the capacity to capture additional contextual/neighbourhood effects (Jones and Bullen, 1993; Orford, 2002) after controlling for locational attributes (such as accessibility and socioeconomic variables).

MLM has gained increasing attention in hedonic studies. Glaesener and Caruso (2015) found significant region-level variations in the impacts of land-use diversity upon the price of residential land in Luxembourg. Law (2017) applied the multi-level hedonic model to estimate the local area effects on housing prices through a case study in Metropolitan London. She found robust evidence of the street-based local area effect on housing prices, which is much stronger than the administrative region-based local area effect. It should be noted that most MLM models estimate neighbourhood (or intergroup) differentiations but ignore the presence of spatial relations between the neighbourhood groups. The impacts of adjacent neighbourhoods on housing prices are expected to be correlated considering their spatial proximity, and such relationship should not be neglected. Dong et al. (2015) extends the classic MLM by considering simultaneously neighbourhood effects and the spatial interactions between the lower-level observations and between the higher-level districts. He proposed a hierarchical spatial autoregressive model (HSAR), which combines the SAR and MLM modelling techniques to decompose the complex spatial effects into different levels. Cellmer et al. (2019) applied the same approach and compared the results of the HSAR model with the classic MLM model and the SAR model, where they found better fit of data by the HSAR model, significant spatial interactions in both levels and significant contextual effects (i.e., price variations across zones). Since a mixture of spatial effects would be present in housing market, the combination of MLM and SAR methods provides a promising way to account for both spatial heterogeneity and spatial dependence in housing data.

2.2.2 Spatio-temporal hedonic modelling

Among the volume of spatial hedonic applications, few have considered the temporal dimension in their analysis (Dubé and Legros, 2014). Spatial hedonic models generally consider the spatial dimension alone and neglect the fact that housing transactions are not only spatially located but occur at a specific time. Real estate data in housing research often consist of a collection of transactions pooled over time. Thus, housing data is spatio-temporal data by nature. Most importantly, the "arrow of time" should not be ignored in housing price modelling (Thanos et al., 2016). Unlike the multidirectional spatial impacts on housing prices, the temporal impacts are expected to be unidirectional, where only the prior sales of neighbouring properties can impact the housing price of each property (Can and Isaac, 1997). In reality, real estate professionals often determine the sales price of a specific property by referring to the "recent comparable sales", which emphasizes both spatial and temporal impacts in housing price determination. As argued by Hyun and Milcheva (2018), probably due to overly optimistic buyers and their herding behaviour, buyers are easily willing to pay housing prices similar to the nearby properties recently transacted, especially in a boom market. Thus, it is crucial to consider the temporal causality underlying the transaction process, especially when dealing with housing data pooled over time.

Lately, increasing attempts have been put on the spatio-temporal hedonic modelling (Thanos et al., 2016). Can and Isaac (1997) might be the first to consider both the space and time dimensions in their hedonic modelling. Although their focus was still on testing spatial dependence specifications and estimation accuracy, they assumed that only the past 6 months' sales have impacts on the housing price of each house. The seminal work of Pace et al. (1998) first systematically introduced the spatio-temporal effects in hedonic modelling, which explicitly incorporated the spatial matrix (S), the temporal matrix (T), the spatio-temporal matrix ST (the product of S and T) and the temporal-spatial matrix TS (the product of T and S) in the autoregressive components of the models. Specifically, it assumed 300 prior observations to define the temporal influence on the price of each house and restricted the spatial influence to 15 neighbouring observations. Their results presented the strong influence of the sales prices of neighbouring properties recently sold. Liu (2013) applied the same approach as Pace et al. (1998) to control both spatial and temporal dependence, and their results showed better model fit and prediction power than the traditional hedonic model that ignored these effects. Füss and Koller (2016) followed the same approach and conducted robust tests by specifying different spatial and temporal lags in models. They found that changing the parameters (i.e., the number of prior observations and the number of neighbouring observations) from the initial values (180 prior sales and 30 neighbouring sales) does not notably change the prediction results. Despite the superiority of such models in prediction power, there is always a struggle to correctly interpret the economic significance of the separate effects, in particular the space-time effects defined by ST and TS.

Recent studies attempt to define spatio-temporal relations through a general spatio-temporal weight matrix (W), which is often referred to as the spatiotemporal autoregressive (STAR) model in the hedonic literature (Thanos et al., 2016). Instead of decomposing the space-time effects into four matrices (S, T, ST, TS), STAR models generally define the spatiotemporal neighboours by one matrix W calculated by a Hadamard product (Dubé et al., 2013) between the spatial and temporal weight matrices, i.e., $W = S \odot T = [s_{ij}] \times [t_{ij}]$. The advantage of such approach lies in the combination of the spatio-temporal closeness and constraints in a unique matrix, as argued by Dubé and Legros (2014). More intuitively, the coefficient ρ of the spatio-temporal lag term in STAR models can be interpreted as the effect of past neighbouring sales on the current prices.

STAR models often start with assumptions on the spatio-temporal relations by determining the spatial and temporal distance cut-offs and the spatial and temporal decay/frictions. Using a huge sample (127,787) of apartment sales between 1990 and 2001 in Paris, Dubé and Legros (2014) constructed STAR models considering various distance cutoffs from 0.5 km to 3 km and compared results with the SAR models. They found that in all cases the STAR specification outperforms the SAR specification in the out-of-sample prediction, and the solely spatial weights matrix in the SAR models produces higher autoregressive coefficient values than the spatio-temporal weights matrix in the STAR models, indicating the upward estimation bias of the dependence parameter. Dubé et al. (2018)

tested the performance of different spatio-temporal specifications in STAR models and found that past transaction information stops contributing to price determination after eight months in Aberdeen, Scotland. Their results also found the dominance of the unidirectional spatio-temporal connections in price determination and thus confirming the influence of the "comparable sales approach" used by the real-estate professionals as a well internalized process for property valuation.

STAR modelling has captured increasing attention in empirical studies. Smith and Wu (2009) developed a STAR model and identified significant evidence of the spatiotemporal neighbours (60-days prior sales within a distance of 3 km) in price determination. Dubé and Legros (2014) not only considered the unidirectional effect of the past 2-4 months' sales, but the multidirectional effect of the same time sales defined as sales occurring in the same month, one month before and one month after. Thanos et al. (2016) further decomposed the spatio-temporal data generating process (DGP) into three components considering the "arrow" of time: the "comparable sales" effect of the recent neighbouring sales (sales over a month), the "contemporaneous spatial peer" effect of the same-time sales (within a month before the sale), and the "sellers' expectations" effect of the future sales (within a quarter of the sale). Their results also indicated the estimation bias of the SAR model and demonstrated three distinct effects in price determination, while the future expectation effect (0.06) was found much less than the prior sales effect (0.33). Based on 30,541 apartment transaction data in Seoul, South Korea between 2006 and 2015, Hyun and Milcheva (2018) built two STAR models for the boom period and the bust period, and they found that the spatial-temporal dependence in housing prices is eight times higher in a boom than a bust. In addition to the STAR models, studies such as Habib and Knockelman (2008), Osland et al. (2016) and Zolnik (2019) also have made unique contributions to the spatiotemporal hedonic modelling literature.

2.3 Methodology

In this section, we describe the four model specifications in our study. As shown in Figure 2.1, we construct four models from the classic SAR model to the spatial multilevel model (SAR+MLM), followed by the STAR model and the STAR multilevel model (STAR+MLM) to control both the spatio-temporal dependence and neighbourhood effects.

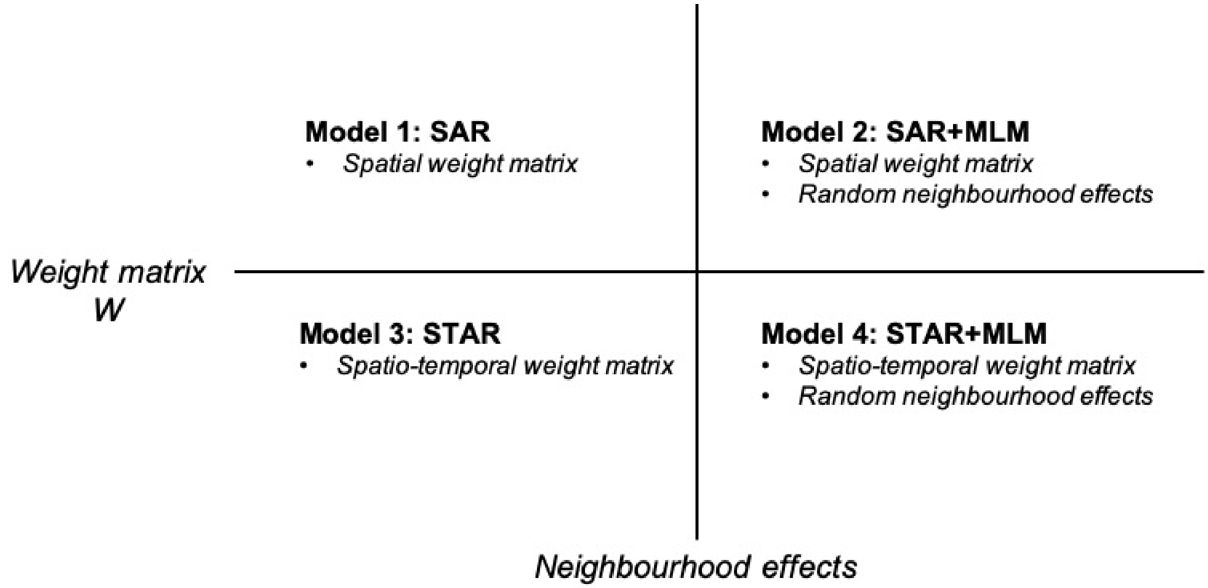


Figure 2.1: Four model specifications by different weight matrices and neighbourhood effect control

2.3.1 Model specifications

SAR model

The spatial autoregressive (SAR) model is the basic model widely applied in spatial econometrics, which explains the spatial dependence through adding a spatial lag term (W_1Y) as in the model below. This model accounts for the spatial interactions at the property level by assuming that each property's sales price tends to be affected by the prices of properties nearby.

$$Y = \rho W_1 Y + \beta X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, \sigma_\varepsilon^2) \tag{2.1}$$

where

Y - vector of the dependent variable

ρ - property-level spatial autoregressive parameter to estimate

W_1 - spatial weight matrix at the property level

β - vector of regression coefficients to estimate

X - matrix of independent variables

ε - vector of an independent, normal distributed error term at the property level

σ_{ε}^2 - property-level variance to estimate

A necessary consideration of building the SAR model is to construct a spatial weight matrix through defining the neighbouring structure and the weight type. For the neighbouring structure, one common way is to set up a distance threshold and define the houses within a certain distance being the neighbours of each house; the other way is to find the k-nearest neighbours (or *knn*) based on the basis of metric distances (Osland et al., 2016). After defining the neighbouring connectivity structure, the spatial weights need to be specified, where the common way is to use the row-standardized weight style, and the other styles include the basic binary scheme and the globally standardized style etc. The specific specification of the matrix is detailed in the STAR model subsection.

SAR+MLM model

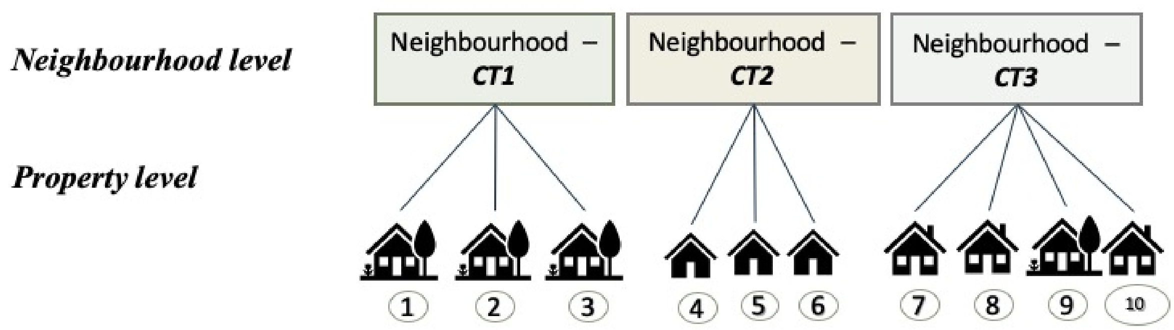


Figure 2.2: A two-level geographically hierarchical housing data structure

This model extends the classic SAR model to a spatial multilevel model, where the nesting housing data structure is considered explicitly. As illustrated in Figure 2.2, houses (Level 1 - property level) are geographically nested within neighbourhoods (Level 2 - neighbourhood level).[1] We use the census tract (CT) as the definition of neighbourhood in our study, and

[1]We do not include a third level, say city level. Babin (2016) using the similar dataset found that city-level controls were insignificant once neighbourhood effects were controlled for.

we expect different impacts of CTs on housing prices even after controlling for the attributes such as neighbourhood sociodemographic attributes (i.e., population density and education rate).

The key motivation of using the multilevel modelling technique in this study is that it clearly identifies the spatial heterogeneity across different neighbourhoods and isolates such neighbourhood/contextual effects from the spatial interactions at the property level. Level 1:

$$Y = \rho W_1 Y + \beta X + \Delta\theta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, \sigma_\varepsilon^2) \tag{2.2}$$

Level 2:

$$\theta = \lambda M \theta + u, \quad u \sim \mathcal{N}(0,\, \sigma_u^2), \quad cov(\varepsilon, \theta) = 0 \tag{2.3}$$

where

Δ - block diagonal design matrix with column vectors of ones for neighbourhoods

θ - vector of the neighbourhood-level random effects that follows a simultaneous autoregressive process

λ - spatial autoregressive parameter indicating strength of dependence at the neighbourhood level

M - spatial weight matrix at the neighbourhood level

u - vector of an independent, normal distributed error term at the neighbourhood level

σ_u^2 - neighbourhood-level variance to estimate

This model is also called the hierarchical spatial autoregressive model (HSAR) by Dong et al. (2015). This particular model specification relaxes the restriction of independence among neighbourhood random effects θ in the standard multilevel modelling literature (Goldstein and Browne, 2002). It assumes θ to be spatially dependent, especially considering that the contextual effect of each neighbourhood may be similar to its adjacent neighbourhoods (Dong et al., 2015). Following this, we define the row-standardized spatial weight matrix M based on the adjacency between each census tract, and assess the extent of spatial interactions at the higher level through parameter λ. The estimated variance σ_u^2 denotes the unexplained variation at the neighbourhood level after we control for the explanatory variables.

The major advantage of this spatial multi-level model lies in its ability to isolate three distinct effects underlying the price determination process: the spatial interactions at the lower property level (ρ); the spatial interactions at the higher neighbourhood level (λ); and spatial heterogeneity across neighbourhoods, i.e., neighbourhood/contextual effects (σ_u^2).

STAR model

We propose this model with a special consideration of the spatio-temporal process underlying the price determination. To be specific, although the "true" data generating process is unknown, both housing theories and empirical studies suggest that the nearby houses sold recently have a large influence on the sales price of a specific house. Therefore, we follow the STAR literature and build a spatio-temporal weight matrix to simultaneously control for the spatial and temporal dependence at the property level. The STAR model is shown below.

$$Y = \rho W_2 Y + \beta X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2) \tag{2.4}$$

In this model, W_2 represents the space-time weight matrix at the property level. To be specific, we first build a spatial weight matrix S and a temporal weight matrix T, and then construct W_2 by calculating a Hadamard product of the two matrices.

To build the spatial weight matrix, the spatial interaction between observations i and j, $s_{i,j}$, is defined by the equation below,

$$s_{i,j} = \begin{cases} exp(-d_{i,j}^2/2d^2), & \text{if } \ i \neq j \ and \ d_{i,j} \leq d \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

where $d_{i,j}$ is the Euclidian distance between properties i and j, and the threshold distance d is set to be 2.5 km. The empirical variograms and the discussion on how we determine the spatial extent for the weights are attached in Appendix A-1. Based on the explorations, we assume that only the properties within 2500 meters of a particular property influence its sales price, and the spatial interaction effect decays exponentially with the distance increase. The spatial weight matrix S is then constructed in a row-standardized way and becomes

$$S = \begin{bmatrix} 0 & s_{1,2} & \cdots & s_{1,n} \\ s_{2,1} & 0 & \cdots & s_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n,1} & s_{n,2} & \cdots & 0 \end{bmatrix}$$

Similarly, we assume the temporal interaction between observations i and j, $\tau_{i,j}$, as defined by the equation below.

$$\tau_{i,j} = \begin{cases} 1/((yymm_i - yymm_j) + 1), & \text{if } yymmdd_i > yymmdd_j \text{ and } 0 \leq (yymm_i - yymm_j) \leq t \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

where $yymmdd_i$ and $yymmdd_j$ are the sales date of property i and property j, respectively. $yymm_i$ denotes the sales year-month of property i, and $yymm_j$ denotes the sales year-month of property j. t represents the temporal interaction threshold. In Kitchener-Waterloo, the realtors tend to refer to the recent 3 months' sales for determining the listing price, and this was concurred by the MPAC experts. Therefore, we constrain the temporal influence up to the past 3 months ($t = 3$), and thus only the past three month's sales j can affect the sales price of property i. The temporal weight matrix T becomes

$$T = \begin{bmatrix} 0 & \tau_{1,2} & \cdots & \tau_{1,n} \\ \tau_{2,1} & 0 & \cdots & \tau_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{n,1} & \tau_{n,2} & \cdots & 0 \end{bmatrix}$$

If property i is sold 2 months after property j, then only the sales price of property j can influence the sales price of property i, not vice versa. Then the temporal weight $\tau_{i,j}$ is equal to 1/3 while $\tau_{j,i} = 0$. Therefore, this matrix captures the unidirectionality of the temporal influence between properties. This restriction on the temporal dimension is the dramatic difference between the STAR model and the classic SAR model where the "arrow of time" is not considered and the influence of property-level interactions is multi-directional.

After defining both the spatial and temporal weight matrices, we construct the spatio-temporal weight matrix W_2 through a Hadamard product as applied in various studies (Smith and Wu, 2009; Thanos et al., 2016).

$$W_2 = S \odot T, \quad w_{i,j} = s_{i,j} \times \tau_{i,j} \tag{2.7}$$

We also use the row-standardized style for the weight matrix W_2 so that the absolute value of the lagged term coefficient ρ ranges from 0 to 1. Considering the large sample and large weight matrices (e.g., for single-detached housing, the spatio-temporal weight matrix is 41,274×41,274 with many zeros defining no relations), we construct sparse weight matrices and run the models through the high-performance server in Compute Canada. Through controlling for the effects of both spatial and temporal distance decay, the STAR model is expected to better represent the economic process of the housing market and more accurately estimate the model parameters. Although examining the influence of different weight matrices on estimation is not a focus of this paper, we acknowledge that sensitivity analysis using different spatial and temporal distance thresholds and different distance decay functions is an area for future work.

STAR+MLM model

This model further extends the STAR model to the hierarchical spatio-temporal model by adding the nesting structure of housing data. We propose this model to test whether the neighbourhood (clustering) effects exist and whether there exists spatial dependence at the neighbourhood level. As discussed in the SAR+MLM model, we simply change the spatial weight matrix to a spatio-temporal weight matrix as in the below equations:

Level 1:

$$Y = \rho W_2 Y + \beta X + \Delta\theta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,\, \sigma_\varepsilon^2) \tag{2.8}$$

Level 2:

$$\theta = \lambda M \theta + u, \quad u \sim \mathcal{N}(0,\, \sigma_u^2), \quad cov(\varepsilon, \theta) = 0 \tag{2.9}$$

2.3.2 Estimation method

We follow the estimation method proposed by Dong et al. (2015) for our model estimations. To be specific, we apply the Bayesian Markov Chain Monte Carlo (MCMC) method, which draws samples sequentially from the conditional posterior distributions for each unknown parameter. To implement the method, we need to first specify prior distributions for all the parameters and then derive their conditional posterior distributions. As shown in the basic Bayesian paradigm,

$$P(\Theta^*|Data) \propto P(Data|\Theta^*) \times P(\Theta^*) \tag{2.10}$$

where the posterior distribution of parameters $\Theta^* = \{\rho, \lambda, \beta, \theta, \sigma_\varepsilon^2, \sigma_u^2\}$ is proportional to the product of the data likelihood $P(Data|\Theta^*)$ and prior distributions $P(\Theta^*)$. To be specific, the posterior distribution for $\Theta^* = \{\rho, \lambda, \beta, \theta, \sigma_\varepsilon^2, \sigma_u^2\}$ is

$$P(\rho, \lambda, \beta, \theta, \sigma_\varepsilon^2, \sigma_u^2|Y) \propto L(Y|\rho, \lambda, \beta, \theta, \sigma_\varepsilon^2, \sigma_u^2) \times P(\rho) \times P(\lambda) \times P(\beta) \times P(\theta) \times P(\sigma_\varepsilon^2) \times P(\sigma_u^2) \tag{2.11}$$

Let the posterior distribution for β be $P(\beta|Y, \rho, \lambda, \theta, \sigma_\varepsilon^2, \sigma_u^2) \sim N(M_\beta, \sum_\beta)$, we are able to derive the posterior distribution for β based on the equation below

$$P(\beta|Y, \rho, \lambda, \theta, \sigma_\varepsilon^2, \sigma_u^2) \propto L(Y|\rho, \lambda, \beta, \theta, \sigma_\varepsilon^2, \sigma_u^2) \times P(\beta) \tag{2.12}$$

The estimation process for the other parameters follows the same approaches as in Dong et al. (2015). The inferences for each model are based on three MCMC chains, and each chain includes 10,000 iterations with a burn-in period of 5,000 to ensure the model convergence. We run our models in R and mainly employ the HSAR package created by Dong et al. (2015).

2.4 Data

2.4.1 Market description

This study examines the housing market dynamics through the spatio-temporal modelling in Kitchener-Waterloo (KW), a medium-size region in Southern Ontario, Canada (Figure 2.3). The region has experienced rapid economic growth in high-tech industries. To accommodate the potential employment and population growth, the regional government proposed a light-rail transit (LRT) line aiming to move people efficiently and revitalize the urban cores through concentrating developments around station areas. The LRT was approved in 2011 and started construction in 2014. Along the transit corridor, an array of high-rise condos and other mixed-use developments have emerged. Not surprisingly, the new LRT investment coupled with the booming high-tech industry, and the international immigrants and migrants from the Greater Toronto Area have all contributed to the increasing attractiveness of the region and the changing demographics, which in turn has heavily shifted the housing market in the region. We have seen a 20.7% increase in the average sales price from 2016 to 2017, compared to an average 3-5% before 2016 (KWAR, 2018). It has been common to see a high buyer-seller ratio and short time-on-market since 2015, and frequent bidding wars occur in the region.

2.4.2 Data preparation

The housing transaction data was provided by the Municipality Property Assessment Company (MPAC) and the Teranet company through a license agreement with the research group.[2] The data contains every residential transaction price between January 2005 and March 2018, along with major housing structural attributes, such as home area, lot size, garage and bedrooms. The original dataset consists of 70,439 transactions. We followed the same data cleaning strategy as in Babin (2016) and removed the non-market rate sales from the transaction dataset through identifying outliers and unexpected observations. After data cleaning, the final dataset used in our analyses becomes a total of 68,258 transactions,

[2]See the detailed data source in Table A-1 in Appendix

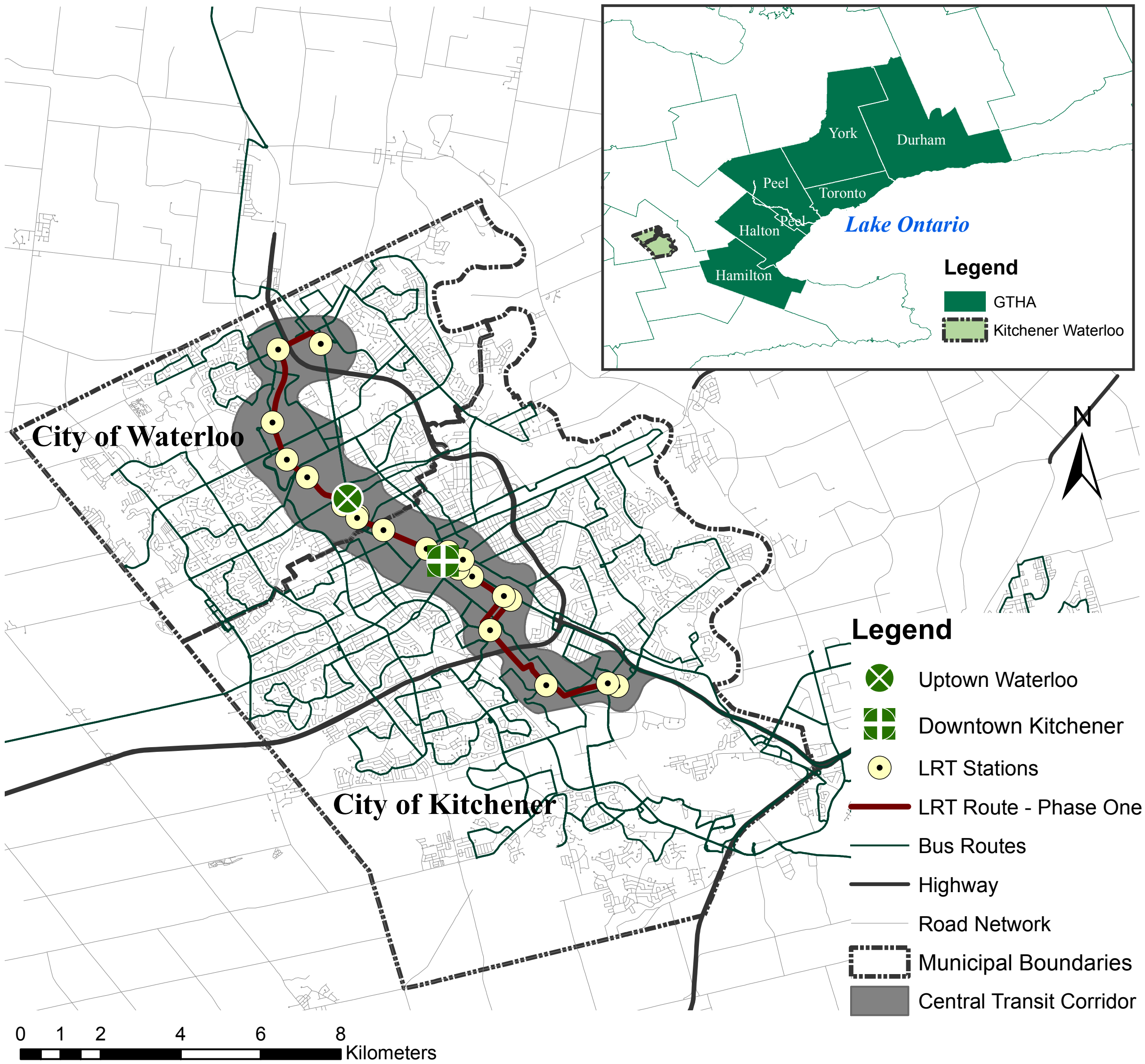


Figure 2.3: Study area and the Central Transit Corridor (CTC)

including 15,364 condominium housing transactions (22.5%), 41,272 single-detached housing transactions (60.5%), 7076 semi-detached/duplex housing transactions (10.4%), and 4546 townhouse transactions (6.7%).

As presented in Table 2.1, the dependent variable in our analyses is the logarithm of the adjusted sales prices. Thus, $(e^{\hat{\beta}} - 1) \times 100$ represents the per cent change of price with one unit increase in each housing attribute. The adjusted sales price was calculated based

Table 2.1: Description of variables

Variable	Description
Dependent variable	
logprice	logarithm of the adjusted sales price [dollars]
Independent variable -	structural attributes
age	age of each house at the sale time [year]
tot_area	total area of each house [1000 sqft]
lot_size	lot size of each house [acre]
baths	number of bathrooms
beds	number of bedrooms
garage	number of garages
fireplace	number of fireplaces
pool	pool - dummy variable [1/0]
Independent variable -	neighbourhood and locational attributes
inter_dense	intersection density [number of intersections within 800 metres]
dis_bus	distance to the nearest bus stop [100 meters]
rd_adj	regional road adjacency - dummy variable [1/0]
os_adj	open space adjacency - dummy variable[1/0]
os_area	total area of open space within 800 meters' access [km^2]
in_ctc	within or without the central-transit corridor [1/0]
edu_rate	post-secondary education percentage in each census tract
pop_dense	population density in each census tract [thousand/km^2]
inter_dense:os_area	interaction term
inter_dense:dis_bus	interaction term
in_ctc:inter_dense	interaction term
in_ctc:os_area	interaction term
Independent variable -	fixed time covariates
sale_year	the sale year - dummy variables
sale_year:in_ctc	interaction term

on the following equation,

$$logprice_i = log\left(\frac{SalesPrice_{it}}{NHPI_t} \times 103.6\right) \tag{2.13}$$

where the sales prices were adjusted to March 2018 dollars using the regional New Housing Price Index (NHPI) from Statistics Canada (2018), and the index value in March 2018 is 103.6. The New Housing Price Index (NHPI) is a monthly measure of new house price changes over time, which is calculated based on the new home builders survey in metropolitan areas across Canada. Figure 2.4 shows the NHPI trend for the Kitchener - Cambridge - Waterloo metropolitan area since 2005. The index follows an almost linear trend. Price adjustment with the NHPI values ensures that our transaction data (including resale market) obtained over years can be reasonably compared after controlling for the regional-scale aggregate price trend.

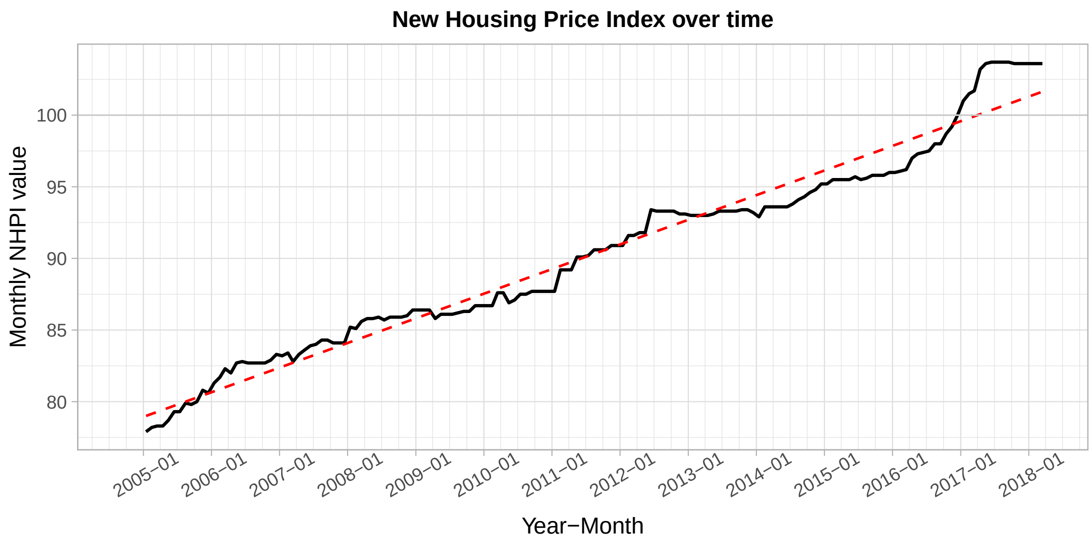


Figure 2.4: The New Housing Price Index for Waterloo Region

Figure 2.5 illustrates the price trends of different housing types before and after adjusting the sales prices, respectively. The curves after price adjustment become flattened to some extent, since the sales prices in the past years have been adjusted to the value of 2018. However, the trend differences across different housing types and the abrupt price surge from 2016 to 2017 are still noticeable even after the price adjustment using the NHPI values. This suggests that the price index alone is not able to control for the price trend variations over time and for different housing types. Due to the influence of economic growth, population growth, and regulation changes over the years, etc., we need to account for the additional temporal heterogeneity in housing prices. To this end, we add the year dummies in the model of each housing type. With 2005 being the reference year, the coefficient of each year dummy variable can be interpreted as the average price difference compared to 2005 after controlling for the observed housing attributes.

This study also controls for the major locational and neighbourhood attributes in our analyses. In particular, we calculated the intersection density in ArcGIS and use it as a

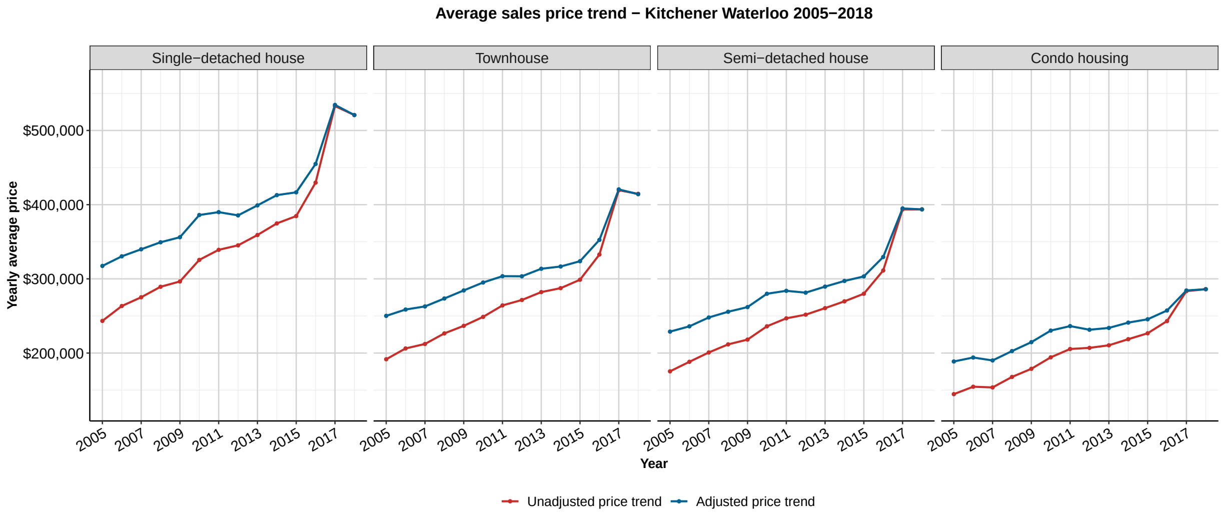


Figure 2.5: The sales price trend for KW from 2005 to 2018 before and after price adjustment by the NHPI

proxy for walkability or street connectivity. For the transit accessibility, since the LRT was still under construction during the study period, we only calculated the distance of each property to the nearest bus stop as an indicator of transit access. Open space amenities nearby including parks, golf course, forests and natural areas are expected to play an important role in price determination. We calculated two related variables to test such impacts following Babin (2016). One is the open space adjacency, which defines whether the property parcel is adjacent to public open space; and the other is the total open space area that can be accessed within 800 meters, which can represent the open space access by walking. In addition, we include a dummy variable *in_ctc* to define whether the property is within the central transit corridor (CTC) with an expectation of a significant price difference between housing within the CTC area and housing outside the CTC area. We also create an interaction term between the sales year dummies and the CTC dummy to capture the sales price difference within and without the CTC in each particular year, after controlling for the observed housing attributes.

Some literature supports the notion that people are willing to pay premiums for a combination of features associated with "compact" development, such as transit accessibility, street design/walkability, open space access and mixed land uses (Krause et al., 2012). Therefore, we include several interaction terms to test the potential synergies. For instance, the interaction term between intersection density and open space area is expected to present a synergic effect on housing prices when people are willing to pay for the improved walking access to the nearby open space; the synergy between the CTC and open space area is also expected considering that people would be willing to pay more for the CTC housing where they have better access to public open space; the interaction term between the CTC and intersection density is expected to be significant as people would be willing to pay more for the housing near transit which also has a better street design for walking. Finally, we include the population density and education rate at the census tract level to control for the socioeconomic qualities of neighbourhoods within the region.

The descriptive statistics of the variables for different housing types are attached in Table A-2 in Appendix. Note that there are 21.7% of condos sold within the CTC area ($n = 3334$) between 2005 and 2018, but only 11.1% of single-detached houses ($n = 4581$), 15% of semi-detached houses ($n = 1061$), and 2.1% of townhouses ($n = 95$) were sold within the CTC during this period.[3]

2.5 Results

2.5.1 Condo housing models

Impact on model performance

From the summary of model statistics for condos in Table 2.2, we see that ignoring the temporal causality in both the SAR model and the SAR+MLM model overestimates the property-level dependence, indicated by the parameter $\hat{\rho}$, which is 0.310 and 0.382, respectively. This is primarily due to the multidirectional spatial relations specified the the SAR models, where not only the past sales but the concurrent sales and the future sales can all influence the sales prices. When only considering the prior sales' influence on properties in the STAR model and the STAR+MLM model, the property-level dependence

[3] Since there are only 95 units of townhouses sold in the CTC from 2005 to 2018 and many of the years include less than 10 observations, we did not include the interaction term between the CTC and time dummies to test the interaction effects from a statistical perspective.

becomes much less, where $\hat{\rho}$ is 0.054 and 0.033, respectively. We also find significant spatial dependence at the neighbourhood level in the STAR+MLM model, indicated by the parameter $\hat{\lambda}$ (0.531). In addition, the higher-level variance ($\hat{\sigma_u^2} = 0.0153$) estimated from the STAR+MLM model confirms the significant contextual effects, i.e., the differences in housing prices across different neighbourhoods. The estimated neighbourhood-level random effects $\hat{\theta}$ from the STAR+MLM model and the SAR+MLM model also show significant neighbourhood correlations in condo prices and identify higher condo prices concentrated at the two urban centres.[4]

Table 2.2: Summary of model statistics - condominium housing

	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\sigma_\varepsilon^2}$	$\hat{\sigma_u^2}$	R^2	DIC
SAR	0.310	-	0.0272	-	0.785	3435314
SAR + MLM	0.382	0.431	0.0214	0.0140	0.821	4395871
STAR	0.054	-	0.0277	-	0.774	714246
STAR + MLM	0.033	0.531	0.0214	0.0153	0.825	950290

DIC: Deviance Information Criterion[5]

The STAR+MLM model explicitly separates three effects: the spatio-temporal dependence at the property level, the spatial dependence at the neighbourhood level, and the spatial heterogeneity across neighbourhoods. The preferred model also presents a better model fit in terms of a much lower DIC value when compared to the SAR models, and it can explain about 82.5% of the total variance in the data. The preferred STAR+MLM model yields a coefficient of 0.033 for the property-level spatio-temporal dependence, suggesting that a $10,000 increase in the average sales prices of neighbouring condo units which are sold within 3 months and are within 2.5 km from a given condo unit will lead to an increase of $330 for the particular condo price. The coefficient of 0.531 for the neighbourhood-level dependence, suggesting that a $10,000 increase in the average sales price of the adjacent neighbourhoods will lead to an increase of $5,310 for the particular condo's neighbourhood price.

[4] In light of the confidentiality terms in our data license agreement, we did not show the maps of neighbourhood random effects in this paper.

Impact on the coefficient estimates

The complete coefficient estimates from the four models are presented in Table 2.3. We do not see much difference in terms of the coefficient estimates for the structural attributes across different models. The total area, the number of baths and garages have significantly positive influence on condo housing prices, while the age of the building and the number of storeys impact the housing price negatively. For different housing types, the mid/high-rise apartments within the CTC show significant positive influence on sales prices. Based on the preferred STAR+MLM model, the sales price of the mid/high-rise apartments within the CTC are $(e^{0.157} - 1) \times 100 \approx 17\%$ higher than the condominium houses within the CTC, and $(e^{0.087} - 1) \times 100 \approx 9\%$ higher than the condo walk-ups (low-rise apartments without elevators) within the CTC. Apart from those relatively consistent estimates, we are more interested to examine which variables' coefficients change significantly across the models. Given that we primarily change the spatio-temporal relations in the models, the spatial and temporal variables are our focus for comparison analyses.

First of all, without controlling for the temporal correlation, models seem to overestimate the time fixed effects (i.e., the coefficients of time dummies shown in Table 2.3). For instance, the coefficient for the year 2013 is estimated to be 0.147 from the STAR model (i.e., condo prices in 2013 are $(e^{0.147} - 1) \times 100 \approx 15.8\%$ higher compared to 2005 after controlling the housing attributes in the STAR model), while it is 0.171 from the SAR model (i.e., condo prices in 2013 are $(e^{0.171} - 1) \times 100 \approx 18.6\%$ higher compared to 2005 after controlling the housing attributes in the SAR model). Similarly, the coefficient is estimated to be 0.166 from the STAR+MLM model, while it is 0.175 from the SAR+MLM model.

Based on the preferred STAR+MLM model estimates, Figure 2.6 plots out the coefficients of the time fixed effects for condos in the CTC and outside the CTC.[6] The figure presents the additional sales price changes over the years after controlling for the major housing attributes for condos within the CTC and outside the CTC, respectively. The two

[6]To obtain the time fixed effects estimates for condos in the CTC, we add (1) the coefficient estimate for each year dummy variable with (2) the coefficient estimate for the interaction term between each year dummy and the CTC dummy.

Table 2.3: Estimates of coefficients from the four models - condo housing

	SAR model		SAR + MLM model		STAR model		STAR + MLM model	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
(Intercept)	7.535***	0.178	6.842***	0.446	10.522***	0.054	11.108***	0.080
age	-0.008***	0.000	-0.009***	0.000	-0.008***	0.000	-0.009***	0.000
tot_area	0.532***	0.007	0.514***	0.006	0.541***	0.007	0.511***	0.007
baths	0.062***	0.004	0.061***	0.003	0.062***	0.004	0.062***	0.003
garage	0.023***	0.003	0.037***	0.003	0.02***	0.003	0.036***	0.003
storey	-0.089***	0.004	-0.099***	0.004	-0.093***	0.004	-0.098***	0.004
fireplace	0.049***	0.004	0.054***	0.004	0.053***	0.004	0.056***	0.004
os_adj	0.021***	0.003	0.017***	0.003	0.024***	0.003	0.017***	0.003
os_area	0.01**	0.004	0.004	0.005	0.003	0.004	0.007	0.005
rd_adj	-0.017***	0.003	-0.037***	0.003	-0.011***	0.003	-0.032***	0.003
in_ctc	0.091***	0.02	0.19***	0.026	0.117***	0.021	0.172***	0.026
dis_bus	-0.005*	0.002	-0.016***	0.003	-0.012***	0.002	-0.017***	0.003
edu_rate	0.005***	0.000	0.002***	0.001	0.0083***	0.000	0.003***	0.001
pop_dense	0.005**	0.002	0.005	0.009	0.0096***	0.002	0.01	0.009
inter_dense	0.001***	0.000	0.001**	0.000	0.001*	0.000	0.001*	0.000
condo walkup	-0.008	0.006	0.023***	0.007	-0.016*	0.007	0.023***	0.007
condo houses	0.123***	0.005	0.164***	0.006	0.121***	0.005	0.163***	0.006
sale_year2006	0.021**	0.008	0.026***	0.007	0.017*	0.008	0.023**	0.007
sale_year2007	0.036***	0.007	0.027***	0.007	0.032***	0.008	0.025***	0.007
sale_year2008	0.093***	0.008	0.092***	0.007	0.09***	0.008	0.088***	0.007
sale_year2009	0.13***	0.008	0.126***	0.007	0.122***	0.009	0.12***	0.008
sale_year2010	0.176***	0.008	0.173***	0.007	0.167***	0.008	0.166***	0.007
sale_year2011	0.176***	0.008	0.17***	0.007	0.163***	0.008	0.161***	0.007
sale_year2012	0.156***	0.008	0.161***	0.008	0.133***	0.008	0.154***	0.008
sale_year2013	0.171***	0.008	0.175***	0.008	0.147***	0.008	0.166***	0.008
sale_year2014	0.185***	0.008	0.2***	0.008	0.162***	0.008	0.191***	0.008
sale_year2015	0.181***	0.008	0.202***	0.008	0.158***	0.008	0.193***	0.008
sale_year2016	0.229***	0.008	0.253***	0.007	0.204***	0.008	0.242***	0.008
sale_year2017	0.364***	0.008	0.384***	0.007	0.338***	0.008	0.372***	0.008
sale_year2018	0.409***	0.013	0.434***	0.012	0.381***	0.014	0.421***	0.012
dis_bus:inter_dense	0.0001	0.000	0.00023**	0.000	0.0004***	0.000	0.00**	0.000
in_ctc:inter_dense	0.001*	0.000	-0.004***	0.000	0.0004	0.000	-0.003***	0.000
os_area:in_ctc	0.086***	0.013	0.082***	0.016	0.115***	0.013	0.105***	0.015
in_ctc:condo walkup	-0.04***	0.011	-0.071***	0.012	-0.065***	0.012	-0.087***	0.012
in_ctc:condo houses	-0.117***	0.01	-0.145***	0.011	-0.145***	0.01	-0.157***	0.011
in_ctc:sale_year2006	-0.009	0.018	-0.021	0.016	-0.005	0.019	-0.017	0.016
in_ctc:sale_year2007	-0.021	0.017	-0.0195	0.015	-0.018	0.018	-0.018	0.015
in_ctc:sale_year2008	-0.03	0.019	-0.0303*	0.017	-0.027	0.019	-0.03*	0.017
in_ctc:sale_year2009	-0.015	0.018	-0.014	0.017	-0.006	0.019	-0.012	0.016
in_ctc:sale_year2010	-0.053*	0.017	-0.047**	0.016	-0.042*	0.018	-0.042**	0.015
in_ctc:sale_year2011	0.016	0.017	0.016	0.016	0.024	0.018	0.018	0.016
in_ctc:sale_year2012	-0.01	0.018	-0.014	0.016	-0.011	0.018	-0.012	0.016
in_ctc:sale_year2013	0.007	0.018	0.016	0.016	0.003	0.018	0.017	0.016
in_ctc:sale_year2014	0.028	0.018	0.029*	0.016	0.026	0.018	0.032*	0.016
in_ctc:sale_year2015	0.052**	0.017	0.029*	0.015	0.047**	0.017	0.028*	0.016
in_ctc:sale_year2016	-0.005	0.016	-0.022	0.015	-0.01	0.017	-0.022	0.015
in_ctc:sale_year2017	-0.039*	0.016	-0.067***	0.015	-0.045**	0.016	-0.068***	0.015
in_ctc:sale_year2018	-0.025	0.026	-0.077**	0.024	-0.018	0.026	-0.08***	0.024
Observations	15364		15364		15364		15364	

Note: *p<0.1; **p<0.01;***p<0.001; S.E. standard error

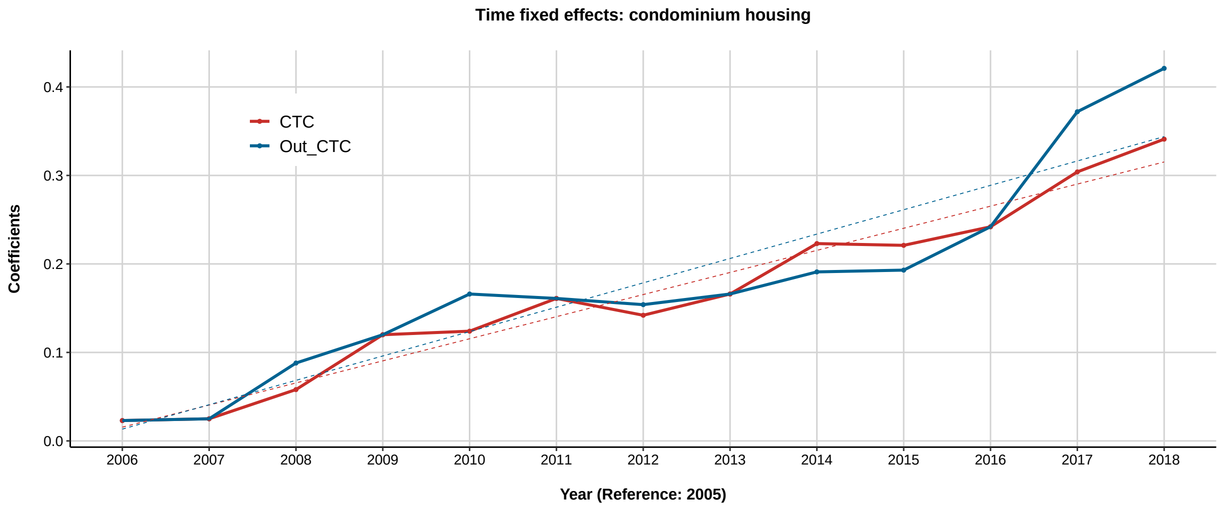


Figure 2.6: Time fixed effects estimates within vs. outside the CTC - condo housing

curves show similar patterns, except for a noticeable price appreciation for condos within the CTC in 2014 when the LRT started construction and a steep price rise for condos outside the CTC from 2016 to 2017.

Table 2.4 presents the transformed parameter estimates for the main spatial and neighbourhood variables across the four models. As expected, all the models observe the positive impacts of the CTC area, being adjacent to open space amenities, better bus transit access and better street connectivity on condo housing prices. When comparing the magnitudes, we see that ignoring the contextual effects seems to have underestimated the added value of the condo housing in the CTC. The sales price of the CTC condo housing is 18.8% higher than the housing outside the CTC based on the STAR+MLM model, while it is only 9.5% from the SAR model and 12.4% from the STAR model, after controlling for the other housing attributes. Open space area alone does not show a significant impact on the condo sales price; however, the interplay between the open space and the CTC presents a synergy effect, where more open space amenities within the CTC significantly increase the condo housing price. The magnitude of the added value of open space in the CTC is higher in the STAR models when controlling for the temporal relations, which is 12.2% and 11.1%, respectively, compared to the SAR models (9.0% and 8.5%, respectively).

For the bus transit access, the condo sales price decreases significantly as the distance to bus stops gets further, indicating that people who buy condos are willing to pay for

Table 2.4: Estimates for the variables of interest from the four models - condo housing

Variables of interest	SAR	SAR+MLM	STAR	STAR+MLM
in_ctc (dummy)	9.5%***	20.9%***	12.4%***	18.8%***
os_adj (dummy)	2.1%***	1.7%***	2.4%***	1.7%***
os_area (km^2)	1.0%**	0.4%	0.3%	0.7%
os_area:in_ctc (km^2)	9.0%***	8.5%***	12.2%***	11.1%***
inter_dense	0.1%***	0.1%**	0.1%*	0.1%*
dis_bus (100 meters)	-0.5%*	-1.6%***	-1.2%***	-1.7%***
in_ctc:inter_dense	0.1%*	-0.4%***	0.0%	-0.3%***
edu_rate (%)	0.5%***	0.2%***	0.8%***	0.3%***
pop_dense ($1000/km^2$)	0.5%**	0.5%	1.0%***	1.0%

Note: *p<0.1; **p<0.01;***p<0.001. The estimates are transformed by $(e^{\hat{\beta}} - 1) \times 100$

better bus transit access. The magnitude of the bus access impact seems to be underestimated in models without controlling for the neighbourhood effects, compared to the preferred STAR+MLM model. The interaction between the CTC and the intersection density shows inconsistent impacts across the models, while our preferred model shows that people are not willing to pay more for the higher intersection density within the CTC especially nearby their condos, suggesting a possibly negative externality effect. With respect to the sociodemographic variables, we see that controlling for the neighbourhood effects significantly decreases both impacts on sales prices, and the population density becomes insignificant in our preferred model. This is possibly due to the correlation between the two neighbourhood variables and the random neighbourhood effects, since the two variables are also defined at the census-tract level.

2.5.2 Single-detached housing models

Impact on the model performances

Table 2.5 summarizes the main model statistics for the single-detached housing. It presents generally similar results as the condo housing, where ignoring the temporal causality in the SAR models overestimates the dependence between properties ($\hat{\rho} = 0.195$ and 0.105, respectively) compared to the STAR models ($\hat{\rho} = 0.040$ and 0.017, respectively). For the higher-level spatial dependence, both the SAR+MLM model and the STAR+MLM model show significant spatial interactions between neighbourhoods ($\hat{\lambda} = 0.699$ and 0.722, respectively). The estimated higher-level variance $\hat{\sigma_u^2}$ (= 0.0036) from the STAR+MLM model for the single-detached housing is much smaller than the lower-level variance $\hat{\sigma_\varepsilon^2}$ (= 0.0154), suggesting that the unexplained housing price variations among the single-detached houses are more attributed to the unobserved property attributes than the unobserved neighbourhood attributes. In addition, the STAR+MLM model presents a better model fit in terms of a much lower DIC value than the SAR models, and it can explain about 85.7% of variation in the data.

Table 2.5: Summary of model statistics - single-detached housing

	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\sigma_\varepsilon^2}$	$\hat{\sigma_u^2}$	R^2	DIC
SAR	0.195	-	0.0177	-	0.838	7831335
SAR + MLM	0.105	0.699	0.0154	0.0034	0.857	9032235
STAR	0.040	-	0.0178	-	0.834	2824525
STAR + MLM	0.017	0.722	0.0154	0.0036	0.857	3341436

DIC: Deviance Information Criterion

The preferred STAR+MLM model yields a coefficient of 0.017 for the spatio-temporal dependence, suggesting that a $10,000 increase in the average housing prices of neighbouring houses which are sold within 3 months and are within 2.5 km from a given house leads to an increase of $170 for the given house price. The coefficient of 0.722 for the neighbourhood-level dependence, suggesting that a $10,000 increase in the average sales price of the adjacent neighbourhoods will lead to an increase of $7,220 for the particular neighbourhood price.

Impact on the coefficient estimates

Table 2.7 presents the coefficients estimated from the four models. As expected, single-detached housing prices tend to significantly increase with the total floor area, lot size, number of bathrooms, garages, fireplace and pool. We do not see much difference in the coefficient estimates for those structural attributes across different models, except for the lot size. The coefficient of the lot size seems to be overestimated in the SAR model (0.515) and the STAR model (0.537) when the neighbourhood effects are not considered, compared to the SAR+MLM model (0.463) and the STAR+MLM (0.468) model. For comparison purpose, we further examine how different models influence the estimates of the spatial and temporal variables.

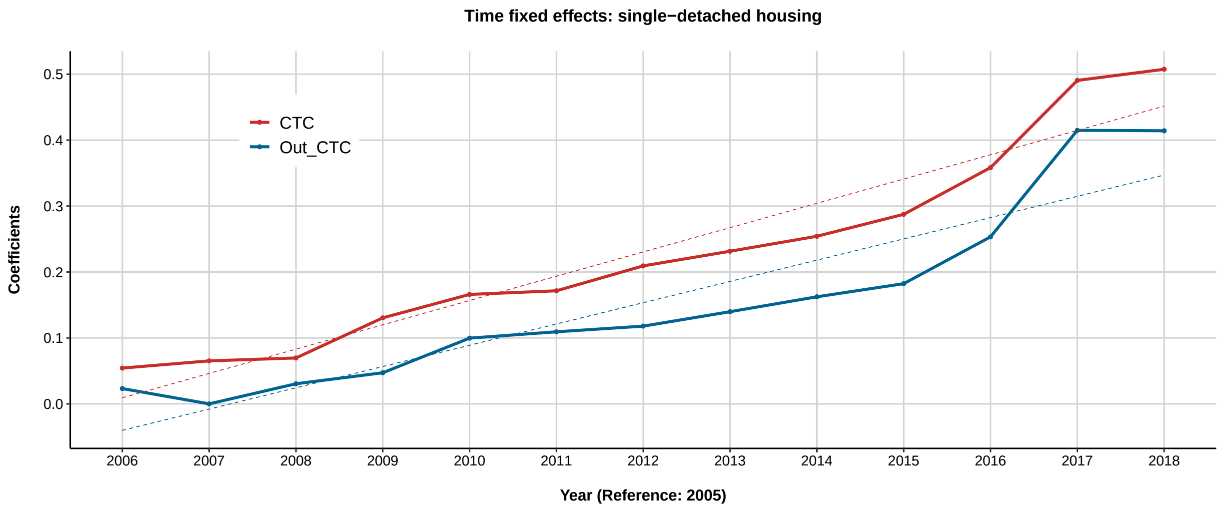


Figure 2.7: Time fixed effects estimates within vs. outside the CTC - single-detached housing

First, without controlling for the temporal relations in the SAR model tends to overestimate the time effects compared to the STAR model, and similarly when we compare the SAR+MLM model with the STAR+MLM model. Based on the estimates from the preferred STAR+MLM model, Figure 2.7 depicts the sales price changes over the years after controlling for the major housing attributes for single-detached houses within the CTC and outside the CTC, respectively. Despite similar patterns, we do see a higher price premium for houses in the CTC over the years, especially after 2011 when the region announced the LRT development.

Table 2.6 presents the transformed estimates of coefficients for the main spatial variables. As expected, sales prices of the single-detached houses are higher when they are outside the CTC area, adjacent to open space or have better access to open space amenities within the walking distance, and have better street connectivity but are not close to bus stops. When comparing the estimates, we find that ignoring the neighbourhood effects seems to underestimate the magnitude of some spatial variables, such as the negative effect of houses in the CTC and the price premium of houses with better open space access, and the synergy between the CTC and the intersection density. The impacts of

Table 2.6: Estimates for the variables of interest from the four models - single-detached houses

Variables of interest	SAR	SAR+MLM	STAR	STAR+MLM
in_ctc (dummy)	-3.8%***	-11.2%***	-2.5%*	-11.0%***
os_adj (dummy)	2.4%***	3.0%***	2.4%***	3.0%***
os_area (km^2)	2.2%***	3.3%***	3.1%***	3.5%***
os_area:in_ctc (km^2)	4.1%***	3.4%***	2.9%***	3.1%***
dis_bus (100 meters)	0.4%***	0.4%***	0.4%***	0.5%***
inter_dense	0.1%***	0.2%***	0.1%***	0.2%***
in_ctc:inter_dense	0.0%	0.2%***	0.0%	0.2%***
edu_rate (%)	0.3%***	-0.1%*	0.4%***	-0.1%*
pop_dense (1000/km^2)	0.0%	-0.2%	-0.2%*	-0.2%

Note: *p<0.1; **p<0.01;***p<0.001. The estimates are transformed by $(e^{\hat{\beta}} - 1) \times 100$

bus transit access and intersection density are not significantly different across models. Again, both education rate and population density become less important in housing price determination when the neighbourhood effects are controlled in models.

Based on the STAR+MLM model results, sales prices of single-detached houses within the CTC are estimated to be 11% less than those outside the CTC after controlling for the other attributes, and houses being 100 meters closer to a bus stop decreases by about 0.5% in sales prices. However, for open space amenities, being adjacent to open space increases sales prices by 3%; 1 more km^2 open space within 800 meters increases by another 3.5%; and being within the CTC adds an extra 3.1% for house with 1 more km^2 open space. When comparing with the condo results in Table 2.4, we find that open space amenities are more valued by the homebuyers of single-detached houses, while being within the CTC and bus transit access are more valued by the condo buyers.

Table 2.7: Estimates of coefficients from the four models - single-detached housing

	SAR model		SAR + MLM model		STAR model		STAR + MLM model	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
(Intercept)	9.47***	0.102	10.848***	0.238	11.4068***	0.037	11.94***	0.05
age	-0.002***	0.000	-0.003***	0.000	-0.0019***	0.000	-0.0027***	0.000
tot_area	0.228***	0.002	0.232***	0.002	0.2274***	0.002	0.2322***	0.002
lot_size	0.515***	0.008	0.463***	0.008	0.5373***	0.008	0.4682***	0.007
baths	0.035***	0.001	0.033***	0.001	0.0369***	0.001	0.0333***	0.001
beds	0.001	0.001	-0.001	0.001	0.0019	0.001	-5.00E-04	0.001
garage	0.045***	0.001	0.044***	0.001	0.0459***	0.001	0.0445***	0.001
fireplace	0.056***	0.001	0.041***	0.001	0.0576***	0.001	0.0409***	0.001
pool	0.056***	0.003	0.055***	0.003	0.0571***	0.003	0.0548***	0.003
os_adj	0.024***	0.002	0.03***	0.002	0.0243***	0.002	0.0302***	0.002
os_area	0.022***	0.003	0.032***	0.003	0.0312***	0.003	0.0339***	0.003
rd_adj	-0.047***	0.003	-0.041***	0.002	-0.0474***	0.003	-0.0405***	0.002
in_ctc	-0.039***	0.011	-0.119***	0.012	-0.0249*	0.011	-0.1166***	0.012
dis_bus	0.004***	0.001	0.004***	0.001	0.0038***	0.001	0.0046***	0.001
inter_dense	0.001***	0.000	0.002***	0.000	0.001***	0.000	0.0016***	0.000
edu_rate	0.003***	0.000	-0.001*	0.000	0.004***	0.000	-6e-04*	0.000
pop_dense	0.000	0.001	-0.002	0.004	-0.0018*	0.001	-0.0018	0.004
sale_year2006	0.023***	0.004	0.024***	0.003	0.0211***	0.004	0.0232***	0.003
sale_year2007	-0.001	0.004	0.002	0.004	-0.0063	0.004	2.00E-04	0.004
sale_year2008	0.027***	0.004	0.033***	0.004	0.021***	0.004	0.0305***	0.004
sale_year2009	0.043***	0.004	0.049***	0.004	0.0369***	0.004	0.0472***	0.004
sale_year2010	0.096***	0.004	0.103***	0.004	0.0868***	0.004	0.0997***	0.004
sale_year2011	0.105***	0.004	0.113***	0.004	0.0956***	0.004	0.1093***	0.004
sale_year2012	0.099***	0.004	0.122***	0.004	0.0858***	0.004	0.1179***	0.004
sale_year2013	0.121***	0.004	0.144***	0.004	0.1071***	0.004	0.1398***	0.004
sale_year2014	0.144***	0.004	0.167***	0.004	0.1284***	0.004	0.1624***	0.004
sale_year2015	0.163***	0.004	0.187***	0.004	0.1476***	0.004	0.1823***	0.004
sale_year2016	0.235***	0.004	0.259***	0.004	0.2165***	0.004	0.2533***	0.004
sale_year2017	0.398***	0.004	0.424***	0.004	0.3736***	0.004	0.4149***	0.004
sale_year2018	0.397***	0.007	0.423***	0.007	0.3742***	0.008	0.4142***	0.007
in_ctc:inter_dense	0.000	0.000	0.002***	0.000	1.00E-04	0.000	0.0015***	0.000
os_area:in_ctc	0.04***	0.008	0.033***	0.009	0.0288***	0.008	0.0312***	0.009
in_ctc:sale_year2006	0.029**	0.01	0.032***	0.009	0.0261**	0.01	0.0311***	0.009
in_ctc:sale_year2007	0.053***	0.01	0.066***	0.009	0.0512***	0.01	0.0652***	0.009
in_ctc:sale_year2008	0.046***	0.011	0.040***	0.01	0.0458***	0.01	0.0391***	0.01
in_ctc:sale_year2009	0.088***	0.01	0.085***	0.01	0.086***	0.01	0.0833***	0.01
in_ctc:sale_year2010	0.07***	0.011	0.068***	0.01	0.0695***	0.011	0.0663***	0.01
in_ctc:sale_year2011	0.064***	0.011	0.063***	0.01	0.0618***	0.011	0.0622***	0.01
in_ctc:sale_year2012	0.065***	0.011	0.093***	0.01	0.058***	0.011	0.0914***	0.01
in_ctc:sale_year2013	0.073***	0.011	0.094***	0.01	0.0639***	0.011	0.0917***	0.01
in_ctc:sale_year2014	0.077***	0.011	0.094***	0.01	0.0671***	0.011	0.0918***	0.01
in_ctc:sale_year2015	0.085***	0.011	0.107***	0.01	0.0746***	0.011	0.1052***	0.01
in_ctc:sale_year2016	0.088***	0.01	0.107***	0.01	0.078***	0.01	0.105***	0.009
in_ctc:sale_year2017	0.068***	0.01	0.077***	0.01	0.0611***	0.01	0.0757***	0.01
in_ctc:sale_year2018	0.077***	0.02	0.095***	0.019	0.0654**	0.02	0.0933***	0.019
Observations	41272		41272		41272		41272	

Note: *p<0.1; **p<0.01;***p<0.001; S.E. standard error

2.5.3 Semi-detached/duplex housing models

Similar to the models for condos and single-detached houses, the STAR+MLM model for semi-detached or duplex housing also finds significant neighbourhood effects and neighbourhood-level dependence ($\hat{\lambda} = 0.634$) and produces better model fit, as reported in Table 2.8. Without considering the temporal causality in the SAR models overestimates the dependence between properties ($\hat{\rho} = 0.298$ and 0.410, respectively). We find significant spatial clustering patterns of the neighbourhood-level effects on semi-detached/duplex housing prices.

Table 2.8: Summary of model statistics - semi-detached/duplex housing

	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\sigma_\varepsilon^2}$	$\hat{\sigma_u^2}$	R^2	DIC
SAR	0.298	-	0.0151	-	0.719	34371862
SAR + MLM	0.410	0.582	0.0125	0.0058	0.758	40419801
STAR	0.007	-	0.0151	-	0.713	588204
STAR + MLM	0.004	0.634	0.0125	0.0063	0.762	718981

DIC: Deviance Information Criterion

With respect to the coefficient estimates in Table 2.9, the four models all find that the sales prices of semi-detached houses or duplexes are positively influenced by the housing area, lot size, bathrooms, garage and pool, while the magnitudes of lot size, garage and open space are significantly less than the single-detached houses. This might indicate that people who buy single-detached houses are willing to pay more for a spacious yard, more garage space and better open space amenities nearby. Better open space access in the CTC area and improved street connectivity (or walkability) also increase the sales prices of semi-detached housing. However, semi-detached housing in the CTC seems to be not significantly different in prices compared to that housing outside the CTC, and bus transit access has no significant influence on the prices of semi-detached housing.

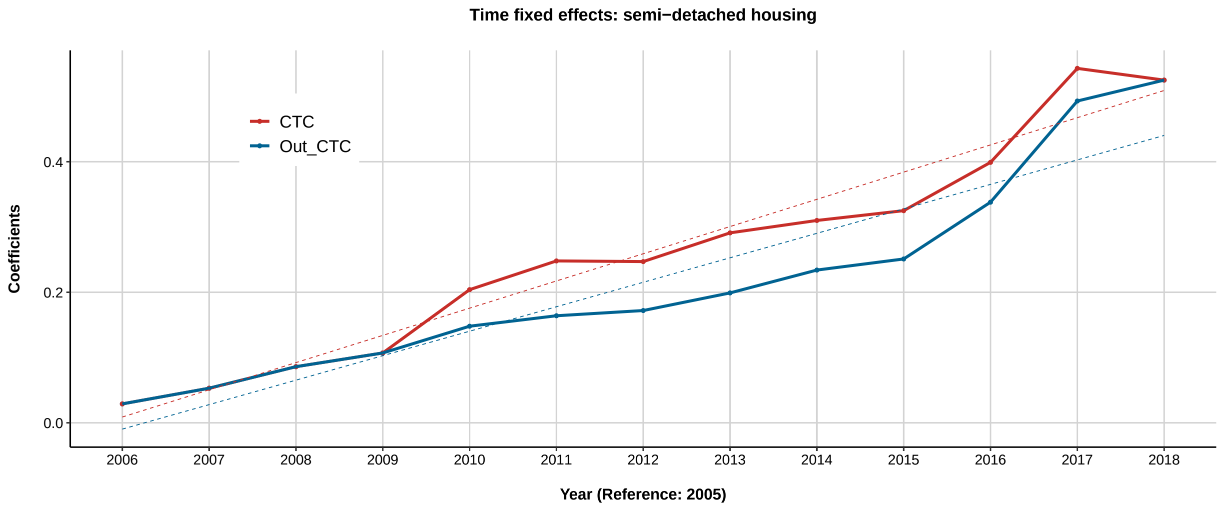


Figure 2.8: Time fixed effects estimates within vs. outside the CTC - semi-detached housing

When comparing the coefficient magnitudes across models, we find that without controlling for the neighbourhood effects seems to underestimate the impact of open space access and the synergy between open space and the CTC. Ignoring the temporal causality seems to have modest impacts on most coefficient estimates. Figure 2.8 presents the additional price variations over the years after controlling for the observed housing attributes in the STAR+MLM model for semi-detached housing. The plot shows a significant price premium for semi-detached houses in the CTC from 2010 to 2017.

Table 2.9: Estimates of coefficients from the four models - semi-detached/duplex housing

	SAR model		SAR + MLM model		STAR model		STAR + MLM model	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
(Intercept)	8.148***	0.507	6.943***	1.117	11.752***	0.039	12.033***	0.061
age	-0.002***	0.000	-0.002***	0.000	-0.002***	0.000	-0.002***	0.000
tot_area	0.227***	0.006	0.226***	0.006	0.233***	0.006	0.227***	0.006
lot_size	0.162***	0.016	0.138***	0.015	0.158***	0.017	0.136***	0.015
baths	0.069***	0.003	0.056***	0.003	0.069***	0.003	0.056***	0.003
garage	0.025***	0.003	0.03***	0.003	0.028***	0.003	0.031***	0.003
fireplace	0.034***	0.004	0.029***	0.004	0.031***	0.004	0.029***	0.004
pool	0.086***	0.02	0.087***	0.018	0.083***	0.02	0.084***	0.019
os_adj	0.012*	0.005	0.012*	0.005	0.011*	0.005	0.012*	0.005
os_area	0.002	0.01	0.057***	0.012	0.006	0.01	0.063***	0.012
rd_adj	-0.002	0.004	-0.011**	0.004	-0.002	0.004	-0.01*	0.004
in_ctc	0.097***	0.022	0.047*	0.029	0.102***	0.022	0.044	0.029
dis_bus	-0.005	0.004	0.002	0.004	-0.005	0.004	0.002	0.004
edu_rate	0.001***	0.000	-0.002***	0.001	0.002***	0.000	-0.002***	0.001
pop_dense	-0.007***	0.002	-0.013*	0.007	-0.01***	0.002	-0.013*	0.007
inter_dense	0.001***	0.000	0.002***	0.000	0.001***	0.000	0.002***	0.000
sale_year2006	0.028***	0.008	0.03***	0.007	0.027***	0.008	0.029***	0.007
sale_year2007	0.053***	0.008	0.054***	0.007	0.051***	0.008	0.053***	0.007
sale_year2008	0.085***	0.008	0.088***	0.007	0.083***	0.008	0.086***	0.007
sale_year2009	0.107***	0.008	0.108***	0.008	0.106***	0.008	0.107***	0.008
sale_year2010	0.15***	0.008	0.15***	0.008	0.146***	0.008	0.148***	0.007
sale_year2011	0.167***	0.008	0.166***	0.007	0.164***	0.008	0.164***	0.007
sale_year2012	0.157***	0.008	0.173***	0.008	0.153***	0.008	0.172***	0.008
sale_year2013	0.19***	0.008	0.201***	0.008	0.185***	0.009	0.199***	0.008
sale_year2014	0.221***	0.008	0.236***	0.008	0.216***	0.008	0.234***	0.008
sale_year2015	0.237***	0.009	0.253***	0.008	0.231***	0.008	0.251***	0.008
sale_year2016	0.326***	0.008	0.34***	0.008	0.319***	0.008	0.338***	0.008
sale_year2017	0.482***	0.008	0.496***	0.008	0.476***	0.008	0.493***	0.008
sale_year2018	0.517***	0.016	0.529***	0.015	0.509***	0.016	0.525***	0.015
in_ctc:inter_dense	-0.002***	0.000	-0.002***	0.000	-0.002***	0.000	-0.002***	0.000
os_area:in_ctc	0.015	0.021	0.094***	0.025	0.012	0.021	0.095***	0.025
in_ctc:sale_year2006	0.003	0.02	0.008	0.018	0.000	0.019	0.009	0.018
in_ctc:sale_year2007	0.006	0.019	0.006	0.017	0.004	0.019	0.008	0.017
in_ctc:sale_year2008	0.012	0.021	0.019	0.019	0.01	0.02	0.021	0.019
in_ctc:sale_year2009	0.01	0.02	0.013	0.018	0.007	0.02	0.013	0.018
in_ctc:sale_year2010	0.04*	0.02	0.054**	0.018	0.041*	0.02	0.056**	0.018
in_ctc:sale_year2011	0.09***	0.021	0.084***	0.019	0.087***	0.021	0.084***	0.019
in_ctc:sale_year2012	0.049*	0.021	0.076***	0.019	0.041*	0.02	0.075***	0.019
in_ctc:sale_year2013	0.08***	0.021	0.092***	0.019	0.076***	0.021	0.092***	0.019
in_ctc:sale_year2014	0.065**	0.021	0.077***	0.019	0.062**	0.021	0.076***	0.019
in_ctc:sale_year2015	0.052*	0.021	0.074***	0.019	0.047*	0.021	0.074***	0.019
in_ctc:sale_year2016	0.029	0.021	0.061**	0.019	0.024	0.021	0.061**	0.019
in_ctc:sale_year2017	0.026	0.02	0.05**	0.018	0.02	0.019	0.05**	0.018
in_ctc:sale_year2018	0.042	0.043	0.000	0.04	0.05	0.044	0.002	0.04
Observations	7076		7076		7076		7076	

Note: *p<0.1; **p<0.01;***p<0.001; S.E. standard error

2.5.4 Townhouse models

For townhouses, the STAR+MLM model finds significant neighbourhood effects. The estimated higher-level variance ($\hat{\sigma_u^2} = 0.0104$) from the STAR+MLM model for townhouses is much larger than the lower-level variance ($\hat{\sigma_\varepsilon^2} = 0.0055$), suggesting that the unexplained housing price variations among townhouses are more attributed to the unobserved neighbourhood differences than the unobserved property differences. The STAR+MLM model also finds the spatial dependence at the neighbourhood-level ($\hat{\lambda} = 0.293$) is much less when compared to the other three housing types. Less spatial clustering patterns of the neighbourhood-level effects on townhouse prices are identified. In addition, the spatio-temporal dependence at the property level becomes less significant and even negative in the STAR+MLM model for townhouses. This might suggest that the prior 3-months' sales of townhouses within 2.5 km do not significantly influence the sales price of a particular townhouse. In general, the STAR+MLM model presents a better model fit in terms of a much lower DIC value than the SAR models, and it can explain about 87.8% of variation in the data.

Table 2.10: Summary of model statistics - townhouses

	$\hat{\rho}$	$\hat{\lambda}$	$\hat{\sigma_\varepsilon^2}$	$\hat{\sigma_u^2}$	R^2	DIC
SAR	0.140	-	0.0068	-	0.849	26108347
SAR + MLM	-0.164	0.319	0.0055	0.011	0.878	31600998
STAR	-0.003	-	0.0069	-	0.849	676338
STAR + MLM	-0.001	0.293	0.0055	0.0104	0.878	863164

DIC: Deviance Information Criterion

Table 2.11 reports the estimates from the four models. Most of the estimates for the structural housing attributes are as expected, where the prices of townhouses increase with larger lot size, more living area, more bathrooms, garages and fireplaces. Better access to open space and better street connectivity can significantly increase the prices of townhouses. When comparing the coefficient magnitudes across models, we find that ignoring the temporal causality seems to have no significant impacts on most coefficient estimates. Figure 2.9 plots out the additional price variations over the years after controlling for the major attributes in the STAR+MLM model and indicates a price surge from 2016 to 2017.

Table 2.11: Estimates of coefficients from the four models - townhouses

	SAR model		SAR + MLM model		STAR model		STAR + MLM model	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
(Intercept)	9.927***	0.246	14.102***	1.231	11.686***	0.023	12.048***	0.054
age	-0.004***	0.000	-0.004***	0.000	-0.004***	0.000	-0.004***	0.000
tot_area	0.214***	0.008	0.199***	0.008	0.222***	0.008	0.199***	0.008
lot_size	0.86***	0.049	0.699***	0.046	0.879***	0.049	0.694***	0.046
baths	0.043***	0.003	0.031***	0.003	0.045***	0.003	0.031***	0.003
garage	0.1***	0.005	0.084***	0.005	0.108***	0.005	0.083***	0.005
fireplace	0.05***	0.004	0.029***	0.004	0.051***	0.004	0.03***	0.004
pool	0.04	0.027	0.05*	0.024	0.04	0.027	0.051*	0.024
os_adj	0.023***	0.003	0.021***	0.003	0.024***	0.003	0.021***	0.003
os_area	0.034***	0.006	0.033***	0.007	0.037***	0.006	0.033***	0.008
rd_adj	0.002	0.005	-0.004	0.004	0.002	0.005	-0.005	0.005
in_ctc	-0.504***	0.047	-0.424***	0.078	-0.559***	0.046	-0.41***	0.078
dis_bus	0.004*	0.002	0.006*	0.002	0.005*	0.002	0.006*	0.002
edu_rate	0.005***	0.000	0.000***	0.001	0.005***	0.000	0.000	0.001
pop_dense	0.009***	0.002	-0.008	0.007	0.005***	0.001	-0.009	0.006
inter_dense	0.001**	0.000	0.002***	0.000	0.001***	0.000	0.002***	0.000
sale_year2006	0.033***	0.007	0.029***	0.006	0.034***	0.007	0.03***	0.006
sale_year2007	-0.04***	0.009	-0.021*	0.008	-0.048***	0.009	-0.02*	0.008
sale_year2008	0.009	0.009	0.025**	0.008	0.003	0.009	0.027**	0.008
sale_year2009	0.04***	0.009	0.056***	0.008	0.034***	0.009	0.058***	0.008
sale_year2010	0.078***	0.009	0.095***	0.008	0.073***	0.009	0.096***	0.008
sale_year2011	0.097***	0.009	0.119***	0.008	0.093***	0.009	0.121***	0.008
sale_year2012	0.087***	0.009	0.129***	0.009	0.081***	0.009	0.131***	0.009
sale_year2013	0.108***	0.009	0.151***	0.009	0.103***	0.009	0.153***	0.009
sale_year2014	0.131***	0.009	0.176***	0.009	0.126***	0.009	0.179***	0.009
sale_year2015	0.158***	0.009	0.201***	0.009	0.153***	0.009	0.203***	0.009
sale_year2016	0.236***	0.008	0.283***	0.009	0.231***	0.009	0.285***	0.009
sale_year2017	0.407***	0.008	0.454***	0.009	0.405***	0.009	0.457***	0.009
sale_year2018	0.428***	0.014	0.472***	0.013	0.427***	0.014	0.474***	0.013
os_area:inter_dense	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000
in_ctc:inter_dense	0.009***	0.001	0.01***	0.001	0.01***	0.001	0.009***	0.001
os_area:in_ctc	0.274***	0.058	0.014	0.096	0.32***	0.058	0.009	0.097
Observations	4546		4546		4546		4546	

Note: *p<0.1; **p<0.01;***p<0.001; S.E. standard error

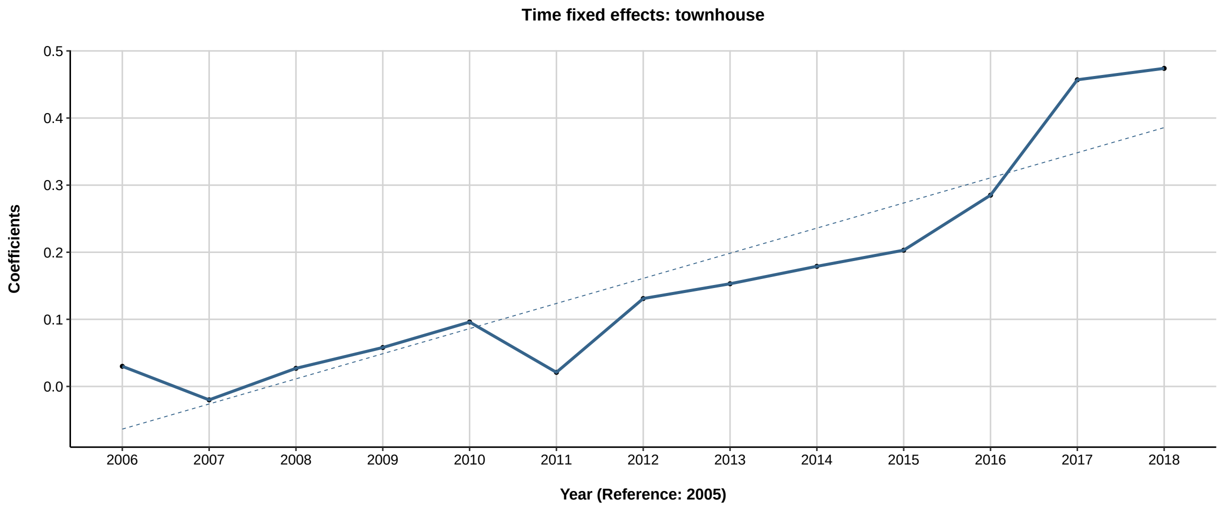


Figure 2.9: Time fixed effects estimates within vs. outside the CTC - townhouses

2.6 Conclusion and discussion

2.6.1 Synthesis of key findings

This paper analyzes the impact of accounting for both spatio-temporal dependence and neighbourhood effects within the setup of traditional spatial autoregressive models (SAR) on model performances and parameter estimates. Using a large housing transaction data set in Kitchener-Waterloo from 2005 to 2018, we specify and estimate models for four housing types: condos, single-detached houses, semi-detached houses/duplexes, and townhouses. The key findings from the models of the four housing types are synthesized as below.

1. Ignoring the spatio-temporal relations in the SAR models (both the SAR model and the SAR+MLM model) overestimates the property-level dependence. Studies using pooled spatial data for hedonic analysis should be cautious of misspecification of the spatial and temporal relationships.

2. Considering the spatio-temporal relationships in the STAR models (both the STAR model and the STAR+MLM model) produces a much lower spatio-temporal dependence at the property level, but generates significantly better model fit. For most

housing types, the impact of the past 3 months' sales of neighbouring properties (within 2.5 km) is significant and positive, confirming the "recent comparable sales" approach in price determination, except for townhouses.

3. Further considering the nesting structure of housing data in the STAR+MLM model, we find significant spatial heterogeneity in price determination across neighbourhoods and significant spatial dependence at the neighbourhood level. The unexplained price variances in condos, semi-detached houses and townhouses are attributed to the unobserved neighbourhood-level differences to some extent. The unexplained price variance in single-detached houses is largely attributed to the unobserved property-level differences.

In brief, this study verifies the proposed three hypotheses. In particular, this study argues the need to take the underlying economic process of housing into hedonic modelling. In other words, hedonic studies need to explicitly put "time" into space (Thanos et al., 2016) and consider the temporal causality in the price determination process. The STAR+MLM model outperforms the other models in particular due to its ability to isolate the lower-level spatio-temporal dependence, the higher-level dependence as well as neighbourhood heterogeneity.

2.6.2 Discussion

For different housing types, the impacts of different model specifications on coefficient estimates are not consistent. Models controlling neighbourhood effects or not are found to produce different estimates for most CTC related variables. Ignoring the temporal causality in models seems to generate inconsistent impacts on different housing types and different variables. For condos and single-detached houses, without considering the spatio-temporal correlations, the SAR models seem to have overestimated the time fixed effects; for semi-detached houses and townhouses, they do not seem to produce significant changes in the magnitude of the time fixed effects.

When focusing on the variables of interest based on the preferred STAR+MLM models, we find that: 1) for condo housing, people are willing to pay 18.8% more for condos that are within the CTC area, and they are willing to pay even more if they have better open space access in the CTC area. Bus transit access also significantly impacts the condo prices, where a 100 meters closer to the nearest bus stop would increase 1.7% of a particular condo price; 2) for single-detached houses, people are not willing to pay more for houses within the CTC area, and the house prices in the CTC are 11% less than the houses outside the CTC area; however, we do find that people are willing to pay for houses in the CTC area when they have better open space access. As expected, the further from bus stops, the higher prices for single-detached house are; 3) for semi-detached houses or duplex, bus access and the CTC area do not seem to significantly influence the housing prices, while people are also willing to pay for the better open space access, and even more so if they are within the CTC area; and 4) for townhouses, better access to open space and better street connectivity can significantly increase the prices of townhouses. In conclusion, although the LRT has not started operation in KW during the study period, we do find the synergy between the CTC and open space access for most housing types, especially for condos. This might indicate that governments should provide both better LRT access and open space amenities so as to attract more residents in the central area and intensify and vibrate the urban cores.

Despite the superiority of the preferred STAR+MLM specification in model performance, this study is not without caveats. First, a common issue related to spatial analyses is the boundary problem (also called edge effect), which originates from ignoring the neighbours outside the boundary. In our paper, despite that we consider all the transaction data in Kitchener Waterloo, the transactions from the surrounded townships of the Region are not considered, which might generate some statistical bias for parameter estimates. Similarly, the boundary issue is also relevant to the spatio-temporal analyses, where not only the spatial boundary matters, but the temporal boundary. As mentioned before, we assume the strict "arrow of time" assumption where only the past three month's sales can affect the current sales price; however, for the first three months' observations in our dataset (i.e., transactions from January 2005 to March 2005), the influence of the past three months' housing transactions is not fully captured, which occurs due to the missing "temporal neighbours" outside our dataset.

For instance, for the spatial-temporal weight matrix (W_st) of single-detached housing (n=41272), the number of neighbours (i.e., the spatio-temporal neighbours filtered by the spatial and temporal distance) is 105 on average; however, when focusing on the first three months' observations (n=444), the average number of neighbours becomes much less (only 25). Even though some research (Thanos et al., 2016) considers all the relations in the past, concurrent and future to avoid the cases with less or no neighbours in the weight matrix, the temporal boundary problem and the potential statistical bias remains unresolved (Higgins et al., 2019). Note that the spatial boundary effect might be diminished when using large samples according to (Anselin, 1988); however, the impact of the temporal boundary problem has not been studied (Higgins et al., 2019).

Second, we did not further conduct sensitivity analysis with different combinations of distance cut-offs and temporal cut-offs in defining the spatio-temporal weight matrix. In particular, the spatial or temporal influence might be not the same for different housing types. Although we choose 2.5 km for the spatial threshold and 3 months for the temporal threshold based on expert views and statistical tests, future work comparing the impact of different spatio-temporal specifications should be done, as in Dubé and Legros (2014). Third, we used the Euclidean distance from bus transit stops as a proxy of transit accessibility instead of the network distance, mainly due to lack of good-quality street network data over time. There are also other intangible variables such as distance to workplaces or commuting time that we did not consider in the model. Fourth, when constructing the weight matrix for condos, we manually geocoded the condo unites in the same building by moving them a bit away from each other to ensure their distance would not be zero. However, a better way to deal with this issue for condos would be using a 3-D distance metric considering the vertical distance as well as the 2-D spatial distance between housing units as proposed by (Higgins et al., 2019). For another, the higher-level "neighbourhoods" in our multi-level modelling are defined by census tracts, while different neighbourhood definitions might also impact the estimation results. The MPAC expert also suggested to use the homogeneous neighbourhoods that they use for property assessment purpose. Lastly, considering the housing market dynamics in KW during the study period, we can further test the period before the housing boom in 2016-2017 and the period during the boom to see whether the underlying spatio-temporal interactions across properties are different in different market conditions, as in Hyun and Milcheva (2018).

Chapter 3

Identifying heterogeneous residential preferences during the construction of a new light-rail transit line

Rosen's(1974) hedonic theory has been extensively applied to various housing studies. Most use the first-stage hedonic model to evaluate the implicit prices of neighbourhood amenities and environmental attributes, but few have further explored households' heterogeneous preferences for those housing attributes. As the urban growth paradigm shifts from sprawl towards intensification through transit investment and compact developments in many North American cities, what houses and locations different households prefer to reside in becomes a key research question. This research proposes a two-stage method to investigate residential preference heterogeneity among different homebuyers from cities of Kitchener and Waterloo, Canada when a new light-rail transit (LRT) line was under construction. Using data from a uniquely designed 2017 housing survey, we aim to uncover the complex relationship between the LRT investment, residential location choices and housing market outcomes.

3.1 Introduction

Heterogeneous preferences for both dwelling and location characteristics are key factors of residential location choices, which can further drive urban social and spatial structure changes (Schirmer et al., 2014). Theoretically, households sort across jurisdictions according to preferences (Tiebout, 1956) and the levels of local public goods (such as school quality, public safety or open space amenities) in different locations (Bayer et al., 2007; Klaiber and Phaneuf, 2010). A better understanding of individual household preferences helps explain the underlying sorting process and the aggregate distribution of residents within a metropolitan area (Kuminoff et al., 2013). Previous studies (Bajari and Kahn, 2005; Bayer et al., 2017; Vasanen, 2012; Massey and Tannen, 2018) have revealed that residential segregation in many cities is partly driven by forces such as income stratification and preferences for racial and ethnicity similarity. Preferences of the middle class for spacious housing in suburbs have also reinforced the process of urban decentralization (Mieszkowski and Mills, 1993). Recently, many cities have transformed urban growth policies from sprawl to intensification, mainly through transit investment and compact developments to attract more people back into urban areas (Dittmar et al., 2004). Up-to-date knowledge on residential preferences can help policymakers to better establish housing plans in station areas in order to satisfy the needs of various households.

Many studies have explored residential preferences of households varying in demographics (Lee et al., 2019), lifecycle stages (Smith and Olaru, 2013), socio-economic status (income, education, ethnicity etc.) (Clark, 2009) and attitudes, values or lifestyle (Ærø, 2006). The 2017 National Community and Transportation Preference Survey shows that 53% of respondents in the 50 largest metro areas of the U.S. prefer walkable, mixed-use urban communities to conventional suburban communities (NAR, 2017). Among them, the younger generation, especially the millennial generation (born from 1981 to 1996), has shown stronger urban preferences, similarly reported by Lee et al. (2019). Retirees also present preferences for urban communities (NAR, 2017). Smaller-size households prefer smaller housing and better access to services in urban centres, while families with children often place value on spacious housing and green space in suburbs (Kim et al., 2005). In addition, studies also show that social connectedness to urban living, underlying value orientations (such as "self-direction"), and other subjective attitudes (such as environmentalism) are also important determinants of housing preferences and location choices (Karsten, 2007; Liao et al., 2015).

These studies mostly use stated-choice experiments or directly ask questions about what residential environments people prefer; however, they only capture preferences for hypothetical communities (known as stated preferences, SP) but not preferences underlying their actual behaviour (known as revealed preferences, RP). An evident inconsistence or mismatch (Vasanen, 2012) between stated preferences and revealed preferences can be seen from the survey report (NAR, 2017), which shows that despite more than half of respondents preferring urban living, the majority (60%) of them currently still live in detached single family houses. This might indicate an undersupply of preferred urban housing units, or trade-offs between dwelling and neighbourhood attributes based on different households' actual needs and preferences (Cao, 2008). Therefore, it is essentially important to understand the heterogeneous preferences underlying location choices of actual movers.

The housing literature (Pan, 2019; Mulley et al., 2018; Cao and Hough, 2012; Duncan, 2011; Billings, 2011; Higgins and Kanaroglou, 2016) has long studied the effects of transit investment on residential property values, using the first-stage hedonic model (Rosen, 1974). The implicit prices estimated from the model are interpreted as willingness to pay (WTP) for housing attributes. However, they do not provide information about the preferences of different households. Ignoring preference heterogeneity limits the ability of empirical studies to understand how transit policies influence residential location choices of different households.

This study applies a two-stage hedonic demand model to examine the housing preferences of different homebuyers during the construction of a new light-rail transit (LRT) project in Kitchener Waterloo (KW), a mid-size urban area in southern Ontario, Canada. We aim to address three questions: (1) During the LRT construction, what dwelling and locational attributes are valued by homebuyers? (2) Are there any differences in willingness to pay for housing attributes across different homebuyer groups? (3) What household characteristics can explain the residential sorting behaviours? We use a specially designed

2017 housing survey to collect the required data. The data are analyzed using a two-stage estimator grounded in Bajari and Benkard (2005) to recover preference parameters. This study is the first to move beyond the first-stage hedonic model to explore the complex relationship between LRT investment, residential location choices and housing market outcomes. With detailed information captured from the survey, we are also able to explore how key socio-demographic and attitudinal factors have contributed to preference heterogeneity.

This paper presents the theoretical foundations of residential location choice modelling, preference identification and estimation methods in Section 2. Section 3 introduces the study area and survey data. Model results are reported in Section 4. The last section summarizes the key findings and caveats.

3.2 Theoretical foundations

3.2.1 A model of residential location choice

According to the seminal work of Rosen (1974), a house is a differentiated product with unique combinations of structural, neighbourhood and locational attributes, $\mathbf{x} = (x_1, x_2, x_3...)$. The price for each house depends on a vector of attributes $\mathbf{x}$, so that a housing market implicitly reveals a function $p(\mathbf{x})$ relating prices and housing attributes. This is the well-known first-stage hedonic price function.

Hedonic equilibrium assumes a market with perfect competition, where all possible combinations of product characteristics are available, and buyers are rational and have full market information. The equilibrium price schedule can be expressed as

$$p_j = \mathbf{p}(\mathbf{x}_j, \xi_j) \tag{3.1}$$

where p_j denotes the housing price of house j, which is determined by the observed housing attributes $\mathbf{x}_j$ and the unobserved characteristics ξ_j. Underlying the equilibrium are consumers with potentially heterogenous preferences and budgets. Following utility maximization theory, each homebuyer h's location choice decision involves a process of maximizing utility subject to a budget constraint

$$\begin{aligned} max \;\; & u_j^h = u^h(\mathbf{x}_j, \xi_j, c) \\ s.t. \;\; & p_j + c \leq y^h \end{aligned} \tag{3.2}$$

The utility u_j^h that house j provides for a given household h, is a function of housing attributes $(\mathbf{x}_j, \xi_j)$ and consumption of a non-housing numeraire good c. y^h represents the household income, which constrains the housing and non-housing expenditure, p_j and c, respectively. Assuming that households are rational utility maximizers, the optimal choice j^* for household h becomes

$$j^*(h) = arg \max_j \;\; u^h(\mathbf{x}_j, \xi_j, c). \tag{3.3}$$

If x_{kj} is continuous, the optimal solution satisfies the following first-order condition, which provides the primary theoretical foundation for residential location choice models,

$$\underbrace{\frac{u^h_{x_{kj^*}}}{u^h_c}}_{\text{Marginal rate of substitution between } x_{kj} \text{ and } c} = \underbrace{\frac{\partial \mathbf{p}(\mathbf{x}_{j^*}, \xi_{j^*})}{\partial x_{kj}}}_{\text{Marginal implicit price of attribute } x_{kj}} \tag{3.4}$$

3.2.2 Preference identification

According to Rosen (1974), by assuming that the observed buyers and sellers are matched in a market equilibria, hedonic pricing functions are able to estimate the implicit market values of housing characteristics. However, Rosen (1974, p.54) notes "...estimated hedonic price-characteristics functions typically identify neither demand or supply". Rather, the hedonic function $p(x)$ represents a joint envelope of a family of demand functions and a family of supply functions. Only when buyers are identical/homogenous, can the hedonic function reveal the demand structure directly. Since households in a city tend to be heterogenous in their preferences, the observed outcome is the result of a complex matching process of households who make tradeoffs among a wide range of both structural and neighbourhood attributes to satisfy their needs (Bayer et al., 2004; Kuminoff et al., 2013). Although a first-stage hedonic method cannot uncover the heterogeneous preferences, combining the properties of market equilibrium from the first-stage hedonic model with household location choice behaviours can do so (Kuminoff et al., 2013). Generally speaking, two fairly broad literature tackling this issue are found in empirical studies.

One emerges from the sorting literature. Equilibrium sorting models build on the intellectual foundations of hedonic models (Rosen, 1974) and discrete-choice models (McFadden, 1978). They use the information provided by an *equilibrium* hedonic price function, together with a formal description of *sorting* behaviour of heterogenous agents, to infer the structure of preferences (Kuminoff et al., 2013). Based on the assumed preference structure, the estimation process involves an iterative procedure equating supply and demand in the market (represented by sorting equilibrium), and follows the discrete-choice modelling approach for preference parameter identification. The sorting framework offers an appealing approach to developing theoretically consistent welfare measures of future policy changes (See Klaiber and Kuminoff (2014) for a review). In particular, it provides a new direction for market simulations, which allow households to re-sort and housing prices to re-equilibrate in responses to "proposed" or "counter-factual" policy changes and unexpected events, such as changes in school quality (Bayer et al., 2007), air quality (Tra, 2010) and open space amenities (Klaiber and Phaneuf, 2010). These models depend on extensive micro data observations and require defining a choice set, basically by aggregating individual houses into housing types within communities. Considering our limited survey sample ($n = 357$) and our main purpose for identifying preference heterogeneity instead of market simulations, we choose not to estimate a sorting model. In particular, the choice set considering combinations of many housing and neighbourhood attributes is likely to be even larger than our sample size, which would make the model unidentifiable.

The second is found in the literature of second stage hedonic demand models. These models use hedonic results from the first stage, and obtain preference parameters of households through a second-stage estimator. One approach for the second-stage demand estimation is based on information from multiple choices of each household type. Repeated choice observations of each household either from panel data (Bishop and Timmins, 2018) or "before" and "after" an exogenous market shock or supply shift (Kuminoff and Pope, 2012), or choices of households from different markets but with common preference structure (Bartik, 1987) can derive demand curves by analyzing the changes in the gradient of hedonic price functions. For example, Poudyal et al. (2009) first defined four submarkets and estimated implicit prices of urban parks and then utilized the price variations across markets to estimate a second-stage demand model for park size. Brasington and Hite (2005) applied the similar approach to estimate demand for environmental quality based on spatial hedonic estimates from six markets. A major problem to be addressed when applying this approach is the endogeneity of the implicit prices in the second stage model.[1] The challenge is to find convincing instruments for implicit prices. This remains an obstacle in housing demand literature.

An alternative approach allows for the second-stage demand estimation through restricting the utility function. These models are also called structural hedonic models (Kuminoff et al., 2013) through restrictions on the shape of demand curves. Chattopadhyay (1999) first applied this approach in air quality analysis. Bajari and Benkard (2005) and Bajari and Kahn (2005) applied similar approach to estimate housing demand for explaining racial segregation in U.S. cities, and von Graevenitz (2013) later used it for environment valuation. We apply this approach in our empirical analysis primarily considering the advantage of its transparent identification strategy based on the functional form specification over the instrument variable approaches.

3.2.3 Estimation method

Our estimation approach is based on Bajari and Kahn (2005)'s three-step models. For the first step, we estimate a hedonic model. In the second step, we estimate household-specific preference parameters based on the hedonic estimates. Lastly, we decompose preference heterogeneity on demographics and attitudinal factors.

[1]Endogeneity of price is a common identification problem when estimating hedonic demand functions. A detailed description of the problem and the resolutions are attached in the Appendix A-2

Step 1: first-stage hedonic regression

For the first-stage hedonic regression, Rosen (1974) points out that a necessary prior condition for estimation of the second-stage demand function is that the first-stage hedonic function should be nonlinear. Ekeland et al. (2004, p. 60) also indicate that "Nonlinearities are generic features of equilibrium in hedonic models and a fundamental and economically motivated source of identification." Bajari and Kahn (2005) argue that when the first-stage hedonic price schedule is nonlinear, variations in the estimates of implicit marginal prices can be obtained, thus adding more information for preference estimation. For those reasons, we build a nonlinear first-stage hedonic model to estimate varying coefficients. The spline fit or quadratic polynomials were considered for estimating the nonlinear model, but they were not able to fit the model with categorical or dummy covariates. Hayfield and Racine (2008) recently developed a nonparametric kernel smoothing methods for mixed data types, which is known as Li-Racine Generalized Kernel Estimation. We apply this method and specify a nonparametric hedonic model in the first step:

$$\begin{aligned}\log(p_j) &= f(\mathbf{x}_j, \xi_j)\\ &= \sum_{k=1}^{11} f(x_{kj}) + \sum_{m=1}^{5} f(x_{mj}) + \xi_j \end{aligned} \tag{3.5}$$

where x_k denotes the continuous housing covariates; x_m refers to the discrete covariates; and ξ accounts for the unobserved housing attributes influencing housing prices. The coefficients to be estimated from the nonparametric model are allowed to vary across observations j. We estimate the nonparametric model specified in equation (3.5), by using the *np* package (Hayfield and Racine, 2008) in R. We selected the adaptive nearest neighbour method for bandwidth selection, and we chose the second-order Gaussian kernel type for the continuous variables and the Li-Racine categorical kernel type for the discrete variables. We then applied the local-constant least squares estimation method for kernel regression. Further, we conducted the kernel regression significance test for each explanatory variable.

Since the dependent variable is $\log(p_j)$, the estimated coefficient $\hat{\alpha_{kj}}$ represents the percentage change of housing price with one unit change of x_k. To obtain the housing price change with one unit change of x_k, that is the implicit hedonic price $\partial\hat{\mathbf{p}}(\mathbf{x}_j, \xi_j)/\partial x_{kj}$, we calculate

$$\frac{\partial\hat{\mathbf{p}}(\mathbf{x}_j, \xi_j)}{\partial x_{kj}} = exp(\hat{\alpha_{kj}} - 1) \cdot p_j \tag{3.6}$$

To make estimates of different variables comparable, we estimate the relative contributions by calculating the standard deviation change of housing price with one standard deviation change of each continuous attribute x_{kj}. For binary variables, for instance single-detached housing, we calculate the standard derivation change of the implicit housing price of single-detached housing compared to non single-detached housing.

Step 2: preference identification

As we argued before, fitting a non-linear first-stage hedonic model is not sufficient for preference estimation, and additional information or assumptions is needed. Therefore, we follow the structural hedonic framework and add parametric assumptions on the utility function form. Chattopadhyay (1999) compares different functional forms of hedonic functions and utility functions, and concludes that the results are robust against different function-form specifications. Accordingly, we assume a quasi-linear utility function in this paper:

$$\begin{aligned} u_j^h &= u^h(\mathbf{x}_j, \xi_j, c) \\ &= \sum_{k=1}^{11} \beta_k^h \log(x_{kj}) + \sum_{m=1}^{5} \beta_m^h x_{mj} + \beta^h \log(\xi_j) + c \end{aligned} \tag{3.7}$$

where utility is log-linear in continuous variables x_k and ξ, and linear in discrete variables x_m and other commodity c. This restrictive assumption implies that the utility is increasing with the housing amenity x_k and becomes concave if people prefer more of this amenity to less. It also supports the properties of diminishing marginal rate of substitution (MRS). β_k^h and β_m^h denote the random preference parameters for the housing attributes. They are assumed to be determined by observed demographics, attitudes and unobserved tastes of households. Thus, the utility of house j provided for household h depends on the housing characteristics and household-specific preferences. To estimate those preference parameters, we discuss the process for both continuous and discrete variables in details.

(1) Continuous variables

To solve equation (3.3), with our utility function specified in equation (3.7), we follow the first-order condition described in equation (3.4) and obtain

$$\frac{u^h_{x_{kj^*}}}{u^h_c} = \frac{\beta^h_k}{x_{kj^*}} = \frac{\partial \mathbf{p}(\mathbf{x}_{j^*}, \xi_{j^*})}{\partial x_{kj}} \tag{3.8}$$

By inverting the above equation and incorporating the estimated marginal price from the first-stage hedonic model, we get

$$\underbrace{\hat{\beta}^h_k}_{\text{Recovered household-specific preference parameter}} = \underbrace{\frac{\partial \hat{\mathbf{p}}(\mathbf{x}_{j^*}, \xi_{j^*})}{\partial x_{kj}}}_{\text{Estimated marginal price from the first-stage hedonic}} \cdot \underbrace{x_{kj^*}}_{\text{Observed value of } x_{kj^*}} \tag{3.9}$$

Equation (3.9) recovers preference parameters for continuous attributes explicitly. It should be noted that the estimated marginal price from the first-stage hedonic model is different for each observation j, thus providing more variation for the preference parameters $\hat{\beta}^h_k$ apart from the difference in the chosen x_{kj^*}. If the first-stage hedonic coefficients are constant, the preference variations can be only explained by the observed values of the attribute, such as the number of bedrooms. However, even those who buy the same number of bedrooms are likely to have different willingness to pay for bedrooms. That being said, what they buy (x_k) reflects only part of their preferences, but how much they pay for that attribute ($\hat{\beta}^h_k$) reflects their underlying preferences. $\hat{\beta}^h_k$ could also be interpreted as the expenditure on the particular housing attribute x_k by household h, and should reflect preference differences across households.

(2) Binary variables

For discrete variables, the first order condition is replaced by a set of inequality constraints as discussed by Bajari and Kahn (2005). We take single-detached housing as an example here to illustrate the preference estimation process. We use $x_m = 1$ to represent a single-detached house and $x_m = 0$ to denote other house types. When household h has chosen a single-detached house j^* (with $x_{mj^*} = 1$), the utility received from this house must be equal to or larger than other types of houses l after controlling for the other housing attributes denoted as $\mathbf{x}_n$, the utility maximization implies that

$$x_{mj^*} = 1 \implies u^h(\mathbf{x}_{j^*}, \xi_{j^*}, c) \geq u^h(\mathbf{x}_l, \xi_l, c)\ \forall l \neq j^* \tag{3.10}$$

Specifically, while controlling for all the other attributes' utility $\bar{u}^h$, the above inequality becomes

$$\beta_m^h + \bar{u}^h + (y^h - p_{j^*}) \geq \bar{u}^h + (y^h - p_l) \tag{3.11}$$

$$\beta_m^h \geq p(x_{mj^*} = 1|\mathbf{x}_n) - p(x_{ml} = 0|\mathbf{x}_n) \tag{3.12}$$

$$\left[x_{mj^*} = 1\right] \implies \left[\beta_m^h \geq \frac{\Delta p}{\Delta x_m}\right] \tag{3.13}$$

This implies that if and only if household h chooses a single-detached house, the preference parameter for single-detached housing exceeds the implicit market price of single-detached house (i.e., $\frac{\Delta p}{\Delta x_m}$).

Step 3: preference regression

Step 1 and Step 2 together provide a way to infer household-specific preference parameters. In step 3, we regress the estimated preference parameters on household demographics and reported attitudes to find household preferences as a function of both types of factors. An early housing demand study by Wheaton (1977) finds that the overt sociodemographic characteristics of households can describe basic differences in housing tastes or preferences. Many other studies show that personal attitudes and latent lifestyles are also key drivers of location choice behaviour (Walker and Li, 2007; Liao et al., 2015; Lewis and Baldassare, 2010). Luckily, our detailed survey data allows us to control for both demographic and attitudinal factors.

(1) Continuous case

We assume the estimated preference parameters to be a linear function of demographics and attitudes. Taking the preference parameter for the attribute x_k as an example, we let

$$\hat{\beta}_k^h = \delta_{k,0} + \delta_{\mathbf{k}} \cdot \mathbf{d}^h + \eta_k^h, \qquad E(\eta_k^h|\mathbf{d}^h) = 0 \tag{3.14}$$

where $\mathbf{d}^h$ denotes a vector of demographics for household h, which in particular refer to household type, household income, homebuyer, age, education and employment status, as well as the reported attitude for that attribute. To estimate $\hat{\delta_{\mathbf{k}}}$ in (3.14), we can simply use ordinary least squares for the regressions, and the residuals are interpreted as the unobserved household-specific taste shocks.

(2) Binary case

We assume that the associated preference parameter is also a linear function of demographics and attitudes,

$$\beta_m^h = \delta_{m,0} + \delta_{\mathbf{m}} \cdot \mathbf{d}^h + \eta_m^h, \qquad \eta_m^h \sim N(0, \sigma^2). \tag{3.15}$$

Since we are not able to identify a specific preference parameter β_m^h, we can not estimate equation (3.15) as a linear regression in (3.14). To estimate $\hat{\delta_{\mathbf{m}}}$, we assume that the error term η_m^h is normally distributed. We already know the underlying condition from equation (3.13), and then we can write the probability of household h choosing to live in single-detached housing as follows,

$$\begin{aligned}
Pr(x_{mj^*} = 1|\mathbf{d}^h) &= Pr\big(\beta_m^h \geq \frac{\Delta p}{\Delta x_m}\big) \\
&= Pr\big(\delta_{m,0} + \delta_{\mathbf{m}} \cdot \mathbf{d}^h + \eta_m^h \geq \frac{\Delta p}{\Delta x_m}\big) \\
&= Pr\Bigg(\eta_m^h \geq \frac{\Delta p}{\Delta x_m} - (\delta_{m,0} + \delta_{\mathbf{m}} \cdot \mathbf{d}^h)\Bigg) \\
&= \Phi\Bigg((\delta_{m,0} + \delta_{\mathbf{m}} \cdot \mathbf{d}^h) - \frac{\Delta p}{\Delta x_m}\Bigg)
\end{aligned} \tag{3.16}$$

Similarly,

$$Pr(x_{mj^*} = 0|\mathbf{d}^h) = 1 - \Phi\Bigg((\delta_{m,0} + \delta_{\mathbf{m}} \cdot \mathbf{d}^h) - \frac{\Delta p}{\Delta x_m}\Bigg) \tag{3.17}$$

Integrating the above cumulative density functions into the likelihood function for the population distribution of preferences for single-detached housing, the coefficients $\hat{\delta_{\mathbf{m}}}$ can be estimated by maximum likelihood estimation (MLE) method.

In brief, by assuming a quasi-linear utility function with random taste coefficients, together with the first-stage hedonic price estimates and the observed household-specific choices and household characteristics, we are able to recover heterogenous tastes for both continuous and dichotomous attributes. This approach differs from recent methods such as the standard logit models which assume homogeneous preference parameters across households.

3.3 Study area and data

3.3.1 Study area

KW is a mid-size urban area located in southern Ontario, Canada. The two municipalities Kitchener and Waterloo, as well as Cambridge and the surrounding townships, collectively make up the Region of Waterloo with a population of 535,154 in 2016 (Statistics Canada, 2017). The region is well known for its high concentration of high-tech industries and rapid economic growth. To transform urban growth from sprawl to intensification, the region proposed a new LRT system aiming to increase intensification of the urban cores and stimulate transit-oriented development (Region of Waterloo, 2019). The 19-km LRT line (Phase One) connecting Kitchener and Waterloo was approved by the Regional Council in 2011 and began construction in 2014.[2]

Geographically, KW is relatively close (roughly 70-120 km) to the Great Toronto and Hamilton Area (GTHA), which is Canada's largest urban region. The GTHA housing market is one of the hottest markets in Canada, where the escalating prices, higher borrowing costs and a new mortgage stress test have prompted some GTHA buyers to seek homes in KW. The more affordable housing in KW, the growing economy and its regional accessibility to the GTHA have made the KW market attractive to the buyers from the GTHA. In fact, the low inventory and unrelenting demand in particular with GTHA buyers migrating to KW have contributed to a housing boom in KW through 2016 to 2017 unexpectedly (KWAR, 2018).

[2]The Phase One LRT line has started services between the two cities since June 2019

3.3.2 Data

We designed a comprehensive housing survey (See the survey questionnaire in Appendix A7) to explore residential location choice behaviours of both home buyers and sellers in KW during the new LRT construction. To find relevant respondents, we requested an address list from Canada Post, which identified 5185 home movers who either bought or sold a home in KW between June 2015 and April 2017. Survey invitations were mailed out to those likely home movers, and 357 buyers (around 10% response rate) and 149 sellers completed the survey via an online survey link or paper survey from June to September of 2017. This paper focuses on results from the homebuyers ($n = 357$) who responded to questions about i) the home buying motivations and characteristics of the homes they bought, ii) the home buying process, iii) stated importance of housing attributes in residential location choices, iv) attitudes towards the new LRT, and v) household characteristics and travel behaviours.

To identify the representativeness of the survey sample, we compared its distribution with the population of transactions obtained from the MPAC company during the same period (from 2015 June to 2017 April, $n = 11692$). To maintain data confidentiality, the spatial distribution maps of the survey sample and the sales dataset are not published. However, we compared the percentage of observations in each census tract for the two dataset (See details in Appendix A2.2). We find that our survey sample has covered most the census tracts as the population dataset, even though we seem to have over-represented the housing units (mainly single-detached houses) in CTs of the suburbs and under-represented the units (mainly condos) in the CTs of the inner urban area.

This study employs the detailed housing and household characteristics from the survey. It should be noted that, we also capture the attitudes of respondents for various housing attributes. In part iii of the survey, we asked respondents to report their perceived importance of structural and neighbourhood attributes in their home decisions. Three options were provided for each attribute, "1 - not important", "2 - somewhat important", and "3 - very important". Those reported attitudes enable us to better understand location choice behaviours, apart from the observed difference in demographics and socio-economics of households. In addition, we have observations of the competitive buyers from the GTHA, which allow us to compare their particular residential preferences for housing in KW to the local buyers. Statistics for housing attributes and household characteristics are summarized below.

Housing attributes

Table 3.1: Descriptive statistics for the variables used in the study

Attribute	Description	Count	Mean	Std.Dev.
Structural attributes				
$SINGLE$	Binary: single detached house	339	0.72	0.45
$BEDM$	Number of bedrooms	339	3.20	0.80
$BATH$	Number of bathrooms	339	2.26	0.75
$GRAG$	Number of garages	340	1.14	0.65
$YARD$	Yard size (square feet)	340	4,091	3,492
BUL_AGE	Building age in 2017	297	30	22
$SIZE$	Categorical: housing size (square feet)	338	NA	NA
	Less than 1,000 square feet	13		
	1,000-1,499 square feet	139		
	1,500-1,999 square feet	102		
	More than 2000 square feet	84		
Locational and neighbourhood attributes				
POP_DENS	Population density (persons/km^2)	327	2,961	2,106
OS_ACES	Open space accessibility	340	42.76	17.84
In_CTC	Binary: in the central transit corridor	340	0.08	0.28
DIS_LRT	Distance to the nearest LRT stop (meters)	340	3,605	1,636
$In\_CTC * DIS\_LRT$	Interaction term - LRT access (meters)	30	844	354
DIS_BUS	Distance to the nearest bus stop (meters)	340	347	310
$POST\_EDU$	Postsecondary education percentage(%)	327	62.35	9.52
OS_ADJ	Binary: open space adjacency	340	0.16	0.37
REG_RD_ADJ	Binary: regional road adjacency	340	0.09	0.29
HP	Housing price (1000$)	327	404	144

Table 3.1 summarizes the descriptive statistics for the structural, locational and neighbourhood attributes used in our study (See the data source of each attribute in the Ap-

pendix A-4). After checking data quality and multicollinearity, we select 16 variables for the first-stage hedonic model, which comprise four binary variables $SINGLE$, In_CTC, OS_ADJ and REG_RD_ADJ, one categorical variable $SIZE$, and 11 continuous variables.[3] The In_CTC variable indicates whether the property is within the central transit corridor (CTC), which is delineated by the Region to represent the areas within a roughly 10-min walk to the LRT stations. We refer to this area as the transit-oriented neighbourhoods in this study. The coefficients of In_CTC identify the price premium provided by the transit-oriented neighbourhoods. The interaction term $In\_CTC * DIS\_LRT$ is added to isolate the impacts of LRT access on the property values apart from the CTC neighbourhood effect.

Household characteristics

Table 3.2 summarizes the characteristics[5] of homebuyers from our survey and compares them with the population of Waterloo Region from Census 2016 (Statistics Canada, 2017). Our sample has a much higher proportion of couple families and a lower proportion of non-family households compared to Census. The median age of the homebuyers in the sample falls into the range of 25-34, compared to 35-54 for the total population of the Region. Thus, most homebuyers in our sample are young couple families. It should be noted that homebuyers who earned more than \$100,000 in 2016 account for a large proportion of our samples (47.9% compared to 35.9% from Census), which suggests that our survey has covered a greater proportion of higher-income households in the region. Given our focus on people looking to buy a home, these differences are not surprising. In addition, 81.1% homebuyers in our sample are local; 11.8% are GTHA buyers; and 7.1% are other

[3]We also tested variables such as number of floors, school quality, average household size, neighbourhood employment rate, neighbourhood average income, safety level and so forth. They were found to be statistically insignificant and not included in the final model specification. Spatial autocorrelation was also tested and it is not significant when the locational attributes are introduced in the model. The partial correlation matrix between the explanatory covariates is also attached in the Appendix A-8

[5]For employment, note that Census 2016 considers the employment status of each household member aged 15 and over, and thus only 39.1% are full-time employed, and others are either not in the labour force or part-time employed. In our survey sample, we classify the households into one group which has at least one full-time job in the household, and the other with no full-time jobs in the household. 87.6% households have at least one full-time job in our sample.

Table 3.2: Summary statistics - homebuyers and comparison with the Census statistics

	Homebuyer survey ($n = 357$)		Census 2016
	Count	Percentage	
Lifecycle characteristics			
Family households	280	84.3%	64.7%
Couple-family with children	131	39.5%	30.6%
Couple-family without children	132	39.8%	24.9%
Lone-parent family	17	5.1%	9.2%
Non-family households	52	15.6%	35.3%
More-persons household	9	2.6%	10.9%
One-person household	43	13.0%	24.4%
Age15 - 24	9	3.5%	16.8%
Age 25 - 34	129	50.6%	17.1%
Age 35 - 54	90	35.3%	34.2%
Age >=55	27	10.6%	31.9%
Socio-demographics			
Less than $50,000	34	10.9%	30.2%
$50,000 - $99,999	129	41.2%	33.9%
$100,000 - $149,999	97	31.0%	20.1%
$150,000 and over	53	16.9%	15.8%
Full-time employed [4]	282	87.6%	39.1%
Not full-time employed	40	12.4%	60.9%
High school	23	7.3%	28.8%
Postsecondary education	151	47.9%	46.54%
Graduate	141	44.8%	5.93%
Other characteristics			
First-time buyers	148	43.8%	NA
Repeat buyers	190	56.2%	NA
GTHA buyer	40	11.8%	NA
Local buyer	274	81.1%	NA
Other buyer	24	7.1%	NA

buyers.[6] More repeat buyers in our sample are observed than the first-time buyers. In addition to the observed household characteristics, attitudes toward the housing attributes are considered.[7] It is worth noting that access to the future LRT stop is a less important factor on average, compared to the other attributes.

3.4 Results

3.4.1 First-stage hedonic results

Following Step 1, we estimate the standard-deviation changes of the housing price with one standard-deviation change of each attribute. We report the mean values as well as the values at the 25th, 50th, and 75th percentiles (labeled $Q1$, $Q2$, and $Q3$) in Table 3.3. The model has a good overall fit with an R^2 above 0.8.[8] The mean relative contributions of the variables show that the structural attributes were the dominant factors determining the housing prices, as expected. Housing type, housing size, the number of bathrooms, garages, yard size, and building age significantly influenced the housing prices in KW, which is consistent with the findings in a hedonic study for KW by Babin (2016).

Among the locational and neighbourhood attributes, education rate had the largest impact on the housing prices. A higher proportion of residents with post-secondary education in a neighbourhood is often associated with a higher income neighbourhood, or "wealthy neighbourhood", which can be a proxy for a higher neighbourhood "quality". On average, people were willing to pay significantly more for the neighbourhoods with a better "quality". The LRT access (CTC_DIS_LRT) and the CTC neighbourhood (In_CTC) did not significantly impact the housing prices in KW during the construction stage. Despite the insignificant price effect of the LRT on the market, we are more interested to under-

[6]Local buyers are from Kitchener, Waterloo, Cambridge, Guelph, London or the surrounded townships or counties, and they are assumed to have better knowledge of the KW housing market. Other buyers are those from other cities in Ontario, or from other provinces of Canada, or international migrants and they are assumed to have least information about the market in KW.

[7]Figure A-7 in the Appendix summarizes the survey responses for each of the attributes used by the homebuyers when choosing a location.

[8]The linear regression was also estimated for the first-stage hedonic model: see the results in the Appendix A-5

Table 3.3: Estimates from the first-stage nonparametric hedonic regression

Attribute	$Q1$	$Q2$	$Q3$	Mean	P value
Structural attributes					
$SINGLE$	0.00000	0.65160	1.28650	0.93270	0.00***
$SIZE$	-0.07031	0.04886	0.21497	0.16145	0.00***
$BEDM$	0.00159	0.00377	0.00760	0.00625	0.08*
$BATH$	0.03436	0.05853	0.09547	0.07646	0.00***
$GRAG$	0.01696	0.03701	0.06665	0.04540	0.00***
$YARD$	0.00512	0.01226	0.02159	0.02177	0.00***
BUL_AGE	-0.13084	-0.05033	0.00157	-0.07505	0.03**
Locational and neighbourhood attributes					
POP_DENS	-0.01531	-0.00421	0.01023	-0.00596	0.61
OS_ACES	-0.00302	0.00317	0.00899	0.00412	0.18
In_CTC	-0.00242	0.00008	0.00094	0.00217	0.66
DIS_LRT	-0.00677	0.00048	0.00689	-0.000001	0.41
CTC_DIS_LRT	-0.00055	0.00001	0.00022	0.00023	0.45
DIS_BUS	-0.00277	0.00006	0.00362	0.00104	0.67
$POST\_EDU$	0.02295	0.08159	0.13689	0.09468	0.01***
OS_ADJ	-0.01664	-0.00013	0.01829	0.01504	0.21
RED_RD_ADJ	-0.01582	-0.00621	-0.00210	-0.01148	0.71
Kernel Regression Estimator				Local-Constant	
Bandwidth Type				Adaptive Nearest Neighbour	
Complete observations				276	
Residual standard error				0.145	
R^2				0.815	

Note: The table reports the mean standard deviation change of the housing price with one standard deviation change of each attribute (based on Step 1), as well as the estimates at the 25th, 50th, and 75th percentiles (labeled $Q1$, $Q2$, and $Q3$). *p<0.1; **p<0.05; ***p<0.01.

stand how has the LRT impacted individual households' location choices. The first-stage estimates are not sufficient to explain the sorting behaviours of different households. The second-stage estimates can help answer this research question.

3.4.2 Heterogeneous residential preferences

We recover preferences through Step 2 and then regress preferences on household characteristics to explore preference heterogeneity following Step 3.

Preferences for the structural housing attributes

Table 3.4: Estimates of the willingness to pay for private yard

	Dependent variable:
	WTP_YARD
Couple-family without children	-5,210 (2,461)**
Lone-parent family household	-3,541(4,695)
More-persons household	-8,465(6,115)
One-person household	-5,566(3,483)
Less than $50,000	1,646(4,194)
$50,000-$99,999	2,365(2,489)
$150,000 and over	10,329(3,045)***
AGE 35-54	-1,313(2,409)
AGE 55+	557(3,440)
Yard size: 2-somewhat important	446(2,845)
Yard size: 3-very important	7,542(3,250)**
Constant	1,528(3,547)
Observations	190
R^2	0.142
Residual Std. Error	13,922
F Statistic	3*** (df = 11)

Note: This is an OLS regression. The dependent variable is the estimated willingness to pay for an increase of yard size from 3000 to 5000 square feet. The omitted category is a couple-family with children, who have $100,000 to $149,999 annual household income, aged 18-34. * p<0.1; ** p<0.05; *** p<0.01.

Table 3.4 presents the willingness to pay differentials for a private-yard size increase from 3000 to 5000 square feet. Let WTP_YARD^h denote the WTP of household h for a yard size increase. Equation 3.7 implies that $WTP\_YARD^h = \hat{\beta}^h(log(5000) - log(3000))$. Given that the random preference parameter $\hat{\beta}^h$ was estimated by equation (3.9), we calculated the measure of WTP_YARD^h and regressed it on household characteristics. The OLS regression results in Table 3.4 show that, as expected, couples with children were willing to pay significantly more for homes with a larger yard. Households with the highest income also demonstrated significantly stronger preferences for homes with a larger yard. In addition, households who had a particularly strong desire for private yard were willing to pay significantly more for that amenity.

Table 3.5: Willingness to pay for the structural housing attributes

	Dependent variable:			
	WTP_SIZE	WTP_BATH	WTP_BEDM	WTP_GRAG
Couple without children	-23,890.6**	-12,646.1**	-1,610.1**	-4,004.7
Lone parent	-28,292.6	-8,281.5	-323.5	-2,808.6
More-persons household	-8,819.1	-33,454.4**	-1,326.6	-10,050.1*
One-person household	-12,285.5	-12,376.6	-1,517.1	-5,822.4*
Less than $50,000	-14,683.1	-10,983.5	-522.8	-2,405.4
$50,000 - $99,999	-10,802.3	-1,954.2	-154.5	-629.3
$150,000 and over	28,479.0**	33,388.0***	1,736.8**	11,584.3***
AGE35-54	5,205.8	5,806.5	297.7	2,211.6
AGE55+	9,846.9	17,543.7*	-319.0	3,931.1
Constant	83,925.0	32,701.8**	3,437.1	10,444.9*
Observations	181	180	181	178
R^2	0.12	0.25	0.11	0.20
Residual Std. Error	60,123.3	28,694.8	3,563.1	12,605.3
F Statistic (df = 17)	1.4	2.8***	1.1	2.4***

Note: Each column presents a separate OLS regression. The dependent variables are the estimated willingness to pay for housing size, an increase from 1 to 2 bathrooms, an increase from 2 to 3 bedrooms and an increase from 1 to 2 garages. The omitted category is a couple-family with children, who have $100,000 to $149,999 annual household income, aged 18-34. * p<0.1; ** p<0.05; *** p<0.01.

Table 3.5 presents results concerning the estimated willingness to pay differentials for the other main structural housing attributes across household groups. Household structure (especially the presence of children) seems to be the major demand shifter for most dwelling attributes. Not surprisingly, families with children, as expected, showed stronger preferences for homes with larger size, more bedrooms and more bathrooms. Household groups with less than $150,000 annual income were not significantly different in their demand for larger homes with more rooms and garages, while the highest income group were willing to pay significantly more for those homes. We also added the attitudinal factors into these models, but the estimates were not statistically significant. This might suggest that household demand for home size and rooms is not heavily influenced by attitudes, but by the real needs considering household structure and income.

Preferences for single-detached housing

Table 3.6 presents the preference differentials across different households for single-detached housing based on the probit estimates. Couples with children were willing to pay significantly more for single-detached housing, compared to other household types; households with lower income were willing to pay significantly less for such housing, especially when comparing the lowest income group (less than $50,000 annual household income) with the reference group ($100,000 to $149,999). Compared to the local buyers, the GTHA buyers in our sample demonstrated significantly stronger preferences for single-detached housing. In addition, possibly because the single-detached houses in the CTC were quite limited and much smaller and older compared to those outside the CTC, no GTHA buyers bought houses within 1,000 meters from the LRT.

After controlling for the household characteristics, those who rated housing type as a very important factor in their location choices were willing to pay significantly more for single-detached housing. Therefore, the GTHA buyers, the higher income households, and the couples with children showed significantly stronger preferences for single-detached houses, and their particular attitudes (possibly related to the dream of owning a single-detached house) further motivated them to buy single-detached houses in KW. The other characteristics such as education and employment status did not significantly differentiate household preferences for single-detached housing.

Table 3.6: Probit estimates of the demand for single-detached housing

	Dependent variable:
	SINGLE
Couple-family without children	-1.05(0.34)***
Lone-parent family	-1.04(0.57)*
More-persons household	-0.02(0.77)
One-person household	-0.79(0.44)*
Less than \$50,000	-1.61(0.53)***
\$50,000-\$99,999	-0.28(0.33)
\$150,000 and over	0.59(0.42)
Age 35-54	-0.01(0.32)
Age 55+	-0.48(0.46)
GTHA buyers	1.43(0.67)**
Other buyers	-0.56(0.46)
Housing type: 3-very important	2.62(0.30)***
Offset	-1.00
Constant	-0.48(0.42)
Observations	250
Log Likelihood	-213.54
Akaike Inf. Crit.	449.08

Note: The dependent variable is the binary variable $SINGLE$. The standardized price for single-detached housing estimated from the first-stage hedonic, is controlled as an offset in the probit model. The omitted category is a couple-family with children, who is a local homebuyer, with \$100,000 to \$149,999 annual household income, aged 25-34, and also with an attitude of not "very important" for the housing type. Standard errors are reported in parentheses. * p<0.1; ** p<0.05; *** p<0.01.

Preferences for the central-transit corridor

Following Step 2 and Step 3, we first estimated preferences for the LRT access and regressed them on homebuyers' characteristics. The OLS regression model based on equation (3.14) shows that all observed household groups were not significantly different in their willingness to pay for the LRT access. It is possibly because the LRT was still under construction that the proximity to the LRT was not yet valued by most homebuyers during this period. Thus, we did not present that model here.

To estimate the demand for the CTC neighbourhood (In_CTC), we constructed probit

Table 3.7: Probit estimates of the demand for the CTC

	Dependent variable: In_CTC	
	(1)	(2)
Couple without children	0.44(0.28)	0.61(0.34)*
Age: 55 and over	1.48(0.33)***	1.75(0.41)***
LRT access: 2-somewhat important	-	1.48(0.41)***
LRT access: 3-very important	-	1.78(0.53)***
Implicit price of the CTC	-1.00	-1.00
Constant	-1.90(0.22)***	-2.94(0.45)***
Observations	204	204
Log Likelihood	-121.82	-109.32
Akaike Inf. Crit.	249.64	228.64

Note: The dependent variable is the binary variable In_CTC. The standardized implicit price for the CTC estimated from the first-stage hedonic, is controlled as an offset in the probit model based on equation (3.16). Observations become smaller due to the incomplete data of household characteristics. Standard errors are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01.

models. Given our sample size and several household characteristics, we first use the Least Absolute Shrinkage and Selection Operator (LASSO) to reduce the number of variables in the probit models. Four dummy variables were finally selected, as shown in Table 3.7. Model 1 considered two lifecycle variables, and Model 2 added two attitudinal variables. Both took the implicit prices for the CTC estimated from the first-stage hedonic model as an offset in the model specification base on equation (3.16).

To compare the two models, an analysis of deviance was conducted through a Chi-squared test ($p < 0.001$) in ANOVA, and we found that Model 2 significantly improved Model 1. Thus, both lifecycle characteristics (mainly the presence of children and age) and attitudes toward the LRT significantly differentiated households' preferences for the CTC and thus their residential choices in KW. In particular, we find that couples without children and seniors aged 55 and over were willing to pay significantly more for the CTC neighbourhoods. For those who rated the LRT access as an important factor in their location choices, they were willing to pay significantly more for the CTC as well. This conforms with residential self-selection theory, which posits that attitudes toward particular neighbourhoods or lifestyles can also influence residential location choices (Van Wee, 2009). Other characteristics such as income, education level and employment status did not significantly impact households' preferences for the CTC.

Preferences for the locational and neighbourhood attributes

Table 3.8: Willingness to pay for the locational and neighbourhood attributes

	Dependent variable:				
	WTP_LRT	WTP_BUS	WTP_OS	WTP_POP_DENS	WTP_EDU
Couple without children	2,218.8	1,197.8**	561.4	2,265.4*	-6,550.9
Loneparent	3,637.9	658.1	173.7	150.3	5,659.1
More-persons household	2,092.7	1,062.3	398.9	2,355.7	-9,576.5
One-person household	1,105.3	1,105.6	-268.1	-512.4	-11,187.1
Less than $50,000	-2,573.1	1,369.1	314.5	1,265.6	-10,013.0
$50,000 - $99,999	-2,685.5	499.9	531.9	-1,589.6	-5,934.1
$150,000 and over	-8,704.1**	-36.9	914.9*	-2,409.3*	18,213.2***
AGE: 35-54	-344.7	135.8	256.8	-114.3	2,376.3
AGE: 55+	1,891.1	795.5	568.2	4,660.6**	6,988.2
Full-time employed	76.7	1,173.5	-741.6	6,985.7***	-1,357.5
Observations	179	179	181	181	181
R^2	0.06	0.06	0.12	0.12	0.15
Residual Std. Error	12,830	2,874	2,252	6,362	29,371
F Statistic (df = 10)	1.2	1.1	1.3	1.3	1.6*

Note: Each column presents a separate OLS regression. The dependent variables are the estimated willingness to pay for moving from 3000 meters to 1000 meters from the nearest LRT stop, moving from 600 to 300 metres from the nearest bus stop, an increase of open space access from 40 to 60, and an increase of population density from 3000 to 5000, and an increase of post-education rate from 60% to 80% in the neighbourhood. The omitted category is a couple-family with children, who have $100,000 to $149,999 annual household income, aged 18-34. *p<0.1; **p<0.05; ***p<0.01.

Table 3.8 presents results concerning the estimated willingness to pay differentials across household groups with respect to the locational and neighbourhood attributes. For the distance to the LRT, only the wealthiest households significantly preferred living further from the LRT. It is important to note that the measure of distance to the LRT is a confounding factor for distance to urban cores, which might suggest that the highest income group were willing to pay more for the suburban neighbourhoods. Couples without children preferred living in denser neighbourhoods, compared to couples with children. The wealthiest households preferred neighbourhoods with better access to open space, less density, and higher "quality" (recalling that a higher post-secondary education rate is often associated with better neighbourhood "quality"). Senior households preferred the denser areas, and as we argued before seniors had stronger preferences for the CTC area. This suggests that during the survey period, the households aged 55 and over demonstrated stronger preferences for the dense urban cores. In addition, households with full-time employee(s) preferred the denser areas, which might imply that they were willing to pay more for living closer to the employment centres which are often located in denser areas. The attitudinal factors did not significantly influence these neighbourhood preferences. To conclude, preference heterogeneity for the locational and neighbourhood attributes was influenced by household structure and income, similar to that for the structural housing attributes.

3.5 Discussion

Using a unique data set, this study applies a two-stage estimation method and a natural experiment offered by the development of a new light rail transit line to identify how the willingness to pay for various attributes of a house vary with the characteristics of households. Together with results from our study, we discuss the relationship between LRT investment and residential location choices and finally present the paper caveats.

3.5.1 LRT development, property value changes and residential location choices

Many cities , like Kitchener Waterloo, which are experiencing economic and population growth have proposed light-rail transit systems to guide "smart growth" (Handy, 2005). LRT is expected not only to provide better accessibility to public transit and services nearby but to support walkable and compact transit-oriented development. Theoretically (Alonso, 1964) households who value the benefits of LRT accessibility and transit-oriented neighbourhoods will relocate to these area and bid up the prices of land and properties close to LRT stations. Empirically, many studies (Knowles and Ferbrache, 2016; Higgins and Kanaroglou, 2016) have examined the relationship between LRT investment and property value changes, which varies by geographic context, economic and external factors (such as location, transport schemes and amenities nearby). The meta-analysis from Mohammad et al. (2013) concludes that LRT systems in North American cities have less impacts on land and housing prices mainly due to the more car dependent and lower-density developments, compared to cities in East Asia and Europe. In fact, to influence a broad range of people to "give up" the American dream housing (i.e., a single-detached house with a big private yard in the suburbs) and reside in more compact homes in transit-oriented neighbourhoods is a key challenge for mid-sized cities in North America to make LRT investment worth.

Our study finds that the new LRT has relatively limited impacts on the KW housing market during the survey period. Both LRT accessibility and transit-oriented neighbourhoods are not significantly capitalized into detached residential property values in KW. One possible reason is that the LRT has not provided any transportation amenity to the nearby locations during its construction. In contrast, due to the construction, the core areas experienced many road detours and closures of retail stores, which might mitigate the possible positive effects of the LRT on the CTC property values. Further, only 17.6% of the survey respondents presented expectations of a housing price increase with the LRT development. This further validates our conclusion of the LRT's modest impacts on the housing market during its construction phase and justifies our focus on the consumption value of other housing attributes to individual households.

Despite the majority of households in our survey living outside the CTC, a small proportion (8%) of them relocated to the CTC. Through a second-stage demand analysis, we find that two specific groups - 1) couple-families without children and 2) seniors aged 55 and over - showed stronger preferences for living in the CTC area. In addition, survey respondents who express particular preferences toward the LRT were also willing to pay more for the transit-oriented neighbourhoods. This suggests that, in addition to the life-cycle factors (mainly age and the presence of children), some homebuyers self-selected into the CTC area based on their particular preferences toward the LRT. Through a comparison of preferences for the main housing attributes across households, we find that couples with children demanded significantly larger home sizes, more bathrooms, bedrooms, and private yards. This might suggest that their choice decisions were more determined by their basic needs for the housing itself, compared to couples without children. A family with children commented that

> "We had the intention to move to a home near LRT. We are actually now farther away from LRT than if we had stayed at our previous home. Size of the home was the main factor in why we moved with LRT accessibility being 2nd."

Another family with children mentioned that

> "One of the many reasons why we moved to Waterloo region was the LRT and the dream at some point in the future to be able to bike/walk to an LRT station and take GO train to Toronto. Also our kids would be able to move independently via active and public transportation (i.e. Bike/walk + LRT). We bought this house as it was [within the] walking distance (4min) to church and school. Also it's close to trails. "

These comments indicate that some families with children did value the future LRT and had an intention to live closer to the LRT. But possibly because of a lack of home options in the CTC to meet their basic housing needs (for instance home size, child-friendly amenities, and trails), they finally did not purchase homes close to the LRT. Supply to target the market of younger families with school-age children, who need larger homes and child-friendly amenities, is possibly "missing" in the CTC. A recent developer survey study shows that few developers have targeted families with children in their mid- and high-rise apartment projects within the urban cores of the Region between 2011 and 2015, but rather developed to cater to young professionals (singles or couples), seniors and students (Tran, 2016). That study also finds that the target market for families has been neglected in the core areas, but concentrated in the suburbs. As a result, despite preferences for the CTC, families with children have no choice but to move to the suburbs. We argue that the future station-area plans should not only target the smaller-size households but also provide homes and amenities (such as safe open space or playground) for families with children. We expect the CTC area to be a complete community with better access to transit and amenities and a mix of affordable home options for a broad range of households.

3.5.2 Caveats

There are several caveats in this study worth mentioning. For the preference identification method, some might criticize the strong restriction on the quasi-linear utility functional form, and suggest other utility function alternatives such as the Cobb-Douglas function or the Constant Elasticity of Substitution (CES) function. However, this remains an empirical question, and without additional information such as observations from multiple markets, this method using the simple quasi-linear utility function provides a theoretically sounded way to recover the unknown true form through a relatively reasonable assumption. Another caveat lies in our survey sample. In particular, the booming condo market in KW area was not accounted for in our analysis due to the small survey sample.

Chapter 4

Who prefers to live in Transit-Oriented Development area? Evidence from a residential location choice survey in Canada

Who prefers to live in Transit-Oriented Development (TOD) areas? This is a central question for many mid-sized cities that strive to promote TOD and intensification in urban cores. Up-to-date knowledge of different demographics' preferences for TOD and their residential location choice behaviour is important to encourage vibrancy in TOD neighbourhoods. This paper uses a detailed residential location choice survey in Kitchener Waterloo, Canada to identify three household groups: (i) Current TOD households, (ii) Potential TOD households, and (iii) Car-dependent households. Through comparing the three groups, we aim to examine whether they have significantly different demographic profiles, residential preferences, and home choices. Our findings will inform planners, policymakers, and developers of the specific market target in TOD housing projects.

4.1 Introduction

Transit-oriented development (TOD) has evolved as the dominant paradigm of urban growth planning (Papa and Bertolini, 2015). TOD is typically defined as a mixed-use, relatively high-density, and pedestrian-friendly development within a radius of 500-800 m from a transit stop (Cervero, 2007). As a promising tool to restrain urban sprawl (Higgins et al., 2014; Staricco and Brovarone, 2018) and stimulate smart growth (Dittmar et al., 2004), TOD has continued its popularity in Europe (Bertolini et al., 2012), Asia (Lyu et al., 2016) and North America (Cervero, 2004; Curtis et al., 2009).

While literature has presented wide support for TOD from scholars, planners, and policymakers, our understanding of the public's acceptance about the new form of community development is still limited (Tian et al., 2015). Tian et al. (2015) argued that "the acceptance of smart growth, however, has not been as fast as expected" (p. 447); Thomas and Bertolini (2014) pointed out that public support for high densities and public transit is a critical factor of TOD success, after synthesizing 11 TOD cases including Toronto, Rotterdam, and Copenhagen. Burchell et al. (2000) found that due to the "market support for sprawl, [and] the automobile's clinging dominance" (p. 821), the success of smart growth policies is far from assured. There is a need to better understand residents' preferences for TOD, to examine who prefers the new mixed-use and transit-friendly communities and who still adheres to the "American dream" for the big single-family houses in conventional suburban communities.

Several survey studies have attempted to assess residents' preferences for compact, mixed-use TOD. The National Association of Realtors (NAR, 2017) surveyed 3000 adults living in the 50 largest metropolitan areas of the U.S. about their preferred communities through the *2017 National Community and Transportation Preference Survey*. The survey found that 53% of respondents prefer walkable, mixed-use communities, while 47% prefer conventional suburban communities. Despite an evident desire for compact and walkable communities (Brookfield, 2017), the majority of residents continue to live in detached-homes and value proximity to highways (NAR, 2017). This conflict indicates a possible mismatch between the preferred neighbourhoods and the actual neighbourhoods (Kumar et al., 2018; Myers and Gearin, 2001). The NAR (2017) survey confirms that one in five who live in a detached home currently would prefer to live in an attached home in a walkable community with a shorter commute. Therefore, there is a need to identify the potential TOD residents and examine the trade-offs they have made in their actual choices, to inform policymakers, lessen the level of residential dissonance and satisfy their housing needs.

In this article, we aim to paint a picture of residential location choices of different households to answer the following questions: 1) Who is currently living in TOD areas? 2) Who shows preferences for TOD but is currently living outside of TOD areas? 3) Who still prefers living in car-oriented suburban areas? We draw on data from a residential location choice survey that was conducted in 2017 to explore households' location choice behaviours during the construction of a new light-rail transit (LRT) line in Kitchener Waterloo (KW). Our data include not only respondents' stated importance of housing and neighbourhood attributes, but also their actual choices and sociodemographic information. We first employ the survey data and conduct latent class analysis (LCA) to identify household groups with different preferences for TOD communities, followed by comparisons of their demographic characteristics, residential preferences and actual location choices.

This study makes contributions in three aspects. Firstly, it advances our understanding of residents' preferences for TOD and other housing attributes in mid-sized cities. Many studies have focused on TODs in large metropolitan areas, but few have provided insight into mid-size cities such as Kitchener and Waterloo, which have seen pervasive core-area decline and extensive decentralization (Bunting et al., 2007). In such municipalities, it might be more challenging to increase transit use and attract significant numbers of residents to TOD areas, since attitudes toward auto-oriented suburbia may remain positive, and travel times are relatively short. Secondly, the role of TOD in triggering gentrification has become a major policy concern (Baker and Lee, 2019; Revington, 2015). Through analyzing demographic profiles, we can figure out whether TOD development has exacerbated social segregation by income. Lastly, we can inform policymakers and developers of any potentially overlooked market targets of housing in TOD areas. This article starts with a literature review on residential location choices and preferences for TOD in Section 2.

Section 3 describes survey data and methods. Findings and discussions are presented in Sections 4 and 5.

4.2 Literature Review

4.2.1 TOD and residential location choices

What are people looking for in TOD neighbourhoods? According to Ewing and Cervero (2010), the built environment of TOD generally refers to five "*D*" elements: *Distance to transit, Design, Destination accessibility, Density, and Diversity.* Several studies have reported that TOD residents generally have a clear preference for walkable neighbourhoods (Brookfield, 2017; Levine and Frank, 2007; Noland et al., 2017), alongside preferences for better access to public transit (Lund, 2006), better street design and connectivity (Song and Knaap, 2003), and better access to public open space (Olaru et al., 2011) and nearby shops/services (also known as residential-commercial land-use mix) (Guo and Bhat, 2007). In addition, some residents having a strong cultural preference for cycling, such as people in the Netherlands (Pojani and Stead, 2015), prefer to have better access to transit stations by cycling and better bike parking facilities in station areas (Puello and Geurs, 2015). With respect to density, few people seem to value high neighbourhood density per se (Dunse et al., 2013). Bramley and Power (2009) analyzed the Survey of English Housing and found that people living in more dense forms are more likely to be dissatisfied with their neighbourhoods, with similar findings from a survey for the city of Leeds, UK (Evans and Unsworth, 2012). These findings suggest that among the five elements of TOD, most seem to be favoured by residents except for density.

4.2.2 TOD and socio-demographics

Who prefers to live in TOD neighbourhoods? Studies have shown that socio-economic factors such as age, income, and children in households often influence neighbourhood preferences (De Vos et al., 2016). The Millennial generation (born in 1981-1996) prefers smaller homes in more walkable communities with a shorter commute and better access to shops and restaurants in more central locations, while the majority of Generation X (born in 1961-1980) are more committed to suburban living with larger lots (NAR, 2017). Older generations, especially retirees looking to downsize, also prefer compact housing in neighbourhoods with better transit accessibility and walkability (Tian et al., 2015). Studies also show that current TODs typically attract smaller households, such as singles and couples without kids (Arrington and Cervero, 2008; Dittmar et al., 2004; Noland et al., 2017). Families with fewer school-age children are more likely to live in compact small housing in TOD areas (Liao et al., 2015).

Lower-income households who have lower car ownership seem to prefer communities with better access to shops and services by active transportation mode (bike/walk/public transit) (Lund, 2006). Based on the results from 2 large-scale surveys in California and four other southwestern states, Lewis and Baldassare (2010) found significant support for compact development from low-income residents, renters, and minorities.

In addition to socio-demographic factors, people with varying lifestyles tend to hold different residential preferences and reside in different neighbourhoods (van Acker et al., 2011). For instance, some people, who hold a strong pro-environmental attitudes, often choose TOD communities to satisfy their travel preferences and lifestyle (Cao et al., 2009; van Wee, 2009; Walker and Li, 2007); and some people self-select into TODs to fulfill a more urban and transit-oriented lifestyle (Noland et al., 2017), where they enjoy walkable and mixed-use environments with better access to mass transit and nonmotorized transportation (Cervero, 2004).

What demographics do current TODs target? Some older TOD projects, such as the transit-village housing in Oakland's Fruitvale, built affordable housing units to serve lower-income households (Jacobson and Forsyth, 2008). Some TODs in metros, such as Chicago, Washington D.C and Vancouver, built more expensive and upscale housing (Arrington and Cervero, 2008). It is important to note that the current trend of new TOD housing is generally to cater to higher-income residents (Arrington and Cervero, 2008). Enhanced desirability and accessibility in station areas often escalate the prices of properties nearby, which might force the lower-income out of TOD areas and induce gentrification (Dawkins and Moeckel, 2016; Revington, 2015). This indicates a possible undersupply of TODs for those in most need of transit-rich neighbourhoods (Levine and Frank, 2007).

4.3 Data and Methods

4.3.1 Research context

The KW region is located in southern Ontario, Canada (See Figure 2.3). The two municipalities Kitchener and Waterloo, as well as Cambridge and the surrounding townships, collectively make up the Region of Waterloo, with a population of 535,154 in 2016 (Statistics Canada, 2017). The region is a fast growing mid-size urban area in Canada, especially with its booming high-tech sector. Under the Provincial Policy Statement (2005) and the Growth Plan for the Greater Golden Horseshoe (2006), the region proposed to manage growth through a new LRT investment and implement intensification objectives through transit-oriented development (Region of Waterloo, 2019). The proposal was approved in 2011, and the Phase One line (a 19-km LRT system connecting the two cities of Kitchener and Waterloo) started construction in 2014 and began its operation in 2019. The Phase Two line is in planning process and will extend to Cambridge. This research focuses on Kitchener and Waterloo, where the Phase One LRT line goes through.

Figure 2.3 illustrates the study area and shows the Central Transit Corridor (CTC), which is a buffer zone around the LRT line where people can reach the LRT stations within an about 10-min walk. The CTC has higher walkability, higher population density, better access to public transit, higher employment access, greater land-use mix, and slightly lower open space access than the areas outside the CTC (Region of Waterloo, 2019). The CTC is also called “TOD areas” or “TOD neighbourhoods” in this paper.

4.3.2 Data collection

We designed a comprehensive housing survey to explore residential location choice behaviours of both home buyers and sellers in KW during the new LRT construction (See the survey questionnaire in the Appendix A7). To find relevant respondents, we requested an address list from Canada Post, which identified 5185 likely home movers who either bought or sold a home in KW between June 2015 and April 2017. Survey invitations were mailed out to those home movers[1], and 357 buyers (around 10% response rate) and 149 sellers completed the survey via an online survey link or paper survey from June to September of 2017. This paper focuses on results from the homebuyers (n = 357) who responded to questions of i) home buying motivations and characteristics of the homes they bought, ii) the home buying process, iii) stated importance of housing attributes in residential location choices, iv) attitudes towards the new LRT, and v) household characteristics and travel behaviours.

This paper first draws on responses from part iii of the survey to explore household preferences for TOD neighbourhoods. Figure 4.1 shows the survey questions (each with three options) regarding the stated importance of attributes in neighbourhood selection, which include physical attributes and accessibility-related attributes. Following the five "*D*" aspects of TOD (Ewing and Cervero, 2010), we extract responses from eight questions which are closely related to TOD elements. To be specific, we use responses from the stated importance of 1) LRT access and 2) bus access to represent preferences for *Distance to transit*, use the stated importance of 3) walkable and 4) bicycle-friendly environment to reflect preferences for neighbourhood *Design*, use the stated importance of 5) accessibility to public open space and 6) accessibility to urban centres to represent preferences for *Destination accessibility*, use the stated importance of 7) land use mix to indicate preferences for *Diversity*, and use the stated importance of 8) density of housing to reflect preferences for *Density*.

In the following sections, we present how we use responses from these eight aspects to identify groups with heterogeneous preferences for TOD through a latent class analysis (LCA). We then compare these groups in terms of demographic profiles, travel mode, moving motivations, residential preferences and their actual location choices based on survey responses. Chi-squared tests are conducted for comparisons between groups.

[1] Canada Post filtered out the addresses of home movers based on our request, but we did not know to what addresses those surveys were mailed. Canada Post directly sent the address list to the mailing service provider at the University of Waterloo who helped us mail out the survey invitations.

Q45-1. Physical neighbourhood

Please rate the importance to your current neighbourhood selection

	Not important	Somewhat important	Very important
Density of housing	1	2	3
Land use mix *	1	2	3
Easy to walk	1	2	3
Bicycle-friendly environment	1	2	3
Traffic noise	1	2	3

* Land use mix: e.g., mix of residential, retail, commercial or employment centre.

Q45-3. Accessibility

Please rate the importance to your current neighbourhood selection

	Not important	Somewhat important	Very important
Commuting time	1	2	3
Commuting cost	1	2	3
Accessibility to… - school	1	2	3
- workplace	1	2	3
- retail and services	1	2	3
- public open space	1	2	3
- urban center	1	2	3
- bus stops			
- future LRT stops	1	2	3
Distance to… - previous neighbourhood	1	2	3
- your family/friends	1	2	3
- highway exits	1	2	3

Figure 4.1: Survey questions of stated importance of physical neighbourhood and accessibility-related attributes

4.3.3 Latent class modelling

Latent class analysis (LCA) is a modelling approach often used to identify clusters from the population based on observed response variables (Morey et al., 2008; Masyn, 2013). The main advantage of the LCA compared to other clustering methods is that LCA is in fact a Finite Mixture Model and derives clusters using a probabilistic model that describes the data distribution (Hagenaars and McCutcheon, 2002), instead of deriving clusters by imposing arbitrary distance measures. The model assumes that the population heterogeneity among a set of response variables results from the existence of latent classes. In our case, we assume that the population consists of different preference classes with respect to TOD neighbourhoods, and each household's preference class is unobserved or latent. We observe each household's set of answers to the eight stated preference questions and expect that households from the same preference class answer similarly. The latent clustering structure (c) denoting underlying preference classes for TOD is assumed to be represented by a set of attitudinal constructs ($u_1 \ldots u_8$) which are the stated importance of TOD features in home decisions, as delineated in Figure 4.2.

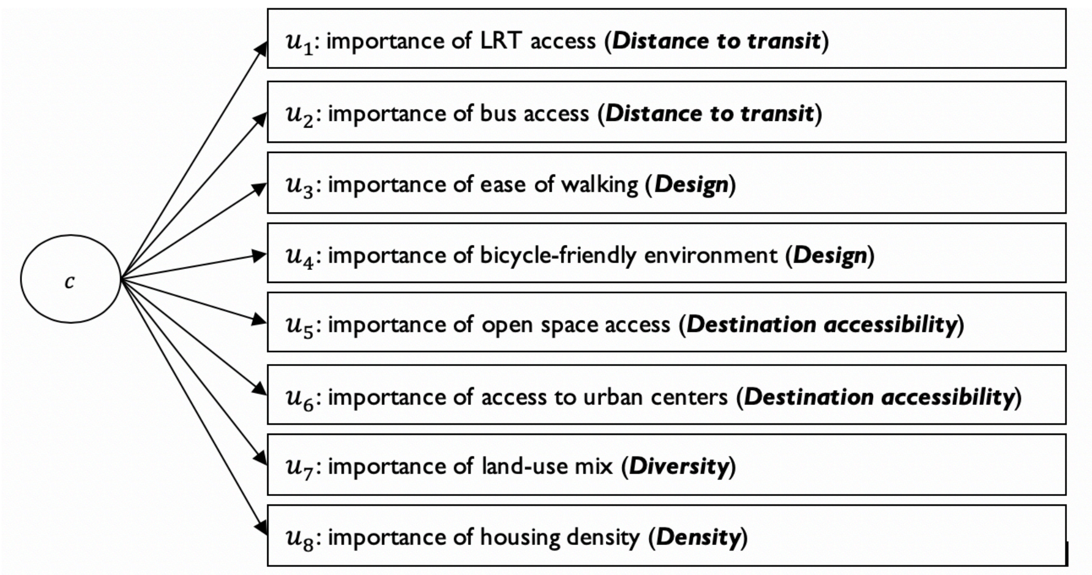


Figure 4.2: Latent class model structure

Note: $u_1 \ldots u_8$ denote observed categorical variables; c denotes the latent class variable; arrow paths represent direct relationships.

The latent class measurement model is as follows,

$$Pr(u_{1i}, u_{2i}, \ldots, u_{Mi}) = \sum_{k=1}^{K} [\pi_k \cdot Pr(u_{1i}, u_{2i}, \ldots, u_{Mi} | c_i = k)] \tag{4.1}$$

$$= \sum_{k=1}^{K} \left[\pi_k \prod_{m=1}^{M} Pr(u_{mi} | c_i = k) \right] \tag{4.2}$$

where we have M categorical latent class indicators, $u_1, \ldots, u_M$ ($M = 8$), and u_{mi} is the observed response to question m for participant i. We assume an underlying unordered categorical latent class variable, denoted by c, with K classes where $c_i = k$ when participant i belongs to class k. π_k denotes the proportion of participants in Class k, i.e., $Pr(c = k)$. $Pr(u_{mi}|c_i = k)$ represents the probability of participant i gives a particular answer to the question u_m conditional on being a member of class k. By maximizing the likelihood function for a K class model with N participants as below,

$$
\begin{aligned}
lnL &= \sum_{i=1}^{N} ln[Pr(u_{1i}, u_{2i}, \ldots, u_{Mi})] && (4.3)\\
&= \sum_{i=1}^{N} ln \sum_{k=1}^{K} \left[\pi_k \prod_{m=1}^{M} Pr(u_{mi}|c_i = k) \right] && (4.4)
\end{aligned}
$$

We can estimate the class-specific response probability $\hat{Pr}(u_{mi}|c_i = k)$ that an individual from a certain class gives a particular answer to that attitudinal question, and also estimate $\hat{\pi}_k$ to get the distribution of the latent class variable (Masyn, 2013)

We use the depmixS4 (Visser and Speekenbrink, 2010) package in R to conduct the latent class analysis through assuming one class, then two classes, three classes, and then four classes. We assess whether adding a class significantly increases the explanatory power of the model through checking the AIC (Akaike information criterion) and BIC (Bayesian information criterion) statistics, which are essentially log-likelihood scores corrected by sample size and number of parameters (Morey et al., 2008). The lower the statistics, the better the model fit (Rid and Profeta, 2011).

After deciding the number of classes, we estimate the class-specific response probabilities as well as the odds ratio between classes as below,

$$
OR_{m|jk} = \frac{\dfrac{\hat{Pr}(u_m|c = j)}{1 - \hat{Pr}(u_m|c = j)}}{\dfrac{\hat{Pr}(u_m|c = k)}{1 - \hat{Pr}(u_m|c = k)}} \tag{4.5}
$$

which is the ratio of the odds of giving a particular answer to question m by members in Class j to the odds in Class k. A large $OR_{m|jk} > 5$ or a small $OR_{m|jk} < 0.2$ indicates a high degree of class separation with respect to the particular response to question m

(Masyn, 2013).

4.4 Findings

In this section, we start with the latent class modelling results and proceed to details of demographic profiles and residential preferences of different classes.

4.4.1 Identifying preference classes

Table 4.1: Statistics for latent class models

Number of classes	Parameters	Log likelihood	AIC	BIC
1	26	-2756.158	5544.315	5605.858
2	54	-2581.828	5233.656	**5368.282**
3	84	-2530.158	5172.316	5387.716
4	116	-2494.363	5146.726	5450.595

Note: AIC means Akaike information criterion; BIC means Bayesian information criterion; The model with 2 classes has the lowest BIC value (as highlighted in the table).

We incrementally built the 1-class, 2-class, 3-class, and 4-class models to inquire into household groups (or "latent classes") with different preferences for TOD. Table 4.1 shows statistics of the four models, where the AIC statistic decreases as the latent class number increases, and the largest difference is seen between the 1-class model and the 2-class model. The BIC statistic decreases from the 1-class model to the 2-class model and then increases as more classes are added. Therefore, the 2-class model most effectively captures the latent clustering structure of preferences especially with the lowest BIC value and shows significant improvement over the 1-class model.

Table 4.2: LCA results with estimated class-specific response probabilities and odds ratios

TOD features	Response	Class 1 (41.5%)	Class 2 (58.5%)	OR_12
1. LRT access	1 - Not important	0.253	0.826	0.07*
	2 - Somewhat important	0.491	0.174	4.58
	3 - Very important	0.257	0.000	345.55*
2. Bus access	1 - Not important	0.151	0.780	0.05*
	2 - Somewhat important	0.487	0.190	4.05
	3 - Very important	0.362	0.030	18.35*
3. Ease of walking	1 - Not important	0.019	0.128	0.13*
	2 - Somewhat important	0.183	0.531	0.20
	3 - Very important	0.797	0.342	7.55*
4. Ease of cycling	1 - Not important	0.081	0.386	0.14*
	2 - Somewhat important	0.292	0.461	0.48
	3 - Very important	0.627	0.153	9.31*
5. Open space access	1 - Not important	0.000	0.224	0.00*
	2 - Somewhat important	0.330	0.591	0.34
	3 - Very important	0.670	0.185	8.94*
6. Access to urban centres	1 - Not important	0.090	0.368	0.17*
	2 - Somewhat important	0.452	0.566	0.63
	3 - Very important	0.458	0.066	11.96*
7. Density of housing	1 - Not important	0.082	0.101	0.80
	2 - Somewhat important	0.526	0.547	0.92
	3 - Very important	0.392	0.352	1.19
8. Land use mix	1 - Not important	0.046	0.159	0.26
	2 - Somewhat important	0.530	0.579	0.82
	3 - Very important	0.423	0.262	2.06

Note: * Odds ratios>5 or <0.2, and they are highlighted to indicate a high degree of preference class separation.

Table 4.2 presents the estimated probabilities of specific responses given the class membership derived from the preferred 2-class model, along with the odds ratios of Class 1 vs. Class 2 calculated based on Equation 4.5. Results show that Class 1 is well separated from Class 2 by the stated preferences for all the main TOD features, except for density and land-use mix. For the importance of LRT access, the probability of a Class 1 member giving a response "1-not important" is 0.253, significantly lower than that of a Class 2 member which is 0.826. Contrarily, the probability of a Class 1 member giving a response "2-Somewhat important" or "3-very important" to the importance of LRT access (0.747) is significantly higher than that of a Class 2 member (0.174). Thus, we can conclude that Class 1 members are more likely to consider the future LRT access as an important factor when deciding where to move compared to Class 2 members. Similarly, Class 1 members are more likely to consider the bus access as an important factor in neighbourhood selection compared to Class 2 members.

With respect to the design and regional accessibility measures, the table shows that members of Class 1 have a much higher probability to rate "3-very important" to ease of walking (0.797), ease of cycling (0.627), access to public open space (0.67), and access to urban centres (0.458) in neighbourhood selection, compared to being a Class 2 member with probabilities being 0.342, 0.153, 0.185, and 0.066, respectively. Regarding housing density and land-use mix, both class members are likely to regard them as somewhat important factors, and the probabilities are not significantly different between the two classes.

The latent class analysis confirms that there are two significant clusters/classes of households who exhibit different preferences for TOD. All respondents are probabilistically in class 1 or 2, with their membership determined by the highest probability.[2] Class 1 represents a group with positive preferences for TOD features. In particular, they value transit accessibility, walkability, bicycle-friendliness, access to open space amenities and urban centres in the neighbourhoods. Class 2 represents a group with a lower preference for TOD but favouring a car-oriented lifestyle (Please also see the justification in the following travel mode section). Based on the LCA results, 41.5% (n=148) of the homebuyers (n=357) are estimated to be in Class 1, while 58.5% (n=209) are estimated to be in Class 2. This indicates a large proportion of residents in KW desire to live a TOD lifestyle.

In our survey, respondents also reported their property addresses, which enable us to identify who currently lives in TOD neighbourhoods (i.e., within the CTC) and who lives outside TOD. Combining with the LCA results, we finally classify our sample into

[2] See the estimated probabilities for each survey respondent in Appendix A-6.

three groups, (i) Current TOD households (8.8%), who live in TOD; (ii) Potential TOD households (36.2%), who live outside TOD but hold preferences for TOD (belonging to Class 1); and (iii) Car-dependent households (55%), who live outside TOD and hold less preferences for TOD (belonging to Class 2). The three groups are illustrated in Figure 4.3.[3]

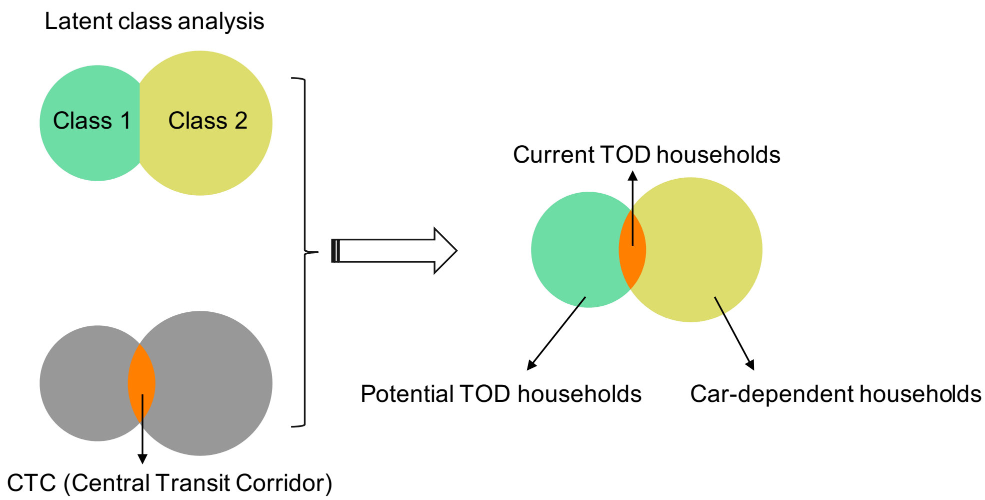


Figure 4.3: Three groups of the total survey sample

The spatial distributions of the three groups are shown in Figure 4.4. The current TOD households in our survey live in the CTC neighbourhoods, and most are close to the two city centres. It should also be noted that, based on our classification results, 63% of the current TOD households belong to Class 1, while 37% belong to Class 2 with a lower TOD preference. To explore why those households who purchased homes in CTC but had no strong preferences for TOD, we checked their moving motivations, which suggests that "the price was much lower than comparable units elsewhere", "established neighbourhood", and "potential value increase" are the main reasons. Thus, the observed choices in the CTC have no direct association with preferences for TOD. The Potential TOD households are observed more in Waterloo and closer to the LRT, while the Car-dependent households are observed more in Kitchener and far from the LRT.

[3]Density plots of class probabilities for the three groups are attached in Appendix A-9

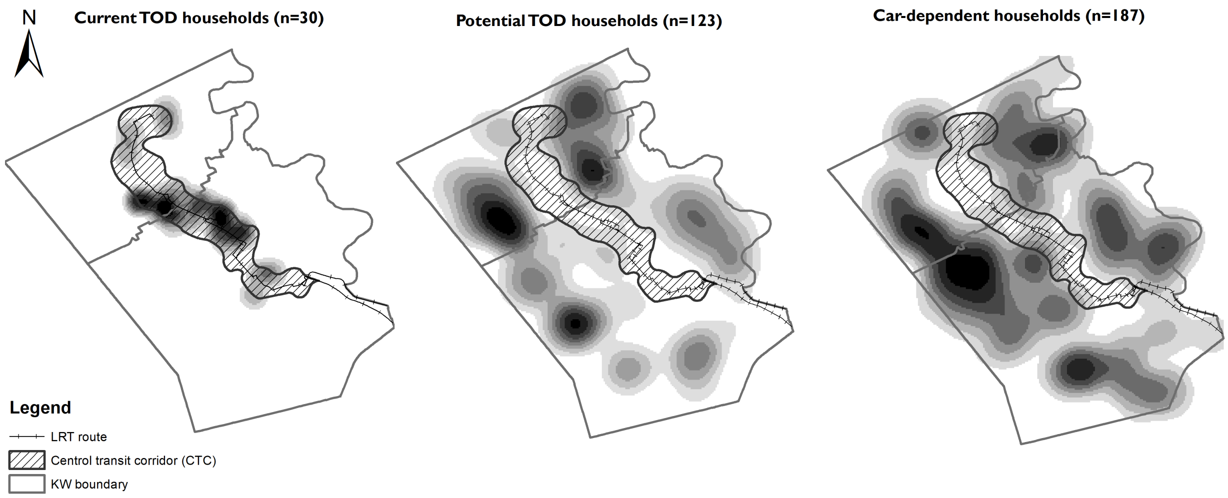


Figure 4.4: Spatial distributions of the three groups
Note: Since we keep the homebuyers' addresses strictly confidential, we show the spatial distributions in terms of kernel densities of the observations. Source: Residential location choice survey in Kitchener Waterloo, Canada, 2017

4.4.2 Travel mode choice

To justify our classification, we look at the car ownership and travel mode choices of the three groups. As shown in Figure 4.5-a, the majority (57%) of the potential TOD households own 1 car or less, while the majority in the total sample owns 2 or more cars (58%); 72% of the Car-dependent households owns 2 or more cars, which is significantly higher than that in the total sample ($\chi^2 = 17.1, p = 0.000$). It should be noted that even the households who currently live in TOD areas own at least one car at home for accommodating their moving activities.

Figure 4.5-b shows the travel mode choice of each group including all the household members. Driving is still the major mode choice for all groups, but the Current TOD households have a significantly higher proportion of people cycling compared to the total sample($\chi^2 = 26.2, p = 0.000$). When comparing the Potential TOD households with the Car-dependent households, we find significant difference in the travel mode choice between the two groups ($\chi^2 = 15.82, p = 0.007$). The Potential TOD group significantly drives less and takes more active transportation modes than the Car-dependent group. These findings further validate our classification.

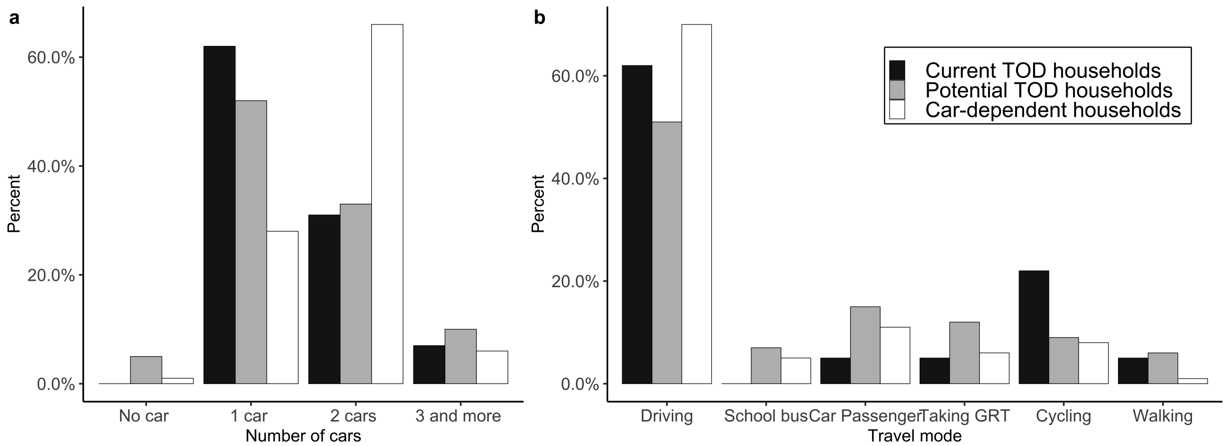


Figure 4.5: Car ownership and travel mode choice of the three groups
Source: Residential location choice survey in Kitchener Waterloo, Canada, 2017.

4.4.3 Demographic profiles

Table 4.3 summarizes the demographic profiles of the three groups, as well as the total sample. Household type and age are significantly different between the Current TOD households and the total sample. A higher proportion of couples without children (55%) and one-person households (21%) have purchased homes in TOD; a higher proportion of seniors aged 55 and over (31%) purchased in TOD as well. However, a higher proportion of younger families (aged 25-54) with children relocated outside TOD, compared to the Current TOD households. When comparing the Potential TOD households and the Car-dependent households, we find no significant differences in all demographic characteristics. When looking closely at the Potential TOD households, the largest proportion is couples with children, aged 25-34, with $50,000-$99,999 annual income.

4.4.4 Moving motivation

Figure 4.6 displays the moving motivations of each group. Seeking better environmental quality and expecting market prices to go up and for the purpose of investment are the common top motivations of moving for all the three groups. The Current TOD households stand out in the LRT-related motivations, 1) expecting price increase due to LRT, and 2) better access to the future LRT stops. Downsizing is another more important factor for the Current TOD households, compared to the other groups. Upsizing is a much more

Table 4.3: Demographic profiles of three groups

	Total sample	C-TOD	P-TOD	COD
Respondents	n = 340 (100%)	n = 30 (8.8%)	n = 123 (36.2%)	n = 187 (55%)
Household type				
Couple with children	40%	17%	41%	42%
Couple without children	40%	55%	34%	40%
Lone-parent family	6%	7%	8%	4%
One person household	12%	21%	14%	11%
Other households	2%	0%	2%	3%
Chi-squared test		χ^2= 7.86	χ^2=2.67	χ^2= 1.39
		p = 0.09	p = 0.62	p = 0.84
Household-head age				
Age 18-24	3%	10%	1%	3%
Age 25-34	51%	38%	48%	55%
Age 35-54	37%	21%	42%	35%
Age 55+	10%	31%	9%	7%
Chi-squared test		χ^2 = 22.8	χ^2 =2.58	χ^2 = 1.92
		p = 0.002	p = 0.45	p = 0.57
Household income				
Less than $50,000	11%	17%	14%	8%
$50,000 - $99,999	42%	41%	42%	41%
$100,000 -$149,999	30%	24%	31%	32%
$150,000 and over	17%	17%	13%	19%
Chi-squared test		χ^2= 1.52,	χ^2=2.31	χ^2= 1.76
		p = 0.68	p = 0.51	p = 0.62

Note: C-TOD is the abbreviation of the Current TOD households; P-TOD denotes the Potential TOD households; and COD represents the Car-oriented households. The Chi-squared tests are conducted to compare the proportion of household demographic in each group with the proportion in the total sample. The significant values are highlighted. Source: Residential location choice survey in Kitchener Waterloo, Canada, 2017.

important factor for both the Potential TOD households and Car-dependent households.

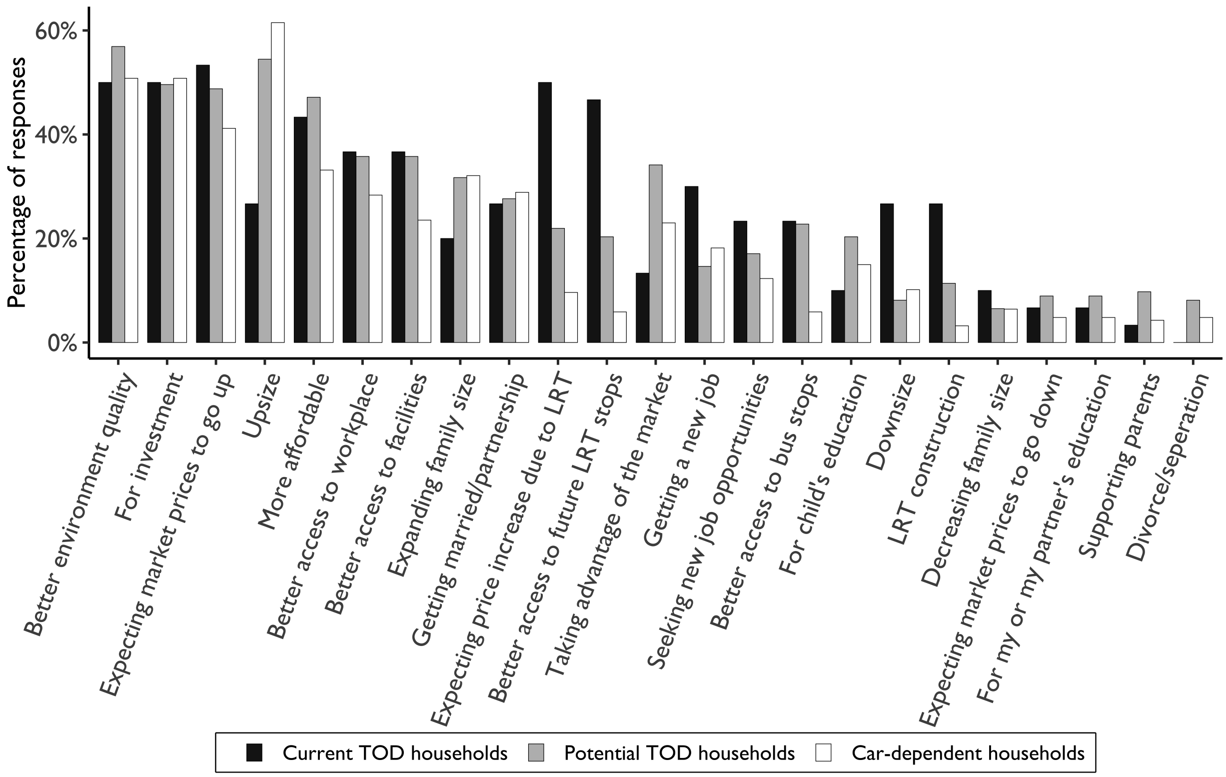


Figure 4.6: Moving motivations

Source: Residential location choice survey in Kitchener Waterloo, Canada, 2017.

4.4.5 Stated preferences for housing attributes

In this section, we examine which housing attributes are important to different household groups. First, we show the preferences for TOD features in Figure 4.7.[4] The Current and Potential TOD households have a stronger preference for a walkable environment, followed by that with ease of cycling, access to open space and urban centres. LRT access and bus access are not very important factors considered in their location choices; however, they are still significantly more important to those favouring TOD than those living in a car-dependent lifestyle.

[4]Note that Table 4.2 has shown the estimated response probabilities of the two classes. Here we report the responses of the three groups in order to better illustrate the preference heterogeneity.

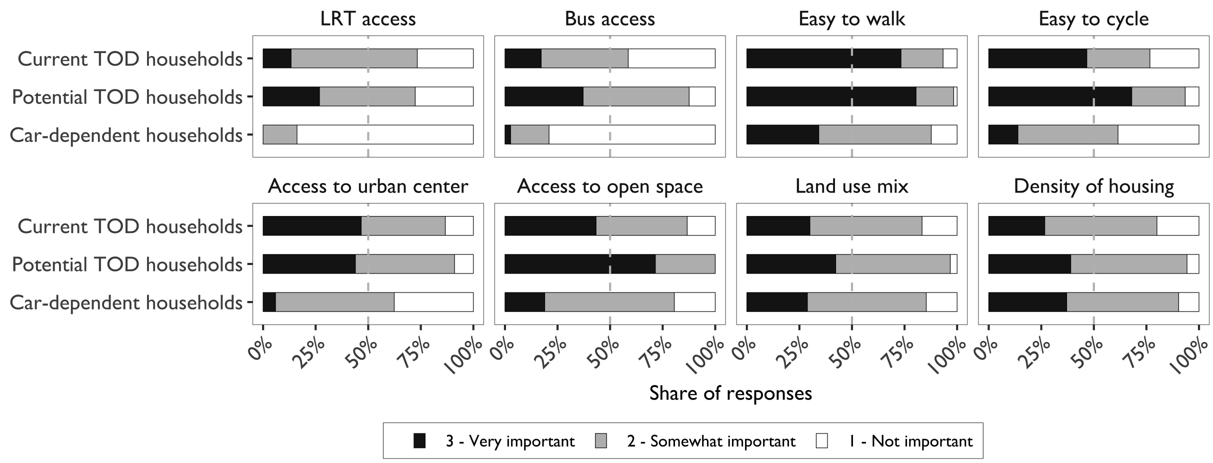


Figure 4.7: Stated preference for TOD features

Note: the figure shows the share of responses for each TOD feature's importance level by each group; Chi-square tests show that LRT access, bus access, ease to walk, ease to cycle, access to urban center, and access to open space are significantly (at the 0.001 level) more important for both current and potential TOD households, than the car-dependent households.

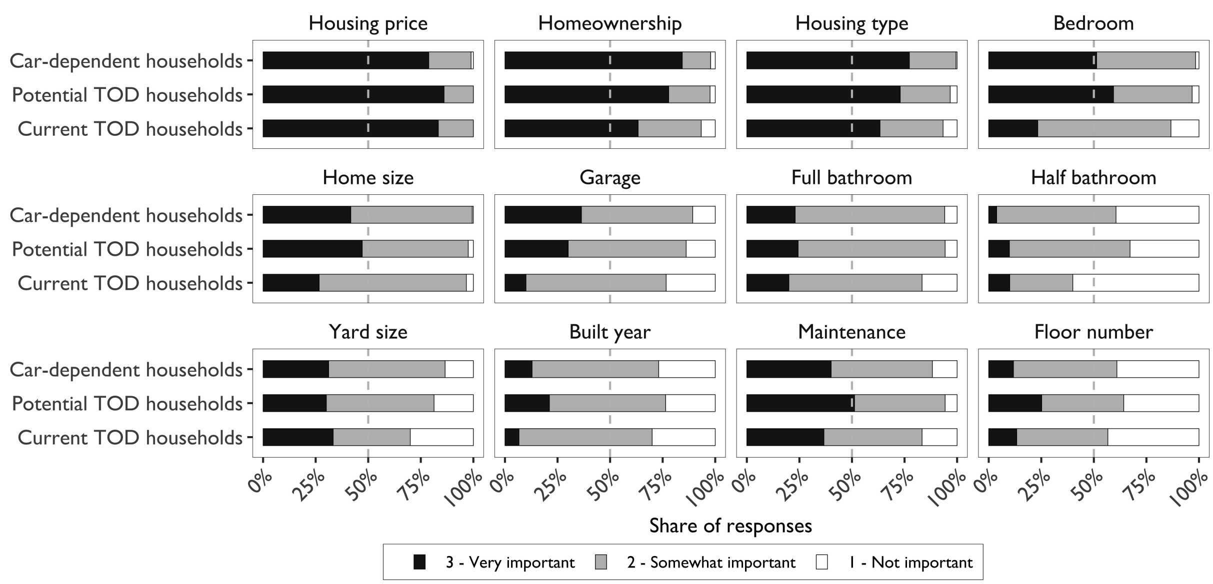


Figure 4.8: Stated preferences for structural housing features

Note: the figure shows the share of responses for each housing feature's importance level by each group; Chi-square tests show that bedroom, home size and garage are significantly (at the 0.001 level) more important for the potential TOD households and the car-dependent households than the current TOD households.

Figure 4.8 shows that housing price, housing type, and homeownership are the three most important factors of residential location choices for all three groups. It also shows that home size, the number of bedrooms and garages are more important to the Potential TOD households than the Current TOD households. This suggests Potential TOD households, primarily families with children, make their choice decisions to meet their needs for larger housing space, more bedrooms and garages. Given that current TOD neighbourhoods in KW are mainly smaller units and new condo developments around stations, which rarely target families with kids (Tran, 2016), potential TOD buyers have no choice but homes far from LRT. This indicates a potential undersupply of housing units in TOD areas for larger households.

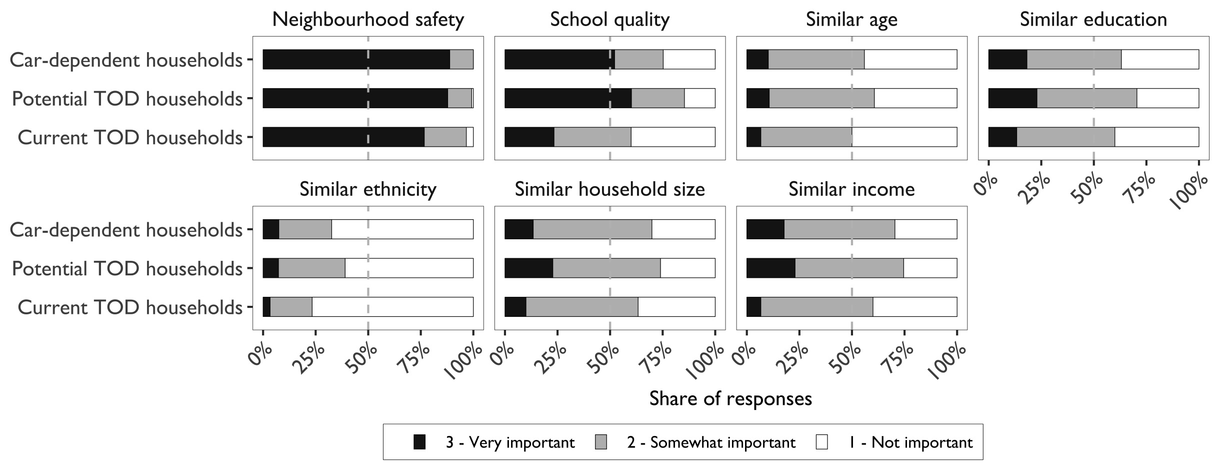


Figure 4.9: Stated preferences for socio-demographic characteristics in neighbourhoods
Note: Chi-square tests shows that school quality is significantly more important to both potential and car-dependent households, compared to the current TOD households at the 0.001 level.

Figure 4.9 shows that neighbourhood safety and school quality are more important than the other socio-demographic characteristics in the neighbourhoods for all three groups. School quality is statistically more important for the Potential TOD households than the Current TOD households. In addition, most households in all groups prefer inclusive neighbourhoods with mixed ethnicity, income, age, education, and household compositions.

4.4.6 Residential location choices

Table 4.4 summarizes the key structural attributes in homes that the three groups have chosen. Housing type, housing size, number of bedrooms, and garages are significantly different between the Current TOD households and the total sample. A higher proportion of Current TOD households have bought high-rise apartments (20%), smaller size with less than 1499 sqft (73%), 1-2 bedrooms (47%), and no garage (33%), compared to that of the

Table 4.4: Residential location choices of three groups

	Total sample	C-TOD	P-TOD	COD
All respondents	$n = 340$	$n = 30$	$n = 123$	$n = 187$
Housing type				
Single-detached house	72%	60%	72%	75%
Semi-detached house	7%	3%	7%	6%
Townhouse/row house	16%	13%	18%	16%
Apartment < 5 storeys	1%	3%	1%	1%
Apartment >= 5 storeys	4%	20%	2%	2%
Chi-squared test		χ^2= 21.8	χ^2=1.39	χ^2= 1.97
		p = 0.005	p = 0.85	p = 0.72
Housing size				
Less than 1000 sqft	4%	13%	3%	3%
1000 – 1499 sqft	40%	60%	38%	40%
1500 – 1999 sqft	30%	13%	35%	30%
2000 – 2499 sqft	18%	7%	17%	18%
2500 – 2999 sqft	5%	0%	5%	6%
More than 2999 sqft	3%	7%	2%	4%
Chi-squared test		χ^2= 18.2,	χ^2=1.90	χ^2= 1.07
		p = 0.004	p = 0.86	p = 0.95
Number of bedrooms				
1-2 bedrooms	13%	47%	12%	8%
3 bedrooms	57%	40%	61%	59%
4 bedrooms	25%	7%	24%	27%
More than 4 bedrooms	5%	7%	3%	5%
Chi-squared test		χ^2= 30.6	χ^2=1.27	χ^2= 4.71
		p = **0.000**	p = 0.73	p = 0.19
Garage				
No garage	10%	33%	7%	9%
1 garage	63%	59%	69%	60%
2 garages	26%	7%	24%	30%
More than 2 garages	1%	0%	1%	1%
Chi-squared test		χ^2= 17.7	χ^2=2.12	χ^2= 2.34
		p = 0.008	p = 0.54	p = 0.50
Number of full bathrooms				
1 full bathroom	32%	40%	32%	32%
2 full bathrooms	50%	50%	52%	47%
More than 2 full bathrooms	18%	10%	16%	21%
Chi-squared test		χ^2= 1.75	χ^2=0.46	χ^2= 0.85
		p = 0.44	p = 0.79	p = 0.64

Note: C-TOD is the abbreviation of the Current TOD households; P-TOD denotes the Potential TOD households; and COD represents the Car-oriented households.The significant values from chi-square tests are highlighted.

total sample (4%, 44%, 13%, and 10%, respectively).

When looking closely at the home choices of the Potential TOD households, we find no significant differences with choices of the Car-dependent households. Most households with latent preferences for TOD bought single-detached homes (72%), 1000-1999 sqft in size (73%), 3-4 bedrooms (85%), and 1 garage (67%), which are not available in current TOD areas of KW.

4.5 Discussion

4.5.1 Current TOD households

Through analyzing the housing survey data in Kitchener Waterloo, we find that single adults, childless couples, and seniors aged 55 and over are likely to have purchased more homes in the TOD neighbourhoods. For household income distribution, the current TOD households are not significantly different from the total sample. However, it is not sufficient to conclude that TODs in KW have not exacerbated gentrification without longitudinal analysis of neighbourhood changes, especially given our limited sample size in TOD areas (with only 8.8% buying in TOD). Therefore, we conclude that residential segregation during the LRT construction is more related to differences in household structure and age based on our survey. In addition, downsizing is one important factor for the Current TOD households to reside in TOD.

4.5.2 Missing target in TOD housing

It is encouraging to see that a significant proportion (36.2%) of households in our sample have latent preferences for TOD, which signifies the potential demand for TOD housing. These Potential TOD households are mainly young families (aged 25-34) with school-age children and \$50,000-\$99,999 annual income . We argue that they represent the missing target of the current housing projects in TOD neighbourhoods.

These Potential TOD households own fewer cars and drive less for commuting than the

Car-dependent households. They have preferences for a TOD lifestyle, in which people walk and cycle more in daily life, take more transit, and embrace better access to public open space and urban facilities. Neighbourhood safety and school quality are also important to these potential TOD residents. We recommend that TODs provide safer environments with enhanced pedestrian and cycling network, address perceptions of school quality and provide better access to urban amenities (such as community centres and parks). Better design and improved destination accessibility in station areas will enable TODs to attract the missing group and accommodate a wide range of residents.

Additionally, upsizing is one important motivation for the relocation of the Potential TOD households. They prefer larger homes and more bedrooms. An undersupply of such homes in current TOD areas might have pushed those residents far from the station areas. We recommend planners and developers take the needs of these families with children into account, in particular, to provide home options with larger space and 3-4 bedrooms. The "missing middle" literature (Webber, 2019) also suggests that medium-density housing types that fall between the scales of single-family homes and mid- to high-rise apartments can also provide desirable home options for larger families. By attracting families with kids, such housing will make a big difference in place-making and building complete and vibrant TOD communities.

4.5.3 Caveats and Future Research

This paper systematically analyzes different demographic groups' preferences for housing in TOD areas by considering both the current and potential demand. We argue that the significant support for transit-oriented development in KW from the missing target would not translate into actual relocation choices, unless local governments and developers do more to produce such housing and neighbourhoods (Lewis and Baldassare, 2010). This paper focuses on the demand side while touching less on the supply side and regulations. We will discuss specific housing and planning policies, as well as developers' behaviours and building strategies in future research. Investigating social changes in TOD neighbourhoods would be another future direction.

Chapter 5

Conclusion

This thesis presents three empirical studies to investigate the housing price dynamics, housing demand and residential preferences in Kitchener-Waterloo, Canada. The first study uses a large transaction dataset through 2005-2018, and first introduces a spatio-temporal multilevel model aiming to control for both spatial and temporal effects on housing price determination. The second study uses a unique-designed survey dataset to explore housing demand of different households during the LRT construction period. The third paper further analyzes the survey data with an emphasis on households' preferences for TOD neighbourhoods. Taking advantage of two unique datasets for housing analyses, this thesis not only sheds light on the overall housing price dynamics over the years but also provides important insights on housing demand and residential preferences along with the LRT construction in the region. The general findings, contributions, planning implications and future work are given below.

5.1 Key findings

Q1: How do the "recent comparable sales" impact the housing prices? How do different neighbourhoods impact housing prices? What are the main advantages of specifying a spatio-temporal multilevel model for housing prices, compared to the purely spatial hedonic model?

The STAR+MLM model results from Chapter 2 provide evidence of three distinct effects on housing price determination: i) the spatio-temporal relations, i.e, the recent comparable sales' impacts; ii) the spatial heterogeneity across neighbourhoods; and iii) the spatial dependence between neighbourhoods.

First, the impact of the past 3 months' sales of neighbouring properties (within 2.5 km) is significant and positive for all housing types except for townhouses, confirming the "recent comparable sales" approach on price determination. The purely spatial models tend to overestimate the dependence between individual properties, mainly due to the strong assumption that not only the past nearby sales but also the concurrent and the future nearby sales can affect the current sales prices. The STAR+MLM model considers the "arrow" of time in spatial relations. Second, neighbourhood heterogeneity plays a significant role in price determination for townhouses, semi-detached houses and condos, while neighbourhood heterogeneity contributes less to the unexplained price variations in single-detached houses, which are more attributed to the unobserved property differences. Third, this thesis identifies significant neighbourhood dependence, where nearby neighbourhoods tend to impact housing prices similarly, especially for condos, single- and semi-detached houses. Those results highlight the importance of considering neighbourhood effects that underly the housing price formation. In addition, the STAR+MLM models produce better model fit.

Q2: What are the associations between housing prices and housing characteristics including the structural and neighbourhood attributes? What trends are seen in the time fixed-effects over the years 2005-2018? What is the relationship between the LRT investment and housing prices of different housing types?

Chapter 2 reports the hedonic estimates for housing characteristics. According to the preferred model results, people are willing to pay 18.8% more for condos in the CTC than condos outside the CTC; however, people are willing to pay 11% less for single-detached houses in the CTC than houses outside the CTC, after controlling for other attributes. For transit, people are willing to pay more for condos with better bus transit access; while people are willing to pay to more for houses being further from bus stops. For public open space, prices are higher with better open space access for most housing types. Further, Chapter 2 finds the significant synergy effect between the CTC and open space access for most housing types, indicating that people are willing to pay even more for housing in the CTC area when it also has better access to public open space.

Estimates for the time fixed effects reflect how the markets have varied since 2005 after controlling for the main attributes and regional housing price inflation. For all the four markets, the housing price has seen a quick rise from 2008 to 2010 after the financial crisis, and experienced an extreme price surge through 2016 to 2017, which echoes the housing boom occurring in KW during this period due to a sudden increase of buyers from the GTHA area.

Model estimates for the interaction between the CTC and time dummies provide insights on the possible impacts the LRT has made on housing prices. After the regional government approved the LRT in 2011, both single-detached and semi-detached houses within the CTC have seen a higher price increase than houses outside the CTC; condo prices in the CTC did not see a premium until 2014 after the LRT started construction. Condo prices in the CTC seem to be less impacted by the LRT investment during the study period, compared to the houses in the CTC. A possible reason might be that most new condos within the CTC were still being constructed and the time lag delayed the influence from the LRT announcement till the construction stage.

Q3: Based on the housing survey analysis, do households have heterogeneous preferences for both dwelling and locational attributes of housing?

Chapter 3 finds that households with children in our survey were willing to pay significantly more for single-detached homes with a larger yard, larger home size, more bedrooms and more bathrooms, as expected. For household income groups, only the wealthiest group with more than $150,000 annual income are willing to pay much more for a larger single-detached house with more space and larger yard size. Therefore, household structure seems to be the major demand shifter for most dwelling attributes. Different household groups do not see much difference in their preferences for the locational and neighbourhood attributes, except that couples without children and seniors aged 55 and over are found willing to pay more for the CTC neighbourhoods, and seniors preferred the denser areas. Household characteristics such as income, education level and employment status do not significantly differentiate housing preferences.

In addition, our results confirm that some people self-selected to certain areas to satisfy their preferences for private open space, single-detached houses, or the CTC neighbourhoods. In other words, people who have certain attitudes toward particular neighbourhoods or lifestyles (such as a transit-oriented lifestyle) can also influence their residential location choices.

Q4: Who is currently living in TOD areas? Who shows preferences for TOD but is currently living outside of TOD areas? Who still prefers living in car-oriented suburban areas? Do the demographic profiles, preferences and home choices of these groups differ significantly?

Chapter 4 further analyzes the survey data with a focus on examining different households' preferences for TOD neighbourhoods. It identifies three groups: 1) Current TOD households; 2) Potential TOD households; and 3) Car-dependent households. It is surprising to see that although 55% of households in our sample stated preferences for the car-oriented neighbourhoods in the suburbs, 36.2% of households showed a strong desire for the TOD neighbourhoods, although they purchased outside the CTC. This indicates a decent proportion of potential demand for housing in the TOD. It should also be noted that most TOD features, including LRT access, bus access, walkability, ease to cycle, access to urban centre and access to open space, are more important factors to the current and potential TOD households, compared to those car-oriented households.

When comparing the three groups in demographic profiles, a higher proportion of smaller size households (singles or couples without children) as well as seniors aged 55 and over are found in the current TOD households, while a larger proportion of couples with children, aged 25-34 with medium household income ($50,000-$99,999) are found in both the potential TOD households and the car-dependent households. With respect to their preferences for structural housing features, bedrooms, home size and garages are found to be more important to those potential TOD buyers than the current TOD households. This was also reflected by comparing the moving motivations, where most current TOD households moved to CTC because of the LRT or downsizing while the potential TOD households moved to the area outside the CTC mainly for upsizing.

5.2 Contributions and planning implications

This thesis makes three contributions to the empirical housing research. First, it contributes to the spatial hedonic modelling literature by providing evidence of the importance to 1) take the time "arrow" into spatial relations in price modelling and 2) consider the higher-level neighbourhood effects in housing price determination. Second, it contributes to the housing demand literature by applying a two-stage demand analysis and recovering heterogeneous preferences under residential location choices. Third, it contributes to the TOD literature by providing evidence of the significant potential demand for housing in TOD neighbourhoods and highlighting the importance of gaining an updated knowledge of various households' preferences and real choices.

These works also provide several implications for urban and regional planning, which are summarized below.

5.2.1 Shaping communities in the corridor by providing a variety of home options

The region has proposed the rapid transit system aiming to not only move people but shape communities. The Regional Official Plan has explicitly devised the goal of planning for "an appropriate range and mix of housing choices for all income groups" to create vibrant urban areas (p.39 Chapter 3, ROP 2015). By looking at the housing market outcomes through 2005-2018, this thesis finds that the majority of homes purchased in the CTC are single-detached homes with an average of 1334 sqft and condo units with one or two bedrooms and an average of 1061 sqft. This might indicate the missing housing types with medium density in the CTC, which fall between the low-density of single-family homes and the high-density of mid- to high-rise condos. According to the 2018 report of "monitoring change in the CTC" conducted by Region of Waterloo (2019), 92.7% of new residential building units from 2011 to 2018 are apartments, and most of the units have only one or two bedrooms. This further confirms that smaller-size apartments will be the primary new supply of homes in the CTC, while the "missing middle" housing options in the CTC will be definitely in undersupply. This thesis suggests governments gear more developments towards the medium-density housing units instead of putting primary focus on the med/high-rise apartments.

It should also be noted that, the existing single-detached houses still account for almost half of housing units in the CTC; however, they are much older and smaller compared to the houses in the suburbs. Based on our model results in Chapter 2, people are willing to pay 11.7% less for houses within the CTC, compared to houses outside the CTC. These low-density single-family houses imply the huge potential for the urban area intensification. The Regional Official Plan has clearly stated that the governments encourage"appropriate, individual lot intensification, such as secondary apartments and garden suites in residential neighbourhoods" (p.40 Chapter 3, ROP 2015) to provide a range and mix of permanent housing. Those improvement measures for single-detached homes and communities are expected to be capitalized into property values and meet the target of density and *reurbanization* in urban cores.

Existing and new condos serve as the other major housing supply in the corridor. According to our results, people are willing to pay 17.2% significantly more for condos within the CTC than condos outside the CTC. These effects can be two-sided. On one hand, governments hope to see property value increase for new developments in the core, which might indicate a vibrant economy but also may be due to investors' speculating purchase. On the other hand, the increased housing prices would reduce the housing affordability and produce gentrification on the other hand. Therefore, continuously supporting higher-end condo developments in the corridor might run counter to the regional's initial plan for creating a wide range of home options for all residents.

5.2.2 Creating complete communities by satisfying various housing demand

According to the Regional Official Plan, the region supports "Transit Oriented Development with a diverse mix of land uses, housing types and open spaces in close proximity to each other" aiming to create *complete communities* (p.19 Chapter 2, ROP 2015). However, "complete communities" should not only refer to the wide range of housing types and neighbourhoods, but should refer to the communities occupied by a range and mix of residents.

The current planning policies encourage medium to high density residential developments in the transit corridor, while most only target the smaller-size households, such as the seniors who seek to downsize and the young singles or couples without children who prefer an urban lifestyle and a shorter commute in the core area. This thesis also finds that during the LRT construction, couples without kids and seniors aged 55 and over are more willing to pay for the CTC housing, compared to the other demographic groups. Families with kids have rarely been targeted by new residential developments in the core area. Chapter 4 in this thesis provides strong evidence of some families with kids indeed holding a strong desire for living in the CTC area. However, they purchased outside the CTC possibly because their basic needs for larger housing space, better schools and kid-friendly neighbourhood amenities could not be satisfied in the CTC. This thesis argues that medium-density housing units with 3-4 bedrooms targeting families with school-age kids should be built for creating complete communities and enhancing urban vibrancy. Suburban detached houses should not be the only choice for families with kids, and appropriate home options with adequate size in the CTC for those families will make a real difference of the region. The Region's planning department has realized the significance of having families with children in the corridor and particular planning policies are waiting to be made.

5.3 Limitations and future work

This thesis was conducted with a particular focus on analyzing the housing market shifts as the region's growth policy changes and providing updated knowledge on housing demand and residential preferences. It answers all the questions initially proposed; however, there are still gaps that need to be bridged in the future.

First, this thesis provides insights on what the price differences are within or outside the CTC area and how the LRT has influenced certain demographics. However, it is not able to derive a definite conclusion about how the LRT announcement and construction have contributed to the housing price increase. To further explore the causality between the transit investment and property value changes, the difference-in-difference (DID) model might serve as an appropriate approach to assessing the policy impacts, as in (Bocarejo et al., 2014; Pilgram and West, 2018). In addition, it might be more meaningful to examine such impacts before and after the LRT operation. For the space-time hedonic modelling, the newly developed method called the Integrated Nested Laplace Approximation (INLA, 2017) can also be implemented for hedonic price modelling. It allows a wide range of different functions of spatial and temporal autocorrelation, and spatiotemporal models. Combining this variety with its computational efficiency, building the INLA space-time hedonic model directs a promising future work.

The second limitation refers to survey data and demand analysis. Despite the fact that our survey was designed to understand the home choices of residents in KW not only for those in the CTC, the obtained sample in the CTC was relatively small, and the analysis in Chapter 3 did not provide many implications for the CTC housing demand. For enhancing that analysis, combining the detailed transaction data with aggregated census-tract household characteristics to conduct the analysis would generate more meaningful results. In addition, a better transit access measure considering transit services (e.g., frequency) and a better walkability measure considering walking access to diverse destinations should also be created for producing more accurate estimation results.

References

Thorkild Ærø. Residential choice from a lifestyle perspective. *Housing, Theory and Society*, 23(2):109–130, 2006. ISSN 14036096. doi: 10.1080/14036090600773139.

William Alonso. *Location and land use: Toward a general theory of land rent.* Cambridge, Massachusetts: Harvard University Press, 1964.

Luc Anselin. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers, 1988. ISBN 90-247-3735-4.

Luc Anselin, R Florax, and Sergio J Rey, editors. *Advances in spatial econometrics: methodology, tools and applications*. Springer Science & Business Media, 2004. ISBN 3540437290.

G. Arrington and Robert Cervero. *Effects of TOD on Housing, Parking, and Travel*. National Academies Press, Washington, DC, 2008. ISBN 9780309117487. doi: https://doi.org/10.17226/14179.

Robert Malcolm Babin. *Estimating Homebuyer Preferences Under Intensification: Hedonic Modelling of Open Space and Multimodal Transit Amenities Preceding Light Rail in Kitchener-Waterloo*. Master's dissertation, University of Waterloo, 2016.

Patrick Bajari and C. Lanier Benkard. Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, 113(6):1239–1276, 2005. ISSN 0022-3808. doi: 10.1086/498586.

Patrick Bajari and Matthew E Kahn. Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities. *Journal of Business & Economic Statistics*, 23(1):20–33, 2005. ISSN 0735-0015. doi: 10.1198/073500104000000334.

Dwayne Marshall Baker and Bumsoo Lee. How Does Light Rail Transit (LRT) Impact Gentrification? Evidence from Fourteen US Urbanized Areas. *Journal of Planning Education and Research*, 39(1):35–49, 2019. ISSN 0739456X. doi: 10.1177/0739456X17713619.

Timothy J Bartik. The Estimation of Demand Parameters in Hedonic Price Models. *Journal of Political Economy*, 95(1):81–88, 1987.

Patrick Bayer, Robert McMillan, and Kim Rueben. An Equilibrium Model of Sorting in an Urban Housing Market. *NBER Working Paper Series*, 2004. doi: 10.3386/w10865.

Patrick Bayer, Fernando Ferreira, and Robert McMillan. A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, 115 (4):588–638, 2007. ISSN 0022-3808. doi: 10.1086/522381.

Patrick Bayer, Marcus Casey, Fernando Ferreira, and Robert McMillan. Racial and ethnic price differentials in the housing market. *Journal of Urban Economics*, 102:91–105, 2017. ISSN 00941190. doi: 10.1016/j.jue.2017.07.004.

Luca Bertolini, Carey Curtis, and John Renne. Station area projects in Europe and beyond: Towards transit oriented development? *Built Environment*, 38(1):31–50, 2012. ISSN 02637960. doi: 10.2148/benv.38.1.31.

Stephen B. Billings. Estimating the value of a new transit option. *Regional Science and Urban Economics*, 41(6):525–536, 2011. ISSN 01660462. doi: 10.1016/j.regsciurbeco.2011.03.013.

Kelly C. Bishop and Christopher D. Timmins. Using Panel Data to Easily Estimate Hedonic Demand Functions. *Journal of the Association of Environmental and Resource Economists*, 5(3):517–543, 2018. ISSN 2333-5955. doi: 10.1086/696981.

Juan Pablo Bocarejo, Ingrid Joanna Portilla, Juan Miguel Velásquez, Mónica Natalia Cruz, Andrés Peña, and Daniel Ricardo Oviedo. An innovative transit system and its impact on low income users: The case of the Metrocable in Medellín. *Journal of Transport Geography*, 39:49–61, 2014. ISSN 09666923. doi: 10.1016/j.jtrangeo.2014.06.018.

Nancy E. Bockstael. Modeling economics and space: the importance of a spatial perspective. *American Journal of Agricultural Economics*, 78(5):1168–1180, 1996.

Glen Bramley and Sinéad Power. Urban form and social sustainability: The role of density and housing type. *Environment and Planning B: Planning and Design*, 36(1):30–48, 2009. ISSN 14723417. doi: 10.1068/b33129.

David M. Brasington and Diane Hite. Demand for environmental quality: a spatial hedonic analysis. *Regional Science and Urban Economics*, 35(1):57–82, 2005.

Katherine Brookfield. Residents' preferences for walkable neighbourhoods. *Journal of Urban Design*, 22(1):44–58, 2017. ISSN 14699664. doi: 10.1080/13574809.2016.1234335.

Chris Brunsdon, A. Stewart Fotheringham, and M. Charlton. Spatial nonstationarity and autoregressive models. *Environment and Planning A*, 30:957–973, 1998.

Trudi Bunting, Pierre Filion, Heidi Hoernig, Mark Seasons, and Jeff Lederer. Density, size, dispersion: Towards understanding the structural dynamics of mid-size cities. *Canadian Journal of Urban Research*, 16(2):27–52, 2007. ISSN 11883774.

Robert W. Burchell, David Listokin, and Catherine C. Galley. Smart growth: More than a ghost of urban policy past, less than a bold new horizon. *Housing Policy Debate*, 11 (4):821–879, 2000. ISSN 10511482. doi: 10.1080/10511482.2000.9521390.

Ayse Can and Megbolugbe Isaac. Spatial dependence and house price index construction. *Journal of Real Estate Finance and Economics*, 14:203–222, 1997.

Xinyu Cao. Is Alternative Development Undersupplied? *Transportation Research Record: Journal of the Transportation Research Board*, 2077:97–105, 2008. ISSN 0361-1981. doi: 10.3141/2077-13.

Xinyu Cao and Jill a. Hough. Hedonic Value of Transit Accessibility: An Empirical Analysis in a Small Urban Area. *Journal of the Transportation Research Forum*, 47(3):171–183, 2012. ISSN 1046-1469. doi: 10.5399/osu/jtrf.47.3.2146.

Xinyu Cao, Patricia L. Mokhtarian, and Susan L. Handy. *Examining the impacts of residential self-selection on travel behaviour: A focus on empirical findings*, volume 29. 2009. ISBN 0144164080253. doi: 10.1080/01441640802539195.

Radosław Cellmer, Katarzyna Kobylinska, and Mirosław Bełej. Application of hierarchical spatial autoregressive models to develop land value maps in urbanized areas. *ISPRS International Journal of Geo-Information*, 8(4), 2019. ISSN 22209964. doi: 10.3390/ijgi8040195.

Robert Cervero. *TOD in the United States: Experiences, Challenges and Prospects*. 2004. ISBN 0309087953.

Robert Cervero. Transit-oriented development. *Environment and Planning A*, 39:2068–2085, 2007. doi: 10.1068/a38377.

Coro Chasco and Julie Le Gallo. Hierarchy and spatial autocorrelation effects in hedonic models. *Economics Bulletin*, 32(2):1474–1480, 2012.

Sudip Chattopadhyay. Estimating the Demand for Air Quality: New Evidence Based on the Chicago Housing Market. *Land Economics*, 75(1):22–38, 1999. ISSN 00237639. doi: 10.2307/3146991.

William A.V. Clark. Changing Residential Preferences across Income, Education, and Age Findings from the Multi-City Study of Urban Inequality. *Urban Affairs Review*, 44(3): 334–355, 2009.

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

Ricardo Crespo and Adrienne Gret-Regamey. Local hedonic house-price modelling for urban planners: Advantages of using local regression techniques. *Environment and*

Planning B: Planning and Design, 40(4):664–682, 2013. ISSN 02658135. doi: 10.1068/b38093.

Carey Curtis, John L. Renne, and Luca Bertolini, editors. *Transit oriented development: making it happen*. Ashgate e-BOOK, 2009. ISBN 0754673154. doi: 10.1068/a38377.

Casey Dawkins and Rolf Moeckel. Transit-Induced Gentrification: Who Will Stay, and Who Will Go? *Housing Policy Debate*, 26(4-5):801–818, 2016. ISSN 2152050X. doi: 10.1080/10511482.2016.1138986.

Jonas De Vos, Veronique Van Acker, and Frank Witlox. Urban sprawl: neighbourhood dissatisfaction and urban preferences. Some evidence from Flanders. *Urban Geography*, 37(6):839–862, 2016. ISSN 02723638. doi: 10.1080/02723638.2015.1118955.

Hank Dittmar, Dena Belzer, and Gerald Autler. An introduction to transit-oriented development. *The New Transit Town: Best Practices In Transit-Oriented Development*, pages 1–18, 2004.

Guanpeng Dong, Richard Harris, Kelvyn Jones, and Jianhui Yu. Multilevel Modelling with Spatial Interaction Effects with Application to an Emerging Land Market in Beijing, China. *PlOS ONE*, 10(6): e0130761, 2015. doi: 10.1371/journal.pone.0130761.

Jean Dubé and Diègo Legros. Spatial econometrics and the hedonic pricing model: what about the temporal dimension? *Journal of Property Research*, 31(4):333–359, 2014. ISSN 14664453. doi: 10.1080/09599916.2014.913655.

Jean Dubé, Marius Thériault, and François Des Rosiers. Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992-2009. *Transportation Research Part A: Policy and Practice*, 54:49–66, 2013. ISSN 09658564. doi: 10.1016/j.tra.2013.07.015.

Jean Dubé, Diègo Legros, and Sotirios Thanos. Past price 'memory' in the housing market: testing the performance of different spatio-temporal specifications. *Spatial Economic Analysis*, 13(1):118–138, 2018. ISSN 17421780. doi: 10.1080/17421772.2018.1395063.

Michael Duncan. The impact of transit-oriented development on housing prices in San diego, CA. *Urban Studies*, 48(1):101–127, 2011. ISSN 00420980. doi: 10.1177/0042098009359958.

Neil Dunse, Sotirios Thanos, and Glen Bramley. Planning policy, housing density and consumer preferences. *Journal of Property Research*, 30(3):221–238, 2013. ISSN 09599916. doi: 10.1080/09599916.2013.795992.

Ivar Ekeland, James J. Heckman, and Lars Nesheim. Identification and Estimation of Hedonic Models. *Journal of Political Economy*, 112(1):60–109, 2004. ISSN 0022-3808. doi: 10.1086/379947.

Alan Evans and Rachael Unsworth. Housing Densities and Consumer Choice. *Urban Studies*, 49(6):1163–1177, 2012. ISSN 00420980. doi: 10.1177/0042098011405692.

Reid Ewing and Robert Cervero. Travel and the Built Environment. *Journal of the American Planning Association*, 76(3):265–294, 2010. ISSN 0194-4363. doi: 10.1080/01944361003766766.

A Stewart Fotheringham and Taylor M Oshan. Geographically weighted regression and multicollinearity : dispelling the myth. *Journal of Geographical Systems*, 18(4):303–329, 2016. ISSN 1435-5949. doi: 10.1007/s10109-016-0239-5.

A. Stewart Fotheringham, Martin E. Charlton, and Chris Brunsdon. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30(11):1905–1927, 1998. ISSN print 0308-518X, electronic 1472-3409. doi: 10.1068/a301905.

Roland Füss and Jan A. Koller. The role of spatial and temporal structure for residential rent predictions. *International Journal of Forecasting*, 32(4):1352–1368, 2016. ISSN 01692070. doi: 10.1016/j.ijforecast.2016.06.001.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.

Arthur Getis. Spatial weights matrices. *Geographical Analysis*, 41(4):404–410, 2009. ISSN 00167363. doi: 10.1111/j.1538-4632.2009.00768.x.

Stephen Gibbons and Henry G. Overman. Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2):172–191, 2012. doi: 10.1111/j.1467-9787.2012.00760.x.

Marie Line Glaesener and Geoffrey Caruso. Neighborhood green and services diversity effects on land prices: Evidence from a multilevel hedonic analysis in Luxembourg. *Landscape and Urban Planning*, 143:100–111, 2015. ISSN 01692046. doi: 10.1016/j.landurbplan.2015.06.008. URL `http://dx.doi.org/10.1016/j.landurbplan.2015.06.008`.

H Goldstein and W Browne. Partioning variation in multilevel models. *Understanding statistics*, 1(4):223–231, 2002.

Harvey Goldstein. *Multi-level models in educational and social research*. Oxford University Press, 1987.

Allen C. Goodman and Thomas G. Thibodeau. Housing Market Segmentation. *Journal of Housing Economics*, 7:121–143, 1998.

Jessica. Guo and Chandra R. Bhat. Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, 15(1): 31–45, 2007. ISSN 09666923. doi: 10.1016/j.jtrangeo.2005.11.001.

Khandker M Nurul Habib and K M Knockelman. Modeling the Choice of Residential Location and Home Type: Recent Movers in Austin, Texas. In *87th Annual Meeting of the Transportation Research Board*, 2008.

Jacques A Hagenaars and Allan L McCutcheon. *Applied latent class analysis*. Cambridge University Press, 2002.

Susan Handy. Smart growth and the transportation-land use connection: What does the research tell us? *International Regional Science Review*, 28(2):146–167, 2005. ISSN 01600176. doi: 10.1177/0160017604273626.

Tristen Hayfield and Jeffrey S Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5):1–32, 2008. ISSN 1548-7660. doi: 10.1198/jasa.2001.s374.

Christopher Higgins and Pavlos Kanaroglou. Forty years of modelling rapid transit s land value uplift in North America moving beyond the tip of the iceberg. *Transport Reviews*, 36(5):610–634, 2016. ISSN 14645327. doi: 10.1080/01441647.2016.1174748.

Christopher Higgins, Mark Ferguson, and Pavlos Kanaroglou. Light Rail and Land Use Change: Rail Transit' s Role in Reshaping and Revitalizing Cities. *Journal of Public Transportation*, 17(2):93–112, 2014. ISSN 1077-291X. doi: 10.5038/2375-0901.17.2.5.

Christopher D. Higgins, Matthew D. Adams, Weeberb J. Réquia, and Moataz Mohamed. Accessibility, air pollution, and congestion: Capturing spatial trade-offs from agglomeration in the property market. *Land Use Policy*, 84:177–191, 2019. ISSN 02648377. doi: 10.1016/j.landusepol.2019.03.002.

Dongwoo Hyun and Stanimira Milcheva. Spatial dependence in apartment transaction prices during boom and bust. *Regional Science and Urban Economics*, 68:36–45, 2018. ISSN 18792308. doi: 10.1016/j.regsciurbeco.2017.11.001.

Yannis M. Ioannides and Jeffrey E. Zabel. Neighbourhood effects and housing demand. *Journal of Applied Econometrics*, 18(5):563–584, 2003. ISSN 08837252. doi: 10.1002/jae.740.

Justin Jacobson and Ann Forsyth. Seven American TODs: Good practices for urban design in Transit-Oriented Development projects. *Journal of Transport and Land Use*, 2:51–88, 2008. ISSN 1938-7849. doi: 10.5198/jtlu.v1i2.67.

Kelvyn Jones and Nina Bullen. A Multi-level Analysis of the Variations in Domestic Property Prices: Southern England, 1980-87. *Urban Studies*, 1993. ISSN 1360063X. doi: 10.1080/00420989320081341.

Lia Karsten. Housing as a way of life: Towards an understanding of middle-class families' preference for an urban residential location. *Housing Studies*, 22(1):83–98, 2007. ISSN 02673037. doi: 10.1080/02673030601024630.

Tae-Kyung Kim, Mark W. Horner, and Robert W. Marans. Life Cycle and Environmental Factors in Selecting Residential and Job Locations. *Housing Studies*, 20(3):457–473, 2005. ISSN 0267-3037. doi: 10.1080/02673030500062335.

Kitchener-cambridge-waterloo Central Transit Corridor (CTC) Monitoring Program. Monitoring Change in the CTC 2018 Report. Technical report, Region of Waterloo, 2019.

H Allen Klaiber and Nicolai V Kuminoff. Equilibrium Sorting Models of Land Use and Residential Choice. In Joshua M Duke and JunJie Wu, editors, *The Oxford Handbook of Land Economics*, chapter 36, pages 352–379. Oxford University Press, 2014. doi: 10.1093/oxfordhb/9780199763740.013.011.

H Allen Klaiber and Daniel J. Phaneuf. Valuing open space in a residential sorting model of the Twin Cities. *Journal of Environmental Economics and Management*, 60(2):57–77, 2010. ISSN 00950696. doi: 10.1016/j.jeem.2010.05.002.

Richard D. Knowles and Fiona Ferbrache. Evaluation of wider economic impacts of light rail investment on cities. *Journal of Transport Geography*, 54:430–439, 2016. ISSN 09666923. doi: 10.1016/j.jtrangeo.2015.09.002. URL `http://dx.doi.org/10.1016/j.jtrangeo.2015.09.002`.

Julia Koschinsky, Nancy Lozano-gracia, and Gianfranco Piras. The welfare benefit of a home's location: an empirical comparison of spatial and non-spatial model estimates. *Journal of Geographical Systems*, 14:319–356, 2012. doi: 10.1007/s10109-011-0148-6.

Andy L. Krause, Christopher Bitter, Andy L Krause, Christopher Bitter, Andy L. Krause, and Christopher Bitter. Spatial econometrics, land values and sustainability: Trends in real estate valuation research. *Cities*, 29:19–25, 2012. ISSN 02642751. doi: 10.1016/j.cities.2012.06.006. URL `http://dx.doi.org/10.1016/j.cities.2012.06.006`.

P Phani Kumar, Ch Ravi Sekhar, and Manoranjan Parida. Residential dissonance in TOD neighborhoods. *Journal of Transport Geography*, 72:166–177, 2018. ISSN 0966-6923. doi: 10.1016/j.jtrangeo.2018.09.005.

Nicolai V. Kuminoff and Jaren C. Pope. A novel approach to identifying hedonic demand parameters. *Economics Letters*, 116(3):374–376, 2012. ISSN 01651765. doi: 10.1016/j.econlet.2012.04.017.

Nicolai V Kuminoff, V Kerry Smith, and Christopher Timmins. The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets. *Journal of Economic Literature*, 51(4):1007–1062, 2013. ISSN 0022-0515. doi: 10.1257/jel.51.4.1007.

KWAR. 2017 A big year for Kitchener Waterloo homes sales. Kitchener Waterloo Association of Realtors® News Release http://www.kwar.ca/2017-big-year-kitchener-waterloo-homes-sales/, 2018.

Stephen Law. Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London. *Cities*, 60: 166–179, 2017. ISSN 02642751. doi: 10.1016/j.cities.2016.08.008. URL `http://dx.doi.org/10.1016/j.cities.2016.08.008`.

Yongsung Lee, Giovanni Circella, Patricia L. Mokhtarian, and Subhrajit Guhathakurta. Heterogeneous residential preferences among millennials and members of generation X in California: A latent-class approach. *Transportation Research Part D: Transport and Environment*, 76:289–304, 2019. ISSN 13619209. doi: 10.1016/j.trd.2019.08.001.

James P Lesage and R Kelley Pace. The biggest myth in spatial econometrics. *Econometrics*, 2(4):217–249, 2014.

Jonathan Levine and Lawrence D. Frank. Transportation and land-use preferences and residents' neighborhood choices: The sufficiency of compact development in the Atlanta region. *Transportation*, 2007. ISSN 00494488. doi: 10.1007/s11116-006-9104-6.

Paul G. Lewis and Mark Baldassare. The Complexity of Public Attitudes Toward Compact Development. *Journal of the American Planning Association*, 76(2):219–237, mar 2010. ISSN 0194-4363. doi: 10.1080/01944361003646471.

Felix Haifeng Liao, Steven Farber, and Reid Ewing. Compact development and preference heterogeneity in residential location choice behaviour: A latent class analysis. *Urban Studies*, 52(2):314–337, 2015. ISSN 1360063X. doi: 10.1177/0042098014527138.

X Liu. Spatial and Temporal Dependence in House Price Prediction. *Journal of Real Estate Finance and Economics*, 47(2):341–369, 2013.

Hollie Lund. Reasons for living in a transit-oriented development and associated transit use. *Journal of the American Planning Association*, 72(3):357–366, 2006. ISSN 01944363. doi: 10.1080/01944360608976757.

Guowei Lyu, Luca Bertolini, and Karin Pfeffer. Developing a TOD typology for Beijing metro station areas. *Journal of Transport Geography*, 55:40–50, 2016. ISSN 0966-6923. doi: 10.1016/j.jtrangeo.2016.07.002. URL `http://dx.doi.org/10.1016/j.jtrangeo.2016.07.002`.

Douglas S. Massey and Jonathan Tannen. Suburbanization and segregation in the United States: 1970–2010. *Ethnic and Racial Studies*, 41(9):1594–1611, 2018. ISSN 14664356. doi: 10.1080/01419870.2017.1312010.

Katherine E. Masyn. Latent Class Analysis and Finite Mixture Modeling. In Todd D. Little, editor, *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis*, pages 551–611. 2013. doi: 10.1093/oxfordhb/9780199934898.013.0025.

Daniel McFadden. Modeling the choice of residential location. *Transportation Research Record*, (673):72–77, 1978. ISSN 0361-1981.

Peter Mieszkowski and Edwin S Mills. The Causes of Metropolitan Suburbanization. *Journal of Economic Perspectives*, 7(3):135–147, 1993. ISSN 0895-3309. doi: 10.1257/jep.7.3.135.

Ministry of Municipal Affairs and Housing. Growth Plan for the Greater Golden Horseshoe. 2006.

Ministry of Municipal Affairs and Housing. Provincial Policy Statement. 2014.

Ministry of Municipal Affairs and Housing. Growth Plan for the Greater Golden Horseshoe. 2017.

Sara I. Mohammad, Daniel J. Graham, Patricia C. Melo, and Richard J. Anderson. A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice*, 50:158–170, 2013. ISSN 09658564. doi: 10.1016/j.tra.2013.01.013. URL `http://dx.doi.org/10.1016/j.tra.2013.01.013`.

Edward Morey, Mara Thiene, Maria De Salvo, and Giovanni Signorello. Using attitudinal data to identify latent classes that vary in their preference for landscape preservation. *Ecological Economics*, 68(1-2):536–546, 2008. ISSN 09218009. doi: 10.1016/j.ecolecon.2008.05.015.

Corinne Mulley, Chi Hong (Patrick) Tsai, and Liang Ma. Does residential property price benefit from light rail in Sydney? *Research in Transportation Economics*, 67:3–10, 2018. ISSN 07398859. doi: 10.1016/j.retrec.2016.11.002.

Dowell Myers and Elizabeth Gearin. Current preferences and future demand for denser residential environments. *Housing Policy Debate*, 12(4):633–659, 2001. doi: 10.1080/10511482.2001.9521422.

NAR. NAR 2017 Community Preference Survey. https://www.nar.realtor/reports/nar-2017-community-preference-survey. Technical report, 2017.

Robert B. Noland, Marc D. Weiner, Stephanie DiPetrillo, and Andrew I. Kay. Attitudes towards transit-oriented development: Resident experiences and professional perspectives. *Journal of Transport Geography*, 60:130–140, 2017. ISSN 09666923. doi: 10.1016/j.jtrangeo.2017.02.015.

Doina Olaru, Brett Smith, and John H E Taplin. Residential location and transit-oriented development in a new rail corridor. *Transportation Research Part A: Policy and Practice*, 45(3):219–237, 2011. ISSN 09658564. doi: 10.1016/j.tra.2010.12.007.

Scott Orford. *Valuing the Built Environment: GIS and House Price Analysis*. London: Routledge, 1999. doi: 10.4324/9781315235134.

Scott Orford. Valuing locational externalities: A GIS and multilevel modelling approach. *Environment and Planning B: Planning and Design*, 29:105–127, 2002. doi: 10.1068/b2780.

Liv Osland. An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, 32(3):289–320, 2010. ISSN 08965803.

Liv Osland, Ingrid Sandvig Thorsen, and Inge Thorsen. Accounting for Local Spatial Heterogeneities in Housing Market Studies. *Journal of Regional Science*, 56(5):895–920, 2016. ISSN 14679787. doi: 10.1111/jors.12281.

R Kelley Pace, Ronald Barry, John M Clapp, and Mauricio Rodriquez. Spatiotemporal Autoregressive Models of Neighborhood Effects. *The Journal of Real Estate Finance and Economics*, 17(1):15–33, 1998. ISSN 0895-5638, 1573-045X. doi: 10.1023/A:1007799028599.

Qisheng Pan. The impacts of light rail on residential property values in a non-zoning city: A new test on the Houston METRORail transit line. *Journal of Transport and Land Use*, 12(1):241–264, 2019. ISSN 1938-7849. doi: 10.5198/jtlu.2019.1310.

Enrica Papa and Luca Bertolini. Accessibility and Transit-Oriented Development in European metropolitan areas. *Journal of Transport Geography*, 47:70–83, 2015. ISSN 0966-6923. doi: 10.1016/j.jtrangeo.2015.07.003.

Clemens A. Pilgram and Sarah E. West. Fading premiums: The effect of light rail on residential property values in Minneapolis, Minnesota. *Regional Science and Urban Economics*, 69(December 2017):1–10, 2018. ISSN 18792308. doi: 10.1016/j.regsciurbeco.2017.12.008.

Dorina Pojani and Dominic Stead. Transit-Oriented Design in the Netherlands. *Journal of Planning Education and Research*, 35(2):131–144, 2015. ISSN 0739456X. doi: 10.1177/0739456X15573263.

Neelam C. Poudyal, Donald G. Hodges, and Christopher D. Merrett. A hedonic analysis of the demand for and benefits of urban recreation parks. *Land Use Policy*, 26(4):975–983, 2009. ISSN 02648377. doi: 10.1016/j.landusepol.2008.11.008.

Lissy La Paix Puello and Karst Geurs. Modelling observed and unobserved factors in cycling to railway stations: Application to transit-oriented-developments in the Netherlands. *European Journal of Transport and Infrastructure Research*, 15(1):27–50, 2015. ISSN 15677141. doi: 10.18757/ejtir.2015.15.1.3057.

Region of Waterloo. Planning Our Future: Regional Growth Management Strategy. 2003.

Region of Waterloo. Regional Official Plan 2031. 2015.

Nick Revington. Gentrification, transit, and land use: Moving beyond neoclassical theory. *Geography Compass*, 9(3):152–163, 2015. ISSN 17498198. doi: 10.1111/gec3.12203.

W. Rid and a. Profeta. Stated Preferences for Sustainable Housing Development in Germany–A Latent Class Analysis. *Journal of Planning Education and Research*, 31 (1):26–46, 2011. ISSN 0739-456X. doi: 10.1177/0739456X10393952.

Sherwin Rosen. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1):34–55, 1974. ISSN 0022-3808. doi: 10.1086/260169.

Håvard Rue, Andrea Riebler, Sigrunn H. Sørbye, Janine B. Illian, Daniel P. Simpson, and Finn K. Lindgren. Bayesian Computing with INLA: A Review. *Annual Review of Statistics and Its Application*, 4(1):395–421, 2017. ISSN 2326-8298. doi: 10.1146/annurev-statistics-060116-054045.

Patrick M. Schirmer, Michael a.B. Van Eggermond, and Kay W. Axhausen. The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*, 7(2):3–21, 2014. ISSN 1938-7849. doi: 10.5198/jtlu.v7i2.740.

Kenneth A Small and Seiji S C Steimetz. Spatial hedonics and the willingness to pay for residential amenities. *Journal of Regional Science*, 52(4):635–647, 2012. ISSN 00224146. doi: 10.1111/j.1467-9787.2012.00774.x.

Brett Smith and Doina Olaru. Lifecycle stages and residential location choice in the presence of latent preference heterogeneity. *Environment and Planning A*, 45(10):2495–2514, 2013. ISSN 0308518X. doi: 10.1068/a45490.

Tony E. Smith and Peggy Wu. A spatio-temporal model of housing prices based on individual sales transactions over time. *Journal of Geographical Systems*, 11(4):333–355, 2009. ISSN 14355930. doi: 10.1007/s10109-009-0085-9.

Yan Song and G. J Knaap. *New urbanism and housing values: a disaggregate assessment*, volume 54. 2003. ISBN 3014056626.

Luca Staricco and Elisabetta Vitale Brovarone. Promoting TOD through regional planning . A comparative analysis of two European approaches. *Journal of Transport Geography*, 66:45–52, 2018. ISSN 0966-6923. doi: 10.1016/j.jtrangeo.2017.11.011. URL `https://doi.org/10.1016/j.jtrangeo.2017.11.011`.

Statistics Canada. Kitchener - Cambridge - Waterloo, Ontario. Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. Ottawa. Released November 29, 2017., 2017.

Statistics Canada. Table 18-10-0205-01 New housing price index, monthly. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810020501, 2018.

Sotirios Thanos, Jean Dubé, and Diègo Legros. Putting Time into Space: The Temporal Coherence of Spatial Applications in the Housing Market. *Regional Science and Urban Economics*, 58:78–88, may 2016. doi: 10.1016/j.regsciurbeco.2016.03.001.

Ren Thomas and Luca Bertolini. Beyond the Case Study Dilemma in Urban Planning: Using a Meta-matrix to Distil Critical Success Factors in Transit-Oriented Development. *Urban Policy and Research*, 32(2):219–237, 2014.

Guang Tian, Reid Ewing, and William Greene. Desire for Smart Growth: A Survey of Residential Preferences in the Salt Lake Region of Utah. *Housing Policy Debate*, 25(3): 446–462, 2015. ISSN 1051-1482. doi: 10.1080/10511482.2014.971333.

Charles M. Tiebout. A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64(5):416–424, 1956. ISSN 0022-3808. doi: 10.1086/257839.

W.R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(sup1):234–240, 1970.

Constant I. Tra. A discrete choice equilibrium approach to valuing large environmental changes. *Journal of Public Economics*, 94(1-2):183–196, 2010. ISSN 00472727. doi: 10.1016/j.jpubeco.2009.10.006.

Jinny Tran. *Understanding Developer's Decision Making in the Region of Waterloo*. Senior honours essay, University of Waterloo, 2016.

Radoslaw Trojanek and Michal Gluszak. Spatial and time effect of subway on property prices. *Journal of Housing and the Built Environment*, 33(2):359–384, 2018. ISSN 15737772. doi: 10.1007/s10901-017-9569-y.

Veronique van Acker, Patricia L. Mokhtarian, and Frank Witlox. *Going soft: On how subjective variables explain modal choices for leisure travel*, volume 11. 2011. ISBN 1530752787.

Bert van Wee. Self-selection: A key to a better understanding of location choices, travel behaviour and transport externalities? *Transport Reviews*, 29(3):279–292, 2009. ISSN 01441647. doi: 10.1080/01441640902752961.

Bert Van Wee. Self-Selection: A Key to a Better Understanding of Location Choices, Travel Behaviour and Transport Externalities? *Transport Reviews*, 29(3):279–292, 2009. ISSN 0144-1647. doi: 10.1080/01441640902752961.

Antti Vasanen. Beyond stated and revealed preferences: The relationship between residential preferences and housing choices in the urban region of Turku, Finland. *Journal of Housing and the Built Environment*, 27(3):301–315, 2012. ISSN 15664910. doi: 10.1007/s10901-012-9267-8.

Ingmar Visser and Maarten Speekenbrink. depmixS4: An R package for hidden markov models. *Journal of Statistical Software*, 36(7):1–21, 2010. ISSN 15487660. doi: 10.18637/jss.v036.i07.

Kathrine von Graevenitz. *Improved Methods for the Valuation of the Environment in an Urban Setting*. PhD thesis, University of Copenhagen, 2013.

Joan L. Walker and Jieping Li. Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9(1):77–101, 2007. ISSN 14355930. doi: 10.1007/s10109-006-0030-0.

Kaitlin Webber. IN-BETWEEN ISSUES : EXPLORING THE " MISSING MIDDLE " IN ONTARIO. PRAGMA DISCUSSION PAPER. Technical report, School of Planning, University of Waterloo, Waterloo, 2019.

William C. Wheaton. A bid rent approach to housing demand. *Journal of Urban Economics*, 4(2):200–217, 1977. ISSN 00941190. doi: 10.1016/0094-1190(77)90023-7.

Danlin Yu, Yehua Dennis Wei, and Changshan Wu. Modeling spatial dimensions of housing prices in Milwaukee, WI. *Environment and Planning B: Planning and Design*, 34(6): 1085–1102, 2007. ISSN 02658135. doi: 10.1068/b32119.

Edmund Zolnik. A longitudinal analysis of the effect of public rail infrastructure on proximate residential property transactions. *Urban Studies*, pages 1–22, 2019. doi: 10.1177/0042098019836564.

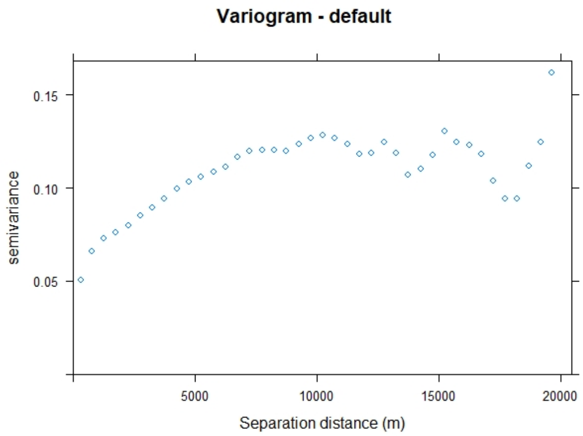
APPENDICES

A1 Appendices in Chapter 2

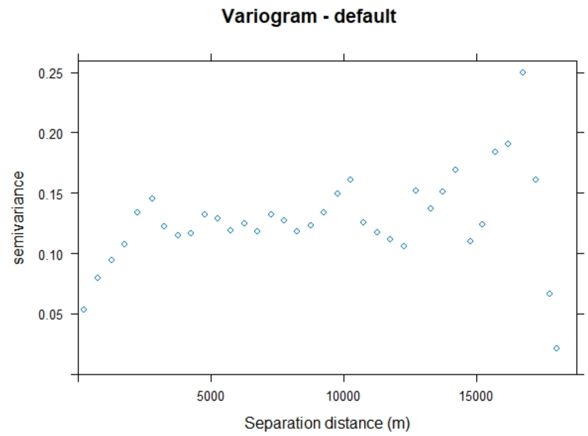
A1.1 Semi-variograms

When constructing a spatial weight matrix, decisions about the spatial extent for the weights are often made empirically by fitting a semi-variogram. This is also referred to as a hybrid "theoretical-empirical" approach by Getis (2009). We computed the empirical semi-variograms of housing prices for different housing types to determine the spatial distance threshold. As illustrated in Figure A-1, each semi-variogram curve depicts the spatial autocorrelation as a function of distance. The increasing pattern reflects the decreasing spatial dependence as the distance increases, and the flattening curve shows the diminishing dependence of the sample points. Based on the curves showing in Figure A-1, the prices of single-detached houses and townhouses influence housing prices of properties more than 5 km away, while the other housing types influence properties around 2-3 km.

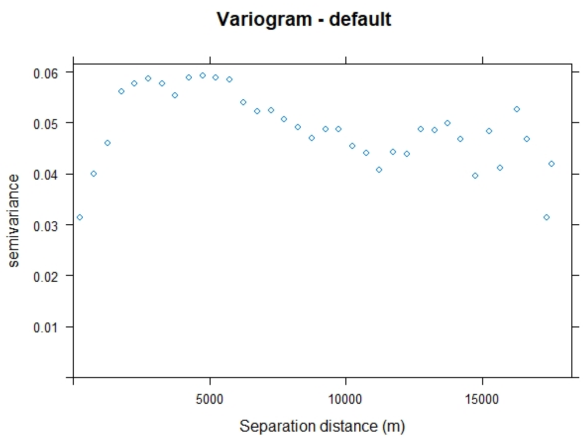
Based on the findings from Dubé and Legros (2014), the spatial over-connection problem would come out when using a large distance threshold. To minimize the risk of introducing bias due to over-connection, this study chooses 2.5 km as the threshold distance mainly to control the number of neighbours for each property defined in the spatial weight matrix. This choice was also approved by the local real estate professionals.



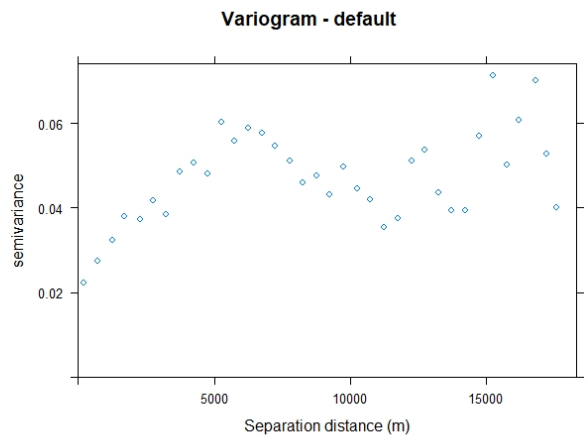
(a) Single-detached housing



(b) Condo housing



(c) Semi-detached housing/duplex



(d) Townhouses

Figure A-1: Empirical semi-variogram fit to transactions of different housing types

Note: This figure shows the range of spatial autocorrelation to be largely contained within a distance of 5km, 3km, 2.5km, and 5km for single-detached houses, condos, semi-detached houses, and townhouses, respectively.

A1.2 Data source

Table A-1: Data source

Input data	Data source and descriptions
Transaction data	*Source*: The MPAC and the Teranet company *Description*: This dataset contains property sales records from Jan 2005 to Mar 2018 in KW, including sales prices and the main housing structural attributes such as square footage, lot size, the number of bedrooms, bathrooms, garage, built year, pool and fireplace.
Property parcel layer	*Source*: The Geospatial Centre at the University of Waterloo *Description*: This layer represents all assessment parcels with unique Assessment Roll Numbers (ARN) in KW in 2018. This polygon layer was joined with the transaction dataset by the same ARN, and was converted to the centroid points of each polygon using the FeatureToPoint tool in ArcGIS.
Bus transit data	*Source*: Region of Waterloo *Description*: The Grand River Transit (GRT) stops dataset includes layers of 2005, 2009, 2011, 2012, 2014, 2015, 2016, 2017, 2018. The Euclidean distance from each parcel (point) to the nearest bus stop was calculated using the spatial join tool in ArcGIS.[a]
Regional roads data	*Source*: The Geospatial Centre at the University of Waterloo *Description*: This dataset includes road layers of the Waterloo Region in 2011, 2012, 2014, 2015, 2016, 2018 and was applied to create the variables of 1) the regional road adjacency and 2) intersection density in ArcGIS.
Open space data	*Source*: The Geospatial Centre at the University of Waterloo *Description*: This dataset contains the park layers of 2011, 2013, 2015 in Waterloo and the park layers from 2009 to 2018 in Kitchener, as well as the regional forests, cemeteries, and golf courses in the Region of Waterloo in 2018. The data was combined and then used to calculate 1) open space adjacency and 2) open space access within 800 metres for each parcel in ArcGIS.
CTC boundary layer	*Source*: Region of Waterloo *Description*: The CTC Analytical Boundary was created by the Region for planning purpose, and was used for this study to create the dummy variable *in_ctc*
Census tract layers	*Source*: Census 2006, Census 2011, Census 2016 from Statistics Canada *Description*: The dataset was used to obtain population density and education rate in each census tract.

[a]Note that we joined the parcel with the transit data based on their transaction year. For instance, the parcels of sales in 2006 were joined with the bus transit data in 2005.

A1.3 Descriptive statistics of variables

Table A-2: Descriptive statistics of variables

Statistic	Mean	St. Dev.	Min	Max
sale_amt_adj	230,364	93,774	13,486	1,313,000
logprice	12.282	0.351	9.509	14.088
age	22.905	12.440	0	61
tot_area	1.133	0.314	0.397	3.302
lot_size	0.002	0.073	0	5
beds	2.330	0.763	0	6
baths	1.534	0.532	0	7.5
garage	0.440	0.569	0	3
storey	1.577	0.601	1	3
fireplace	0.177	0.417	0	3
inter_dense	32.312	15.610	6.963	77.091
dis_bus	1.648	1.391	0.108	14.459
rd_adj	0.467	0.499	0	1
os_adj	0.347	0.476	0	1
os_area	1.098	1.044	0.024	7.220
in_ctc	0.217	0.412	0	1
edu_rate	51.711	8.285	25.075	72.825
pop_dense	2.561	1.077	0.265	5.175
sale_year	2,011	3.951	2,005	2,018

(a) Condo housing (N = 15,364)

Statistic	Mean	St. Dev.	Min	Max
sale_amt_adj	394,867	154,603	68,403	3,553,269
logprice	12.828	0.328	11.133	15.083
age	32.109	24.460	0	201
tot_area	1.597	0.599	0.463	7.570
lot_size	0.143	0.097	0.001	2.970
beds	3.148	0.621	1	7
baths	1.980	0.702	1	7.5
garage	1.046	0.815	0	5
fireplace	0.546	0.617	0	6
pool	0.054	0.225	0	1
inter_dense	32.612	11.263	1	82
dis_bus	2.401	2.441	0.077	15.000
rd_adj	0.075	0.263	0	1
os_adj	0.172	0.377	0	1
os_area	0.949	0.858	0.000	20.418
in_ctc	0.111	0.314	0	1
edu_rate	54.182	8.309	25.075	72.825
pop_dense	2.276	1.118	0.124	5.175
sale_year	2,011	3.895	2,005	2,018

(b) Single-detached housing (N = 41,272)

Statistic	Mean	St. Dev.	Min	Max
sale_amt_adj	284,701	74,820	71,150	1,200,000
logprice	12.531	0.229	11.173	13.998
age	32.674	24.534	0	165
tot_area	1.280	0.304	0.672	4.166
lot_size	0.096	0.091	0.003	4.933
beds	3.110	0.746	1	6
baths	1.670	0.561	0	5
garage	0.484	0.617	0	5
fireplace	0.128	0.346	0	2
pool	0.005	0.073	0	1
inter_dense	33.915	11.245	9	82
dis_bus	1.546	1.295	0.076	15
rd_adj	0.159	0.365	0	1
os_adj	0.108	0.310	0	1
os_area	0.833	0.709	0.004	4.245
in_ctc	0.150	0.357	0	1
edu_rate	51.378	7.846	0.000	71.366
pop_dense	2.715	1.123	0.124	5.175
sale_year	2,010	3.917	2,005	2,018

(c) Semi-detached/duplex housing (N = 7,076)

Statistic	Mean	St. Dev.	Min	Max
sale_amt_adj	313,430	72,086	95,192	898,465
logprice	12.632	0.214	11.464	13.708
age	10.405	9.195	0	112
tot_area	1.383	0.195	0.750	3.051
lot_size	0.062	0.028	0.013	0.627
beds	2.949	0.351	0	5
baths	1.865	0.547	0	4
garage	0.941	0.441	0	2
fireplace	0.145	0.366	0	2
pool	0.002	0.047	0	1
inter_dense	29.496	9	7	76
dis_bus	2.485	2.569	0.102	15
rd_adj	0.083	0.276	0	1
os_adj	0.197	0.398	0	1
os_area	1.155	0.873	0.183	3.720
in_ctc	0.021	0.142	0	1
edu_rate	58.063	7.275	37.1	71.0
pop_dense	1.845	1.203	0.124	4.640
sale_year	2,011	3.834	2,005	2,018

(d) Townhouses (N = 4,546)

A1.4 Correlation matrix

Table A-3: Correlation Matrix for the single-detached housing observations

	logprice	age	tot_area	lot_size	beds	baths	garage	fireplace	pool	os_adj	os_area	rd_adj	in_ctc	dis_bus	inter_dense	edu_rate	pop_dense
logprice	1	-0.40	0.77	0.33	0.38	0.64	0.60	0.45	0.18	0.21	0.32	-0.09	-0.19	0.38	-0.30	0.52	-0.32
age	-0.40	1	-0.41	0.11	-0.14	-0.46	-0.34	-0.12	-0.003	-0.13	-0.40	0.11	0.52	-0.34	0.55	-0.40	0.32
tot_area	0.77	-0.41	1	0.26	0.51	0.68	0.51	0.42	0.15	0.22	0.31	-0.06	-0.16	0.41	-0.25	0.43	-0.33
lot_size	0.33	0.11	0.26	1	0.13	0.16	0.16	0.27	0.16	0.10	0.01	0.08	-0.05	0.12	-0.16	0.01	-0.08
beds	0.38	-0.14	0.51	0.13	1	0.38	0.22	0.23	0.11	0.09	0.10	-0.04	-0.06	0.14	-0.11	0.16	-0.11
baths	0.64	-0.46	0.68	0.16	0.38	1	0.48	0.40	0.12	0.18	0.28	-0.06	-0.20	0.30	-0.28	0.40	-0.27
garage	0.60	-0.34	0.51	0.16	0.22	0.48	1	0.29	0.10	0.13	0.19	-0.05	-0.19	0.27	-0.25	0.36	-0.22
fireplace	0.45	-0.12	0.42	0.27	0.23	0.40	0.29	1	0.18	0.15	0.10	-0.04	-0.13	0.11	-0.15	0.17	-0.09
pool	0.18	-0.003	0.15	0.16	0.11	0.12	0.10	0.18	1	0.04	0.02	0.01	-0.05	0.02	-0.06	0.02	-0.02
os_adj	0.21	-0.13	0.22	0.10	0.09	0.18	0.13	0.15	0.04	1	0.13	-0.07	-0.09	0.09	-0.14	0.11	-0.09
os_area	0.32	-0.40	0.31	0.01	0.10	0.28	0.19	0.10	0.02	0.13	1	-0.03	-0.23	0.28	-0.37	0.47	-0.38
rd_adj	-0.09	0.11	-0.06	0.08	-0.04	-0.06	-0.05	-0.04	0.01	-0.07	-0.03	1	0.02	-0.12	0.04	-0.04	0.07
in_ctc	-0.19	0.52	-0.16	-0.05	-0.06	-0.20	-0.19	-0.13	-0.05	-0.09	-0.23	0.02	1	-0.14	0.44	-0.16	0.05
dis_bus	0.38	-0.34	0.41	0.12	0.14	0.30	0.27	0.11	0.02	0.09	0.28	-0.12	-0.14	1	-0.25	0.24	-0.39
inter_dense	-0.30	0.55	-0.25	-0.16	-0.11	-0.28	-0.25	-0.15	-0.06	-0.14	-0.37	0.04	0.44	-0.25	1	-0.17	0.37
edu_rate	0.52	-0.40	0.43	0.01	0.16	0.40	0.36	0.17	0.02	0.11	0.47	-0.04	-0.16	0.24	-0.17	1	-0.36
pop_dense	-0.32	0.32	-0.33	-0.08	-0.11	-0.27	-0.22	-0.09	-0.02	-0.09	-0.38	0.07	0.05	-0.39	0.37	-0.36	1

A2 Appendices in Chapter 3

A2.1 Hedonic demand identification

The earlier work on hedonic demand dates back to the two-stage model proposed by Rosen (1974) to estimate the demand function for each housing attribute x.

Stage 1 - estimates a nonlinear hedonic function $P = P(x, \epsilon)$, and then calculates the implicit marginal price for each housing attribute x, $P_x = \partial P / \partial x$, which is equal to the marginal willingness to pay ($MWTP$) for that attribute on equilibrium, $MWTP = P_x$.

Stage 2 - regresses the estimated $MWTPs$ of all buyers on (i) the quantities (or quality level) of the attributes that they consume, (ii) income and (iii) taste related variables, $MWTP = f(x, income, tastes, \varepsilon)$

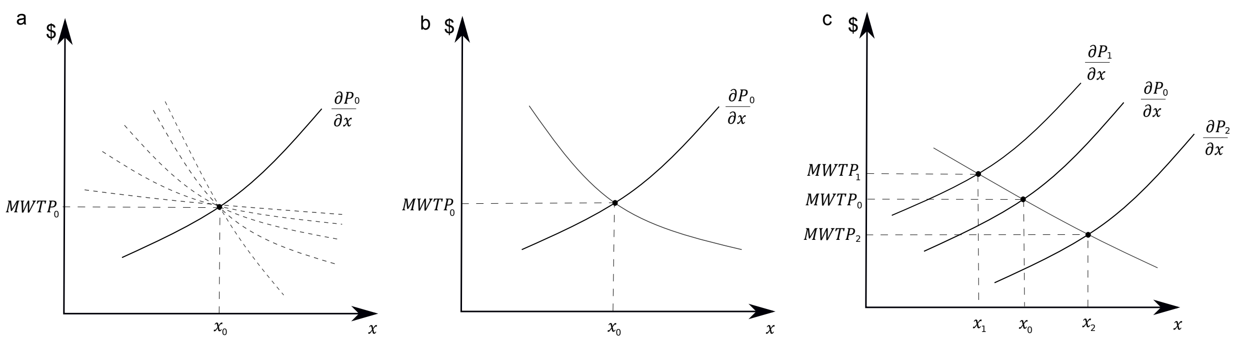


Figure A-2: Illustration of hedonic demand identification

Note: (a) shows that only one observation in a single market is not sufficient for deriving the demand curve of a household, since an infinite set of curves might go through the point. (b) shows the structural hedonic method that we employ in this paper, which imposes restrictions on the shape of demand curve or structure of preferences. (c) shows the identification methods when multiple points in the demand curve are available.

This approach seems to provide a straightforward way to estimate a global demand function for each attribute. However, on one hand, an endogeneity problem makes the identification challenging. When the hedonic price schedule is nonlinear, the error term ε in the second stage regression, which represents unexplained variations in $MWTP$ or unexplained tastes, is likely to be correlated with quantities of the attribute x chosen by households. Thus, x would be endogenous in demand function $MWTP = f(x, income, tastes, \varepsilon)$. Implementing Rosen's two-stage model often involves finding valid instruments to address the endogeneity problem. On the other hand, as shown in Figure A-2-a, only one choice of each household in a single market is not sufficient to derive the demand curve for that household. Even if a nonlinear hedonic price equation (such as quadratic hedonic model) provides varying marginal prices (as shown in the solid line of Figure A-2-a), an infinite set of demand curves (the dash lines) might interact with the marginal price curve through the observed point x_0. Without additional information about demand, it is not possible to recover demand curves by using Rosen's two-stage model.

Figure A-2-b and A-2-c illustrate the two methods that are applied in recent hedonic demand studies. Figure A-2-b shows the method that recovers demand by restricting the shape of demand curves: referred to as structural hedonic models (Kuminoff et al., 2013). Bajari and Benkard (2005) propose a 3-step approach for estimating the structural hedonic model, who assume a known parametric utility form (quasi-linear in their paper) to identify preference parameters. Bajari and Kahn (2005) then apply the same method for housing demand estimation. This method is particularly applicable to the case that only one choice is observed for each household in a single market.

While the preference parameters recovered from structural hedonic models depend on the function-form assumption, Bajari and Benkard (2005) point out that those assumptions can be relaxed when multiple purchase choices of each individual are available. Figure A-2-c exactly shows the identification method that collects multiple choices for each household, so as to trace out linear or nonlinear demand curves. Repeated choice observations of each household either from panel data (Bishop and Timmins, 2018) or "before" and "after" an exogenous market shock or supply shift (Kuminoff and Pope, 2012), or observing choices of households from different markets but with common preference structure (Bartik, 1987) can derive demand curves by analyzing the changes in the gradient of hedonic price functions.

Considering that we only observe one location choice for each homebuyer, this study employs the structural hedonic demand model and follows Bajari and Kahn (2005)'s approach to recover preference parameters. Although this method is different from Rosen (1974)'s original two-stage demand model, it still builds on the first-stage hedonic estimates and then moves further to a second-stage demand model for preference identification.

A2.2 Survey sampling

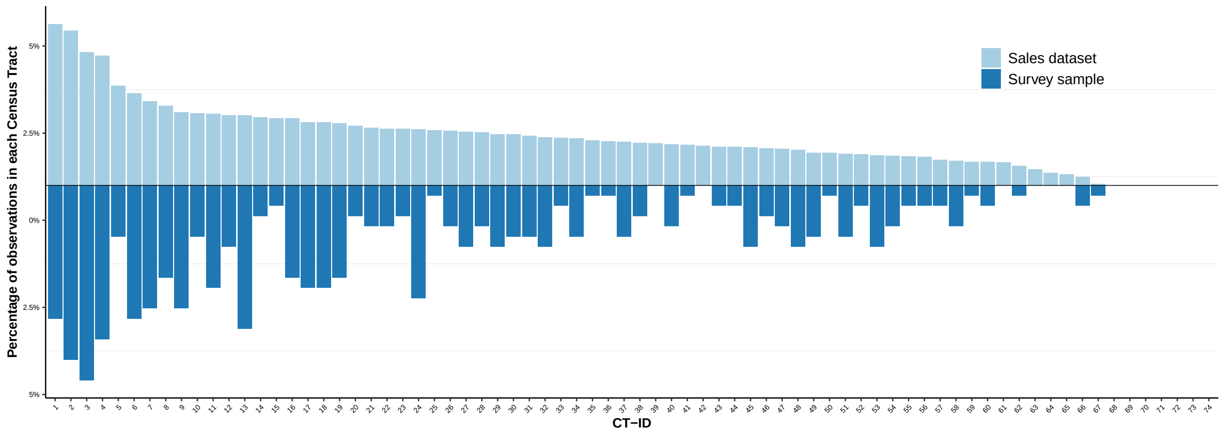


Figure A-3: The percentage of observations in each census tract from the survey sample and the sales dataset, respectively
Source: 1) Residential location choice survey in Kitchener Waterloo, Canada. 2) MPAC sales transaction dataset

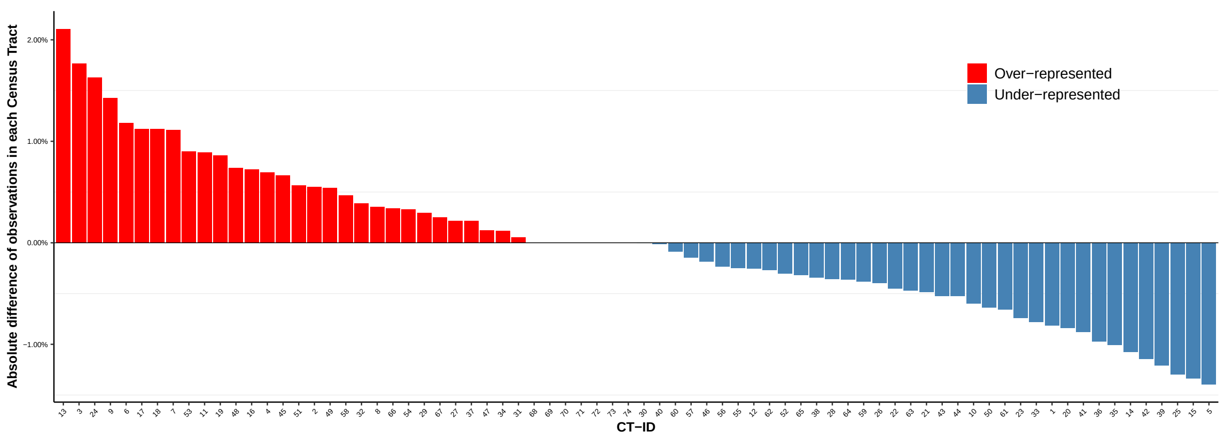


Figure A-4: The absolute difference of the percentage of observations in each census tract between the survey sample and the sales dataset
Source: 1) Residential location choice survey in Kitchener Waterloo, Canada. 2) MPAC sales transaction dataset

Figure A-3 compares the observation percentage in each census tract between the survey sample (the half down) and the sales dataset (the half up). Even though our survey sample captures observations in most CTs in KW as the population dataset, we still notice the distribution differences between them. Figure A-4 shows that our sample has over-represented almost half of the CTs and under-represented the other half. To further explore where the differences are, we plot out the map in Figure A-5, which shows that most of the under-represented CTs are concentrated in the inner urban area, and most of the over-represented CTs are mainly in the suburban area. This is further confirmed by Figure A-6, where single-detached houses are over-represented in the survey sample, while condominium housing units are under-represented. Further, the average sales price in our survey sample is a bit higher than the sales dataset, indicating that our survey has captured fewer housing units of lower-income households.

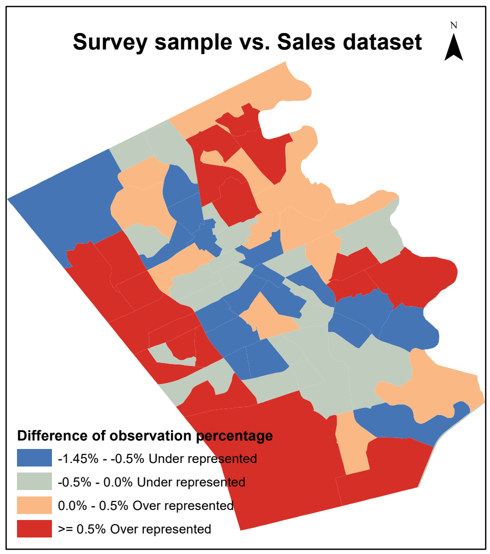


Figure A-5: The spatial distribution of the absolute difference of the observation percentage in each census tract between the survey sample and the sales dataset

Source: 1) Residential location choice survey in Kitchener Waterloo, Canada. 2) MPAC sales dataset

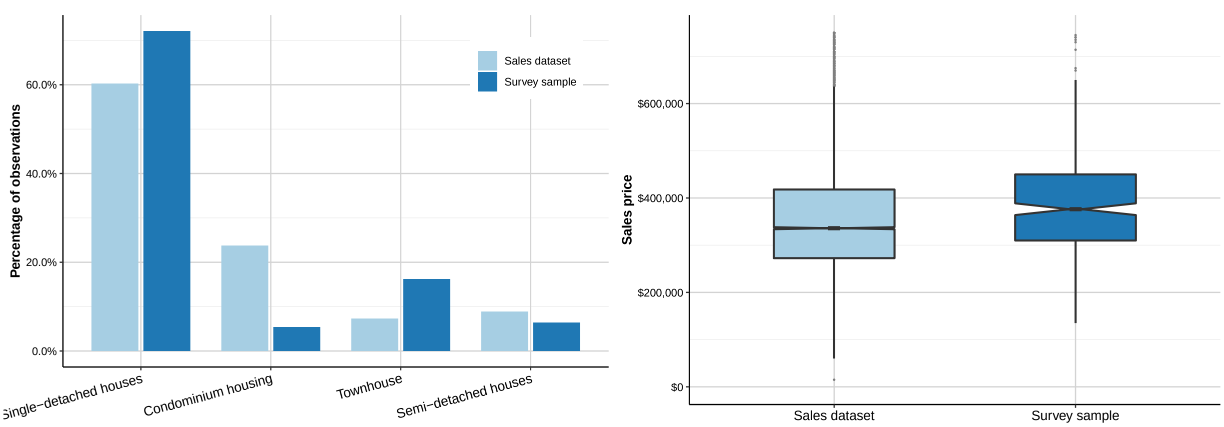


Figure A-6: Survey sample vs. sales dataset with respect to housing types and housing prices
Note:: The left figure shows the percentage of observations of different housing types from the survey sample and the sales dataset; the right figure shows the box plots of sales prices from the two dataset. *Source*: 1) Residential location choice survey in Kitchener Waterloo, Canada. 2) MPAC sales transaction dataset

A2.3 Data source

Table A-4: Variables and data source

Variables	Data source
Structural attributes (Homebuyer survey)	
$SINGLE$	Self-reported in homebuyer survey
$BEDM$	Self-reported in homebuyer survey
$BATH$	Number of full bathrooms plus half number of half-bathrooms
$GRAG$	Self-reported in homebuyer survey
$YARD$	Subtracting the property footprint from the lot size
BUL_AGE	Self-reported in homebuyer survey
$SIZE$	Self-reported in homebuyer survey
Locational and neighbourhood attributes	
OS_ACES	Calculated based on the gravity model as Babin (2016)
In_CTC	Provided by Region of Waterloo
DIS_LRT	Calculated in ArcGIS by network analysis
$In\_CTC * DIS\_LRT$	Calculated in ArcGIS by network analysis
DIS_BUS	Calculated in ArcGIS by network analysis
$POST\_EDU$	National Housing Survey at the DA level (Statistics Canada, 2016)
POP_DENS	National Housing Survey at the DA level (Statistics Canada, 2016)
OS_ADJ	Calculated in ArcGIS as explained in Babin (2016)
REG_RD_ADJ	Calculated in ArcGIS as explained in Babin (2016)
HP	Self-reported in homebuyer survey

A2.4 Linear regression results

Table A-5: Linear regression results

	log(HP)
SINGLE	0.183*** (0.032)
SQFT 1000 - 1499	0.128** (0.058)
SQFT 1500 - 1999	0.146** (0.063)
SQFT 2000 - 2499	0.187*** (0.071)
SQFT 2500 - 2999	0.323*** (0.084)
SQFT More than 2999	0.441*** (0.104)
BDMS	0.048** (0.019)
BATH	0.084*** (0.022)
GRAG	0.082*** (0.022)
YARD	0.00001*** (0.000)
BUL_AGE	-0.005** (0.002)
BUL_AGE^2	0.00004** (0.00002)
POP_DENS	-0.00000 (0.00001)
OS_ACES	0.0002 (0.001)
In_CTC	0.173 (0.124)
CTC_DISLRT	-0.0002 (0.0001)
DIS_LRT	-0.00001 (0.00001)
DIS_BUS	-0.00000 (0.00005)
POST_EDU	0.005*** (0.001)
OS_ADJ	-0.017 (0.032)
REG_RD_ADJ	-0.067 (0.041)
Constant	11.899*** (0.135)
Observations	276
R^2	0.715
Residual Std. Error	0.176 (df = 254)
F Statistic	30.391*** (df = 21; 254)
Note: *p<0.1; **p<0.01;***p<0.001	

A2.5 Stated preference from the housing survey

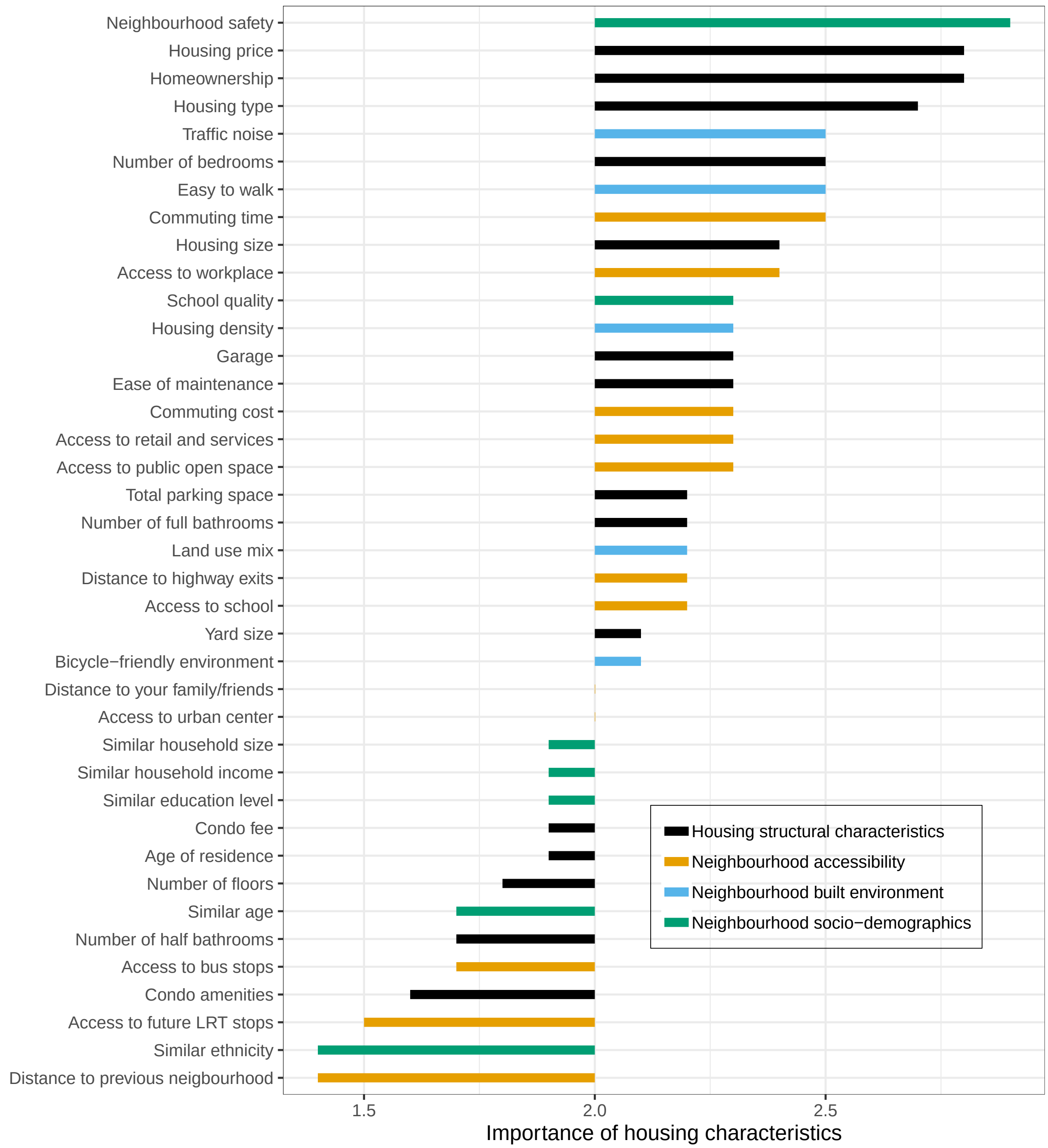


Figure A-7: Stated importance of housing attributes in location choices

Note:: This figure shows the mean importance value of each housing attribute in the sample based on the reported attitudes. "1-not important"; "2-somewhat important"; "3-very important" . *Source*: Residential location choice survey in Kitchener Waterloo, Canada, 2017.

A2.6 Correlation matrix

	SINGLE	BDMS	BATH	GRAG	YARD	BUL_AGE	POP_DENS	OS_ACES	ln_CTC	DIS_LRT	CTC_DISLRT	DIS_BUS	POST_EDU	OS_ADJ	REG_RD_ADJ
SINGLE	1.000														
BDMS	0.152	1.000													
BATH	0.127	0.245	1.000												
GRAG	0.212	0.124	0.257	1.000											
YARD	0.111	0.169	0.130	0.060	1.000										
BUL_AGE	0.406	0.054	-0.282	-0.197	0.236	1.000									
POP_DENS	-0.206	-0.050	0.107	0.029	-0.044	0.022	1.000								
OS_ACES	0.083	-0.052	0.063	-0.010	-0.064	-0.086	0.057	1.000							
ln_CTC	-0.058	-0.105	0.137	-0.042	-0.075	0.196	0.162	-0.022	1.000						
DIS_LRT	0.134	0.092	-0.020	-0.056	-0.001	-0.458	0.011	-0.148	-0.128	1.000					
CTC_DISLRT	0.047	0.067	-0.101	0.013	0.049	-0.117	-0.112	-0.020	0.914	0.001	1.000				
DIS_BUS	-0.060	-0.104	0.068	0.008	0.236	-0.065	-0.194	-0.218	0.085	0.294	-0.044	1.000			
POST_EDU	0.011	-0.018	0.248	0.015	0.027	-0.189	-0.189	0.212	0.088	-0.122	-0.019	0.193	1.000		
OS_ADJ	-0.236	-0.009	0.082	0.094	0.085	0.155	-0.190	0.138	0.024	0.031	-0.037	0.011	-0.038	1.000	
REG_RD_ADJ	-0.187	-0.165	-0.034	0.057	0.039	0.057	-0.014	0.030	-0.105	0.057	0.129	0.015	-0.062	0.017	1.000

Figure A-8: Partial correlation matrix of explanatory variables

A3 Appendices in Chapter 4

A3.1 Estimated class probabilities from the LCA

Table A-6: Estimated class probabilities of each survey respondent - LCA results

Id	Class	Class 1 probability	Class 2 probability
1	Class2	0.000	1.000
2	Class2	0.048	0.952
3	Class1	0.957	0.043
4	Class2	0.000	1.000
5	Class2	0.064	0.936
6	Class1	0.865	0.135
7	Class2	0.000	1.000
8	Class2	0.000	1.000
9	Class2	0.028	0.972
10	Class1	0.990	0.010
11	Class2	0.070	0.930
12	Class1	1.000	0.000
13	Class1	0.986	0.014
14	Class2	0.004	0.996
15	Class2	0.253	0.747
16	Class2	0.037	0.963
17	Class1	0.976	0.024
18	Class2	0.055	0.945
19	Class1	0.995	0.005
20	Class1	0.584	0.416
21	Class2	0.070	0.930
22	Class1	1.000	0.000
23	Class1	0.990	0.010
24	Class2	0.003	0.997
25	Class2	0.081	0.919
26	Class1	0.943	0.057
27	Class2	0.008	0.992
28	Class2	0.017	0.983
29	Class2	0.413	0.587
30	Class2	0.004	0.996
31	Class2	0.018	0.982
32	Class2	0.025	0.975

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
33	Class1	0.709	0.291
34	Class1	1.000	0.000
35	Class2	0.001	0.999
36	Class2	0.014	0.986
37	Class2	0.004	0.996
38	Class1	0.976	0.024
39	Class2	0.000	1.000
40	Class2	0.006	0.994
41	Class2	0.227	0.773
42	Class2	0.000	1.000
43	Class2	0.001	0.999
44	Class1	1.000	0.000
45	Class2	0.003	0.997
46	Class2	0.000	1.000
47	Class2	0.253	0.747
48	Class1	1.000	0.000
49	Class2	0.320	0.680
50	Class1	0.994	0.006
51	Class1	0.993	0.007
52	Class1	0.951	0.049
53	Class2	0.145	0.855
54	Class2	0.004	0.996
55	Class2	0.003	0.997
56	Class2	0.213	0.787
57	Class1	1.000	0.000
58	Class1	0.999	0.001
59	Class2	0.049	0.951
60	Class2	0.024	0.976
61	Class2	0.000	1.000
62	Class1	0.909	0.091
63	Class1	1.000	0.000
64	Class1	1.000	0.000
65	Class2	0.051	0.949
66	Class1	0.714	0.286
67	Class2	0.000	1.000
68	Class1	0.844	0.156
69	Class1	0.761	0.239

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
70	Class1	0.960	0.040
71	Class2	0.000	1.000
72	Class2	0.032	0.968
73	Class2	0.000	1.000
74	Class2	0.142	0.858
75	Class2	0.489	0.511
76	Class2	0.095	0.905
77	Class2	0.004	0.996
78	Class2	0.133	0.867
79	Class1	0.999	0.001
80	Class2	0.001	0.999
81	Class2	0.012	0.988
82	Class2	0.001	0.999
83	Class2	0.000	1.000
84	Class2	0.000	1.000
85	Class1	1.000	0.000
86	Class2	0.138	0.862
87	Class1	1.000	0.000
88	Class2	0.001	0.999
89	Class1	0.999	0.001
90	Class2	0.003	0.997
91	Class2	0.270	0.730
92	Class2	0.000	1.000
93	Class2	0.000	1.000
94	Class1	0.923	0.077
95	Class1	0.847	0.153
96	Class2	0.005	0.995
97	Class1	0.968	0.032
98	Class2	0.001	0.999
99	Class1	0.814	0.186
100	Class1	0.631	0.369
101	Class2	0.330	0.670
102	Class2	0.008	0.992
103	Class1	1.000	0.000
104	Class2	0.001	0.999
105	Class2	0.427	0.573
106	Class1	0.983	0.017

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
107	Class1	0.609	0.391
108	Class1	1.000	0.000
109	Class1	0.996	0.004
110	Class2	0.000	1.000
111	Class2	0.123	0.877
112	Class1	0.969	0.031
113	Class1	0.998	0.002
114	Class2	0.145	0.855
115	Class2	0.198	0.802
116	Class2	0.000	1.000
117	Class2	0.004	0.996
118	Class1	0.992	0.008
119	Class2	0.000	1.000
120	Class2	0.007	0.993
121	Class2	0.000	1.000
122	Class2	0.004	0.996
123	Class1	0.846	0.154
124	Class2	0.379	0.621
125	Class2	0.094	0.906
126	Class1	0.935	0.065
127	Class1	0.991	0.009
128	Class1	1.000	0.000
129	Class2	0.001	0.999
130	Class2	0.000	1.000
131	Class1	0.998	0.002
132	Class2	0.002	0.998
133	Class2	0.315	0.685
134	Class2	0.009	0.991
135	Class1	1.000	0.000
136	Class2	0.000	1.000
137	Class1	0.509	0.491
138	Class2	0.008	0.992
139	Class1	0.948	0.052
140	Class1	0.986	0.014
141	Class1	0.994	0.006
142	Class1	0.835	0.165
143	Class2	0.025	0.975

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
144	Class1	0.841	0.159
145	Class2	0.005	0.995
146	Class2	0.002	0.998
147	Class2	0.315	0.685
148	Class2	0.048	0.952
149	Class2	0.448	0.552
150	Class2	0.123	0.877
151	Class1	0.992	0.008
152	Class1	1.000	0.000
153	Class2	0.320	0.680
154	Class2	0.228	0.772
155	Class1	1.000	0.000
156	Class1	0.852	0.148
157	Class2	0.010	0.990
158	Class2	0.000	1.000
159	Class1	1.000	0.000
160	Class1	0.968	0.032
161	Class1	0.994	0.006
162	Class2	0.216	0.784
163	Class1	0.885	0.115
164	Class2	0.320	0.680
165	Class2	0.000	1.000
166	Class1	0.959	0.041
167	Class1	1.000	0.000
168	Class1	0.999	0.001
169	Class1	0.994	0.006
170	Class1	0.994	0.006
171	Class1	0.761	0.239
172	Class2	0.001	0.999
173	Class2	0.000	1.000
174	Class2	0.000	1.000
175	Class2	0.028	0.972
176	Class2	0.025	0.975
177	Class1	1.000	0.000
178	Class1	1.000	0.000
179	Class2	0.004	0.996
180	Class1	1.000	0.000

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
181	Class2	0.059	0.941
182	Class2	0.011	0.989
183	Class2	0.459	0.541
184	Class2	0.000	1.000
185	Class2	0.065	0.935
186	Class2	0.000	1.000
187	Class2	0.000	1.000
188	Class2	0.005	0.995
189	Class2	0.315	0.685
190	Class2	0.000	1.000
191	Class1	0.963	0.037
192	Class2	0.068	0.932
193	Class1	0.688	0.312
194	Class2	0.016	0.984
195	Class2	0.002	0.998
196	Class2	0.000	1.000
197	Class2	0.028	0.972
198	Class2	0.004	0.996
199	Class2	0.000	1.000
200	Class2	0.116	0.884
201	Class2	0.017	0.983
202	Class1	1.000	0.000
203	Class2	0.462	0.538
204	Class2	0.000	1.000
205	Class1	1.000	0.000
206	Class1	1.000	0.000
207	Class2	0.002	0.998
208	Class1	0.614	0.386
209	Class2	0.001	0.999
210	Class2	0.002	0.998
211	Class2	0.254	0.746
212	Class1	1.000	0.000
213	Class1	1.000	0.000
214	Class2	0.000	1.000
215	Class2	0.025	0.975
216	Class1	0.976	0.024
217	Class2	0.000	1.000

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
218	Class2	0.315	0.685
219	Class1	0.562	0.438
220	Class1	0.849	0.151
221	Class2	0.000	1.000
222	Class2	0.004	0.996
223	Class2	0.008	0.992
224	Class1	1.000	0.000
225	Class2	0.000	1.000
226	Class2	0.000	1.000
227	Class1	1.000	0.000
228	Class1	0.972	0.028
229	Class1	1.000	0.000
230	Class1	0.996	0.004
231	Class2	0.011	0.989
232	Class2	0.466	0.534
233	Class2	0.142	0.858
234	Class2	0.000	1.000
235	Class1	0.873	0.127
236	Class2	0.029	0.971
237	Class2	0.029	0.971
238	Class2	0.000	1.000
239	Class2	0.024	0.976
240	Class2	0.000	1.000
241	Class1	0.970	0.030
242	Class1	1.000	0.000
243	Class2	0.320	0.680
244	Class2	0.000	1.000
245	Class1	1.000	0.000
246	Class2	0.001	0.999
247	Class1	1.000	0.000
248	Class2	0.001	0.999
249	Class1	0.968	0.032
250	Class1	0.998	0.002
251	Class2	0.001	0.999
252	Class2	0.413	0.587
253	Class1	1.000	0.000
254	Class2	0.490	0.510

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
255	Class1	1.000	0.000
256	Class2	0.051	0.949
257	Class2	0.441	0.559
258	Class2	0.213	0.787
259	Class1	0.996	0.004
260	Class2	0.044	0.956
261	Class1	0.915	0.085
262	Class1	0.998	0.002
263	Class1	0.852	0.148
264	Class2	0.000	1.000
265	Class2	0.036	0.964
266	Class2	0.008	0.992
267	Class1	0.942	0.058
268	Class2	0.496	0.504
269	Class2	0.143	0.857
270	Class1	0.710	0.290
271	Class2	0.068	0.932
272	Class2	0.000	1.000
273	Class1	0.996	0.004
274	Class1	0.937	0.063
275	Class1	1.000	0.000
276	Class1	0.754	0.246
277	Class2	0.000	1.000
278	Class1	0.950	0.050
279	Class2	0.000	1.000
280	Class1	0.998	0.002
281	Class2	0.003	0.997
282	Class1	0.935	0.065
283	Class2	0.000	1.000
284	Class2	0.028	0.972
285	Class2	0.228	0.772
286	Class1	1.000	0.000
287	Class2	0.000	1.000
288	Class1	0.953	0.047
289	Class2	0.000	1.000
290	Class2	0.016	0.984
291	Class2	0.000	1.000

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
292	Class2	0.015	0.985
293	Class1	0.953	0.047
294	Class1	1.000	0.000
295	Class2	0.001	0.999
296	Class1	0.977	0.023
297	Class1	1.000	0.000
298	Class1	0.754	0.246
299	Class1	1.000	0.000
300	Class1	1.000	0.000
301	Class1	1.000	0.000
302	Class1	0.951	0.049
303	Class1	0.993	0.007
304	Class1	1.000	0.000
305	Class2	0.000	1.000
306	Class2	0.001	0.999
307	Class2	0.008	0.992
308	Class1	0.909	0.091
309	Class2	0.008	0.992
310	Class1	0.715	0.285
311	Class2	0.018	0.982
312	Class1	0.578	0.422
313	Class2	0.007	0.993
314	Class2	0.010	0.990
315	Class2	0.008	0.992
316	Class2	0.015	0.985
317	Class1	0.935	0.065
318	Class1	0.971	0.029
319	Class1	0.999	0.001
320	Class2	0.000	1.000
321	Class2	0.404	0.596
322	Class2	0.000	1.000
323	Class1	0.950	0.050
324	Class1	0.525	0.475
325	Class1	0.990	0.010
326	Class1	1.000	0.000
327	Class2	0.000	1.000
328	Class1	0.976	0.024

Continued on next page

Table A-6 – continued from previous page

Id	Class	Class 1 probability	Class 2 probability
329	Class2	0.049	0.951
330	Class2	0.001	0.999
331	Class2	0.457	0.543
332	Class2	0.079	0.921
333	Class2	0.048	0.952
334	Class1	0.775	0.225
335	Class2	0.088	0.912
336	Class2	0.227	0.773
337	Class2	0.001	0.999
338	Class2	0.000	1.000
339	Class2	0.004	0.996
340	Class1	0.957	0.043
341	Class2	0.003	0.997
342	Class1	1.000	0.000
343	Class2	0.003	0.997
344	Class1	0.520	0.480
345	Class1	0.968	0.032
346	Class1	0.999	0.001
347	Class1	0.844	0.156
348	Class2	0.037	0.963
349	Class1	0.903	0.097
350	Class2	0.049	0.951
351	Class1	1.000	0.000
352	Class2	0.008	0.992
353	Class2	0.009	0.991
354	Class1	1.000	0.000
355	Class1	0.584	0.416
356	Class2	0.009	0.991
357	Class1	0.947	0.053

A3.2 Density plots of class probabilities

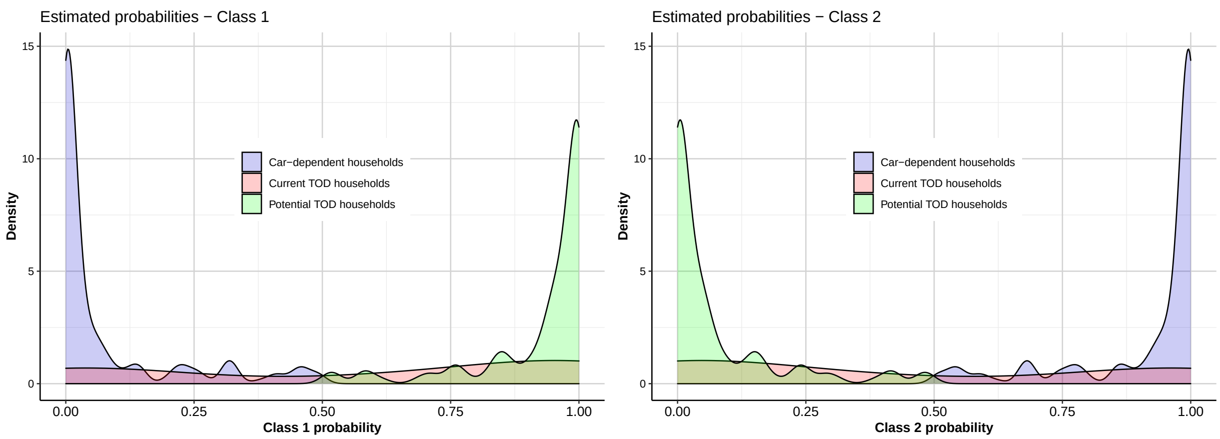


Figure A-9: Density plots of class probabilities for the three groups

Note: The two plots show the kernel density estimates of the class probabilities for households of the three groups. Not surprisingly, the car-dependent households are estimated to have higher probabilities of being in Class 2; the potential TOD households have higher probabilities of being in Class 1; and the current TOD households show a mix of probabilities being Class 1 or 2.

A4 R code for Chapter 2

```
setwd("C:/Users/y377huan/Desktop/HSAR/1.Single")

library(spdep)
library(HSAR)

#1.Read data, n = 41437
single_pnt <- readRDS(file = "single_pnt.RDS")

# Order by ctuid - important for ct and parcel index connection
single_pnt <- single_pnt[order(single_pnt$CTUID_2016),]

#delete NAs, n = 41272 (delete 165 rows)
single_pnt_new <- single_pnt[single_pnt$inst_num_add %notin% delete_instNum,]
saveRDS(single_pnt_new, file = "single_pnt_new.RDS")

#2. Create weight matrix
## 2.1 Create Delta matrix (n*m)
ct_ply_all <- st_read("C:/Users/y377huan/Desktop/inla_all/data/ct_poly_all.shp",
    stringsAsFactors =FALSE)
ct_ply_all <- ct_ply_all[order(ct_ply_all$CTUID),]## 74 CTs in total
ct_ply_all <- as(ct_ply_all, "Spatial") # transfer to polygon points
```

```
plot(ct_ply_all,border="green")
saveRDS(ct_ply_all, "ct_ply_all.RDS")

ct_ply <- readRDS(file="C:/Users/y377huan/Desktop/HSAR/RDS_data_all/ct_ply.RDS")
plot(ct_ply,border="green")
plot(single_pnt_new,col="red",pch=16, cex=0.1, add=TRUE)

# 69 CTs: length(unique(single_pnt_new$CTUID_2016))
MM <- as.data.frame(table(single_pnt_new$CTUID_2016))
Utotal <- dim(MM)[1]
Unum <- MM[,2]
Uid <- rep(c(1:Utotal),Unum)

deletCT <- as.character(MM[MM$Freq==0,]$Var1) ## delete "5410101.02"

MM <- MM[!MM$Var1 %in% deletCT,]
Utotal <- dim(MM)[1] ##69 CTs
Unum <- MM[,2]
Uid <- rep(c(1:Utotal),Unum)

## ct and id connection
# ct_id <- cbind(ct_index=as.vector(c(1:Utotal)), MM)

#Delta matrix for random effects
n <- nrow(single_pnt_new)
Delta <- matrix(0,nrow=n,ncol=Utotal)
for(i in 1:Utotal) {
  Delta[Uid==i,i] <- 1
}
rm(i)
Delta <- as(Delta,"dgCMatrix") ## 41272 * 69
saveRDS(Delta, "Delta_single.RDS")

## 2.2 Create weight matrix - M
# extract the CT-level spatial weights matrix using the queen's rule
ct_ply_single <- ct_ply[!ct_ply$CTUID %in% deletCT,] ## 69 features
saveRDS(ct_ply_single, "ct_ply_single.RDS")

nb.ct <- poly2nb(ct_ply_single)
list.ct <- nb2listw(nb.ct,style = "W", zero.policy = TRUE)
mat.ct <- listw2mat(list.ct)

M <- as(mat.ct,"dgCMatrix") #69*69
saveRDS(M, "M_single.RDS")
```

```r
## 2.3 Create weight matrix Ws
# check spatial lag (distance)
v <- gstat::variogram(log(sale_amt_adj)~1, single_pnt_new, cutoff=20000, width=20000/40)
plot(v, main="Variogram - default", xlab = "Separation distance (m)")
m.sph <- gstat::vgm(psill = 0.1, model = "Sph", range = 5000, nugget = 0.05)
gstat::fit.variogram(v, gstat::vgm(c("Exp", "Mat", "Sph")))

# find nearest 10 neighbours
nb10 <- knn2nb(knearneigh(single_pnt_new_only, k=10))
saveRDS(nb10, "nb10_single.RDS")
list10 <- nb2listw(nb10, style="W", zero.policy = TRUE)
library(Matrix)
Ws10 <- as(list10,"CsparseMatrix") ##41272 * 41272
saveRDS(Ws10, "Ws10_single.RDS")

########below code in Compute Canada server###########

# find neighbours within a radius of 2500 meters
nb2500 <- dnearneigh(single_pnt_new_only, 0, 2500)
saveRDS(nb2500, "nb2500_single.RDS")

# calculate distance
dlist0 <- nbdists(nb2500,single_pnt)
saveRDS(dlist0, "dlist0_single.RDS")
dlist1 <- dlist0

# Ws - distance decay (exponentially decay function)
dlist1 <- lapply(dlist0, function(x) exp(-0.5*(x/2500)^2))
saveRDS(dlist1, "dlist1_single.RDS")

# row-standardized spatial weight matrix
list1 <- nb2listw(nb2500, glist=dlist1, style="W", zero.policy = TRUE)
Ws <- as(listw2mat(list1),"dgCMatrix")
saveRDS(Ws, "Ws_single.RDS")

## 2.4 get sales time data
saledate <- single_pnt_new$sale_YearMon
saleyear <- single_pnt_new$sale_Year
salemonth <- single_pnt_new$sale_Month
instNum <- single_pnt_new$inst_num_add
ct_id <- single_pnt_new$CTUID_2016

saletime <- cbind(instNum,saleyear,salemonth, ct_id) %>%
      as.data.frame()
saveRDS(saletime, "saletime_new_single.RDS")
```

```
## 2.5 Create space-time matrix based on space matrix
# calculate distance
dlist2 <- dlist1

for (i in seq(along=nb2500))
dlist2[[i]] <- ifelse(instNum[i] > instNum[nb2500[[i]]] &
                        ((saleyear[i]-saleyear[nb2500[[i]]])*12 + (salemonth[i] -
                            salemonth[nb2500[[i]]])) <=3 &
                        ((saleyear[i]-saleyear[nb2500[[i]]])*12 + (salemonth[i] -
                            salemonth[nb2500[[i]]])) >=0,
                      dlist1[[i]]/((saleyear[i]-saleyear[nb2500[[i]]])*12 + (salemonth[i]
                          - salemonth[nb2500[[i]]]) + 1), 0)

saveRDS(dlist2, "dlist2_single.RDS")

# row-standardized spatial weight matrix
list2 <- nb2listw(nb2500, glist=dlist2, style="W", zero.policy = TRUE)
Wst <- as(listw2mat(list2),"dgCMatrix")
saveRDS(Wst, "Wst_single.RDS")

table(round(rowSums(Wst)))
#######above code in server##########

#3.Prepare data for modelling

single_dat <- data.frame(
  instNum = single_pnt_new$inst_num_add,
  Sale_amt_adj = single_pnt_new$sale_amt_adj,
  logPrice = log(single_pnt_new$sale_amt_adj),
  sale_Year = single_pnt_new$sale_Year,
  age = single_pnt_new$sale_Year - single_pnt_new$yrblteff,
  tot_area = single_pnt_new$area_tot/1000,
  lot_size = single_pnt_new$eff_ltsz_add,
  frontage = single_pnt_new$eff_fr_add,
  beds = single_pnt_new$bedrooms,
  baths = single_pnt_new$baths,
  garage = single_pnt_new$gara_add,
  str_quality = single_pnt_new$quality,
  fireplace = single_pnt_new$fireplcs,
  pool = single_pnt_new$pool_add,
  inter_dense = single_pnt_new$INTERSEC_DENS,
  ave_roa = single_pnt_new$Ave_ROA,
  dis_bus = single_pnt_new$Dis_bus/100,
  rd_adj = single_pnt_new$Rd_adj,
```

```r
  os_adj = single_pnt_new$Os_adj,
  os_area = single_pnt_new$Os_area/1000000,
  in_ctc = single_pnt_new$InCTC,
  edu_rate = single_pnt_new$Edu_rate,
  pop_dense = single_pnt_new$Pop_dense/1000,
  intDens_os = single_pnt_new$interDens_os/1000000,
  intDens_bus = single_pnt_new$interDens_bus/100,
  intDens_ctc = single_pnt_new$interDens_ctc,
  ctc_os = single_pnt_new$ctc_os/1000000)
single_dat$age2 <- single_dat$age*single_dat$age

saveRDS(single_dat, "single_dat.RDS")
table(complete.cases(single_dat)) ##41272

######Run linear models######
lm_model <- lm(formula = f1,data = single_dat)
s <- summary(lm_model)
save(s, file="lm_model_single.RData")

# test spatial dependence based on 10 nearest neigbhours - 0.29
moran_single <- lm.morantest(lm_model, listw = list10)

# test spatial dependence based on space-time matrix
listw2 <- readRDS("listw2.RDS")
moran_single <- lm.morantest(lm_model, listw = listw2)
saveRDS(moran_single, file="moran_single.RDS")

#4.Run models using the HSAR package

# model formula
f1 <- logPrice ~ age + tot_area + lot_size + baths + beds + garage + fireplace + pool + os_adj +
  os_area + rd_adj + in_ctc + dis_bus + inter_dense + edu_rate + pop_dense +
  inter_dense*os_area + inter_dense*dis_bus + inter_dense*in_ctc + in_ctc*os_area +
  factor(sale_Year) + in_ctc*factor(sale_Year)

# read weight matrix
single_dat <- readRDS("single_dat.RDS")
Ws <- readRDS("Ws_single.RDS") # 41272*41272
Wst <- readRDS("Wst_single.RDS") # 41272*41272
M <- readRDS("M_single.RDS") # 69*69
Delta <- readRDS("Delta_single.RDS") #41272*69

# parameters
betas= coef(lm(formula=f1, data=single_dat))
```

```
pars_SAR=list( rho = 0.5, sigma2e = 2.0, betas = betas)
pars_HSAR=list( rho = 0.5,lambda = 0.5, sigma2e = 2.0, sigma2u = 2.0, betas = betas)

# Model 1 - SAR
res11_single_3chain <- HSAR::sar(f1, data = single_dat, W=Ws,
                                 burnin=5000, Nsim=10000, thinning = 3, parameters.start=
                                    pars_SAR)
summary(res11_single_3chain)
saveRDS(res11_single_3chain, file = "res11_single_3chain.RDS")

# Model 2 - SAR+MLM
res12_single_3chain <- HSAR::hsar(f1, data = single_dat, W=Ws,M=M,Delta=Delta,
                                  burnin=5000, Nsim=10000, thinning = 3, parameters.start=
                                     pars_HSAR)
summary(res12_single_3chain)
saveRDS(res12_single_3chain, file = "res12_single_3chain.RDS")

# Model 3 - STAR
res21_single_3chain <- HSAR::sar(f1, data = single_dat, W=Wst,
                                 burnin=5000, Nsim=10000, thinning = 3, parameters.start=
                                    pars_SAR)
summary(res21_single_3chain)
saveRDS(res21_single_3chain, file = "res21_single_3chain.RDS")

# Model 4 - STAR+MLM
res22_single_3chain <- HSAR::hsar(f1, data = single_dat, W=Wst,M=M,Delta=Delta,
                                  burnin=5000, Nsim=10000, thinning = 3, parameters.start=
                                     pars_HSAR)
summary(res22_single_3chain)
saveRDS(res22_single_3chain, file = "res22_single_3chain2.RDS")

# report results

#STAR + MLM
x <- as.numeric(res22_single$Mus)
ct_ply_single$Mus_star <- x
ct_single_sf <- st_as_sf(ct_ply_single)
#plot(ct_single_sf)

library(RColorBrewer)
pal <- brewer.pal(4,"OrRd")

plot(ct_single_sf["Mus_star"],
     main = "STAR + ML model",
     breaks = "quantile", nbreaks = 4, border="grey40",
```

```
     pal = pal)

# SAR + MLM
x <- as.numeric(res12_single$Mus)
ct_ply_single$Mus_sar <- x
ct_single_sf <- st_as_sf(ct_ply_single)
#plot(ct_single_sf)

library(RColorBrewer)
pal <- brewer.pal(4,"OrRd")

plot(ct_single_sf["Mus_sar"],
     main = "SAR + ML model",
     breaks = "quantile", nbreaks = 4, border="grey40",
     pal = pal)

# calculate p values of t test
xbar <- res22_single$Mbetas
se <- res22_single$SDbetas
t_stat <- xbar/se
pvalue <- 2*pt(-abs(t_stat), df=41271)
round(pvalue,4)

# calculate confidence interval for rho and lamda
t_stat_rho <- res22_single$Mrho/res22_single$SDrho
pvalue_rho <- 2*pt(-abs(t_stat_rho), df=41271)

t_stat_lambda <- res22_single$Mlambda/res22_single$SDlambda
pvalue_lambda <- 2*pt(-abs(t_stat_lambda), df=41271)
```

A5 R code for Chapter 3

```
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(dev = 'pdf')
libs <- c('tidyverse','ggplot2','dplyr', 'ggpubr', 'latticeExtra', 'gridExtra', 'MASS',
          'colorspace', 'plyr', 'Hmisc', 'scales', 'lattice','ggthemes','gmodels',
          'magrittr','stargazer','tidyr','scales', 'graphics', 'sjPlot', "corrplot", "np")
lapply(libs, require, character.only = T)

setwd("/Users/yukeysha/Desktop/Paper1/2018-May")
#gis-referenced 340
file1 <- read.csv("Survey-joinNHS__LRT_16April2018.csv", na.strings = c("", " ", "NA", "
    Other, please specify..."))
```

```
tb1 <- as.tibble(file1)
tb1$internal.id <- tb1$Survey_Internal.ID
#buyer-all 357
file2 <- read.csv("buyers_all.csv", na.strings = c("", " ", "NA", "Other, please specify
    ..."))
tb2 <- as.tibble(file2)
##left-join, since we need buyer's information for 340 samples
join <- left_join(tb1, tb2, by = "internal.id")
tb <- join

# Database setup
#Property structural attributes
#colnames(tb)
tb_str <- tb[,c("internal.id", "HP", "TYPE", "BDMS", "FBTH", "HBTH", "GRAG", "SIZE", "STRY
    ", "BLT_YEAR",
                "buy.impt.house.N.of.bedroom", "buy.impt.house.N.of.full.bath",
                "buy.impt.house.N.of.covered.parking",
                "buy.impt.house.yard.size", "buy.impt.house.type")]
str(tb_str)

#Square footage
addmargins(table(tb_str$SIZE))
tb_str$SQFT <- tb$SIZE
#house size - i.e., living area (= building footprint * storeys as in Robert's thesis)
tb_str$SIZE <- recode(tb_str$SIZE,
                      "Less than 1000" = 749,
                      "1000 - 1499" = 1249,
                      "1500 - 1999" = 1749,
                      "2000 - 2499" = 2249,
                      "2500 - 2999" = 2749,
                      "More than 2999" = 3249)
tb_str$SIZE <- as.integer(tb_str$SIZE)
#TYPE
tb_str$TYPE <- recode(tb_str$TYPE,
                      "Apartment in a building with 5 or more storeys" = "APT",
                      "Apartment in a building with fewer than 5 storeys" = "APT",
                      "Apartment or flat in a duplex (with an upper and lower unit in same
                          house)" = "APT",
                      "Single-detached house" = "SING",
                      "Semi-detached house" = "SEMI",
                      "Townhouse/row house" = "ROW")
#CREATE SINGLE-FAMILY HOUSE DUMMY VARIABLE
tb_str$SINGLE <- ifelse(tb_str$TYPE == "SING", 1, 0)
table(tb_str$SINGLE)
#built_year
```

```
tb_str$BUL_AGE <- (2017 - tb_str$BLT_YEAR)
tb_str$BUL_AGE <- as.integer(tb_str$BUL_AGE)
#bedroom
tb_str$BDMS <- ifelse(tb_str$BDMS == "0", NA, tb_str$BDMS)
#bathrooms
tb_str$FBTH[tb_str$FBTH == 0] <- NA
## combine full-bath and half-bath into one
tb_str <- mutate(tb_str,BATH = FBTH + (HBTH*0.5))

#ADD yard-size
#calculated based on GIS erase, 340 observations
file_yard_size <- read.csv(file = "/Users/yukeysha/Desktop/Paper1/Joined-Lot-Size/Yard
    size-Aug16-2018-YH.csv")
tb_yard_size <- as.tibble(file_yard_size)
tb_yard_size$internal.id <- tb_yard_size$X...internal_id
tb_yard_size$YARD <- tb_yard_size$Yard.size
tb_yard_size <- select(tb_yard_size, internal.id, YARD)
#Join yard size to the tibble
tb_str <- left_join(tb_str, tb_yard_size, "internal.id")

#------locational attributes-----
tb_local <- tb[,c("OS_ACES", "OS_ADJ", "REG_RD_ADJ", "DIS_LRT", "DIS_BUS", "SCHQ")]
tb_local$SCHQ <- as.numeric(levels(tb_local$SCHQ))[tb_local$SCHQ]
tb_local$DIS_BUS <- (-tb_local$DIS_BUS)

#------nghd attribities-----
##colnames(tb)
tb_nghd <- tb[,c("AVE_HHSIZE", "AVE_AGE", "AVE_DWVALUE", "POP_DENS", ##neighbourhood
    average size, age, house value, population density
                 "CM_TOTAL", "CM_TRANS",  ##transit commuters proportion
                 "CDD15", "CDD15_POSTSEC", ##education rate by calucaluting the proportion
                     of postsecondary
                 "CDD25", "CDD25_POSTSEC",
                 "CITIZENS", "CITIZEN_CAN", "CITIZEN_NOTCAN", ##immigrants
                 "IMM_TOTAL", "NON.IMMIGRANTS", "IMMIGRANTS",
                 "EMPL_RATE", "UNEMPL_RATE",##employment rate
                 "AVE_INCOM", ##nghd average income
                 "buy.impt.ngbh.access.LRT",
                 "buy.impt.ngbh.access.bus",
                 "buy.impt.ngbh.access.open.space",
                 "buy.impt.house.N.of.covered.parking",
                 "In_CTC")]
tb_nghd <- mutate(tb_nghd,
                  TRANS_CMT = (CM_TRANS/CM_TOTAL)*100,
                  POST_EDU = (CDD15_POSTSEC+CDD25_POSTSEC)/(CDD15+CDD25)*100,
```

```r
                   NON_CAN = CITIZEN_NOTCAN/CITIZENS*100,
                   IMM = IMMIGRANTS/(IMMIGRANTS + NON.IMMIGRANTS)*100)
tb_nghd <- tb_nghd[,c("AVE_HHSIZE", "AVE_AGE", "AVE_DWVALUE", "POP_DENS",
                       "TRANS_CMT", "POST_EDU", "NON_CAN", "IMM",
                       ## proportion of public transit commuters, post-secondary, non-canadian citizens, immigrants
                       "EMPL_RATE", "UNEMPL_RATE",
                       "AVE_INCOM",
                       "buy.impt.ngbh.access.LRT",
                       "buy.impt.ngbh.access.bus",
                       "buy.impt.ngbh.access.open.space",
                       "In_CTC")]

#------Household characteristics-----

#0.buyers from KW or GTHA
tb_hhld <- select(tb, starts_with("HH_"))
# 244 lived in KW before, 40 from GTHA, 56 from other places (only 8 from other countries originally)
tb_hhld$HH_KW <- ifelse(tb$buy.before.lived.in.KW == "Yes", "KW", "Other")
## Immigrants: not born in Canada - 83 / 257 born-in Canada
tb_hhld$HH_IMM <- ifelse(is.na(tb$hhld.born.in.canada.province), 1, 0)
# 8 from Toronto
tb_hhld$HH_TRT <- ifelse(stringr::str_detect(tb$buy.before.lived.in.other, "ronto") == TRUE, 1, 0)
tb_hhld$HH_TRT[is.na(tb_hhld$HH_TRT)] <- 0
# 40 from GTHA (including Toronto)
tb_hhld$HH_GTHA <- tb$buy.before.lived.in.other
# GTHA list from our survey responses
tb_hhld$HH_GTHA <- ifelse(tb_hhld$HH_GTHA %in% GTHA_buyers, "GTHA buyers",
                          ifelse(tb_hhld$HH_GTHA %in% Local_buyers, "Local buyers", "Other"))

## combine KW and GTHA into one variable
tb_hhld$HH_BUYER <- paste(tb_hhld$HH_KW, tb_hhld$HH_GTHA)
tb_hhld$HH_BUYER <- recode(tb_hhld$HH_BUYER,
                           "KW Other" = "Local buyers", ## 244
                           "Other GTHA buyers" = "GTHA buyers", ## 40
                           "Other Local buyers" = "Local buyers", ## 30
                           "Other Other" = "Other", #24
                           "NA Other" = "NA") %>% #2
  as.factor()
tb_hhld$HH_BUYER[tb_hhld$HH_BUYER=="NA"] <- NA
tb_hhld$HH_BUYER <- factor(tb_hhld$HH_BUYER) ##use factor() to delete the NA level
addmargins(table(tb_hhld$HH_BUYER))
```

```
## 244 KW buyers, 30 buyers around KW => totally 274 are Local buyers
## 40 GTHA buyers
## 24 from other cities, provinces or countries (who have little information for the housing market)

#-------structure of the dataset-----
##1.hhld ethnicity
tb_hhld$HH_WHITE <- ifelse(tb_hhld$HH_ETHN == "White", 1, 0)
addmargins(table(tb_hhld$HH_WHITE))
##2.hhld income
addmargins(table(tb_hhld$HH_INCM))
tb_hhld$HH_INCM <- ordered(tb_hhld$HH_INCM,
                           levels = c("Less than $29,999","$30,000-$49,999",
                                      "$50,000-$74,999", "$75,000-$99,999",
                                      "$100,000-$149,999","$150,000-$249,999",
                                      "$250,000-$499,999"))
##3.hhld type
addmargins(table(tb_hhld$HH_TYPE))
##4.hhld employment status
addmargins(table(tb_hhld$HH_EMPL))
levels(tb_hhld$HH_EMPL)[levels(tb_hhld$HH_EMPL) == "Full time"] <- "Full-time employed"
levels(tb_hhld$HH_EMPL)[levels(tb_hhld$HH_EMPL) == "Part time"] <- "Part-time employed"
levels(tb_hhld$HH_EMPL)[levels(tb_hhld$HH_EMPL) == "Student"] <- "Student, unemployed or other"
levels(tb_hhld$HH_EMPL)[levels(tb_hhld$HH_EMPL) == "Unemployed"] <- "Student, unemployed or other"
levels(tb_hhld$HH_EMPL)[levels(tb_hhld$HH_EMPL) == "Other"] <- "Student, unemployed or other"
addmargins(table(tb_hhld$HH_EMPL))

##CREATE FULL_EMPLOYMENT DUMMY VARIABLE
table(tb_hhld$HH_EMPL)
##(Full-time employed 282)  (Student, unemployed or other 11)   (Part-time employed 11)  (Retired 18)
tb_hhld$HH_FULL_EMPL <- ifelse(tb_hhld$HH_EMPL == "Full-time employed", 1, 0)
table(tb_hhld$HH_FULL_EMPL) ## 0-40; 1-282

tb_hhld$HH_INCM <- factor(tb_hhld$HH_INCM, ordered = FALSE)
tb_hhld$HH_TYPE <- relevel(tb_hhld$HH_TYPE, ref = "Couple without children")
tb_hhld$HH_INCM <- relevel(tb_hhld$HH_INCM, ref = "$100,000-$149,999")
tb_hhld$HH_EDU <- relevel(tb_hhld$HH_EDU, ref = "Postsecondary")
tb_hhld$HH_ETHN <- relevel(tb_hhld$HH_ETHN, ref = "White")
tb_hhld$HH_EMPL <- relevel(tb_hhld$HH_EMPL, ref = "Full-time employed")
tb_hhld$HH_BUYER <- relevel(tb_hhld$HH_BUYER, ref = "Other") #24 other

```

```
##Combine those dataframes
full_data <- cbind(tb_str, tb_local, tb_nghd, tb_hhld)
full_data$HP[full_data$HP == 0] <- NA
full_data <- mutate(full_data, LNHP = log(full_OLS$HP))
str(full_data) ##340 observations

#ADD first-time buyers
#from buyer.analysis.all, 357 observations
file_1st_buyer <- read.csv( file = "/Users/yukeysha/Desktop/Yu.Survery/4.Analysis/R.
    analysis/Buyer.analysis.all/Number_homes_bought.csv")
tb_1st_buyer <- as.tibble(file_1st_buyer)
tb_1st_buyer$internal.id
tb_1st_buyer$HH_homes_before <- tb_1st_buyer$buy.N.of.homes.bought
str(tb_1st_buyer) ## 3 variables
tb_1st_buyer <- select(tb_1st_buyer, internal.id, HH_homes_before)
addmargins(table(tb_1st_buyer$HH_homes_before))
## 0-160; 1-92; 2-55; >=2 - 47; sum 354

str(full_data) ## 340 * 52
full_data <- left_join(full_data, tb_1st_buyer, "internal.id") ## 340 * 53
str(full_data)
addmargins(table(full_data$HH_homes_before))
#          0           1           2    More than 2          Sum
#        148          90          55             45          338
## create HH_FIRST to define the first-time buyer
full_data <- mutate(full_data, HH_FIRST = ifelse(full_data$HH_homes_before == "0", 1, 0))
full_data$HH_FIRST <- factor(full_data$HH_FIRST)

############Summary of counts for each hhld characteristic -- needs to add first-time
    buyers; GTHA buyers,...

full_data <- mutate(full_data, HH_AGE_RANGE = ifelse(full_data$HH_AGE %in% 18:24, "18-24",
                                                     ifelse(full_data$HH_AGE %in% 25:34, "
                                                         25-34",
                                                            ifelse(full_data$HH_AGE %in%
                                                                35:54, "35-54",
                                                                   ifelse(full_data$HH_AGE
                                                                       %in% 55:100, "55+"
                                                                       , NA)))))
t1 <- table(full_data$HH_TYPE)
t1.p <- prop.table(t1)
cb1 <- cbind(t1, t1.p)

t2 <- table(full_data$HH_FIRST)
t2.p <- prop.table(t2)
```

```r
cb2 <- cbind(t2, t2.p)

t3 <- table(full_data$HH_BUYER)
t3.p <- prop.table(t3)
cb3 <- cbind(t3, t3.p)

t4 <- table(full_data$HH_INCM)
t4.p <- prop.table(t4)
cb4 <- cbind(t4, t4.p)

t5 <- table(full_data$HH_IMM)
t5.p <- prop.table(t5)
cb5 <- cbind(t5, t5.p)

t6 <- table(full_data$HH_AGE_RANGE)
t6.p <- prop.table(t6)
cb6 <- cbind(t6, t6.p)

t7 <- table(full_data$HH_FULL_EMPL)
t7.p <- prop.table(t7)
cb7 <- cbind(t7, t7.p)

t8 <- table(full_data$HH_EDU)
t8.p <- prop.table(t8)
cb8 <- cbind(t8, t8.p)

rb <- rbind(cb1, cb2, cb3, cb4, cb5, cb6, cb7, cb8) %>%
  as.data.frame.matrix()
colnames(rb) <- c("count", "percentage")

writeLines(capture.output(stargazer(rb, summary=FALSE,rownames=TRUE)),
           "Demographic_summary_Aug22.tex")

#1.First-stage regression
hed_data <- full_data ## hedonic data 340 observations
ggdensity(hed_data, x = "DIS_LRT", add = "mean", ru = TRUE)+
  theme_pubr()
ggdensity(hed_data, x = "DIS_BUS", add = "mean", ru = TRUE)+
  theme_pubr()
ggdensity(hed_data, x = "OS_ACES", add = "mean", ru = TRUE)+
  theme_pubr()
table(hed_data$REG_RD_ADJ) ## 31 - 1; 309 - 0
table(hed_data$OS_ADJ) ## 55 - 1; 285 - 0
table(hed_data$SINGLE) ## 245 - 1; 94 - 0
table(hed_data$In_CTC) ## 30 -1; 310 - 0
```

```
table(hed_data$SIZE)
##   749 1249 1749 2249 2749 3249
##    13  139  102   56   17   11

#-----Correlation tests------
library(car)
vif(lm(LNHP ~ SINGLE + SQFT + BDMS + BATH + GRAG + YARD
            + BUL_AGE +  POP_DENS + OS_ACES + In_CTC + DIS_LRT + CTC_DISLRT
            + DIS_BUS + POST_EDU + OS_ADJ + REG_RD_ADJ, data = hed_data))

data1 <- dplyr::select(hed_data, SINGLE, SQFT, BDMS, BATH, GRAG, YARD
          , BUL_AGE, BUL_AGE2, POP_DENS, OS_ACES, In_CTC, DIS_LRT, CTC_DISLRT
          , DIS_BUS, POST_EDU, OS_ADJ, REG_RD_ADJ, HP)
X <- data1[,-c(2,8)] #exclude SQFT, and BUL_AGE2
X <- drop_na(X) ## 285 observations of 16 variables
library(corpcor)
pcor <- cor2pcor(cov(X)) ## partial correlation
write.csv(pcor, "partial_correlation.csv")

lm_fit <- lm(log(HP) ~ SINGLE + SQFT + BDMS + BATH + GRAG + YARD +
            + BUL_AGE + BUL_AGE2 + POP_DENS + OS_ACES + In_CTC + CTC_DISLRT + DIS_LRT
            + DIS_BUS + POST_EDU + OS_ADJ + REG_RD_ADJ, data = test_data)
summary(lm_fit)

#-----Non-parametric regression------
library(np)

bw <- npregbw(log(HP) ~ factor(SINGLE) + SQFT + BDMS + BATH + GRAG + YARD +
            + BUL_AGE + POP_DENS + OS_ACES + In_CTC + DIS_LRT + CTC_DISLRT
            + DIS_BUS + POST_EDU + OS_ADJ + REG_RD_ADJ, data = test_data,
             bwtype = "adaptive_nn", bwmethod = "cv.aic", ukertype = "liracine")
fit.np <- npreg(bw, gradients = TRUE, residuals = TRUE)
summary(fit.np)
mean_np <- colMeans(fit.np$grad)
for (i in 1:16) {
  mean_np[i] <- mean(fit.np$grad[,i][fit.np$grad[,i] !=0])}
mean_np
coef(fit.lm)

#significance test
sig.np <- npsigtest(fit.np)

##summary of quantiles
summary_np <- summary(fit.np$grad)

```

```
#summary of the first variable
summary_np[,1] <- summary(fit.np$grad[,1][fit.np$grad[,1] !=0])
## summary of other variables
for (i in 2:16) {
  summary_np[,i] <- summary(fit.np$grad[,i])}
#### I manually saved the summary into excel and then csv file so as to frame a tibble
    here
file <- read.csv("Summary_np_Jan.csv")
tb <- as.data.frame(file)
tb[,2:5] <- round(tb[,2:5],2)
stargazer(tb, title = "First-stage np-hedonic regression results", type = "latex", out = "
    First-stage np-hedonic_Aug22.tex",
          digits = 2, align = TRUE, single.row = TRUE, summary = FALSE, rownames = FALSE,
              no.space = TRUE)

#2. Preference estimates
pref_data <- hed_data
fit.np.grad <- as.data.frame(fit.np$grad)
rows.omit <- fit.np$rows.omit ## 64 omited due to NAs from np-hedonic
pref_data <- filter(pref_data, !(rownames(pref_data) %in% rows.omit))
str(pref_data) ##276 observations

##combine the derived gradient with the original survey data
pref_data <- cbind(pref_data,fit.np.grad)
View(pref_data[,61:77])
pref_data <- pref_data %>%
  mutate(PREF.SING = (exp(V1)-1)*HP)%>%
  mutate(PREF.SQFT = (exp(V2)-1)*HP)%>%
  mutate(PREF.BEDM = (exp(V3)-1)*HP*BDMS)%>%
  mutate(PREF.BATH = (exp(V4)-1)*HP*BATH)%>%
  mutate(PREF.GRAG = (exp(V5)-1)*HP*GRAG)%>%
  mutate(PREF.YARD = (exp(V6)-1)*HP*YARD)%>%
  mutate(PREF.BUL = (exp(V7)-1)*HP*BUL_AGE)%>%
  mutate(PREF.POP = (exp(V8)-1)*HP*POP_DENS)%>%
  mutate(PREF.OS = (exp(V9)-1)*HP*OS_ACES)%>%
  mutate(PREF.CTC = (exp(V10)-1)*HP)%>%
  mutate(PREF.LRT = (exp(V11)-1)*HP*DIS_LRT)%>%
  mutate(PREF.CTC.LRT = (exp(V12)-1)*HP*CTC_DISLRT)%>%
  mutate(PREF.BUS = (exp(V13)-1)*HP*DIS_BUS)%>%
  mutate(PREF.EDU = (exp(V14)-1)*HP*POST_EDU)%>%
  mutate(PREF.OS.ADJ = (exp(V15)-1)*HP)%>%
  mutate(PREF.REG.ADJ = (exp(V16)-1)*HP)

## First, calculate the relative contribution, by (exp(b1)-1)*y*sd(x1)/sd(y). This is
    important.
```

```
## this means that one sd of x change, how much sd of y changes
pref_data <- pref_data %>%
  mutate(Z.grad.SING = (exp(V1)-1)*HP/sd(HP))%>%
  mutate(Z.grad.SQFT = (exp(V2)-1)*HP/sd(HP))%>%
  mutate(Z.grad.BEDM = (exp(V3)-1)*HP*sd(BDMS)/sd(HP))%>%
  mutate(Z.grad.BATH = (exp(V4)-1)*HP*sd(BATH)/sd(HP))%>%
  mutate(Z.grad.GRAG = (exp(V5)-1)*HP*sd(GRAG)/sd(HP))%>%
  mutate(Z.grad.YARD = (exp(V6)-1)*HP*sd(YARD)/sd(HP))%>%
  mutate(Z.grad.BUL = (exp(V7)-1)*HP*sd(BUL_AGE)/sd(HP))%>%
  mutate(Z.grad.POP = (exp(V8)-1)*HP*sd(POP_DENS)/sd(HP))%>%
  mutate(Z.grad.OS = (exp(V9)-1)*HP*sd(OS_ACES)/sd(HP))%>%
  mutate(Z.grad.CTC = (exp(V10)-1)*HP/sd(HP))%>%
  mutate(Z.grad.LRT = (exp(V11)-1)*HP*sd(DIS_LRT)/sd(HP))%>%
  mutate(Z.grad.CTC.LRT = (exp(V12)-1)*HP*sd(CTC_DISLRT)/sd(HP))%>%
  mutate(Z.grad.BUS = (exp(V13)-1)*HP*sd(DIS_BUS)/sd(HP))%>%
  mutate(Z.grad.EDU = (exp(V14)-1)*HP*sd(POST_EDU)/sd(HP))%>%
  mutate(Z.grad.OS.ADJ = (exp(V15)-1)*HP/sd(HP))%>%
  mutate(Z.grad.REG.ADJ = (exp(V16)-1)*HP/sd(HP))

## single - o/1 for probit model
pref_data$SINGLE <- as.factor(pref_data$SINGLE)
pref_data$In_CTC <- as.factor(pref_data$In_CTC)
pref_class <- right_join(class, pref_data, by = "internal.id")
summary(pref_data[,94:109]) ## per 1sd of x change, the change of sd of y

#3. Preference regression
#---change classifications and reference levels
pref_data$HH_TYPE <- recode(pref_data$HH_TYPE,
                            "Couple without children" = "Couple-family without children",
                            "Couple with children" = "Couple-family with children",
                            "One-person household" = "Non-family households",
                            "More-persons household" = "Non-family households")
pref_data$HH_TYPE <- relevel(pref_data$HH_TYPE, "Couple-family without children")
pref_data$HH_INCM <- recode(pref_data$HH_INCM,
                            "Less than $29,999" = "Less than $50,000",
                            "$30,000-$49,999" = "Less than $50,000",
                            "$50,000-$74,999" = "$50,000-$99,999",
                            "$75,000-$99,999" = "$50,000-$99,999",
                            "$150,000-$249,999"= "$150,000 and over",
                            "$250,000-$499,999" = "$150,000 and over")
pref_data$HH_INCM <- relevel(pref_data$HH_INCM, "$50,000-$99,999")
pref_data$HH_BUYER <- relevel(pref_data$HH_BUYER, "Local buyers")
pref_data$HH_AGE_RANGE <- recode(pref_data$HH_AGE_RANGE,
                                 "18-24" = "18-34",
                                 "25-34" = "18-34",
```

```
                                  "35-54" = "35-54",
                                  "55+" = "55+") %>%
  as.factor()
pref_data$HH_AGE_RANGE <- relevel(pref_data$HH_AGE_RANGE,"18-34")
pref_data$HH_FIRST <- recode(pref_data$HH_FIRST,
                              "0" = "Experienced homebuyer",
                              "1" = "First-time homebuyer")
pref_data$HH_FIRST <- relevel(pref_data$HH_FIRST, "First-time homebuyer")

#---probit model for single dummy---
single.price <- pref_data$PREF.SING/100000
pref_single <- glm(data = pref_data,
                  SINGLE ~ HH_TYPE + HH_FIRST + HH_BUYER + HH_INCM + HH_AGE_RANGE + HH_
                      FULL_EMPL + HH_EDU + offset(-single.price),
                  family = binomial(link='probit'))
summary(pref_single)
multiply.by.100000 <- function(x) (x * 100000)

prefsingle <- stargazer(pref_single,
            title = "Estimates of the Willingness to Pay for Single-Detached House",
            covariate.labels = c("Couple with children", "Lone-parent family", "More-
                persons household", "One-person household",
                                  "First time purchase", "Other buyers", "GTHA buyers",
                                  "Less than 29,999", "30,000-49,999", "50,000-74,999", "
                                      75,000-99,999", "150,000-249,999",
                                  "250,000-499,999", "Age", "Full-time employed", "Graduate
                                      ", "High school"),
            single.row = TRUE, align = TRUE, no.space = TRUE,
            model.names = FALSE, report = "vcs*",
            apply.coef=multiply.by.100000, apply.se=multiply.by.100000,
            dep.var.labels = c("WTP for single-detached house"),
          type = "text", out = "WTP_1)_single_with_age_Sep26.tex")

#---probit model for CTC-----

pref_CTC <- pref_class
addmargins(table(pref_CTC$HH_TYPE))
pref_CTC$HH_TYPE <- dplyr::recode(pref_CTC$HH_TYPE,
                            "Loneparent family household" = "Other households",
                            "More-persons household" = "Other households")
pref_CTC$HH_TYPE <- relevel(pref_CTC$HH_TYPE, "Couple-family with children")
addmargins(table(pref_CTC$HH_TYPE))
#Data matrix preparation
dataset_ctc <- dplyr::select(pref_CTC, c("In_CTC", "PREF.CTC", "HH_TYPE","HH_AGE_RANGE", "
    buy.impt.ngbh.access.LRT"))
```

```
matrix <- model.matrix(~., dataset_ctc)
y <- matrix[,2]
x <- matrix[,-c(2,3)]
z <- matrix[,3]*(-1)
testx <- as.data.frame(x)
testy <- as.data.frame(y)
testz <- as.data.frame(z)
test <- cbind(testx, testy, testz)
test$CTC <- test$y
test$couple_without_child <- test$`HH_TYPECouple-family without children`
test$one_person <- test$`HH_TYPEOne-person household`
test$age55 <- test$`HH_AGE_RANGE55+`
test$impt2 <- test$`buy.impt.ngbh.access.LRT2 - Somewhat important`
test$impt3 <- test$`buy.impt.ngbh.access.LRT3 - Very important`
test$offsetz <- test$z
test <- as.tibble(test)
## no importance and demographic characteristics
nullmod <- glm(data=test, y ~ 1,  offset = scale(offsetz), family = binomial(link = "
    probit"))
## only dempgraphic characteristics
glm1 <- glm(data=test, y ~ couple_without_child +
      age55, offset = scale(offsetz), family = binomial(link = "probit"))
summary(glm1)
## both importance and demographic characteristics
glm2 <- glm(data=test, y ~ couple_without_child +
      age55 + impt2 + impt3, offset = scale(offsetz), family = binomial(link = "probit"))
summary(glm2)

#McFadden's pseudo-R squared
1-(logLik(glm1)/logLik(nullmod))
1-(logLik(glm2)/logLik(nullmod))
anova(glm1, glm2, test="Chisq")
fit.ctc <- stargazer(glm1, glm2,
            title = "Binomial regressions for the CTC",
            covariate.labels = c("Couple without children", "Age: 55 and over",
                                 "LRT access: 2-somewhat important", "LRT access: 3-very
                                     important"),
            single.row = TRUE, align = TRUE, no.space = TRUE,
            model.names = FALSE, report = "vcs*",
            dep.var.labels = c("In_CTC"),  digits = 2,
            type = "text",
            out = "/Users/yukeysha/Desktop/Paper1/2019/Paper_1_2019/Latex_tables/CTC_Mar28
                .tex")

#-----Calculate WTP models------
```

```
pref_class$buy.impt.house.N.of.full.bath <- relevel(pref_class$buy.impt.house.N.of.full.
    bath, ref =  "2 - Somewhat important")
## from 1 to 2 bath
pref_class$WTP_BATH_1_2 <- pref_class$PREF.BATH*(log(2)-log(1))

WTP_BATH_1_2 <- lm(data = pref_class,
                  WTP_BATH_1_2 ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_EDU
                      + HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.house.N.of.full.bath+pref_class$buy.impt.house.N.
                      of.half.bath)
summary(WTP_BATH_1_2)
## from 2 to 3 bedrooms
pref_class$WTP_BED_2_3 <- pref_class$PREF.BEDM*(log(3)-log(2))

WTP_BED_2_3 <- lm(data = pref_class,
                  WTP_BED_2_3 ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_EDU +
                       HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.house.N.of.bedroom)
summary(WTP_BED_2_3)
## wtp for sqft
WTP_sqft <- lm(data = pref_class,
                  PREF.SQFT ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_EDU +
                      HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.house.size)
summary(WTP_sqft)
## from 1 to 2 garage
pref_class$WTP_GRAG_1_2 <- pref_class$PREF.GRAG*(log(2)-log(1))

WTP_GRAG_1_2 <- lm(data = pref_class,
                  WTP_GRAG_1_2 ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_EDU
                      + HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.house.N.of.covered.parking)
summary(WTP_GRAG_1_2)
## from 3000 to 5000 sqft of yard size (mean= 4231, sd = 3696)
pref_class$WTP_YARD_3K_5K <- pref_class$PREF.YARD*(log(5000)-log(3000))
WTP_YARD_3K_5K <- lm(data = pref_class,
                  WTP_YARD_3K_5K ~ HH_TYPE + HH_INCM + HH_AGE_RANGE
                  + pref_class$buy.impt.house.yard.size)
summary(WTP_YARD_3K_5K)
wtp_YARD<- stargazer(WTP_YARD_3K_5K,
            title = "Estimates of the Willingness to Pay",
            # covariate.labels = c("Couple-family without children", "Lone-parent family",
                 "More-persons household", "One-person household",  "Age35-54", "Age55+",
                "Attitude: 2-somewhat important", "Attitude: 3-very important"),
```

```r
            single.row = TRUE, align = TRUE, no.space = TRUE,
            model.names = FALSE, report = "vcs*",
            dep.var.labels = c("WTP"),  digits = 0,
            type = "text", out = "/Users/yukeysha/Desktop/Paper1/2019/Paper_1_2019/Latex_
                tables/WTP_YARD.tex")

wtp1 <- stargazer(WTP_sqft, WTP_BATH_1_2, WTP_BED_2_3, WTP_GRAG_1_2, WTP_YARD_3K_5K,
            title = "Willingness to pay for the structural attributes",
            digits = 1,
            single.row = TRUE, align = TRUE, no.space = TRUE, report = "vc*",
            dep.var.labels = c("SIZE", "BATH", "BEDM", "GRAG", "YARD"),
          type = "text", out = "/Users/yukeysha/Desktop/Paper1/2018-Sep/Latex_tables/WTP_
              (1)_Jan28.tex")
## from 1.5KM to 3KM from the LRT (sd = 1644, mean = 3668)
pref_class$WTP_LRT_1.5K_3K <- pref_class$PREF.LRT.total*(log(3000)-log(1500))
WTP_LRT_1.5K_3K <- lm(data = pref_class,
                  WTP_LRT_1.5K_3K ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_
                      EDU + HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.ngbh.access.LRT)
summary(WTP_LRT_1.5K_3K)

pref_class$WTP_LRT_1K_2K <- pref_class$PREF.LRT.total*(log(2000)-log(1000))
WTP_LRT_1K_2K <- lm(data = pref_class,
                  WTP_LRT_1K_2K ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_EDU
                       + HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.ngbh.access.LRT)
summary(WTP_LRT_1K_2K)
### output _ LRT access
pref_class$WTP_LRT_CTC_1k_500 <- pref_class$PREF.CTC.LRT*(log(500)-log(1000))
WTP_LRT_CTC_1k_500 <- lm(data = pref_class,
                  WTP_LRT_CTC_1k_500 ~ HH_TYPE + HH_INCM + HH_AGE_RANGE)
summary(WTP_LRT_CTC_1k_500)
pref_class$WTP_LRT_3K_1K <- pref_class$PREF.LRT*(log(1000)-log(3000))
WTP_LRT_3K_1K <- lm(data = pref_class,
                  WTP_LRT_3K_1K ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL)
summary(WTP_LRT_3K_1K)
wtpLRT <- stargazer(WTP_LRT_3K_1K,
            title = "Estimates of the Willingness to Pay for moving from 3km to 1 km away
                from the nearest LRT stop",
            # covariate.labels = c("Couple-family without children", "Lone-parent family",
                 "More-persons household", "One-person household", "Age35-54", "Age55+", "
                Attitude: 2-somewhat important", "Attitude: 3-very important"),
            single.row = TRUE, align = TRUE, no.space = TRUE,
            model.names = FALSE, report = "vcs*",
            dep.var.labels = c("WTP for the LRT access"),  digits = 1,
```

```r
            type = "text", out = "/Users/yukeysha/Desktop/Paper1/2018-Sep/Latex_tables/WTP_LRT_Jan27.tex")
## from 600 to 300 meters from the bus stop (mean= 342, sd = 310)
pref_class$WTP_BUS_600_300 <- pref_class$PREF.BUS*(log(300)-log(600))

WTP_BUS_600_300 <- lm(data = pref_class,
                  WTP_BUS_600_300 ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL)
summary(WTP_BUS_600_300)
## from 40 to 60 of open space amenity (mean= 42.18, sd = 17)
pref_class$WTP_OS_40_60 <- pref_class$PREF.OS*(log(60)-log(40))
WTP_OS_40_60 <- lm(data = pref_class,
                  WTP_OS_40_60 ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_EDU + HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.ngbh.access.open.space)
summary(WTP_OS_40_60)
## from 3000 to 5000 of population density (mean= 2959, sd = 2116)
pref_class$WTP_POP_3K_5K <- pref_class$PREF.POP*(log(5000)-log(3000))
WTP_POP_3K_5K <- lm(data = pref_class,
                  WTP_POP_3K_5K ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_EDU + HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.ngbh.density)
summary(WTP_POP_3K_5K)
## from 60 to 80 of education rate (mean= 62.35, sd = 9)
pref_class$WTP_EDU_60_80 <- pref_class$PREF.EDU*(log(80)-log(60))
WTP_EDU_60_80 <- lm(data = pref_class,
                  WTP_EDU_60_80 ~ HH_TYPE + HH_INCM + HH_AGE_RANGE + HH_FULL_EMPL + HH_EDU + HH_FIRST + HH_BUYER
                  + pref_class$buy.impt.ngbh.similar.education)
summary(WTP_EDU_60_80)
wtp2 <- stargazer(WTP_BUS_300_600, WTP_OS_40_60, WTP_POP_3K_5K, WTP_EDU_60_80,
            title = "Willingness to pay for the locational and neighbourhood attributes",
            digits = 1,
            single.row = TRUE, align = TRUE, no.space = TRUE, report = "vc*",
            dep.var.labels = c("WTP_BUS", "WTP_OS", "WTP_POP_DENS", "WTP_EDU"),
          type = "text", out = "/Users/yukeysha/Desktop/Paper1/2018-Sep/Latex_tables/WTP_(2)_Jan28.tex")
```

A6 R code for Chapter 4

```r

knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(dev = 'pdf')
libs <- c('tidyverse','ggplot2','dplyr', 'ggpubr', 'latticeExtra', 'gridExtra', 'MASS',
```

```r
          'colorspace', 'plyr', 'Hmisc', 'scales', 'lattice','ggthemes','gmodels',
          'magrittr','stargazer','tidyr','scales', 'graphics', 'sjPlot', "corrplot", "np")
lapply(libs, require, character.only = T)

# latent class analysis
library(depmixS4)
set.seed(1)
#write.csv(buyer_all,file="/Users/yukeysha/R_workspace/Latent_paper/buyer_all.csv")
#buyer_all <- read.csv("/Users/yukeysha/R_workspace/Latent_paper/buyer_all.csv")

# 1-class
mod1 <- depmix(list(buy.impt.ngbh.density~1, buy.impt.ngbh.walking~1, buy.impt.ngbh.cycling~1,
                    buy.impt.ngbh.mix~1,buy.impt.ngbh.access.LRT~1, buy.impt.ngbh.access.bus~1,
                    buy.impt.ngbh.access.open.space~1,buy.impt.ngbh.access.center~1),
               data=buyer_all, # the dataset to use
               nstates=1, # the number of latent classes
               family=list(multinomial("identity"),multinomial("identity"),multinomial("identity"),
                           multinomial("identity"),multinomial("identity"),multinomial("identity"),
                           multinomial("identity"),multinomial("identity")),
               respstart=runif(24))
fm1 <- fit(mod1, emc=em.control(rand=FALSE))
summary(fm1)

# 2-class
mod2 <- depmix(list(buy.impt.ngbh.density~1, buy.impt.ngbh.walking~1, buy.impt.ngbh.cycling~1,
                    buy.impt.ngbh.mix~1,buy.impt.ngbh.access.LRT~1, buy.impt.ngbh.access.bus~1,
                    buy.impt.ngbh.access.open.space~1, buy.impt.ngbh.access.center~1),
               data=buyer_all, # the dataset to use
               nstates=2, # the number of latent classes/states
               family=list(multinomial("identity"),multinomial("identity"),multinomial("identity"),
                           multinomial("identity"),multinomial("identity"),multinomial("identity"),
                           multinomial("identity"),multinomial("identity")),
               respstart=runif(48))
fm2 <- fit(mod2, emc=em.control(rand=FALSE))
fm2
summary(fm2)
#Return the posterior states for a fitted (dep-)mix object.
```

```r
#In the case of a latent class or mixture model these are the class probabilities.
posterior.states <- depmixS4::posterior(fm2)
table(posterior.states$state) ## 2 classes; 1-148; 2-209
fm2@posterior

# 3-class
mod3 <- depmix(list(buy.impt.ngbh.density~1, buy.impt.ngbh.walking~1, buy.impt.ngbh.cycling~1,
                    buy.impt.ngbh.mix~1,buy.impt.ngbh.access.LRT~1, buy.impt.ngbh.access.bus~1,
                    buy.impt.ngbh.access.open.space~1,buy.impt.ngbh.access.center~1),
               data=buyer_all, # the dataset to use
               nstates=3, # the number of latent classes
               family=list(multinomial("identity"),multinomial("identity"),multinomial("identity"),
                           multinomial("identity"),multinomial("identity"),multinomial("identity"),
                           multinomial("identity"),multinomial("identity")),
               respstart=runif(72))
fm3 <- fit(mod3, emc=em.control(rand=FALSE))
fm3
summary(fm3)
posterior.states3 <- depmixS4::posterior(fm3)
table(posterior.states3$state)## 1-193;2-76;3-88

#llratio(fm2,fm3)
# 4 class (n = 357)
mod4 <- depmix(list(buy.impt.ngbh.density~1, buy.impt.ngbh.walking~1, buy.impt.ngbh.cycling~1,
                    buy.impt.ngbh.mix~1,buy.impt.ngbh.access.LRT~1, buy.impt.ngbh.access.bus~1,
                    buy.impt.ngbh.access.open.space~1,buy.impt.ngbh.access.center~1),
               data=buyer_all, # the dataset to use
               nstates=4, # the number of latent classes
               family=list(multinomial("identity"),multinomial("identity"),multinomial("identity"),
                           multinomial("identity"),multinomial("identity"),multinomial("identity"),
                           multinomial("identity"),multinomial("identity")),
               respstart=runif(96))
fm4 <- fit(mod4, emc=em.control(rand=FALSE))
fm4
summary(fm4)
posterior.states4 <- depmixS4::posterior(fm4)
table(posterior.states4$state)## 1-49; 2-142;3-69;4-97
```

```r

buyer_all_class <- cbind(buyer_all, posterior.states$state) # 357*679
buyer_all_class$class2 <- buyer_all_class$`posterior.states$state`
table(buyer_all_class$class2)

## Identify three groups
library("dplyr")
table(buyer_all_class$In_CTC)
table(buyer_all_class$class2)

#three groups
buyer_all_class <- mutate(buyer_all_class, group = ifelse(buyer_all_class$In_CTC == 1, "
    Current TOD households",
                                                          ifelse(buyer_all_class$class2 ==
                                                              1, "Potential TOD
                                                              households",
                                                                 "Car-dependent households
                                                                 ")))
# cc <- dplyr::select(buyer_all_class, c("In_CTC", "class2", "group"))
# View(cc)
table(buyer_all_class$In_CTC,buyer_all_class$class2)
table(buyer_all_class$In_CTC,buyer_all_class$group)
table(buyer_all_class$class2,buyer_all_class$group)

buyer_all_class$group <- factor(buyer_all_class$group, levels = c("Current TOD households"
    , "Potential TOD households","Car-dependent households"))

# 30 current TOD; 123 Potential TOD homebuyers; 187 car-dependent
addmargins(table(buyer_all_class$group)) # sum = 340

## Demographic profiles - chi-square tests for demographics
## test whether each group has different demographic distribution compared to the total
## household type
ppt <- prop.table(table(buyer_all_class$hhld.type))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    hhld.type))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$hhld.type))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$hhld.type))%>%
  as.data.frame()

```

```
p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$hhld.type)
    %>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$hhld.type)
    %>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$hhld.type)
    %>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$hhld.type),1),2)

## household age
ppt <- prop.table(table(buyer_all_class$hhld.age.range))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    hhld.age.range))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$hhld.age.range))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$hhld.age.range))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$hhld.age.
    range)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$hhld.age.
    range)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$hhld.age.
    range)%>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$hhld.age.range),1),2)

```

```
## household income
ppt <- prop.table(table(buyer_all_class$hhld.income))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$hhld.income))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$hhld.income))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$hhld.income))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$hhld.income)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$hhld.income)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$hhld.income)%>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq)
chisq.test(p3$Freq, p=ppt$Freq)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$hhld.income),1),2)

## homebuyer
ppt <- prop.table(table(buyer_all_class$hhld.buyer))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$hhld.buyer))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$hhld.buyer))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$hhld.buyer))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$hhld.buyer)%>%
  as.data.frame()
```

```
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$hhld.buyer
    )%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$hhld.buyer)
    %>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$hhld.buyer),1),2)

## buying experience (no difference at all compared to the total)
ppt <- prop.table(table(buyer_all_class$hhld.first))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    hhld.first))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$hhld.first))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$hhld.first))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$hhld.first)
    %>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$hhld.first
    )%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$hhld.first)
    %>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq)
chisq.test(p2$Freq, p=ppt$Freq)
chisq.test(p3$Freq, p=ppt$Freq)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$hhld.first),1),2)

#demographic graphs
addmargins(table(buyer_all_class$hhld.income,buyer_all_class$group),1)
addmargins(prop.table(table(buyer_all_class$hhld.income,buyer_all_class$group),2))
```

```r

chisq.test(table(buyer_all_class$hhld.income,buyer_all_class$group))
chisq.test(table(buyer_all_class$hhld.type,buyer_all_class$group))
chisq.test(table(buyer_all_class$hhld.age.range,buyer_all_class$group))

chisq.test(table(buyer_all_class$hhld.first,buyer_all_class$group))
chisq.test(table(buyer_all_class$hhld.buyer,buyer_all_class$group))

a1 <- prop.table(table(buyer_all_class$group, buyer_all_class$hhld.type),2)%>%
  as.data.frame()
a1$Var <- c("Household type")
levels(a1$Var2)[levels(a1$Var2)=="Other households with 2 or more persons"] <- "Other
    households"
levels(a1$Var2)[levels(a1$Var2)=="Loneparent family household"] <- "Loneparent family"
a2 <- prop.table(table(buyer_all_class$group,buyer_all_class$hhld.income),2)%>%
  as.data.frame()
a2$Var <- c("Household income")
a3 <- prop.table(table(buyer_all_class$group,buyer_all_class$hhld.age.range),2)%>%
  as.data.frame()
a3$Var <- c("Household head age")
levels(a3$Var2)[levels(a3$Var2)=="15-24"] <- "Age 18-24"
levels(a3$Var2)[levels(a3$Var2)=="25-34"] <- "Age 25-34"
levels(a3$Var2)[levels(a3$Var2)=="35-54"] <- "Age 35-54"
levels(a3$Var2)[levels(a3$Var2)=="55+"] <- "Age 55+"
a4 <- prop.table(table(buyer_all_class$group,buyer_all_class$hhld.first),2)%>%
  as.data.frame()
a4$Var <- c("Buying experience")
a5 <- prop.table(table(buyer_all_class$group,buyer_all_class$hhld.buyer),2)%>%
  as.data.frame()
a5$Var <- c("Homebuyers")

lc_a <- rbind(a1,a2,a3,a5)
lc_a$Var1 <- ordered(lc_a$Var1, levels = c("Current TOD households", "Potential TOD
    households","Car-dependent households"),
                     labels = c("Current TOD households", "Potential TOD households","Car-
                         dependent households"))
lc_a$Var <- ordered(lc_a$Var, levels = c("Household type","Household income","Household
    head age","Homebuyers"),
                    labels = c("Household type","Income","Age","Homebuyers"))

 ggplot(data=lc_a)+
  geom_bar(stat = "identity", aes(x=Var2, y = Freq, fill = Var1), colour="black", size
      =0.2,
           width = 0.5, position = position_stack(reverse = TRUE))+
  scale_fill_grey(start = 0.1, end = 1)+
```

```
  facet_grid(.~Var,scales="free_x", space="free", switch = "y", as.table = FALSE)+
  scale_y_continuous(labels = scales::percent)+
  xlab("")+
  ylab("Share of responses")+
  theme_bw(base_size = 12, base_family = "Gill Sans MT" )+
  theme(panel.border = element_rect(size=0.3),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        strip.background = element_blank(),
        legend.margin=margin(t=0.25,r=0.25,b=0.25,l=0.25,unit = 'cm'),
        legend.position = c(0.45, -1),
        legend.direction = "horizontal",
        legend.justification = "center",
        legend.text=element_text(size=10),
        legend.title = element_blank(),
        legend.box.background = element_rect(size=0.3),
        legend.key.width = unit(0.3, "cm"),
        legend.key.height = unit(0.3, "cm"),
        legend.spacing.x = unit(0.5,'cm'),
        axis.text=element_text(size=12),
        axis.title = element_text(size = 12),
        plot.margin = margin(t=0.1,r=0.1,b=0.5,l=0.1,unit = 'cm'))+
  rotate_x_text(60)+
  ggsave("/Users/yukeysha/R_workspace/Latent_paper/hhld.profile.jpeg", width = 7, height =
      4, dpi = 1200)

#Plot out the motivations
#Motivations of moving into the current house
motivation <- c("For investment", "Getting a new job",
                       "Seeking new job opportunities", "Getting married/partnership",
                       "Divorce/seperation", "Expanding family size",
                       "Decreasing family size", "Supporting parents",
                       "For my or my partner's education", "For child's education",
                       "Better environment quality", "More affordable",
                       "Upsize", "Downsize",
                       "Taking advantage of the market", "Expecting market prices to go
                           down",
                       "Expecting market prices to go up", "Better access to workplace",
                       "Better access to facilities", "Better access to bus stops",
                       "LRT construction", "Better access to future LRT stops",
                       "Expecting price increase due to LRT")
### n total = 340

####1. n = 30
CTOD <- filter(buyer_all_class, buyer_all_class$group == "Current TOD households")
```

```
movin <- dplyr::select(CTOD, grep("^move.in.",cn, value = TRUE))

df_min <- data.frame(count = apply(movin, 2, sum)) ## 2 - column
#row.names(df_min)
df_min$motivation <- motivation
df_min <- as_tibble(rownames_to_column(df_min)) ## rowname to the first coloumn
df_min_CTOD <- arrange(df_min, desc(count)) ## sort by descending count
df_min_CTOD$group <- "Current TOD households"
df_min_CTOD$percent <- (df_min_CTOD$count)/30

####2. n = 123
PTOD <- filter(buyer_all_class, buyer_all_class$group == "Potential TOD households")
movin <- dplyr::select(PTOD, grep("^move.in.",cn, value = TRUE))

df_min <- data.frame(count = apply(movin, 2, sum))
df_min$motivation <- motivation
df_min <- as_tibble(rownames_to_column(df_min))
df_min_PTOD <- arrange(df_min, desc(count))
df_min_PTOD$group <- "Potential TOD households"
df_min_PTOD$percent <- (df_min_PTOD$count)/123

####2. n = 187
COD <- filter(buyer_all_class, buyer_all_class$group == "Car-dependent households")
movin <- dplyr::select(COD, grep("^move.in.",cn, value = TRUE))

df_min <- data.frame(count = apply(movin, 2, sum))
df_min$motivation <- motivation
df_min <- as_tibble(rownames_to_column(df_min))
df_min_COD <- arrange(df_min, desc(count))
df_min_COD$group <- "Car-dependent households"
df_min_COD$percent <- (df_min_COD$count)/187

### combine all the three groups
df_movin <- rbind(df_min_CTOD,df_min_PTOD,df_min_COD)
df_movin <- as.tibble(df_movin)

df_movin$group <- factor(df_movin$group, levels = c("Current TOD households","Potential
    TOD households", "Car-dependent households"))

library(extrafont)
ggplot(data = df_movin, aes(x=reorder(motivation,desc(percent)), y = percent)) +
  geom_bar(aes(fill = group),
    stat = "identity", width = 0.8, position = "dodge", color = "black", size = 0.2) +
  scale_fill_grey(start = 0.1, end = 1)+
  scale_y_continuous(labels = scales::percent)+
```

```r
  xlab("")+
  ylab("Percentage of responses") +
  labs(title = "") +
  #guides(fill=guide_legend("Three groups"))+
  theme_bw(base_family = "Gill Sans MT" ,base_size = 18)+
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), axis.line = element_line(colour = "black"),
        legend.position = c(0.5,-1.12),
        legend.direction = "horizontal",
        legend.margin=margin(t=0.25,r=0.25,b=0.25,l=0.25,unit = 'cm'),
        legend.text=element_text(size=15),
        legend.title = element_blank(),
        #legend.title = element_text(size = 15),
        legend.box.background = element_rect(size=0.8),
        legend.key.width = unit(0.4, "cm"),
        legend.key.height = unit(0.3, "cm"),
        legend.justification = "center",
        legend.spacing.x = unit(0.5,'cm'),
        axis.text=element_text(size=18),
        axis.title = element_text(size = 18),
        plot.margin = margin(t=0.1,r=0.1,b=0.6,l=0.1,unit = 'cm'))+
  rotate_x_text(angle = 70)+
  ggsave("/Users/yukeysha/R_workspace/Latent_paper/motivation.jpeg", width = 12, height =
      8, dpi = 1200)

## Residential preferences
#TOD preferences
r1 <- prop.table(table(buyer_all_class$buy.impt.ngbh.density, buyer_all_class$group),2)%>%
  as.data.frame()
r1$Var <- c("Density of housing")

r2 <- prop.table(table(buyer_all_class$buy.impt.ngbh.mix, buyer_all_class$group),2)%>%
  as.data.frame()
r2$Var <- c("Land use mix")

r3 <- prop.table(table(buyer_all_class$buy.impt.ngbh.cycling, buyer_all_class$group),2)%>%
  as.data.frame()
r3$Var <- c("Bicycle-friendly environment")

## chi-square test
## data:  c(0.233, 0.3, 0.467)
## X-squared = 0.49761, df = NA, p-value = 0.3198
chisq.test(c(0.233,0.300,0.467), p=c(0.066,0.254,0.680), simulate.p.value = TRUE)
## data:  c(0.233, 0.767)
## X-squared = 0.45242, df = NA, p-value = 0.05847
```

```r
chisq.test(c(0.233,0.767), p=c(0.066,0.934), simulate.p.value = TRUE)

r4 <- prop.table(table(buyer_all_class$buy.impt.ngbh.walking, buyer_all_class$group),2)%>%
  as.data.frame()
r4$Var <- c("Easy to walk")

r5 <- prop.table(table(buyer_all_class$buy.impt.ngbh.access.LRT, buyer_all_class$group),2)%>%
  as.data.frame()
r5$Var <- c("LRT access")

r6 <- prop.table(table(buyer_all_class$buy.impt.ngbh.access.bus, buyer_all_class$group),2)%>%
  as.data.frame()
r6$Var <- c("Bus access")

r7 <- prop.table(table(buyer_all_class$buy.impt.ngbh.access.open.space, buyer_all_class$group),2)%>%
  as.data.frame()
r7$Var <- c("Access to public open space")
## data:  c(0.133, 0.433, 0.434)
## X-squared = 17.611, df = NA, p-value = 0.0004998
chisq.test(c(0.133, 0.433, 0.434), p=c(0.001,0.285,0.714), simulate.p.value = TRUE)

r8 <- prop.table(table(buyer_all_class$buy.impt.ngbh.access.center, buyer_all_class$group),2)%>%
  as.data.frame()
r8$Var <- c("Access to urban centers")

pre_r <- rbind(r1,r2,r3,r4,r5,r6,r7,r8)
pre_r <- as.data.frame(pre_r)

pre_r$Var1 <- ordered(pre_r$Var1, levels = c("3 - Very important", "2 - Somewhat important", "1 - Not important"))
pre_r$Var2 <- ordered(pre_r$Var2, levels = c("Car-dependent households", "Potential TOD households", "Current TOD households"))

pre_r$Var <- ordered(pre_r$Var,
                     levels = c("LRT access", "Bus access", "Easy to walk", "Bicycle-friendly environment",
                                "Access to urban centers", "Access to public open space","Land use mix",
                                "Density of housing"),
                     labels = c("LRT access", "Bus access", "Easy to walk", "Easy to cycle",
```

```r
                                              "Access to urban center", "Access to open space
                                                  ","Land use mix",
                                              "Density of housing"))

ggplot(data=pre_r)+
  geom_bar(stat = "identity", aes(x=Var2, y = Freq, fill = Var1), colour="black", size
      =0.2,
           width = 0.5, position = position_stack(reverse = TRUE))+
  geom_hline(yintercept=0.5, linetype="dashed", size=0.7, color = "grey")+
  scale_fill_grey(start = 0.1, end = 1)+
  facet_grid(.~Var,scales="free_y", space="free", switch = "y", as.table = FALSE)+
  facet_wrap(~Var, ncol = 4)+
  scale_y_continuous(labels = scales::percent)+
  xlab("")+
  ylab("Share of responses")+
  theme_bw(base_size = 15, base_family = "Gill Sans MT")+
  theme(panel.border = element_rect(size=0.3),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        strip.background = element_blank(),
        strip.text = element_text(size=15),
        legend.margin=margin(t=0.25,r=0.25,b=0.25,l=0.25,unit = 'cm'),
        legend.position = c(0.45, -0.4),
        legend.direction = "horizontal",
        legend.justification = "center",
        legend.text=element_text(size=12),
        legend.title = element_blank(),
        legend.box.background = element_rect(size=0.3),
        legend.key.width = unit(0.3, "cm"),
        legend.key.height = unit(0.3, "cm"),
        legend.spacing.x = unit(0.5,'cm'),
        axis.text=element_text(size=15),
        axis.title = element_text(size = 15),
        plot.margin = margin(t=0,r=0,b=1.4,l=0,unit = 'cm'))+
  coord_flip()+
  rotate_x_text(45)+
  ggsave("/Users/yukeysha/R_workspace/Latent_paper/TOD.pref_revised.jpeg", width = 12,
      height = 4.5, dpi = 1200)

#stated importance for structural attributes
s1 <- prop.table(table(buyer_all_class$buy.impt.house.price, buyer_all_class$group),2)%>%
  as.data.frame()
s1$Var <- c("Housing price")

```

```
s2 <- prop.table(table(buyer_all_class$buy.impt.house.howeownership, buyer_all_class$group
    ),2)%>%
  as.data.frame()
s2$Var <- c("Homeownership")

s3 <- prop.table(table(buyer_all_class$buy.impt.house.type, buyer_all_class$group),2)%>%
  as.data.frame()
s3$Var <- c("Housing type")

s4 <- prop.table(table(buyer_all_class$buy.impt.house.N.of.bedroom, buyer_all_class$group)
    ,2)%>%
  as.data.frame()
s4$Var <- c("Bedroom")

s5 <- prop.table(table(buyer_all_class$buy.impt.house.size, buyer_all_class$group),2)%>%
  as.data.frame()
s5$Var <- c("Home size")

s6 <- prop.table(table(buyer_all_class$buy.impt.house.N.of.total.parking, buyer_all_class$
    group),2)%>%
  as.data.frame()
s6$Var <- c("Garage")

s7 <- prop.table(table(buyer_all_class$buy.impt.house.N.of.full.bath, buyer_all_class$
    group),2)%>%
  as.data.frame()
s7$Var <- c("Full bathroom")

s8 <- prop.table(table(buyer_all_class$buy.impt.house.yard.size, buyer_all_class$group),2)
    %>%
  as.data.frame()
s8$Var <- c("Yard size")

s9 <- prop.table(table(buyer_all_class$buy.impt.house.built.year, buyer_all_class$group)
    ,2)%>%
  as.data.frame()
s9$Var <- c("Built year")

s10 <- prop.table(table(buyer_all_class$buy.impt.house.maintaneance, buyer_all_class$group
    ),2)%>%
  as.data.frame()
s10$Var <- c("Maintenance")

s11 <- prop.table(table(buyer_all_class$buy.impt.house.N.of.half.bath, buyer_all_class$
    group),2)%>%
```

```
  as.data.frame()
s11$Var <- c("Half bathroom")

s12 <- prop.table(table(buyer_all_class$buy.impt.house.N.of.floor, buyer_all_class$group),2)%>%
  as.data.frame()
s12$Var <- c("Floor number")

pre_s <- rbind(s1,s2,s3,s4,s5,s6,s7,s8,s9,s10,s11,s12)
pre_s <- as.data.frame(pre_s)

pre_s$Var1 <- ordered(pre_s$Var1, levels = c("3 - Very important", "2 - Somewhat important", "1 - Not important"))
pre_s$Var2 <- ordered(pre_s$Var2, levels = c("Current TOD households", "Potential TOD households","Car-dependent households"))
pre_s$Var <- ordered(pre_s$Var,
                     levels = c("Housing price", "Homeownership", "Housing type", "Bedroom","Home size","Garage",
                                "Full bathroom", "Half bathroom", "Yard size", "Built year", "Maintenance", "Floor number"),
                     labels = c("Housing price", "Homeownership", "Housing type", "Bedroom","Home size","Garage",
                                "Full bathroom", "Half bathroom", "Yard size", "Built year", "Maintenance", "Floor number"))

ggplot(data=pre_s)+
  geom_bar(stat = "identity", aes(x=Var2, y = Freq, fill = Var1), colour="black", size=0.2,
           width = 0.5, position = position_stack(reverse = TRUE))+
  geom_hline(yintercept=0.5, linetype="dashed", size=0.7, color = "grey")+
  scale_fill_grey(start = 0.1, end = 1)+
  facet_grid(.~Var,scales="free_y", space="free", switch = "y", as.table = FALSE)+
  facet_wrap(~Var, ncol = 4)+
  scale_y_continuous(labels = scales::percent)+
  xlab("")+
  ylab("Share of responses")+
  theme_bw(base_size = 15, base_family = "Gill Sans MT")+
  theme(panel.border = element_rect(size=0.3),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        strip.background = element_blank(),
        strip.text = element_text(size=15),
        legend.margin=margin(t=0.25,r=0.25,b=0.25,l=0.25,unit = 'cm'),
        legend.position = c(0.45, -0.25),
        legend.direction = "horizontal",
```

```r
        legend.justification = "center",
        legend.text=element_text(size=12),
        legend.title = element_blank(),
        legend.box.background = element_rect(size=0.3),
        legend.key.width = unit(0.3, "cm"),
        legend.key.height = unit(0.3, "cm"),
        legend.spacing.x = unit(0.5,'cm'),
        axis.text=element_text(size=15),
        axis.title = element_text(size = 15),
        plot.margin = margin(t=0,r=0,b=1.5,l=0,unit = 'cm'))+
  coord_flip()+
  rotate_x_text(45)+
  ggsave("/Users/yukeysha/R_workspace/Latent_paper/Structural.pref_revised.jpeg", width =
      12, height = 5.8, dpi = 1200)

#social economics
sd1 <- prop.table(table(buyer_all_class$buy.impt.ngbh.similar.hhld.size, buyer_all_class$
    group),2)%>%
  as.data.frame()
sd1$Var <- c("Similar household size")

sd3 <- prop.table(table(buyer_all_class$buy.impt.ngbh.similar.age, buyer_all_class$group)
    ,2)%>%
  as.data.frame()
sd3$Var <- c("Similar age")

sd4 <- prop.table(table(buyer_all_class$buy.impt.ngbh.similar.hhld.income, buyer_all_class
    $group),2)%>%
  as.data.frame()
sd4$Var <- c("Similar income")

sd5 <- prop.table(table(buyer_all_class$buy.impt.ngbh.similar.education, buyer_all_class$
    group),2)%>%
  as.data.frame()
sd5$Var <- c("Similar education")

sd6 <- prop.table(table(buyer_all_class$buy.impt.ngbh.similar.ethn, buyer_all_class$group)
    ,2)%>%
  as.data.frame()
sd6$Var <- c("Similar ethnicity")

sd7 <- prop.table(table(buyer_all_class$buy.impt.ngbh.similar.school.quality, buyer_all_
    class$group),2)%>%
  as.data.frame()
sd7$Var <- c("School quality")
```

```
sd2 <- prop.table(table(buyer_all_class$buy.impt.ngbh.similar.safety, buyer_all_class$
    group),2)%>%
  as.data.frame()
sd2$Var <- c("Neighbourhood safety")

pre_sd <- rbind(sd1,sd2,sd3,sd4,sd5,sd6,sd7)
pre_sd <- as.data.frame(pre_sd)

pre_sd$Var1 <- ordered(pre_sd$Var1, levels = c("3 - Very important", "2 - Somewhat
    important", "1 - Not important"))
pre_sd$Var2 <- ordered(pre_sd$Var2, levels = c("Current TOD households", "Potential TOD
    households","Car-dependent households"))

ggplot(data=pre_sd)+
  geom_bar(stat = "identity", aes(x=Var2, y = Freq, fill = Var1), colour="black", size
      =0.2,
           width = 0.5, position = position_stack(reverse = TRUE))+
  geom_hline(yintercept=0.5, linetype="dashed", size=0.7, color = "grey")+
  scale_fill_grey(start = 0.1, end = 1)+
  facet_grid(.~Var,scales="free_y", space="free", switch = "y", as.table = FALSE)+
  facet_wrap(~Var, ncol = 4)+
  scale_y_continuous(labels = scales::percent)+
  xlab("")+
  ylab("Share of responses")+
  theme_bw(base_size = 15, base_family = "Gill Sans MT")+
  theme(panel.border = element_rect(size=0.3),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        strip.background = element_blank(),
        strip.text = element_text(size=15),
        legend.margin=margin(t=0.25,r=0.25,b=0.25,l=0.25,unit = 'cm'),
        legend.position = c(0.45, -0.38),
        legend.direction = "horizontal",
        legend.justification = "center",
        legend.text=element_text(size=12),
        legend.title = element_blank(),
        legend.box.background = element_rect(size=0.3),
        legend.key.width = unit(0.3, "cm"),
        legend.key.height = unit(0.3, "cm"),
        legend.spacing.x = unit(0.5,'cm'),
        axis.text=element_text(size=15),
        axis.title = element_text(size = 15),
        plot.margin = margin(t=0,r=0,b=1.3,l=0,unit = 'cm'))+
  coord_flip()+
```

```
  rotate_x_text(45)+
  ggsave("/Users/yukeysha/R_workspace/Latent_paper/Social.pref_revised.jpeg", width = 12,
      height = 4.5, dpi = 1200)

# tests for importance level - homeownership
## housing ownership
ppt <- prop.table(table(buyer_all_class$buy.impt.house.howeownership))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    buy.impt.house.howeownership))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$buy.impt.house.howeownership))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$buy.impt.house.howeownership))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$buy.impt.
    house.howeownership)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$buy.impt.
    house.howeownership)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.impt.
    house.howeownership)%>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
addmargins(round(prop.table(table(buyer_all_class$buy.impt.house.howeownership,buyer_all_
    class$group),2),2),1)

#tests for importance level - housing type
ppt <- prop.table(table(buyer_all_class$buy.impt.house.type))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    buy.impt.house.type))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$buy.impt.house.type))%>%
  as.data.frame()
```

```
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.impt.house.type))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$buy.impt.house.type)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$buy.impt.house.type)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.impt.house.type)%>%
  as.data.frame()
ppt
addmargins(round(prop.table(table(buyer_all_class$buy.impt.house.type,buyer_all_class$group),2),2),1)
chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)

## Residential choices
table(buyer_all_class$buy.house.size)
table(buyer_all_class$buy.house.type)
# housing density type
### define High-density (Apartment with 5 or more storeys);
### in-between (apartments less than 5 storeys); and low-density housing
#Low density defined as single detached houses on medium to large properties
#High density defined as small detached dwellings, townhouses, condominiums and apartments
buyer_all_class <- mutate(buyer_all_class,
                  buy.house.type.density= ifelse(buy.house.type == "Single-detached house", "Low-density housing",
                                                 ifelse(buy.house.type == "Apartment with 5 or more storeys",
                                                   "High-density housing", "Medium-density housing")))
buyer_all_class$buy.house.type.density <- factor(buyer_all_class$buy.house.type.density,
                                                 levels = c("Low-density housing", "Medium-density housing", "High-density housing"))
table(buyer_all_class$buy.house.type.density)

# residence size
# small - less than 1000; medium - 1001-2500; large - greater than 2500
buyer_all_class <- mutate(buyer_all_class,
```

```r
                   buy.house.size.class= ifelse(buy.house.size == "Less than 1000", "Small
                       housing",
                                                    ifelse(buy.house.size == "2500 - 2999" |
                                                        buy.house.size == "More than 2999",
                                                           "Large housing", "Medium housing")
                                                               ))

buyer_all_class$buy.house.size.class <- factor(buyer_all_class$buy.house.size.class,
                                                    levels = c("Small housing", "Medium
                                                        housing", "Large housing"))
table(buyer_all_class$buy.house.size.class)


ch1 <- prop.table(table(buyer_all_class$group,buyer_all_class$buy.house.type.density),2)
    %>%
  as.data.frame()
ch1$Var <- c("Housing type")

ch2 <- prop.table(table(buyer_all_class$group,buyer_all_class$buy.house.size.class),2)%>%
  as.data.frame()
ch2$Var <- c("Housing size")
#write.csv(buyer_all_class, "/Users/yukeysha/R_workspace/Latent_paper/buyer_all_class.csv
    ")

# chi-tests for home choices
## housing type
ppt <- prop.table(table(buyer_all_class$buy.house.type))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    buy.house.type))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$buy.house.type))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$buy.house.type))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$buy.house.
    type)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$buy.house.
    type)%>%
  as.data.frame()
```

```
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.house.
    type)%>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$buy.house.type),1),2)

## housing size
ppt <- prop.table(table(buyer_all_class$buy.house.size))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    buy.house.size))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$buy.house.size))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$buy.house.size))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$buy.house.
    size)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$buy.house.
    size)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.house.
    size)%>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$buy.house.size),1),2)

## bedroom
ppt <- prop.table(table(buyer_all_class$buy.house.bedroom))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    buy.house.bedroom))%>%
  as.data.frame()
```

```
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$buy.house.bedroom))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$buy.house.bedroom))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$buy.house.
    bedroom)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$buy.house.
    bedroom)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.house.
    bedroom)%>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$buy.house.bedroom),1),2)

## full bath
ppt <- prop.table(table(buyer_all_class$buy.house.full.bath))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    buy.house.full.bath))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$buy.house.full.bath))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$buy.house.full.bath))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$buy.house.
    full.bath)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$buy.house.
    full.bath)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.house.
    full.bath)%>%
  as.data.frame()
```

```
chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$buy.house.full.bath),1),2)

## half bath
ppt <- prop.table(table(buyer_all_class$buy.house.half.bath))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    buy.house.half.bath))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$buy.house.half.bath))%>%
  as.data.frame()
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households"
    ,]$buy.house.half.bath))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$buy.house.
    half.bath)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$buy.house.
    half.bath)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.house.
    half.bath)%>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$buy.house.half.bath),1),2)

## garage
ppt <- prop.table(table(buyer_all_class$buy.house.garage))%>%
  as.data.frame()
pp1 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$
    buy.house.garage))%>%
  as.data.frame()
pp2 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Potential TOD households"
    ,]$buy.house.garage))%>%
  as.data.frame()
```

```r
pp3 <- prop.table(table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.house.garage))%>%
  as.data.frame()

p1 <- table(buyer_all_class[buyer_all_class$group=="Current TOD households",]$buy.house.garage)%>%
  as.data.frame()
p2 <- table(buyer_all_class[buyer_all_class$group=="Potential TOD households",]$buy.house.garage)%>%
  as.data.frame()
p3 <-table(buyer_all_class[buyer_all_class$group=="Car-dependent households",]$buy.house.garage)%>%
  as.data.frame()

chisq.test(p1$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p2$Freq, p=ppt$Freq, simulate.p.value = TRUE)
chisq.test(p3$Freq, p=ppt$Freq, simulate.p.value = TRUE)
ppt
round(prop.table(table(buyer_all_class$group,buyer_all_class$buy.house.garage),1),2)

# number of cars}
car <- addmargins(table(buyer_all_class$group,buyer_all_class$travel.N.of.cars),2)
mode1 <- addmargins(table(buyer_all_class$group, buyer_all_class$travel.previous.person.1.mode),2)
mode2 <- addmargins(table(buyer_all_class$group, buyer_all_class$travel.previous.person.2.mode),2)
mode3 <- addmargins(table(buyer_all_class$group, buyer_all_class$travel.previous.person.3.mode),2)
car
mode1
mode2
mode3
p.car <- addmargins(round(prop.table(table(buyer_all_class$group,buyer_all_class$travel.N.of.cars),1),2),2)
p.mode1 <- addmargins(round(prop.table(table(buyer_all_class$group, buyer_all_class$travel.previous.person.1.mode),1),2),2)
p.mode2 <- addmargins(round(prop.table(table(buyer_all_class$group, buyer_all_class$travel.previous.person.2.mode),1),2),2)
p.mode3 <- addmargins(round(prop.table(table(buyer_all_class$group, buyer_all_class$travel.previous.person.3.mode),1),2),2)
p.car
p.mode1
p.mode2
p.mode3
```

A7 Housing survey questionnaire

HOME BUYER AND HOME SELLER

SURVEY QUESTIONNAIRE

I. Motivations for Moving

- As a recent home mover, we first ask about your motivations for moving **(Page1-2)**

Then, this survey will mainly ask information on your home buying and home selling experience.

II. Home Seller Survey

- If you recently sold a home in Kitchener-Waterloo, please complete **Part A, B, C** (about 20 min)

Part A	Features of the home you sold (**Page3**)
Part B	Your home selling experience (Page5)
Part C	LRT and home selling (Page7)

III. Home Buyer Survey

- If you recently bought a home in Kitchener-Waterloo, please complete **Part D, E, F, G** (about 25 min)

Part D	Features of your new home (**Page10**)
Part E	Residential location choice (Page12)
Part F	Your home buying experience (Page20)
Part G	LRT and location choice (Page24)

IV. Household Characteristics and Travel Behaviour

- For all of you, please complete Part H (about 15 min)

Part H	Household characteristics (H1); travel behavior (H2); LRT and travel (H3) **(Page27)**

First, we would like to ask about your motivations for moving

Motivations for Moving

When you move,

- some factors might **push you out of your old home** (for instance, it's too small for your household size);
- some might **pull you to a new home** (for instance, you really love the parks and open space in the area, or it is close to friends or family);
- something may be **both a push and pull factor** (for example, your old house was too expensive, but the new one is very affordable).

The next question asks you to tell us which **push/pull factors** were important in your case.

Q1. What motivated you to move? (Please select all that apply)

Motivations		Why did you leave your previous home?	Why did you move to your new home?
Investment	For investment	☐	☐
Job change	Getting a new job	☐	☐
	Seeking new job opportunities	☐	☐
Life stage change	Getting married/partnership	☐	☐
	Separation/divorce	☐	☐
	Expanding family size	☐	☐
	Decreasing family size	☐	☐
	Supporting my or my partner's parents	☐	☐

Motivations		Why did you leave your previous home?	Why did you move to your new home?
Education	For my or my partner's education	☐	☐
	For child’s education/childcare	☐	☐
Neighbourhood	Environmental quality	☐	☐
House	Affordability	☐	☐
	Upsize	☐	☐
	Downsize	☐	☐
Market	Taking advantage of a buyer or seller's market	☐	☐
	Expect home prices to go down	☐	☐
	Expect home prices to go up	☐	☐
Accessibility	Accessibility to my or my partner’s workplace	☐	☐
	Accessibility to facilities (shopping and services)	☐	☐
	Accessibility to bus stops	☐	☐
LRT	LRT construction	☐	☐
	Accessibility to future LRT stops	☐	☐
	Anticipating future price increase due to LRT	☐	☐
Other	Please specify_____________________	☐	☐
Other	Please specify_____________________	☐	☐

Home Seller Survey

- Please complete Part A, B, C, if you recently sold a home in Kitchener-Waterloo.
- We will ask you about the features of the home you sold and the process of selling it.

PART A – First we will ask you about the features of the home you sold

Q2. What is the address of the home you sold?

Unit No. [ ]

House No. [ ]

Street Name [ ]

City [ ]

Postal Code [ ]

Q3. What type of home did you sell?

- ○ Single-detached house
- ○ Semi-detached house
- ○ Townhouse/row house
- ○ Apartment or flat in a duplex (with an upper and lower unit in same house)
- ○ Apartment in a building with fewer than 5 storeys
- ○ Apartment in a building with 5 or more storeys
- ○ Other, please specify... ______________________

Q4. When was this home built (approximately)?

[ ] Year **or** ☐ I don't know

Q5. Concerning the home that you sold, was it:

- ○ Freehold (you outright own the house and the land)
- ○ A cooperative (you own a share of the entire building)
- ○ A condominium (you own the unit and share ownership of common elements)

Q6. What is the approximate square footage of the home that you sold?

- ○ Less than 1000
- ○ 1000 -1499
- ○ 1500 -1999
- ○ 2000-2499
- ○ 2500-2999
- ○ More than 2999

Q7. Have you done major repairs or renovations since you bought the home (Please select all that apply)?

- ☐ No, only regular maintenance (cleaning, painting, furnace, etc.)
- ☐ Yes, minor repairs (missing or loose floor tiles or bricks, defective steps or sidings, etc.)
- ☐ Yes, major repairs (roof, electrical, plumbing, heating or structural repairs, etc.)
- ☐ Yes, I/We rebuilt the house.
- ☐ Repairs/renovations were done to get the house ready to sell.

Q8. What type of heating did you use in the home?

- ○ Electric
- ○ Gas
- ○ Oil
- ○ Other, please specify... ______________________

Q9. Please provide the number of each facility in your home

Bedrooms	
Full bathrooms (sink, toilet and shower/tub)	
Half bathrooms (sink and toilet)	
Floors (basement and attic/loft excluded)	
Garage or other covered parking spaces	
Other parking spaces	

Q10. If the home you sold is a condo, how much was your condo fee in the most recent month?

$ ☐ **or** ☐ I don't know

PART B – We will ask you about your home selling experience

Q11. Was this your (your household's) first experience selling a home?

○ Yes
○ No

Q12. When was your home sold? (Month/Year, e.g., "Jan. 2016 = 01/2016")

[ ]

Q13. When selling the home, I/we… (Please select all that apply)

☐ Used a REALTOR® or real estate agent
☐ Listed it by myself
☐ Sold it without listing (Please go to **Q18**)
☐ Other, please specify... ______________________

Q14. How did you or your REALTOR® decide on the list price of your home? (Please select all that apply)

☐ Using comparable sales
☐ Using historical trends
☐ Need to receive a minimum amount from the sale
☐ I don't know
☐ Other, please specify... ______________________

Q15. Did you revise your list price during the selling process?

○ Yes, I revised once
○ Yes, I revised more than once
○ No (Please go to **Q17**)

Q16. What was your initial list price?

$ [ ]

Q17. What was your (final) list price?

$ [ ]

Q18. How long did it take to find the buyer?

[ ] months, [ ] weeks, [ ] days

Q19. How many people sent offers to you?

[ ]

Q20. Which offer did you accept? (Please select all that apply)

- ☐ Highest price
- ☐ The first received (Please go to **Q25**)
- ☐ The first above asking price
- ☐ No contingencies (Common buyer's contingencies include inspection, appraisal, financing, and insurance.)
- ☐ Buyer's ability to close fastest
- ☐ Other, please specify... ______________________

Q21. When the first offer from the winning buyer came, …

- ○ I/We accepted the offer (Please go to **Q25**)
- ○ I/We made a counter-offer

Q22. When making the counter-offer, …

- ○ I/We countered the price of the offer, with all the other terms being unchanged
- ○ I/We countered other terms of the offer, with the price being unchanged
- ○ I/We countered both the price and the other terms in the offer

Q23. Was your counter-offer accepted?

- ○ Yes (Please go to **Q25**)
- ○ No

Q24. How many bids did you receive from the winning buyer?

- ○ Two
- ○ Three
- ○ More than three

Q25. What was the selling price of your home?

$ [ ]

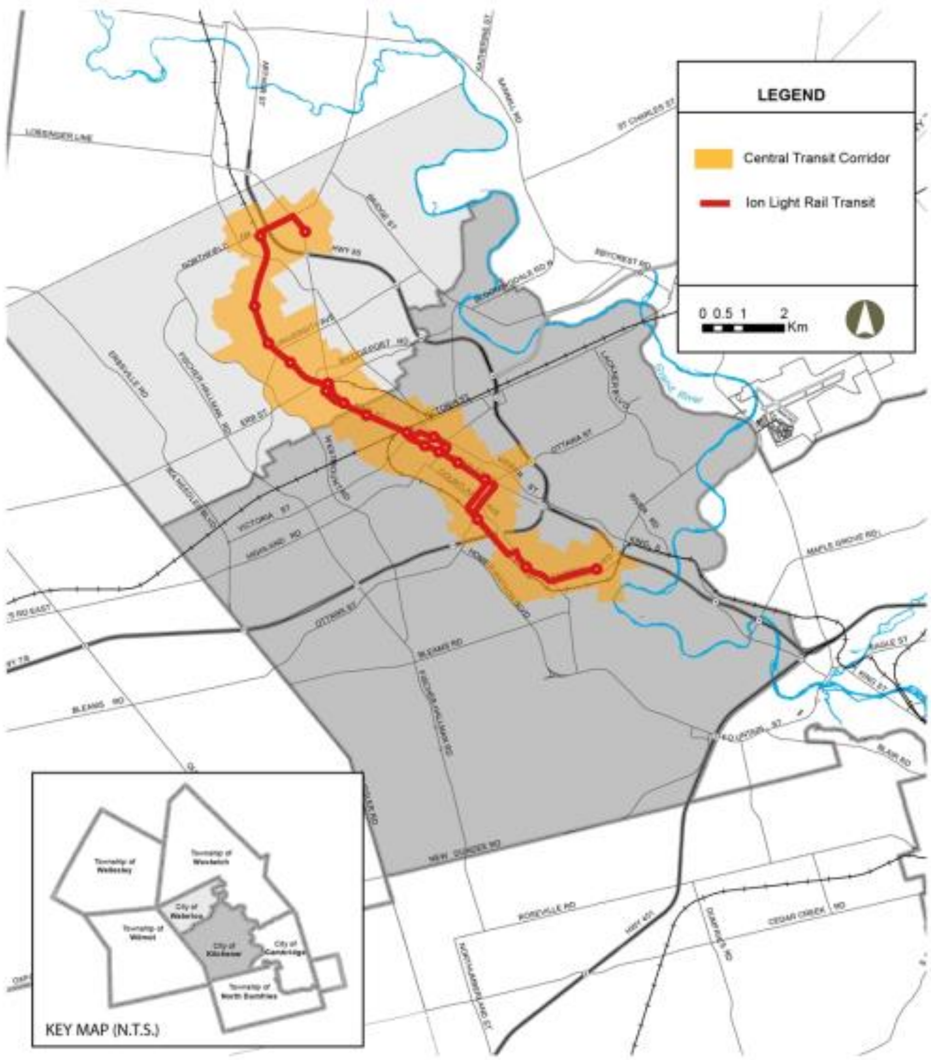
Q26. Do you agree to give us permission to access to your Realtor.ca listing? (This is not required to get the gift card.)

Kind reminder: we would like to learn more about your listing information to improve our studies of the Kitchener-Waterloo housing market, and we will keep your information strictly confidential.

○ I agree ○ I disagree

PART C - LRT and Home Selling

- ***As you may know, a 19-km light rail transit (LRT)*** line connecting Fairview Park Mall and Conestoga Mall is being built in Kitchener-Waterloo and is expected to begin service in early 2018.
- ***The map of the LRT line*** with future stops and the Central Transit Corridor (CTC) area is shown below (source: Region of Waterloo).
- ***The Central Transit Corridor (CTC)*** is defined as the area **within around 800 meters** or roughly a **10-minute walk distance** from the future LRT stops.



Q27. To what extent has the LRT influenced your selling decision?

○ Not important ○ Somewhat important ○ Very important

Q28. Was the home you sold inside the CTC area?

Note: there is a web-based lookup tool for you to check whether this home is inside the CTC area or not by clicking http://research.wici.ca/survey/ctc.html.

○ Yes
○ No (Please go to **Q30**)

Q29. Did any of the factors below influence your decision to sell the home inside the CTC area? (Please select all that apply)

☐ LRT construction
☐ Potentially heavier traffic in CTC area
☐ Potential crowding in CTC area
☐ Less safety in CTC area
☐ Less cleanness in CTC area
☐ More noise in CTC area
☐ Inconvenience for parking, driving, travelling with young children or doing groceries etc.
☐ The chance to profit from price increase
☐ Other, please specify... ______________________

Q30. After selling your home, did you buy another home?

○ Yes, I/we bought a home (Please go to **Q32**).
○ No, I/we rented a home.
○ No, I/we moved to another home that I/we previously bought (Please go to **Q32**)
○ Other, please specify... ______________________

Q31. Why did you choose renting instead of buying? (Please select all that apply)

☐ Can't afford mortgage/down payment
☐ Not being able to keep up with monthly payments
☐ Short term housing needs
☐ Convenience of renting process versus buying process
☐ Less responsibility (e.g. repairs and maintenance)
☐ No debt
☐ Easy to move
☐ Other, please specify... ______________________

Q32. Have you moved out of Kitchener-Waterloo, or are you planning to soon?

○ Yes
○ No (Please go to **Q34**)

Q33. Would you please state where you have moved or plan to move?

(then please go to **Part H on page27)**

Q34. Did you recently buy a home in Kitchener-Waterloo?

- ○ Yes (then we kindly invite you to answer the **homebuyer survey** starting from **PART D, and you can get two gift cards after completion**)
- ○ No (then Please go to **Part H on page27**)

Home Buyer Survey

- Please complete Part D, E, F, G, if you recently bought a home in Kitchener-Waterloo.
- We will ask you about the features of your new home and the process of finding it.

PART D - First we will ask you about the features of the home you bought

Q35. What is the address of your new home?

Unit No. ☐

House No. ☐

Street Name ☐

City ☐

Postal Code ☐

Q36. What type is your new home?

- ○ Single-detached house
- ○ Semi-detached house
- ○ Townhouse/row house
- ○ Apartment or flat in a duplex (with an upper and lower unit in same house)
- ○ Apartment in a building with fewer than 5 storeys
- ○ Apartment in a building with 5 or more storeys
- ○ Other, please specify... ______________________

Q37. When was the home built (approximately)? (Year)

☐ Year **or** ☐ I don't know

Q38. Concerning the home that you bought, is it ...?

- ○ Freehold (you outright own the house and the land)
- ○ A cooperative (you own a share of the entire building)
- ○ A condominium (you own the unit and share ownership of common elements)

Q39. Does your new home need any repairs? (Please select all that apply)

- ☐ No, only regular maintenance (cleaning, painting, furnace, etc.)
- ☐ Yes, minor repairs (missing or loose floor tiles or bricks, defective steps or sidings, etc.)
- ☐ Yes, major repairs (roof, electrical, plumbing, heating or structural repairs, etc.)
- ☐ Yes, I'm planning major renovation soon.
- ☐ Yes, I'm planning to rebuild it.

Q40. What type of heating do you use in your new home?

- ○ Electric
- ○ Gas
- ○ Oil
- ○ Other, please specify... ______________________

Q41. Please provide the number of each facility in your new home

Bedrooms	☐
Full bathrooms (sink, toilet, shower/tub)	☐
Half bathrooms (sink and toilet)	☐
Floors (basement and attic/loft excluded)	☐
Garage or other covered parking spaces	☐
Other parking spaces	☐

Q42. What is the approximate square footage of your new home?

- ○ Less than 1000
- ○ 1000 -1499
- ○ 1500 -1999
- ○ 2000-2499
- ○ 2500-2999
- ○ More than 2999

Q43. If your new home is a condo, how much was your condo fee in the most recent month?

$ ☐ **or** ☐ I don't know

Part E – Choosing your new home

Home Choice Decisions

When you choose a new home,

- you might choose the **house itself (Q44)** and the **neighbourhood (Q45)**
- there are factors that might be most **important** to you (for instance an easy commute or being near children's school). These factors influence your search and decision to buy a home.
- at the same time, buying a home involves trade-offs, and the home you buy might not quite be your **ideal** home.

The next set of questions ask you first about what is **most important** to you and next what your **ideal home and neighbourhood** would be like **in your case.**

Q44. Choosing the house itself

Q44-1. Please rate the importance of each feature in your home buying decision.

Residential features	Not important	Somewhat important	Very important
EXAMPLE	1	2	3√
Housing price	1	2	3
Housing type (e.g., single detached, townhouse, apartment)	1	2	3
Homeownership (e.g., freehold, condominium)	1	2	3
Housing size	1	2	3

Residential features	Not important	Somewhat important	Very important
Yard size	1	2	3
Age of your residence	1	2	3
Number of bedrooms	1	2	3
Number of full bathrooms	1	2	3
Number of half bathrooms	1	2	3
Number of floors	1	2	3
Garage or covered parking spaces	1	2	3
Total parking spaces	1	2	3
Ease of maintenance	1	2	3
Condo fee	1	2	3
Condo amenities	1	2	3
Other 1, please specify____________________	1	2	3
Other 2, please specify____________________	1	2	3

Q44-2. Please indicate your ideal home

- When looking for a new home, we sometimes make trade-offs depending on our budget or other considerations.
- Perhaps, you ideally wanted a single-detached home, but bought a row house after considering trade-offs and current opportunities.

***Please tell us the home that you desired most when buying your home, without considering your budget and any other trade-offs*.**	
1. Your ideal home **- Housing type**	○ Single-detached house ○ Semi-detached house ○ Townhouse/row house ○ Apartment or flat in a duplex (with an upper and lower unit in same house) ○ Apartment in a building with fewer than 5 storeys ○ Apartment in a building with 5 or more storeys ○ Other, please specify______________________
2. Your ideal home **- Own or rent**	○ Own ○ Rent
3. Your ideal home **- Homeownership**	○ Freehold (you own the house and the land) ○ Cooperative (you own a share of the entire building) ○ Condominium (you own the unit and share ownership of common elements)
4. Your ideal home **- Square footage**	○ Less than 1000 ○ 1000 -1499 ○ 1500 -1999 ○ 2000-2499 ○ 2500-2999 ○ More than 2999

5. Your ideal home - **Yard size**	○ No outdoor space ○ Patio or deck or balcony ○ Small yard (area of 0-4 single car garages) ○ Medium yard (area of 5-9 single car garages) ○ Large yard (area of 10-16 single car garages) ○ Very large yard (area of 17+ single car garages)
6. Your ideal home - **Built year range** **(please select all that apply to you)**	☐ No preference ☐ 2010-2016 ☐ 2005-2009 ☐ 2000-2004 ☐ 1990-1999 ☐ 1980-1989 ☐ 1970-1979 ☐ 1960-1969 ☐ 1950-1959 ☐ 1940-1949 ☐ 1930-1939 ☐ 1920-1929 ☐ before 1920
7. Your ideal home - **Number of bedrooms**	
- **Number of full bathrooms**	
- **Number of half bathrooms**	
- **Number of floors**	
- **Garage spaces or covered parking spaces**	
- **Total parking spaces**	

Other features, please specify____

Other features, please specify____

Q45. Choosing the neighbourhood where the house is

- Thinking about “neighbourhood” as the area within a **ten-minute walk** (or **1 KM**) of your house.
- Again, we ask you to tell us about the **importance of neighbourhood features** in your decision to buy and then to tell us what your **ideal neighbourhood** would be like.

Q45-1. Physical neighbourhood

- ***First, please rate the importance to your current neighbourhood selection***

IMPORTANCE	Not important	Somewhat important	Very important
Density of housing	1	2	3
Land use mix *	1	2	3
Easy to walk	1	2	3
Bicycle-friendly environment	1	2	3
Traffic noise	1	2	3

* Land use mix: e.g., mix of residential, retail, commercial or employment centre.

- ***Then, please indicate your ideal physical neighbourhood***

IDEAL NEIGHBOURHOOD	Low level	Medium level	High level
Density of housing	1	2	3
Land use mix *	1	2	3
Easy to walk	1	2	3
Bicycle-friendly environment	1	2	3
Traffic noise	1	2	3

* Land use mix: e.g., mix of residential, retail, commercial or employment centre.

Q45-2. Social neighbourhood

- ***First, please rate the importance to your current neighbourhood selection***

IMPORTANCE	Not important	Somewhat important	Very important
Similarity of … to yourself - household size	1	2	3
- household income	1	2	3
- education level	1	2	3
- age	1	2	3
- ethnicity	1	2	3
Safety level	1	2	3
School quality	1	2	3

- ***Then, please indicate your ideal social neighbourhood***

IDEAL NEIGHBOURHOOD	Low level	Medium level	High level
Similarity of … to yourself - household size	1	2	3
- household income	1	2	3
- education level	1	2	3
- age	1	2	3
- ethnicity	1	2	3
Safety level	1	2	3
School quality	1	2	3

Q45-3. Accessibility

- ***First, please rate the importance to your current neighbourhood selection***

IMPORTANCE	Not important	Somewhat important	Very important
Commuting time	1	2	3
Commuting cost	1	2	3
Accessibility to… - school	1	2	3
- workplace	1	2	3
- retail and services	1	2	3
- public open space	1	2	3
- urban center	1	2	3
- bus stops			
- future LRT stops	1	2	3
Distance to… - previous neighbourhood	1	2	3
- your family/friends	1	2	3
- highway exits	1	2	3

- ***Then, please indicate your ideal accessibility levels***

IDEAL NEIGHBOURHOOD	Low	Medium	High
Commuting time	1	2	3
Commuting cost	1	2	3
Accessibility to… - school	1	2	3
- workplace	1	2	3
- retail and services	1	2	3
- public open space	1	2	3
- urban center	1	2	3
- bus stops	1	2	3
- future LRT stops	1	2	3
Distance to… - previous neighbourhood	1	2	3
- your family/friends	1	2	3
- highway exits	1	2	3

Q46. Were any other selling points important to you that we did not list here?

PART F – Home Buying Experience

Q47. Were you (your household) renting before buying this home?

○ Yes ○ No

Q48. What are the main reasons that you (your household) chose buying instead of renting? (Please select all that apply)

- ☐ Build home equity
- ☐ Stability (stay in your home as long as you want)
- ☐ Liberty (free to make customizations or renovations to home)
- ☐ Take advantage of the low interest rate
- ☐ Investment
- ☐ Due to LRT
- ☐ The pride of being a home owner
- ☐ More affordable
- ☐ Lack of availability of suitable rental
- ☐ Other, please specify... ____

Q49. How many home(s) have you (your household) bought before buying this home?

- ○ 0
- ○ 1
- ○ 2
- ○ More than 2

Q50. When did you buy this home? (Month/Year, e.g., "Jan. 2016= 01/2016")

☐

Q51. Did you buy a newly constructed home or a previously owned home?

- ○ A newly constructed home
- ○ A previously owned home (Please go to **Q53**)

Q52. Why did you buy a newly constructed home instead of a previously owned home? (Please select all that apply)

- ☐ Quality construction
- ☐ Great home design/custom designs
- ☐ Easy maintenance
- ☐ Well-designed neighbourhood (open space, trails, schools, etc.)
- ☐ Warranty

- ☐ Energy-efficiency
- ☐ Indoor air quality
- ☐ Safety
- ☐ Reputable builder
- ☐ Investment (price appreciation)
- ☐ Other, please specify... ______________________

Q53. When buying the home, I/We… (Please select all that apply)

- ☐ Used a REALTOR® or real estate agent.
- ☐ Bought it directly from a seller.
- ☐ Bought it directly from a developer.
- ☐ Bought it from a friend.
- ☐ Inherited it.
- ☐ Other, please specify... ______________________

Q54. When deciding on the preferred location for your home, which source(s) of information did you rely on? (Please select all that apply)

- ☐ REALTORS®
- ☐ Realtor.ca
- ☐ Other websites (e.g., Kijiji, FSBO (For Sale By Owner))
- ☐ Social media
- ☐ Friends/family
- ☐ Newspaper
- ☐ Personal experience
- ☐ Other, please specify... ______________________

Q55. How long were you searching for a home before the final transaction?

☐ months, ☐ weeks, ☐ days

Q56. How did you choose the homes to visit? (Please select all that apply)

- ☐ By geographical area (familiar neighbourhood/desired neighbourhood)
- ☐ By price range
- ☐ By housing type/features (single detached homes/condos)
- ☐ Near a certain school/workplace
- ☐ Suggested by my REALTOR®
- ☐ Other, please specify... ______________________

Q57. If applicable, please list names of the neighbourhoods/areas (for example Mary-Allen, St. Mary's hospital, etc.) or names of the intersections (for example King/Victoria, Erb St W/Fischer-Hallman Rd N, etc.) where you have looked for homes.

Q58. How many homes did you visit before buying one of them?

- ○ 0 - 5
- ○ 6 -10
- ○ 11 - 20
- ○ More than 20

Q59. How many other homes in the same neighbourhood of your new home did you look at before buying the home?

Neighbourhood: the area within ***a ten-minute walk (or 1 KM)*** *of your house*

- ○ 0
- ○ 1-2
- ○ 3-4
- ○ More than 4

Q60. How many other homes did you bid on unsuccessfully before buying your new home?

Q61. What was the approximate budget for buying your new home?

$

Q62. What was the asking price of your new home?

$

Q63. Do you know how long your new home had been on the market before you bought it?

months, weeks, days

or ☐ I do not know

Q64. How did you decide the amount of your initial offer for the home which you bought? (Please select all that apply)

- ☐ As suggested by REALTORS®
- ☐ Comparable sales
- ☐ A fixed percentage below asking price
- ☐ A fixed percentage above asking price
- ☐ The maximum allowed by my budget
- ☐ Other, please specify... ______________________

Q65. For the home that you bought, what price did you offer initially/first?

$ ☐

Q66. When you sent the first offer, the seller…

- ○ Accepted the offer (Please go to **Q70**)
- ○ Made a counter-offer

Q67. When making the counter-offer, the seller…

- ○ Countered the price of the offer, with all the other terms being unchanged
- ○ Countered other terms of the offer, with the price being unchanged
- ○ Countered both the price and the other terms in the offer

Q68. Did you accept the counter-offer?

○ Yes (Please go to **Q70**) ○ No

Q69. How many offers did you send to the seller before the final transaction?

- ○ 1
- ○ 2
- ○ 3
- ○ More than 3

Q70. What was the final selling price of your new home?

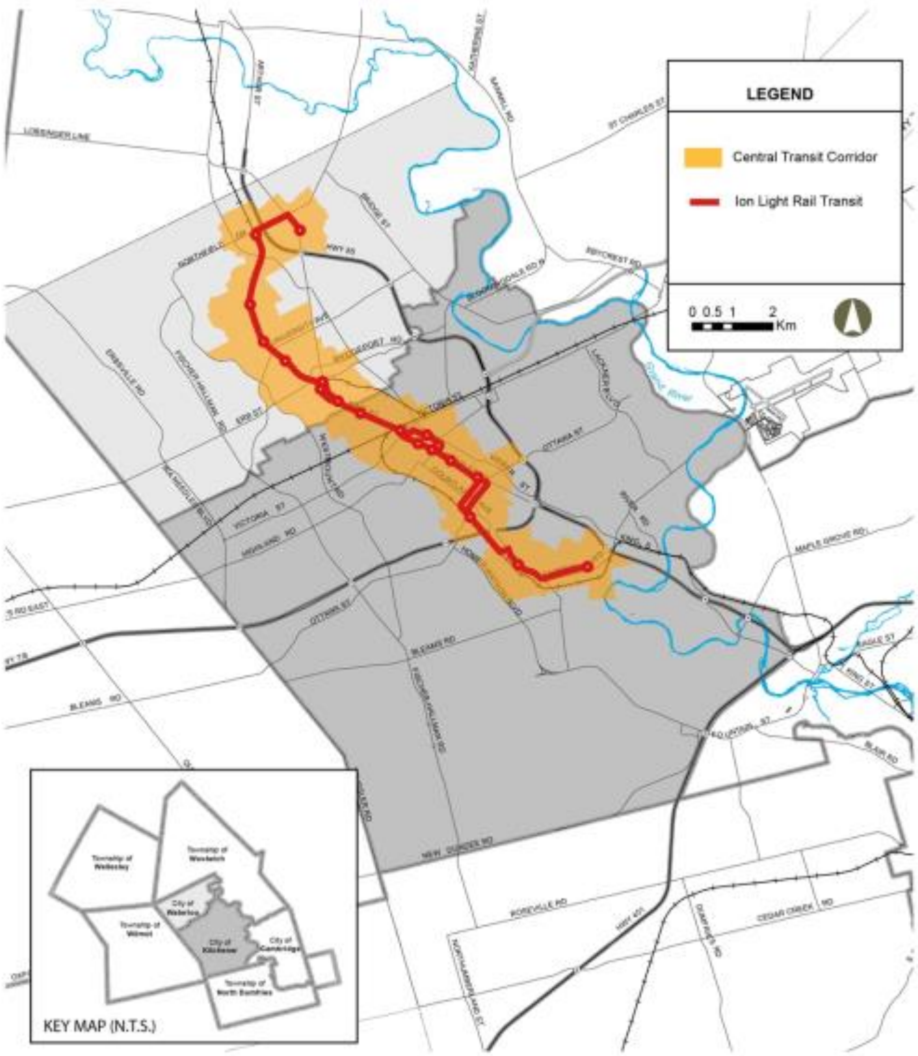
$ ☐

Q71. How much is your monthly mortgage payment if applicable?

$ ☐

PART G – LRT and Location Choice

- ***As you may know, a 19-km light rail transit (LRT)*** line connecting Fairview Park Mall and Conestoga Mall is being built in Kitchener-Waterloo and is expected to begin service in early 2018.
- ***The map of the LRT line*** with future stops and the Central Transit Corridor (CTC) area is shown below (source: Region of Waterloo).
- ***The Central Transit Corridor (CTC)*** is defined as the area **within around 800 meters** or roughly a **10-minute walk distance** from the future LRT stops.



Q72. To what extent has the LRT influenced your location choice decision?

○ Not important ○ Somewhat important ○ Very important

Q73. Is your new home inside the CTC area?

Note: there is a web-based lookup tool for you to check whether this home is inside the CTC area or not by clicking http://research.wici.ca/survey/ctc.html.

○ Yes
○ No (Please go to **Q75**)

Q74. What features of LRT, if any, have influenced your decision to buy your home inside the CTC area? (Please select all that apply; and then Please go to Q76)

☐ Faster than buses
☐ Quieter than buses
☐ More reliable than buses (on-time performance)
☐ Safer than buses
☐ More comfortable than buses
☐ Able to be productive during commuting
☐ Able to avoid traffic congestion
☐ Safer than driving
☐ Lower cost than driving (saving gas costs and parking rates)
☐ No need for finding parking
☐ Freeing up household car
☐ Environment-friendly
☐ Saving travel time
☐ Potential housing price increase due to LRT
☐ Other, please specify... ______________________

Q75. Did any of the factors below contribute to your decision to buy your home outside the CTC area? (Please select all that apply)

☐ LRT construction
☐ Potentially heavier traffic in CTC area
☐ Potential crowding in CTC area
☐ Less safety in CTC area
☐ Less cleanness in CTC area
☐ More noise in CTC area
☐ Inconvenience for parking, driving, travelling with young children or doing groceries etc.
☐ Not economical (higher housing price within CTC area)
☐ Other, please specify... ______________________

Q76. Did you live in Kitchener Waterloo before you bought your new home?

- ○ Yes
- ○ No, then please state where you lived before

Q77. Have you moved into your new home in Kitchener Waterloo?

- ○ Yes (Please go to **Part H**)
- ○ No, I am not moving in, because I bought this home for investment. (Please go to **Part H**)
- ○ Not yet, but I am planning to move in. (Please go to **Q78**)
- ○ Other, please specify…_______________________

Q78. When will you move to your new home in Kitchener Waterloo approximately? (Month/Year, e.g., "Jan. 2016= 01/2016")

PART H - Household Characteristics and Travel Behaviour

- A household is a person or group of persons living in the same residence. They do not have a usual place of residence elsewhere in Canada or abroad.
- This part includes **H1-**Household Characteristics, **H2-**Travel behaviour and **H3**-LRT and travel.

H1. Household Characteristics

Q79. Would you describe yourself as________? (Please select all that apply)

☐ Aboriginal (First Nations (North American Indian), Métis or Inuk (Inuit))
☐ White
☐ South Asian (e.g., East Indian, Pakistani, Sri Lankan, etc.)
☐ Chinese
☐ Black
☐ Filipino
☐ Latin American
☐ Arab
☐ Southeast Asian (e.g., Vietnamese, Cambodian, Malaysian, Laotian, etc.)
☐ West Asian (e.g., Iranian, Afghan, etc.)
☐ Korean
☐ Japanese
☐ Other, please specify... ______________________

Q80. If you were born in Canada, please select the province or territory in which you were born in. (Please go to Q82)

○ Newfoundland
○ Prince Edward Island
○ Nova Scotia
○ New Brunswick
○ Quebec
○ Ontario
○ Manitoba
○ Saskatchewan
○ Alberta
○ British Columbia
○ Yukon
○ North West Territories
○ Nunavut

Q81. If you were not born in Canada, how long have you lived in Canada?

[ ] years, [ ] months

Q82. What is the range of your household income before taxes (Gross income of all members) for year 2016?

- ○ Less than $29,999
- ○ $30,000-$49,999
- ○ $50,000-$74,999
- ○ $75,000-$99,999
- ○ $100,000-$149,999
- ○ $150,000-$249,999
- ○ $250,000-$499,999
- ○ $500,000 and over
- ○ Prefer not to answer

Q83. How many people are in your household including yourself?

[ ]

Q84. Please describe each of your household members

Note: If there is more than one person having the same relationship to you, please indicate them separately with a number. e.g., if you have 3 children, please enter **child 1**; **child 2**; **child 3** into the "**Relationship to you**" box.

Relationship to you	Sex			Age	Highest education				Labour force status						Transit pass	Driving license
	Male	*Female*	*Other*	Years	*Lower than high school*	*High school*	*Post-secondary*	*Graduate*	*Full time*	*Part time*	*Student*	*Retired*	*Unemployed*	*Other*		
EXAMPLE: ***My Father***	√			*45*			√		√							√
Yourself																

H2. Travel Behaviour

Q85. How many cars does your household currently own or lease?

________Cars

Q86. Are you a member of any car-share organization? (For example, Community CarShare, Student CarShare)

○ Yes ○ No

Q87. Compared to 3 years ago, have there been changes in your travel habits? (Please select all that apply)

	More	Less
I drive my car	○	○
I use public transit	○	○
I walk	○	○
I cycle	○	○
Other, please specify...	○	○

Q88. Please rank the following seven types of activities in terms of its priority when your family makes decisions on its household travel schedule,

where **1** is the **highest** priority, and **7** is the **lowest** priority activity type that may be deferred to another day.

_____ School / Work Activities

_____ Service Activities (e.g. visiting banks or other services)

_____ Grocery Shopping/Farmer's Market

_____ Chaperone Activities (e.g. accompanying others to their own activities)

_____ Social Activities (e.g. meeting with friends or family, attending events, or helping others)

_____ Recreational Activities (e.g. exercising, playing team sports, or visiting parks)

_____ Other Shopping Activities (e.g. shopping for housewares, clothing or other personal items)

H2-1. Current Travel Behaviour

If you (your household) are currently living in Kitchener-Waterloo, please describe each of your household members' current travel behaviour from **Q89 to Q92**; otherwise, please go to **Q93**.

Q89. Please indicate your (your household) current travel behaviour

Relationship to you	Workplace/school location (postal code or name)	Commuting time - one way (min)	Current main commuting mode						Other commuting mode (please specify)
			Driving	*Car passenger*	*Walking*	*Cycling*	*School bus*	*Taking GRT*	
EXAMPLE: ***My Father***	*University of waterloo*	*15*	√						
Yourself									

Q90. How important is each of these factors in influencing your household's current commuting mode choice?

Factors	Not important	Somewhat important	Very important
Shortest commuting time	1	2	3
Cheapest commuting cost	1	2	3
Shortest waiting time	1	2	3
Reliable time schedule	1	2	3
Availability of owning car and travel by car	1	2	3
Vehicle that is environmental friendly	1	2	3
Safety of the travel mode	1	2	3
Healthy travel mode	1	2	3
Workplace or school is close to transit stop	1	2	3
Home is close to transit stop	1	2	3
Flexible schedule	1	2	3
Comfort/ freedom	1	2	3
Factors that influence driving (such as low traffic volume)	1	2	3

Q91. How does traffic congestion influence your current daily commute?

- ○ Not seriously
- ○ Somewhat seriously
- ○ Very seriously

Q92. In a typical week, how many days do you use public transit?

- ○ Every day (7 days)
- ○ Every weekday (5 days)
- ○ 3-4 days
- ○ 1-2 days
- ○ Rarely or never

H2-2. Previous Travel Behaviour

If you (your household) previously lived in Kitchener-Waterloo, please tell us each of your household members' travel behaviour at that time from **Q93 to Q96**; otherwise, please go to **Q97**.

Q93. Please indicate your (your household) previous travel behaviour

Relationship to you	Workplace/school location (postal code or name)	Commuting time - one way (min)	Previous main commuting mode						Other commuting mode (please specify)
			Driving	*Car passenger*	*Walking*	*Cycling*	*School bus*	*Taking GRT*	
EXAMPLE: ***My Father***	*University of waterloo*	*15*	√						
Yourself									

Q94. How important is each of these factors influencing your household's previous commuting mode choice?

Factors	Not important	Somewhat important	Very important
Shortest commuting time	1	2	3
Cheapest commuting cost	1	2	3
Shortest waiting time	1	2	3
Reliable time schedule	1	2	3
Availability of owning car and travel by car	1	2	3
Vehicle that is environmental friendly	1	2	3
Safety of the travel mode	1	2	3
Healthy travel mode	1	2	3
Workplace or school is close to transit stop	1	2	3
Home is close to transit stop	1	2	3
Flexible schedule	1	2	3
Comfort/ freedom	1	2	3
Factors that influence driving (such as low traffic volume)	1	2	3

Q95. How did traffic congestion influence your previous daily commute?

- ○ Not seriously
- ○ Somewhat seriously
- ○ Very seriously

Q96. In a typical week, how many days did you use public transit approximately when you lived at your previous home?

- ○ Every day (7 days)
- ○ Every weekday (5 days)
- ○ 3-4 days
- ○ 1-2 days
- ○ Rarely or never

H3. LRT and Travel

Q97. What is your general attitude towards the LRT system in K-W Region?

- ○ Very positive
- ○ Positive
- ○ Neutral
- ○ Negative
- ○ Very negative

Q98. Among the following features of the future LRT services, which might be important to you? (Please select all that apply)

- ☐ Transit fare
- ☐ Hours of operation
- ☐ Facilities for people with mobility restrictions
- ☐ Service frequency
- ☐ Shelter/Station facilities
- ☐ On time performance
- ☐ Convenience for walking to the ION stations
- ☐ Convenience for bus connections and transfers
- ☐ Availability of scheduling information
- ☐ Availability of mobile updated information
- ☐ Having helpful staff
- ☐ Crowdedness/comfort
- ☐ Wi-Fi
- ☐ Other, please specify... ______________________

Q99. I plan to use the LRT system for …? (Please select all that apply)

- ☐ School / work activities
- ☐ Chaperone activities (e.g. accompanying others to their own activities)
- ☐ Grocery shopping activities
- ☐ Farmer’s market activities
- ☐ Other shopping activities (e.g. shopping for housewares, clothing or other personal items)
- ☐ Service activities (e.g. attending medical appointments, visiting banks or other services)
- ☐ Social activities (e.g. meeting with friends or family, attending events, or helping others)
- ☐ Recreational activities (e.g. exercising, playing team sports, or visiting parks)
- ☐ I will not use LRT for any purpose
- ☐ Other, please specify... ______________________

Q100. How did you hear of this survey? (Please select all that apply)

- ☐ I received your survey package by mail
- ☐ I was contacted by REALTORS® in Kitchener-Waterloo
- ☐ I was contacted by Kitchener-Waterloo Neighbourhood Associations
- ☐ I was contacted directly by the researchers
- ☐ Other, please specify... ______________________

Q. Would you like to submit the survey?

- ○ Yes, I want to submit the survey.
- ○ No, I want to withdraw from the survey.

Thank you for your participation.

Please indicate below whether you would like to receive further updates on this project and an invitation to attend a briefing session on the results of this study, and whether you would like to receive a gift card based on the amount of the study that you have completed.

- ☐ **Yes**, I would like to receive further updates.
- ☐ **Yes**, I would like to receive an Amazon gift card.
- ☐ **Yes**, I would like to receive a Home Hardware gift card.
- ☐ **No**, I would not like to receive further updates or a gift card.

If you choose **Yes**, please enter your **email address** __________________**,** or **provide your name** ______ **and mailing address**________________________. We will send/email a feedback letter and/or your preferred gift card to you in the next step. Please refer to our study webpage (http://research.wici.ca/blogs/yu/home-buyer-and-seller-survey/) to check updates as well.

If you have any questions or concerns, please contact Yu Huang (yu.huang@uwaterloo.ca), Prof. Dawn Parker (dcparker@uwaterloo.ca) or Prof. Jeff Casello (jcasello@uwaterloo.ca) at the University of Waterloo, or you can fill out the additional comments box below.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#19555#). However, the final decision about participation is yours. If you have questions for the Committee contact the Chief Ethics Officer, Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

Additional Comments:

____

____

____