SEGMENTATION OF REAL ESTATE DEVELOPERS IN EXTENSION 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

Residential space supply in extension areas emphasises the segmentation of the market in order to attract different types of households. The proposed model uses a latent class approach to identify the importance of different project characteristics in the probability of belonging to one of the two proposed classes, and given the class, the importance of different spatial attributes in the utility of the project. Projects unit value, built space, unit size, accessibility and location attributes were used to estimate the model. Results show expected parameter values, classifying projects according to their "exclusivity".

Key words: Location choice model, Real Estate, Latent class.

1. INTRODUCTION

In the last few decades, there has been an extensive development of land use and transport interaction models (LUTI), mostly from the transport engineering discipline, but also from other fields, such as geography and urban economics. Current LUTI models emphasize on the agent's economic location decision based on his characteristics and the alternative attributes, with transport-related variables playing a significant role.

Wegener (2014) identifies 3 generations in LUTIs, in which the third generation addresses the heterogeneity in preferences, as different agents will value attributes differently according to their own characteristics. This issue requires a higher resolution in the data used, but allows a richer interpretation of parameters and higher adjustment of the model, when the subsets of agents are correctly defined. Exogenous definitions of subsets (by race, or gender, for example) may group agents which doesn't necessarily have the same preferences, and or divide agents which do have similar preferences. This problem can be addressed with Latent class analysis, in which the parameters of the probability of each agent of being part of a subset are estimated in order to maximize the model's likelihood of reproducing the observed decisions.

The present paper explores the different approaches that developers may have towards spatial attributes when evaluating the location of a new residential project in expansion areas. Segmentation in real estate market is an important issue as an strategy of taking household market shares according to their desire of differentiation from other types of households, by income or other characteristics. The model presented jointly estimates both the importance of different characteristics in the projects in the probability of belonging to an specific class of projects, and given that class, the importance in the project utility of different spatial attributes.

The paper is structured as follows. Section 2 provides an overview of the literature in the field of location choice, urban sprawl and agent heterogeneity. Section 3 presents our case study: Santiago (Chile) and Section 4 describes the data collection and processing effort. In Section 5, a mathematical model for the behavior the residential real estate developer is proposed, with estimation results (for both the base model and the model including the new metrics) being shown and discussed in Section 6. Section 7 concludes the paper.

2. LITERATURE REVIEW

This review starts with an introduction to classic models of location and supply of built space, highlighting the theoretical models and current applications. Discussion in the area of urban sprawl is presented, as it's relevant to define the spatial attributes used as input for the model. A more methodological state of art in latent class modelling for location choice is presented in the last section.

2.1. Models that explain residential demand and supply for location

Since the models of Von Thunen and Alonso (Alonso, 1964), residential location has generally been modeled as a bid among different agents (households) in which the location is taken by the highest bidder, which evaluates the utility of a location depending on his trade-off between living space (housing attribute) and commuting costs (location attribute).

McFadden (1978) identified that the heterogeneity of location choice depends on a set of attributes of the house and its location (accessibility, dwelling characteristics, public services, etc.), and of the household (age, income, etc). Fujita (1989) argues that location decision is based in three basic factors: Accessibility, space and environmental amenities. The first is related to potential of activities, the second is related to the characteristics of the dwelling, and the third is related to zone attributes such as views, schools, etc. It has also been shown that accessibility and neighborhood characteristics are highly correlated, and that socioeconomic and demographic attributes may be more important than others, as they capitalize accessibility (Weisbrod, Lerman, & Ben-Akiva, 1980). Also, dwelling attributes is correlated to accessibility (houses vary their characteristics depending on their location) (Waddell, 1993).

As for housing supply, it has not been studied as much as housing demand, because of the complexity of the decision-making process and the multiplicity of products involved (Smith, 1976; Di Pasquale, 1999). Much of the literature for housing supply has focused on exogenous attributes such as macroeconomic conditions, investment and the elasticity between quantity provided and unit price, in the context of utility maximization processes (see for example DiPasquale & Wheaton, 1992; Smith, 1976). Because of its focus on macroeconomic aspects, studies in this area normally don´t address disaggregate spatial issues (accessibility and neighborhood characteristics) (Haider & Miller, 2004).

Supply models based on microeconomic principles are included in models of land use and transport interaction (LUTI) such as Urban Sim (Waddell, 2000), MUSSA (Martínez, 1996) or ILUTE (Salvini & Miller, 2005).

Urban Sim models supply based on the information of current buildings in each cell (150 x 150 meters), and uses the year built data to identify the supply history of the cell. A multinomial logit for the probability of a cell to develop certain land use is estimated considering location attributes in different areas: site (existing development, etc.), urban design scale (proximity to highways, etc.), regional accessibility (access to jobs, etc.), and market conditions (vacancy rates) (Waddell et als, 2003).

MUSSA models supply based on the historic data of developer's behavior disaggregated in zones and type of dwelling. The changes in supply levels depend on the rent variations for each zone and dwelling type, which reflects the willingness to pay function. As an equilibrium model, supply is constrained to equal demand.

Haider and Miller (2004) propose a model to be used in ILUTE, which focuses on real estate developers as heterogeneous agents (uses data from residential projects with specific builder's information). Attributes aggregated in zones and accessibilities to opportunities calculated in a geographical information system (SIG) are used to estimate a logit model for different typologies of development. From the results of this model. the authors propose the concept of "spatial inertia", as the attraction of existing type of uses in a zone to similar uses.

2.2. Drivers of urban sprawl.

Abundant research has been made about urban growth from an economic perspective (Duranton & Puga, 2013), but the spatial characteristics of the development (sprawl or compact) has received fewer attention. Burchfield et al. (2006) related the extent of sprawl in USA cities to urban, topographic and political issues. These authors conclude that sprawl is related to cities built around automobile, lower historical population growth rates, uncertainty about future development (speculation), water aquifers in urban fringe, colder temperature (less value of open space), among others.

According to Brueckner (2001), the principal reasons of sprawl are the population growth, income rise, and decrease in commuting cost. The author identifies three market failures behind the expansion of sprawl: failure to account for benefits of open space, for congestion costs, and for infrastructure costs.

Some authors (Mieszkowski & Mills, 1993) perceive the sprawl as part of natural development of the city, derived from the raise of incomes of certain groups, which can afford larger houses, and transport advance which lowers travel times to the CBD. Other authors (Echenique et als, 2012; Glaeser & Kahn, 2003) evaluate sprawl as equally efficient as compact development, based on the problems of overcrowding and reduced housing choices, and addressing a better quality of life in the suburbs. However, these authors seem to underestimate the effects of sprawl in increase of car use and the resulting increase in congestion (Ewing et al. 2003).

2.3. Vectors of Growth from consolidated urban areas

The localization of households in the city can be explained by decisions based on certain attributes that can be directly measured on a zone or buffer around the location. However, there are other attributes that may not be as straight forward, but could be linked to a more intuitive approach to agent decisions. The agents not only evaluate the disaggregate attributes of a bundle of localization candidates, they also perceive the city as a whole and evaluate the relative location of a candidate in relation to the city zones and its general structure.

One of the first models of city structure was proposed by Park and Burguess (1925), in which the different activities are arranged in concentric rings surrounding the Central Business District (CBD), following the bid-rent curve. Hoyt (1939) proposed a more elaborated model derived from the empirical review of 64 cities in the United States, from which emerged a sector model based on stripes of homogeneous zones connecting the center to the outer areas. The reasoning is that the sectors of the city grow outwards from the preexisting sectors, assisted by radial transport lines. The racial element and poverty zones play a key role in defining these stripes, which reach the center avoiding undesirable zones.

A third model that comes to complement the last two, is the “multiple nuclei” model proposed by Ullman and Harris (1945), whom recognize the importance of the concentric and sector models, but observe that cities don't have just one center but many. The authors elaborate on the factors that create centers but also that separate them because of incompatibilities and specialization.

2.4. Heterogeneity in location choice

Location choice models have represented heterogeneity via dividing agents in different subgroups, each with different valuation parameters. This segmentation traditionally has been done with an exogeneous criteria based on agent’s characteristics (by income, age, etc). Endogenous segmentation has been done with latent classes; the first model to use it for location choice was designed by Walker & Li (2007), where the latent classes were the lifestyle of the household. Other models have been formulated since then (Ettema, 2010, Olaru et al., 2011, Liao et al., 2014, Lu et al., 2014, among others). As far as we know, the only model that applies latent class to urban developer´s types is Glumac, Han, Schaefer (2014), in which different agents involved in the development of brownfields are characterized.

3. MODEL OF DEVELOPER BEHAVIOR

A revealed preferences approach is used. This implies less flexibility in recollecting agent’s data, but is important as we want to measure real urban structure attributes. This approach also has the problem of dealing with a high number of alternatives of location, that normally are spatially correlated. We used a random sampling of alternatives to correct this problem[1].

The model was formulated as a discrete choice model (Domencich & McFadden, 1975; McFadden, 1978), in which there’s a decision maker (the developer), who has a set of alternatives (possible locations in the study area, where he might develop a project), each one

[1] In this version of the model a uniform sampling distribution was used, future versions will explore different sampling strategies.

with different attributes. The model assumes that the decision maker perceives a utility of developing his project in each alternative depending on the attributes of the location, and that determines the probability of taking that alternative.

We divide the developers in two classes. To estimate which projects belong to each class, we used latent class approach (Henry & Lazarsfeld, 1968, Kamakura & Russel, 1989). In simple terms, a latent class model allows the recognition of heterogeneity in the agents that make decisions, introducing in an election model a probability that the decision maker belongs to a specific class, each class having different patterns of behavior or preferences (with different parameters in its utility functions). Some models without latent classes introduce this heterogeneity exogenously (using systematic taste variation or assigning classes *a priori*); however, the latent class methodology estimates the probability of belonging to a class together with the probability of choosing an alternative. For choice models, a model with latent classes has many advantages over a multinomial logit model and mixed logit models, especially in the interpretability of parameters and its relationship with agent attributes (Hess et al, 2009).

In our model, the probability of belonging to one of the two classes of developer depends on the average unit price, average plot size and number of units in the project:

$$P(s|Price_n\,, Plot\ Size_n\,, Units_n\,) \tag{3}$$

The profit π_i^s obtained from building (and selling) a project in a specific location (alternative) i for a developer of class s is:

$$\pi_i^s \quad = \quad B_i^s - c_i^s \quad c_i^s \quad = \quad (\beta_{i1}^s * X_{i1} + \cdots + \beta_{iK}^s * X_{iK}) - \gamma^s * r_i \tag{4}$$

Where B_i^s is the benefit of choosing location i for the developer of class s and c_i^s is the cost of developing the project in zone i. Since price is predefined, the benefit comes mostly from sale-speed of the project and can be interpreted as the (expected) present value of the revenue at location i. Since construction costs must be highly correlated with the type of project, we simplify the cost component, making it only dependent on the land value (r_i). For the same reasons, the expected benefit depends only on location, and not dwelling attributes.

We assume that, given a class of project s, the probability of developing a particular location i is proportional to the expected profit on that location, compared with all other possible alternatives. We assume that a random termϵ_{si}, accounting for unobserved attributes and behavior of the developer, can be associated to the profit for each type of project in each location.

$$\underline{\pi_i^s} = \pi_i^s + \epsilon_{si} \tag{6}$$

Assuming an Extreme Value distribution for the error term, renders a multinomial logit expression for the probability of developing a unit in a particular zone:

$$P_i^s = \frac{exp(\pi_i^s)}{\sum_{i \in \Omega} \quad exp(\pi_i^s)} \qquad \forall\, h \tag{7}$$

where Ω is the set of all possible locations. Because estimating a logit model with a choice set as large as 30.625 alternatives (number of cells in the study area) would be inefficient and too expensive in computational terms, a sampling strategy was used, where 9 alternatives were randomly sampled from Ω , to conform a choice set of 10 alternatives, where the remaining one is the chosen location. Since the sampling probability is uniform across all alternatives, the sampling correction term cancels out and, hence, the choice probability is:

$$P_i^s = \frac{exp(\pi_i^s)}{\sum_{i=1}^{10} \quad exp(\pi_i^s)} \qquad \forall\, h \tag{8}$$

To sample the alternatives, the other nine un-chosen location were randomly selected from the locations that fulfill the following conditions: are located within 10 km. of any built project in the database, are located outside the principal outer ring of Santiago (Américo Vespucio), a slope lower than 3 (in a scale from 1 to 5, where 5 is the steeper), and had enough remaining capacity for development.

Since the latent class approach implies simultaneously estimating, for each observed project, the probability of choosing a location i given its class, and the probability of belonging to each class s, we estimate the final probability of a given project of choosing a location *i*:

$$P(X_n\,, \pi_i) = \sum_{s=1}^{S} \quad P(\pi_i, s) * P(X_n) \tag{9}$$

The logic behind this model is not far from observed behavior in the real estate sector, where developers have a well-defined niche (class that defines their preferences). As developers perceive a latent demand for their product in next years, they define a number of projects (also according to their capital) and seek for the location that maximizes their utility, given the type (class) of the project

4. APPLICATION: SANTIAGO CASE STUDY

Santiago is the main city of Chile in administrative, commercial and demographic aspects, among others, with a population that raises to approximately 6'158'080 habitants. (INE, 2007). To accommodate this population, the expansion of the city has not only been produced by continuous diffusion, but also for the development of already existing and, especially, new satellite locations near Santiago. Regulatory conditions (or the lack of them) largely explain the types of developments that have been generated. The current scenario is based on a policy called "conditioned urban planning"[2] (formulated on the Metropolitan Plan for Santiago of 1997 and 2003), according to which new areas of development, outside the city limits, can be proposed by private real estate agents, whom must build infrastructure and other facilities to mitigate the

[2] In the current instruments this policy is applied in the Conditioned Urban Development Zones (ZDUC, in Spanish), and Conditioned Urban Development Projects(PDUC, in Spanish).

effects of the new urbanization. This condition of liberalization of the areas of future development of the city makes that governmental or regulatory decisions (which are difficult to model) have less importance, and opens a good opportunity for land use researchers to propose, estimate and validate market driven models. Globalization and capital flow has promoted a continuous expansion of the highway system, and other "artifacts" associated with sprawl, such as malls, high standard industrial parks (de Mattos, 2010), and of course mega gated communities (which encompass the projects studied here).

5. DATA ASSEMBLY

The input data used for the model comes from three sources: a private survey and official data. The first type is a database of all residential projects from 2004 to 2014, provided by the consulting firm Inciti[3]. The second type is mainly census data (2002) and network topology. From this data we derived accessibility measures.

Explanatory variables or attributes (x_{ik}) used in this modelling effort are described in Table 1.

Type	Attribute	Description	Source
Cell Attributes	*Density*	*Houses per Hectare*	*Census 2002*
	Land Value	*average UF per square meter, from 2006-2009.*	*Transsa*
Long Distance Accessibility	*Accesib. to high income households*	*Average of Negative exponential of travel time (beta=-0.05) to 110 districts with higher socioeconomic proportion of households(1).*	*own calculation based on openstreetmap roads and census 2002.*
Short Distance Accessibility	*Accesibility to Outer Ring (A. Vespucio)*	*Negative exponential of travel time (beta=-0.05)*	*own calculation based on openstreetmap roads*
	Accesib. to nearest consolidated areas	*Negative exponential of travel time (beta=-0.03) to nearest cell with density higher than 7 households/ha(census 2002 and new residential*	*own calculation based on openstreetmap roads, census 2002 and new residential projects (inciti.cl)*
	Accessibility to nearest highway	*Negative exponential of travel time (beta=-0.03) to nearest highway.*	*own calculation based on openstreetmap roads*
	Accesib. to nearest Industrial zone	*Negative exponential of travel time (beta=-0.03) to nearest cell with industrial land use*	*own calculation based on openstreetmap roads and industrial land use (Ministry of Housing)*
	Average Distance to hillsides	*Average of distances to nearest hillside in every 360 directions.*	*own calculation*
Developer´s project characteristics	*Project Unit Price*	*Average price of units in UF/m2 (2)*	*Inciti and GFK project database*
	Terrain Size	*Average, in square meters.*	*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. Socioeconomic proportion is based on a stratification methodology by Adimark (2014), where households are divided in five classes according to education and material belongings. Socioeconomic proportion of a district is the cuotient between number of households in the two higher educated classes and number of households in the two lower educated.
(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

Detailed information about data sources and calculation methodology is described below.

5.1. New Residential Projects

We use a database of 1.833 residential projects (one family detached houses) built between the years of 2004 and 2014, in the suburban and expansion area of Santiago (out of the main ring of the city, Americo Vespucio). Those projects account for a total of 89,422 new housing units.

It is not easy to quantify the relative weight of this type of developments in total urban growth of the city, but it is possible to make some estimates. According to demographic projections of the National Institute of Statistics (INE), the metropolitan area grew by 717,561 inhabitants in 2004-2013. In turn, if we consider a range of between 3 and 4 persons per household (INE (2007)

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

indicates an average of 3.5 persons per household), it can be estimated that the projects studied are equivalent to a range de 357.688 to 268,266 people in a similar period, which means approximately 37.4% and 49.8% of the new supply had to be produced. It´s difficult to accurately quantify this impact (by differences in the study area, changes in the size of households, demographics, etc.), but we can ensure that the weight of these projects is important within the range of the city.

Preliminary analysis of the location of these projects, and the process of construction of new highways, reveals a pattern of sprawl materialized in growing travel times for the households that choose to live there (see Figure 3).

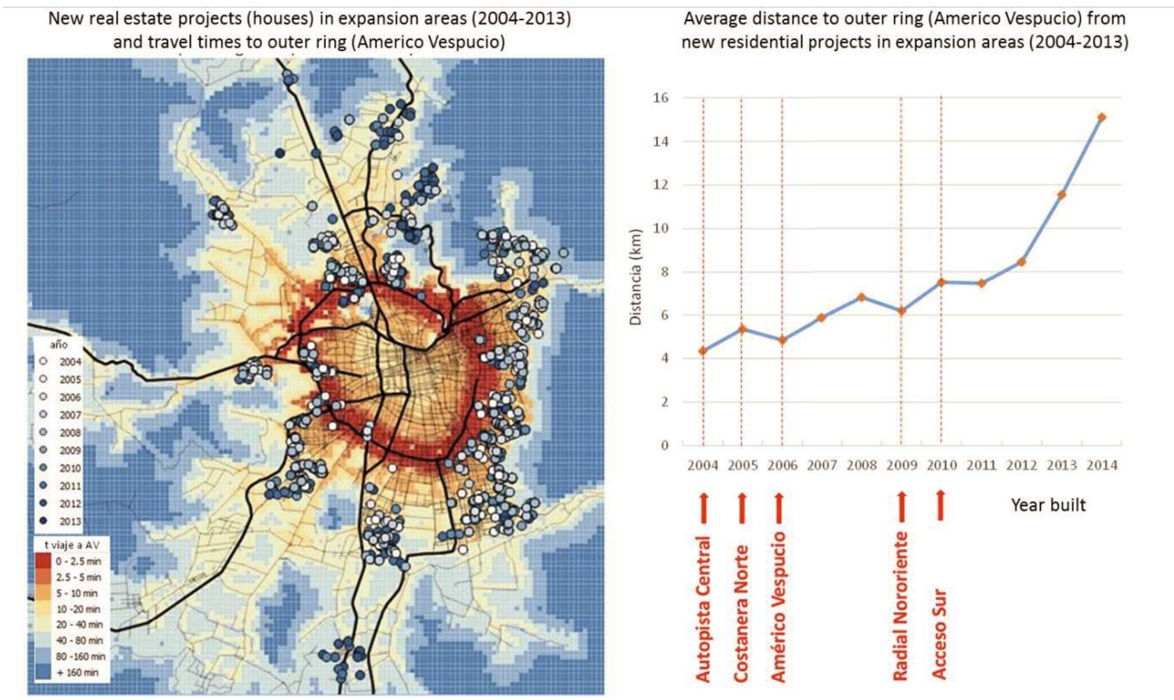

Figure 3: Location of Residential Projects and travel time to outer ring (Americo Vespucio). Price of projects in scale from blue (low price) to red (high price). Highways are highlighted in black. Source: own elaboration, based on data of inciti.cl.

5.2. Official data: Household Census

Beside new residential projects data, the model considered the preexistence of consolidated residential areas, provided by the National Census (INE, 2002). Importance was given to the localization of households by socioeconomic group (GSEs). GSEs are rated by marketing firm Adimark (2000) on the basis of the information provided by the National Census, which considers two fundamental areas: household possessions and level of education of the household head. According to these parameters a score is assigned, usually grouping household into five segments: ABC1, C2, C3, D and E (which can be related to income). This data was aggregated to districts, some of which were modified by the team to fit some differences in rural-urban areas that were not included in the original districts.

Santiago has a strong concentration of higher income groups in the north east zone of the city. The south and west central core of Santiago has big areas of lower income. The data shows that some areas of medium concentration of higher income groups appear out of north east zone, in expansion areas. This is consistent with the latest pattern of segregation in Santiago, which indicates that this variable is decaying in a macro scale but growing in the micro scale (Sabatini, Cáceres & Cerda, 2001).

5.3. Accessibility and distance

A shape of roads is extracted from *Openstreetmap*[4], which is the input for accessibility metrics calculated for this model. The methodology for this attribute is described in model methodology.

In this particular case, we use accessibilities as proposed by (Ingram, 1971), defined by the following equation.

$$Accesibility_{ij} = exp\,(-\beta * travel\_time_{ij}) \qquad (5)$$

With β as a decay parameter. This β parameter was estimated by plotting an histogram of number of trips by duration of the trip (SECTRA, 2015), and adjusting an negative exponential to the histogram, which estimated a beta parameter of -0.05 to fit the data. This is a proxy of accessibility decay along travel time or accessibility-travel time elasticity. Travel time (in minutes) is obtained using a methodology of cost surface analysis for travel time (for example, see Van Leusen, 1999). This methodology is set on a grid where each cell has an impedance or friction, which represents the cost involved in traversing the cell. For this case, we take time (minutes) as the cost, calculated from the average speed at which the cell can be crossed. The impedance is based on the *openstreetmap* data of roads, for which an average speed is assigned depending the hierarchy of the road. With a Dijkstra algorithm (1959), least path from a destination to all cells can be found, with its corresponding travel time. This accessibility measure is not intended as a full transportation model, but it suits our research as a simple and intuitive regional scale indicator of proximity. Figure 3 shows an example of such metric.

Average distance to hillsides was calculated for each cell, measuring the distance to the closest hillside in each 360 directions separated by 1°, and making a simple average. The result is a map of more open and enclosed spaces.

6. ESTIMATION RESULTS

We estimate the models through maximum likelihood using the statistical software Biogeme (Bierlaire, 2003).

From the class membership model results (bottom of Table 2), Class 1 can be interpreted as developers of less "exclusive" projects, as they sell at lower price, with smaller plot size and higher number of units in the project.

---

[4] www.openstreetmap.com

Final log-likelihood	-1397.064	
Likelihood ratio test	4956.374	
Variable	Coefficient (t-test)	
	Class 1	Class 2
Density (House/Ha)	-0.00154 (-10.08)	-0.00366 (-6.29)
Land Value	-0.217 (-8.43)	-0.266 (-7.98)
Accesib. to high income households	-5.6 (-4.06)	18.5 (9.71)
Accesibility to Outer Ring (A. Vespucio)	0.00501 (9.13)	0.00242 (2.45)
Accesib. to nearest consolidated areas	9.15 (25.51)	5.09 (10.1)
Accessibility to nearest highway	-0.00448 (-12.38)	-0.00157 (-3.27)
Accesib. to nearest Industrial zone	0.00125 (4.84)	-0.0039 (-6.4)
Average distance to hillsides	0.0000281 (2.07)	-0.000156 (-7.81)
Class Membership Variables	Class 1	Class 2
Intercept	13.3 (8.09)	
Average Unit Price	-0.283 (-6.3)	
Average plot size	-0.0147 (-3.95)	
Number of units in project	0.0204 (2.48)	

Table 2: Estimation results.

In the model, the signs for the location attributes perform as expected. Plot value (cost of the land) is negative for both classes. Both classes try to choose locations with less density (more open space), and this is more important for high price projects. Accessibility to nearest highway is negative for both, which make sense as this type of infrastructure acts as a NIMBY for nearby projects, independent of the accessibility they provide to other destinations. Accessibility to nearest industrial zones is positive for low price projects, but negative for high price projects. This reflects that industry is not always a negative amenity, but also is a source of opportunities. Accessibility to nearest consolidated area is positive for both classes, and shows the tendency towards agglomeration (which is higher for lower price projects).

Accessibility to high income households is positive for high price projects and negative for low price projects. We would expect that this attribute would be positive for both type of projects, so we can assume there´s an omitted variable. Observations show that actually low price projects tend to locate far from high income zones, but this probably correspond to other attributes (normally land value, which is simplified in our model due to data availability) and not to the desire of being far from high income households. We assume there's a problem of endogeneity in this variable.

An interesting issue with endogeneity emerged during the estimation process. The same presented model, without the "Average distance to hillsides" attribute showed a negative parameter in the attribute of ·Accessibility to Outer ring", for high income class. We assume that those projects tend to locate far from this ring, but not because the accessibility to the ring itself, but because of the presence of certain amenities. When the distance to hillsides was introduced, the endogeneity problem was solved.

Other interpretation to this attribute of is exclusivity. Average distance to hillsides shows to be significant for both classes, being negative for high price and positive for low price projects. This means that high price projects value enclosed locations which are, maybe, more “exclusive”. On the other hand, low price projects value open locations, maybe as they have more space for continuous development and potential for connecting to other places.

7. CONCLUSIONS

Model results showed the important difference in valuation of different spatial attributes for different classes of projects, and also showed that project price and other characteristics are relevant to segment projects according to their valuation of space.

It’s interesting that introducing a spatial attribute we could correct an endogeneity problem that was observed in a previous model (that high price projects show a negative parameter for accessibility to city outer ring). As we introduce this new attribute, this parameter turned to positive sign. With this result, we can interpret the average distance to hillsides as a centrifuge force that guide projects far away from the city, searching for more secluded and “exclusive” locations. Projects looking for small average distance to hillsides locations is linkable to the concept of “gated communities”, as they use topographic delimitation as one more barrier to keep out “not desired” elements (in topographic scale, lower price projects; as in gated communities is crime).

We observed another endogeneity problem in the base model that couldn´t be corrected. We expected that both classes of projects show a positive parameter for accessibility to high income areas. But low price projects showed a negative parameter. Normally, this problem is solved when land value is introduced, but this wasn’t the case. We think that more precision in the land value data is needed to correct this endogeneity.

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), Complex Engineering Systems Institute, ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816), Project AccesoBarrio – A web-based platform to evaluate and model urban accessibility (FONDECYT N°1150239), and CONICYT Phd Scholarship.

BIBLIOGRAPHY

Adimark GfK & Thompson, W. (2000). El nivel socioeconómico. Manual de aplicación.

Alonso, W. (1964). Location and land use. Toward a general theory of land rent. *Location and land use. Toward a general theory of land rent.*

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. In *Swiss Transport Research Conference* (No. TRANSP-OR-CONF-2006-048).

Brueckner, J. K. (2001). Urban sprawl: Lessons from urban economics. Brookings-Wharton papers on urban affairs, 2001(1), 65-97.

Burchfield, M., Overman, H. G., Puga, D., & Turner, M. A. (2006). Causes of sprawl: A portrait from space. The Quarterly Journal of Economics, 587-633.

De Mattos, C. A. (2010). Globalización y metamorfosis metropolitana en América Latina: De la ciudad a lo urbano generalizado. *Revista de Geografía Norte Grande*, (47), 81-104.

DiPasquale, D. (1999). Why don't we know more about housing supply? The Journal of Real Estate Finance and Economics, 18(1), 9-23.

DiPasquale, D., and Wheaton, W. C. (1992). "The Markets for Real Estate Assets and Space." Real Estate Econ. 20(2), 181–197.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1), 269-271.

Domencich, T. and McFadden, D. (1975). Urban travel demand: a behavioral analysis, Amsterdam : North-Holland Publishing Company.

Duranton, G., ; Puga, D. (2013). The growth of cities.

Echenique, M. H., Hargreaves, A. J., Mitchell, G., & Namdeo, A. (2012). Growing cities sustainably: does urban form really matter?. *Journal of the American Planning Association*, *78*(2), 121-137.

Ettema, D. F. (2010). The impact of telecommuting on residential relocation and residential preferences, a latent class modeling approach. *Journal of Transport and Land Use*, 3, 7-24.

Ewing, R., Pendall, R., Chen, D. (2003). Measuring sprawl and its transportation impacts. Transportation Research Record: Journal of the Transportation Research Board, (1831), 175-183.

Fujita, M. (1989). *Urban economic theory: land use and city size*. Cambridge university press.

Glaeser, E. L., Kahn, M. E. (2004). Sprawl and urban growth. Handbook of regional and urban economics, 4, 2481-2527.

Glumac, B., Han, Q., & Schaefer, W. F. (2014). Actors' preferences in the redevelopment of Brownfield: Latent class model. *Journal of Urban Planning and Development*, 141(2), 04014017.

Haider, M., & Miller, E. (2004). Modeling location choices of housing builders in the Greater Toronto, Canada, Area. Transportation Research Record: Journal of the Transportation Research Board, (1898), 148-156.

Harris, C. D., & Ullman, E. L. (1945). The nature of cities. *The Annals of the American Academy of Political and Social Science*, 242(1), 7-17.

Hess, S., Ben-Akiva, M., Gopinath, D., & Walker, J. L. (2009). *Taste heterogeneity, correlation, and elasticities in latent class choice models* (No. 09-2428).

Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities*. Federal Administration, Washington, D.C.

Ingram, D. R. (1971). The concept of accessibility: a search for an operational form. *Regional studies*, 5(2), 101-107.

Instituto Nacional de Estadísticas (INE), Chile. (2007). *Ingresos de hogares y personas 2006*. Santiago.

Instituto Nacional de Estadísticas (Chile). (2002). *Censo 2002 resultados*. INE.

Kamakura, W. A., & Russell, G. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research*, Vol. 26, No. 4 (Nov., 1989), pp. 379-390

Lazarsfeld, P. F., Henry, N. W., & Anderson, T. W. (1968). *Latent structure analysis* (Vol. 109). Boston: Houghton Mifflin.

Martinez, F. (1996). MUSSA: land use model for Santiago city. Transportation Research Record: Journal of the Transportation Research Board, (1552), 126-134.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

Mieszkowski, P., Mills, E. S. (1993). The causes of metropolitan suburbanization. *The Journal of Economic Perspectives*, *7*(3), 135-147.

Liao, F. H., Farber, S., & Ewing, R. (2015). Compact development and preference heterogeneity in residential location choice behaviour: A latent class analysis. *Urban Studies*, 52(2), 314-337.

Lu, Z., Southworth, F., Crittenden, J., & Dunhum-Jones, E. (2015). Market potential for smart growth neighbourhoods in the USA: A latent class analysis on heterogeneous preference and choice. *Urban Studies*, 52(16), 3001-3017.

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

Park, R. E., Burgess, E. W., & McKenzie, R. D. (1984). *The city*. University of Chicago Press.

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.

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.

SECTRA. (2015). Actualización y recolección de información del sistema de transporte urbano, IX Etapa: Encuesta Origen Destino Santiago 2012. Report by Observatorio Social, Universidad Alberto Hurtado. Santiago: Biblioteca de SECTRA. Available at: http://datos.gob.cl/datasets/ver/31616 Accessed 20.04.15.

Smith, B. A. (1976). The supply of urban housing. The quarterly journal of economics, 389-405.

Van Leusen, M. (1999). Viewshed and cost surface analysis using GIS (cartographic modelling in a cell-based GIS II). *BAR International Series*, 757, 215-224.

Von Thünen, J (1826). Der isolierte Staat. Beziehung auf Landwirtschaft und Nationalökonomie.
Waddell, P., Berry, B. J., & Hoch, I. (1993). Housing price gradients: The intersection of space and built form. *Geographical Analysis*, *25*(1), 5-19.

Waddell, P., Berry, B. J., & Hoch, I. (1993). Housing price gradients: The intersection of space and built form. Geographical Analysis, 25(1), 5-19.

Waddell, P. (2000). A behavioral simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. Environment and planning B: Planning and Design, 27(2), 247-263.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., & Ulfarsson, G. (2003). Microsimulation of urban development and location choices: Design and implementation of UrbanSim. Networks and Spatial Economics, 3(1), 43-67.

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

Weisbrod, G., Lerman, S. R., & Ben-Akiva, M. (1980). Tradeoffs in residential location decisions: Transportation versus other factors. *Transport Policy and Decision Making*, *1*(1), 13-26.