SUBDIVIDING THE SPRAWL: ENDOGENOUS IDENTIFICATION OF HOUSING SUBMARKETS IN EXPANSION AREAS OF SANTIAGO, CHILE

Preprint · January 2019

DOI: 10.13140/RG.2.2.30587.77603

2 authors:

Ricardo Hurtubia
Pontificia Universidad Católica de Chile

Tomás Cox
University of Chile

Working Paper

SUBDIVIDING THE SPRAWL: ENDOGENOUS IDENTIFICATION OF HOUSING SUBMARKETS IN EXPANSION AREAS OF SANTIAGO, CHILE

Tomás Cox [a,b]. tcox2@uc.cl
Ricardo Hurtubia [a,c]. rhurtubia@uc.cl

(a) Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile.
(b) Department of Urbanism, Universidad de Chile.
(c) School of Architecture, Pontificia Universidad Católica de Chile.

ABSTRACT

Urban sprawl is a phenomenon observed in most cities around the globe and especially in Latin America, it appears to be associated to socioeconomic segregation. The housing industry usually target specific segments of the population, generating projects that can accentuate the segregation process. In his paper we build a model to identify housing submarkets as a function of projects characteristics and location choice patterns. The modelling framework is based on discrete-choice, latent-class models, allowing to jointly identify market segmentations, and their particular location choice preferences, without the need of arbitrary 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. Each submarket behaves differently in terms of location differences, reproducing the socioeconomic segregation already observed in the consolidated city and the Chilean society.

Key words**:** Location Choice Model, Housing Submarkets, Latent Classes.

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 et al., 2007). 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, 2003a). This fragmentation is expressed in the shift from a continuous development of the city to the emergence of a system of so called "globalization artifacts", such as private mega-residential projects, shopping malls, airports, and business/light-industrial parks, all linked through highway networks (Borsdorf et al., 2007). 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. 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. Sprawl in Santiago has been based mostly on large residential real estate projects, which are the focus of this research.

This real estate development pattern seems to be consistent with the existence of housing submarkets (Palm, 1978; Schnare and 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 and Denton, 1988).

This paper proposes a model to understand location choice patterns of residential projects and their membership to different housing submarket. The modelling approach, based on latent class models (Kamakura and Russell, 1989) and location choice models (McFadden, 1978), allows identifying 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 and Lizam, 2016; Schnare and Struyk, 1976).

Unlike most models found on the housing segmentation literature, the objective of the proposed approach is not only to understand price formation, but to understand location preferences of different developers who are trying to satisfy a heterogeneous, but subdividable, housing demand. For this, we assume developers are rational agents, who attempt to maximize profits by choosing the location where the difference between expected selling price and development costs (land and construction) is larger. The expectation of the selling price is a function of location attributes, but this function will be different depending on the market segment (or submarket) they target. Since submarkets are not directly observable, a latent class model (LCM) approach is taken, where the characteristics of the project define its probability of being demanded by each market segment and, consequently, defines the class or "latent submarket" of the project.

The model presented in this paper belongs to a larger family of location choice models (Antoniou and Picard, 2015; Pagliara et al., 2010) which are a key component of Integrated Land Use and Transport (LUTI) Models (Wegener, 2014), an increasingly used tool for strategic transport and urban planning, and hence could also be a contribution to this field. To our knowledge, the model presented here is the first 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 and Li, 2007).

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 2014 (accounting for 1,833 projects and 89,422 units). Estimation results confirm a very clear market segmentation, with significantly different 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.

The paper is structured as follows. Section 2 provides an overview of the literature in the field of housing submarkets, location choice and agent heterogeneity; section 3 presents the proposed model; section 4 presents the model implementation, introducing Santiago as a case study, describing the data assembly and showing the estimation results. Section 5 concludes the paper.

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 good (housing units) that usually have very high transaction costs. As most markets, it can be subdivided into submarkets, but there are some key differences that are relevant to this work.

First, housing is a product that is consumed by the whole society, therefore demand (and, in consequence, supply) is heterogeneous. Second, housing is a product that has a unique and fixed location in space, which derives in the consolidation of neighborhoods with defined characteristics (e.g. Schelling, 1978; Galster, 2001). Due to spatial dependence (Anselin, 2013), these neighborhoods tend to show internal homogeneity in terms of dwelling characteristics and sociodemographic variables. Finally, high production costs and economies of scale have consolidated a project-based industry, in which the production of new housing is done mainly in batches of nearly similar units. However, these batches are usually customized (in terms of location and physical design) to serve different segments of the market demand.

Therefore, 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, called housing submarkets.

2.1. Empirical housing submarkets

Segmentation of housing markets requires understanding how dwelling and location attributes distribute within each segment, with selling price being one of the most relevant. Empirical studies on the relation between housing attributes and selling prices are usually based on hedonic price models (Rosen, 1974). Earlier studies assumed that price behavior could be defined by a single hedonic function for all dwellings across a city. This meant that the effect in price for the increase of an attribute (e.g. built surface) is constant across the city and across the different types of units.

Instead of a single hedonic function, different studies (Schnare and Struyk, 1976; Adair, Berry and McGreal, 1996; Galster, 1996; Goodman and Thibodeau, 1998; Watkins, 2001; Rosmera and Lizam, 2016, *inter alia*) have shown that multiple housing submarkets coexist within a city. This has been achieved by running linear regressions within subsets of similar housing units, obtaining different (and significant) parameter estimates for each subset, while reaching a better fit (R^2) than an aggregated model.

The previously cited literature shows consensus on the existence of housing submarkets, but how to identify them is a matter of ongoing discussion. Two main types of attributes can be used: spatial or structural. Spatial segmentation refers to using different city zones as different submarkets. Structural segmentation refers to using dwelling characteristics such as quality, space or type (detached, etc.) to identify submarkets. Exogenous methods have been used to identify segments according to predefined attributes, such as administrative areas or housing types (Adair et al., 1996; Palm, 1978; Schnare and Struyk, 1976) but, starting from Bourassa *et al.* (1999), statistical methods such as principal component, factor analysis, and clustering techniques have been used to identify submarkets. These methods identify groups of dwellings by seeking for minimum internal variance and maximizing variance between groups, in terms of housing characteristics.

The approach based on hedonic prices, while providing highlights of market characteristics and information of average marginal valuation of attributes, is limited in the regard of understanding demand in a more disaggregate way. For this purpose, discrete location-choice models seem to be an adequate tool.

2.2. Behavior-based modelling of markets: Location choice models

Location choice models are based on the seminal work of McFadden (1978), who proposed that different type of agents h (households, which can be divided by income or other relevant feature) face a set L of locations i (dwellings) as discrete alternatives, each one reporting a utility U based on a vector of location attributes Z_i, a selling price r_i, and a vector of preference parameters β_h:

$$U_i^h = V(Z_{i,}, r_i, \beta_h) + \varepsilon_{hi} \quad (1)$$

where ε is an error term accounting for unobserved attributes of the location or idiosyncratic behavior of the decision maker. This, according to Random Utility Theory (McFadden, 1973), allows to estimate the location choice probabilities. If the error term follows a IID Gumbell distribution, the probability of household h choosing location i is:

$$P_i^h = \frac{\exp\left(\mu \cdot V(Z_{i,}, r_i, \beta_h)\right)}{\sum_{j \in L} \exp\left(\mu \cdot V(Z_{j,}, r_j, \beta_h)\right)} \tag{2}$$

The scale parameter μ cannot be identified with this specification and, therefore, is usually assumed to be equal to one. For simplicity we omit it in all remaining equations.

In the mathematical formulation of these models, the location attributes are generally an undetermined vector of variables to be specified later in the implementation stage. These attributes can be divided into accessibility, zonal (densities, amenities, etc.) and unit attributes such as built space or selling price (Hurtubia et al., 2010). This approach has been implemented in operational LUTI models like UrbanSim (Waddell, 2011) and MUSSA (Martínez and Donoso, 1996, 2010), mainly to predict how different agents locate in the city and how their preferences change due to transport projects, subsidies/taxes, and/or changes in land use regulation. These models not only characterize how different household or firms value urban amenities (which is important for policy design), but also can simulate future growth of the city and, therefore, demand for infrastructure.

In this formulation for housing demand, price is assumed to be exogenous. When modelling housing supply (developers' location decisions), location depends on an expected selling price, which is not exogenous and can be calculated as a function of project characteristics and location attributes. Location of housing supply can be modeled using a mathematical structure similar to the one used for housing demand, but some differences apply. Developers are profit maximizers instead of utility maximizers (they will choose the location which reports maximum profit), and their profit is the difference between the dwellings' expected selling price R_{in} and development costs C (mainly, land and construction costs). As the selling price is a function of demand, most operational models have treated supply as an answer to demand. Therefore, profit π_i^n for a given project n in a location i can be written as:

$$\pi_i^n = R_{in}(\beta_n, Z_i, X_n) - C(X_n, R_i) + \xi_{ni} \tag{3}$$

where Z_i is a vector of location attributes, X_n is a vector of project characteristics, β_n is a vector of marginal attribute valuation and R_i is the land cost. From equation (3), and assuming error term ξ_{ni} follows a IID Gumbel distribution, the probability of project n being built at location i can be written as in equation (2), but replacing the utility V with the profit function π. This formulation is also common in operational LUTI models such as TRANUS (De la Barra, 1989), MUSSA (Martínez and Donoso, 1996, 2010) and UrbanSim (Waddell, 2011).

2.3. Addressing submarkets through heterogeneity in location choice models.

The idea that a single aggregate function can´t describe the behavior of housing prices across a city (and therefore it must be modelled in terms of submarkets), is analogous to what is known as "heterogeneity" in location choice models, which is the explicit differentiation of preference parameters by type of agent. This differentiation is usually defined exogenously using decision maker characteristics, such as income, car ownership and households' size. For a review of residential location choice models following this approach see Schirmer, Van Eggermond, & Axhausen (2014). Models for the location of residential supply considering heterogeneity of the developers can be found in Haider and Miller (2004) and Zöllig and Axhausen (2015).

Exogenous definition of types of agents (and hence heterogeneity) does not 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 and Russel, 1989). These models estimate the probability to belong 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 3.

Latent class models have been widely used to model heterogeneity in residential location choice characteristics (Ettema, 2010; Liao et al., 2014; Lu et al., 2014; Olaru et al., 2011; Walker and Li, 2007).

Latent class models applied to the problem of location choice for residential supply are not reported in the literature, to the extent of the authors' knowledge. The model presented next aims at filling this gap.

3. A MODEL FOR ENDOGENOUS SEGMENTATION OF HOUSING SUBMARKETS

We propose a model where 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. Therefore, developers attempt to maximize their profit by choosing the best location for each project, depending on the submarket they belong 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.

We model the profit of a project n belonging to a submarket s following equation (3), but decomposing the cost between development costs and land acquisition.

$$\max_{i \in L} \pi_n(i|s) = R_{nis}(Z_i, X_n) - D_n(X_n) - L_i \cdot q_n \tag{4}$$

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 function of project characteristics. 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). We assume that profit of the location decision for each project is independent from the other projects' location decisions.

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 ρ_s and β_s, which are submarket-specific, represent the marginal price of dwelling characteristics and location attributes respectively.

$$R_{nis} = \beta_s \cdot Z_i + \rho_s \cdot X_n \tag{5}$$

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 (Domenich and McFadden, 1975; McFadden, 1978), we can account for these uncertainties if we assume that the profit of each location alternative has a random error (as seen in equation 3) following an IID Gumbel distribution, and treating the decision process under an stochastic approach. Then, 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:

$$P_n(i|s) = \frac{\exp\left(N_n \cdot \pi_n(i|s)\right)}{\sum_{j=1}^{N} \exp\left(N_n \cdot \pi_n(j|s)\right)} \quad \forall\, n, i, s \tag{6}$$

Where N_n is the number of units within project n. If we replace (4) and (5) in (6), the location choice probability is:

$$P_n(i|s) = \frac{\exp\left(N_n \cdot (\beta_s \cdot Z_i + \rho_s \cdot X_n - D_n(X_n) - L_i \cdot q_n)\right)}{\sum_{j=1}^{N} \exp\left(N_n \cdot \left(\beta_s \cdot Z_j + \rho_s \cdot X_n - D_n(X_n) - L_j \cdot q_n\right)\right)} \quad \forall\, n, i, s \tag{7}$$

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:

$$P_n(i|s) = \frac{\exp(\beta_s \cdot Z_i - L_i \cdot q_n)}{\sum_{j=1}^{N} \exp\left(\beta_s \cdot Z_j - L_j \cdot q_n\right)} \quad \forall\, n, i, s \tag{8}$$

This location probability is conditional to the 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 $W_{ns}(X_n, \theta_s)$, where project characteristics (X_n) and membership parameters (θ_s) define how much a project n fits into a submarket s. Making similar assumptions about unobserved attributes and stochastic behavior as in (6), the probability of a project n belonging to submarket s is:

$$P_n(s \mid X_n) = \frac{exp(W_{ns}(X_n, \theta_s))}{\sum_{g \in S} exp(W_{ng}(X_n, \theta_g)} \quad \forall\, s, n \tag{9}$$

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:

$$P_n(i) = \sum_s P_n(i|s) \cdot P_n(s|X_n) \qquad \forall\, i, n \tag{10}$$

Using equation (10), a maximum likelihood estimation of parameters β_s and θ_s can be estimated from observations of 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.

4. SANTIAGO CASE STUDY: PROJECT-BASED EXPANSION

In order to explore the properties of the model, we propose as a case study the development of residential projects in the expansion areas of Santiago, Chile. With 6,123,000 habitants (Instituto Nacional de Estadísticas, 2018), Santiago is by a large extent the main city of Chile, concentrating administrative power, services and commerce. Unlike most of other Chilean cities (somewhat compact with well-defined sectors), Santiago has evolved from a traditional compact city to a fragmented and globalized city (Borsdorf, 2003b).

Since the mid-1970s' economic reform, market-oriented policies opened the real estate industry for capital investment and stimulated the deregulation of land markets in Chile. According to the World Bank, GDP *per capita* in Chile has raised from US$954 in 1970 to US$13,793 in 2016, which correlates with higher levels of car ownership and access to mortgage credit. In the mid-1990s, private investment in transport infrastructure was incorporated by a special legal frame of concessions, which promoted a highway boom all along the country and especially in Santiago[1], connecting expansion areas to the historic city center and the CBD.

This improvement in connectivity and accessibility, together with the 1997 modification to the Metropolitan Plan for the city, triggered a significant shift in location patterns in Santiago and its surroundings. Within the regulation changes, a policy called "conditioned urban planning zoning[2]" allowed an intensive development of expansion areas. Figure 1 shows the "centrifugal "evolution of the location of new real estate projects in the outskirts of the city from 2004 to 2014, and how this correlates to the construction of urban highways.

While the conditioned urban planning policy somehow ordered the expansion process, the definition of the development areas seems arbitrary and it is likely it was influenced by market forces and the existence of scheduled projects in those areas (Heinrichs et al., 2009) This means that real estate location is more the outcome of the developers' decisions than of discretional regulations; making this case study particularly suitable to be approached through econometric models.

The urban sprawl of Santiago has led to the raise of new residential zones, most of them with a configuration of gated communities (Borsdorf et al., 2007), mainly near the north exit of the city (i.e., *Chicureo*), which has become a new high income area of the city, where approximately 500 hectares were built from 2002 to 2013, and more than 6,000 pre-approved hectares remain to be developed (Transsa Consulting, 2013). Other zones in the south exit of Santiago are being developed under other types of well-delimited urban expansion areas contiguous to consolidated small towns outside Santiago (referred to as *satellites* in this research).

[1] The main city of Chile in administrative, commercial and demographic aspects, among others, and according to INE (2007), with a population that raises to approximately 6'158'080 habitants.
[2] This instrument (Zonas de Desarrollo Urbano Condicionado, or ZODUC in Spanish) has nowadays evolved into a project oriented version of itself (Proyectos de Desarrollo Urbano Condicionado, or PDUC in Spanish).

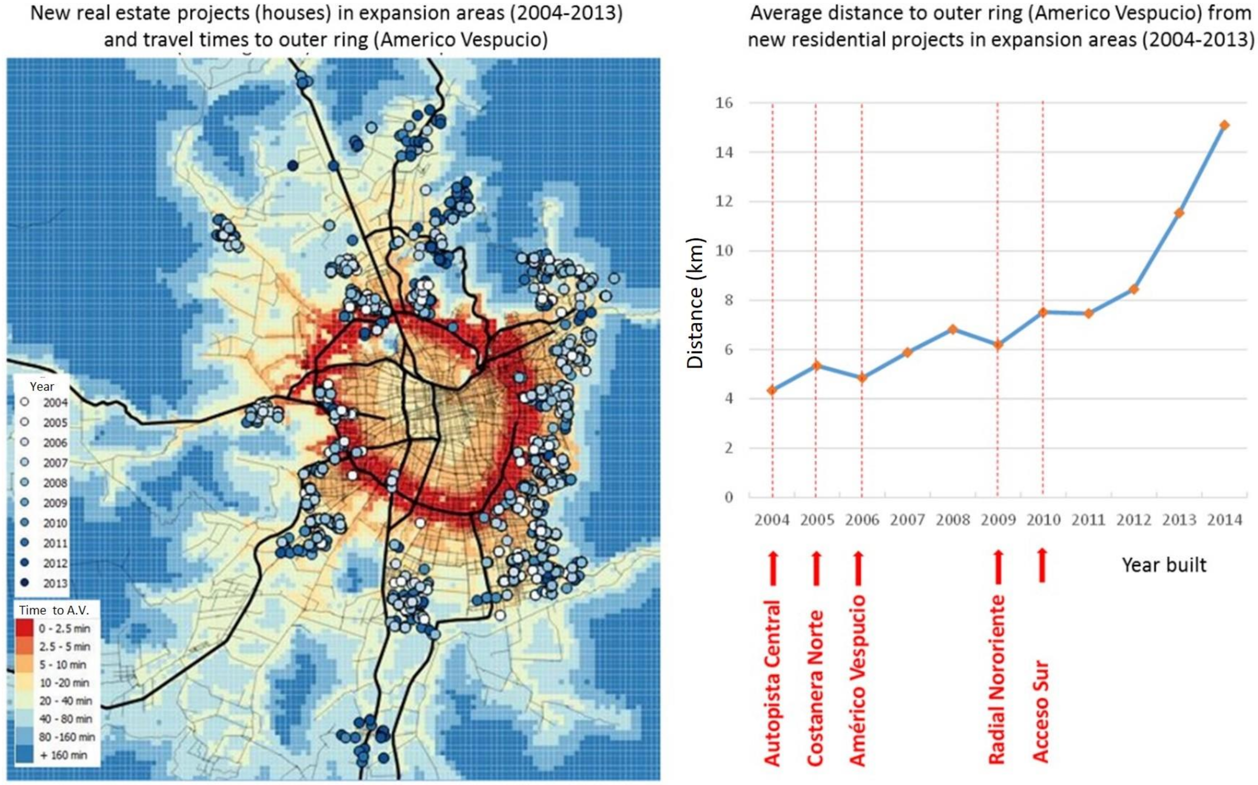

Figure 1: Location of Residential Projects (left) and average travel time to the Americo Vespucio ring (right). Highways are highlighted in black on the left side map and their names and construction year are displayed in red on the righ. Source: own elaboration, based on data from the consulting firm Inciti [3].

4.1. MODEL IMPLEMENTATION AND DATA

We applied the model described in Section 3, to a database of residential projects in expansion areas. The class membership function W_{ns} depends on the plot size, number of units in the project, and initial asking price, following a linear specification. The price function R (see equation 5) depends on a set of location attributes Z_i. 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.

[3] http://www.inciti.com/

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 fulfilled the following conditions: were located within 10 km. of any built project in the database, were located outside of the principal road ring of Santiago (Américo Vespucio), had a feasible terrain slope, and had enough remaining capacity for development.

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 2014[4]. This database describes 1,833 residential projects (one family detached houses) 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 (Instituto Nacional de Estadísticas, 2002, 2018). Demographic attributes of the cells are obtained from the National Census (Instituto Nacional de Estadísticas, 2002) and a socioeconomics segmentation provided by Adimark (2000). A road network topology, obtained from Open Street Map[5] was used to compute accessibility measures. Travel time is obtained thorugh cost surface analysis (for example, see Leusen, 1997) which is an adaptation of Dijkstra's (1959) shortest path algorithm. Average distance to hillsides was calculated for each cell, measuring the distance to the closest hillside every direction (using a 1° step).

All variables describing location attributes (Z_i) and project characteristics (X_n) used in the model are described in Table 1. Figure X shows the distribution for the main characteristics of housing projects in the database.

[4] provided by the consulting firm Inciti http://www.inciti.com/
[5] https://www.openstreetmap.org/

Type	Attribute	Description	Sources
Cell (location) Attributes Z_i	Density	Number of households in a cell for each year divided by surface	Census 2002 (Instituto Nacional de Estadísticas, 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 (Instituto Nacional de Estadísticas, 2002)
	Land acquisition cost ($L_i \cdot q_n$)	Average plot size multiplied by average land value in the cell (UF/m²) (2)	Transsa Consulting (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 (Instituto Nacional de Estadísticas, 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/m²) (2)	Inciti and GFK project database
	Plot Size	Average plot size of the units in the project (m²)	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 1: Attributes used in proposed models

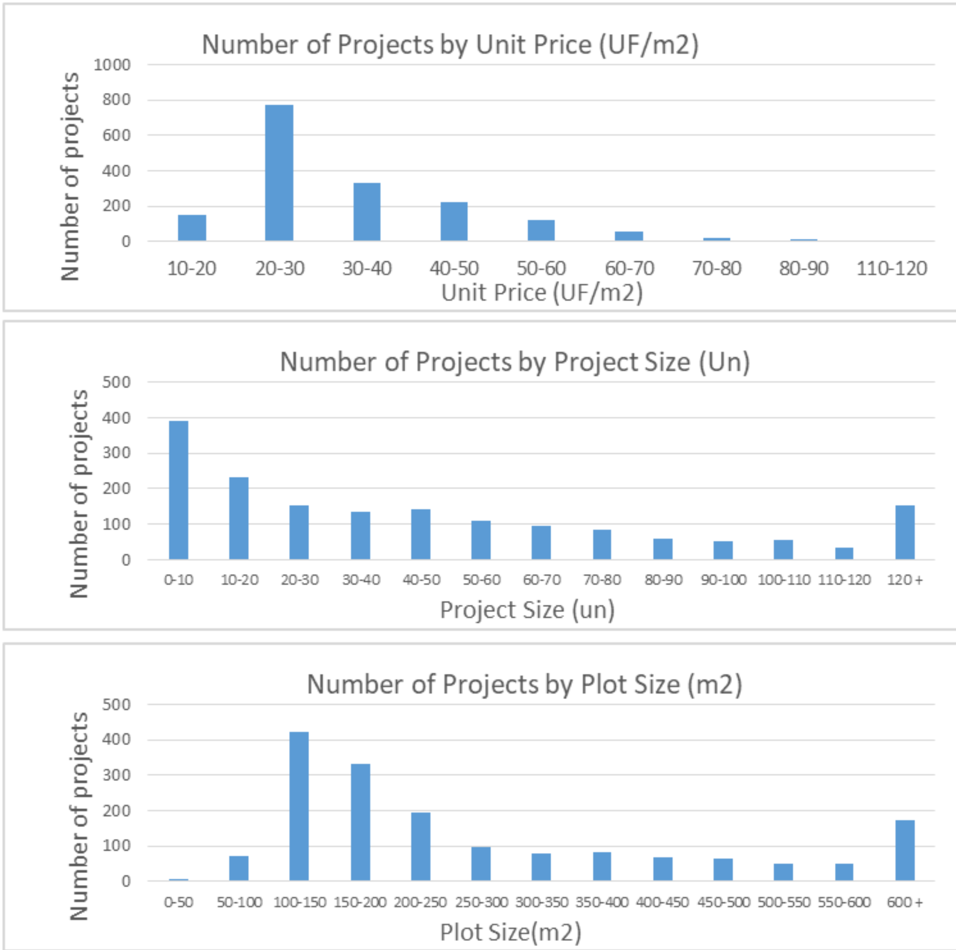

Figure 2: Histogram of number of projects for each range of value in 3 characteristics of the project database.

4.2. ESTIMATION RESULTS

The model described in section 3 is estimated through maximum likelihood using the statistical software PythonBiogeme (Bierlaire, 2003; Bierlaire and Fetiarison, 2009).

For comparison purposes, two models were estimated: a base model with no latent classes, where the expected price –and therefore the location decision- is indifferent to the submarket of the project. The other is the latent class model described in equations 8 to 10, 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. Results are shown in Table 2.

	BASE MODEL (NO CLASSES)	SUBMARKET MODEL (2 CLASSES)	
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	
Attribute parameters (β)	Coefficient (t-test)		
		Class 1 ("Massive")	Class 2 ("Exclusive")
Travel time to high price projects	-0.0519 (-13.92)	-0.0303 (-3.63)	-0.0609 (-14.25)
Travel time to low price projects	-0.0255 (-5.59)	-0.0591 (-6.07)	0.0118 (2)
Land acquisition cost	-0.0000668 (-3.38)	-0.000215 (-6.02)	
Travel time CBD	-0.0272 (-4.19)	-0.0712 (-5.34)	0.00564 (0.75)
Density	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	-0.0151 (-1.81)	0.0851 (4.78)	-0.0568 (-5.71)
Travel Time to nearest Highway	0.129 (15.81)	0.24 (14.27)	0.0407 (4.6)
Travel time to nearest industrial zone	-0.122 (-17.32)	-0.19 (-12.63)	-0.0946 (-9.7)
Travel time to nearest satellite	-0.00233 (-0.61)	-0.0376 (-5.11)	0.0209 (4.71)
Class Membership parameters (θ)		Class 1	Class 2
Intercept		63.6 (2.07)	
Unit Price		-1.29 (-2.08)	
Plot size		-0.13 (-1.88)	
# Units		0.0775 (1.6)	

Table 2: Estimation results.

In both models, most parameters were significant to 95% certainty, and signs and magnitudes are as expected, with a few exception that will be analized later. The latent class model considerably outperforms the base model.

The class membership model (bottom of Table 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 sell at lower price, with smaller plot size and a higher number of units in the project. In contrast, class 2 projects can be labeled as "exclusive" submarket.

Several parameter in the latent class model change significantly with respect to the base 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 base 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 and Hidalgo, 2008; Webster et al., 2002), but in this case instead of crime, protecting 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 tendency showing that 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 project, 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 perspective.

4.3. SPATIAL SEGREGATION OF SUB-MARKETS

Using the class membership parameters (θ), the probability of belonging to the Massive or Exclusive sub-markets can be computed for every project in the database. Figure 3 (upper left) 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 3 (upper right) shows the empirical 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 a 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 3, upper left). 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 this two submarkets. This pattern is coherent with the strong social segregation and inequality observed in Chilean society (Programa de las Naciones Unidas Para el Desarrollo, 2017)

For a given class of project, the location choice probability in all feasible space can be calculated. Figure 3 (bottom) shows the spatial distribution of these probabilities considering cell attributes as observed in 2014. As expected, Exclusive projects replicate the segregation pattern, showing higher location probabilities towards the northeast of Santiago, which agglomerates high income households, and seemingly trying to “move away” from the city. Massive projects have a more even location distribution, concentrating along the outer ring (*Americo Vespucio*) and clearly preferring more central locations.

Probabilities for projects in the Massive submarket are not so dependent on the location of previous massive projects, and are even extended to areas normally associated with exclusive projects. This may be due to the high price project-seeking behavior of this submarket, which is confirmed by the parameter estimated for Travel time to high price projects. This is a phenomena well reported in the literature (Sabatini et al., 2001) where low income households try to locate near high income areas, in order to profit from better urbanization, services and –mainly- status.

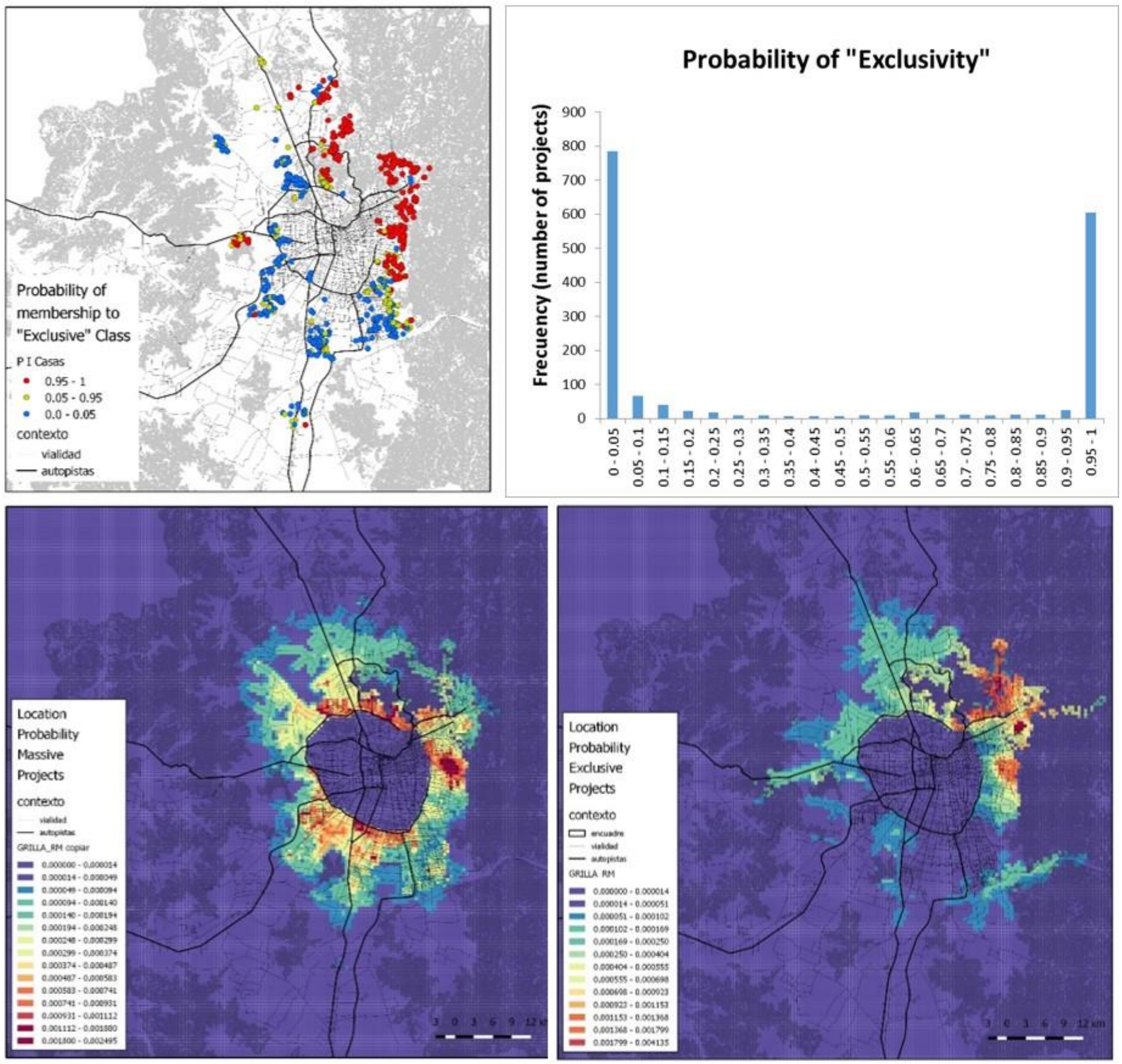

Figure 3: map (upper left) with location of projects and segmentation according to probability of membership to "exclusive" submarket. Histogram (upper right) with number of projects in each range of probability of membership to "exclusive" submarket. Spatial distribution of location probabilities for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

5. CONCLUSIONS

A model for location choice of real estate project in expansion areas is proposed. The modelling framework makes possible to simultaneously identify the parameters of a submarket classification function and the parameters of different expected price functions for each submarket, using a location decision model with latent class structure.

This model structure can also reveal the 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 formation (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. It is interesting to see that this extreme polarization is not due to extreme differences in project characteristics (see Figure 2). Prices, number of units per project and plot surface are not so polarized, but show a rather normal distribution. The differences are, therefore, mostly the product of how each submarket valuates spatial attributes and urban externalities. On one hand, urban agglomerations can be seen as a space of opportunity, where proximity and direct interaction is a positive externality of household and firm location; but on the other hand, agglomeration can also be seen as congestion (not only traffic, but also environmental), undesired social interactions, high cost of private living space and therefore the lack of it. This opposing view of cities is reflected in both submarkets, as the Massive submarket has a positive preference for density and proximity to the city center, while the exclusive submarket tends to search for open space far from consolidated urbanized areas (not only the CBD, but also external urban satellites).

ACKNOWLEDGEMENTS

This article was originated in the context of the research project “Modelo de oferta residencial en el área periurbana de Santiago”; (FAU Project No. 04/13, Universidad de Chile). The authors thank Marcelo Bauza, of Inciti (http://www.inciti.com/), for his help with data collection. The authors also acknowledge the support provided by the Center for Sustainable Urban Development (CEDEUS, CONICYT/FONDAP 15110020), the Complex Engineering Systems Institute, ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816), FONDECYT (projects 1150239 and 1180605) and the CONICYT PhD Scholarship for corresponding author (2015-2019).

REFERENCES

Adair AS, Berry JN and McGreal WS (1996) Hedonic modelling, housing submarkets and residential valuation. Journal of Property Research 13(1): 67–83. DOI: 10.1080/095999196368899.

Adimark GfK and W Thompson (2000) El nivel socioeconómico. Manual de aplicación. Santiago.

Anselin L (2013) Spatial Econometrics: Methods and Models. Springer Science & Business Media.

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

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

Bierlaire M and Fetiarison M (2009) Estimation of discrete choice models: extending BIOGEME. Swiss Transport Research Conference (STRC) (September).

Borsdorf A (2003a) Como modelar el desarrollo y la dinamica de la ciudad latinoamericana. EURE (Santiago) 29: 37–49.

Borsdorf A (2003b) Como modelar el desarrollo y la dinâmica de la ciudad latinoamericana. EURE XXIX(86): 37–49.

Borsdorf A and 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. DOI: 10.1016/j.landusepol.2007.04.001.

Borsdorf A, Hidalgo R and 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.

Bourassa SC, Hamelink F, Hoesli M, et al. (1999) Defining housing submarkets. Journal of Housing Economics 8(2): 160–183. DOI: 10.1006/jhec.1999.0246.

De la Barra T (1989) Integrated Land Use and Transport Modeling. Cambridge: Cambridge University Press.

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numerische mathematik 1(1): 269–271.

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

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.

Galster G (1996) William Grigsby and the analysis of housing sub-markets and filtering. Urban Studies 33(10): 1797–1805. DOI: 10.1080/0042098966376.

Galster G (2001) On the Nature of Neighbourhood. Urban Studies 38(12): 2111–2124.

Goodman A and Thibodeau T (1998) Housing Market Segmentation. Journal of Housing Economics 7(2): 121–143. DOI: 10.1006/jhec.1998.0229.

Haider M and 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.

Heinrichs D, Nuissl H and Rodríguez YC (2009) Dispersión urbana y nuevos desafíos para la gobernanza. EURE (Santiago) XXXV: 29–46.

Hurtubia R, Gallay O and Bierlaire M (2010) Attributes of Households , Locations and Real-Estate Markets for Land Use Modeling. (SustainCity deliverable). Technical Report Transport (Available at http://transp-epfl.ch/documents/technicalReports/sustaincity_WP2_7.pdf).

Instituto Nacional de Estadísticas (2002) Resultados Censo 2002. Santiago. Available at: http://www.ine.cl/docs/default-source/FAQ/síntesis-de-resultados-censo-2002.pdf?sfvrsn=2.

Instituto Nacional de Estadísticas (2018) Resultados Censo 2017. Available at: http://www.censo2017.cl/wp-content/uploads/2017/12/Presentacion_Resultados_Definitivos_Censo2017.pdf.

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

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.

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

Lu Z, Southworth F, Crittenden J, et al. (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.

Martínez F and Donoso P (1996) MUSSA : a behavioural land use equilibrium model with location externalities, planning regulations and pricing policies.

Martínez F and Donoso P (2010) The MUSSA II land use auction equilibrium model. In: Francesca P, Preston J, and Simmonds D (eds) Residential Location Choice: Models and Applications. Berlin: Springer, pp. 99–113.

Massey DS and Denton N a. (1988) The Dimensions of Residential Segregation. Social Forces 67(2): 281–315. DOI: 10.1093/sf/67.2.281.

McFadden D (1973) Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics.

McFadden D (1978) Modelling the choice of residential location. In: Karlqvist A, Lundqvist F, Snickars F, et al. (eds) Spatial Interaction Theory and Planning Models. Amsterdam, North Holland, pp. 75–96.

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

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

Pagliara F, Preston J and Simmonds D (2010) Residential Location Choice. Models and Applications. Springer Science & Business Media. DOI: 10.1007/978-3-642-12788-5.

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

Programa de las Naciones Unidas Para el Desarrollo (2017) Desiguales. Orígenes, Cambios y Desafíos de La Brecha Social En Chile. Uqbar (ed.) Mycological Research. Santiago de Chile.

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

Rosmera NA and Lizam M (2016) Housing market segmentation and the spatially varying house prices. Social Sciences (Pakistan) 11(11).

Sabatini F, Cáceres G and 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.

Schelling TC (1978) Micromotives and Macrobehavior. New York: WW Norton.

Schirmer PM, Van Eggermond M a. B and Axhausen KW (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 AB and Struyk RJ (1976) Segmentation in Urban Housing Markets. Journal of Urban Economics 3(2): 146–166. DOI: 10.1016/0094-1190(76)90050-4.

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

Waddell P (2011) Integrated land use and transportation planning and modelling: Addressing challenges in research and practice. Transport Reviews 31(2): 209–229. DOI: 10.1080/01441647.2010.525671.

Walker JL and Li ÆJ (2007) Latent lifestyle preferences and household location decisions. Journal of Geographical Systems 9(1): 77–101.

Watkins CA (2001) The definition and identification of housing submarkets. Environment and Planning A 33(12): 2235–2253. DOI: 10.1068/a34162.

Webster C, Glasze G and Frantz K (2002) The global spread of gated communities. Environment and Planning B: Planning and Design 29(3): 315–320. DOI: 10.1068/b12926.

Wegener M (2014) Land-Use Transport Interaction Models. In: Springer Berlin Heidelberg (ed.) Handbook of Regional Science, pp. 741–758.

Zöllig C and Axhausen KW (2015) A conceptual, agent-based model of land development for UrbanSim. In: Bierlare M, De Palma A, Hurtubia R, et al. (eds) Integrated Transport and Land Use Modeling for Sustainable Cities. Lausanne: EPFL Press. Available at: http://www.sustaincity.org/publications/SC_Conceptual_Agent_Based_Model.pdf.