A location choice model for residential real estate projects in expansion areas of Santiago, Chile

Tomás Cox [a,b] and Ricardo Hurtubia [a,c]

[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

A discrete choice model for localization of residential projects (categorized by unit price-range) in expansion areas of Santiago is presented. Cell attributes (density, land use, etc.) are used as explanatory variables, but also accessibility (calculated with cost surface analysis) to previously built projects and to households divided by social groups are relevant explanatory variables for the choice decision process. The use of these variables is relevant in order to represent the social segregation patterns observed in Santiago and how these patterns are replicated in the sprawling process.

As expected, estimation results show strong positive correlations between location of high priced projects and accessibility to similar projects, and strong negative correlation for these projects to accessibility to low income social groups (and *viceversa*). Discontinuous expansion is attributed to centrifugal drivers associated to influence areas of lower income social groups, rather than the explanation of trade-off between accessibility and land price usually found on classic urban economic theory.

Cell attributes are not as relevant as relational attributes (accessibility), as modelling expansion areas is not so dependent on near amenities but to be associated to "catchment areas" of certain zones in the consolidated city structure.

Working paper, available online May 2017

1. Introduction

The model presented is formulated to answer the question about which are the drivers or attractors for the localization of residential projects in expansion areas. Considering the amplitude of the problem of residential location, we focus the research area in the rural-urban fringe of the city, in which location and real estate development has a different approach than in consolidated areas.

The discontinuous and fragmented pattern of development of residential projects, which are big in extension but far from urban opportunities raises a more particular question, about the centrifugal forces, different from land value and which we hypothesize are linked to complex dynamics relating to influence areas of different social groups and among projects.

The study of the behavior of this ecology of projects is particularly interesting in Chile, due to its very market driven and quite unregulated city growth, added to a society in which status and differentiation reflects on residential developments as social positioning products.

In the next sections a discrete choice model for the developer´s decision of localization of residential projects in expansion areas of Santiago is formulated, constructed on a database of 1,833 projects from 2004 to 2014, accounting for 89,422 developed units. A literature review is presented, to show some of the main drivers for localization that are commonly considered. Then, some facts on the particular case study of Santiago are described briefly. In the methodology section, we propose a microeconomic model for the developer's choice of location for a particular type of project. The model is based on spatial attributes that are taken from official sources (mainly localized or cell specific attributes) and relational attributes such as accessibility to other projects or specific types of households in the consolidated city. The methodology to calculate this attributes is also described in this section. The data processed and generated through this methodology is analyzed prior to formulate the model, to understand some of the patterns in a more intuitive way. Model results are shown and parameter estimates are commented. We conclude the paper by discussing the results and exploring the next steps to properly simulate future scenarios of urban sprawl using the models proposed here.

The model is similar to the one proposed by Haider and Miller (2004), and shares various aspects with the supply models used in Urban Sim (Waddell et als, 2003) and MUSSA (Martínez, 1996). Our main contribution is the exploration of variables that are relevant in expansion areas (and could be extrapolated to inner city models), such as accessibility to similar social groups and projects, which is also very relevant in the Latin American context. This endogeneity is an extension to the concept of "spatial inertia", applied by Haider and Miller to zones, but in the model presented is extended to gravitational accessibility (long distance).

2. Literature review

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, 1997). 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 (Miller & Salvini, 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.

3. 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"[1] (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 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 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).

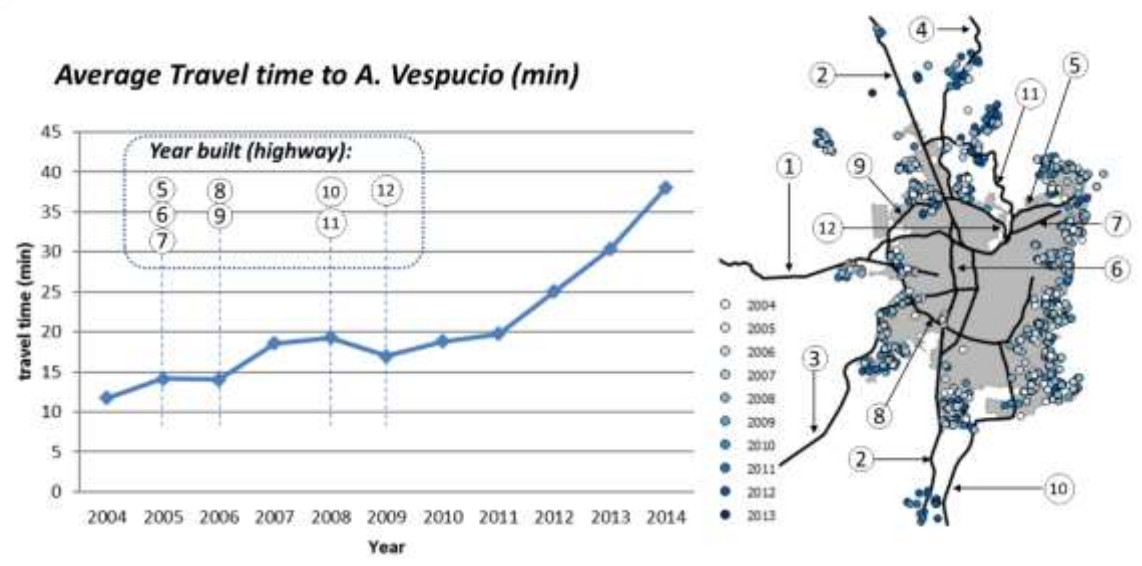

Figure 1: Average travel time to outer ring (A. Vespucio), for new residential projects in expansion areas. The year of construction for each highway is shown (see Annex 2). Source: own elaboration.

[1] 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).

Figure 1 shows, based on the studied projects, the continuous increase of (average) travel time from new units to the consolidated city, as a result of this pattern of discontinuous growth.

4. Methodology

Some of the attributes in the cells are extracted from official sources (Census), but accessibilities are calculated especially for this research. The accessibility measures are based on cost surface analysis, which can inform of general accessibilities based on the road network, avoiding the need for transport models. This is an advantage in terms of simplicity of the accessibility calculation process, although it produces access measures that do not account for congestion dynamically. However, these simplified measurements can be good enough to explain the decision making process because, at the moment of development of these new projects, congestion is usually not an issue (although it may become one afterwards)

Land value is extracted from the Inner Taxation Service (SII) database for average land value for each taxation zone in the region, as of year 2010. The appraisal process for every zone done by this bureau is based on market comparing criteria (survey of offered values in the zone). This type of variables is not commonly included because of the difficulty of obtaining reliable data (Haider & Miller, 2004). In this case, data has its limitations because of aggregation in zones and that is not specific to different type of terrains. Also, land value should be dynamic in time, and for simulation stages it should include a price formation model. For the model presented, it´s used as a control variable and conclusions about its results should be taken carefully.

4.1. Proposed model

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 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. Taking this in to account, we also include systematic taste variation, dividing the developers in 5 groups according to price range of the projects they develop.

The profit π_i^h of a location (alternative) i for a project of type h is:

$$\pi_i^h \quad = \quad B_i^h - c_i^h \quad = \quad \left(\beta_i^h * X_i + \cdots + \beta_I^H * X_I\right) - (\gamma^h * r_i) \tag{1}$$

Where B_i^h is the benefit of choosing location i for the project and c_i^h 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 type of project h 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 term ϵ_{hi}, accounting for unobserved attributes and behavior of the developer, can be associated to the profit for each type of project in each location. 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:

$$P_i^h = \frac{exp(\pi_i^h)}{\sum_{i \in \Omega} exp(\pi_i^h)} \quad \forall\, h \tag{2}$$

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[2], the sampling correction term cancels out and, hence, the choice probability is

$$P_i^h = \frac{exp(\pi_i^h)}{\sum_{i=1}^{10} exp(\pi_i^h)} \quad \forall\, h \tag{3}$$

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), have a slope lower than 3 (in a scale from 1 to 5, where 5 is the steeper), and had enough remaining capacity for development.

The logic behind this model is not far from observed behavior in the real estate sector, where developers have a well-defined niche (range of prices). 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 price of the project.

4.2. Grid and Accessibility measures

The model is based on a discrete grid of 87.5 x 87.5 km, with cells of 500 x 500 m. (175 x 175 cells), in which every cell is a possible alternative and has attributes specific to its location.

The attributes on each cell are divided in zonal (density, socioeconomic, etc.) and relational (accessibility). The zonal attributes are extracted from diverse sources (described below), and the relational are calculated with the following methodology.

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

The calculation of accessibility for each possible location to a destination set is based on the methodology of cost surface analysis (implemented in different spatial analysis programs, such as IDRISI, algorithm known as *cost surface analysis*). The following describes the steps taken to this calculation:

1) Each cell has an impedance or friction, which represents the cost involved in crossing the cell. For this case, we take time (minutes) as the cost, calculated from the average speed at which the cell can be crossed.
2) The average speed is based on the street network in the area, depending its hierarchy: Highway: 80 km/hr; other urban streets: 20 km/hr (normal for congested areas), rural roads: 50 km/hr. If there is no road, walking speed is assigned (3 km/hr, 1 km/hr in sloped areas).
3) Based on this costs, travel time from each cell to one cell can be calculated with Dijkstra algorithm (1959), considering the grid as a regular network where each cell is a node and the cost in the arc between two nodes is the average of the cost in the two corresponding cells.
4) This methodology calculates the travel time from any cell to the rest of the cells. From this, a gravitational measure (Ingram, 1971) can be obtained to account for accessibility in a cell I to a set of destinations J (for example, to schools), with the following functional form:

$$Accesibility_i = \sum_{j=1}^{J} O_j * \exp(-\beta * time_{ij}) \tag{4}$$

where β is an elasticity factor that determines the decay of accessibility in function of time, and O_j is the importance of destination j.

This measure can be classified as gravitational type (Geurs & Van Wee, 2004), because it considers the cost associated with moving to a set of destinations, weighted by the importance or weight of those destinations. The aim of this work is not focused on developing a full measure of accessibility, but rather an approach that meets reporting proximities and distances between locations on a regional scale. Therefore we consider only two of the four components of the accessibility measure proposed Geurs and Van Wee (2004) -the land use (location of the destinations) and the transport system (beats per network) -, and does not consider the time factor (time availability of people) and individual (availability of transport modes).

For this model, we calculated accessibility to different types of destinations:

1) New projects by type: calculated for each year, where the set of destinations are the projects built till that year, of certain type (A to E). Gravitational measure with negative exponential function (beta= -0.05. This elasticity makes accessibility decay to 50% at 14 min. and to 5% at 60 min., considering that being in the destination gives 100% accessibility) was used. The importance of each destination is the number of units of each project. The road network also is modified according to the existent highways.
2) To preexisting households by socioeconomic group: Same as 1), but the set of destinations are the centroids of 2002 census districts, separated by social group. The importance of each destination is the number of households of that group.
3) Other destinations: to nearest highway, to nearest cell with more than 7 households per hectare (urban), to nearest industrial cell.

5. Data Collection and analysis

In this section, official source data is described and analyzed together with ad-hoc calculations of accessibility to ground some first approach to the important attributes behavior, prior to estimate the model.

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, but it is possible to make some projections. 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) 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.

The houses were divided into types (Table 1) by unit prices (UF/m^2)[3]. The classification was made with Jenks (1967) Natural Breaks, which searches through successive iterations on a data range, grouping observations that most resemble each other and in turn differ from the rest.

Project Type	Number of project	Total number of built units	Min Built Surface Price (UF/m^2)	Max Built Surface Price (UF/m^2)
A	62	1190	63.81	109.72
B	231	5140	46.74	63.8
C	368	14361	33.91	46.73
D	565	34265	24.6	33.9
E	607	34466	11.7	24.59

Table 1: Project Classification. Source: own elaboration.

Projects had little inner variation in terms of prices, which fits the hypothesis of low social mixture within specific projects.

[3] UF: *Unidad de Fomento*, a national price unit adjustable by inflation equivalent to $US 39,4 (21-10-2016)

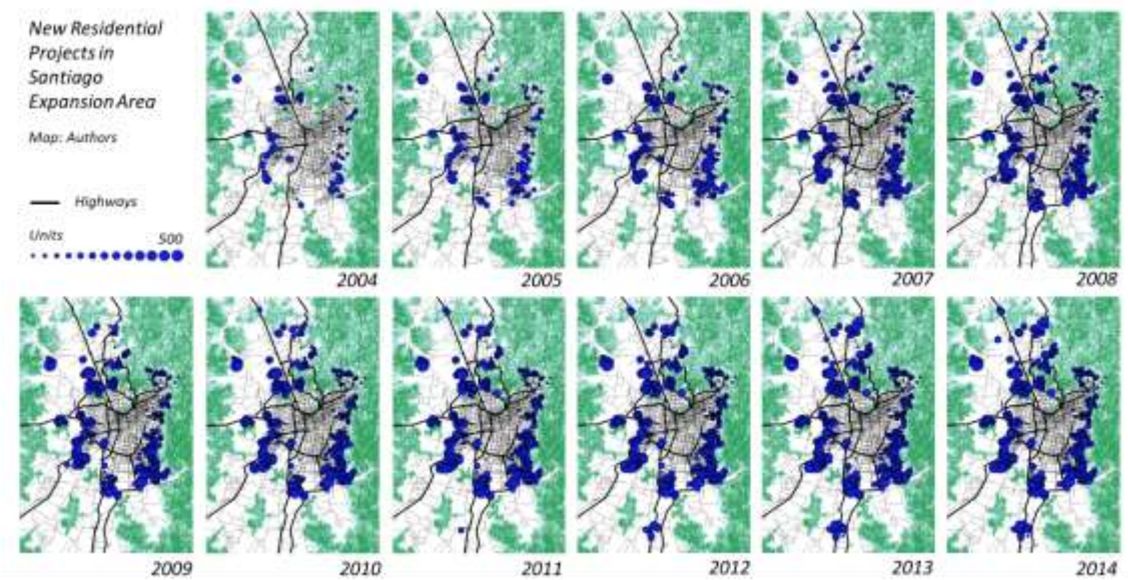

Figure 2: Evolution of residential project locations in Santiago by year (accumulated). New highways are also shown. Source: own elaboration.

5.2. Household Census data

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 office 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 (from higher to lower socioeconomic status). 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.

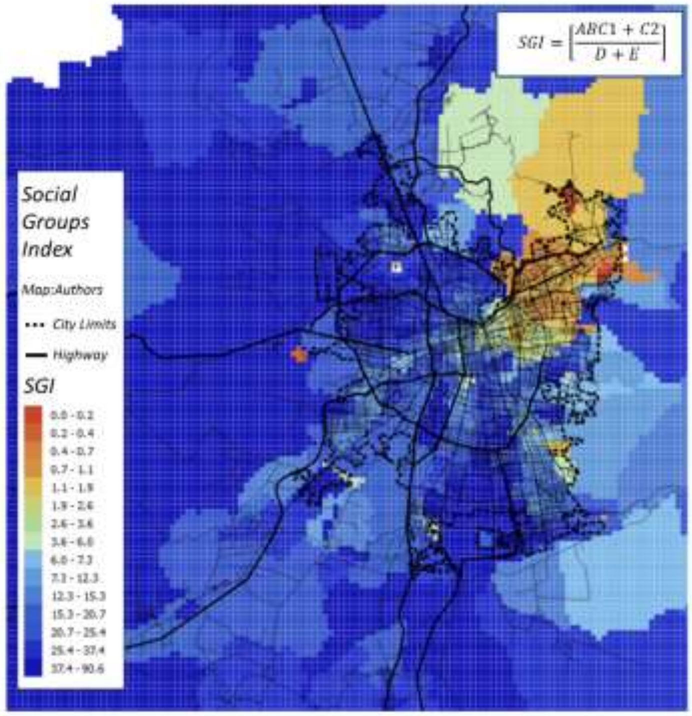

Figure 3: Social Groups Index, consisting of the ratio between high and low income groups in each cell. Source: own elaboration.

Figure 3 shows the strong concentration of higher social 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 social 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

As it was expected, accessibility to higher price projects (A) is concentrated towards the north east, but also there is an interesting medium accessibility area in Chicureo, a recent development area in the north part of Santiago. This value of accessibility is explained because of two mountain roads that connect this area to the north east of the consolidated city.

The accessibility to lower price projects (E) is much more spread in the rest of the city and its expansion area, which confirms the common observation that lower income groups are located much more deconcentrated that higher income groups.

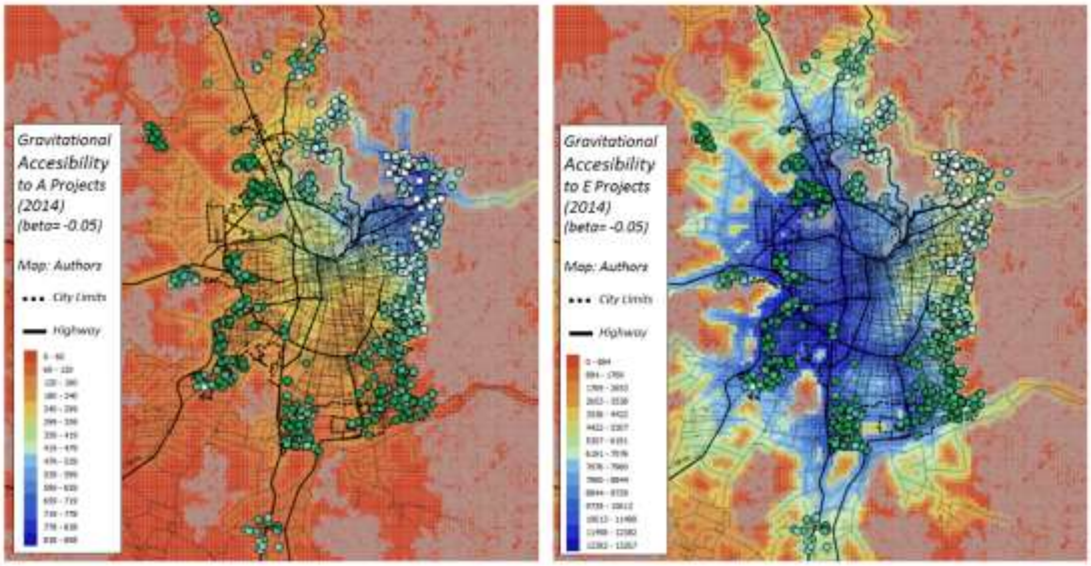

Figure 4: Accessibility to projects of type A and E. Source: own elaboration.

The probability of developing a cell not only relates to its relation to socioeconomic aspects, but also to structural conditions of the city. Figure 5 (left) shows the accessibility to nearest urban cell (above 7 households by hectare, according to 2002 census. This variable is updated with the incoming projects). This accessibility is calculated with a strong decay in function of travel time (beta = -0.3. This elasticity makes accessibility decay to 50% at 2.3 min. and to 5% at 10 min., considering that being in the destination gives 100% accessibility), so it can show the condition of being in the “growing frontier” of the city or smaller urban satellites. At first sight, there is a strong connection between developed cells and its “frontier” status.

Figure 6 (right) shows the travel time (minutes) to the nearest highway. It´s also clear that projects are attracted to this kind of infrastructure, as it ensures fast access to the consolidated city. But also there are differences according to type of development, as higher price projects tend to search for locations on a reasonable distance to the nearest highway.

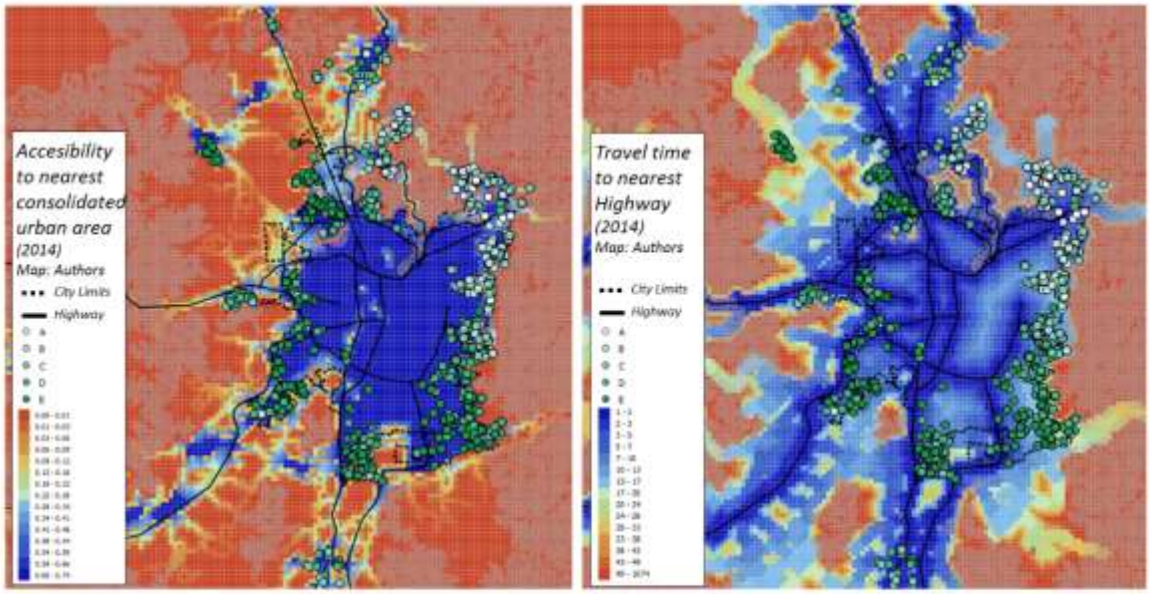

Figure 5: Accessibility to nearest consolidated urban area (cells with more than 7 Households/Ha) and travel time (minutes) to nearest highway. Source: own elaboration.

5.4. Price and accessibility

Classic urban economics theory states that housing distribution along the city follows a tradeoff between accessibility and land price (Alonso, 1964). Therefore, the question of what drives households to locate in projects far away from the city (as we see in the case of Santiago) could be answered by identifying a low price per land-unit in those projects. But instead, as we see in Figure 6, price and accessibility do not correlate in a strong opposite way. The correlation is insignificant ($r^2 = 0.015$) and we can see that some projects with low accessibility have higher prices than projects with good accessibility. We use accessibility to the outer ring Américo Vespucio (the main avenue-highway of the city, which connects all the suburbs around Santiago) as a proxy of accessibility to opportunities for locations in expansion areas.

The aggregation of projects into "neighborhoods" (clusters of new projects) doesn't generate distortion (ecological fallacy), as they share almost identical accessibility, and prices have small inner variation (autosegregation).

Figure 6: Relation of accessibility to consolidated city and average price of units in projects. Source: own elaboration.

As the main hipothesis states, the price can be better explained by the relation of the project to social groups and other existing projects. For example, comparing the project price to accessibilities to high price projects (A&B), and to accessibility to lower social groups (D&E), a strong relation appears. Figure 7 shows increasing price as accessibility to A&B projects rises, and alike, decreasing accesibility to D&E social groups. The correlation between price and the ratio between both accesibilities is significant ($r^2 = 0.78$).

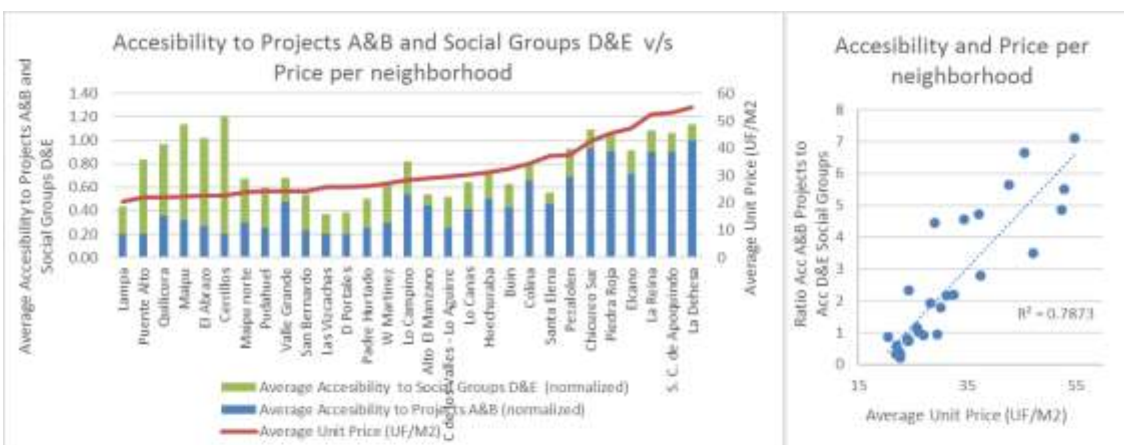

Figure 7: Relation of accessibility to projects A and B, to social groups D and E (bars), and average price of units in projects (line). Source: own elaboration.

6. Estimation Results

The model described in section 5.1 is estimated through maximum likelihood with Biogeme (Bierlaire, 2003).

Before describing the resulting parameter values, we can outline the structure of the model by identifying 3 categories in which we can divide the selected variables:

1) Gravitational Long Distance Relations: the relation of the cell (location) to the system of projects and preexisting households in the region, considering high price projects and social groups of lower income (which showed low correlation, see Annex 1).
2) Cell Attributes: characteristics of the location itself (500 x 500 m. cell), which include density (households per hectare) and plot value. Social composition in the cell was discarded as it was highly correlated to long distance relations and show strange signs in the coefficients.
3) Short distance nearest destination relation: Normally these variables are used for neighborhood amenities such as schools or malls, and also for NIMBYs (Not In My Back Yard) such as contaminant industries or water treatment. As expansion areas are not consolidated and new residential development in this areas doesn´t normally install near commerce, schools, etc. but *viceversa*, here we used accessibilities to structural aspects of the city such as highways, urban frontier, and consolidated macro industrial areas.

From values in table 2, some conclusions can be made about the valuation that each type of developer gives to the variables:

MODEL RESULTS			Final log-likelihood	-84176		
			Likelihood ratio test	214959		
Variable Type	Variable	Project Type	Parameter Name	Value	t-test	
Gravitational Long Distance Relations	Acc_DE	A	*Acc_DE_pr_A*	-8.98000	-1.45	**
		B	*Acc_DE_pr_B*	-3.26000	-1.85	**
		C	*Acc_DE_pr_C*	0.46700	0.45	
		D	*Acc_DE_pr_D*	1.38000	1.41	**
		E	*Acc_DE_pr_E*	2.57000	2.73	*
	Acc_Pr_A+B	A	*Acc_Pr_A+B_pr_A*	0.00216	3.63	*
		B	*Acc_Pr_A+B_pr_B*	0.00068	2.6	*
		C	*Acc_Pr_A+B_pr_C*	-0.00022	-0.97	
		D	*Acc_Pr_A+B_pr_D*	-0.00307	-7.28	*
		E	*Acc_Pr_A+B_pr_E*	-0.00683	-8.46	*
Cell Attributes	Density	A	*Density_pr_A*	-0.00835	-3.46	*
		B	*Density_pr_B*	-0.00473	-7.02	*
		C	*Density_pr_C*	-0.00372	-9.87	*
		D	*Density_pr_D*	-0.00138	-6.16	*
		E	*Density_pr_E*	-0.00030	-1.51	**
	Plot_Value	A	*Plot_Value_pr_A*	0.00006	0.88	
		B	*Plot_Value_pr_B*	-0.00006	-1.56	**
		C	*Plot_Value_pr_C*	-0.00014	-3.1	*
		D	*Plot_Value_pr_D*	-0.00081	-6.09	*
		E	*Plot_Value_pr_E*	-0.00088	-3.93	*
Short distance nearest destination relation	Acc_Near_Urban	A	*Acc_Near_Urban_pr_A*	10.50000	5.27	*
		B	*Acc_Near_Urban_pr_B*	7.08000	12.78	*
		C	*Acc_Near_Urban_pr_C*	5.58000	16.59	*
		D	*Acc_Near_Urban_pr_D*	7.92000	19.86	*
		E	*Acc_Near_Urban_pr_E*	7.01000	17.26	*
	t_t_Near_Hway	A	*t_t_Near_Hway_pr_A*	-0.01530	-0.88	
		B	*t_t_Near_Hway_pr_B*	-0.02450	-2.67	*
		C	*t_t_Near_Hway_pr_C*	-0.01910	-3.51	*
		D	*t_t_Near_Hway_pr_D*	-0.00652	-1.83	**
		E	*t_t_Near_Hway_pr_E*	-0.00349	-1.14	
	Acc_Near_Industry	A	*Acc_Near_Industry_pr_A*	-0.00413	-1.78	**
		B	*Acc_Near_Industry_pr_B*	-0.00264	-4.22	*
		C	*Acc_Near_Industry_pr_C*	-0.00070	-2.34	*
		D	*Acc_Near_Industry_pr_D*	0.00048	1.6	**
		E	*Acc_Near_Industry_pr_E*	0.00170	5.54	*

*Table 2: Estimated parameters. * Significant at 95%. ** Significant at 85%*

1) As expected, relational variables show an attraction of high price developers to locate near preexisting high price projects, and far from lower income social groups. For lower price projects, coefficients show attraction to lower income groups and the opposite to high price preexisting projects. This last result is not so intuitive, as one

would think that, controlling by land value as we do, lower price projects would valuate to locate near high price projects, even though they don´t because of high land value.

2) Controlling by other variables, all type of projects prefers lower densities, and this valuation is bigger for higher price projects. This is intuitive as developers look for open spaces where is easier to buy big plots, much of which will not be built immediately, but reserved for future demand.
3) Plot value parameters have the correct sign and values, and indicates that location probability for higher price projects depends less on land value. In firsts estimations, we used land value (per square meter), but it showed that lower price projects tend to locate in cells with higher land value than high price projects. As we corrected this variable by the average size of the plot, parameters improved. Coefficients show that land price is almost not significant for high price projects.
4) The most significant variables are accessibilities to the urban frontier, which shows how important for projects to cluster and locate nearby already built cells. This variable is dynamic in time and accounts for cells with urban density not only in the continuous consolidated area of Santiago, but also for small satellites and recently developed cells. It could be interesting to divide this variable depending if the accessibility is taken to bigger consolidated or smaller recent urban areas.
5) Travel time to Highways shows a negative value (longer distance to the nearest highway makes a location less attractive). This result makes sense, and also the fact that for projects of type A is not significant (they search for more exclusive locations, not near the highway, that can be related to industrial or other land uses).
6) Accessibility to industrial areas shows positive values for lower price projects, and negative for higher price projects. A recent trend in residential real estate companies is to develop big projects where there´s coexistence of industry and housing (mid income), and the first uses the lots that face the highway and the residential locates in second line. This type of industry has to be non-pollutant and built with high urbanization standards, many times mixed with administrative offices.

6.1. Mapping Probabilities

Using equation (3), we compute the choice probabilities for each type of project, accounting for all locations eligible for developer's choice set. Figure 9 shows a map of the probability of developing each type of project in each cell. These maps can be read as describing the attractiveness of potential locations for new real estate developments.

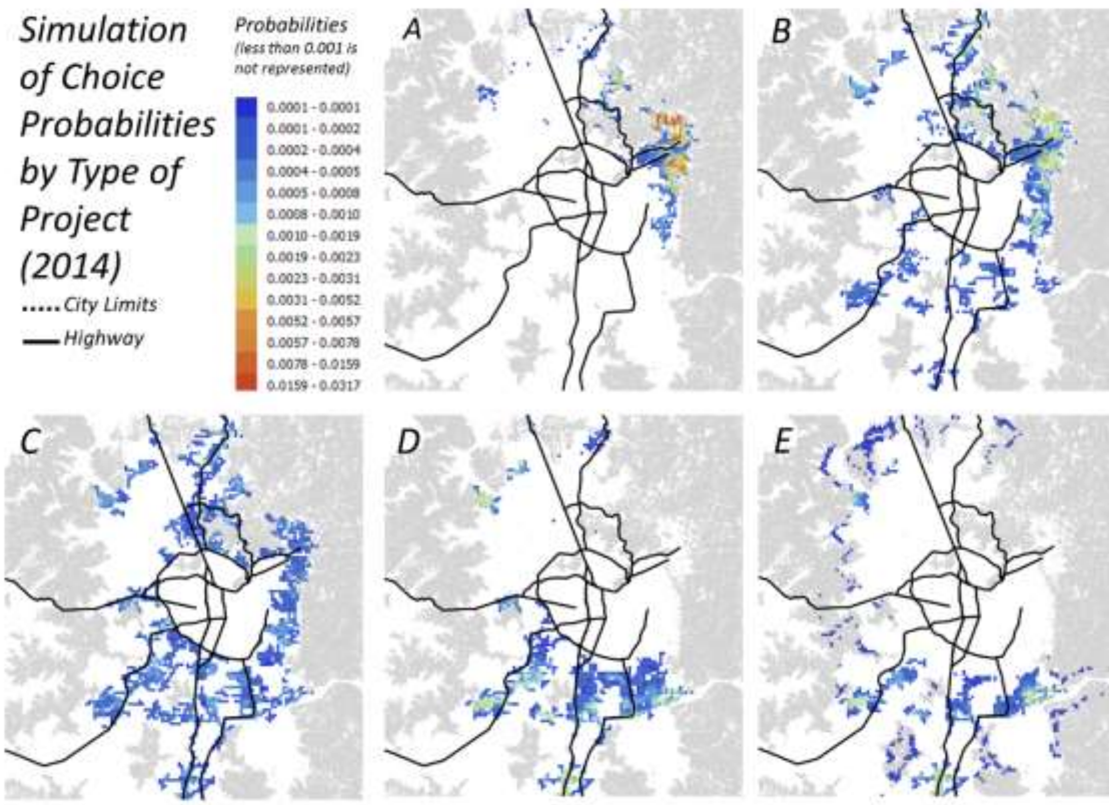

Figure 8: Probabilities of developing a project of each type in each cell. Cells with probability less than 0.0001 are not represented (colored in white). Source: own elaboration.

General structure of the segregation pattern of Santiago is reproduced (high price projects in north east area, middle income groups spread around more centric locations (around outer ring A. Vespucio), and lower price projects in south-west peripheries), but also the “spillage” of high price projects from north east zone to Chicureo (upper north area in the map).

We can see that higher price projects have more certainty of where to locate (they concentrate in fewer cells, each with higher probability), whereas middle price projects would show a flatter probabilities curve.

Low price projects are simulated mostly in low accessibility areas, deep in the valleys, but we observe that this kind of locations are mostly taken by high price projects. Including a variable that accounts for the proximity to mountain bases could solve this, but could also be an idiosyncratic variable (only adjusted to Santiago case).

Hoyt´s sector model (1939) can be clearly seen in the map, as higher probability of some type of projects distribute mostly in strips that grow outwards. It´s interesting to see that in areas like the northwestern quadrant the probabilities are very low (below 0.0001) for cells near the city, but going further away the probabilities increase. This shows the effects of proximity to lower income social groups, which dissipates with distance.

7. Conclusions

When modelling sprawl, zone-specific (or local) attributes become less relevant than distant relational attributes (modelled here through accessibility), mainly because these developments take place in areas that are not consolidated yet. Residential market is the main growth motor in metropolitan areas, and typical zonal amenities (schools, commerce) will follow its development, and not *viceversa*.

In this relational attributes, the role of accessibility to distant social elements (socioeconomic groups and other projects) is relevant, as the city has an underlying structure of vectors of growth of different well defined sectors, and is important for a residential project to be in "catchment areas" of certain social groups. Therefore, while the trade-off between land price and distance to CBD (or consolidated urban areas) explains sprawl to some extent, there are other important drivers such as search for an isolated or "exclusive location", away from influence areas of certain activities or socioeconomic groups (for mid-price projects), and while its possible, the location in the influence area of high price projects or higher status social groups (for higher price projects). The relations explored here are consistent with the assert that housing developers tend to be very conservative; even if they are trying to locate in unconsolidated places, they try to be near other similar projects that have proved to be successful.

The relevance of accessibility (in this case not only to the CBD but to more distributed destinations) and how much this attribute is determined by the presence of road infrastructure confirms the widely-analyzed trend of the relation between increase in road capacity and urban. The highway typology, in this case, particularly favors the sprawl drivers presented, as its enclosure helps to bypass non-desired attributes of the immediate context, but connecting with areas that are desired (high income and well developed zones). These results confirm that development of new transport infrastructure, particularly highways and urban freeways, becomes a significant driver of sprawl since it decreases travel costs to the periphery, as already described in the literature (Brueckner and Fansler, 1983; Glaeser & Khan, 2004; McGrath, 2005; Baum-Snow, 2007; Su & DeSalvo, 2008; Wassmer, 2008).

The complex ecology of forces determining the location of projects is interesting in the Chilean case because land use regulations are not particularly strong, therefore the resulting built environment resembles quite closely the free forces of an unregulated market. This presents the opportunity to research these issues and contrast the Chilean experience with other countries.

On the land value variable, is interesting to note that the model didn´t work well with the variable alone. It was more significant when it was multiplied by the average house plot size for each type of project. This makes sense as it is possible that low price projects pay more for land, but as they develop in high density, they can distribute the price in more units. So is better to use an average plot price to account for the real land price paid per unit.

Further work includes validation of the model though reproduction of the assignation process from basis year based on the choice probabilities, which includes also exploration of path-dependency. Subsequently, different scenarios of infrastructure and regulation can be simulated to explore infrastructure impact on localization considering the accessibility to social groups and types of projects addressed here.

8. Acknowledgements

This article was originated in the context of the research "Model of residential supply location in Santiago peri-urban space" (project funded FAU No. 04/13), which aims to produce a model to explain and predict development trends of housing supply in the peri-urban area of Santiago. The authors thank Marcelo Bauza, of Inciti (http://www.inciti.com/), for his valuable help with data collection.

The authors also acknowledge the support provided by the Center for Sustainable Urban Development (CEDEUS) (National Commission for Scientific and Technological Research CONICYT / Financing Fund Research Centres in Priority Areas - FONDAP 15110020), and the Project AccesoBarrio – A web-based platform to evaluate and model urban accessibility (FONDECYT N°1150239).

9. 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.*

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

Baum-Snow, N. (2007). Did highways cause suburbanization?. *The Quarterly Journal of Economics*, 775-805.

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.

Brueckner, J. K., & Fansler, D. A. (1983). The economics of urban sprawl: Theory and evidence on the spatial sizes of cities. *The Review of Economics and Statistics*, 479-482.

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

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.

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.

Geurs, K. T., & Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography*, *12*(2), 127-140.

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

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*, 7-17.

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

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

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).

McGrath, D. T. (2005). More evidence on the spatial scale of cities. *Journal of Urban Economics*, *58*(1), 1-10.

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.

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.

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

Su, Q., & DeSalvo, J. S. (2008). The effect of transportation subsidies on urban sprawl. *Journal of Regional Science*, *48*(3), 567-594.

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.

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.

Wassmer, R. W. (2008). Causes of urban sprawl in the United States: auto reliance as compared to natural evolution, flight from blight, and local revenue reliance. *Journal of Policy Analysis and Management*, *27*(3), 536-555.

Annex 1: Correlation analysis

A common problem in urban variables is the monocentric structure of some cities (which is particularly true for the case of Santiago), leading to a strong correlation among gravitational accessibilities as destinations tend to be concentrated. In our case, having a big study area with expansion areas, this correlation could be higher.

Another potential issue is the use of variables related to new projects and variables related to socioeconomic groups, which could be correlated.

VARIABLES	Acc_Pr_A+B	Acc_Pr_D+E	Acc_ABC1	Acc_DE	t_t_Near_Hway	Acc_Avesp	Acc_Near_Industry	acc_Near_Urban	Density	%_ABC1C2	%_C3	Land_Value
Acc_Pr_A+B	1.00											
Acc_Pr_D+E	0.31	1.00										
Acc_ABC1	0.79	0.72	1.00									
Acc_DE	0.36	0.87	0.80	1.00								
t_t_Near_Hway	-0.27	-0.53	-0.44	-0.49	1.00							
Acc_Avesp	0.55	0.91	0.90	0.90	-0.52	1.00						
Acc_Near_Industry	0.20	0.64	0.59	0.71	-0.27	0.67	1.00					
acc_Near_Urban	0.46	0.56	0.73	0.74	-0.31	0.67	0.54	1.00				
Density	0.26	0.35	0.45	0.51	-0.16	0.45	0.26	0.72	1.00			
%_ABC1C2	0.47	-0.07	0.26	-0.04	0.10	0.07	-0.06	0.11	0.02	1.00		
%_C3	-0.02	0.33	0.19	0.40	-0.11	0.30	0.30	0.40	0.45	-0.21	1.00	
Land_Value	0.65	0.24	0.65	0.37	-0.15	0.47	0.21	0.52	0.48	0.35	0.13	1.00

Table 3: Correlation among main variables for year 2014. Correlations higher than |0.75| are highlighted.

Table 3 shows a significant correlation between Accessibility to projects and Accessibility so social groups. The only pair in these variables that shows low correlation is Accessibility to DE social groups and Accessibility to AB projects. So these two variables could be used together to represent both attractive and not attractive forces on the probability of locating a project.

Accessibility to the main outer ring (Américo Vespucio, *Acc_Avesp*) is strongly correlated to both accessibilities to projects and to households by social group, as in both cases they tend to represent where people mainly live.

Annex 2: Highway by year built

ID	Highway Name	Date Built	
		Year	Month
1	RUTA 68 PAJARITOS TUNEL ZAPATA	pre-2000	0
2	RUTA 5	pre-2000	0
3	AUTOPISTA DEL SOL	2001	6
4	SANTIAGO-LOS ANDES	2001	12
5	COSTANERA NORTE	2005	4
6	AUTOPISTA CENTRAL	2005	4
7	KENNEDY	2005	4
8	A VESPUCIO SUR	2006	4
9	AV. AMERICO VESPUCIO EXTERIOR	2006	8
10	ACCESO SUR	2008	2
11	RADIAL NORORIENTE	2008	2
12	TUNEL SN CRISTOBAL	2009	6

Table 4: Highway by built year.