PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

# LATENT SEGMENTATION IN RESIDENTIAL LOCATION CHOICES

**TOMAS COX OETTINGER**

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences

Advisor:

**RICARDO HURTUBIA GONZALEZ**

Santiago de Chile, September 2020

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

# LATENT SEGMENTATION IN RESIDENTIAL LOCATION CHOICES

**TOMÁS COX OETTINGER**

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences

Members of the Committee:

**RICARDO HURTUBIA** DocuSigned by: EA0535DF546F4FD...

**FRANCISCO MARTÍNEZ** DocuSigned by: Francisco Martínez C. D4FFBFF51DA74A7...

**ERIC MILLER** DocuSigned by: Eric Miller 8724C0A69A364C1...

**JUAN DE DIOS ORTÚZAR** DocuSigned by: Juan de Dios Ortúzar 376BE4D3F7A24BE...

**SEBASTIÁN RAVEAU** DocuSigned by: Sebastián Raveau F. 95311A329DF1442...

**GUSTAVO LAGOS** DocuSigned by: Gustavo Lagos C. B558E6FD3B4641F...

Santiago de Chile, September 2020

*¡Tierra a la vista!*

Elena

**AGRADECIMIENTOS**

Para mis papas, de quienes aprendí sobre persistencia y prioridades, y siempre han estado ahí para apoyarme.

Para Paola. En las buenas y en las malas, siempre creíste en mí, y en tus palabras encontré la fuerza para seguir.

Para Elena, Gaspar y Florencia, por toda la alegría y belleza.

Para Ricardo. Por toda la paciencia y buenos consejos, tanto en lo académico como en lo personal.

**CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

# SEGMENTACION LATENTE EN LAS DECISIONES DE LOCALIZACION RESIDENCIAL

Tesis enviada a la Dirección de Postgrado en cumplimiento parcial de los requisitos para el grado de Doctor en Ciencias de la Ingeniería.

TOMAS COX OETTINGER

## RESUMEN

La forma en la que crece la ciudad está definida en gran medida por cómo diferentes actores eligen dónde localizarse en la ciudad, de lo cual depende en gran parte la sustentabilidad del desarrollo urbano. Por esto es importante modelar estas decisiones, entendiéndolas y cuantificando el rol jugado por distintas variables explicativas, con el objeto de alimentar políticas públicas bien informadas.

Para entender el comportamiento de estos actores, ya sean hogares, desarrolladores inmobiliarios, comercio, u otros, es importante reconocer la heterogeneidad de preferencias que subyacen en sus decisiones, evitando modelos demasiado generales. Esta heterogeneidad puede estar ligada tanto a las diferentes características de estos tomadores de decisiones, como también a la multiplicidad de contextos urbanos en los cuales se emplazan los inmuebles a elegir.

De acuerdo a esto, el objetivo principal de esta tesis es explorar y proponer métodos adecuados para definir segmentaciones que reflejen la diversidad del mercado de la vivienda, con aplicaciones al contexto chileno, centrándose en la interacción entre la heterogeneidad en las preferencias de los hogares y las características de la oferta de espacio construido y atributos urbanos de su contexto.

Del objetivo principal se desprenden tres objetivos específicos, los cuales se asocian a las tres partes de esta tesis.

**En la primera parte** se implementa un modelo de elección de localización bajo un enfoque *choice* con clases latentes, en el cual el tomador de decisión es el desarrollador inmobiliario, enfrentándose a un conjunto de alternativas para localizar un proyecto con características dadas previamente. El modelo se enfoca en los proyectos en áreas de expansión de Santiago, buscando modelar cómo cambian las preferencias de localización dado que el proyecto pertenezca a submercados diferentes, los cuales son definidos como clases latentes. Los resultados del modelo indican una importante polarización del mercado, según el cual hay dos estrategias bien definidas al elegir la localización: una de ellas asociada a proyectos más "exclusivos", los cuales rehúyen la densidad y se atraen mutuamente; y otra asociada a proyectos más "masivos", que se desarrollan cercanos a áreas consolidadas y satélites. El método probabilístico de clasificación permite observar que la gran mayoría de los proyectos son clasificados en uno u otro submercado, sin muchos proyectos en el intermedio, lo cual permite confirmar la polarización existente.

**En la segunda parte** se explora un modelo de clases latentes aplicado a la elección de localización, pero en el marco de un enfoque tipo *bid-auction*. Esta combinación, no reportada antes en la literatura y que denominamos "clases latentes espaciales", permite que la segmentación de las clases latentes se efectúe sobre los inmuebles rematados, lo cual implica poder generar clases de sectores en la ciudad. De esta forma, se propone un nuevo método para incluir heterogeneidad espacial en las preferencias. Este método es implementado para modelar la elección de localización en una muestra de hogares de la Encuesta Origen Destino de Santiago, mostrando un mejor ajuste que otros métodos y una segmentación de la ciudad coherente, que permite una mejor interpretación de los resultados.

**En la tercera parte** se aplica el enfoque de clases latentes espaciales, propuesto en la segunda parte, para explorar su capacidad de modelar la forma en que las preferencias de los hogares varían de acuerdo a si están observando un inmueble en una zona endógenamente clasificada como de Desarrollo Compacto o no. Se exploran algunas aplicaciones, como la posibilidad de calibrar un índice de Desarrollo Compacto a partir de la función de clasificación del modelo de clases latentes. Complementariamente, se constata que la fórmula de elasticidad reportada en la literatura no es aplicable a los modelos bid ni para clases latentes, por lo que se deriva y aplica una fórmula de elasticidad para estos casos. El modelo permite observar que el nivel socioeconómico del entorno tiene mucho peso en la decisión de localización, siendo esto más relevante todavía en casos de Desarrollo Compacto. Esto muestra que las políticas de integración social en densidad son complejas de implementar y por lo tanto deben ser cuidadosamente diseñadas.

Miembros de la Comisión de Tesis Doctoral

Ricardo Hurtubia

Gustavo Lagos

Francisco Martínez

Eric Miller

Juan de Dios Ortúzar

Sebastián Raveau

Santiago, Julio 2020

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

**LATENT SEGMENTATION IN RESIDENTIAL LOCATION CHOICES**

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences

TOMAS COX OETTINGER

## ABSTRACT

Urban growth is largely defined by different actors choosing where to locate in the city, on which the sustainability of urban development strongly depends. For this reason, it is important to model these decisions —understanding them and quantifying their variables— to better inform urban public policy.

To understand the behavior of these actors, whether they are households, real estate developers, commerce, or others, it is important to recognize the heterogeneity of preferences that underlie their decisions and avoid generalizing models. This heterogeneity can be linked both to the different characteristics of these decision makers, as well as the variety of urban contexts where available properties are located.

Accordingly, the main objective of this research is to explore and propose suitable methods to define segmentations that reflect the diversity of the housing market, with case studies in the Chilean context, focusing on the interaction between heterogeneity in household preferences and the characteristics of the built space and attributes of its urban context.

Three particular objectives emerge from the main objective, which are associated with the three parts of this thesis:

**In the first part,** a location choice model with latent classes is implemented, in which the decision maker is the real estate developer who faces a set of alternatives to locate a project with given characteristics. The model focuses on projects in expansion areas of Santiago. The objective is to understand how the location preferences change depending on the project belonging to different sub-markets, which are defined as latent classes. The results of the model indicate an important polarization of the market, according to which there are two well-defined strategies when choosing a location. One strategy is associated with more “exclusive” projects, which avoid density and attract similar projects, while the other is associated with more "massive" projects, which are developed close to consolidated areas and urban satellites. The probabilistic classification method allows us to observe that the vast majority of projects are classified into either one or the other of these two submarkets with remarkably high probabilities, without many projects in between, which confirms the existence of a strong polarization.

**In the second part,** a latent class approach for location choice is again explored but within the framework of a bid-auction model. This combination, not previously reported in the literature and which we call “latent spatial classes”, allows latent class segmentation to be carried out on the auctioned properties, which generates classes of locations in the city. Thus, a new method is proposed to include spatial heterogeneity in preferences. This method is implemented to model the location choice in a sample of households from the Santiago Origin Destination Survey, showing a better fit than other methods and a coherent segmentation of the city that permits a clearer interpretation of the results.

**In the third part,** the latent spatial classes approach proposed in the second part is applied to explore its ability to model the way in which household preferences vary according to whether or not they are observing a property in an area classified as Compact Development. Some applications are explored, such as the possibility of calibrating a Compact Development index from the latent class model classification function. In addition, it is found that the elasticity formula for logit models reported in the literature is not applicable to a bid model or to latent classes, so an elasticity formula is derived and applied for these cases. The model reveals the importance of an area’s socioeconomic level in the location decision, this being even more relevant in cases of Compact Development. This result shows that any policy aiming for social mixing in dense areas is complex to implement and therefore must be carefully designed.

Members of the Doctoral Thesis Committee:

Ricardo Hurtubia

Gustavo Lagos

Francisco Martínez

Eric Miller

Juan de Dios Ortúzar

Sebastián Raveau

Santiago, July 2020

# 1. INTRODUCTION

"*The diversity that is generated by cities rests on the fact that in cities so many people are so close together, and among them contain so many different tastes, skills, needs, supplies, and bees in their bonnets*."

Jane Jacobs, The death and life of great American cities (1961)

Cities can be thought of as some of the most complex artifacts created by humans, and their complexity relies on the emergent nature of their origins and development, which results in dynamics that continuously defy our understanding.

People and built space make the city; its creators become part of their creation. *Flesh and stone*, as Richard Sennett (1997) would say. The continuous interrelation between these two systems is what makes cities strong and durable. People shape built space and built space shapes people's lives and relationships. To accommodate space to our activities, people build homes, workspaces, and infrastructure to move and socialize.

Stone is resistant and hard to shape; it takes the effort of multiple people to build a neighborhood and buildings may last much longer than a person's lifetime. Cities are built by an entire society, and people are continuously looking for ways to inhabit the outcome of years of building and rebuilding space.

This thesis, in a broad sense, is about understanding how people's needs and tastes determine how they compete for the diverse supply of existing built space. In this sense, the principal focus of this thesis is to acknowledge heterogeneity when modelling the interaction between people and built space.

The specific method to be used in this thesis is location choice modeling, which uses a microeconomic approach to characterize the behavior of agents when choosing a location. As described in the following section, these models relate observed household characteristics and unobserved variables such as preferences of households, to the chosen location, which possesses observed urban and built attributes. Econometric methods are used to identify the unobserved preferences, considering the observed decisions and attributes.

## 1.1 Location choice models

Location choice modeling is a tool increasingly used to understand and forecast demand for residential and other land uses in different areas of cities, in current or simulated scenarios, accounting for changes in transport infrastructure or building regulations, among others. They contribute to the efficient use of resources when locating infrastructure such as public transport and other services, by understanding the drivers of city growth, be it expansion or densification.

The importance of modelling location decisions has been recognized from Von Thünen (1826) onwards. These studies examine city distributions based on different types of agents seeking to maximize their utility by making a trade-off between the cost of transportation and the price or rent of a location, depending on individual characteristics such as income. This model was extended to the urban case by Wingo (1961), Alonso (1964), Muth (1969) and Mills (1967). These models correspond to what Wegener (2014) identifies as the first generation of location models, which were deterministic and focused on technical aspects such as urban growth and densities.

McFadden (1978) introduced stochasticity into location choice, through the use of discrete choice models, which gave a strong theoretical and mathematical basis to the discipline. Most location choice models presented nowadays are based on this framework, usually referred to as the “choice approach”.

The bidding approach of Alonso and stochasticity in McFadden were the basis for Ellickson´s (1981) bid-auction approach, where locations are assigned to the best bidder among a set of competing agents.

Location choice models are implemented mostly as a part of Transport and Land Use Interaction (LUTI) models. LUTI assumes that location choice depends on accessibility of the location, which is an output of the behavior of the transport system, which itself depends on the located demand for trips.

According to Wegener (2014), the 1990s brought a revival of LUTI models, as a second generation came about that concerned environmental, equity, and segregation issues. Now, the development of a third generation is underway which incorporates other dimensions in preferences, mainly related to the heterogeneity of agents, the modelling of latent characteristics, and greater spatial and temporal detail.

## 1.2 Latent classes for agent heterogeneity in location choice models

To better reproduce the observed behavior of a group of agents, one must address the diversity of preferences guiding their decisions. Normally, this heterogeneity is included in models as a two-step process: first, agents are divided into segments or types of agents using differences in agent characteristics as criteria (age, income, etc.); second, the model is estimated in order to identify a specific set of parameters for each type of agent. This sequential method induces estimation bias as it assumes that segmentation is error-free (Ben-Akiva et al., 2002), and it does not assure that agents, in the first step, are segmented in groups of agents with similar preferences, as preferences are only identified afterwards, in the second step. Another method to introduce heterogeneity is the definition of interactions between agent characteristics and location attributes. This technique does not present the aforementioned bias, but renders results that are more difficult to interpret and does not allow to identify types of decision-makers

Latent class models (Kamakura & Russell, 1989) can overcome this issue, as they use a one-step which jointly estimates classes and preferences. Therefore, class identification is informed by differences in preferences of agents. In simple terms, a latent class model introduces in the model a probability for the decision maker belonging to a specific class or segment, with each class having different preference parameters in their utility functions. Most implementations of latent class models in the transportation and location choice literature use a logit formulation for the class-membership probability, which depends on the characteristics of the decision maker and a set of parameters to be estimated jointly with preference parameters.

Although latent class models have been applied in different decision contexts, the first work that applies this methodology to location choice, to the extent of our knowledge, was presented by Walker and Li (2007). These authors propose a model in which the latent classes correspond to the different lifestyles of households choosing their location.

In more recent applications, the classes to which the decision makers belong are based not only on socioeconomic characteristics (income, age, education, etc.), but also on agents' attitudes regarding certain aspects of the city or other areas (Olaru, Smith, & Taplin, 2011); lifestyle as attitudes towards transportation and accessibility based on the development of transit-oriented-development (TOD) in the area (Meng, Taylor, & Scrafton, 2016; Smith & Olaru, 2013); preference for compact neighborhoods (Liao, Farber, & Ewing, 2014), smart growth neighborhoods (Lu, Southworth, Crittenden, & Dunhum-Jones, 2014), trade-off between having a private garden and park space in the neighborhood (Tu, Abildtrup, & Garcia, 2016), or predisposition towards telework and the effect of a more distant location (Ettema, 2010).

## 1.3 Spatial heterogeneity

Spatial heterogeneity is a special case of heterogeneity in general, in which model parameters are not stable in space (Anselin, 1999). For decision models, this means that preference parameters of agents are different depending on the location.

We can find an early example of spatial heterogeneity in Quandt (1958), who used different functions in a linear regression for different subsets of observations. Identifying the structure of this variation of parameters can be approached with different methods. A direct approach is to use zones defined exogenously to the model; for example, by administrative limits or by predominance of certain land uses, where each zone will have a different set of preference parameters that fit the observations within the zone. Exogenous zonification, however, presents the shortcoming of the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984), which recognizes that zone-based spatial analysis can have different outcomes depending on the zonification used. This means that, since exogenous segmentation can be arbitrary, so can be the results.

Some techniques address this issue by estimating location-specific parameters, using observations within a given distance (Chica-Olmo, 1995; Dubin, 1992) or using decreasing weights for observations based on distance to location (Fotheringham, Brunsdon, & Charlton, 2002; Páez, Long, & Farber, 2008). These approaches use “sliding neighborhoods”, where each observation is explained from values within a given distance, but are not adapted to address the spatial structure of the city.

In hedonic price models (Rosen, 1974), where the price of real estate is modelled as a function of location attributes, the issue of spatial heterogeneity has been dealt-with by defining zones as submarkets. These zones have been sometimes identified with methods such as the definition of functional zones like, for example, center and peripheries (Jang & Kang, 2015). Other models use two-step methods such as Principal Component Analysis or

Clustering to identify homogenous areas in terms of spatial attributes or homogenous sets of housing units in terms of their built characteristics (Bourassa, Hamelink, Hoesli, & Macgregor, 1999; Rosmera & Lizam, 2016). These methods, however, may introduce some bias as explained in section 1.2. As the objective of these techniques is to identify zones where same parameters apply, defining these zones from differences in built characteristics does not assure differences in estimated parameters, as they are only known in a second step.

### 1.4 Heterogeneity, segregation and the Chilean context

As pointed out extensively by Jane Jacobs (1969) and several other authors (Alexander, 1965; Batty, 2008; Glaeser, 2011), diversity is perhaps the most important feature in a city's origins and stability over time, and the source of its economic power and capacity to generate innovation. Yet, a growing amount of literature describes how people with different characteristics sort into segregated communities in the same city (Massey, 2016; Quillian, 2012; Sabatini, 2003), undermining the possibility to interact and generate innovation. This "diversity-segregation conundrum" as described by Florida (2017), is a characteristic feature of large and dense cities, and much more research is needed to understand the interdependence between the two.

In Chile, income and material wealth has increased in recent decades, but inequalities in aspects such as income, urban segregation and social mobility are a hard burden that persist over time (PNUD, 2017). Spatial segregation, associated with inequalities in access to opportunities and unequal built environment standards, is the most visible outcome of this dynamic in Chile's main city, Santiago (Sabatini, Cáceres, & Cerda, 2001). High income households live mainly in the northeast quarter of Santiago, with a residential "garden city" type of urbanization, while the rest of the city, except for the CBD, is characterized by extended peripheries of low-income households and industrial areas. In the last decades, this sectorial distribution has been broken down by the location of residential projects for medium and high income households in expansion areas not necessarily connected to high income areas, mostly under the typology of gated-communities (Sabatini, 2015).

## 1.5 Research opportunities

Inequalities and socio-spatial segregation are determinant in the spatial distribution of land uses, households and real estate supply, especially in developing countries. To understand the residential location choices in this context, it is important to acknowledge the complex definition of and interaction among different segments of households and also how the city is perceived as segmented in different zones or neighborhoods.

Two areas for possible contribution are identified, related to methodological and case study aspects:

i) Methodological: Spatial segmentation of the city in location choice models has only been addressed with two-step methods, which can induce estimation-bias, but this could be overcome by segmenting locations applying Latent Class Models to space.

ii) Case Study (evidence): The use of latent segmentation allows exploration of preferences in the Chilean context considering the definition and interaction of segments. This method helps to identify aspects such as the polarization of submarkets and the differences in preferences according to these submarkets. The approach taken in this thesis helps to build a better understanding of the emergence of the segmented urban spatial structure of Santiago.

## 1.6 General objective

The main objective of this research is to explore proper methods to define segmentations of the housing market in the Chilean context, focusing on the interaction of heterogeneity in household preferences with project supply characteristics and spatial attributes.

## 1.7 Specific objectives

The thesis is composed of three main stages, each focusing on a specific objective:

i) Model the heterogeneity in location strategies of housing projects in expansion areas, measuring their attraction to certain spatial attributes or urban elements and characterizing the possible polarization of projects into specific submarkets.

ii) Explore a novel method to treat heterogeneity in location choice models, using "latent spatial classes". This allows to endogenously define spatial submarkets, where households may have different preferences for urban attributes.

iii) Explore the application of “latent spatial classes” to report evidence on a specific Chilean policy, allowing one to observe the interaction between the latent spatial class of Compact Development Zones, and preferences of different types of households in these types of zones.

Each of these objectives is studied in three different models. The conceptual ground, methods, data and results are reported in each of the next chapters.

# 2. SUBDIVIDING THE SPRAWL: ENDOGENOUS SEGMENTATION OF HOUSING SUBMARKETS IN EXPANSION AREAS OF SANTIAGO, CHILE[1]

Tomás Cox and Ricardo Hurtubia

## ABSTRACT

Urban sprawl is a phenomenon observed in most cities around the globe and especially in Latin America, where it is associated to socioeconomic segregation. In the case of Chile, sprawl has been generally based on large real estate projects. Developers target their projects to different types of consumers, which translates into submarkets with a broad range of housing-unit's characteristics, but also different location strategies. This heterogeneity has been analyzed and measured in the literature, but quantitative studies have used exogenous or sequential methods to identify submarkets, leading to potential bias in the segmentation. In this chapter we propose an econometric model to measure location drivers for different types of real estate projects that fills this gap. The modelling framework is based on discrete-choice and latent-class models, allowing us to simultaneously identify market segmentations, and their particular location choice preferences, without the need of arbitrary

---

[1] This chapter is published online as a paper in journal "Environment and Planning B: Urban Analytics and City Science". August 25, 2020. https://doi.org/10.1177/2399808320947728.

or ex-ante definitions of submarkets. The model is applied to the city of Santiago, Chile. Results reveal two clearly different approaches taken by developers to produce housing, with one submarket of “exclusive” and more sprawling projects, and another submarket of “massive” and more density driven projects. Location strategies are very different between submarkets, reproducing the socio-spatial segregation already observed in the consolidated city.

## 2.1. Introduction

The horizontal growth of some contemporary cities, based on scattered private projects of single-family detached houses, has been a trend observed not only in Anglo-Saxon countries, with a long suburban tradition, but also in Latin American metropolitan areas in the last decades (Borsdorf, Hidalgo, & Sánchez, 2007; Webster, Glasze, & Frantz, 2002). This pattern in Latin American cities is the latest stage of the evolution from an originally compact shape, to a sectorial distribution in the last century and, finally, to a fragmented structure in recent decades (Borsdorf, 2003). In this scenario, most residential projects in expansion areas are built as “gated communities,” with emphasis on vigilance/security, social homogeneity and marketing campaigns based on the image of a suburban, high-standard lifestyle (Coy & Pöhler, 2002). As we will present later, the Chilean case (especially in Santiago) is no exception to this trend, although amplified by the existence of some market-oriented land use policies.

Originally, private projects in expansion areas of Chilean cities were associated to high income groups searching for a “garden city” life-style, but recent authors (Borsdorf et al., 2007; Borsdorf, Hildalgo, & Vidal-Koppmann, 2016) have pointed out the broad spectrum of households locating in these projects, from high-income to low-income groups, with each project being targeted to specific segments. While some authors have studied how the location of these projects produce accessibility and environmental conditions that often imply a burden to middle and low-income households living in them (Cáceres-Seguel, 2015, 2017; Gainza & Livert, 2013; Romero et al., 2012), there has not been much attention paid to understanding the heterogeneity in this market, especially in terms of location strategies. Although these authors have described the location of projects in terms of accessibility, spatial and geographical variables, the analysis is generally case-oriented and there has not been a systematic effort to measure differences in location drivers among different types of projects.

This real estate development pattern seems to be consistent with the existence of housing submarkets (Palm, 1978; Schnare & Struyk, 1976), although defined not only by product similarity (units) but also by spatial attributes, as proposed by Watkins (2001). Identifying and characterizing these submarkets is relevant to understand the logic behind the production of built space and the emergence of spatial and structural (i.e. housing characteristics) segmentations in expansion areas of the city, which can be one of the causes of fragmented urban sprawl and residential segregation (Massey & Denton, 1988).

This chapter proposes a model to understand location choice patterns of residential projects and their membership to different housing submarkets. The modelling approach, based on latent class models (Kamakura & Russell, 1989) and location choice models (McFadden, 1978), allows for identification of housing submarkets from the observed location data of residential projects through simultaneous estimation of location choice and market segmentation parameters. This is a contribution to the housing submarkets literature, where the problem has been generally analyzed following a two-step fashion, with market segments being defined prior to the estimation of location preferences or hedonic price parameters (Bourassa et al., 1999; Rosmera & Lizam, 2016; Schnare & Struyk, 1976).

To our knowledge, the model presented here is the first housing supply location choice model using latent classes to segment real estate projects according to their characteristics and location choice. Latent class models have been used before in location choice, but mostly to segment households according to their characteristics (Ettema, 2010; Liao et al., 2014; Lu et al., 2014; Olaru et al., 2011; Walker & Li, 2007).

While the model can be applied to understand location choices in any part of the territory, we believe it can be particularly useful to understand location strategies in areas where submarkets are not already well defined, such as expansion areas.  Therefore, the proposed modelling approach is applied to the case of Santiago, Chile using data describing all new real estate projects built in expansion areas between years 2004 and 2013 (accounting for 1,833 projects and 89,422 units). Estimation results confirm a very clear market segmentation, with significantly different housing location preferences between submarkets of projects. We argue this reflects social segregation, which clearly manifests spatially in consolidated areas of Santiago, and is now replicated in the sprawl.

This chapter is structured as follows. Section two provides an overview of the literature in the field of housing submarkets, location choice and agent heterogeneity. Section three presents the proposed model. Section four presents the model implementation, introducing Santiago as a case study, describing the data assembly and showing the estimation results. Finally, conclusions are presented.

### 2.2. Housing submarkets and location choice models

Housing markets are different from other markets for several reasons (Galster, 1996; O´Sullivan, 2012). In particular, they deal with heterogeneous quasi-unique goods (housing units) that usually have very high transaction costs. As most markets, they can be subdivided into submarkets, but there are some key differences that are relevant to this work.

Because of demographic, spatial and production factors, new housing products are heterogeneous but can be grouped in clusters or subgroups of nearly similar products with some internal variance, which has been studied as housing submarkets (Schnare & Struyk, 1976; Adair, Berry & McGreal, 1996; Galster, 1996; Goodman & Thibodeau, 1998; Watkins, 2001; Rosmera & Lizam, 2016, *inter alia*).

These submarkets can be correlated with social segregation patterns (Daniels, 1975; Hwang, 2015) by contributing to the emergence of homogenous neighborhoods. While spatial segregation, understood as the physical separation of two or more groups of agents into different areas of the city (Massey & Denton, 1988) is the result of individual location preferences with respect to the location of other groups or types of agents (Clark, 1991; Schelling, 1978), most theoretical and applied approaches trying to measure or describe segregation are based on exogenous definitions of types or groups of agents. While this makes these approaches intuitive and easily transferred to public policy, exogenous and/or fixed definition of groups has been criticized in the literature since this is clearly a complex process that depends on multiple variables (Wright, 2000). This is also the case in the discussion about segregation in Latin America and particularly in Chile (Ruiz-Tagle & López-Morales, 2014). We believe that the use of latent submarkets, as proposed in this chapter, can help to tackle this issue by using an endogenous segmentation process that helps not only to identify groups that tend to agglomerate (or segregate from each other) but also to measure their location preferences and, therefore, the drivers of segregation.

McFadden (1978) proposed modelling the residential location as a discrete decision, in which each household is a decision maker facing a set of locations (dwellings) as alternatives. Each alternative reports a utility to the household, which is a function of location attributes, dwelling price and household preferences. Alternatives with higher utility have a higher probability of being chosen (stochasticity is given by a random error term accounting for unobserved attributes and idiosyncratic behavior).

In location choice models, heterogeneity is the explicit differentiation of preference parameters by type of decision maker. This differentiation is usually defined exogenously, based on decision maker characteristics, such as income, car ownership and households' size (for a review see Schirmer, Van Eggermond, & Axhausen, 2014). Models for the location of residential supply considering heterogeneity of the developers are reviewed by Haider and Miller (2004) and Zöllig and Axhausen (2015).

Exogenous definitions of types of agents (and hence heterogeneity) cannot ensure an adequate and representative clustering of decision makers with similar preferences. To tackle this problem, endogenous segmentation techniques can be used. The most common approach for endogenous segmentation in location choice models is latent class modeling (Kamakura & Russell, 1989). These models estimate the probability of belonging to a certain class of decision maker as a function of her characteristics, while simultaneously estimating the preference parameters for each of the classes considered in the model. This approach is explained with more detail in section 2.3.

Latent class models have been used to account for heterogeneity in residential location choice (Ettema, 2010; Glumac, Han, & Schaefer, 2014; Ibraimovic & Hess, 2017; Liao et al., 2014; Lu et al., 2014; Smith & Olaru, 2013; Tu et al., 2016; Walker & Li, 2007), allowing for a better characterization of behavior. Latent class models applied to the problem of location choice for residential supply are not reported in the literature, to the extent of our knowledge.

## 2.3. A model for endogenous segmentation of housing submarkets

We propose a model where the decision makers are real estate developers. We assume each developer produces one project with given characteristics. Each developer chooses where to locate their project from all feasible locations in the study area, and their location preferences vary depending on the project characteristics (i.e. the submarket it targets).

In our model, submarkets are endogenously identified as a function of the project characteristics and location patterns. We assume each submarket targets a different type of consumer, whose willingness to pay for a dwelling in a specific location defines the price. Similar to households maximizing utility in standard location choice models as proposed by McFadden (1978), real estate developers are profit maximizers. Therefore, developers attempt to maximize their profit by choosing the best location for each project, depending on the submarket the project belongs to. However, submarket segmentation is not explicit and must be identified. We do this by assuming submarkets can be treated as latent classes, with each project belonging to a "latent submarket" with a probability, which is a function of its characteristics. The set of possible submarkets ($S$) is unknown to the analyst before estimation.

We model the profit of a project $n$ belonging to a submarket $s \in S$, decomposing the cost between development costs and land acquisition. The developers profit maximization problem then is:

$$\max_{i \in L} \pi_n(i|s) = R_{nis}(Z_i, X_n) - D_n(X_n) - L_i(\overline{Z}_i) \cdot q_n \tag{2.1}$$

where $\pi_n(i|s)$ is the expected profit per unit built in location $i$, given that the project containing it ($n$) belongs to the submarket $s$. This profit is a function of the expected price of a unit in project $n$ if it is built in location $i$, given that it belongs to submarket $s$ ($R_{nis}$). This price is a function of a vector of characteristics of the project $X_n$ and location attributes $Z_i$. We assume all units within a project have the same characteristics and, therefore, the same price. Development cost for a unit within project $n$ ($D_n$) is also a function of project characteristics ($X_n$). The development cost function may account for economies of scale due to the total number of units in the project, an attribute that can be included in $X_n$. The land acquisition cost is the product of land price per surface unit at location $i$ ($L_i$) and the amount of land required to build one dwelling within the project ($q_n$). The land price is also a function of location attributes ($\overline{Z}_i$) but in a different period, so we assume it to be exogenous in the rest of the formulation. We assume that the profit for each project is independent from the other projects' location decisions and, therefore, our model is not accounting for agglomeration economies.

The expected selling price ($R_{nis}$) is modeled using a linear-in-parameters specification, similarly to what is usually done in hedonic price models, where parameter-vectors $\rho_s$ and

$\beta_s$, which are submarket-specific, represent the marginal price of dwelling characteristics and location attributes respectively.

$$R_{nis} = \beta_s \cdot Z_i + \rho_s \cdot X_n \tag{2.2}$$

The estimated selling price, as well as development and land costs, may be subject to uncertainties derived from imperfect information, unobserved attributes or non-rational behavior. According to random utility theory (Domencich & McFadden, 1975), we can account for these uncertainties if we assume that the profit associated to each location alternative has a random error following an IID Gumbel distribution, and treating the decision process under an stochastic approach. This assumption, which renders a Multinomial Logit model (MNL), is frequently used in the location choice literature (see for example Hurtubia & Bierlaire, 2014; Martínez & Henríquez, 2007; McFadden, 1978; Walker & Li, 2007). A reason for this is the "Independence of Irrelevant Alternatives" property of the MNL, which allows to estimate a model using a sample of alternatives instead of the complete choice set, usually very large in this type of problem (Antoniou & Picard, 2015). Additionally, the MNL has the advantage of having a closed form, something that is particularly convenient for models with a latent class structure, and therefore computationally expensive to estimate, such as the one proposed in this work.

Therefore, the probability that a location alternative i reports the maximum profit among all alternatives, conditional to a particular submarket $s$, and therefore being chosen to build a project $n$ is:

$$P_n(i|s) = \frac{\exp\big(\mu \cdot \pi_n(i|s)\big)}{\sum_{j=1}^{J} \exp\big(\mu \cdot \pi_n(j|s)\big)} \quad \forall\, n, i, s \tag{2.3}$$

where $\mu$ is a scale parameter. Replacing (2.1) and (2.2) in (2.3), the location choice probability is:

$$P_n(i|s) = \frac{\exp\big(\widehat{\beta_s} \cdot Z_i + \widehat{\rho_s} \cdot X_n - \mu \cdot D_n(X_n) - \mu \cdot L_i \cdot q_n\big)}{\sum_{j=1}^{J} \exp\big(\widehat{\beta_s} \cdot Z_j + \widehat{\rho_s} \cdot X_n - \mu \cdot D_n(X_n) - \mu \cdot L_j \cdot q_n\big)} \quad \forall\, n, i, s \tag{2.4}$$

With some algebra, we can see that terms that are not specific to the location (number of units, development costs and the project characteristics in the expected price) can be cancelled-out:

$$P_n(i|s) = \frac{\exp(\beta_s \cdot Z_i - \mu \cdot L_i \cdot q_n)}{\sum_{j=1}^{N} \exp\big(\beta_s \cdot Z_j - \mu \cdot L_j \cdot q_n\big)} \quad \forall\, n, i, s \tag{2.5}$$

Therefore, as development costs are not part of the profit function in the location choice, any economies of scale due to number of units in the project are not relevant for modelling this particular decision. Economies of scale could be considered when defining the size of projects, but this decision is previous and exogenous to this model. It should also be noticed that development costs could have some variation across the city for the same project, but for modelling purposes we assume this variable to be constant across space.

The location probability of (2.5) is conditional to submarket $s$. We assume the membership of a project to a particular submarket is latent. However, following Kamakura and Russell (1989), this relation can be described through a class membership function $W_s(X_n, \theta_s)$. We assume this membership can be described by project characteristics ($X_n$) and their corresponding parameters ($\theta_s$), explaining how much a project $n$ fits into a submarket $s$. Making similar assumptions about unobserved attributes and stochastic behavior as in (2.3), the probability of a project $n$ belonging to submarket $s$ is:

$$P_n(s \mid X_n) = \frac{exp(W_{ns}(X_n, \theta_s))}{\sum_{g \in S} exp(W_{ng}(X_n, \theta_g)} \quad \forall\, s, n \tag{2.6}$$

Where $S$ is the set of possible project submarkets. Finally, following the latent class approach, the unconditional probability of choosing a location alternative $i$ is:

$$P_n(i) = \sum_{s} P_n(i|s) \cdot\ P_n(s|X_n) \qquad \forall\, i, n \tag{2.7}$$

Using equation (2.7), parameters $\beta_s$ and $\theta_s$ can be estimated through maximum likelihood using observed project location decisions, without requiring any information regarding submarket structure. This approach avoids an ex-ante definition of the membership of projects to submarkets and, instead, infers how developers perceive projects as part of a submarket, according to their characteristics and expected profit in different locations.

The estimation results allow the modeler to label each class according to the magnitudes and signs of parameters $\beta_s$ and $\theta_s$, assigning a “recognizable adjective” to each class, as it is done in the case study. The number of classes is defined exogenously, although the optimum number of classes could be found with an iterative and exploratory estimation process.

### 2.4. Santiago case study: Project-based expansion

To test our model, we propose as a case study the development of residential projects in the expansion areas of Santiago, Chile. With 6,123,000 habitants (INE, 2018), Santiago is by a large extent the main city of Chile, concentrating administrative power, services and commerce.

In this case study, we will focus on private residential projects built in suburban and expansion areas (outside the outer ring road) from 2004 to 2013. During this period, several urban highways were built, facilitating the development in areas that were previously hard to reach. Figure 2-1 shows the "centrifugal" evolution through time of the location of new real estate projects in the outskirts of the city, and how this correlates with the construction of urban highways.

These projects were regulated under a policy called "conditioned urban development zones[2]" which, from 1997 to present day, allows developers to urbanize rural areas, if certain basic requirements of connectivity and amenities are met. This means that real estate location is

[2] Called ZODUC, due to their acronym in Spanish

more the outcome of the developers' decisions than of discretional regulations; making this case study particularly suitable to be approached through econometric models.

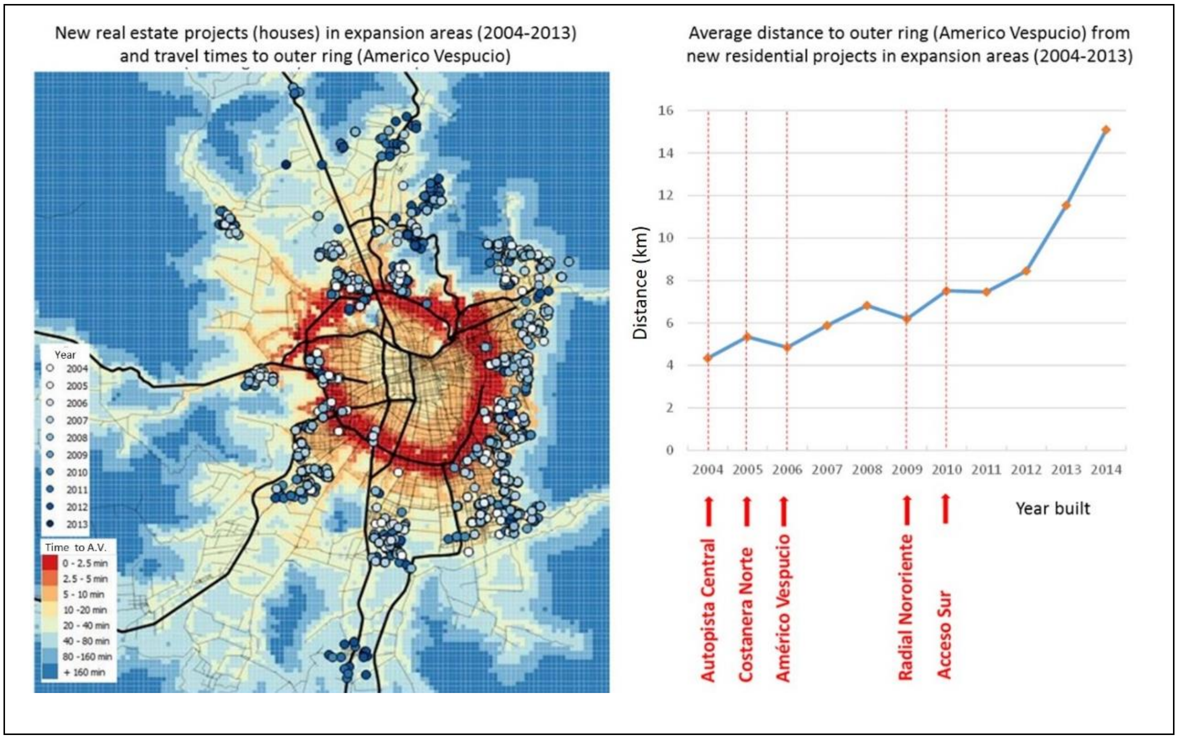



Figure 2-1: Location of residential projects (left) and average travel time to the outer ring road (right).

### 2.4.1. Model implementation and data

We applied the model described in section 2.3, to a database of residential projects in expansion areas. The class membership function $W_{ns}$ depends on intrinsic observed characteristics of the projects. Due to data limitations we had only a few characteristics: the number of units in the project, the average plot size and average listed asking price (we use average because units in the same project may vary in their characteristics, having slightly different prices and sizes). Although few, these variables are among the most relevant attributes describing a housing project (Hurtubia, Gallay, & Bierlaire, 2010). Due to the evolution of the urban growth process, some of the location attributes are updated for each year. The location model for a specific project will depend on attributes for the same period when it was built.

We divided the study area into a 175 x 175 grid, resulting in 30,625 cells of 500 by 500 meters. Each cell is a valid alternative in the location decision process but, because estimating a logit model with such a large choice-set (30,625 alternatives) would be inefficient and too expensive in computational terms, a sampling strategy was used following McFadden (1978). We use the observed location of the project as the chosen alternative while nine un-chosen alternatives were randomly sampled from all locations that were feasible.

Project data comes from a private cadaster of all residential developments built in Santiago's expansion areas (out of the main ring of the city, *Americo Vespucio*) from 2004 to 2013[3]. This database describes 1,833 residential projects accounting for a total of 89,422 new housing units. These projects represent approximately 26% of the total new housing supply in this period, according to own calculations based on intercensus growth (INE, 2002, 2018). Demographic attributes of the cells are obtained from the National Census (INE, 2002) and a socioeconomics segmentation provided by Adimark (2000). Land cost is available at an

[3] provided by the consulting firm Inciti (http://www.inciti.com/ )

aggregate spatial zoning for year 2014[4]. A road network topology, obtained from Open Street Map, was used to compute accessibility measures. Travel time is obtained through cost surface analysis (see Leusen, 1997). All variables describing location attributes ($Z_i$) and project characteristics ($X_n$) used in the model are described in Table 2-1.

### 2.4.2. Estimation results

The model described in section 2.3 is estimated using the statistical software Biogeme (Bierlaire, 2003) and considering two classes. Models with more classes were estimated, but the parameters were not significant, which can be interpreted as evidence of this market being polarized into two well-defined submarkets. For comparison purposes a base model with no latent classes (i.e. all projects have the same location preferences) was also estimated. Results are shown in Table 2-2.

In both models, most parameters were significant at the 95% confidence level, and signs and magnitudes are as expected, with a few exceptions that will be analyzed later. The latent class model considerably outperforms the basic model in terms of fit.

[4] provided by the consulting firm Transsa (http:// http://www.transsa.com/)

| Type | Attribute | Description | Sources |
|---|---|---|---|
| Cell (location) Attributes $\boldsymbol{Z_i}$ | Density | Number of households in a cell for each year divided by surface | Census 2002 (INE, 2002), Inciti and GFK project database(new projects) |
| | Location Socioeconomic Index | Ratio between the number of high income (ABC1 and C2) and low income households (D and E) in the cell (1) | Adimark (Market consulting) classification (ABC1: higher income, E: lower income) with Census 2002 data (INE, 2002) |
| | Land acquisition cost $(\boldsymbol{L_i \cdot q_n})$ | Average plot size multiplied by average land value in the cell (UF/m$^2$) (2) | Transsa Consulting ((2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)(2013)2013) |
| | Distance to hillsides | Average distance (km) of the cell to the nearest hillside in a 360° parse. | own calculation. |
| Travel Time (TT) | Travel time to CBD | Travel time (min) to nearest point in CBD (Alameda-Providencia axis) | own calculation based on openstreetmap roads. |
| | Travel time to nearest industrial zone | Travel time (min) to nearest industrial zone. | own calculation based on openstreetmap roads. |
| | Travel Time to nearest Highway | Travel time (min) to nearest existing highway for each year. | own calculation based on openstreetmap road. |
| | Travel time to nearest *satellite* | Travel time (min) to nearest cell with density of more than 7 Households per Há, outside the Santiago main continuous urbanized area. | own calculation based on openstreetmap roads and census 2002 (INE, 2002). |
| | Travel time to high price projects | Weighted average of travel time (min) to the 10% of highest price residential projects built the year before. Number of units is used as weight. | own calculation based on openstreetmap road (for travel time). Inciti and GFK project database (for new projects). |
| | Average travel time to low price projects | Weighted average of travel time (min) to the 10% of lowest price residential projects built the year before. Number of units is used as weight. | own calculation based on openstreetmap road (for travel time). Inciti and GFK project database (for new projects). |
| Developer´s project characteristics | Project Unit Price | Average price of units in the project (UF/m$^2$) (2) | Inciti and GFK project database |
| | Plot Size | Average plot size of the units in the project (m$^2$) | Inciti and GFK project database |
| | Number of Units | Number of houses built in the project. | Inciti and GFK project database |
| (1) Santiago Metropolitan Region has 541 censal districts. This index is based on a stratification methodology by Adimark (2014), where households are divided in five classes (ABC1, C2, C3, D and E) according to education and material belongings. (2) UF: Unidad de Fomento. Monetary unit that is re-adjustable according to inflation, which is equivalent to 42 dollars (August 2017) | | | |

Table 2-1: Attributes used in proposed models.

| | BASE MODEL (NO CLASSES) | SUBMARKET MODEL (2 CLASSES) | |
|---|---|---|---|
| Attribute parameters ($\boldsymbol{\beta}$) | Coefficient (t-test) | | |
| | | Class 1 ("Massive") | Class 2 ("Exclusive") |
| Travel time to high price projects (min) | -0.0519 (-13.92) | -0.0303 (-3.63) | -0.0609 (-14.25) |
| Travel time to low price projects (min) | -0.0255 (-5.59) | -0.0591 (-6.07) | 0.0118 (2) |
| Land acquisition cost (UF) | -0.0000668 (-3.38) | -0.000215 (-6.02) | |
| Density (Households/Hectare) | 0.000543 (5.9) | 0.00154 (10.11) | -0.000295 (-1.87)* |
| Location Socioeconomic Index | 0.0088 (2.41) | 0.00903 (2.03) | 0.0217 (3.42) |
| Distance to hillsides (m.) | -0.0151 (-1.81)* | 0.0851 (4.78) | -0.0568 (-5.71) |
| Travel time to CBD (min) | -0.0272 (-4.19) | -0.0712 (-5.34) | 0.00564 (0.75)** |
| Travel Time to nearest Highway (min) | 0.129 (15.81) | 0.24 (14.27) | 0.0407 (4.6) |
| Travel time to nearest industrial zone (min) | -0.122 (-17.32) | -0.19 (-12.63) | -0.0946 (-9.7) |
| Travel time to nearest satellite (min) | -0.00233 (-0.61)** | -0.0376 (-5.11) | 0.0209 (4.71) |
| Class Membership parameters ($\boldsymbol{\theta}$) | | Class 1 | Class 2 |
| Intercept | - | 63.6 (2.07) | - |
| Average unit asking price (UF/m2) | - | -1.29 (-2.08) | - |
| Plot size (m2) | - | -0.13 (-1.88)* | - |
| # Units (un) | - | 0.0775 (1.6)** | - |
| Null model log-likelihood | -3875.25 | -3875.25 | |
| Final log-likelihood | -2425.09 | -1926.59 | |
| Likelihood ratio test (against null model) | 2900.33 | 3897.32 | |

Table 2-2: Estimation results.

The class membership model (bottom of Table 2-2) shows parameters that affect the probability of belonging to class 1. By interpreting the signs of these parameters, class 1 can be labeled as a submarket of more "massive" projects, as they have a lower asking price[5], with smaller plot size and a higher number of units in the project. In contrast, class 2 projects can be labeled as belonging to a more "exclusive" submarket.

Several parameters in the latent class model change significantly with respect to the basic one. This is because the class-specific parameters are describing a much more coherent behavior. For example, travel time to low price projects, to the CBD and to the nearest satellite are all negative in the basic model but become positive for class 1 (massive projects) and remain negative for class 2 (exclusive projects) in the latent class model. A similar change is observed for density and distance to hillsides.

The interpretation of the parameters becomes much more intuitive in the latent class model. For example, both massive and exclusive projects prefer to locate near high price projects, but this is much more important for the exclusive projects while, at the same time, the exclusive projects try to locate as far as possible from low price projects (which is not the case for the massive ones). The case of the distance to hillsides variable is interesting, showing that high income households prefer to locate in enclosed or "protected" places, which can be interpreted as an extension in a topographic scale of the typology of gated communities (Borsdorf & Hidalgo, 2008; Webster et al., 2002), but in this case instead of

[5] This asking price is not the same as the expected selling price ($R_{nis}$) of equation (2.1)

crime, protecting themselves against new “undesirable” projects locating nearby. Travel time to CBD is significant for massive projects, which seem to prefer locations with good accessibility to employment centers. However this variable becomes irrelevant for exclusive projects, which is consistent with the observed trend where this type of development (usually associated to households with higher car ownership) tend to locate farther away from the consolidated city.

Although both classes value to have low travel times to certain amenities or desirable opportunities (e.g. high income projects, industry, CBD), which clearly benefit from the presence of highways connecting them, they also prefer locations far from the highways themselves. This, although seems to be contradictory, reflects how urban highways provide benefit to peripheral locations (by increasing their accessibility) but, simultaneously, are not desirable from a public space and externalities perspective.

The parameter for land acquisition cost is the scale parameter, following equation (2.5), but it cannot be confidently interpreted as such since the available data only provides a very coarse approximation for this attribute. Due to several unknown factors, such as the amount of time passed between the purchase of land and the construction of the project, or the interest rates involved in the transactions, the land cost attribute can be only interpreted as a proxy of the opportunity cost of developing that parcel.

### 2.4.3. Spatial distribution of sub-markets

Using the class membership parameters ($\theta$), the probability of belonging to the massive or exclusive sub-markets can be computed for every project in the database. Figure 2-2 shows the location of projects and their probability of belonging to the exclusive class. The spatial segregation is evident, with the north east part of the city clearly dominated by the Exclusive submarket (in red) and only one satellite in the west exit of the city breaking this pattern.

The histogram in Figure 2-3 (top) shows the empirical membership probability distribution for exclusive projects. Most of the projects can be clearly classified in one of the two submarkets. 47% of the projects fall in the 0 to 0.05 range of probability of being classified as exclusive (so they can be labeled as Massive), 36% in the range of 0.95 to 1 (clearly exclusive) and only 17% are in the wide intermediate range of 0.05 to 0.95 (yellow dots in Figure 2-2). This pattern shows that there is not a smooth transition from the exclusive to the massive submarkets, and that real estate decision makers strongly divide their location choices according to these two submarkets. This pattern is coherent with the strong social segregation and inequality observed in Chilean society (PNUD, 2017).

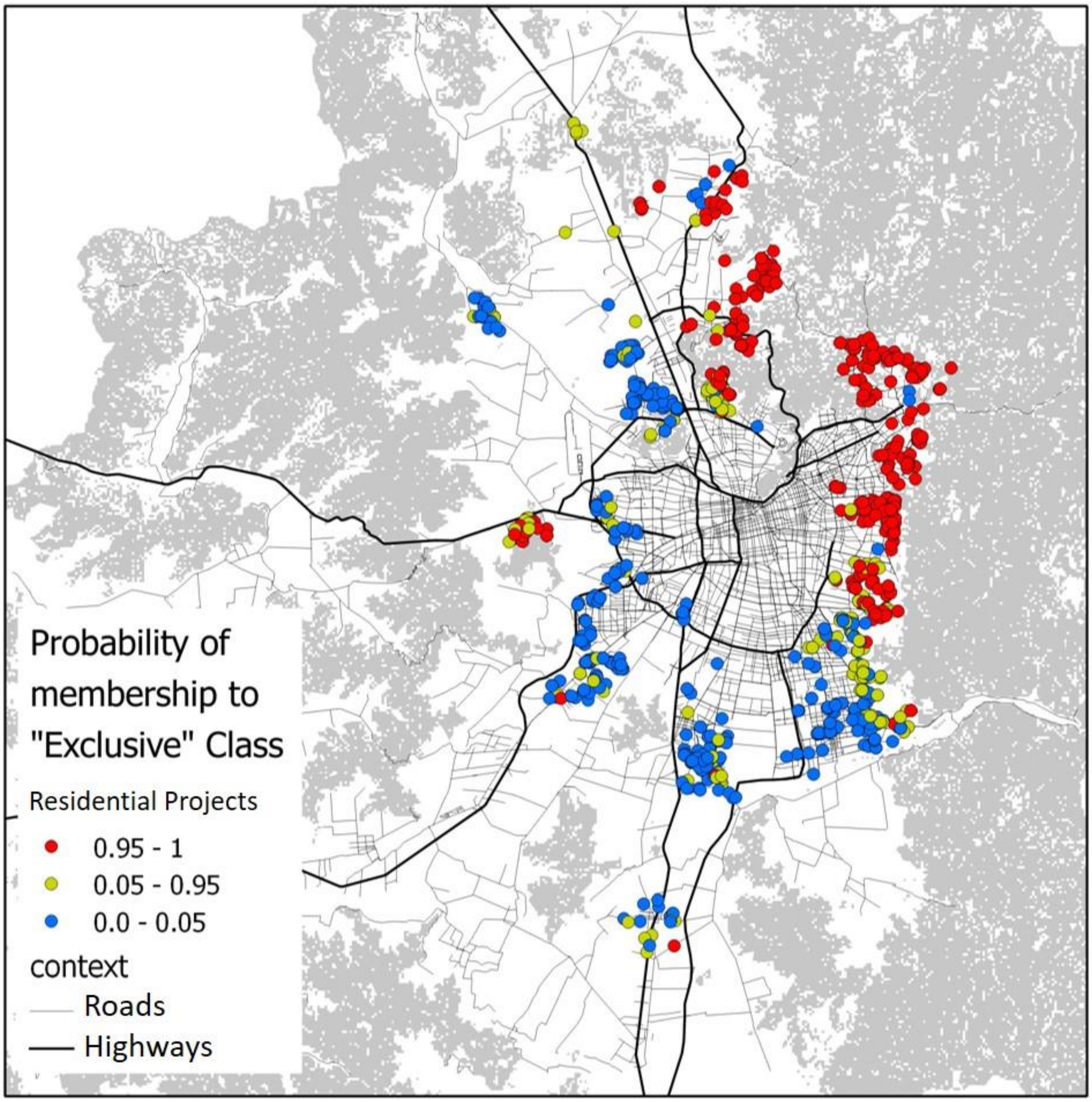



Figure 2-2: Location of projects and segmentation according to probability of membership to “exclusive” submarket.

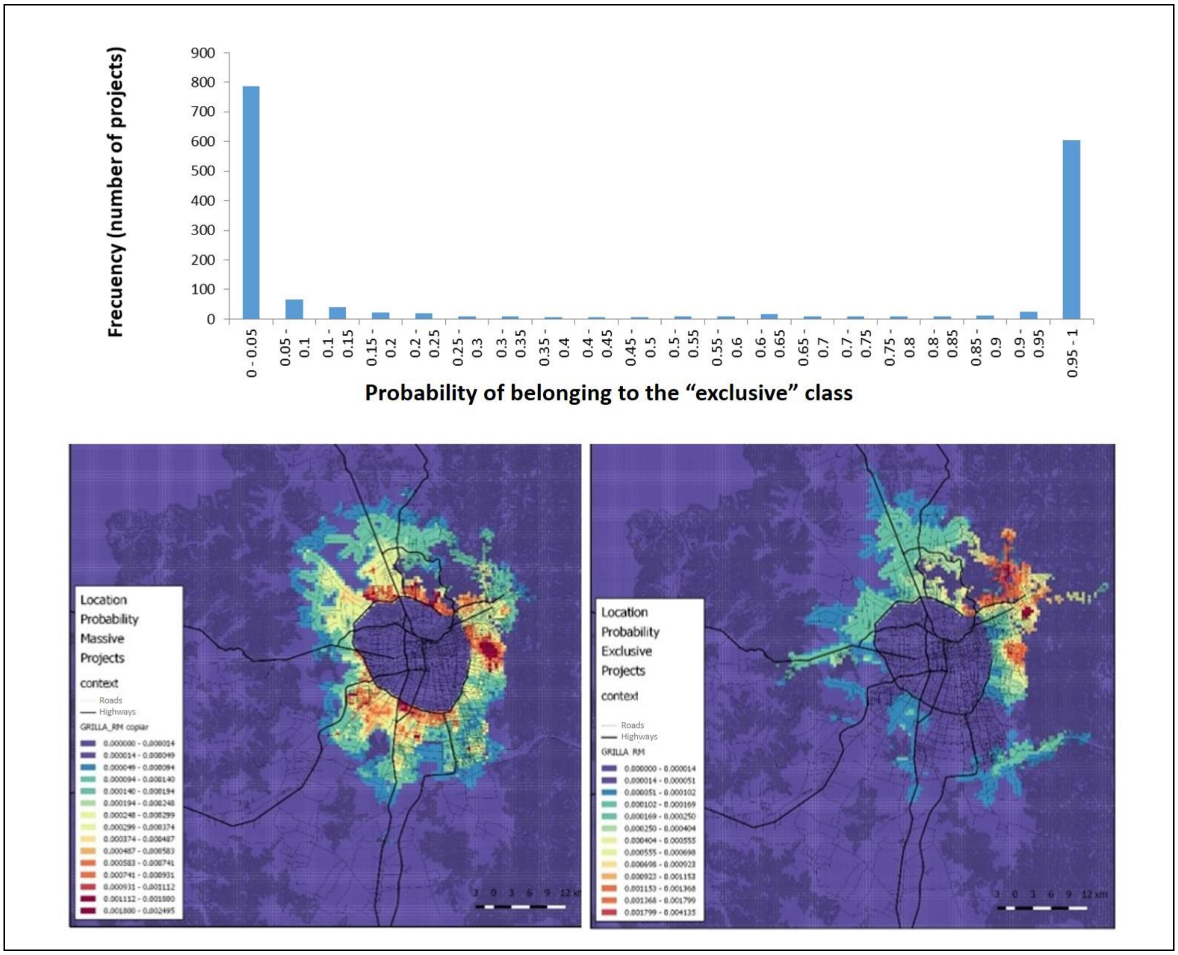



Figure 2-3: Histogram (top) with number of projects in each range of probability of membership to "exclusive" submarket. Spatial distribution of location probabilities for Massive (bottom left) and Exclusive (bottom right) projects.

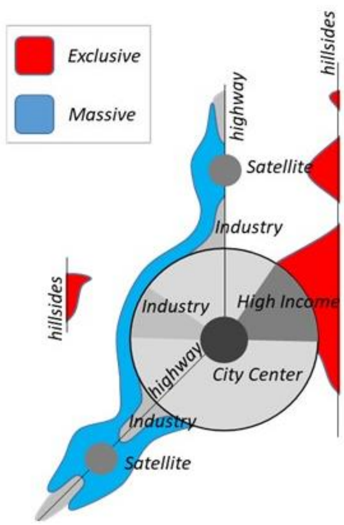
The extreme segmentation of projects into submarkets, with very different location strategies, is a clear result of deregulation and market-oriented land use policies implemented in Santiago, something well discussed in previous literature (Borsdorf & Hidalgo, 2008; Cox & Hurtubia, 2016; Heinrichs, Nuissl, & Rodríguez, 2009). Loose regulations allowed developers to produce housing targeted at very specific segments of the population, differentiating their products not only through unit or project characteristics, but also through location.

### 2.4.4. Location elasticities to urban elements

We calculate the aggregate elasticities for location choice probabilities with respect to each location attribute conditional to each project class. Depending on the sign and magnitude of the elasticities, shown in Figure 2-4, the attributes can be interpreted as "attractors" or "repellers" of location for each submarket. All the distance or travel time attributes are attractors if their sign is negative.

The most relevant attributes attracting the location of "massive" projects are low travel times to similar projects, to the city center and to industry areas. On the other hand, the most repulsive attributes for this submarket are low travel times to the nearest highway and closeness to hillsides. In the case of "exclusive" projects, the most relevant attractors are low travel times to similar projects and to industry areas. The most relevant attributes that make a location unattractive for this submarket are low travel times to satellite urban areas and high land costs.

In general, attributes related to accessibility play a much more relevant role in the location choice process than intrinsic attributes of the location itself (other than access). This quantification of "attraction and repulsion forces" for each type of project allows us to draw a schematic model of project-based urban expansion, which is shown to the left in Figure 2-4. This diagram represents two simplified location behaviors: While massive projects have a continuous and "attached" expansion from the city, exclusive projects expand mainly from the existing high income area in a "furtive" manner, or in isolated areas with their "backs against the slope".




| Massive | | | Exclusive | | |
|---|---|---|---|---|---|
| Elasticity | Attraction | Urban Element | Attraction | Elasticity | Attribute |
| -0.91 | (+) | High price projects | (+) | -1.06 | (Average travel time to high price projects) |
| -2.24 | (+) | Low price projects | (-) | 0.54 | (Average travel time to low price projects) |
| -1.49 | (+) | City Center | (-) | 0.11 | (Travel time City Center) |
| -0.94 | (+) | Satellites | (-) | 0.72 | (Travel time to nearest satellite) |
| 2.04 | (-) | Highways | (-) | 0.42 | (Travel Time to nearest Highway) |
| -0.94 | (+) | Industries | (+) | -0.81 | (Travel time to nearest industrial zone) |
| 1.10 | (-) | Hillsides | (+) | -0.44 | (Average distance to hillsides) |
| -0.25 | (-) | Plot Cost | (-) | -1.11 | (Average Plot Cost) |
| 0.58 | (+) | Households | (-) | -0.04 | (Density) |
| 0.03 | (+) | High SEL Households | (+) | 0.29 | (Location Socioeconomic Index) |

Figure 2-4: Diagram of location of projects according to attraction to urban elements (left), and model elasticities (right).

## 2.5. Conclusions

A model for location choice of real estate projects in expansion areas is proposed. The modelling framework makes it possible to simultaneously identify the parameters of a submarket classification function and the parameters of different expected price (and, therefore, profit) functions for each submarket, using a location decision model with latent class structure and, therefore, not requiring ex-ante definitions of market segments (although, the characteristics that are submarkets classifiers must be exogenously defined by the analyst). Thanks to a better representation of heterogeneity in the developers' preferences, the proposed model outperforms a basic location choice model in terms of fit, simultaneously allowing for a better understanding of urban growth drivers.

Estimation results confirm there are two clearly different classes of projects in expansion areas of Santiago de Chile, according to their characteristics and location preferences. This reveals an inherent link between the spatial (location) and structural (unit characteristics) segmentation that emerges in the housing production process, given a developer that tries to find the most profitable location for a project of certain characteristics. This is consistent with the submarket definition proposed by Watkins (2001), which asserts that structural and spatial attributes are both relevant dimensions in the market segmentation of housing.

Among the main findings is the clear distinction between expected price/profit (and therefore location preferences) for both submarkets. The polarization of the market is also a relevant finding, showing that the great majority of projects (83%) belong to one of the two market classes with more than a 95% probability. This seem to reflect segregation and inequality patterns observed in many other aspects of Chilean society and it is mostly the product of deregulation and market-oriented land use policies (such as the "conditioned urban development zones or ZODUC) which permit developers to target submarkets with large differences on their valuation of urban externalities and willingness to pay for spatial attributes.

**References** for all chapters are presented in a specific chapter after conclusions.

# 3. LATENT SEGMENTATION OF URBAN SPACE THROUGH RESIDENTIAL LOCATION CHOICE[6]

Tomás Cox and Ricardo Hurtubia

## ABSTRACT

Understanding the preferences of households in their location decisions is key for residential demand forecast and urban policy making. Accounting for preference heterogeneity across agents is useful for the modelling process but not enough to completely describe location choice behavior. Due to place-specific conditions, the same agent may have different preferences depending on the sector of the city considered as potential location, a phenomena known as spatial heterogeneity. Segmenting the city by defining zones where agents are supposed to behave similarly has been a common modelling solution, assigning different zonal preference parameters in the estimation process. This has been usually done with two-step methods, where spatial segmentation is done independently of the location choice process, something that could bias estimation results. We propose and test a one-step model for simultaneous estimation of location preference parameters and spatial segmentation,

[6] This chapter is under second round of revision for publication as a paper at journal "Networks and Spatial Economics".

therefore accounting for heterogeneity across agents and space. The model is based on Ellickson´s bid-auction approach for location choice and latent class models. We test our model with a case study in Santiago, Chile and compare it with other models for spatial segmentation. In terms of predictive power, our approach outperforms a model with no zones, a model with zones defined exogenously, and a clustering-based two-step model. This novel approach allows for a better conceptual ground for urban predictive models with spatial segmentation.

## 3.1. Introduction

Modelling households' location decisions is key to understand past and future patterns of urban growth and change, which helps to plan transport and services infrastructure, and to design policies that guide towards a better and more sustainable urban development.

The work by Alonso (1964) was the first to model the spatial distribution of different types of households according to their specific willingness to pay (WP) for a location as a function of its attributes, such as accessibility and built surface. In this model, each location is assumed to be auctioned, and the household with the highest bid (correlated with its WP) "wins" the location. This defines both the spatial distribution of households and the prices of real estate goods. Most present models of location choice are based on later formulations by McFadden (1978) and by Ellickson (1981). In both approaches, households evaluate each location in terms of its attributes and dwelling characteristics, and their probability of choosing the location depends on this evaluation (a utility function in McFadden's and a WP function in Ellickson's).

In location choice models, WP functions are generally based on the interaction between a vector of location attributes and agent characteristics, and a vector of unobserved parameters that represent the marginal contribution of each attribute and characteristic to the WP. These parameters are specific to each type of agent and are modelled as to represent their preferences in the choice process. Besides demographic heterogeneity we can also find heterogeneity across (types of) places, known as spatial heterogeneity. As demographic or agent heterogeneity can be included by segmenting agents according to their characteristics, also spatial heterogeneity can be included by segmenting locations according to their spatial attributes and assigning segment-specific preferences.

In this chapter we argue that spatial segmentation and how it affects valuation of attributes (preferences) is a complex issue, as it is part of the spatial cognition of city dwellers. If we want to achieve a spatial segmentation that maximizes the likelihood that parameters are representative of preferences in each zone, we cannot predefine zones only from differences in built or urban attributes as it is done, for example, in cluster analysis (Jain, 2010; MacQueen, 1967). Location preferences should play an active role in the spatial segmentation process, something that can't be achieved when segmentation is defined before the estimation of preference parameters, but can be done with a joint estimation of both spatial segmentation and preferences.

In order to do so, we propose a model based on the bid-auction approach, where agents not only have a WP function, but also have an heterogeneous perception of urban space which can be described by a spatial segmentation function, with parameters that are estimated jointly with the WP parameters, following a latent class modelling approach (Kamakura & Russell, 1989).

This model presents a novel simultaneous approach to spatial segmentation and location choice, which we affirm is ahead of previous (two-step) zone-based segmentation methodologies, thanks to zones defined by a classification function based on parameters that are estimated in order to maximize the likelihood of reproducing the phenomenon. This method is also behavior-based, in a model formulation that follows agent segmentation process and is consistent with microeconomic theory. We apply the model to household location data for Santiago de Chile and compare results with those obtained when using a model with no segmentation, and other two models where segmentation is done in a first step (exogenous zones and cluster-based zones). Model comparison is done using a validation subsample.

The chapter is structured in five sections. After this introduction, section two discusses the issue of spatial heterogeneity in general terms; section three details the proposed modelling framework; section four explains the mathematical formulation of the bid-auction localization model. Section five presents the proposed model, conceptually and mathematically. Section six presents the data and implementation of the proposed model.

Section seven presents the results and the comparison with other approaches. Finally, conclusions are presented.

## 3.2. Sources and methods for spatial heterogeneity

In location choice models, heterogeneity is dealt with when the modeler accounts for different behavior for different types of agents, understanding that people have a complex nature and that a general rule or set of drivers is not enough. The differences in behavior under the same conditions, for different types of people, is also observed across space. Spatial heterogeneity means that model parameters are not stationary across space (Anselin, 1988) which means that, for different reasons, the same individual facing the same conditions will behave different in different parts of the territory.

### 3.2.1. Sources of spatial heterogeneity

Fotheringham et al. (2002) identify three reasons for non-stationary parameters in space. Two of them belong to modelling shortcomings: one related to the possibility of data samples being different across space and the other to the existence of non-observed variables that are correlated with spatial variations. The third one relates to the spatial phenomena itself, related to contextual effects that affect the valuation of location attributes by individuals. The fact that the same person could react in different ways to the same stimuli, depending on the location of the city where she or he stands, provides evidence for the existence of some underlying qualitative aspects of places that interact with observed attributes. Neighborhood effects (Becker & Murphy, 2009; Durlauf, 2003; Sampson, Morenoff, &

Gannon-Rowley, 2002) may explain the synergies among certain urban attributes that can act as a multiplier of their effects on behavior. As traditional model parameters represent the marginal effect of one additional unit of an attribute in the level of the explained variable, it is natural to think that this effect is not constant, as it can be sensitive (or relative) to the levels of other urban attributes in the same location.

Different levels of urban attributes can represent states of saturation or scarceness of that attribute. In places where the attribute is abundant, it is possible that the valuation of that attribute is lower than in places where the attribute is scarce, meaning the preference parameter for that attribute should vary across space. For example, the same additional square meter of green area can have a greater effect (larger parameter value) if the location in question has a high built density (and therefore scarcer green areas) than if it's a low density location. In a complex system as a city, attributes interact in complex ways, so it is reasonable to assume that the effect of urban attributes on behavior is not isolated from the magnitude of other attributes (Abbott, 1997).

One of the first systematic works exploring how neighborhoods are recognized and affect behavior was the research by Lynch (1960), which surveyed inhabitants of three cities in the US to obtain maps of how they perceived their neighborhoods. Lynch identified five elements (zones, barriers, paths, milestones, nodes) that people can recognize as characteristic of a city and that are related to the reading that people make to orient themselves and be able to "navigate" through the city; a concept that was later investigated as mental maps applied to space (Gould & White, 1974). Other research (Nasar, 1990;

Salesses, Schechtner, & Hidalgo, 2013) has identified how people not only identify sectors, but also apply different valuations based on their urban attributes. This is consistent with general theories from psychology, for example, Gestalt (Wertheimer & Riezler, 1944), fragmentation or chunking (Gobet et al., 2001; Miller, 1956), and mental models (Johnson-Laird, 2010), which indicates that people tend to group or add information in elements to simplify the abundant information of the context, and be able to handle it efficiently.

### 3.2.2. Methods for identifying spatial heterogeneity

Spatial heterogeneity is a special case of heterogeneity in general, for which we can find an early example in Quandt (1958), who used different functions in a linear regression for different subsets of observations.

There are different technics to structure the variation of parameters across space. The simplest one, that can be usually seen in hedonic price (Rosen, 1974) and location choice models, is to use an exogenous zonification (administrative or functional mainly), where each zone has a different set of parameters corresponding to the observations in the zone. Exogenous zonification, however, presents the shortcoming of the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984), which recognizes that zone-based spatial analysis can have different outcomes depending on the zonification used. This means that, since exogenous segmentation can be arbitrary, so can be the results.

Some techniques address this issue by estimating location-specific parameters, running a regression for each zone or area only using observations within a distance (Moving windows regression, MWR) (Chica-Olmo, 1995; Dubin, 1992) or using decreasing weights for observations depending on distance to location (Geographically weighted regression, GWR) (Fotheringham et al., 2002; Páez et al., 2008). These methods have the advantage of not having to rely on arbitrary zones (but the size of the moving window and the decreasing weights function can be arbitrary).

### 3.2.3. Housing submarkets as a form of spatial heterogeneity

In order to define zones, or submarkets, that really group similar preferences, a simultaneous estimation method (one step) has to be used. The proposed simultaneous estimation is based on defining two sets of parameters: submarket-specific preference parameters, and submarket classification function parameters. Both sets of parameters are estimated jointly to better capture the phenomena and reduce bias (Ben-Akiva et al., 2002).

We propose that this joint estimation can be achieved for location choice models by using Latent Class Models (LCM) (Kamakura & Russell, 1989). These models estimate the probability of individuals belonging to a certain class of decision maker as a function of her characteristics, while simultaneously estimating the preference parameters for each of the classes considered in the model.

LCM have been widely used to model heterogeneity in preferences for location choice across decision makers (Cox & Hurtubia, 2019b; Ettema, 2010; Liao et al., 2014; Lu et al., 2014; Olaru et al., 2011; Walker & Li, 2007). The cited literature uses latent classes to identify classes of households and real estate developers or, in general terms, agents that search for locations, based in a traditional choice framework as described by McFadden (1978). LCM allows estimating a different set of preference parameters for each class of agent. The probability of choosing a location is a total probability that considers the probabilities of being part of each class, and the probabilities of choosing that location conditional on belonging to a particular class.

In the authors' knowledge, LCM have only been applied in location choice models under a "traditional" choice approach, in which the classes segment households (or other agents such as firms) and give a different set of parameters to each class. However, the framework has not been applied in the context of a bid auction approach which, although mathematically analogous to the choice approach in its formulation, has a totally different interpretation and allows for the introduction of endogenous heterogeneity in preferences across space. In this matter, the authors have found only two previous examples of spatial segmentation using LCM (Oliva, Galilea, & Hurtubia, 2018; Sarrias, 2019) but not directly applied to location choice. Interestingly, the bid-auction formulation, which enables applying LCM to spatial data, imposes a structure where agents have to be exogenously classified in order to compute a WP for each of them. This opens the question of whether there is a trade-off in this regard, or if there are methods allowing to classify locations and agents simultaneously without exogenous definitions, which should be addressed in further research.

### 3.3. Problem and proposed model: simultaneous estimation

We argue that in order to define zones, or submarkets, that really group similar preferences, a simultaneous estimation method (one step) has to be used. The proposed simultaneous estimation is based on defining two sets of parameters: submarket-specific preference parameters, and submarket classification function parameters. Both sets of parameters are estimated jointly to better capture the phenomena and reduce bias (Ben-Akiva et al., 2002).

We propose that this joint estimation can be achieved for location choice models by using Latent Class Models (LCM) (Kamakura & Russell, 1989). These models estimate the probability of individuals belonging to a certain class of decision maker as a function of her characteristics, while simultaneously estimating the preference parameters for each of the classes considered in the model.

LCM have been widely used to model heterogeneity in preferences for location choice across decision makers (Cox & Hurtubia, 2019b; Ettema, 2010; Liao et al., 2014; Lu et al., 2014; Olaru et al., 2011; Walker & Li, 2007). The cited literature uses latent classes to identify classes of households and real estate developers or, in general terms, agents that search for locations, based in a traditional choice framework as described by McFadden (1978). LCM allows estimating a different set of preference parameters for each class of agent. The probability of choosing a location is a total probability that considers the probabilities of being part of each class, and the probabilities of choosing that location conditional on belonging to a particular class.

In the authors' knowledge, LCM have only been applied in location choice models under a "traditional" choice approach, in which the classes segment households and give a different set of parameters to each class. However, the framework has not been applied in the context of a bid auction approach which, although mathematically analogous to the choice approach in its formulation, has a totally different interpretation and allows for the introduction of endogenous heterogeneity in preferences across space. In this matter, the authors have found only one reference of spatial segmentation using LCM (Sarrias, 2019), who evaluated changes in subjective evaluations of well-being, but not directly applied to location choice and with far lesser detail in spatial attributes used for segmentation.

### 3.4. Bid-auction approach in location choice models

Location choice models are based on the assumption of agents facing a set of location alternatives. Each alternative reports a utility, which depends on the alternative attributes, agent characteristics, and a set of preference parameters. Building on McFadden's (1978) choice model, and Alonso's (1964) work, Ellickson (1981) proposed the bid auction approach which is appropriated for location decisions in a households and real estate market interaction.

The bid auction model can be derived from a utility maximization problem, in which an agent $h$ chooses a location $i$ from a set $\Omega$ of different locations in the city, trading off with consumption of other goods $(x)$. Besides consumed goods, the agent's utility depends on a vector $\beta_h$ of preference parameters and a vector $Z_i$ of attributes of each location $i$. A budget constraint is added, in which rent $r_i$ for location plus expenditure in other goods (priced at $p$) have to be equal or less than the agents´ available income $I_h$.

$$\max_{x,i\in\Omega} U_{hi}(x, Z_i, \beta_h) \tag{3.1}$$

$$s.a. \quad px + r_i \leq I_h$$

Assuming equality to clear $x$ from the budget constraint and replacing it in $U$, we obtain an indirect utility function ($V$) and the utility maximization problem simplifies to choosing the location that maximizes $V$:

$$\max_{i\in\Omega} V_{hi}(I_h - r_i, Z_i, p, \beta_h) \tag{3.2}$$

Considering a fixed referential maximum utility level $\overline{U_h}$ (expected by the agent), we can clear the rent $r_i$ which, if assumed endogenous, represents the willingness to pay (WP) of the agent for that location, in order to reach the reference utility (Jara-Díaz & Martínez, 1999). Endogenous rent then becomes the willingness to pay agent h for location $i$ ($WP_{hi}$):

$$WP_{hi} = I_h - V_{hi}(\overline{U_h}, Z_i, p, \beta_h) \tag{3.3}$$

If the utility function of (3.1) has a (quasi) linear form, the willingness to pay function can be simplified and written in terms of two components. One component is specific to the agent, related to the income level and expected maximum utility, and other is related to the preferences the agent has for attribute location (Martinez, 2000):

$$WP_{hi} = b_h + f(Z_i, \beta_h) \tag{3.4}$$

Assuming an i.i.d Gumbel distributed error term associated to the $WP_{hi}$ (accounting for unobserved attributes) the probability that agent $h$ is the highest bidder for location $i$, and therefore locates there, is defined by a logit function where $\mu$ is non-identifiable a scale parameter (McFadden, 1973):

$$P(h|i) = \frac{\exp(\mu WP_{hi}(b_h, Z_i, \beta_h))}{\sum_{g \in H} \exp(\mu WP_{gi}(b_h, Z_i, \beta_h))} \tag{3.5}$$

From a sample of located agents (segmented by type) and the attributes of their location, and using maximum likelihood estimation, this model can identify, for each type of agent, the marginal WP for each attribute considered in the WP function.

The bid auction approach has been used in several Transport and Land Use Interaction (LUTI) models such as MUSSA (F. Martínez, 1996), ILUTE (Salvini & Miller, 2005) and IRPUD (Wegener, 2011). Several research papers on the literature about location choice also use this approach (Chattopadhyay, 1998; Gross, Sirmans, & Benjamin, 1990; Hurtubia & Bierlaire, 2014; Hurtubia, Martinez, & Bierlaire, 2019; Muto, 2006)).

### 3.5. Proposed latent spatial-segmentation model

The proposed model applies a latent class model framework (LCM) to a bid-auction location choice model. Mathematically, applying LCM to bid-auction framework is relatively similar to doing so for a choice framework, but the interpretation and application differs in substantial aspects.

In a bid-auction model with latent classes, the class membership function applies to the location, understood as the agent that receives the bids (e.g the owner of the property or the land). Therefore, the class-specific preference parameters can be interpreted as the location-seeking agents (households and firms) having a different valuation of urban attributes conditional to the class of the location.

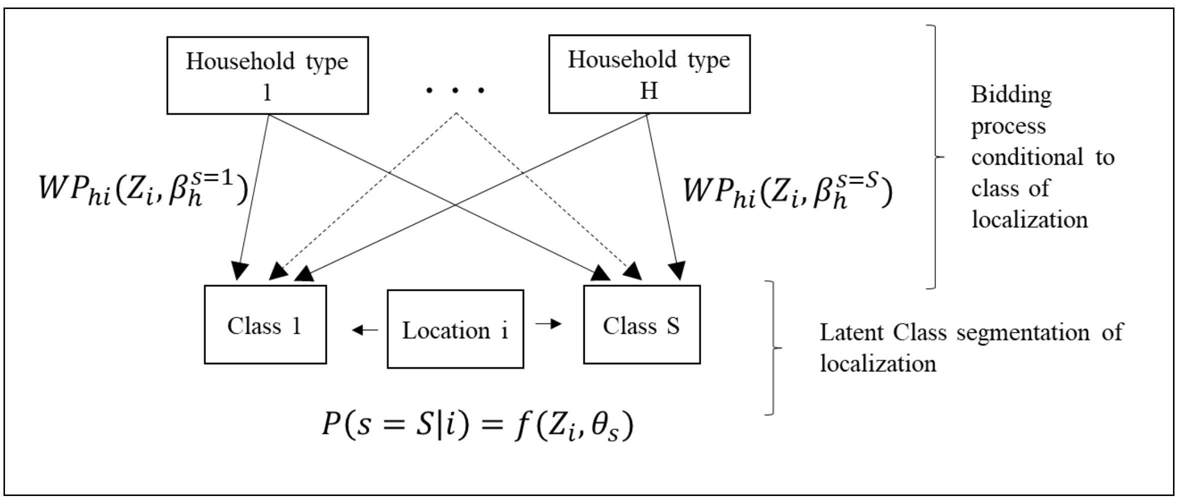


Figure 3-1: Latent classes applied in a bid-auction framework.

In simple terms, an agent will valuate differently an amenity depending on the class of the location. If the distribution of the magnitudes of the attributes is somehow continuous in the city (which indeed happens, due to spatial dependence; see Anselin, 1988) classes of locations can be related to neighborhoods.

Departing from the bid-auction model presented above, we modify equation (3.5) so the probability $P(h|i, s)$ is conditional to each class $s$ of locations or submarkets:

$$P(h|i,s) = \frac{\exp(\mu_s WP_{hi}^s(b_h, Z_i, \beta_h^s))}{\sum_{g \in H} \exp(\mu_S WP_{gi}^s(b_h, Z_i, \beta_g^s))} \tag{3.6}$$

Each agent will have a different $WP$ depending on the submarket or class of the location where they are bidding, because the $WP$ is function of a set of preferences parameters $\beta_h^s$ which are conditional to the class of location.

Simultaneously, each location will have a probability of belonging to a class which, according to the standard formulation of LCMs, is a logit probability based on a class membership function $W_{is}$ for which we assume an error term $\epsilon$ i.i.d Gumbel and a non-identifiable scale parameter $\gamma$:

$$P(s|i) = \frac{exp(\gamma W_{is}(Z_i, \theta_s))}{\sum_{g \in S} exp(\gamma W_{ig}(Z_i, \theta_g)} \tag{3.7}$$

As we are segmenting space into zones or neighborhoods, the class membership function $W_{is}$ depends on location attributes $Z_i$, instead of agents characteristics, unlike previous applications of LCM to location choice (see for example, Walker and Li, 2007, and Hoshino, 2011). A vector of parameters $\theta_s$ is estimated, which represent the marginal contribution of each location attribute to the probability of belonging to a class.

Given the probability that agent $h$ gets the location $i$, conditional to the class of the location, for all agents and locations (equation 3.6), and also the probability that location $i$ belongs to class $s$, for all locations and classes (equation 3.7), the probability that agent h gets location $i$, unconditional to class membership is:

$$P(h|i) = \sum_{s} P(h|i, s) \cdot P(s|i) \tag{3.8}$$

Using equation (3.8), maximum likelihood estimation can be used to identify parameters $\beta_s$ and $\theta_s$ from observations of agents' location decisions, without requiring any information regarding submarket structure. This approach avoids an ex-ante definition of the membership of locations to submarkets and, instead, infers how agents' perceive locations as part of a submarket, and accordingly variate their preferences.

In this specific model implementation, we use an estimation method proposed by Lerman & Kern (1983), where the maximum likelihood is not only targeting to reproduce the actual localizations of agents, but also minimizing the difference between winning WP and observed price (in this case monthly rent) paid by the located agent. This method is similar to other methods to discrete-continuous choices (see for example Bhat, Astroza, Bhat, & Nagel, 2016), although strictly formulated for location choice. We use this method because it allows to identify the scale parameter for each class ($\mu_s$) and, therefore, enables the identification of parameters for all type of agents (otherwise parameters for one type of agent have to be fixed and the other parameters estimated relative to them). This scaling of parameters renders estimated values of WP with the same magnitude as observed prices.

As we are including latent classes, first we have to specify a Lerman & Kern likelihood function specific to each class, and then specify a final likelihood function considering the probabilities of each class.

Equation (3.9) shows the likelihood function $\mathcal{L}$ conditional to each class $s$, where $R_i$is the observed rent in the location $i$, $WP_{h'i}^{s}$ is the highest bid modelled for location $i$ conditional to that location being part of class $S$, and $\omega^s$ is the scale parameter of the class-specific logit function. This function is calculated for each location $i$.

$$\mathcal{L}_i^s = \omega^s \cdot \exp\left(-\omega^s \cdot \left(R_i - WP_{h'i}^s\right)\right) \cdot \exp\left(-exp\left(-\omega^s \cdot \left(R_i - \left(\frac{1}{\omega^s}\right) \cdot \ln \sum_{\substack{g \in H \\ g \neq h'}} exp\left(\omega^s \cdot WP_{gi}^s\right)\right)\right)\right) \tag{3.9}$$

The final likelihood function to be maximized, not conditional to class, is:

$$\mathcal{L} = \prod_{i \in \Omega} \left( \sum_{s \in S} [\mathcal{L}_i^s \cdot P(s|i)] \right) \tag{3.10}$$

The likelihood function is basically maximizing the joint probability that, for each location, the winning agent has the highest bid, and that the highest bid is equal to the observed rent.

### 3.6. Application to Santiago case study

The proposed model was tested with a database of households from the 2012 Origin Destination Survey for Santiago (SECTRA, 2015), each with socio economic variables and exact georeference. Location attributes were calculated for each location using a

Geographical Information System (GIS). With this information, besides the proposed model, a base model (with no spatial heterogeneity) and two other alternative approaches (cluster-based zones and administrative zones) were estimated for comparison purposes. Direct log-likelihood was measured with a validation sample, for the proposed model and the alternative approaches, in order to compare predictive power.

### 3.6.1. Urban structure of Santiago

The spatial structure of Santiago depends on its particular history and national hierarchy. With 6,123,000 inhabitants (INE, 2018), it is the main city of Chile in terms of population, economic activity and administrative power. Santiago has evolved from a traditional compact city to a fragmented and globalized city (Borsdorf, 2003), where both densification and expansion development patterns have been observed in recent decades (Cox & Hurtubia, 2016; Vergara Vidal, 2017)

In terms of urban structure, the city of Santiago answers to the latest stage of the model described by Borsdorf (2003), where there is a main Central Business District (CBD) based on the historical center, from which departs a wedge of high income residential areas (towards the north-east in the case of Santiago), with an spine in its central axis of more modern commercial and office areas (*Providencia* and *Avenida Las Condes*). In the case of Santiago, this commercial spine is becoming more relevant in later years (Suazo, 2017). Borsdorf's model describes the location of high income households as very differentiated from low income households, which locate in broad areas beside industrial corridors. In the case of Santiago, this areas correspond to north-west, west and south peripheral areas of the city. Also described in this model, and observed by other authors (de Mattos, 1999; Sabatini & Salcedo, 2010) is the later fragmentation of this sectorial model towards a more network-based urbanization, with growing system of highways that connect so-called “globalization artifacts”, such as malls, airport, gated communities and industrial parks, sprawling on the fringes of the city, many times inserted in but not actually connected with low income areas. Although this relatively new “leap-frog” urbanization hasn´t followed the sectorial residential segregation seen in past decades, the segregation is still being reproduced but in a lower scale (Sabatini, 2015).

If the proposed model adequately reproduces how people perceive city areas when choosing location, this depiction of the city coming from urban geography should be somehow observed when mapping the resulting spatial segmentation of the model.

### 3.6.2. Data

Household data was extracted from the Santiago 2012 Origin-Destination Survey (SECTRA, 2015), accounting for 18,624 observations, from which 14,172 were used for estimation (exclusion was based on lack of some key attributes for some observations). The survey considers 790 zones as its basic spatial analysis unit, we use these zones to compute some of the attributes describing each location (such as average income and accessibility measures). Households were segmented into three categories according to the educational level (EL) of the head of household. Low EL correspond to 0 to 11 years of education[7], middle EL to 12 to 15 years of education and high EL to 16 or more years. The map in Figure 3-2 presents the spatial distribution of households, as well as the general structure of the city while Table 3-1 characterizes these types of households. Table 3-2 describes the attributes used to describe each location and their sources.

One limitation of the database is the lack of information on built surface for the dwelling of each observed household´s location. As a proxy, we use the average built surface in the zone of the dwelling.

| COD | Level | Years of Education | Number of Households | % |
|---|---|---|---|---|
| Lo-EL | Low Educational Level | 0 to 11 | 6620 | 37.1% |
| Mid-EL | Middle Educational Level | 12 to 15 | 7774 | 43.6% |
| Hi-EL | High Educational Level | 16 + | 3436 | 19.3% |
| | | TOTAL | 17830 | |

Table 3-1: Segmentation of households according to educational level (EL).

[7] In Chile, having 12 years of education implies finishing the compulsory high school degree. However, a large part the population does not achieve this educational level.

| Variable | Unit | Description | Source | Mean | Min | Max |
|---|---|---|---|---|---|---|
| Monthly Rent | Million CLP | Monthly rent paid by household in million chilean pesos (CLP) | Origin Destination Survey (SECTRA, 2012) | 0.19 | 0.01 | 5 |
| Accessibility to Industry (transit) | - | Gravitational with negative exponential function weighted by industry surface in destination zone. | Own calculation based on Internal Revenue Service (2014) and SECTRA (2015) | 1807 | 33 | 4536 |
| Accessibility to commerce (transit) | - | Gravitational with negative exponential function weighted by commerce surface in destination zone. | Own calculation based on Internal Revenue Service (2014) and SECTRA (2015) | 2262 | 46 | 6096 |
| Accessibility to Industry (car) | - | Gravitational with negative exponential function weighted by industry surface in destination zone. | Own calculation based on Internal Revenue Service (2014) and SECTRA (2015) | 5082 | 1031 | 6934 |
| Accessibility to commerce (car) | - | Gravitational with negative exponential function weighted by commerce surface in destination zone. | Own calculation based on Internal Revenue Service (2014) and SECTRA (2015) | 5894 | 1048 | 8583 |
| Distance to closest Subway Station | km | Euclidian distance from household to closest subway station as of 2012 | Own calculation in QGIS | 4.74 | 0.03 | 49.84 |
| Distance to closest highway exit | km | Euclidian distance from household to closest highway exit as of 2012 | Own calculation in QGIS | 2.04 | 0.03 | 13.27 |
| Zonal average income | Million CLP | Average income of the households in the OD Zone. | Origin Destination Survey (SECTRA, 2012) | 0.66 | 0.14 | 4.95 |
| Built surface | m2 | Average built surface of residential units in the block of the household | Internal Revenue Service (2014) | 31 | 49.59 | 207.3 |
| Built density in zone | coef | Total built surface divided by zone area. | Internal Revenue Service (2014) | 0.38 | 0 | 4.59 |

Table 3-2: Variables evaluated for the model. *CLP: Chilean Peso.

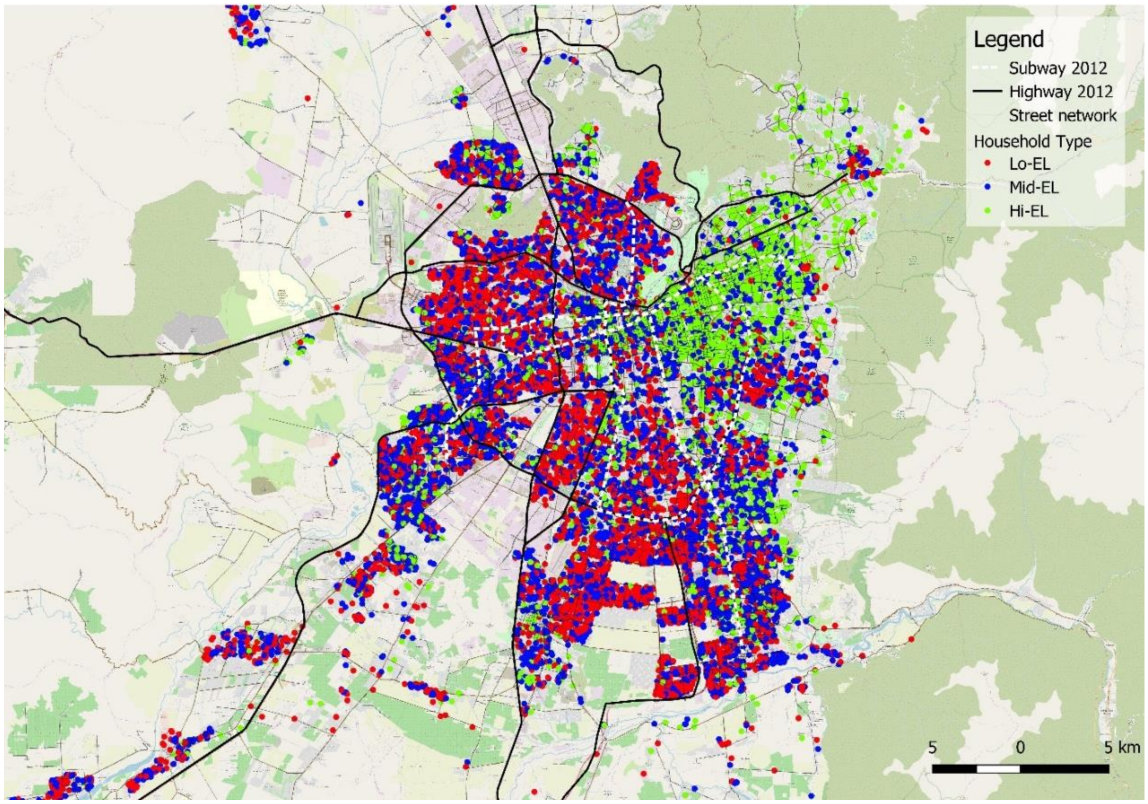



Figure 3-2: Location of survey households according to their educational level.

Accessibility for each zone $i$ in mode $m$ was calculated with a gravitational measure (Hansen, 1959), following:

$$Acc_i^m = \sum_j O_j \cdot exp(-\beta \cdot tv_{ij}^m) \quad (3.11)$$

where $O_j$ is the amount of opportunities (e.g. built surface of commerce, or industry) in each of the possible destination zones, $tv_{ij}^m$ is the travel time in mode $m$ (car or transit) between pair of zones $i$ and $j$, and $\beta$ is an impedance parameter (we used $\beta = 0.05$, which gave the higher significance and log-likelihood in the estimation stage, while also reproducing observed distributions of trip-lengths). For accessibility to commerce and industry, built surface of each land use in all of the 790 OD Survey zones was extracted from the Internal Revenue Service registry for 2014. Travel times by car and transit between each OD Survey zone was obtained from the strategic transport model for Santiago (ESTRAUS, SECTRA, 2016) which is calibrated with the same OD Survey travel data.

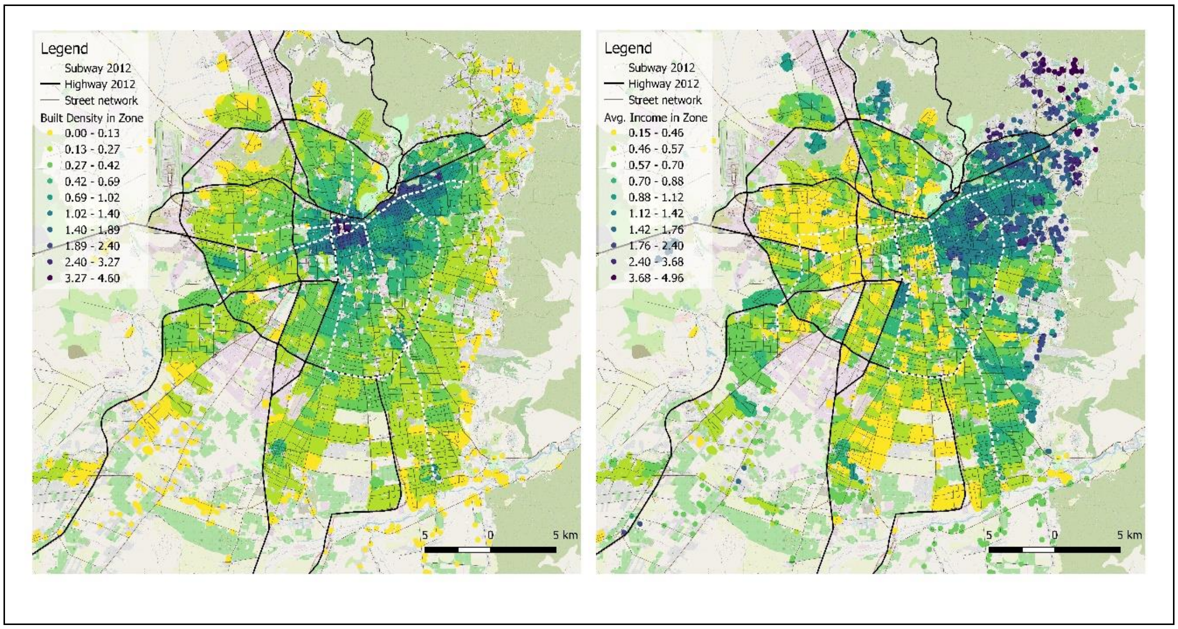



Figure 3-3: Maps showing attributes used for the spatial segmentation function (latent class model) and the clusterization. The attributes are built density in zone (left) and average income in zone (right).

### 3.6.3. Estimation results

Several specifications for the Willingnes to Pay ($WP_{hi}$) and Class Membership ($W_{is}$) functions were explored. The results for a model with two classes are reported; models with three and more classes could not be estimated as parameters could not be identified. We interpret this as a consequence of the structure of the data, confirming that Santiago is strongly polarized into two well-defined classes of locations. Also a benchmark model was estimated, parallel to the estimation of the proposed model. The benchmark (base) model uses the same variables, household types and approach (bid-auction and Lerman & Kern estimation) as the proposed model, but has no segmentation of locations. Table 3-3 shows estimation results for both models, where coefficients can be interpreted as the marginal willingness to pay for each attribute. The t-test for each estimate is presented between parentheses. Estimation was made using the econometric software Biogeme (Bierlaire, 2020).

For this implementation of the proposed model, the zonal average income and zonal built density (see Figure 3-3) were used to classify the city locations in two different types (classes) of locations. In the base model these attributes were included directly in the WP function. From the sign of the class membership parameters, the probability of membership to Class One improves in zones with higher income and density. Therefore Class One can be labelled as wealthy and dense locations. By opposition, class two corresponds to sparse and low income locations.

Parameters $\mu_1$ and $\mu_2$ are the scale parameters for the choice model of each class, identifiable thanks to the Lerman & Kern (1983) estimation method we used. For comparison purposes between models, attributes with non-significant parameters were not excluded.

| | | BASE MODEL (NO CLASSES) | PROPOSED MODEL (2 CLASSES) | |
|---|---|---|---|---|
| Observations | | 17830 | 17830 | |
| Null model log-likelihood | | -258759 | -258759 | |
| Final log-likelihood | | -77213 | -72709 | |
| Attribute | Household Type | Coefficient (t-test) | | |
| | | | Class 1 (Dense/Wealthy) | Class 2 (Sparse/Low Income) |
| Accessibility to Commerce by transit | Low-EL | -0.0000965 (-0.36)* | -0.00438 (-1.51)* | 0.000613 (3.65) |
| | Mid-EL | 0.00351 (14.57) | 0.00329 (3.16) | 0.00255 (13.54) |
| | Hi-EL | 0.00741 (18.74) | 0.0019 (2.05) | 0.0054 (8.45) |
| Accessibility to Industry by transit | Low-EL | 0.000398 (1.06)* | 0.00801 (1.94)* | -0.000453 (-1.89)* |
| | Mid-EL | -0.00429 (-12.49) | -0.0064 (-4.33) | -0.00263 (-9.79) |
| | Hi-EL | -0.0112 (-19.67) | -0.00542 (-3.87) | -0.0054 (-5.81) |
| Accessibility to Commerce by car | Low-EL | 0.00109 (3.41) | 0.00855 (2.08) | -0.000142 (-0.68)* |
| | Mid-EL | -0.00203 (-7.05) | -0.000289 (-0.22)* | -0.00257 (-11.18) |
| | Hi-EL | 0.005 (10.37) | 0.00983 (7.68) | -0.00607 (-6.19) |
| Accessibility to Industry by car | Low-EL | -0.000958 (-2.42) | -0.0144 (-2.89) | 0.000561 (2.18) |
| | Mid-EL | 0.00333 (9.02) | 0.00158 (1.02)* | 0.00355 (12.57) |
| | Hi-EL | -0.00186 (-2.86) | -0.00929 (-5.77) | 0.00766 (6.79) |
| Distance to nearest subway station | Low-EL | 0.0166 (0.91)* | 0.0215 (0.07)* | 0.0374 (3.37) |
| | Mid-EL | 0.0994 (5.61) | 0.167 (1.33)* | 0.0691 (5.8) |
| | Hi-EL | 0.299 (8.23) | 0.868 (8.51) | 0.137 (3.92) |
| Distance to nearest highway exit | Low-EL | -0.0385 (-0.74)* | -1.6 (-2.13) | 0.0372 (1.19)* |
| | Mid-EL | 0.0000959 (0)* | -0.188 (-0.82)* | 0.0528 (1.56)* |
| | Hi-EL | -0.00546 (-0.06)* | -0.459 (-2.32) | 0.033 (0.29)* |
| Average Built surface in zone | Low-EL | 0.00696 (1.26)* | 0.127 (3.2) | 0.0144 (4.48) |
| | Mid-EL | 0.0597 (13.23) | 0.128 (9.09) | 0.0458 (16.02) |
| | Hi-EL | 0.146 (27.74) | 0.238 (22.88) | 0.0627 (8.18) |
| Average Income in Zone | Low-EL | -0.0482 (-0.09)* | | |
| | Mid-EL | 7.68 (30.99) | | |
| | Hi-EL | 7.61 (33.93) | | |
| Built density in zone | Low-EL | -2.77 (-6.62) | | |
| | Mid-EL | 0.657 (3.07) | | |
| | Hi-EL | 0.554 (2.4) | | |
| Household type constant | Low-EL | 5.92 (7.06) | 5.36 (10.43) | |
| | Mid-EL | -4.65 (-6.04) | 2.13 (3.97) | |
| | Hi-EL | -28.7 (-22.04) | -8.66 (-5.92) | |
| Class Membership Variables | | | Class 1 | Class 2 |
| Intercept | | | -9.22 (-31.85) | |
| Average Income in Zone | | | 10.8 (26.64) | |
| Built density in zone | | | 1.86 (5.87) | |
| $\mu_1$ | | 0.164 (169.84) | 0.0907 (65.96) | |
| $\mu_2$ | | | 0.28 (116.92) | |

Table 3-3: Estimation parameters.

A first thing to notice is how parameters that appear as non-significant in the base model can become significant when using the proposed approach. For example, the parameters for distance to the nearest highway are all non-significant in the base model but become significant for Low-EL and High-EL in the dense/wealthy class. In this case, the negative sign can be interpreted as those types of households having a higher WP for locations that are close to highways.

The model fit (rho-squared) is significantly higher for the latent class model. When applying a likelihood ratio test between models (LR= 9008) it is confirmed that the latent class model provides more information about perception even when being penalized for using more parameters (Ben-Akiva & Lerman, 1985).

Built surface (as said before, we use average built surface in the zone of the dwelling as a proxy for this variable) is, as expected, a relevant attribute. From the base model, we observe that higher EL households are willing to pay more than other households for additional built space. But from the latent class model we can also see that these households are willing pay as much as almost four times more for an additional square meter, if the location is in a wealthier and denser zone. We acknowledge that not using the exact built surface of the dwelling could induce endogeneity in the estimation and other techniques, such as instrumental variables, could be used to overcome this issue (Guevara & Ben-Akiva, 2006).

Accessibility to commerce, by transit and car, is always positive in the base model but, when including latent classes, we see some differences. Accessibility by car increases willingness to pay in wealthy areas, but decreases it in low income/density areas for all types of households (except mid-EL households which are indifferent to this attribute in dense and wealthy areas). This can be interpreted as households assigning a positive value to the type of commerce usually observed in wealthy areas, with the opposite occurring in low income areas. This result is consistent with our expectations, considering the fact that high income municipalities are capable to minimize or mitigate the negative externalities of commercial activities (congestion, garbage production, landscape impact of buildings), while lower income municipalities usually don't have enough resources to control this.

Accessibility to industry has a mixed effect, although mostly negative, for both car and public transport in the base model. In the model with spatial classes, accessibility to industry becomes a positive attribute for the sparse/low-income class and a negative one for the dense/wealthy class. This is a good example of how the proposed approach can capture preferences that diverge and that, otherwise, would lead to a misinterpretation of the drivers behind location behavior.

Distance to subway stations is always positive, which is counterintuitive but expected considering we are including other variables that account for accessibility through public transport. In the latent class model we see that being far from subway stations is more important in wealthier zones than in low income zones, and this difference is more critical for High EL households. These may be due to reasons similar to those exposed for the commerce case, considering the negative externalities metro stations can produce in their immediate surroundings

This differences in parameters between classes is consistent with our hypothesis that spatial heterogeneity in preferences can be captured by using latent classes as different zones in the city. The latent class model has a significantly better log-likelihood the benchmark model, which indicates that this new dimension of heterogeneity introduced helps to better reproduce the location choice phenomenon.

### 3.6.4. Spatial distribution of class membership of locations

Once the class membership parameters of the model are estimated, they can be used to evaluate the probability of each location of belonging to each spatial class all over the city. The map in Figure 3-4 shows the spatial distribution of these probabilities, indicating a clear segmentation of the city, consistent with the well-known socioeconomic segregation patterns of Santiago (Sabatini, 2003). Class 1 locations (blue), related to dense and high-income zones, are clearly correlated with what is known as the "high income wedge" of Santiago. However, there are several places with a high probability of belonging to class 1 outside of this wedge, where a combination of density and relatively high income is observed. Most of the city has a high probability of belonging to class 2 (yellow), which are mostly the extended peripheries, with low density and medium and low income.

Some isolated zones with high probability of belonging to class 1 are newer private developments, where medium high income households have located. These projects answer to the typology of gated communities, locating outside the high income area of the city and being mainly disconnected to their immediate lower income context (Borsdorf et al., 2007) and many times surrounded by hills to reinforce their “enclave” condition.

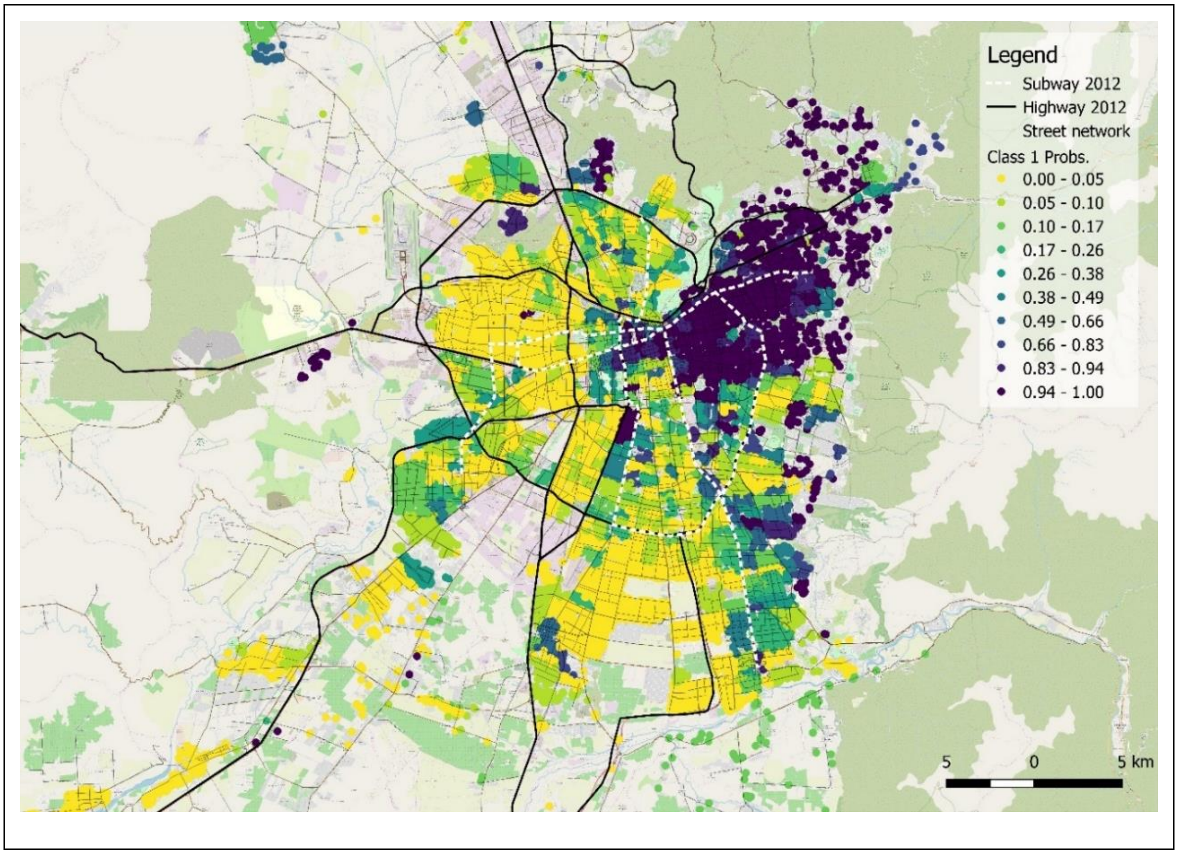



Figure 3-4: OD survey households´ locations and their probability of membership to a wealthy high density zone (class 1 in the model) according to the proposed model.

It can be seen that few locations are neither yellow nor blue. As it is shown in histogram in figure 3-5, most of the locations (55%) fall below the 1% probability of belonging to class 1 while 16% do so for class 2. This is evidence that perception or valuation of urban attributes in Santiago not only variates for different zones, but also that this differences have clear cuts, building strong perceptual urban limits.

As the location classification in this model is estimated simultaneously with the model for the location decision process of each household, this can be interpreted as households perceiving the city as a two clearly different sets of zones, where they will apply different valuations of urban attributes.

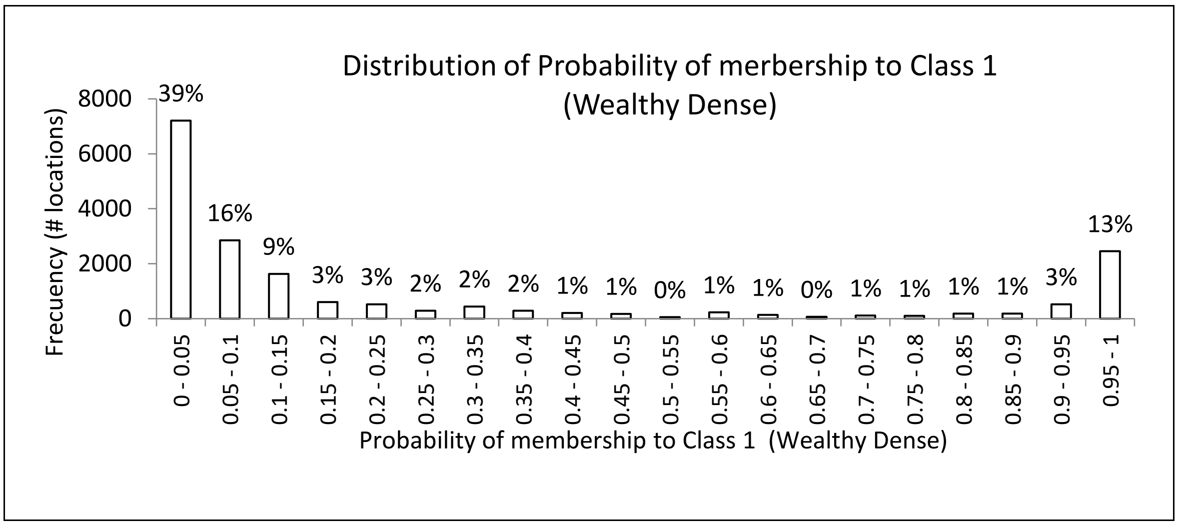



Figure 3-5: Histogram of the probability of membership to class 1 for every location (households' location in ODS 2012).

### 3.6.5. Comparison to alternative models: Exogenous zones and attribute-based clusters

There are different approaches to include spatial heterogeneity in a location choice model. In order to compare the effectiveness of the proposed method, we compare its results with those of two alternative approaches: a model with exogenous zones, and two models with zones based on clustering of location attributes. We used the same attributes as in the base and proposed model. The same two attributes that were used for the segmentation function in the latent class model were used for the clustering process in the cluster models.

For the model of exogenous zones, we use the seven macro-zones of the Origin Destination Survey that we have used in this work, and we estimated the location model with preference parameters specific to each zone. Map in figure 3-6 shows the households colored according to the zone where they belong.

Table 7.1 (in Annex) shows the estimated preference parameters of each household for each attribute in each zone, for the exogenous zone model. The principal result to notice is that the log-likelihood (-76.511) is lower than the one obtained with the proposed model while, as it can be expected, it´s better than the log-likelihood for the base model (with no zones). Parameters are mostly significant at 95%, and with significantly different values for each zone, indicating that these zones, although defined with no explicit market considerations, are still capable to define submarkets where preferences are different. Accessibilities by car where excluded as the model was not able to be estimated with all the attributes, which is probably due to the high amount of parameters involved.

As for the cluster-based models, they are expected to have better performance than the exogenous zones model, as locations are grouped following differences in some of their attributes. We used a k-means method (MacQueen, 1967) for the cluster-based models, where each possible location was assigned to one cluster (we estimated two models: one with seven clusters, in order to compare with exogenous zones model, and one with two clusters, similar to the number of classes in the proposed model). The assignation criteria was to group locations with similar level of two attributes: Zonal Average Income and Built Density (the same attributes used in the proposed model to generate the latent classes). Figure 3-7 shows the spatial distribution of the clusters. Tables 7.2 and 7.3 (in Annex) shows the estimation results for this model. The log-likelihood of these two models is higher than the one for the exogenous model and the base model (no zones), but still is not higher than the one for the proposed model. Most of the parameters are significant at a 95% level and show differences between zones, indicating they do represent different zones.

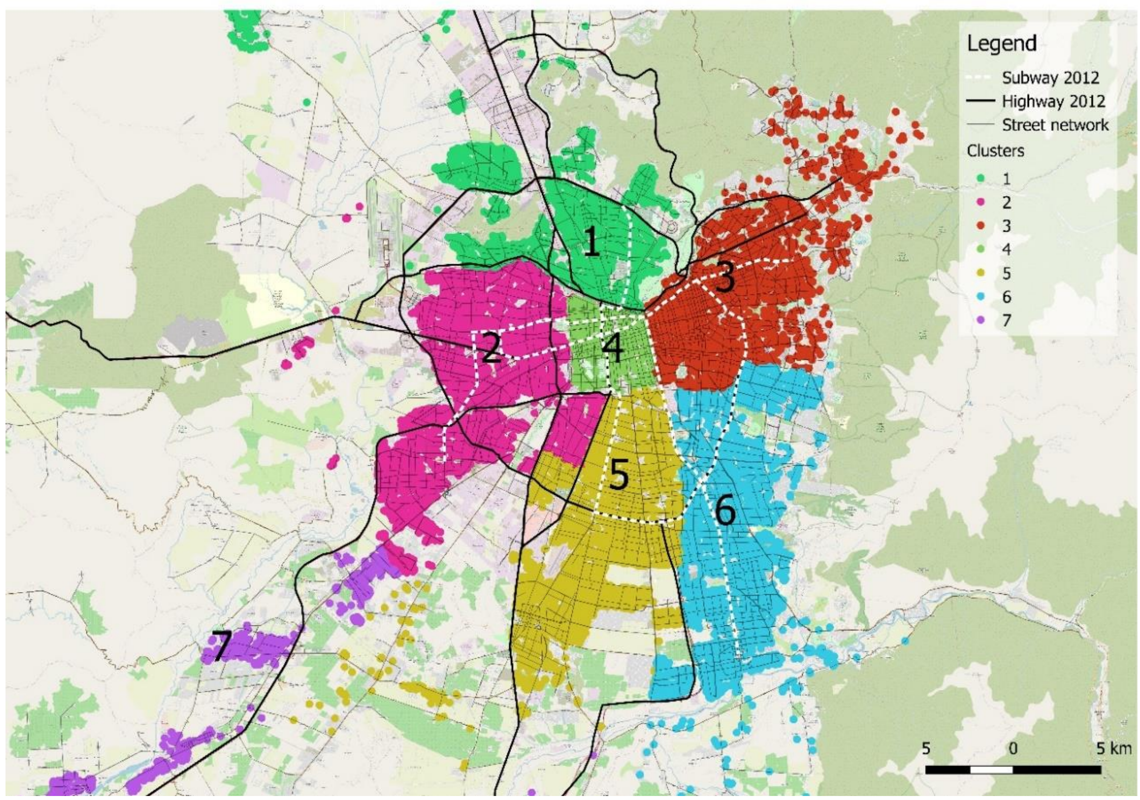



Figure 3-6: Survey households colored according to the survey zone they belong (zones used to estimate the exogenous zones model).

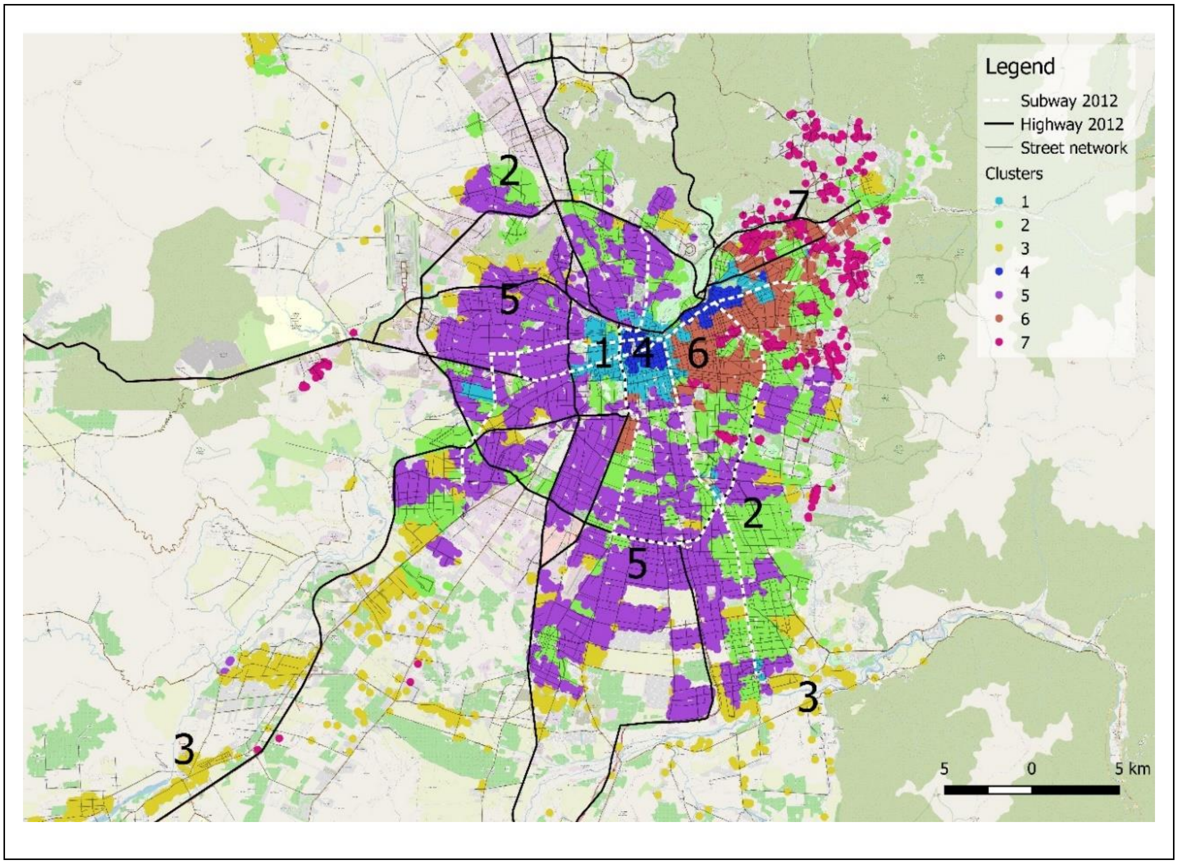



Figure 3-7: Survey households colored according to the cluster they belong (clusters used to estimate the seven clusters model).

To validate the proposed model against the other models, we reestimated all the models with a random sample of 90% of the locations, and then with the remaining 10% we calculated the probability that the observed winning household also had won the bid in the model, for each location in the validation sample.

Table 3-4 shows a summary of model-fit and information statistics for all the estimated models using the validation sample.

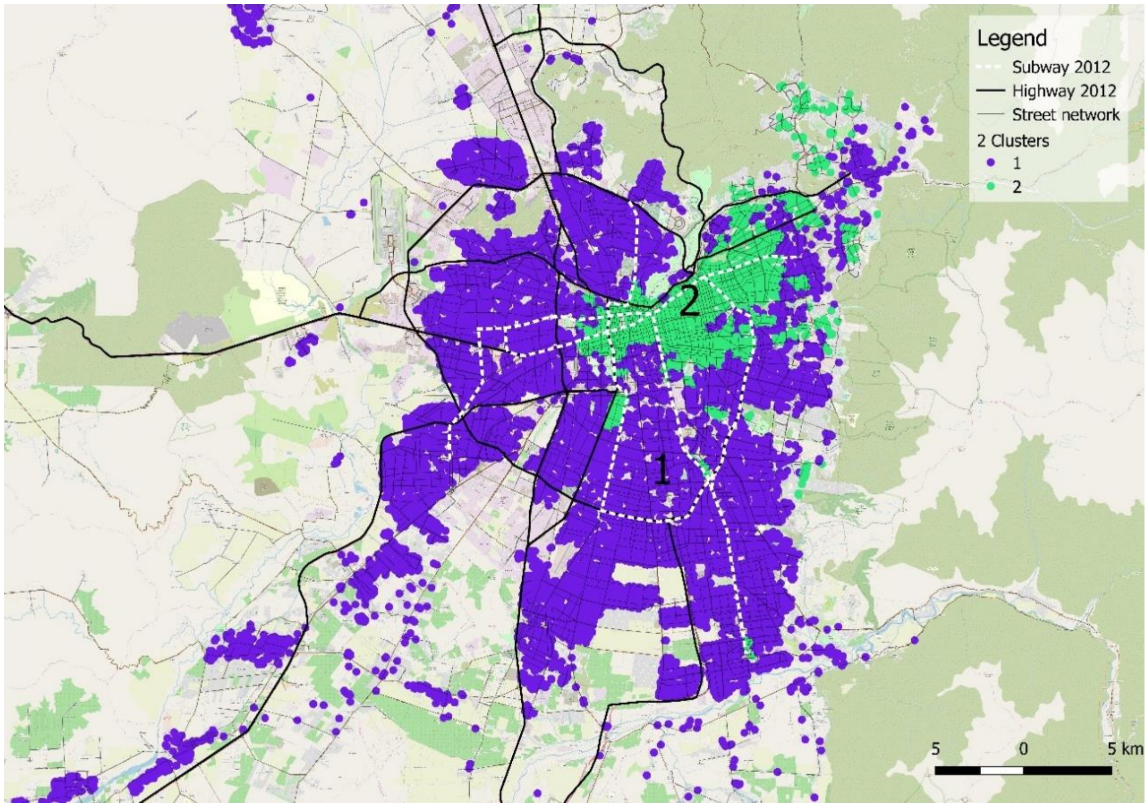



Figure 3-8: Survey households colored according to the cluster they belong (clusters used to estimate the two clusters model).

| Model | Log-Likelihood | Number of parameters | % significant parameters (95%) | AIC | BIC |
|---|---|---|---|---|---|
| No spatial heterogeneity | -7,608 | 31 | 74% | 15,278 | 15,448 |
| 7 Exogenous zones | -7,534 | 109 | 55% | 15,287 | 15,885 |
| 2 Cluster-based zones | -7,494 | 56 | 57% | 15,100 | 15,408 |
| 7 Cluster-based zones | -7,450 | 193 | 46% | 15,285 | 16,344 |
| Endogenous zones (proposed model) | -7,216 | 50 | 76% | 14,532 | 14,806 |

**Table** 3-4: Model log-likelihood comparison.

The proposed model shows better log-likelihood than the other models, and also better percentage of significant parameters. To account for the relation between the number of parameters and log-likelihood, we also calculate the Akaike Information Criterion (AIC) (Akaike, 1998) and Bayesian Information Criterion (BIC) (Schwarz, 1978) for each model, where the proposed model also has a better performance.

### 3.7. Conclusions and discussion

We propose a discrete choice model that allows to include spatial heterogeneity with probabilistic zones (fuzzy limits among zones) that are defined endogenously, following a one-step estimation of location preferences and zone segmentation parameters.

The main conclusion is that the proposed location model, with endogenous spatial heterogeneity (LCM applied to the bid auction approach), outperforms other common approaches in terms of direct log-likelihood when applied to data from a hold-out sample, indicating better forecasting abilities. Also, the proposed model is parsimonious, as it has better log-likelihood with fewer parameters than the other models. This is confirmed by the Aikaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) as is shown in table 4.

When we apply the segmentation function to the map of Santiago, two clear zones emerge, which are closely related to the income stratification of this city. The proposed method helps not only to identify zones, but also to observe how fuzzy or clear-cut are the boundaries among zones, which can relate to perceived barriers or to smooth transitions between zones.

This modelling approach provides a significant contribution by including spatial segmentation from a behavioral perspective. As the segmentation method is integrated in the microeconomic modelling of the decision process, segmentation parameters can be interpreted as the decision maker criteria of segmentation of the city that assist him in the location choice process. It can be interpreted as individuals dealing with qualitative (classes) and quantitative (WPs) aspects in their decision process.

Besides the extension to traditional location choice models in order to better forecast residential demand, the proposed model has other useful public policy applications as it can provide city planners with new insight about which spatial attributes are affecting location patterns. For example, if one class of locations can be identified as “desirable” (for instance, the attributes used in the classification function are related to compact or sustainable development), then the probability of belonging to that class, for each location, can be used as an urban index to measure the status that location with regards to the desired objective.

The proposed approach can be easily extended to hedonic price models, usually found in the real estate submarkets literature (Bourassa et al., 1999; Rosmera & Lizam, 2016). The approach can also be applied to other behavior taking place in the urban context that could be influenced by spatial attributes, such as mobility patterns and their relation to the built environment (see Oliva et al., 2018, for an early example of this).

**References** for all chapters are presented in a specific chapter after conclusions.

# 4. COMPACT DEVELOPMENT AND PREFERENCES FOR SOCIAL MIXING IN LOCATION CHOICES: RESULTS FROM REVEALED PREFERENCES IN SANTIAGO, CHILE[8]

Tomás Cox and Ricardo Hurtubia

## ABSTRACT

Even though densification and social mixing are declared objectives of many nowadays urban planning paradigms, their simultaneous implementation is usually questioned by different actors and is not frequent in practice. In a market economy, understanding potential demand for this class of development, from different types of households, is essential to define public policy oriented to achieve both compact development and social mixture. In order understand the preferences of households and potential demand, we design and implement a location choice model based on bid-rent theory, using census data and location attributes. The model uses latent spatial classes to endogenously segment locations according to their attributes, and identifies households' preferences specific to each spatial class. We specify the model to identify spatial classes related to compact development (CD);

[8] This chapter is under revision for publication as a paper at "Journal of Regional Science".

by doing this, we can measure how households preferences for urban attributes change between zones classified as CD or suburban.

We apply the model to Santiago de Chile, where social mixing in dense and well-located areas is being intensely discussed. We find strong differences in households' valuation of attributes between spatial classes. Results show that social mixing is more difficult in dense, well-connected areas than in suburban areas, because higher income households are more sensitive to the socioeconomic context of the location in compact areas.

We derive the elasticity formulation for bid-auction location choice models, which allows to quantify the importance of location attributes in location probability. Latent spatial classes also shows an interesting methodology to generate a behaviorally-based CD index for urban areas.

## 4.1. Introduction

The interaction of different social groups in the city, encouraged by living closer to each other, helps towards an equal access to opportunities and a vibrant heterogeneity that allows innovation, as pointed out by Jacobs (1969) and other, more recent, works (Bettencourt, Lobo, Helbing, Kühnert, & West, 2007; Cohendet, Grandadam, & Simon, 2010; Stolarick & Florida, 2006). But most cities normally tend to show socio-spatial segregation (Kempen & Ozüekren, 1998; Massey, 2016), which shows that social mixing is hardly natural to our societies, and therefore it has been a recurrent, although hard to achieve, objective in urban planning policies.

Different types of social mixing policies have been applied in recent decades mainly in Europe, United States and Australia (Galster, 2013). This includes mechanisms such as redevelopment of low-income areas, where some original families remain located, and the rest is sold to higher income households; vouchers for low income families to locate in higher income areas; and imposing a percentage of social housing in exchange for allowing a higher density for the project. This last mechanism has been implemented in Colombia, Brazil and United States, and now it is being discussed in Chile (Dohnke, Heinrichs, Kabisch, Krellenberg, & Welz, 2015; Ruiz-Tagle & Romano, 2019).

Adding density to social housing not only makes these projects more attractive for private developers (as they can sell more units), but also gives urban space a more efficient use. Dense development is consistent with more central locations, where low income households are more likely to improve their access to different oportunities, and also helps the global objective of greater sustainability (United Nations, 2018). Densification of urban areas has several potential positive outcomes, such as shifts to more sustainable transport behavior (Ewing & Cervero, 2017), savings in service provision, reduction in energy use (Ewing, Bartholomew, Winkelman, Walters, & Chen, 1997; Ewing & Rong, 2008), work productivity (Cervero, 2001), innovation (Carlino, Chatterjee, & Hunt, 2011), social mobility (Ewing, Hamidi, Grace, & Wei, 2016); but it has also been associated with an increase in land values, (Downs, 2005), congestion, overcrowding, unaffordability of housing, lack of open spaces, among others (Burton, 2000).

Achieving compact cities is an essential part in many of nowadays urban planning paradigms (such as smart growth, transit oriented development and compact development). Nevertheless, it is restricted in most urban regulations due to its potential negative outcomes (Mills, 2005) and has difficulties in being implemented (Downs, 2005), although in some contexts there is a latent demand for it (Myers & Gearin, 2001) and may bring more satisfaction than suburban living (Lovejoy, Handy, & Mokhtarian, 2010).

Despite the discussion of positive and/or negative effects of compact development (CD) and social mixing, the success of its implementation depends strongly on the reception by the potential household demand. The literature has shown that socioeconomic attributes of neighborhoods are within the most relevant features that defines households' location decision (Schirmer et al., 2014; Waddell, 2006). Understanding (and forecasting) preferences of households for the attributes of these developments is crucial, especially when there is few past experiences in a given context.

In order to add more evidence to the discussion, we propose a model for residential location choice, based on census data for Santiago, which allows to explore the revealed preferences of different types of households towards locating in dense and well-located areas (which we associate to compact development areas), and more precisely, to understand how the socioeconomic level of the area influences their location choices in contexts of compact development, compared to suburban areas.

In our case study, we segment households into nine types, according to three educational levels and three life cycle stages. Then we use the observed locations of these households and, based on a bid-auction framework (Ellickson, 1981) and discrete choice models, we infer how much each type of household values different urban attributes, with average education of the head of the household within the zone (a proxy of the socioeconomic level) being one of them.

To analyze the interaction between compact development and the valuation of socioeconomic level, we use a latent class modeling framework (LCM) (Kamakura & Russell, 1989), that allows us to simultaneously understand how households classify (or "tag") different areas of the city, and how their valuation of location attributes vary according to this classification. This method, originally proposed by Cox and Hurtubia (2019a), allows for an endogenous and data driven segmentation of urban space, avoiding ex-ante or arbitrary subdivisions, and achieving better model fit.

We identify four main contributions in this chapter. First, we report the first application of latent spatial classes in a case study for urban policy making, exploring the benefits of capturing spatial heterogeneity of location preferences. Second, this chapter explores the spatial class membership probability as a method to calculate a behavior-based index of compact development, which could be extended to other urban dimensions. Third, we report evidence on households' demand for urban and socially sustainable development in the Chilean context, which can be useful for urban policy making. Fourth, we derive the formulation for elasticities to be used in a bid-auction location choice framework which, due to the use of willingness to pay instead of utility functions, deviates from traditional elasticities and has not yet been explored in previous literature.

This chapter is structured as follows: section two presents a literature review on compact development and location preferences of households; section three presents the methods used in this research, including the econometric model and formulation of elasticities; section four presents the case study; section five presents the data for the case study; section six presents and analyzes modeling results, and finally conclusions are presented.

## 4.2. Literature review

The following literature review focuses on two main aspects that provide a conceptual framework to our work and which are relevant to contrast and validate the model and the results presented in this chapter. First, we analyze the trends of household location in urban central areas, in order to better understand the type of behavior we are modelling. Second, we explore the literature on social mixture in dense areas, understanding when and where such policies work and therefore providing insight on the potential relevance of our model as an input for public policy design.

### 4.2.1. Compact development and household location preferences

The positive and negative effects of sprawling and compact cities are an ongoing and intensively discussed subject in the urban studies literature (Bruegmann, 2008; Crane, 2008; Ewing & Hamidi, 2015). Compact cities may encourage to travel shorter distances and drive less, but may also lead to excessive crowding and other density-related problems. The most frequent and outspoken paradigm in the last decades has been promoting cities with higher densities, walkability, accessibility to transit, mixing of land uses and types of households, among other principles. Sustainability, in a broad sense, is underlying in all these principles, which have been comprised under Neo-traditional planning with concepts such as Smart Growth, New Urbanism and Transit Oriented Development (Downs,2005; Sharifi, 2017).

Strongly related is the concept of compact development (CD), which is usually associated with " the five Ds": density, diversity, design, distance to transit and destination accessibility (Cervero & Kockelman, 1997; Ewing & Cervero, 2010). Different studies find that density, land-use diversity, and pedestrian-oriented designs generally reduce trip rates and encourage non-auto travel in statistically significant ways (Ewing & Cervero, 2017), although some find its contribution is marginal (Stevens, 2017).

While attributes associated to CD are appealing to planners and environmentalists, they are not very frequent in practice, mostly because of longstanding low density traditions (Downs, 2005), especially in North American cities. This is why an important quantity of research has focused on household preferences for this type of development.

Classical urban economics models explain location as a function of income and a trade-off between dwelling surface and transportation costs, under the assumption of a monocentric city (Alonso, 1964; Mills, 1967; Muth, 1969). In the United States context, suburbs have been traditionally associated to higher income households, as they could afford motorized vehicles and bigger homes. Since the 1970, as cars where more affordable, some high income households no longer saw the advantages of living in suburbs and, also associated to a higher valuation of time, started moving towards central areas of cities (LeRoy & Sonstelie, 1983).

Along with income, lifecycle has an important role in location, as new households usually demand smaller dwellings closer to the city center. For example, as a family grows, more space and other amenities associated to suburban living are demanded (schools, parks, etc.) and, when children leave, there is a tendency to return to smaller dwellings with better accessibility (Doling, 1976). This helps to understand the "return to the city center" phenomena described by LeRoy & Sonstelie (1983), as nowadays a higher proportion of households chose not to raise children, so there is a lower interest in migrating to suburbs.

Recent research based on survey data and choice models, mainly in the United States, have found correlations between households' characteristics and their location choices in respect to CD. While high income households are increasingly gaining interest in locating in central neighborhoods, low income households are still correlated to CD (Cao, 2008; Lewis & Baldassare, 2010; Liao et al., 2014; Olaru et al., 2011; Smith & Olaru, 2013; Walker & Li, 2007).

The return to the center of the city by high income households, sometimes related to gentrification, implies that preferences for central locations are not linear or polarized. Some research find correlations of CD with lower education (Cao, 2008; Olaru et al., 2011) and other with professionals (Walker & Li, 2007)**,** which is consistent with this non-linear and context-dependent dynamic. Lifecycle is central in this aspect, as preferences for CD are both strong in elderly (Cao, 2008; Smith & Olaru, 2013; Walker & Li, 2007) and also young (Smith & Olaru, 2013) and households defined as “non-family” (*i.e.* no children) (Olaru et al., 2011; Walker & Li, 2007).

Research based on generational segmentation found that suburbs are correlated with middle class home-owners, while CD tends to have more millennials and also affluent and highly-educated people (Lee, Circella, Mokhtarian, & Guhathakurta, 2019). CD preferences are also associated with people that appreciate social heterogeneity and have less concern for privacy (Liao et al., 2014), are less likely to own their dwellings (Lewis & Baldassare, 2010; Liao et al., 2014), have fewer cars or don’t drive (Cao, 2008; Olaru et al., 2011), are part of minorities (Lewis & Baldassare, 2010), and have higher awareness for sustainability (Rid & Profeta, 2011).

### 4.2.2. Density and social mixing

Social mixing policies are based on the premise that spatial closeness among different social groups may produce sharing of social capital and therefore higher social sustainability (Bolt & van Kempen, 2013). This argument has been contrasted with evidence from different contexts, showing that integration among social groups is not guaranteed as a function of spatial closeness, and often conflicts may appear (Atkinson & Kintrea, 2000; Mugnano & Palvarini, 2013).

Some evidence shows that concerns about social mixing from medium and high income households are higher if mixing is at the block level than at the neighborhood level (Perrin & Grant, 2014), but can be lowered when compensated with other variables such as design quality (Arundel & Ronald, 2017; Bretherton & Pleace, 2011). Other studies show that even when units in the same project are non-distinguishable, social divisions still appear (Tersteeg & Pinkster, 2016). Research in the Australian context showed that a negative response to social mixing is more related to the densification it implies, than to mixing itself (Nematollahi, Tiwari, & Hedgecock, 2016). As reported by research in Chile, social mixing is also resisted by real estate developers, even in the presence of strong incentives (Greene, Mora, Figueroa, Waintrub, & Ortúzar, 2017; Waintrub, Greene, & Ortúzar, 2016).

When measuring the impact of densification on social diversity, studies in different contexts show that there is a non-linear effect. An important study in 318 U.S. metropolitan areas showed that medium density areas are more segregated than low and high density areas, but changes in density normally increase segregation (Pendall & Carruthers, 2003). Another study in the Brazilian context had similar results, indicating that density in major and polycentric cities increases integration, but in midsize monocentric increases segregation (García-López & Moreno-Monroy, 2016). Other studies have found that densification projects do not assure increasing diversity in the neighborhood (Bramley & Power, 2009; Kim, 2016; Kupke, Rossini, & McGreal, 2012).

The hypothesis that densification of well-located areas may reduce segregation is disputed and different case studies vary in their results. Therefore further evidence and theory is needed to understand the complex relation between household's preferences and levels of density. We believe this chapter contributes on this topic.

## 4.3. Methods

We propose a location choice model following a bid-auction approach (Ellickson, 1983), where households compete for dwellings by bidding their willingness to pay, which is based on their preferences for location attributes. We include latent classes, following Cox and Hurtubia (2019a), where each location in the city belongs to a class with a certain probability, which is a function of zonal attributes. This formulation allows to estimate a different set of preference parameters for each class which, combined with the probability of a location belonging to a class, render willingness to pay functions that are specific to each type of household and each location.

The following sections explain the modelling framework in detail, starting with the mathematical formulation of the (classic) bid-auction model and continuing with the adaptation of the latent class framework for segmentation of urban space.

### 4.3.1. Model framework: Real estate market as a bid-auction model

The bid-auction model proposed by Ellickson (1983) is based on the assumption that a set of households $H$ face a set of dwellings or locations alternatives $I$. Each location ($i$) has a vector of attributes $Z_i$ (built surface, accessibility, among others). Households are segmented in categories ($h$), and each category has a vector of preference parameters $\beta_h$ associated to the valuation of attributes of the location, and a parameter $b_h$ that comprises household structural attributes such as income and expected maximum utility (Jara-Díaz & Martínez, 1999). Based on these assumptions, each type of household $h$ has a Willingness to Pay (WP) for each location $i$:

$$WP_{hi} = b_h + f(Z_i, \beta_h) \tag{4.1}$$

The real estate market is modelled considering that all types of households enter an auction for each location, bidding their WP, and that the type of household with the maximum WP wins the location.

To account for unobserved attributes of locations and heterogeneity in preferences, an i.i.d Gumbel distributed error term can be added to $WP_{hi}$, rendering a multinomial logit model (McFadden, 1973) for the probability of a household $h$ being the highest bidder for location $i$:

$$P(h|i) = \frac{\exp(\mu WP_{hi}(b_h, Z_i, \beta_h))}{\sum_{g\in H} \exp(\mu WP_{gi}(b_h, Z_i, \beta_h))} \tag{4.2}$$

where $\mu$ is a scale parameter. From a sample of located households (segmented by type) and the attributes of their location, using maximum likelihood estimation, we can identify, for each category of household, the parameters $\beta_h$ associated to each attribute and the $b_h$ constant. Equation (4.2) can be interpreted as the probability of the owner of the location or dwelling $(i)$ choosing the highest bidder from the set $H$. This is not the same as the probability of the “choice” approach (McFadden, 1978) which maximizes the likelihood of a household $h$ choosing a location $i$ from a set of alternatives $I$, although Martinez (1992) showed that they are equivalent if some conditions are fulfilled.

The bid auction approach has been used in several Transport and Land Use Interaction (LUTI) models such as MUSSA (Martínez, 1996), ILUTE (Salvini & Miller, 2005) and IRPUD (Wegener, 2011). Several research papers on the literature about location choice also use this approach (Chattopadhyay, 1998; Gross et al., 1990; Hurtubia & Bierlaire, 2014; Hurtubia et al., 2019; Muto, 2006).

### 4.3.2. Latent spatial classes within bid-auction models

Accounting for heterogeneity (segmentation of agents with different preferences) allows for a better interpretation of parameters and model fit. For modelers, a good segmentation is not trivial to define, as preferences of agents are not known a priori. To overcome this issue, Latent Classes (LC) (Kamakura & Russell, 1989) is a technique used in discrete choice models that allows to endogenously segment the decision makers into classes, each with class-specific preference parameters. With LC models, the analyst does not impose a segmentation, but relies on a classification function with parameters to be estimated simultaneously with the preference parameters.

Since Walker and Li (2007) we can find several applications of LC in location choice models (Cox & Hurtubia, 2019b; Ettema, 2010; Liao et al., 2014; Lu et al., 2014; Olaru et al., 2011) These authors do not use LC in a bid-auction model, but in the standard McFadden´s (1978) choice model, where latent classes are applied to segment households. Because of their mathematical structure and interpretation, in bid-auction models the segmentation can be applied to locations (Cox & Hurtubia, 2019a).

Departing from the bid-auction model presented above, we modify equation (4.2) so the probability $P(h|i,s)$ is conditional to each class $s$ of locations:

$$P(h|i,s) = \frac{\exp(\mu_s \cdot WP_{hi}^s(b_h, Z_i, \beta_h^s))}{\sum_{g\in H} \exp(\mu_S \cdot WP_{gi}^s(b_h, Z_i, \beta_g^s))} \tag{4.3}$$

Each agent has a different $WP$ depending on the class of the location where they are bidding, because the $WP$ is function of a set of preferences parameters $\beta_h^s$, which are now conditional to the class of the location.

Simultaneously, each location will have a probability of belonging to a class which, according to the standard formulation of LC models, is a multinomial logit probability based on a classification function $W_{is}$ for which we assume an additive i.i.d Gumbel distribution error term, and a non-identifiable scale parameter $\gamma$:

$$P(s|i) = \frac{exp(\gamma \cdot W_{is}(\widehat{Z}_\iota, \theta_s))}{\sum_{g\in S} exp(\gamma \cdot W_{ig}(\widehat{Z}_\iota, \theta_g)} \tag{4.4}$$

As we are segmenting locations into classes, the class membership function $W_{is}$ depends on a set of location attributes $\widehat{Z}_\iota$, instead of agents characteristics, as in previous applications of LC to location choice (see for example Walker & Li, 2007, and Hoshino, 2011). A vector of parameters $\theta_s$ is estimated, which represent the marginal contribution of each location attribute to the probability of belonging to a spatial class.

Given the probability that agent $h$ gets location $i$, conditional to the class of the location (equation 4.3), and also the probability that location $i$ belongs to class $s$ (equation 4.4), the probability that agent $h$ is the highest bidder for (and therefore gets) location $i$, unconditional to class membership is:

$$P(h|i) = \sum_{s} P(h|i,s) \cdot P(s|i) \tag{4.5}$$

Using equation (4.5), maximum likelihood estimation can be used to identify parameters $\beta_s$ and $\theta_s$ from observed location decisions. This approach avoids an ex-ante definition of the membership of locations to spatial classes and, instead, infers how agents' perceive locations as part of a spatial class, and accordingly variate their preferences.

#### 4.3.3. Calculation of elasticities in a bid-auction location choice model

Traditional formulation of elasticities in logit models is based on how the probability of choosing an alternative varies with respect to the variation of an specific attribute (see, for example, Domencich & McFadden, 1975 and Ortúzar & Willumsen, 2011). This is calculated for attributes that vary only for one of the alternatives (for example, travel time of buses).

In bid models, the attributes included in the WP functions are not specific to households (which are the "alternatives"), but are proper to the location (accessibility, for example). Therefore, a variation in location attributes affects not only one but every alternative (the bidding households). This condition makes the traditional formula for direct elasticity in (choice) logit models not valid for bid-auction location models. We derive the logit function for the particular case of bid-auction models (see annex 7.5), finding that elasticity ($E_{P(h|i),z_i^k}$) for the location probability with respect to the *k-th* location attribute, for a household type $h$ is:

$$E_{P(h|i),\, z_i^k} = \beta_h^k \cdot z_i^k \cdot \left(1 - P(h|i)\right) - \sum_{g \neq h} \left[P(g|i) \cdot \beta_g^k \cdot z_i^k\right] \tag{4.6}$$

where $P(h|i)$ is the location probability of household $h$ in location $i$, $\beta_h^k$ is the preference parameter of the *k-th* location attribute, for household $h$, and $z_i^k$ is the value of the *k-th* attribute in the location $i$.

In the model presented in this chapter, we calculate this elasticity for each spatial class. This only implies that the bid elasticity conditional to a spatial class has to be calculated using the parameters estimated for that spatial class.

In this case, the direct point bid-elasticity can be understood as the percentage change in the probability of household $h$ being the best bidder for location $i$, with respect to a marginal change in attribute $z_i^k$.

The aggregate elasticity ($E_{P(h|s),\, z_i^k}$) for a household type $h$, is calculated by adapting the equation of Domencich & McFadden (1975), and equation (4.6), so it can be specific to a spatial class:

$$E_{P(h|s),\, z_i^k} = \frac{\sum_i \left[ E_{P(h|i,s),\, z_i^k} \cdot P(h|i,s) \cdot P(s|i) \right]}{\sum_i [P(h|i,s) \cdot P(s|i)]} \tag{4.7}$$

### 4.4. Case study: social mixing policies in Chile

The discussion about densification has been abundant and intense in recent years in Chile, especially in its capital, Santiago, as it has observed an important raise in the proportion of built apartments with respect to houses (according to census data, 20.4% of dwellings were apartments in 2002, while this proportion increased to 30.2% in 2017). Central areas of the city have concentrated this type of development, which can be associated to a nation-wide demographic transition, based on an increase of the number of single-person households and couples with no kids, a decline in the number of "traditional" households, together with an important immigration flow and higher life expectancy (Diaz Franulic, 2017). These trends, together with higher transport costs in general, increase demand for more central and smaller dwellings.

Densification has also come into public discussion with a recent law project called *Zonas de Integración Social* (Social Integration Zones), which would allow to define zones in well-connected areas where developers can increase density beyond pre-established limits, conditional to the inclusion of a percentage of affordable (and subject to subsidies) housing.

This mechanism has been already implemented in Colombia, Brazil and the United States (Budds & Teixeira, 2005; Hananel, 2014; Lobo, 2015). After long deliberations, the law project was finally put on hold by the congress, as density incentives where not well defined and it could open a door for unsustainable density levels and excessive developers´ benefit. This mistrust can be rooted in elements such as the stigmatization of dense low-cost apartments in neighborhoods with poor urban standard (one emblematic area was famously dubbed as “vertical ghettos” by an authority), and the undermining of historic low-density neighborhoods by high-rise residential towers.

Social mixing policies in Chile are part of an optimistic approach towards the effects of social mixing, that has been present in urban planning since at least two hundred years (Galster, 2013) but, lately, academia seems to be more skeptic about its real effects. These policies are based on the bridging of social capital, but social interaction between inhabitants of these projects is rare, and the positive effects are mostly about improvement of physical quality of neighborhood (Bolt & van Kempen, 2013).

The mechanism that has been discussed in Chile (private developers being allowed to build higher density in well-located areas, in exchange of integrating social housing), requires an evaluation of how the market may react to these projects. In a strongly market-driven urban development, as it is the case of Chile, the success or failure of these projects depends, among other variables, on the acceptance of the population towards living in socially-diverse projects and, in consequence, on the interest of private developers in building them. The question is more complex when density is included, as Chile has a long tradition of single-family housing.

Using the model presented in the methodology section, we can characterize the potential demand for CD and its drivers. CD zones can be treated as a latent spatial class, so we can identify how household's preferences vary from places that are classified as belonging to a CD class, compared with other spatial classes (e.g. suburban), putting special attention to the variation of preferences with respect to the socioeconomic level in the location. Due to the model structure, we can measure this effects for different types of households.

### 4.4.1. Data

In general terms, the model is fed with household data from the 2017 national census, and land use data extracted from different sources, including own calculations using a Geographical Information System.

### 4.4.2. Household data and segmentation

We use the 2017 national census data to identify 454,570 households that relocated between 2012 and 2017 in the Metropolitan Region of Santiago, from a total of 2,241,551 households living in that area in 2017. There is a data limitation in this aspect, because census only asks each household if they lived the same municipality in 2012, therefore we do not observe households that moved within the municipality.

To estimate a bid-auction model, households have to be segmented into categories. Wealth or resource availability and household structure are main internal determinants of location choice (Doling, 1976). According to available data (census does not ask for income), education of head of household and life cycle of household was used to produce segments.

| Educational Level (years of formal education) | |
|---|---|
| Low-EL | from 1 to 8 years |
| Mid-EL: | from 9 to 12 years |
| HI-EL: | more than 13 years |
| **Life Cycle (age of integrants of household)** | |
| Indep: | All between 18 and 65 years |
| Senior: | No one below 18 years and at least one above 65 years |
| wChild: | At least one below 18 years |

Table 4-1: Household segmentation criteria.

| | Indep | Senior | wChild | TOTAL |
|---|---|---|---|---|
| Low-EL | 20218<br>4% (7%) | 10423<br>2% (8%) | 18294<br>4% (9%) | 48935<br>10% (25%) |
| Mid-EL | 72287<br>15% (14%) | 11445<br>2% (6%) | 72581<br>15% (20%) | 156313<br>33% (40%) |
| Hi-EL | 162977<br>34% (16%) | 13740<br>3% (4%) | 92605<br>20% (15%) | 269322<br>57% (36%) |
| TOTAL | 255482<br>54% (37%) | 35608<br>8% (18%) | 183480<br>39% (44%) | 474570<br>100% |

Table 4-2: Number of households by type.

Both criteria explain different aspects of households, as education level is related to the capacities or availability of resources, and life cycle is more related to needs and restrictions of the household.

Description of segmentation criteria and resulting proportions is shown in tables 4-1 and 4-2. Table 4-2 shows number of households by segment. Second line shows the proportion of each segment in the estimation dataset (movers), and the proportion of the segment in the population (between parentheses). As can be seen, and confirming intuition, independent households with high education are overrepresented among movers with respect to their total population in the study area, and seniors in general (but more in low education) are underrepresented.

### 4.4.3. Urban context data

Each dwelling in the census is georeferenced at a zonal level (there are 1,630 zones in the study area, with an average surface of 47 Ha. each but significantly smaller in central areas). Urban context attributes, such as accessibility (distance to city center and to nearest subway station), zonal land use (percentage of commerce, built density, land use mix), socioeconomic level (% of households with high education in the zone) and a proxy of dwelling size (average size of dwellings in the zone) are calculated for each zone.

Since these attributes are used to explain relocations that took place between 2012 and 2017, they are calculated for years 2012 and 2014, depending on the source data availability, which can be closer to what households observed when taking the decision.

| Attribute | Mean | Min | Max | Description | Year | Source |
|---|---|---|---|---|---|---|
| Distance to Nearest Subway (km) | 2.25 | 0.12 | 12.43 | Euclidian Distance (in GIS) | 2012 | Own Calculation |
| Distance to City Center (km) | 9.9 | 0.22 | 23.66 | Euclidian Distance (in GIS) | 2012 | Own Calculation |
| % Commerce | 0.04 | 0 | 0.48 | Percentage of Built Surface for commerce | 2014 | SII |
| Avg Unit Built Surface | 54.6 | 18.9 | 230.9 | Average of the surface of residential units in census | 2014 | SII |
| Land Use Entropy | 0.59 | 0 | 0.98 | Index of Diversity in the zone | 2014 | Own Calculation with SII Data |
| Built Density | 0.48 | 0 | 5.86 | Ratio between built surface and total area of census | 2014 | SII |
| % Hi-EL Households | 0.36 | 0 | 0.96 | Percentage of heads of households with aobove hig | 2012 | Census |

Table 4-3: Statistics, description and source of urban attributes.

Land use entropy is a measure of diversity, and is calculated following Turner, Gardner, & O'Neill (2001), and corrected considering the overall proportions of each land use in the study area (Song, Merlin & Rodriguez, 2013) as seen in equation (4.8).

$$Entropy_i = -\frac{\sum_u \left[\left(\frac{\mathrm{r_{ui}}}{\sum_v r_{vi}}\right) \cdot ln(\frac{\mathrm{r_{ui}}}{\sum_v r_{vi}})\right]}{ln(U)} \tag{4.8}$$

where $U$ is the number of land uses (five in our case study) and $r_{ui}$ is:

$$\mathrm{r_{ui}} = \frac{p_{ui}}{p_u} \tag{4.9}$$

where $p_{ui}$ is the percentage of land use $u$ in zone $i$ and $p_u$ is the percentage of land use $u$ in all the study area.

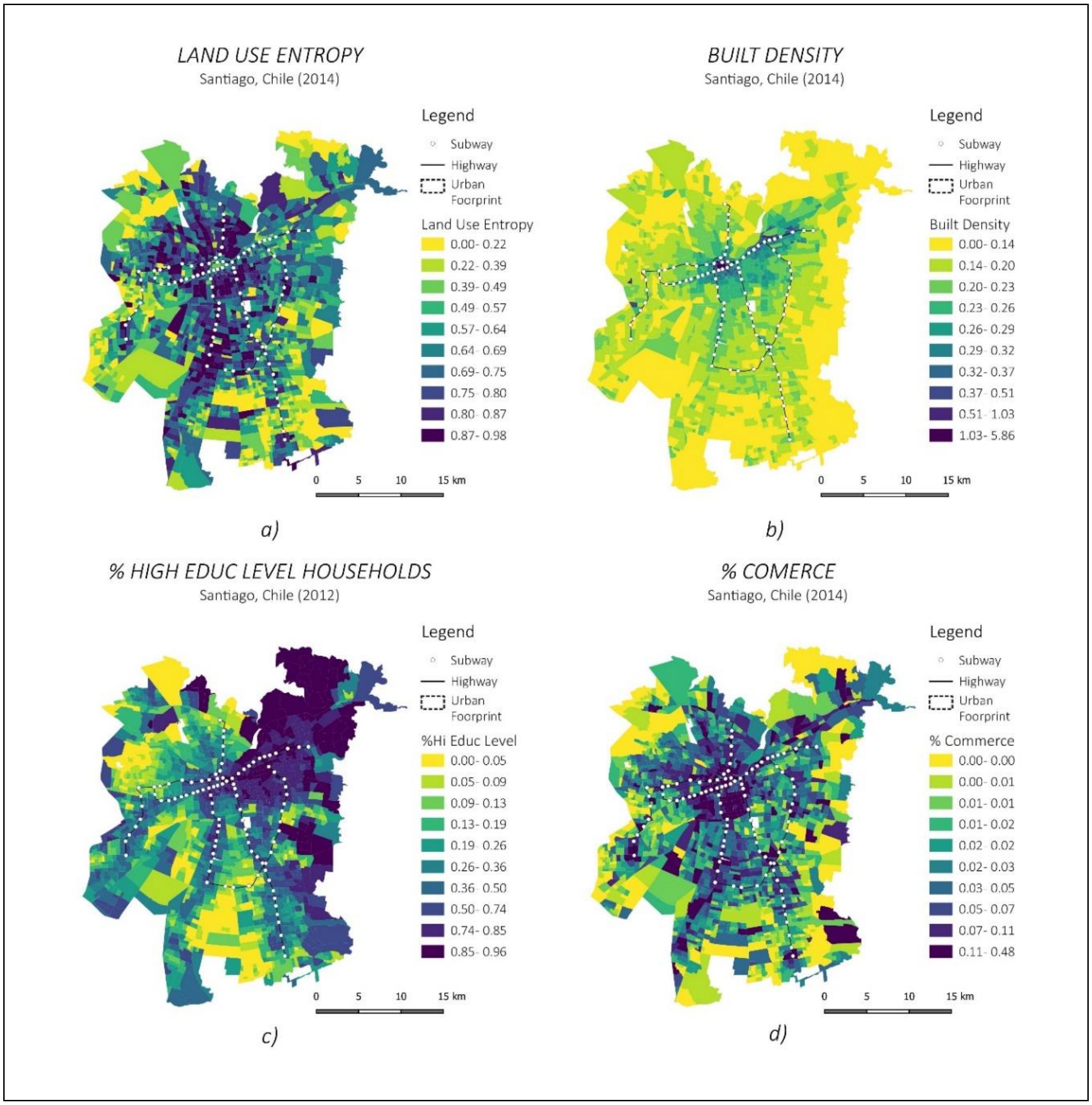



Figure 4-1: Distribution of urban attributes for the location choice model

The entropy index for a zone $i$ assumes the value of 1 in a census zone when the proportion of each land use in that zone is equal to the proportion of the same land use type in the complete study area, and 0 if there is only one type land use in the whole zone. Entropy in this case study is calculated using the built surface of five land uses: residential, commercial, office, industry and services (educational, sports, religious, health), which are taken from the Internal Revenue Service (SII, for its name in Spanish) database (Servicio de Impuestos Internos, 2014). Figure 4-1 shows the spatial distribution of this and other land use attributes for the study area.

### 4.5. Results and analysis

Estimated parameters for the WP function of each type of household for each class of location (see equation 4.3) allow to understand the relative value of each urban attribute in their location decisions, enabling comparison across households. In the other hand, parameters estimated for the spatial segmentation function (see equation 4.4) allow to characterize the segmentation criteria and the labeling of each class as CD or Suburban.

With both functions we calculate an aggregate location probability of each type of household in each of the two spatial classes, so we can compare which type of household is more likely to locate in CD zones.

The spatial segmentation function, with its parameters, allows to map each area of Santiago according to their probability of being classified as CD, which reports a spatial distribution of this segmentation. We calculate elasticities in the segmentation function to measure the weight of density, distance to subway and diversity in the probability of a location being classified as CD.

We also calculate the aggregate elasticity of each location attribute, for each type of household in each spatial class. This is central to our analysis as it allows to report, for each type of household, how the relevance of socioeconomic level in the area varies depending on that area being CD or not.

### 4.5.1. Estimation results and elasticities

Due to the fact that, in a logit model, the effect of the WP of a household with respect to the choice probability is relative to the value of the WP of the other households, we had to fix to zero the parameters of one arbitrary household type (we choose Low-EL with Children), otherwise the likelihood maximization problem is indeterminate, with multiple possible solutions. Therefore, all the values of the parameters are relative to those of this household, and the signs do not necessarily represent that the household has a negative or positive valuation of the attribute, but only if its valuation is higher or lower than the valuation of the reference household type. For the same reasons, we also fixed in zero the parameters of the segmentation function of one of the spatial classes.

Table 4-4 shows the parameters of both the WP functions specific to household type and spatial class (two central columns), and the parameters of segmentation function (in the lower part of the table). Except for eight estimates, all parameters are significant at the 95% confidence level. Columns to the right show the aggregate elasticities of each attribute with respect to location or class membership probability.

The parameters of the segmentation function (bottom of Table 4-4), show that, for class one, the parameter for built density is negative, positive for distance to nearest subway station, and negative for land use entropy. Therefore, spatial class one can be labeled as "suburban zones" and, by opposition, spatial class two can be labeled as compact development (CD). The calculation of aggregate elasticities of these three attributes for the probability of being classified as CD indicates that built density (0.13) is less relevant than distance to subway (-0.18) and land use entropy (0.27). Figure 4-2 shows the spatial distribution of the probability of belonging to the CD class.

Regarding location preferences we find, as expected, significant differences in elasticities between households. Moreover, for each type of household, there are significant differences in parameters between spatial classes, which validates the introduction of spatial heterogeneity through the identification of latent classes.

Comments on location probability elasticities for the socioeconomic level of the location is discussed in a special section below.

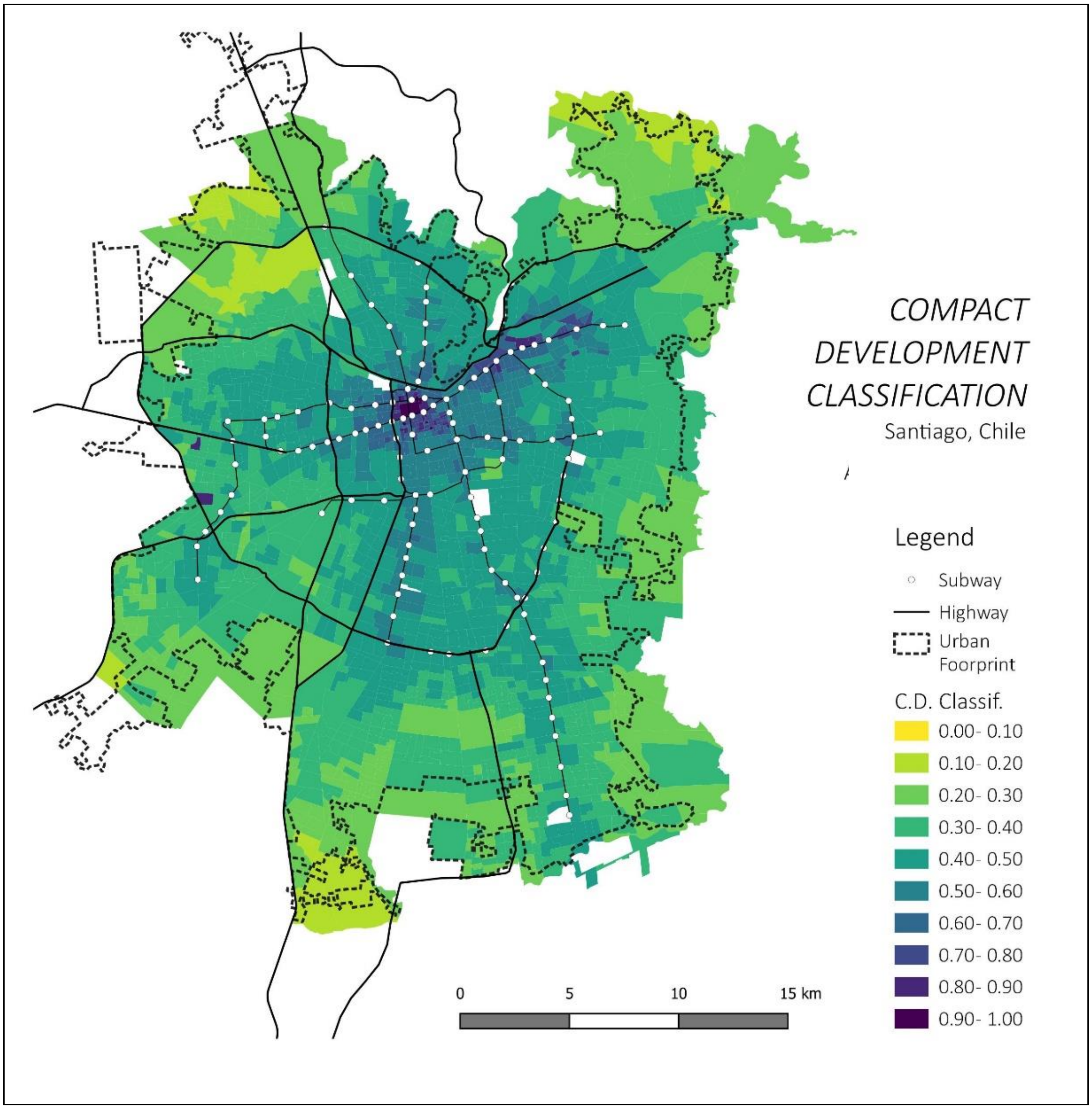



Figure 4-2: Probability of membership to compact development class in Santiago.

| Location Attribute | Household Types | | Class Specific Coefficients (and t-test) | | Location Probability Elasticity | |
|---|---|---|---|---|---|---|
| | Education Level | Life Cycle | Compact Development | Suburban | Compact Development | Suburban |
| Constant | Low-EL | Indep | 1.11 (5.2) | -0.927 (-9.43) | | |
| | | Senior | 0.656 (3.33) | -2.26 (-10.9) | | |
| | | wChild | 0 | 0 | | |
| | Mid-EL | Indep | 2.3 (12.57) | -0.6 (-6.68) | | |
| | | Senior | -2.6 (-11.5) | -1.52 (-9.26) | | |
| | | wChild | 2.16 (10.94) | 0.378 (5.16) | | |
| | Hi-EL | Indep | -0.224 (-1.21) | 0.351 (4.47) | | |
| | | Senior | -3.58 (-15.89) | -3.13 (-18.79) | | |
| | | wChild | 0.364 (1.73) | -1.53 (-21.22) | | |
| Distance to City Center (km) | Low-EL | Indep | -0.283 (-15.81) | 0.0486 (9.76) | -0.97 | 0.10 |
| | | Senior | -0.0817 (-9.41) | 0.00403 (0.48) | 0.42 | -0.35 |
| | | wChild | 0 | 0 | 1.37 | -0.38 |
| | Mid-EL | Indep | -0.239 (-24.38) | 0.0606 (13.33) | -0.78 | 0.01 |
| | | Senior | -0.0155 (-1.74) | -0.0665 (-7.5) | 1.00 | -0.86 |
| | | wChild | -0.236 (-16.67) | 0.067 (17.7) | -0.79 | 0.33 |
| | Hi-EL | Indep | -0.0794 (-10.09) | -0.226 (-40.83) | 0.19 | -1.21 |
| | | Senior | -0.012 (-1.35) | -0.0939 (-11.2) | 0.87 | -0.87 |
| | | wChild | -0.0969 (-9.68) | 0.0456 (12.36) | 0.18 | 0.21 |
| % Hi-EL Households | Low-EL | Indep | 12.8 (13.18) | -1.92 (-14.93) | -0.78 | -0.59 |
| | | Senior | 11.3 (11.74) | 2.27 (10.15) | -0.53 | -0.03 |
| | | wChild | 0 | 0 | -0.91 | -0.40 |
| | Mid-EL | Indep | 13.7 (14.28) | 0.972 (9.98) | -0.65 | -0.32 |
| | | Senior | 16.1 (16.5) | 2.17 (11.43) | -0.19 | -0.23 |
| | | wChild | 12.4 (12.87) | 0.786 (9.13) | -0.68 | -0.30 |
| | Hi-EL | Indep | 18.2 (18.82) | 4.2 (42.66) | 0.65 | 0.31 |
| | | Senior | 17.8 (18.26) | 4.56 (20.54) | 0.38 | 0.50 |
| | | wChild | 15.7 (16.44) | 4.71 (58.21) | -0.18 | 0.65 |
| % Comerce | Low-EL | Indep | 18.7 (7.45) | -1.16 (-3.08) | 0.08 | 0.01 |
| | | Senior | 16.3 (6.42) | -2.8 (-3.64) | -0.02 | -0.04 |
| | | wChild | 0 | 0 | -0.25 | 0.05 |
| | Mid-EL | Indep | 18.2 (7.31) | -0.822 (-2.83) | 0.04 | 0.21 |
| | | Senior | 17.3 (6.87) | -6.42 (-6.89) | 0.00 | -0.17 |
| | | wChild | 18.3 (7.3) | -3.04 (-11.24) | 0.05 | -0.04 |
| | Hi-EL | Indep | 17.3 (7.02) | 2.04 (8) | -0.01 | 0.18 |
| | | Senior | 18.6 (7.42) | -8.7 (-10.62) | 0.06 | -0.26 |
| | | wChild | 17.2 (6.76) | -1.93 (-8.17) | -0.01 | -0.03 |
| Avg Unit Built Surface (m2) | Low-EL | Indep | -0.00942 (-2.77) | 0.0177 (13.05) | 0.00 | 0.38 |
| | | Senior | -0.0206 (-5.51) | 0.00709 (3.66) | -0.47 | -0.12 |
| | | wChild | 0 | 0 | 0.36 | -0.43 |
| | Mid-EL | Indep | -0.00871 (-2.65) | 0.0105 (9.12) | 0.03 | 0.00 |
| | | Senior | -0.000425 (-0.12) | 0.0125 (5.05) | 0.42 | 0.09 |
| | | wChild | -0.014 (-4.18) | 0.00447 (4.51) | -0.21 | -0.24 |
| | Hi-EL | Indep | -0.00531 (-1.63) | 0.00859 (8.73) | 0.11 | -0.18 |
| | | Senior | 0.000965 (0.29) | 0.0247 (17.58) | 0.52 | 0.83 |
| | | wChild | -0.0228 (-5.57) | 0.017 (18.83) | -0.68 | 0.28 |
| Class Membership Attribute | | | | | | |
| Intercept | | | 0 | 0.927 (26.42) | | |
| Built Density | | | 0 | -0.66 (-35.62) | 0.13 | -0.26 |
| Distance to Closest Subway | | | 0 | 0.101 (29.66) | -0.18 | 0.07 |
| Land Use Entropy | | | 0 | -0.852 (-29.94) | 0.27 | -0.26 |

Table 4-4: Model estimates, for WP and spatial class segmentation functions.

By comparing the observed and the predicted proportions of households of each type in each zone, the model has a fit ($R^2$) of 0.88.

### 4.5.2. Location probability shift by household type in compact development neighborhoods.

To evaluate which type of households are more likely to locate in each spatial class, we calculate an aggregate location probability, following the equation:

$$P(h|s) = \frac{\sum_{i \in I}[P(h|i,s) \cdot P(s|i) \cdot H_i]}{\sum_{i \in I}[P(s|i) \cdot H_i]} \tag{4.10}$$

where $H_i$ is the total number of households observed in location $i$.

This aggregate probability indicates which household types are more attracted by the CD spatial class in their location decisions. Table 4-5 shows some intuitive patterns but also some interesting variations. Independent households are more likely to locate in CD, except when they have a low educational level. Seniors are also more likely to locate in CD, but this declines when their education level increases (less educated senior households have a low probability of living in suburbs, but for high education seniors the probability is not so different than living in CD). Households with children have higher probabilities of living in suburbs, regardless their educational level.

| Education Level | Life Cycle | Compact Development | Suburban | Relative difference (CD over Suburban) |
|---|---|---|---|---|
| Low-EL | Indep | 3.3% | 5.5% | -40% |
| | Senior | 3.2% | 0.4% | 759% |
| | wChild | 2.0% | 5.6% | -64% |
| Mid-EL | Indep | 16.7% | 12.1% | 38% |
| | Senior | 3.1% | 2.0% | 57% |
| | wChild | 7.9% | 22.2% | -65% |
| Hi-EL | Indep | 52.8% | 19.1% | 176% |
| | Senior | 3.7% | 2.4% | 58% |
| | wChild | 7.3% | 30.7% | -76% |
| | | 100% | 100% | |

Table 4-5: Aggregate location probability.

### 4.5.3. Elasticities for socioeconomic level in compact development neighborhoods

Looking at elasticities in Table 4-4 (last two columns of highlighted rows), for the variable of % of Hi-EL households (which accounts as a proxy of socioeconomic level in the neighborhood of a location), we can see that an increase in this variable affects very differently the probabilities of location, depending on the household type. As expected, Mid-EL and Low-EL households have a lower location probability when the socioeconomic level in the zone increases, whereas the opposite happens with Hi-EL households, with one exception. Hi-EL households with Children decrease their probabilities of locating in CD when the socioeconomic level in the area increases, which seems counterintuitive, especially considering their positive parameter for this attribute. We hypothesize that this is an effect of competition with other Hi-EL households, especially Hi-EL Independent households, with strong preferences for these locations and more disposable income, therefore outbidding other households.

We can identify significant differences between spatial classes in location elasticities for this attribute. An important difference is seen for Hi-EL Independent households, for which the socioeconomic level of the location affects the double when the zone is CD. This difference is important, as this type of household has the higher overall probability of locating in CD zones.

When looking at elasticities of Mid-EL and Lo-EL households, for all of them except one, an increase in socioeconomic level of the location has a stronger negative effect when the zone is CD than when is Suburban. This means that, while a raise in socioeconomic level in a zone implies a higher difficulty for Lo-EL and Mid-EL households to locate, this difficulty is even higher in CD zones.

### 4.5.4. Compact development classification

Beside the analysis of location, the estimated model allows to report an endogenous classification of location areas into CD or Suburban, based on the probabilistic classification of the $W_s$ function. If we apply this function with its estimated parameters to each zone, we can report a probability of CD for each one. Figure 4-2 shows the spatial distribution of this probability.

As it can be expected, the probability is high along the subway lines, and gets its maximum value in the city center, which may not have the highest value in land use diversity, but has a high built density.

Figure 4-3 shows the amount of surface in the city that can be classified as CD according to the class membership probability. This graph shows that only a small area (0.54%) of the city has a probability above 75% of being CD, while 17.7% of the city has a probability above 75% of being suburban. If we apply a strict classification of the city into both classes, cutting in the 50% probability, only 8.5% of the city would belong to the CD class.

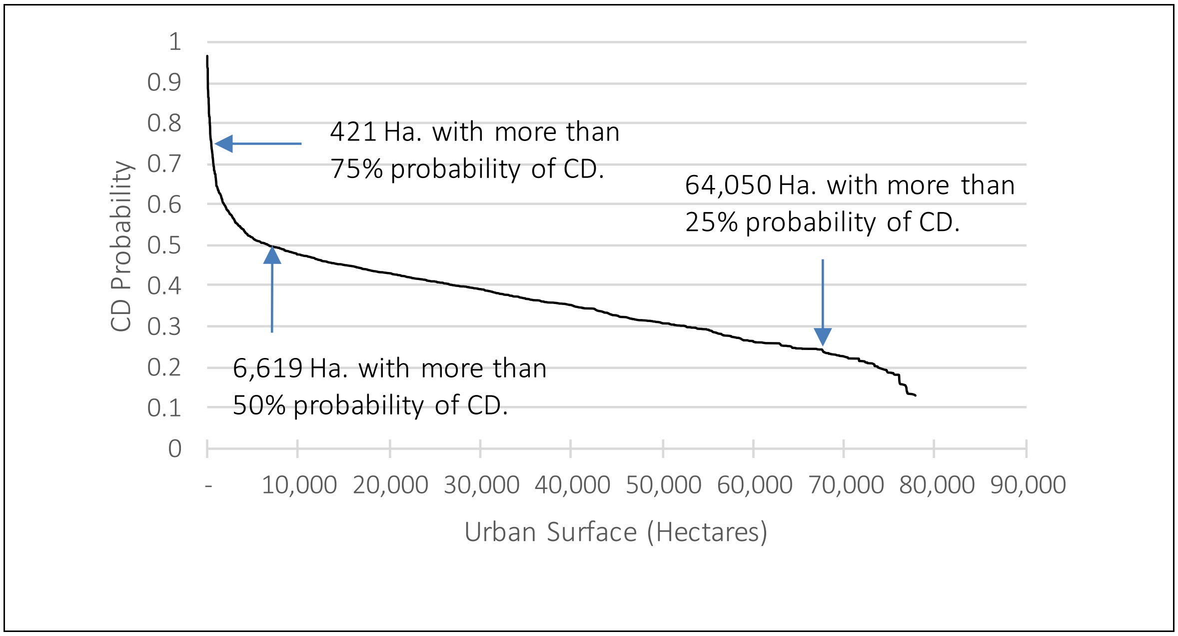


Figure 4-3: Cumulative probability of being classified as CD for all zones of the city.

Results from the classification function also allow to determine thresholds of the urban attributes at which the CD probability reaches values closer to one or to zero. This sensitivity analysis could allow planners to determine the combination of attributes that are necessary for an area to be perceived as CD.

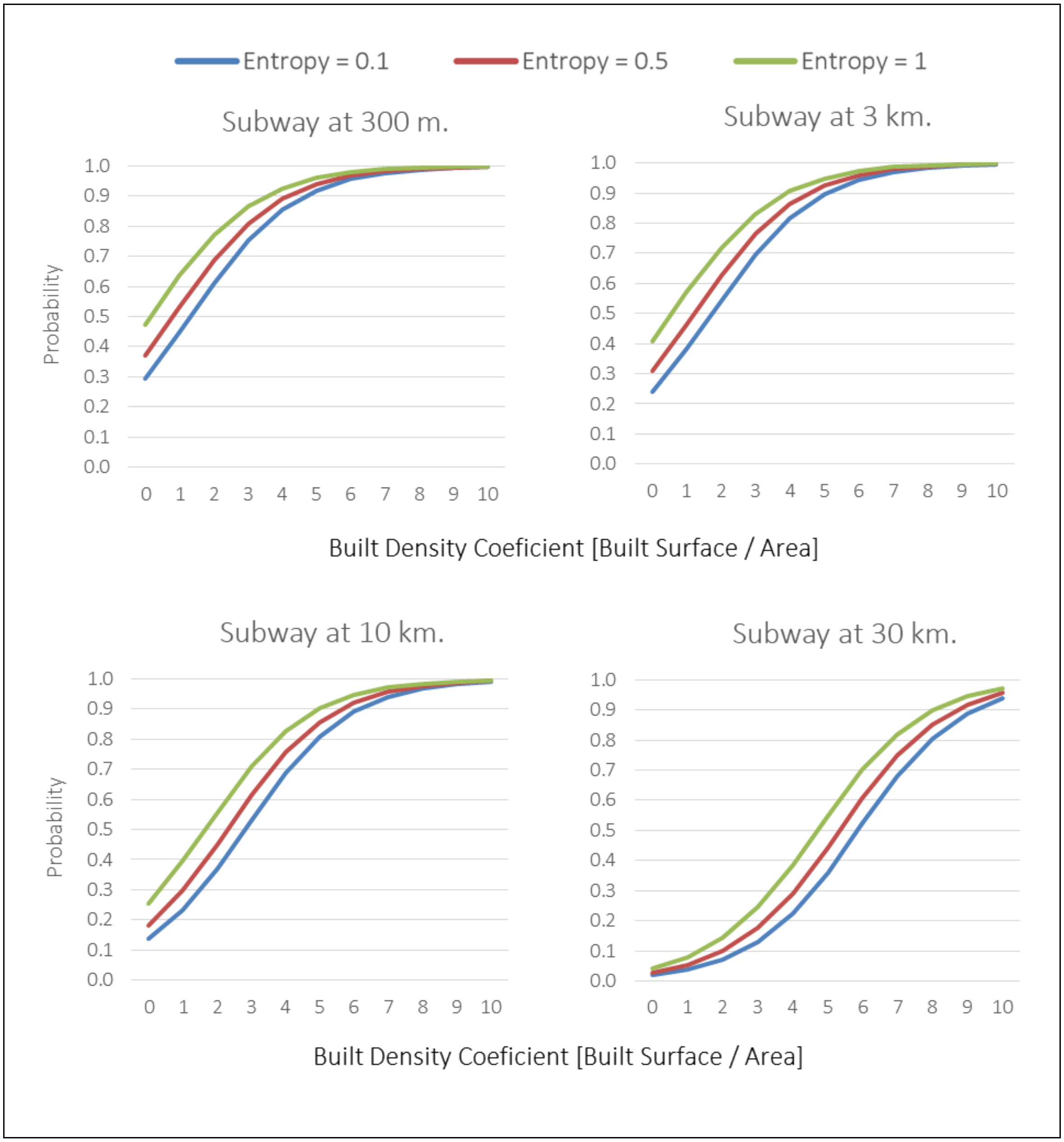



Figure 4-4: Charts showing the probability curve of a zone being classified as a CD neighborhood, depending on built density, distance to subway and land use entropy.

As shown by the plots in Figure 4-4, in an area somewhat close to a subway station (300 m. upper-left), the probability of CD reaches 0.95 when built density is around 5[9]. Graphs show that if entropy is lower, built density has to be higher to compensate. When subway stations are farther away (see figures at the bottom), higher densities are needed (e.g. built density coefficient of around 10 when subway is at a distance of 30 km).

### 4.6. Conclusions

Results show that, for Santiago, compact development areas are attractive to independent and highly educated households which, due to their higher income, tend to be the best bidders for dwellings in those locations. This, at first glance, could be interpreted as a positive trend, because it makes compact development more economically viable, triggering urban renewal and favoring a more compact city. However, results also show that, in general, households are more sensitive to the socioeconomic level in the neighborhood when considering a CD location. While for low and mid education levels, an increase in the socioeconomic level of the location implies a strong reduction of their likelihood to be able to locate there, this is the opposite for high education households. This is not only a problem due to the reduced probability of lower income households locating in CD areas, but also an issue if social

[9] This coefficient means that in a block with total terrain surface *x*, the total built area is *5x* (this would led, for example, to a 10 story building with a footprint of half of the terrain).

housing is built in such areas and triggers the relocation of higher income households, therefore defeating the purpose of such policies if social integration was between their objectives.

Nevertheless, this difference is not so strong in suburban zones, indicating that social mixture can be more easily achieved in this class of locations. Suburban zones, with less pressure and more available land, are less prohibitive for Lo-EL households even when the neighborhood socioeconomic level increases. This is consistent with the findings of Pendall and Carruthers (2003) who showed that increases in density may exacerbate segregation due to the higher WP of households with high income and no kids. These results are also relevant for explaining gentrification dynamics, which may quickly turn the social composition of a neighborhood and exclude lower income households, therefore increasing social segregation.

Spatial segmentation into latent classes shows that CD zones are scarce in Santiago. At the same time, demand for dwelling in these locations come primarily from the types of households that are most quickly growing in Chilean society: Independents and Seniors (Diaz Franulic, 2017). Both elements contribute to a higher market pressure over CD areas.

These results mean that social mixing, which is already difficult in the Chilean context (and apparently, elsewhere), could be even more difficult in dense and well-located projects. CD is positive in terms of sustainability and innovation, but it may not be recommendable to be implemented together with social mixing unless accompanied by complementary policies.

Results show that, unlike what it is usually believed, zones with a higher probability of belonging to the suburban class are more likely to accommodate social mixing with success. It is important to notice, however, that no zone belongs absolutely to one class or the other, and that a high probability of belonging to the suburban class can still be achieved with moderately high densities or with good (although probably not excellent) distance to metro and land use diversity.

The latent spatial segmentation method used in this work, allows to identify CD zones as a function of built density, distance to subway stations and land use diversity. The classification function and the subsequent logit probability of belonging to the CD class, can be interpreted as a Compact Development Index, which goes from 0 to 1. This indicator, with parameters that are estimated from observed behavior, can be a useful contribution for building urban indexes, which are usually hard to construct due to the difficulty in calibrating their parameters (Maclaren, 1996; Oliva et al., 2018). The proposed model also allows to find which combinations of values, for certain urban variables, offer a higher probability of a zone being classified as CD. These findings are particularly useful for informing public policy and definition of built environment regulations, when the objective is to generate more CD areas.

Another relevant contribution of this work is the derivation of the elasticity formula for location choice models using the bid-auction approach which, to the extent of our knowledge, has not been proposed in the literature before. This formulation is also necessary to estimate elasticities for the classification probability in any discrete choice model with latent classes. It is also useful in regular discrete choice models, to estimate elasticities for characteristics of the decision maker that are typically included in the specification of the utility function (e.g. age or gender in mode choice models) but for which, so far, no elasticity formula has been proposed.

**References** for all chapters are presented in a specific chapter after the conclusions.

# 5. CONCLUSIONS

Contributions provided by this thesis range from direct methodological improvements and evidence from case studies, to wider understanding of modelling urban phenomena.

## 5.1. Direct contributions

The direct contributions of this thesis can be summarized in four main aspects. We will focus on methodological aspects, as contributions about evidence are mostly summarized in each chapter.

A first contribution is the formulation of latent spatial classes, which is a clear advance in treating spatial heterogeneity with a strong behavior basis. In this thesis, latent spatial classes is used in the context of location choice models, but it can be extended to be used in parallel with other discrete choice models, including mode choice in transport modelling, and also to continuous models, such as linear regressions estimated with ordinary least squares. Therefore, a wide range of phenomena can be modelled including this technique, opening a vast area for future work.

A second contribution is the use of latent classes in location choices as a method to explore polarization in space, which in this thesis has been related to spatial segregation patterns. Related to this, a third contribution is the use of spatial latent classes to provide a robust method to measure a specific characteristic of urban areas. Probability of membership to a latent spatial class can be used for this purpose, if the class can be labelled with certain relevant urban characteristic (compact development, using the study case for this thesis). This probability, being measured from 0 to 1, resembles a normalized index. Traditionally, indexes are based in a set of attributes, each being associated to a weight or its marginal contribution to the final value of the index. Using latent spatial classes allows to find, with a behavioral basis, the weights for each attribute, and also to measure if they are significant.

A fourth contribution is the formulation of a “bid elasticity”, which was not part of original objectives, and deserves a more extensive future analysis of its applications and properties. It is important to notice that its application is extensive to any discrete choice based on a logit function, in which the variable to be analyzed is part of the decision maker.

**5.2. Contribution to urban modelling**

Modelling is not only about reproducing observed data, but also implies a deeper understanding of the essential dynamics of a phenomena. The approach presented in this thesis is intended as a step towards a more comprehensive representation of people´s behavior in the city.

Traditional modelling of location choices, and also of travel behavior, normally assumes that urban attributes can be directly measured, represented in variables such as distances, travel times, zonal densities or built surface. The econometric effort is mainly focused in identifying unobservable variables from the decision maker side, normally preferences.

But the city is not only about plain built space that can be measured transparently. There is an underlying structure of the city as perceived by people in everyday life, which has been explored in other disciplines such as urban geography and urbanism.

Kevin Lynch (1960), a seminal author in this area, studied the urban structure from characteristic of a city that people can identify (in interviews) as important in their "navigation" of the city.

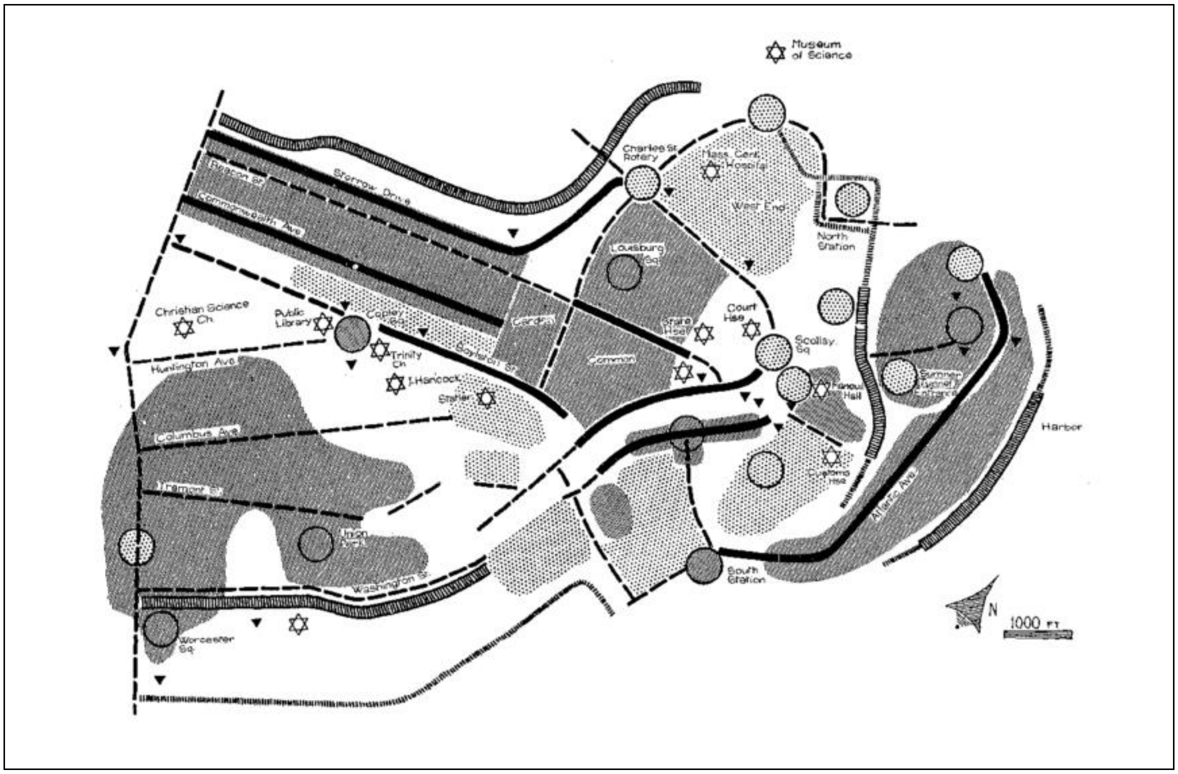



Figure 5-1: Lynch´s map of Boston.

Actually, people perceive limits, nodes, hierarchies, zones, and others elements, which do not necessary relate to administrative boundaries. It is reasonable to think that their behavior answers to these perceptions, therefore models of residential location also should.

Nowadays, measuring this underlying structures can be approached with GIS, through heuristics such as Space Syntax or clustering methods, based on built elements and or natural barriers. First approaches for the models presented in this thesis were built upon this type of methods, but they lacked a behavioral basis, as they were based on pure physical elements. So the resulting structures reflected differences in physical attributes, but didn't necessarily had a relation to people´s behavior.

The main concern in this thesis was to design a model of people´s behavior in which urban underlying structures play a role. The formulation of the model had to allow that these structures could not be observable, but identifiable from peoples´ decisions, same as preferences. Modelling behavior with latent classes resulted as a very consistent way to simultaneously identifying both preferences and also distinctive zones in the city.

Probably the main contribution of this thesis is to state that, even though urban attributes can be measured, the way they shape macrostructures that people perceive is not known but can be unveiled observing how people behave. So unobserved or latent aspects of these models are not only on decision maker's side, but also in the urban attribute side.

Because of estimation limitations, the maps that resulted from the models presented here are still rough, being able only to segment the city into few macrozones. In this sense, there is still a long way to capture the rich complexity of a city´s urban structure. Nevertheless, it opens the door to explore other behavior-based methods, which may be suitable to identify more complex urban structures.

Considering that they depart from a solid behavioral basis, these models seem to be a solid start and a good contribution in crossing disciplines between econometrics and urban geography.

## 6. ACKNOWLEDGMENTS

The authors would like to acknowledge the support provided by the Center for Sustainable Urban Development (CEDEUS, CONICYT/FONDAP 15110020), the Complex Engineering Systems Institute, ISCI (CONICYT PIA/BASAL AFB180003), FONDECYT (project N°1180605), and the CONICYT PhD Scholarship for the corresponding author (2015-2019).

## 7. REFERENCES


Abbott, A. (1997). Of time and place: the contemporary relevance of the Chicago School. *Social Forces*, *75*(4), 1149–1182.

Adair, A. S., Berry, J. N., & McGreal, W. S. (1996). Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research*, *13*(1), 67–83.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer.

Alexander, C. (1965). A city is not a tree. *Architectural Forum*, *122*(1), 58–62.

Alonso, W. (1964). *Location and land use*. Cambridge: Harvard University Press.

Anselin, L. (1988). *Spatial econometrics: methods and models*. Boston: Springer Science & Business Media.

Anselin, L. (1999). *Spatial econometrics*. https://doi.org/10.1016/j.regsciurbeco.2006.11.009

Antoniou, C., & Picard, N. (2015). Econometric methods for land use microsimulation. In M. Bierlare, A. De Palma, R. Hurtubia, & P. Waddell (Eds.), *Integrated Transport and Land Use Modeling for Sustainable Cities* (pp. 95–112). Lausanne: EPFL Press.

Arundel, R., & Ronald, R. (2017). The role of urban form in sustainability of community:

The case of Amsterdam. *Environment and Planning B: Urban Analytics and City Science*, *44*(1), 33–53.

Atkinson, R., & Kintrea, K. (2000). Owner-occupation, social mix and neighbourhood impacts. *Policy and Politics*, *28*(1), 93–108.

Batty, M. (2008). The size, scale, and shape of cities. *Science*, *319*(5864), 769–771.

Becker, G., & Murphy, K. (2009). *Social Economics: Market Behavior in a Social Environment*. Cambridge: Harvard University Press.

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. Cambridge: MIT Press.

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. (2002). Integration of choice and latent variable models. *Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges*, 431–470.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(17), 7301–7306.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. In *3rd Swiss Transportation Research Conference*. Ascona, Switzerland.

Bierlaire, M. (2020). *A short introduction to PandasBiogeme*. *TRANSP-OR 200605.*

*Transport and Mobility Laboratory, ENAC, EPFL*. Retrieved from https://transp-or.epfl.ch/documents/technicalReports/Bier20.pdf

Bolt, G., & van Kempen, R. (2013). Introduction Special Issue: Mixing Neighbourhoods: Success or Failure? *Cities*, *35*, 391–396.

Borsdorf, A. (2003). Como modelar el desarrollo y la dinámica de la ciudad latinoamericana. *EURE, XXIX*(86), 37–49.

Borsdorf, A., & Hidalgo, R. (2008). New dimensions of social exclusion in Latin America: From gated communities to gated cities, the case of Santiago de Chile. *Land Use Policy*, *25*(2), 153–160.

Borsdorf, A., Hidalgo, R., & Sánchez, R. (2007). A new model of urban development in Latin America: The gated communities and fenced cities in the metropolitan areas of Santiago de Chile and Valparaíso. *Cities*, *24*(5), 365–378.

Borsdorf, A., Hildalgo, R., & Vidal-Koppmann, S. (2016). Social segregation and gated communities in Santiago de Chile and Buenos Aires. A comparison. *Habitat International*, *54*, 18–27.

Bourassa, S. C., Hamelink, F., Hoesli, M., & Macgregor, B. D. (1999). Defining housing submarkets. *Journal of Housing Economics*, *8*(2), 160–183.

Bramley, G., & Power, S. (2009). Urban form and social sustainability: the role of density and housing type. *Environment and Planning B: Planning and Design*, *36*(1), 30–48.

Bretherton, J., & Pleace, N. (2011). A difficult mix: Issues in achieving socioeconomic diversity in deprived UK neighbourhoods. *Urban Studies, 48*(16), 3433–3447.

Bruegmann, R. (2008). Point: Sprawl and Accessibility. *Journal of Transport and Land Use*, *1*(1), 5–11.

Budds, J., & Teixeira, P. (2005). Ensuring the right to the city: Pro-poor housing, urban development and tenure legalization in Saõ Paulo, Brazil. *Environment and Urbanization*, *17*(1), 89–113.

Burton, E. (2000). The compact city: Just or just compact? A preliminary analysis. *Urban Studies*, *37*(11), 1969–2001.

Cáceres-Seguel, C. (2015). Ciudades satélites periurbanas en Santiago de Chile: paradojas entre la satisfacción residencial y precariedad económica del periurbanita de clase media. *Revista INVI*, *30*(85), 83–110.

Cáceres-Seguel, C. (2017). Vivienda social periurbana en Santiago de Chile : la exclusión a escala regional del trasurbanita de Santiago de Chile. *Economía, Sociedad y Territorio*, *17*(53), 171–198.

Cao, X. (2008). Is alternative development undersupplied? Examination of residential preferences and choices of Northern California movers. *Transportation Research Record*, *2077*(1), 97–105.

Carlino, G. A., Chatterjee, S., & Hunt, R. M. (2011). Urban Density and the Rate of Invention. *SSRN Electronic Journal*, (January).

Cervero, R. (2001). Efficient urbanisation: Economic performance and the shape of the metropolis. *Urban Studies*, *38*(10), 1651–1671.

Cervero, R., & Kockelman, K. (1997). Travel Ridership and the 3Ds: Density, Diversity and Design. *Transportation Research D*, *2*(3), 199–219.

Chattopadhyay, S. (1998). An empirical investigation into the performance of Ellickson's random bidding model, with an application to air quality valuation. *Journal of Urban Economics*, *43*(2), 292–314.

Chica-Olmo, J. (1995). Spatial Estimation of Housing Prices and Locational Rents. *Urban Studies*, *32*(8), 1331–1344.

Clark, W. A. (1991). Residential preferences and neighborhood racial segregation: a test of the Schelling segregation model. *Demography*, *28*(1), 1–19.

Cohendet, P., Grandadam, D., & Simon, L. (2010). The anatomy of the creative city. *Industry and Innovation*, *17*(1), 91–111.

Cox, T., & Hurtubia, R. (2016). Vectores de expansión urbana y su interacción con los patrones socioeconómicos existentes en la ciudad de Santiago. *EURE*, *42*(127), 185–207.

Cox, T., & Hurtubia, R. (2019a). *Latent segmentation of urban space through residential location choice. Working Paper*. https://doi.org/10.13140/RG.2.2.32329.77922

Cox, T., & Hurtubia, R. (2019b). *Subdividing the sprawl: Endogenous identification of housing submarkets in expansion areas of Santiago, Chile. Working Paper*. https://doi.org/10.13140/RG.2.2.30587.77603

Coy, M., & Pöhler, M. (2002). Gated communities in Latin American megacities: Case studies in Brazil and Argentina. *Environment and Planning B: Planning and Design*, *29*(3), 355–370.

Crane, R. (2008). Counterpoint: Accessibility and Sprawl. *Journal of Transport and Land Use*, *1*(1), 13–19.

Daniels, C. B. (1975). The influence of Racial Segregation on Housing Prices. *Journal of Urban Economics*, *2*(2), 15–122.

de Mattos, C. a. (1999). Santiago de Chile, globalización y expansión metropolitana: lo que existía sigue existiendo. *EURE (Santiago)*, *25*(76), 43–62.

Diaz Franulic, C. (2017). Migración internacional, envejecimiento poblacional y segunda transición demográfica ¿hacia dónde va Chile? *Notas de Población*, (105).

Dohnke, J., Heinrichs, D., Kabisch, S., Krellenberg, K., & Welz, J. (2015). Achieving a Socio-Spatial Mix? Prospects and Limitations of Social Housing Policy in Santiago de Chile.

*Housing Studies*, *30*(6), 839–857.

Doling, J. (1976). The family life cycle and housing choice. *Urban Studies*, *13*(1), 55–58.

Domencich, T., & McFadden, D. (1975). Urban travel demand: a behavioural approach. Amsterdam: North-Hollan Publishing Co.

Downs, A. (2005). Smart growth: Why we discuss it more than we do it. *Journal of the American Planning Association*, *71*(4), 367–378.

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

Durlauf, S. (2003). Neighborhood Effects. *Observatory*, *4*, 2173–2242.

Ellickson, B. (1981). An alternative test of the hedonic theory of housing markets. *Journal of Urban Economics*, *9*(1), 56–79.

Ettema, D. (2010). The impact of telecommuting on residential relocation and residential preferences. A latent class modeling. *The Journal of Transport and Land Use*, *3*(1), 7–24.

Ewing, R., Bartholomew, K., Winkelman, S., Walters, J., & Chen, D. (1997). *Growing cooler: The evidence on urban development and climate change*. *Renewable Resources Journal* (Vol. 25). Urban Land Institute.

Ewing, R., & Cervero, R. (2010). Travel and the built environment. *Journal of the American*

*Planning Association*, *76*(3), 265–294.

Ewing, R., & Cervero, R. (2017). “Does Compact Development Make People Drive Less?” The Answer Is Yes. *Journal of the American Planning Association*, *83*(1), 19–25.

Ewing, R., & Hamidi, S. (2015). Compactness versus Sprawl: A Review of Recent Evidence from the United States. *Journal of Planning Literature*, *30*(4), 413–432.

Ewing, R., Hamidi, S., Grace, J. B., & Wei, Y. D. (2016). Does urban sprawl hold down upward mobility? *Landscape and Urban Planning, 148*, 80–88.

Ewing, R., & Rong, F. (2008). The impact of urban form on U.S. residential energy use. *Housing Policy Debate*, *19*(1), 1–30.

Florida, R. (2017). The Diversity–Segregation Conundrum. *American Journal of Community Psychology*, *59*(3–4), 272–275.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression. The analysis of spatially varying relationships*. John Wiley & Sons.

Gainza, X., & Livert, F. (2013). Urban form and the environmental impact of commuting in a segregated city, Santiago de Chile. *Environment and Planning B: Planning and Design*, *40*(3), 507–522.

Galster, G. (1996). William Grigsby and the analysis of housing sub-markets and filtering. *Urban Studies*, *33*(10), 1797–1805.

Galster, G. (2013). Neighborhood social mix: Theory, evidence, and implications for policy and planning. *Policy, Planning, and People: Promoting Justice in Urban Development*, (January 2013), 308–336.

García-López, M. Á., & Moreno-Monroy, A. I. (2016). Income Segregation and Urban Spatial Structure: Evidence From Brazil. *CAF-Development Bank of Latin America*.

GFK Adimark. (2000). *El nivel socioeconómico. Manual de aplicación*. Report. Santiago. October.

Glaeser, E. (2011). *Triumph of the city: How urban spaces make us human*. Pan Macmillan.

Glumac, B., Han, Q., & Schaefer, W. (2014). Actors' Preferences in the Redevelopment of Brownfield: Latent Class Model. *Journal of Urban Planning and Development*, *140*(3), 402–408.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking Mechanisms in Human Learning. *Trends in Cognitive Sciences*, *5*(6), 236–243.

Goodman, A., & Thibodeau, T. (1998). Housing Market Segmentation. *Journal of Housing Economics*, *7*(2), 121–143.

Gould, P., & White, R. (1974). *Mental Maps*. England and New York: Penguin.

Greene, M., Mora, R. I., Figueroa, C., Waintrub, N., & Ortúzar, J. de D. (2017). Towards a sustainable city: Applying urban renewal incentives according to the social and urban characteristics of the area. *Habitat International*, *68*, 15–23.

Gross, D. J., Sirmans, C. F., & Benjamin, J. D. (1990). An empirical evaluation of the probabilistic bid-rent model: The case of homogenous households. *Regional Science and Urban Economics*, *20*(1), 103–110.

Guevara, C., & Ben-Akiva, M. (2006). Endogeneity in Residential Location Choice Models. *Transportation Research Record: Journal of the Transportation Research Board*, *1977*(1), 60–66.

Haider, M., & Miller, E. (2004). Modeling Location Choices of Housing Builders in the Greater Toronto Area. *Transportation Research Record: Journal of the Transportation Research Board*, (1898), 148–156.

Hananel, R. (2014). Can centralization, decentralization and welfare go together? The case of Massachusetts Affordable Housing Policy (Ch. 40B). *Urban Studies*, *51*(12), 2487–2502.

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, *25*(2), 73–76.

Heinrichs, D., Nuissl, H., & Rodríguez, Y. C. (2009). Dispersión urbana y nuevos desafíos para la gobernanza. *EURE (Santiago)*, *XXXV*, 29–46.

Hoshino, T. (2011). Estimation and analysis of preference heterogeneity in residential choice behaviour. *Urban Studies (Edinburgh, Scotland)*, *48*(2), 363–382.

Hurtubia, R., & Bierlaire, M. (2014). Estimation of Bid Functions for Location Choice and Price Modeling with a Latent Variable Approach. *Networks and Spatial Economics*, *14*(1), 47–65.

Hurtubia, R., Gallay, O., & Bierlaire, M. (2010). Attributes of Households , Locations and Real-Estate Markets for Land Use Modeling. *(SustainCity Deliverable). Technical Report*. Retrieved from http://transp-or.epfl.ch/documents/technicalReports/sustaincity_WP2_7.pdf

Hurtubia, R., Martinez, F. J., & Bierlaire, M. (2019). A quasi-equilibrium approach for market clearing in land use microsimulations. *Environment and Planning B: Urban Analytics and City Science*, *46*(3), 445–468.

Hwang, S. (2015). Residential Segregation, Housing Submarkets, and Spatial Analysis: St. Louis and Cincinnati as a Case Study. *Housing Policy Debate*, *25*(1), 91–115.

Ibraimovic, T., & Hess, S. (2017). A latent class model of residential choice behaviour and ethnic segregation preferences. *Housing Studies*, *3037*, 1–21.

INE. (2002). *Resultados Censo 2002*. Santiago. Retrieved from http://www.ine.cl/docs/default-source/FAQ/síntesis-de-resultados-censo-2002.pdf?sfvrsn=2

INE. (2018). *Resultados Censo 2017*. Retrieved from http://www.censo2017.cl/wp-content/uploads/2017/12/Presentacion_Resultados_Definitivos_Censo2017.pdf

Jacobs, J. (1961). *The death and life of great American cities*. Vintage Books.

Jacobs, J. (1969). *The Economy of Cities*. New York: Vintage Books.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666.

Jang, M., & Kang, C. D. (2015). Retail accessibility and proximity effects on housing prices in Seoul, Korea: A retail type and housing submarket approach. *Habitat International*, *49*, 516–528.

Jara-Díaz, S. R., & Martínez, F. J. (1999). On the specification of indirect utility and willingness to pay for discrete residential location models. *Journal of Regional Science*, *39*(4), 675–688.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243–18250.

Kamakura, W., & Russell, G. (1989). A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research*, *26*(4), 379–390.

Kempen, R. Van, & Ozüekren, A. S. (1998). Ethnic Segregation in Cities : New Forms and. *Urban Studies*, *35*(10), 1631–1657.

Kim, J. (2016). Achieving mixed income communities through infill? The effect of infill housing on neighborhood income diversity. *Journal of Urban Affairs*, *38*(2), 280–297.

Kupke, V., Rossini, P., & McGreal, S. (2012). Measuring the impact of higher density housing development. *Property Management*, *30*(3), 274–291.

Lee, Y., Circella, G., Mokhtarian, P. L., & Guhathakurta, S. (2019). 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.

Lerman, S., & Kern, C. (1983). Hedonic Theory, Bid Rents, and Willingness-to- Pay: Some Extensions of Ellickson's Results. *Journal of Urban Economics*, *13*(3), 358–363.

LeRoy, S. F., & Sonstelie, J. (1983). Paradise lost and regained: Transportation innovation, income, and residential location. *Journal of Urban Economics*, *13*(1), 67–89.

Leusen, M. Van. (1997). Viewshed and Cost Surface Analysis Using GIS (Cartographic Modelling in a Cell-Based GIS II). *BAR International Series*, (757), 215–224.

Lewis, P. G., & Baldassare, M. (2010). The complexity of public attitudes toward compact development: survey evidence from five states. *Journal of the American Planning Association*, *76*(2), 219–237.

Liao, F. H., Farber, S., & Ewing, R. (2014). Compact development and preference heterogeneity in residential location choice behaviour: A latent class analysis. *Urban Studies*, *52*(2), 1–24.

Lobo, B. (2015). Planning for urban renewal: Public/private initiatives in Bogota, Colombia, 2000--2014. *Journal of Urban Regeneration & Renewal*, *8*(3), 245–259.

Lovejoy, K., Handy, S., & Mokhtarian, P. (2010). Neighborhood satisfaction in suburban versus traditional environments: An evaluation of contributing characteristics in eight California neighborhoods. *Landscape and Urban Planning*, *97*(1), 37–48.

Lu, Z., Southworth, F., Crittenden, J., & Dunhum-Jones, E. (2014). Market potential for smart growth neighbourhoods in the USA: A latent class analysis on heterogeneous preference and choice. *Urban Studies*, *52*(16), 3001–3017.

Lynch, K. (1960). *The image of the city*. MIT press.

Maclaren, V. W. (1996). Urban sustainability reporting. *Journal of the American Planning Association*, *62*(2), 184–202.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297).

Martinez, F. (2000). Towards a land use and transport interaction framework. In K. Hensher, D, Button (Ed.), *Handbook of transport modelling* (Vol. 1, pp. 393–407). Elsevier

Amsterdam.

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

Martinez, F. J. (1992). The bid-choice land-use model : an integrated economic framework. *Environment and Planning A*, *24*(6), 871–885.

Martínez, F. J., & Henríquez, R. (2007). A random bidding and supply land use equilibrium model. *Transportation Research Part B: Methodological*, *41*(6), 632–651.

Massey, D. (2016). Segregation and the perpetuation of disadvantage. *The Oxford Handbook of the Social Science of Poverty*, *16*, 369–405.

Massey, D., & Denton, N. (1988). The Dimensions of Residential Segregation. *Social Forces*, *67*(2), 281–315.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press.

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlqvist, F. Lundqvist, F. Snickars, & J. Weibull (Eds.), *Spatial Interaction Theory and Planning Models* (pp. 75–96). Amsterdam, North Holland.

Meng, L., Taylor, M. A. P., & Scrafton, D. (2016). Combining Latent Class Models and GIS Models for Integrated Transport and Land Use Planning – A Case Study Application. *Urban*

*Policy and Research*, *34*(4), 305–329.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2).

Mills, E. S. (1967). An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review*, *57*, 197–210.

Mills, E. S. (2005). Why do we have urban density controls? *Real Estate Economics*, *33*(3), 571–585.

Mugnano, S., & Palvarini, P. (2013). "Sharing space without hanging together": A case study of social mix policy in Milan. *Cities*, *35*, 417–422.

Muth, R. F. (1969). *Cities and Housing*. Chicago, Illinois: University of Chicago Press.

Muto, S. (2006). Estimation of the bid rent function with the usage decision model. *Journal of Urban Economics*, *60*(1), 33–49.

Myers, D., & Gearin, E. (2001). Current preferences and future demand for denser residential environments. *Housing Policy Debate*, *12*(4), 633–659.

Nasar, J. L. (1990). The Evaluative Image of the City. *Journal of the American Planning Association*, *56*(1), 41–53.

Nematollahi, S., Tiwari, R., & Hedgecock, D. (2016). Desirable Dense Neighbourhoods: An

Environmental Psychological Approach for Understanding Community Resistance to Densification. *Urban Policy and Research*, *34*(2), 132–151.

O´Sullivan, A. (2012). *Urban Economics* (8th ed.). New York: McGraw-Hill/Irwin.

Olaru, D., Smith, B., & Taplin, J. H. E. (2011). Residential location and transit-oriented development in a new rail corridor. *Transportation Research Part A: Policy and Practice*, *45*(3), 219–237.

Oliva, I., Galilea, P., & Hurtubia, R. (2018). Identifying cycling-inducing neighborhoods : A latent class approach. *International Journal of Sustainable Transportation*, *12*(10), 701–713.

Openshaw, S. (1984). *Concepts and techniques in modern geography number 38: the modifiable areal unit problem*. Norwick: Geo Books.

Ortúzar, J. de D., & Willumsen, L. (2011). *Modelling Transport* (4th ed.). New York: John Wiley & Sons.

Páez, A., Long, F., & Farber, S. (2008). Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques. *Urban Studies*, *45*(8), 1565–1581.

Palm, R. (1978). Spatial Segmentation of the Urban Housing Market. *Economic Geography*, *54*(3), 210–221.

Pendall, R., & Carruthers, J. I. (2003). Does density exascerbate income segregation. *Housing Policy Debate*, *14*(4), 541–589.

Perrin, L., & Grant, J. L. (2014). Perspectives on mixing housing types in the suburbs. *Town Planning Review*, *85*(3), 363–385.

PNUD. (2017). *Desiguales. Orígenes, cambios y desafíos de la brecha Social en Chile*. Santiago de Chile: Uqbar.

Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association*, *53*(284), 873–880.

Quillian, L. (2012). Segregation and Poverty Concentration: The Role of Three Segregations. *American Sociological Review*, *77*(3), 354–379.

Rid, W., & Profeta, A. (2011). Stated preferences for sustainable housing development in Germany—a latent class analysis. *Journal of Planning Education and Research*, *31*(1), 26–46.

Romero, H., Vásquez, A., Fuentes, C., Salgado, M., Schmidt, A., & Banzhaf, E. (2012). Assessing urban environmental segregation (UES). the case of Santiago de Chile. *Ecological Indicators*, *23*, 76–87.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure

Competition. *Journal of Political Economy*, *82*(1), 34–55.

Rosmera, N. A., & Lizam, M. (2016). Housing market segmentation and the spatially varying house prices. *Social Sciences (Pakistan)*, *11*(11).

Ruiz-Tagle, J., & López-Morales, E. (2014). El estudio de la segregación residencial en Santiago de Chile: Revisión crítica de algunos problemas metodológicos y conceptuales. *Eure*, *40*(119), 25–48.

Ruiz-Tagle, J., & Romano, S. (2019). Mezcla Social e Integración Urbana: Aproximaciones teóricas y discusiones del caso. *Revista INVI*, *34*(may), 45–69.

Sabatini, F. (2003). La segregación social del espacio en las ciudades de América Latina. *Cuadernos de Instituto de Estudios Urbanos, Universidad Católica de Chile, Serie Azul*, *35*, 59–70.

Sabatini, F. (2015). La ruptura del patrón de segregación y su significado teórico y práctico. In A. G. Aguilar & I. Escamilla (Eds.), *Segregación urbana y espacios de exclusión. Ejemplos de México y América Latina*. (pp. 25–46). Universidad Nacional Autónoma de México.

Sabatini, F., Cáceres, G., & Cerda, J. (2001). Segregación residencial en las principales ciudades chilenas: Tendencias de las tres últimas décadas y posibles cursos de acción. *Eure (Santiago)*, *27*(82), 21–42.

Sabatini, F., & Salcedo, R. (2010). Gated communities and the poor in Santiago, Chile: Functional and symbolic integration in a context of aggressive capitalist colonization of lower-class areas. *Housing Policy Debate*, *18*(3), 577–606.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS ONE*, *8*(7).

Salvini, P., & Miller, E. J. (2005). ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, *5*(2), 217–234.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing “Neighborhood Effects”: Social Processes and New Directions in Research. *Annual Review of Sociology*, *28*(1), 443–478.

Sarrias, M. (2019). Do monetary subjective well-being evaluations vary across space? Comparing continuous and discrete spatial heterogeneity. *Spatial Economic Analysis*, *14*(1), 53–87.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York: WW Norton.

Schirmer, P. M., Van Eggermond, M. a. B., & Axhausen, K. W. (2014). The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*, *7*(2), 3–21.

Schnare, A. B., & Struyk, R. J. (1976). Segmentation in Urban Housing Markets. *Journal of*

*Urban Economics*, *3*(2), 146–166.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

SECTRA. (2015). *Encuesta de movilidad del Gran Santiago 2012*. Retrieved from http://www.sectra.gob.cl/biblioteca/detalle1.asp?mfn=3253

SECTRA. (2016). *Análisis estratégico de proyectos de transporte urbano, etapa V*. Retrieved from http://www.mtt.gob.cl/

Sennett, R. (1997). *Carne y piedra: El cuerpo y la ciudad en la civilización occidental*. Madrid: Alianza.

Servicio de Impuestos Internos. (2014). Catastro de Bienes Raices. Retrieved from https://www4.sii.cl/sismunInternet/?caller=DETALLE_CAT_Y_ROL_COBRO

Sharifi, A. (2017). From Garden City to Eco-urbanism : The quest for sustainable neighborhood development. *Sustainable Cities and Society*, *20*(September 2015), 1–16.

Smith, B., & Olaru, D. (2013). Lifecycle stages and residential location choice in the presence of latent preference heterogeneity. *Environment and Planning A*, *45*(10), 2495–2514.

Stevens, M. R. (2017). Does Compact Development Make People Drive Less? *Journal of the American Planning Association*, *83*(1), 7–18.

Stolarick, K., & Florida, R. (2006). Creativity, connections and innovation: A study of linkages in the Montréal Region. *Environment and Planning A*, *38*(10), 1799–1817.

Suazo, G. (2017). *Actividades en Santiago de Chile en 1990-2015 : Impacto en los tiempos de viaje en la ciudad (Master Thesis)*. Pontificia Universidad Católica de Chile.

Tersteeg, A. K., & Pinkster, F. M. (2016). "Us Up Here and Them Down There": How Design, Management, and Neighborhood Facilities Shape Social Distance in a Mixed-Tenure Housing Development. *Urban Affairs Review*, *52*(5), 751–779.

Transsa Consulting. (2013). *Estudio de inserción en mercado, PDUC en Colina*. Not Published.

Tu, G., Abildtrup, J., & Garcia, S. (2016). Preferences for urban green spaces and peri-urban forests: An analysis of stated residential choices. *Landscape and Urban Planning*, *148*, 120–131.

United Nations. (2018). *Sustainable Development Goals. 11: Sustainable cities and communities*. Retrieved from https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-11-sustainable-cities-and-communities.html Accesed May 22

Vergara Vidal, J. E. (2017). Verticalización. La edificación en altura en la Región Metropolitana de Santiago (1990-2014). *Revista INVI*, *32*(90), 9–49.

Waddell, P. (2006). *Reconciling Household Residential Location Choices and*

*Neighborhood Dynamics*. *Working Paper*. Retrieved from http://www.urbansim.org/pub/Research/ResearchPapers/smr-urbansim.pdf

Waintrub, N., Greene, M., & Ortúzar, J. de D. (2016). Designing incentive packages for increased density and social inclusion in the neighbourhood of mass transit stations. *Habitat International*, *55*, 133–147.

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

Watkins, C. A. (2001). The definition and identification of housing submarkets. *Environment and Planning A*, *33*(12), 2235–2253.

Webster, C., Glasze, G., & Frantz, K. (2002). The global spread of gated communities. *Environment and Planning B: Planning and Design*, *29*(3), 315–320.

Wegener, M. (2011). The IRPUD model. *Spiekermann & Wegener in Dortmund. Available Online: Http://Www. Spiekermann-Wegener. Com/Mod/Pdf/AP_1101_IRPUD_Model. Pdf (Accessed on 1 December 2011)*.

Wertheimer, M., & Riezler, K. (1944). Gestalt Theory. *Social Research*, *11*, 77–99.

Wright, E. O. (2000). *Class Counts*. Cambridge University Press.

Zöllig, C., & Axhausen, K. W. (2015). A conceptual, agent-based model of land development for UrbanSim. In M. Bierlare, A. De Palma, R. Hurtubia, & P. Waddell (Eds.), *Integrated Transport and Land Use Modeling for Sustainable Cities*. Lausanne: EPFL Press.

# 8. ANNEX

**Table 7.1:** Estimation parameter of the exogenous zones model

| **EXOGENOUS ZONES MODEL** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Observations | | 17830 | | | | | | |
| Null model log-likelihood | | -258759 | | | | | | |
| Final log-likelihood | | -76511 | | | | | | |
| Attribute | Household Type | | | | | | | |
| | | Zone 1 | Zone 2 | Zone 3 | Zone 4 | Zone 5 | Zone 6 | Zone 7 |
| Accessibility to Commerce by transit | Low-EL | -0.000337 (-1.46)* | 0.000264 (1.1)* | -0.000766 (-1.63)* | -0.000837 (-1.49)* | 0.000475 (2.32) | 0.000804 (3.09) | 0.000885 (0.62)* |
| | Mid-EL | -0.000476 (-2.15) | -0.00105 (-4.51) | -0.0000457 (-0.21)* | 0.00165 (4.93) | 0.000754 (3.53) | 0.000798 (3.85) | 0.0026 (1.64)* |
| | Hi-EL | | | | | | | |
| Accessibility to Industry by transit | Low-EL | | | | | | | |
| | Mid-EL | | | | | | | |
| | Hi-EL | | | | | | | |
| Accessibility to Commerce by car | Low-EL | | | | | | | |
| | Mid-EL | | | | | | | |
| | Hi-EL | -0.00143 (-4.51) | -0.00129 (-4.68) | 0.00119 (6.63) | 0.00203 (5.08) | -0.00234 (-7.5) | -0.000322 (-1.21)* | -0.00222 (-2.79) |
| Accessibility to Industry by car | Low-EL | | | | | | | |
| | Mid-EL | | | | | | | |
| | Hi-EL | | | | | | | |
| Distance to nearest subway estation | Low-EL | -0.151 (-4.04) | 0.18 (1.09)* | -0.154 (-0.71)* | 2.03 (1.51)* | -0.148 (-2.04) | -0.165 (-1.18)* | 0.00224 (0.11)* |
| | Mid-EL | 0.0322 (1.07)* | -0.879 (-6) | -0.848 (-5.28) | 1.1 (1.23)* | -0.116 (-1.54)* | -0.765 (-5.84) | 0.0438 (1.85)* |
| | Hi-EL | -0.355 (-5.26) | -0.719 (-3.05) | -1.52 (-11.31) | -3.11 (-2.48) | -0.304 (-2.15) | -1.62 (-8.3) | 0.0745 (1.63)* |
| Average Built surface in zone | Low-EL | 0.00394 (0.4)* | -0.000768 (-0.04)* | 0.0858 (2.44) | 0.014 (0.41)* | -0.0258 (-1.68)* | -0.0258 (-1.68)* | 0.0267 (1.5)* |
| | Mid-EL | 0.0447 (6.15) | 0.0655 (3.6) | 0.225 (18.53) | 0.0272 (1.4)* | 0.0679 (4.28) | 0.0402 (2.71) | -0.0203 (-1)* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Hi-EL | 0.0468 (2.69) | 0.0867 (2.69) | 0.244 (28.63) | -0.0675 (-2.1) | 0.163 (5.35) | 0.125 (6.17) | 0.11 (3.54) |
| Average Zonal Income | Low-EL | 1.03 (0.95)* | -3.05 (-2.44) | -0.524 (-0.37)* | 1.67 (0.58)* | -0.153 (-0.1)* | -4.21 (-3.26) | -3.5 (-1.86)* |
| | Mid-EL | 9.2 (10.83) | 17.8 (16.66) | 1.36 (2.52) | 0.398 (0.22)* | 7.71 (5.18) | 9.8 (12.86) | 5.66 (2.8) |
| | Hi-EL | 24.9 (14.45) | 25.1 (11) | 5.83 (15.94) | 10.4 (4.65) | 28.8 (14.81) | 16.3 (18.78) | 10.3 (3.45) |
| Built Density in Zone | Low-EL | 0.203 (0.13)* | -0.207 (-0.17)* | -2.79 (-1.64)* | -1.42 (-1.42)* | -1.92 (-0.96)* | -5.06 (-2.79) | -0.716 (-0.12)* |
| | Mid-EL | 4.88 (3.63) | -0.737 (-0.63)* | 0.539 (0.91)* | 0.496 (1.21)* | -8.71 (-4.19) | 0.435 (0.32)* | 14.3 (1.99) |
| | Hi-EL | 5.3 (1.82)* | 0.971 (0.38)* | 0.298 (0.82)* | 0.441 (0.95)* | -6.23 (-1.5)* | -0.547 (-0.22)* | -34.1 (-2.26) |
| Household constant | Low-EL | 8.4 (16.32) | 8.4 (16.32) | 8.4 (16.32) | 8.4 (16.32) | 8.4 (16.32) | 8.4 (16.32) | 8.4 (16.32) |
| | Mid-EL | 0.64 (1.54)* | 0.64 (1.54)* | 0.64 (1.54)* | 0.64 (1.54)* | 0.64 (1.54)* | 0.64 (1.54)* | 0.64 (1.54)* |
| | Hi-EL | -9.34 (-6.99) | -9.34 (-6.99) | -9.34 (-6.99) | -9.34 (-6.99) | -9.34 (-6.99) | -9.34 (-6.99) | -9.34 (-6.99) |
| $\mu_1$ | | 0.168 (169.54) | | | | | | |

**Table 7.2**: Estimation parameters of the seven clusters model.

| **ATTRIBUTE-BASED 7 CLUSTER MODEL** | |
|---|---|
| Observations | 17830 |
| Null model log-likelihood | -258759 |
| Final log-likelihood | -75556 |

| Attribute | Household Type | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 7 |
| Accessibility to Commerce by transit | Low-EL | 0.00198 (1.02)* | -0.000495 (-0.82)* | 0.000771 (0.78)* | 0.00291 (0.45)* | 0.00028 (0.84)* | 0.0021 (0.38)* | -0.00882 (-1.15)* |
| | Mid-EL | -0.00276 (-2.54) | 0.00268 (6.13) | 0.0017 (1.72)* | -0.000372 (-0.13)* | 0.00111 (3.16) | -0.00653 (-3.04) | 0.00839 (1.74)* |
| | Hi-EL | 0.000356 (0.25)* | 0.00436 (7.18) | 0.00456 (1.84)* | -0.00277 (-1.21)* | 0.00212 (2.37) | -0.00466 (-3.19) | 0.00975 (4.11) |
| Accessibility to Industry by transit | Low-EL | -0.00401 (-1.76)* | 0.00092 (1.22)* | 0.000858 (0.82)* | -0.00639 (-0.98)* | -0.000119 (-0.26)* | -0.00168 (-0.21)* | 0.0238 (1.64)* |
| | Mid-EL | 0.00253 (1.72)* | -0.00425 (-7.66) | -0.00209 (-1.94)* | -0.0012 (-0.45)* | -0.000528 (-1.1)* | 0.00985 (3.34) | -0.0207 (-2.32) |
| | Hi-EL | -0.000514 (-0.24)* | -0.00622 (-7.76) | -0.00497 (-1.81)* | 0.000268 (0.11)* | -0.000639 (-0.53)* | 0.00656 (3.26) | -0.0194 (-4.4) |
| Accessibility to Commerce by car | Low-EL | -0.00225 (-1.05)* | 0.0018 (2.6) | 0.00109 (1.17)* | -0.0104 (-0.93)* | 0.000284 (0.71)* | 0.000177 (0.02)* | 0.029 (2.51) |
| | Mid-EL | 0.00684 (5) | -0.00285 (-5.85) | -0.00199 (-2.18) | 0.00413 (0.87)* | -0.00315 (-7.65) | 0.0224 (6.56) | 0.00411 (0.64)* |
| | Hi-EL | 0.0116 (5.56) | -0.000168 (-0.24)* | -0.00533 (-2.24) | 0.0191 (4.91) | -0.00579 (-5.48) | 0.0297 (12.83) | 0.012 (4) |
| Accessibility to Industry by car | Low-EL | 0.00359 (1.45)* | -0.00214 (-2.8) | -0.00157 (-1.54)* | 0.0116 (0.95)* | 0.000148 (0.3)* | -0.000692 (-0.06)* | -0.037 (-2.46) |
| | Mid-EL | -0.00844 (-4.97) | 0.00457 (8.63) | 0.0029 (2.87) | -0.00285 (-0.54)* | 0.00278 (5.57) | -0.0257 (-6.27) | 0.00183 (0.22)* |
| | Hi-EL | -0.0144 (-5.62) | 0.00217 (2.63) | 0.00653 (2.5) | -0.0179 (-3.93) | 0.00435 (3.48) | -0.0323 (-11.67) | -0.00475 (-1.22)* |
| Distance to nearest subway estation | Low-EL | 0.21 (0.16)* | 0.0308 (0.3)* | 0.00572 (0.28)* | 6.41 (1.38)* | 0.0484 (0.77)* | 3.04 (1.44)* | 1.56 (1.92)* |
| | Mid-EL | 1.5 (1.55)* | -0.12 (-1.76)* | 0.0593 (2.66) | -0.429 (-0.2)* | -0.0551 (-0.84)* | 1.3 (1.77)* | 0.338 (0.73)* |
| | Hi-EL | 4.68 (3.54) | 0.0818 (0.88)* | 0.0545 (1.09)* | -2.26 (-1.4)* | -0.444 (-2.35) | 3.05 (6.74) | 1.46 (6.91) |
| | Low-EL | -0.285 (-0.33)* | -0.361 (-2.27) | 0.0214 (0.33)* | 7.79 (2.23) | -0.0726 (-0.71)* | -2.33 (-1.94)* | -0.116 (-0.08)* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Distance to nearest highway exit | Mid-EL | 1.18 (2.2) | 0.292 (2.75) | -0.0681 (-0.94)* | 1.74 (1.16)* | -0.0841 (-0.77)* | -1.37 (-2.89) | -2.26 (-4.21) |
| | Hi-EL | -1.53 (-2.31) | 0.113 (0.79)* | -0.265 (-1.61)* | 0.163 (0.16)* | -0.244 (-0.83)* | -2.51 (-7.55) | -2.31 (-9.51) |
| Average Built surface in zone | Low-EL | 0.00786 (0.19)* | 0.00036 (0.02)* | 0.0035 (0.51)* | 0.301 (2.64) | -0.0027 (-0.22)* | -0.0027 (-0.22)* | 0.0472 (0.67)* |
| | Mid-EL | 0.141 (6.41) | 0.0581 (5.4) | 0.00533 (0.7)* | 0.186 (4.1) | 0.131 (10.39) | 0.0957 (3.78) | 0.0458 (1.31)* |
| | Hi-EL | 0.161 (6.18) | 0.109 (9.16) | 0.0163 (0.98)* | 0.138 (3.86) | 0.242 (7.58) | 0.0578 (3.68) | 0.0265 (1.99) |
| Average Zonal Income | Low-EL | -0.834 (-0.18)* | 1.93 (1.09)* | 3.04 (1.88)* | -13.1 (-1.24)* | 0.132 (0.11)* | 2.64 (0.43)* | -1.85 (-0.46)* |
| | Mid-EL | 2.92 (1.21)* | 9.95 (9.3) | 14.7 (7.96) | -1.06 (-0.26)* | 19.5 (15.8) | 2.51 (1.22)* | 1.35 (1.01)* |
| | Hi-EL | 14.1 (5.25) | 18.9 (16.2) | 29.8 (6.68) | 7.6 (2.35) | 33.4 (10.56) | 9.39 (8.32) | 3.17 (6.38) |
| Built Density in Zone | Low-EL | 0.0226 (0.01)* | 2.84 (1.13)* | -5.51 (-1.76)* | 0.699 (0.48)* | -3.34 (-2.23) | -6.47 (-1.51)* | -21 (-1.09)* |
| | Mid-EL | 10.5 (5.51) | 2.85 (1.74)* | 11.4 (3.74) | 0.318 (0.55)* | -2.04 (-1.29)* | -4.63 (-3.2) | -8.89 (-1.26)* |
| | Hi-EL | 12.1 (6.44) | -4.7 (-2.24) | -16.8 (-2.37) | -0.829 (-1.38)* | -3.53 (-0.86)* | -6.3 (-7.4) | -26.2 (-8.58) |
| Household constant | Low-EL | 5.72 (6.24) | 5.72 (6.24) | 5.72 (6.24) | 5.72 (6.24) | 5.72 (6.24) | 5.72 (6.24) | 5.72 (6.24) |
| | Mid-EL | -4.91 (-5.59) | -4.91 (-5.59) | -4.91 (-5.59) | -4.91 (-5.59) | -4.91 (-5.59) | -4.91 (-5.59) | -4.91 (-5.59) |
| | Hi-EL | -21.3 (-13.44) | -21.3 (-13.44) | -21.3 (-13.44) | -21.3 (-13.44) | -21.3 (-13.44) | -21.3 (-13.44) | -21.3 (-13.44) |
| $\mu_1$ | | 0.175 (167.84) | | | | | | |

**Table 7.3**: Estimation parameters of the two clusters model.

| **ATTRIBUTE-BASED 2 CLUSTER MODEL** | | | |
|---|---|---|---|
| Observations | | 17830 | |
| Null model log-likelihood | | -258759 | |
| Final log-likelihood | | -76170 | |
| Attribute | Household Type | | |
| | | Cluster 1 | Cluster 2 |
| Accessibility to Commerce by transit | Low-EL | 0.000447 (1.22)* | -0.0022 (-0.89)* |
| | Mid-EL | -0.00263 (-7.59) | -0.000142 (-0.11)* |
| | Hi-EL | -0.0065 (-10.01) | -0.000636 (-0.51)* |
| Accessibility to Industry by transit | Low-EL | -0.0000768 (-0.3)* | 0.00187 (1.02)* |
| | Mid-EL | 0.00252 (10.4) | -0.00115 (-1.2)* |
| | Hi-EL | 0.00518 (11.44) | -0.00171 (-1.94)* |
| Accessibility to Commerce by car | Low-EL | 0.00103 (3.36) | -0.00137 (-0.48)* |
| | Mid-EL | -0.00279 (-10) | 0.0104 (6.56) |
| | Hi-EL | -0.000588 (-1.14)* | 0.0251 (17) |
| Accessibility to Industry by car | Low-EL | -0.000896 (-2.35) | 0.000997 (0.28)* |
| | Mid-EL | 0.00354 (9.97) | -0.0101 (-5.15) |
| | Hi-EL | 0.00232 (3.4) | -0.0239 (-13.26) |
| Distance to nearest subway estation | Low-EL | 0.0194 (1.1)* | 0.178 (0.26)* |
| | Mid-EL | 0.0594 (3.34) | 0.193 (0.63)* |
| | Hi-EL | 0.155 (4.2) | 0.798 (3.86) |
| Distance to nearest highway exit | Low-EL | | |
| | Mid-EL | -0.028 (-0.59)* | -1.26 (-3.91) |
| | Hi-EL | 0.106 (1.17)* | -2.34 (-10.87) |
| Average Built surface in zone | Low-EL | -2.31 (-3.14) | -0.87 (-0.93)* |
| | Mid-EL | 2.72 (4.25) | 0.175 (0.45)* |
| | Hi-EL | 1.9 (1.52)* | -1.3 (-3.79) |
| Average Zonal Income | Low-EL | 0.0279 (0.1)* | -0.00536 (-0.01)* |
| | Mid-EL | 14.5 (34.11) | 0.786 (0.97)* |
| | Hi-EL | 24.7 (43.52) | 5.66 (12.93) |
| Built Density in Zone | Low-EL | -2.31 (-3.14) | -0.87 (-0.93)* |
| | Mid-EL | 2.72 (4.25) | 0.175 (0.45)* |
| | Hi-EL | 1.9 (1.52)* | -1.3 (-3.79) |
| Household constant | Low-EL | 5.75 (7.53) | |
| | Mid-EL | -5.21 (-6.65) | |
| | Hi-EL | -28.2 (-19.99) | |
| $\mu_1$ | | 0.171 (168.87) | |

**Annex 7.5: Bid elasticities**

We present the derivation of the "bid elasticity", which is a variation on the elasticity for choice models. As explained in the body of chapter four, traditional choice elasticity accounts for variations in the probability of choosing an alternative with respect to a variation in a particular attribute of that alternative.

For this case, the particular attribute is part of the utility function (or WP in bid-auction models) of <u>all the alternatives</u>, with a different parameter in each alternative. This implies that the elasticity is not only direct, but also crossed with the probability of all other alternatives.

The elasticity $E_{P(h|S),\, z_i^k}$ of the location probability $P(h|i)$, of household $h$ in location $i$, with respect to a location attribute $z_i^k$, is:

$$E_{P(h|S),\, z_i^k} = \frac{\partial P(h|i)}{\partial z_i^k} \cdot \frac{z_i^k}{P(h|i)} \tag{7.1}$$

Where $P(h|i)$ is the probability of a household $h$ being the best bidder for location $i$, defined by:

$$P(h|i) = \frac{\exp(WP_{hi})}{\sum_{g \in H} \exp(WP_{hi})} \tag{7.2}$$

The derivative of $P(h|i)$ with respect to $z_i^k$ is (using the quotient rule for deriving divisions):

$$\frac{\partial P(h|i)}{\partial z_i^k} = \frac{\exp(WP_{hi}) \cdot \beta_h^k \cdot \sum_{g\in H} \exp(WP_{hi}) - \exp(WP_{hi}) \cdot \sum_{g\in H}\left[\exp(WP_{hi}) \cdot \beta_g^k\right]}{\left[\sum_{g\in H} \exp(WP_{hi})\right]^2} \tag{7.3}$$

Then, from (7.1) and (7.3), bid elasticity is:

$$E_{P(h|s),\, z_i^k} = \frac{\exp(WP_{hi}) \cdot \beta_h^k \cdot \sum_{g\in H} \exp(WP_{hi}) - \exp(WP_{hi}) \cdot \sum_{g\in H}\left[\exp(WP_{hi}) \cdot \beta_g^k\right]}{\left[\sum_{g\in H} \exp(WP_{hi})\right]^2} \cdot \frac{z_i^k}{\frac{\exp(WP_{hi})}{\sum_{g\in H} \exp(WP_{hi})}} \tag{7.4}$$

Simplifying with some algebra:

$$E_{P(h|s),\, z_i^k} = \frac{\beta_h^k \cdot \sum_{g\in H} \exp(WP_{hi}) - \sum_{g\in H}\left[\exp(WP_{hi}) \cdot \beta_g^k\right]}{\sum_{g\in H} \exp(WP_{hi})} \cdot z_i^k \tag{7.5}$$

$$E_{P(h|s),\, z_i^k} = \left[\frac{\beta_h^k \cdot \sum_{g\in H} \exp(WP_{hi})}{\sum_{g\in H} \exp(WP_{hi})} - \frac{\sum_{g\in H}\left[\exp(WP_{hi}) \cdot \beta_g^k\right]}{\sum_{g\in H} \exp(WP_{hi})}\right] \cdot z_i^k \tag{7.6}$$

$$E_{P(h|s),\, z_i^k} = \left[\beta_h^k - \frac{\sum_{g\in H}\left[\exp(WP_{hi}) \cdot \beta_g^k\right]}{\sum_{g\in H} \exp(WP_{hi})}\right] \cdot z_i^k \tag{7.7}$$

$$E_{P(h|s),\, z_i^k} = \beta_h^k \cdot z_i^k - \sum_{g\in H}\left[P(g|i) \cdot \beta_g^k \cdot z_i^k\right] \tag{7.8}$$

$$E_{P(h|s), z_i^k} = \beta_h^k \cdot z_i^k - \sum_{g \neq h} \left[ P(g|i) \cdot \beta_g^k \cdot z_i^k \right] + P(h|i) \cdot \beta_g^k \cdot z_i^k \tag{7.9}$$

$$E_{P(h|s), z_i^k} = \beta_h^k \cdot z_i^k \cdot \left(1 - P(h|i)\right) - \sum_{g \neq h} \left[ P(g|i) \cdot \beta_g^k \cdot z_i^k \right] \tag{7.10}$$

Finally, the formula for the bid elasticity is composed by the "traditional" direct choice elasticity, plus the sum of the cross elasticities for all the other households.