Discrete choice and microsimulation methods for agent-based land use modeling

THÈSE N^{O} 5554 (2012)

PRÉSENTÉE LE 2 NOVEMBRE 2012

À LA FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT
LABORATOIRE TRANSPORT ET MOBILITÉ
PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Ricardo Daniel HURTUBIA GONZÁLEZ

acceptée sur proposition du jury:

Prof. T. Mountford, président du jury
Prof. M. Bierlaire, directeur de thèse
Prof. K. W. Axhausen, rapporteur
Prof. A. de Palma, rapporteur
Prof. F. Martínez, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2012

From the point of ignition
To the final drive
The point of the journey
Is not to arrive
—Neil Peart

To Olivia and "Evento", who are starting the journey…

and

To Francisca, who walks with me…

Acknowledgements

I would first like to thank my supervisor Michel Bierlaire for guiding me during this time, always with a smile and good humor. He is not only a great scientist, professor and beer connoisseur, but the living proof that you can be very good and serious at what you do while still being a fun and cheerful person. It was both an honor and a pleasure to work with him.

I'm thankful to the members of my thesis committee for their commentaries and contributions which helped to improve the quality of this document. Special thanks go to Francisco Martínez for introducing me to the fascinating world of research (more than ten years ago already) and for being a mentor and a good friend during all this time. Moreover, I would like to thank all my former professors at the Transport Division of the University of Chile, I wouldn't have reached this achievement without their original teachings and encouragement.

This thesis work was partially financed by the SustainCity project, funded by the European Commission, and the SUPat project, funded by the Swiss National Science Foundation. I would like to thank my colleagues in both projects for the inspiring discussions and exchange of ideas.

I would like to thank all my (former and current) colleagues at the Transport and Mobility Laboratory, for countless hours of lunch-time discussion, work-related discussion and SAT discussions. Thomas Robin and Javier Cruz were my first office-mates and provided great help at the beginning of my PhD with code-writing, model-estimation and practical hints for a newcomer in Lausanne. Bilge Atasoy also had the "privilege" of sharing an office with me and was always tolerant (noise-proof), friendly and of great help when I didn't know something (quite often) or needed feedback for an idea. Jingmin Chen was always there to patiently help me with anything computer-related. Special thanks go to Gunnar Flötteröd, who helped me through a significant part of my thesis work and who was a good friend and "wingman" during many evenings where brainstorming was performed under the influence of catalyzing beverages.

Bilal Farooq provided critical help by generating the synthetic population used in the models of Chapters 5 and 6. Amanda Stathopoulos was of great help by discussing with me the economic justification of some of the ideas shown in Chapter 5. Dimitris Efthymiou collaborated with me in the implementation of the UrbanSim model and the generation of the maps for Chapter 6 and I'm grateful for his help.

Marianne Ruegg has always been the most helpful person since day 0 at EPFL. I can't thank her enough for making my life (and the life of everyone else at the lab) easier.

Despite the distance, my life-long Chilean friends were an indirect, but important, emotional and psychological support during this time. I thank them all for their friendship.

My parents are the prime-movers of all my achievements. I would like to thank my father for transmitting to me his strong sense of self-improvement and hard-work ethics. My mother has been my support and backup all my life, despite me not always being the most considerate and thankful son. She is an example of determination, hard work and selflessness. I'm eternally grateful for her love and patience. I owe her everything I am.

I would like to thank Francisca for being my partner and supporting me all this time. She is the engine that moves me and that gives purpose to everything I do. Without her encouragement I would have never started many of the projects that now help me to be a happier person.

Finally I would like to thank my daughter Olivia, for just existing and being so incredibly awesome. You are the best thing that has ever happened to me.

Lausanne, October 2012

Abstract

This thesis describes methods to model the land use component of an urban system. Specifically, it proposes methods to model and simulate the location choice of agents (households or firms) and the formation of prices for real estate goods in a city. These methods are based on the application of two main tools: discrete choice models and microsimulation.

Modeling urban systems is extremely relevant for project evaluation and policy making, due to the expensive, large and often irreversible nature of interventions at the urban scale. However, this is a complex task because it involves several sub-systems (land use, transport and energy among others) together with a large number of heterogeneous, interacting agents. Location choice and price models are a fundamental component of land use models because they describe the dynamics of the city and the spatial distribution of agents and activities. These models are complex because of the large nature of the problem, the presence of nonlinearities due to location externalities and the quasi-unique nature of the traded goods (locations or dwellings). In general, land use models are often hard to implement due to the large amount of data required to model each and every sub-system, even more if the modeling approach is agent-based or disaggregated.

This thesis contributes to the field of land use modeling in four aspects. First, an analysis of the formation of the choice set in problems with a large number of alternatives. The analysis is focused on comparing methods to model the availability of each alternative with explicit choice set formation models. Results show that availability-modeling heuristics are useful when dealing with large choice sets, but may significantly deviate from the explicit model.

Second, a model for simultaneous estimation of location choice and real estate price is proposed. The model considers that each good in the market is traded in a latent auction, where the potential willingness to pay of all agents determines the transaction price. The proposed approach has the advantage of explaining prices as a function of the willingness to pay of the agents, therefore not being determined by the market conditions of the estimation period, as it happens with hedonic price models. Another advantage comes from the latent nature of the auction, which allows to estimate the model even when detailed data on transaction prices is not available. The model is estimated for the city of Brussels using a double measurement equation approach, allowing to estimate the location choice and the price model simultaneously. The proposed approach is compared with other methods, showing better results, especially when available price data is aggregated.

Third, a market clearing method for agent-based models is proposed. The approach takes into account the expectations of bidding agents as they observe (and react to) the real estate

market conditions. The model assumes that, after adjustment of their expectations, agents bid simultaneously for the available locations. This produces a higher level market clearing that determines the real estate price. The proposed method does not require to solve a fixed point problem to find an equilibrium (or market clearing prices) and thus does not require to group agents in clusters. This makes possible to calculate the expectation adjustment at an individual level, therefore making the approach suitable for application in a microsimulation framework. The proposed model is implemented for the city of Brussels and a simulation is performed for the 2001-2008 period. Results show that the model is able to reproduce the observed trends of spatial distribution of agents and real estate prices.
Finally, a case study of a full, integrated land use and transport microsimulation model is presented. The model, implemented in the urban simulation platform UrbanSim and the traffic microsimulator MATSim, is estimated and applied to the city of Brussels. The analysis of the case study is focused on the requirements and difficulties of implementing a full scale land use microsimulator, with a special focus on data collection, data processing, model estimation and calibration of the system. An analysis of the trade-off between level of details, implementation costs and quality of the results is also provided, identifying the major difficulties when implementing large scale urban microsimulation models.

Keywords: location choice, land use, market clearing, real estate, discrete choice, microsimulation.

Résumé

Cette thèse décrit plusieurs méthodes pour modéliser l'utilisation du sol dans un système urbain. En particulier, elle propose des méthodes de modélisation et de simulation du choix résidentiel des ménages et du choix de lieux d'implantation des entreprises, ainsi que de la formation des prix des biens immobiliers dans une ville donnée. Ces méthodes s'appuient sur l'application de deux outils principaux : les modèles de choix discrets et la microsimulation.

La modélisation des systèmes urbains est très importante pour l'évaluation de projet et la mise en place de politiques publiques. Ceci est dû au coût, à la taille et à la nature souvent irréversible des interventions à l'échelle urbaine. Par ailleurs, cette tâche est complexe car elle implique plusieurs sous-systèmes (utilisations du sol, transport, énergie, etc.) associés à un grand nombre d'agents hétérogènes et interagissant entre eux. Les modèles de choix résidentiel et d'implantation industrielle et de prix représentent un élément fondamental des modèles d'utilisation du sol en cela qu'ils décrivent les dynamiques de la ville et la distribution spatiale des agents et des activités. Ces modèles sont complexes par leur taille, par la présence de non-linéarité liées aux externalités de localisation et par à la nature quasi unique des biens échangés (terrains ou habitations). En général, les modèles d'utilisation du sol sont souvent difficiles à implémenter car beaucoup de données sont nécessaires pour chacun des sous-modèles. Ceci est encore plus vrai lorsque l'approche est désagrégée ou basée sur les agents.

Cette thèse contribue à la recherche sur les modèles d'utilisation du sol de quatre manières. D'abord par une analyse de la formation des ensembles de choix dans des problèmes où le nombre d'alternatives est élevé. L'analyse se concentre sur la comparaison de méthodes pour modéliser la disponibilité de chaque alternative à l'aide de modèles explicites de formation d'ensemble de choix. Les résultats montrent que les heuristiques utilisées dans la modélisation de la disponibilité sont utiles lorsqu'il s'agit de gérer des ensembles de choix de grande taille, mais peuvent significativement dévier du modèle explicite.

Deuxièmement, un modèle pour l'estimation simultanée des choix résidentiel/d'implantation industrielle et des prix immobiliers est proposé. Le modèle considère que chaque bien sur le marché est négocié lors d'enchères latentes, où la volonté potentielle de payer de chacun des agents détermine le prix de transaction. L'approche proposée a l'avantage d'expliquer les prix comme une fonction de la volonté de payer des agents, et non par les conditions de marché de la période d'estimation, comme c'est le cas pour les modèles de prix hédoniques. Un autre avantage provient de la nature latente de l'enchère, permettant d'estimer le modèle même lorsque les données détaillées sur les prix de transaction ne sont pas disponibles. Le

modèle est estimé pour la ville de Bruxelles à l'aide d'une approche par double équation de mesure, permettant d'estimer le choix résidentiel/d'implantation industriel et le modèles de prix simultanément. L'approche proposée est comparée avec d'autres méthodes et montre de meilleures résultats, particulièrement quand les données de prix disponibles sont agrégées.
Troisièmement, une méthode d'ajustement des prix du marché pour les modèles basés sur les agents est proposée. L'approche prend en compte les attentes des acheteurs potentiels lorsqu'ils observent et réagissent à l'état du marché immobilier. Le modèle fait l'hypothèse qu'après ajustement de leurs attentes, les acheteurs potentiels enchérissent simultanément pour les terrains et les habitations disponibles. Ceci produit un ajustement à un niveau supérieur des prix du marché qui détermine les prix immobiliers. La méthode proposée ne nécessite pas de résoudre un problème du point fixe pour trouver un équilibre (ou un prix ajusté du marché) et ainsi n'a pas besoin de regrouper les agents en groupe. Cela rend possible d'évaluer l'ajustement des attentes à un niveau individuel, et ainsi rend la méthode adaptée au cadre de la microsimulation. Le modèle proposé est mis en oeuvre pour la ville de Bruxelles et la période 2001-2008 a été simulée. Les résultats montrent que le modèle est capable de reproduire les tendances observées tant pour la distribution des agents que pour les prix immobiliers.
Finalement, une étude de cas d'un modèle complet et intégré de microsimulation de l'utilisation du sol et des transport est présenté. Le modèle, implémenté sur la plateforme de simulation urbaine UrbanSim et le microsimulateur de trafic MATSim, est estimé et appliqué à la ville de Bruxelles. L'analyse de l'étude de cas se concentre sur les conditions requises et les difficultés de la mise en oeuvre d'un microsimulateur de l'utilisation du sol à grande échelle, avec un accent particulier sur la collecte et le traitement des données, l'estimation du modèle et la calibration du système. Une analyse du compromis entre niveau de détail, coût de mise en oeuvre et qualité des résultats est aussi fournie, identifiant les difficultés principales de la mise en oeuvre d'un modèle de microsimulation urbain à grande échelle.

Mots-clés : choix résidentiel, utilisation du sol, prix du marché, biens immobiliers, choix discret, microsimulation.

Contents

List of Tables

List of Figures

1 Introduction

In the year 2010, for the first time in history, there was more people living in cities than in rural areas and, for the next four decades, all the worlds population growth is expected to take place in urban areas (United Nations, 2011). Thanks to agglomeration economies, cities are the places where most economic, social, cultural and political activities and interactions take place. However, the very same activities and interactions also generate externalities like congestion, pollution, crowding, crime and social segregation. Cities also consume a large amount of resources like land, energy and, because of the need to travel, time from the people living in them. The existence of externalities and the (many times inefficient) use of important resources makes the study of cities a relevant subject.

Understanding urban systems is extremely relevant for project evaluation and policy making, due to the expensive, large and often irreversible nature of interventions at the urban scale. However, this is a complex task because it involves several sub-systems (land use, transport and energy among others) together with a large number of heterogeneous, interacting agents. Economics and mathematical modeling are useful tools for this task because of their analytical and quantitative nature.

Land use models are an increasingly used tool for analyzing and forecasting the evolution of cities and evaluating the potential effects of urban interventions such as real estate developments, modifications to the transport system and changes in urban or transport policy. They are of particular relevance for the field of transport modeling, since long-term travel demand is explained in a large amount by the spatial distribution of agents and activities in a region. The distribution and agglomeration of agents (households and firms) describes trip generation and attraction in spatially disaggregate terms and can be used to generate origin-destination trip matrices or, in the case of agent-based travel models, can be used to define the location of the activities involved in the tours of traveling agents

Location choice and price models are fundamental components of land use models because they describe the dynamics of the city and the spatial distribution of agents and activities. These models are complex because of the large nature of the problem, the presence of nonlinearities due to location externalities and the quasi-unique nature of the traded goods (locations or dwellings).

The relevance of location of agents and activities within a city has been recognized by economics since von Thünen (1826) and evolved to become a main element in urban economics after the works of Alonso (1964), Mills (1967) and Muth (1969), among others, who described the spatial allocation process as the outcome of an equilibrated auction market in what is now known as the bid-rent or bid-auction theory. One of the major breakthroughs in the field of urban location modeling came with the work of McFadden (1978), which introduced the use of discrete choice analysis to model the choice of residential location, providing a robust theoretical framework, based on utility maximization assumptions, and the econometric tools required to estimate location choice models. Ellickson (1981), expanded McFadden's choice approach and applied it to a bid-auction framework, where the willingness to pay of agents define both their location and the prices of the real estate goods.

Most existing land use models use discrete choice analysis (either under the bid or choice approach) to model location choice (Wegener, 2004; Hunt et al., 2005). However, they differ in many aspects like the treatment of time (dynamic or cross-sectional), the level of aggregation in the representation of demand and supply (aggregate or agent-level) and, most important, in how market clearing is treated.

Cross sectional models are usually consistent with the equilibrium approach of traditional urban economics but they are only able to forecast for long term future scenarios where a steady state is assumed to be reached. This is a somehow constraining feature of these type of model since cities are dynamic systems where information and goods flow at different speeds and observing what happens while they evolve may provide relevant information. Dynamic models are able to capture the inter-temporal dependency of decisions and changes in the urban context and are, therefore, more realistic and detailed.

Models with an aggregate representation of supply or demand have the advantage of being simple to implement in computational terms and easier to reconcile with (equilibrium driven) urban economic theory. However, aggregation implies lack of representation of the heterogeneity found in both attributes and behavior of supply and demand. Agent-based models are capable of describing agents and their behavior with the highest level of details and, more important, to account for interactions at the individual level, therefore being potentially able to represent phenomena like individual driven dynamics, social-network effects or activity-based behavior. However, agent-based model are expensive to implement in computational terms and usually require more disaggregate data and more detailed sub-models to describe the behavior of agents.

Market clearing refers to the process through which the conflicts resulting from interaction between agents competing for locations are solved. This usually implies finding the vector of prices that allows for discrimination and matching between agents and locations. Market clearing can be modeled through the solution of equilibrium problems but this requires aggregation of agents and strong assumptions like perfect matching between supply and demand or perfect information of decision makers. Disaggregate models can deal with the market clearing problem in many ways but they are usually very expensive in computational terms (individual simulation of the negotiation process of all transactions) or unrealistic (ignoring market effects such as demand or demand surplus).

This thesis concerns several aspects of location choice modeling in the urban context with, a focus on applications that are dynamic, disaggregate and properly accounting for market clearing.

1.1 Contributions

This thesis contributes to the field of land use modeling in four aspects. First, an analysis of the formation of the choice set in problems with a large number of alternatives. The analysis is focused on comparing methods to model the availability of each alternative with explicit choice set formation models. Results show that availability-modeling heuristics are useful when dealing with large choice sets, but may significantly deviate from the explicit model by estimating different choice probabilities.

Second, a bid-auction model for location choice is proposed. The model considers that each good in the market is traded in a latent auction, where the potential willingness to pay of all agents determines the transaction price. The proposed approach has the advantage of explaining prices as a function of the willingness to pay of the agents, therefore not being exclusively determined by the market conditions of the estimation period, as it happens with hedonic price models. Another advantage comes from the latent nature of the auction, which allows to estimate the model even when detailed data on transaction prices is not available. The model is estimated for the city of Brussels using a double measurement equation approach, allowing to estimate the location choice and the price model simultaneously. The proposed approach is compared with other methods, showing better results when available price data is aggregate.

Third, a quasi-equilibrium framework for bid-auction models is proposed. The approach takes into account the expectations of bidding agents as they observe (and react to) the real estate market conditions. The model assumes that, after adjustment of their expectations, agents bid simultaneously for the available locations. This produces a higher level market clearing that determines the real estate price. The proposed method does not require to solve a fixed point problem to find an equilibrium (or market clearing prices) and thus does not require to group agents in clusters. This makes it possible to calculate the expectation adjustment at an individual level, therefore making the approach suitable for application in a microsimulation, agent-based, framework. The proposed model is implemented for the city of Brussels and simulations are run for the 2001-2008 period. Results are compared with observed data, showing that the framework is able to predict trends in real estate prices and socioeconomic dynamics.

Finally, a case study of a comprehensive integrated land use and transport microsimulation model is presented. The model, implemented in the urban simulation platform UrbanSim (Waddell, 2002) and the traffic microsimulator MATSim (Rieser et al., 2007), is estimated and applied to the city of Brussels. The analysis of the case study is focused on the requirements and difficulties of implementing a full scale land use microsimulator, with a special focus on data collection, data processing, model estimation and simulation. As a validation exercise, results are analyzed and compared with real data and with the results obtained with the model proposed in Chapter 5 of this thesis.

1.2 Thesis outline

The thesis is organized in 7 chapters including this introduction. The remaining chapters are described next including (when necessary) a reference to the corresponding publication:

- Chapter 2 reviews the application of random utility models to the particular problem of location choice. It focuses on the microeconomic derivation of the problem, the different possible approaches that can be taken and the management of technical difficulties like dealing with large choice sets.

- Chapter 3 analyzes and compares different methods for choice set formation in discrete choice models. The analysis is focused in comparing explicit methods with heuristics that approximate the choice set formation process. This chapter and the preliminary analysis that leads to it has been published as:
  Bierlaire, M., Hurtubia, R. and Flötteröd, G. (2009). A comparative analysis of implicit and explicit methods to model choice set generation. *Proceedings of the 9th Swiss Transport Research Conference*, Ascona, Switzerland.
  Bierlaire, M., Hurtubia, R. and Flötteröd, G. (2010). Analysis of the implicit choice set generation using a Constrained Multinomial Logit model, *Transportation Research Record* 2175: 92-97.

- Chapter 4 proposes a method to estimate bid function for price estimation and location choice modeling under a bid approach. The method uses a latent variable approach to model the real estate price as the expected maximum bid of auctions that determine the location of agents and prices simultaneously. Estimation results are validated and compared with those obtained using other method. This chapter and the preliminary work that lead to it has been published as:
  Hurtubia R., Martinez F. and Bierlaire, M. (2011). Bid auction model for simultaneous determination of location and rent in land use microsimulation models. *Proceedings of the XV Congreso Chileno de Ingeniería de Transporte*, Santiago, Chile.
  Hurtubia, R., and Bierlaire, M. (2012). Estimation of bid functions for location choice and price modeling with a latent variable approach. *Technical report TRANSP-OR 120206.* Transport and Mobility Laboratory, Ecole Polytechnique Fédérale

- Chapter 5 proposes a modeling framework for market clearing in location choice problems considering agents at the individual level. A preliminary version of this model has been published in
  Hurtubia, R., and Bierlaire, M. (2011). Bid rent model for simultaneous determination of location and rent in land use microsimulations. *Proceedings of the 11th Swiss Transport Research Conference*, Ascona, Switzerland.
  Hurtubia, R., Bierlaire, M., and Martinez, F. (2012). Dynamic microsimulation of location choices with a quasi-equilibrium auction approach. *Proceedings of the 12th Swiss Transport Research Conference,* Ascona, Switzerland.

- Chapter 6 describes the implementation of a comprehensive integrated land use model for the city of Brussels. The chosen platform for this is the agent-based land use simulator UrbanSim (Waddell, 2002) coupled with the agent-based transport simulator MATSim (Rieser et al., 2007). The chapter is focused on describing the required efforts to implement such a model and compares the obtained results with those achieved with the model proposed in Chapter 5.

- Chapter 7 concludes the thesis and identifies possible further research.

2 Random utility models applied to location choice

In the urban context, location choice can be described as the process of matching agents (households, and firms) with locations (dwellings, buildings or real estate in general). The discrete nature of the choice process makes random utility models (also known as discrete choice models) an appropriate tool to estimate, understand and forecast location choice. Domencich and McFadden (1975) first proposed to use discrete choice analysis to model residential location as a way to predict the availability of different transport modes in the context of travel demand modeling. Applications to real case studies of residential location came soon after (Quigley, 1976; Lerman, 1977; Anas, 1981), proving the feasibility and usefulness of the approach for a broader set of purposes. The work of McFadden (1978) is of particular relevance because it provides a comprehensive theory of stochastic housing location choice and the tools to model it.

Location choice is different from other choice problems in two main aspects: the particular nature of the involved demand and supply and the large size of the problem.

The location choice problem is characterized by an almost inelastic demand: all agents need to locate somewhere in order to live or perform their productive activities and, in general, can not afford to opt-out of the market. Moreover, the choice of location is generally a long-term decision, involving large amounts of resources meaning that, at least in the residential context, agents usually "consume" only one unit. From the supply side, the quasi-unique attributes of locations (two locations can't have the same position in space) and the high cost of development makes real estate goods scarce by nature.

Depending on the focus given to the problem (supply or demand oriented), the location choice can be understood under two main paradigms: the Choice approach and the Bid-auction approach. Under the Choice approach (McFadden, 1978; Anas, 1982), agents select the location that maximizes their utility under the assumption that they behave as price takers. The Bid-auction approach (Ellickson, 1981) assumes that real estate goods are traded in an auction market, where the best bid for a particular location determines both the located agent and the price or rent of the good. Each approach has advantages and disadvantages that are analyzed later in this chapter.

Besides the unique nature of its supply and demand, the location choice problem is a large one. In the urban context the number of possible locations and agents looking for them is generally big. This represents a modeling difficulty from a technical point of view because of the difficulties of estimating discrete choice models with a large set of alternatives. The large nature of the problem also rises questions regarding the behavior of decision makers and how many alternatives they take into account when choosing a location: it is unlikely that, for example, a household will consider each and every possible dwelling before relocating.

This chapter reviews how random utility models are applied to the specific problem of location choice. It analyzes the Choice and Bid-auction approaches in sections 2.1 and 2.2 respectively and compares them in Section 2.3. Section 2.4 reviews methods to deal with problems where the number of alternatives is large as it is the case of location choice. Finally, section 5.5 concludes the chapter by identifying the advantages and disadvantages of each of the discussed approaches and methods.

2.1 The Choice approach to location choice

The location choice problem can be expressed as the classical microeconomic problem of a consumer choosing the location of maximum utility, given income constraints:

$$\max_{x,i} U(x, z_i) \tag{2.1}$$

$$s.t.\; px + r_i \leq I$$

In the previous problem, the consumer maximizes his utility by choosing a vector of continuous goods (x) and a discrete location (i), described by a set of attributes (z_i). The budget constraint states that the total amount spent in goods (with price p) plus the price of the selected location (r_i) must be smaller that the consumer's available income (I). Solving the problem on x and assuming equality in the budget constraint, the problem can be re-written as (Rosen, 1974):

$$\max_{i} V(p, I - r_i, z_i) \tag{2.2}$$

where V is the indirect utility function, conditional on the the location. Notice how the utility depend on both the location attributes and the location price. If we consider that consumers (h) have different preferences and that location alternatives (i) have unobserved attributes, we can rewrite the utility of (2.2) as $V_{ih} + \varepsilon_i$, where ε is an stochastic error term that account for the unobserved attributes of locations. This allows to express the probability of an agent h choosing a particular location i as the probability of that location being the one providing maximum utility:

$$P(i|h) = Prob\left\{V_{ih} + \varepsilon_i > V_{jh} + \varepsilon_j, \forall j \neq j\right\} \tag{2.3}$$

Assuming an iid Extreme Value distribution for the error term of the utility function, the probability of a household h choosing a location i is (McFadden, 1978):

$$P(i|h,S) = \frac{\exp(\mu V_{ih})}{\sum\limits_{j \in S} \exp(\mu V_{jh})} \tag{2.4}$$

where S is the set of available locations and μ is a postive scale parameter. The logit probability of (2.4) describes the spatial distribution of agents under the choice approach and is the main equation behind most location choice models in the literature (Pagliara and Wilson, 2010).

Given the dependency on price described by (2.2), the main behavioral assumption under the choice approach is that decision makers are price takers. Real estate prices or rents (r_i) can be modeled as a function of attributes of the location, following the hedonic framework proposed by Rosen (1974) or, when possible, as the result of the interaction (equilibrium) between supply and demand (Anas, 1982). Because prices depend on the attributes of the locations it is likely that they will be correlated with other variables of the utility function, generating endogeneity problems. Many empirical estimations of location choice models have reported this problem (Quigley, 1976; Waddell, 1992; Bhat and Guo, 2004; Guevara and Ben-Akiva, 2006) finding price parameters that are not significant or even positive.

2.2 The Bid-auction approach to location choice

Since Alonso (1964), the real estate market has been understood as an auction market, where households bid their willingness to pay for a particular good (residential unit) which is assigned to the best bidder. This process simultaneously defines the price of the good, understood as the maximum bid in the auction process.

The willingness to pay, from an economic point of view, can also be derived from the consumer's problem of maximum utility given income constraints. Following the derivation of the previous section (see equation 2.1), the consumer's problem is reduced to choosing the location (i) that provides maximum utility:

$$\max_i V(p, I - r_i, z_i) \tag{2.5}$$

Conditional on a fixed level of maximum utility that can be obtained ($\overline{U}$), the indirect utility can be inverted in the rent variable (Jara-Díaz and Martínez, 1999):

$$r_i = I - V^{-1}(\overline{U}, p, z_i) \tag{2.6}$$

Under the auction market assumption, the rent variable can be understood as the willingness to pay for a particular location, therefore the bid function B can be expressed as:

$$B_{hi} = I_h - V_h^{-1}(\overline{U_h}, p, z_i) \tag{2.7}$$

In the bid function, the index h has been included to take into account heterogeneity in preferences within different households. The bid, or bid-rent, function can then be understood as the maximum rent (or price) a household (or any other agent) can pay for a particular dwelling (location), while enjoying a fixed utility level $\overline{U_h}$ (Fujita, 1989). If we assume fixed prices of other goods (p) and a quasi-linear structure of the inverse utility function it is possible to separate the bid function into components that are specific to the decision maker and the location (Martinez, 2000):

$$B_{hi} = I_h - f_h^1(\overline{U_h}) - f_h^2(z_i) \tag{2.8}$$

This form of the bid function explicitly shows that the willingness to pay will be higher for households of higher income, but it will also depend on the overall utility level the household expects to achieve ($\overline{U_h}$) and the attributes of the location (z_i). This result is important because it indicates that, in equilibrium and for a given utility level $\overline{U_h}$, the household will be indifferent on choosing any location, as long as the price paid for it is B_{hi}. In other words, the household will be indifferent to locate in any place where he is the best bidder. It also indicates that, if necessary, the household can homogeneously increase (or decrease) his bid for all locations at the expense (or benefit) of his perceived utility. For example, if a household is unable to win any auction, it may reduce the desired utility level and increase all bids for all alternatives, in order to ensure being located somewhere. The opposite should happen if a household is systematically the best bidder in several locations.

Ellickson (1981) showed that the bid defined by (2.7) can also be written directly as a function of the location attributes (z_i) and proposed to account for the unobserved heterogeneity in preferences across households by adding a random error term:

$$\tilde{B_{hi}} = B_h(z_i) + \varepsilon_h = B_{hi} + \varepsilon_h \tag{2.9}$$

It is relevant to notice that under the choice approach the error term accounts for unobserved attributes in the location while the error terms of equation (2.9) does it for unobserved characteristics of decision makers (households).

The probability of a residential unit or location i being occupied by h is the probability of that particular household being the best bidder for the location among all the other bidding households:

$$P_{h/i} = Prob\left\{ B_{hi} + \varepsilon_h > B_{h'} + \varepsilon_{h'},\, \forall h' \neq h \right\}$$

If the error terms follow an iid Extreme Value distribution, the best bid probability can be expressed as a Logit model:

$$P(h|i,H) = \frac{\exp(\mu B_{hi})}{\sum\limits_{g \in H} \exp(\mu B_{gi})} \tag{2.10}$$

where H is the set of bidding households. Under the auction market assumption, the price or rent (r_i) of a good will be the maximum bid which, in a stochastic setting, can be expressed as the following expectation:

$$r_i = E\left(\max_h(\tilde{B_{hi}})\right) \tag{2.11}$$

The Extreme Value distribution assumption allows to express the expected maximum bid for a particular location as the logsum of the bids (Ben-Akiva and Lerman, 1985):

$$r_i = \frac{1}{\mu} \ln \left(\sum_{g \in H} \exp(\mu B_{gi}) \right) + \frac{\gamma}{\mu} \tag{2.12}$$

where γ is Euler's constant. In estimation, the bids can be identified only up to an (unknown) constant. This happens because the Logit model is under-identified and, while relative bids are enough to calculate the best bidder probability of (2.10), they do not necessarily relate to (absolute) real prices or rents. This means the term $\frac{\gamma}{\mu}$ will be absorbed by this constant and, therefore, equation (2.12) can be re-written as:

$$r_i = \frac{1}{\mu} \ln \left(\sum_{g \in H} \exp(\mu B_{gi}) \right) + C \tag{2.13}$$

where C is an unknown constant indicating that the absolute value of the bids cannot be measured from the observation of located households (best bidders).

An important feature of the use of the expected maximum bid as the price is the possibility of writing the maximum bid probability of (2.10) as a function of the price, following:

$$P(h|i) = \exp\left(\mu(B_{hi} - r_i + \gamma)\right) \tag{2.14}$$

The relationship defined by equations (2.13) and (2.14) is one of the main advantages of the bid approach: prices are explicitly dependent on the preferences of all households involved in the market and vice versa. Therefore, the price formation process is explicit and may be affected by market conditions like the levels of supply or demand surplus. A hedonic model (usually used to describe prices in a choice approach) estimates prices by assigning a marginal value to each attribute of a location, independent of the involved agents and therefore being less sensitive to market conditions like demand or supply surplus.

2.3 The bid-choice equivalence

It is possible to demonstrate that, under the assumption of an auction market and equilibrium conditions, the location where the agent is the highest bidder is also that of the maximum surplus or maximum utility (Martinez, 1992, 2000). This assures that the auction outcome yields an allocation consistent with maximum utility behavior of consumers. The consumer surplus is defined as the difference between the willingness to pay for a good and the actual price of the good. If the utility is written in terms of consumer surplus it will take the following form (Martinez, 1992):

$$V_{hi} = B_{hi} - r_i \tag{2.15}$$

Replacing (2.15) in (2.4), the probability of a household h choosing a location i is:

$$P(i|h) = \frac{\exp(\mu(B_{hi} - r_i))}{\sum\limits_{j \in S} \exp(\mu(B_{hj} - r_j))} \tag{2.16}$$

If prices are the outcome of an auction process and the market clears, the distribution of households across locations obtained through (2.16) will be the same as the distribution obtained from (2.10). Assuming that the scale parameter is the same for the bid and choice probabilities and replacing (2.14) into (2.16), we obtain the following relationship:

$$P(i|h) = \frac{P(h|i)}{\sum\limits_{j \in S} P(h|j)} \tag{2.17}$$

Under classic economic equilibrium conditions (supply satisfying demand), the location of each and every agent is guaranteed. This translates into the sum of the best bid probabilities over every location being equal to one for each agent, or $\sum\limits_{j} P(h|j) = 1 \, \forall h$. Under this conditions, the bid-choice equivalence is achieved because, from (2.17), $P(i|h) = P(h|i) \, \forall h, i$.

The same equivalence can be directly obtained from Bayes theorem by expressing one conditional probability as a function of the other:

$$P(h|i) = \frac{P(i|h)P(h)}{P(i)} \tag{2.18}$$

The marginal probabilities $P(h)$ and $P(i)$ can be interpreted as the probability of a household h being located somewhere and the location i being used by someone, respectively. Again, classic equilibrium conditions ensure a perfect match between supply and demand, meaning that $P(h) = P(i) = 1$ and, therefore $P(h|i) = P(i|h)$. The advantage of this derivation of the equivalence is that it does not require to impose the same value for the scale parameter in the bid-auction and choice approaches.

This result is important because it indicates that the spatial distribution of agents can be properly modeled with either approach and, therefore, the selection of one particular method should be based on other considerations, like data availability or the desired approach for price modeling.

2.3.1 Hedonic price versus expected maximum bid

Besides the equivalence in terms of spatial distribution, the estimated rents obtained under the choice or bid approach should also be equivalent under equilibrium conditions (Wheaton, 1977). Following Rosen (1974)'s approach, real estate prices or rents can be expressed as a function of the attributes of the location ($r_i = f(z_i)$). The most common form for a hedonic price model is a linear in parameters function:

$$r_i = \sum_k \alpha_k z_i^k \tag{2.19}$$

where k is an index for the k^{th} attribute of the location. The parameters in a hedonic prices model can be interpreted as the market value of each of the attributes:

$$\alpha_k = \frac{\partial r_i}{\partial z_i^k} \tag{2.20}$$

Under the assumption of an auction market (bid approach), the market value for each of the attributes (that is, the price at which this attribute would be bought) can be expressed as the derivative of the logsum (equation 2.13) with respect to the attribute. Since the attributes appear in the bid function of each household, the derivative takes the following form

$$\frac{\partial r_i}{\partial z_i^k} = \sum_{h \in H} \left(\frac{\partial \left(\ln \left(\sum_g \exp(B_{gi}) \right) \right)}{\partial B_{hi}} \cdot \frac{\partial B_{hi}}{\partial z_i^k} \right) \tag{2.21}$$

If the bid function is also linear in parameters ($B_{hi} = \sum_k \beta_h^k z_i^k$) we have:

$$\frac{\partial r_i}{\partial z_i^k} = \sum_{h \in H} \left(P(h|i) \beta_h^k \right) \tag{2.22}$$

Therefore, if the prices are the outcome of an auction, the standard hedonic model will be an approximation of the maximum expected bid, where the parameter α tries to reproduce a weighted average of the individual households preferences β. Assuming that land is sold in auctions we conclude that there is a direct mapping between the consumer's utility functions and the corresponding hedonic rent functions. This implies first, that only a subset of functions are supported by the bid-auction theory and can be used for hedonic rent models and, second and perhaps more constraining, that once a utility function has been chosen to model location probabilities, the corresponding hedonic rent function is identified; any other function is not supported by the theory.

Moreover, the equivalence is only true when estimating hedonic or bid functions for a cross-sectional price dataset and, therefore, under the assumption of equilibrium. Even if equivalent at estimation, hedonic and maximum bid models will generate different forecast for prices. An analysis of the differences in prices obtained when modeling with each approach can be found in Hurtubia et al. (2010).

2.4 Managing large choice sets

One of the essential characteristics of the location choice problem is the large number of alternatives involved. The number of possible locations in the urban context is usually big as is the number of agents looking for a location. This problem has different implications and solutions depending on the point of view: behavioral or technical

From a behavioral point of view, it can be assumed that decision makers do not consider the universal choice set when selecting an alternative but, through a choice set formation process, they reduce it to a subset of alternatives. From a technical point of view, the analyst needs to deal with the difficult problem of estimating and applying a model with a large choice set. In this case techniques like sampling are useful to reduce the computational complexity. Both choice set formation and sampling of alternatives techniques are described next.

2.4.1 Choice set formation

In smaller size discrete choice problems the analyst is assumed to be able to identify which alternatives are not available for each decision maker and remove them from the choice set. In the location choice problem, reducing the choice set to only feasible alternatives is difficult because the long term nature of the decision makes most alternatives feasible. However, the decision maker could still consider only a subset of the universal choice set. Some experimental evidence suggests that decision makers usually deal with large choice sets by aggregating or eliminating alternatives until a manageable size is reached (Miller, 1956; Tversky, 1972) and that humans can't consider an extremely large number of alternatives without suffering information overload (Malhotra, 1982). This has motivated the search of methods to model the choice set formation.

Manski (1977) proposed to model the consideration of different choice sets in a probabilistic way, enumerating the choice sets and estimating the choice probability conditional on each possible set of alternatives. This approach is difficult to implement in the context of location choice because, due to combinatorics, the large number of alternatives generates an even larger number of possible choice sets. Because of this, many heuristics that reduce the size of the choice set by eliminating alternatives have been implemented (for an extensive review of existing models see Shocker et al., 1991). The main problem of methods based on elimination alternative is that, besides obvious (and explicit) capacity or budget constraints, the set of (non)considered alternatives is not observable and, therefore, the definition of rules to remove alternatives is arbitrary. Some methods propose to estimate the thresholds at which the value of an attribute triggers the exclusion of the alternative from the consideration set (Cascetta and Papola, 2001; Swait, 2001b; Martínez et al., 2009). However, these models are difficult to estimate because they introduce additional nonlinearities and correlations in the utility function and it's not clear if they can be directly applied as an alternative to Manski's framework. Some evidence coming from the route choice literature (Frejinger et al., 2009) suggests that, in order to get unbiased results, it is better to take the universal choice set as consideration set and sample from it.

2.4.2 Sampling of alternatives

Even if the universal choice set has been reduced, it is still likely that the number of alternatives in the consideration set for a location choice problem will be large. Estimation of a discrete choice model over a large set of alternatives can be computationally expensive and sampling may be required.

McFadden (1978) proved that a logit model, like those of (2.4) and (2.10), can be consistently estimated on a subset of alternatives by adding an alternative-specific correction factor. Following this, the probability that a decision maker n chooses an alternative i given a subset D_n (defined by the modeler) of the universal choice set is:

$$P(i|D_n) = \frac{e^{\mu V_{in} + \ln \pi(D_n|i)}}{\sum_{j \in D_n} e^{\mu V_{jn} + \ln \pi(D_n|j)}} \tag{2.23}$$

where $\pi(D_n|i)$ the probability of constructing the subset D_n given that alternative i was chosen. Depending on the modeling approach, Bid-auction or Choice, D_n will be a subset of H or S respectively.

It is important to notice the difference between choice set generation and sampling of alternatives. While choice set generation reduces the number of alternatives in the final choice set by eliminating alternatives under the assumption that they are not considered by the decision maker, sampling of alternatives reduces the number of alternatives in the estimation process but still assumes that the full choice set is considered by the decision maker. While choice set generation implies behavioral assumptions, sampling of alternatives does not.

2.5 Conclusions

This chapter analyzed the different approaches and techniques that can be used to apply discrete choice models to the particular problem of location choice. Regarding the modeling approach (Bid or Choice) we conclude that the Bid approach is more appealing for the following reasons:

- It provides an explicit explanation of the price formation process as the outcome of an auction. The expected maximum bid is also a function of the preferences of all agents and, therefore, will depend on the socioeconomic composition and size of demand.
- Data collection efforts may be lower: a sample of agents can be more representative than an equally large sample of locations because it does not require to cover the whole space. This is due to the higher heterogeneity observed across locations than across individuals. In practical terms, this means that building a representative choice set under the Bid approach (a sample of households or firms) may be easier than doing it for the Choice approach
- Because price is not an explanatory variable in the bid-auction probability, it does not produce the price endogeneity problems usually found when using the choice approach.

Regarding management of large choice sets, it is clear from Section 2.4.1 that choice set formation is a complex problem that requires further analysis. This is done in the next chapter.

3 Methods for choice set formation: a comparative analysis

In standard choice models, it is assumed that the alternatives considered by the decision maker can be deterministically specified by the analyst. The choice set is characterized by deterministic rules based on the characteristics of the decision maker and the choice context. For example, locations that are too far are not considered in a destination choice context, car is not considered as a possible transportation mode if the traveler has no driver license, or no car.

There are, however, many situations where the deterministic choice set generation procedure is not satisfactory, or even possible. Data may be unavailable (the availability of car is unknown to the analyst), or rules are fuzzy by nature. For instance, train is not considered as a transportation mode if it involves a long walk to reach the train station. But how long is a "long walk"?

The situation is even more complex in the case of location choice models, due to the large nature of the involved choice sets. Households may not consider locations that are in some particular zones of the city or may reduce their consideration set to alternatives that are close to their current location or that have specific facilities required by them. For example, a family with children may not want to locate too far from schools but, again, the definition of when a location is "too far to be considered" is not clear. Since the set of considered alternatives is not observable, the analyst can only impose arbitrary or guessed rules to exclude alternatives from the choice set in a deterministic way.

Stochastic modeling of the choice set formation was first proposed by Manski (1977), who prosed to estimate the probability of considering each and every possible choice set. This approach is mathematically robust but very hard to implement in problems with more than a few alternatives. This motivated the development of heuristics that simulate the choice set formation by independently removing alternatives from the choice set when an attribute exceeds a threshold value. This type of model is feasible to implement in large choice sets but it is not clear if it is consistent with the explicit approach of Manski.

This chapter analyzes and compares models for choice set generation. Section 3.1 describes existing methods to do this while Section 3.2 describes a particular heuristic, the Constrained Multinomial Logit or CMNL (Martínez et al., 2009) that will be used for comparison purposes. In Section 3.3, we compare the CMNL with the theoretical framework of Manski (1977), first through a simple example and, second, by estimating both models on synthetic data. Finally, section 3.4 concludes the chapter.

3.1 Models for choice set formation

Modeling explicitly the choice set generation process involves a combinatorial complexity, which makes the models intractable except for some specific instances. (Manski, 1977) defines the theoretical framework in a two stage process, where the probability that decision maker n chooses alternative i is given by

$$P_n(i) = \sum_{\mathscr{C}_m \subseteq \mathscr{C}} P_n(i|\mathscr{C}_m) P_n(\mathscr{C}_m) \tag{3.1}$$

where $P_n(i|\mathscr{C}_m)$ is the probability for individual n to choose alternative i conditional on the choice set $\mathscr{C}_m$ and $P_n(\mathscr{C}_m)$ is the probability for individual n to consider choice set $\mathscr{C}_m$. The sum runs on every possible subset $\mathscr{C}_m$ of the universal choice set $\mathscr{C}$.

Swait and Ben-Akiva (1987) and Ben-Akiva and Boccara (1995) build on this framework and use explicit random constraints to determine the choice set generation probability. The probability of considering a choice set $\mathscr{C}_m$ is a function of the consideration of the different alternatives in the universal choice set:

$$P_n(\mathscr{C}_m) = \frac{\prod_{i \in \mathscr{C}_m} \phi_{in} \prod_{j \notin \mathscr{C}_m} (1 - \phi_{jn})}{1 - \prod_{k \in \mathscr{C}} (1 - \phi_{kn})} \tag{3.2}$$

where ϕ_{in} is the probability that alternative i is considered by user n, which may be modeled by a binary logit model that depends on the attributes of the alternative. Note that 3.2 assumes independence of the consideration probabilities across alternatives, which is a restrictive assumption since there can be correlation in the consideration criteria of different alternatives.

Swait (2001a) proposes to model the choice set generation as an implicit part of the choice process in a multivariate extreme value (MEV) framework, requiring no exogenous information. Here, choice sets are not separate constructs but another expression of preferences. The probability of considering a choice set is defined as the probability for that choice set to correspond to the maximum expected utility for an individual n:

$$P_n(\mathscr{C}_m) = \frac{e^{\mu I_{n,\mathscr{C}_m}}}{\sum_{\mathscr{C}_k \subseteq \mathscr{C}} e^{\mu I_{n,\mathscr{C}_k}}} \tag{3.3}$$

where μ is the scale parameter for the higher level decision (choice set selection) and $I_{n,\mathscr{C}_m}$ is the inclusive value (the "logsum" or expected maximum utility) of choice set $\mathscr{C}_m$ for decision maker n:

$$I_{n,\mathscr{C}_m} = \frac{1}{\mu_m} \ln \sum_{j \in \mathscr{C}_m} e^{\mu_m V_{nj}}. \tag{3.4}$$

Here, μ_m is the scale parameter and V_{nj} is the deterministic utility of alternative i for decision maker n. Swait's probabilistic choice set generation approach does not require assumptions by the analyst about which attributes affect an alternative's availability. Note that Swait's model also needs to account for every possible subset $\mathscr{C}_m$ of the universal choice set $\mathscr{C}$.

Clearly, these methods are hardly applicable to medium to large scale choice problems due to the computational complexity that arises from the combinatorial number of possible choice sets. If the number of alternatives in the universal choice set is J, the number of possible choice sets is $(2^J - 1)$.

In the context of route choice, Frejinger et al. (2009) assume that all decision makers consider the universal choice set, so that $P_n(\mathscr{C}_m) = 0$ when $\mathscr{C}_m \neq \mathscr{C}$, and only one term remains in (3.1). Results show that, by doing so, unbiased results are obtained. However, this may not be appropriate in other contexts, where searching costs may be high or decision makers have a limited search capacity.

Therefore, various heuristics have been proposed in the literature that derive tractable models by approximating the choice set generation process.

In the quantitative marketing literature, the use of heuristics to model the construction of the choice set (or consideration set) has been a usual practice; a review of existing models can be found in Hauser et al. (2009). Many heuristics are based on lexicographic preferences rules (Dieckmann et al., 2009), where the choice set is determined by key attributes of the alternatives on which the consumers base the construction of their consideration set. This approach is similar to the elimination by aspects heuristic, proposed by Tversky (1972). Models like the one proposed by Gilbride and Allenby (2004) consider the construction of the choice set as a two-stage process, which is consistent with Manski's approach but solves the choice set enumeration issue by using Bayesian and Monte Carlo estimation methods.

Other heuristics use a one stage approach (see for example Elrod et al. (2004)) where the choice set generation process is simulated through direct alternative elimination. This is done by setting the alternative's utility to minus infinity when certain attributes reach a threshold value. The alternative-elimination approach implies a different behavioral assumption from the two-stage approach, where the individual does not observe choice sets explicitly but, instead, makes a compensatory choice between all the alternatives belonging to a unique choice set of available or "possible" alternatives, which is a sub-set of the universal choice set.

Following the same one-stage approach, other heuristics assume that the elimination of the alternatives is not deterministic. These are based on the use of penalties in the utility functions, and have been proposed by Cascetta and Papola (2001) (the Implicit Availability/Perception (IAP) model) and expanded by Martínez et al. (2009) (the Constrained Multinomial Logit or CMNL model). In the next section, we briefly describe the CMNL model and provide its theoretical background in the context of choice set generation.

3.2 Choice set generation with the CMNL model

Assuming that $\mathscr{C}_n$ is the choice set that the decision maker is actually considering, the choice model is given by

$$P_n(i|\mathscr{C}_n) = \Pr\left(U_{in} \geq U_{jn}, \forall j \in \mathscr{C}_n\right), \tag{3.5}$$

where U_{in} is the random utility associated with alternative i by decision maker n. If $\mathscr{C}_n$ is known to the analyst, it can be characterized by indicators of the consideration of each alternative by the decision maker:

$$A_{in} = \begin{cases} 1 & \text{if alternative } i \text{ is considered by individual } n, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

The choice model can be equivalently written as

$$\begin{aligned} P_n(i|\mathscr{C}_n) &= \Pr\left(U_{in} \geq U_{jn}, \forall j \in \mathscr{C}_n\right) \\ &= \Pr\left(U_{in} + \ln A_{in} \geq U_{jn} + \ln A_{jn}, \forall j \in \mathscr{C}\right). \end{aligned} \tag{3.7}$$

For an unconsidered alternative, this adds $\ln 0 = -\infty$ to its utility, so that the choice probability is 0, whereas the addition of $\ln 1 = 0$ has no effect on the utility of a considered alternative.

In the case of a logit model, the choice probabilities are

$$P_n(i) = \frac{e^{V_{in}+\ln A_{in}}}{\sum_{j\in C} e^{V_{jn}+\ln A_{jn}}}. \tag{3.8}$$

In order to avoid taking the logarithm of zero, the previous equation can also be written as:

$$P_n(i) = \frac{A_{in}e^{V_{in}}}{\sum_{j\in C} A_{in}e^{V_{jn}}}. \tag{3.9}$$

The heuristics proposed by Cascetta and Papola (2001) and Martínez et al. (2009) consist in replacing the indicators A_{in} by the probability ϕ_{in} that individual n considers alternative i.

Cascetta and Papola (2001) introduce the IAP model as a way to incorporate awareness of paths into route choice modeling without requiring an explicit choice set generation step. A similar approach that penalizes the utilities of "dominated" alternatives is proposed by Cascetta et al. (2007).

Martínez et al. (2009) expand the IAP idea and propose the CMNL model. The functional form for ϕ_{in} is assumed to be a binary logit, considering that the availability of an alternative is related with bound constraints on its attributes. For example, if X_{ink} is the kth variable of alternative i for decision maker n that influences the consideration of i, we have

$$\phi^u_{in}(X_{ink}; u_k, \omega_k) = \frac{1}{1+\exp(\omega_k(X_{ink}-u_k))} \tag{3.10}$$

where the u_k parameter is the value at which the constraint is most likely to bind, and ω_k is the scale parameter of the binary logit. For instance, X_{ink} may be the walking distance to the train station, and u_k may be the maximum distance that individual n is willing to walk. Both u_k and ω_k are to be estimated or (if possible) defined by the analyst based on observable constraints. The intuition is that when the attribute X_{ink} exceeds u_k, the consideration probability ϕ^u_{in} tends to zero, while this availability tends to one when the value of the attribute is below u_k.

Expression 3.10 represents an upper value cut-off, where u_k represents the maximum value that the attribute X_{ink} can have in order for alternative i to be considered. To model a lower value cut-off, we only need to invert the sign of the scale parameter ω_k:

$$\phi^{\ell}_{in}(X_{ink};\ell_k,\omega_k) = \frac{1}{1+\exp(-\omega_k(X_{ink}-\ell_k))}. \tag{3.11}$$

Functions 3.10 and 3.11 can be generalized to account for more than one constraint, allowing for several upper and lower bounds to be included simultaneously:

$$\phi_{in}(X_{in};\ell,u,\omega) = \prod_k \phi^{u}_{in}(X_{ink};u_k,\omega_k)\phi^{\ell}_{in}(X_{ink};\ell_k,\omega_k). \tag{3.12}$$

The CMNL approach has an operational advantage over Manski's framework since it does not require enumerating the choice sets, which makes it easier to specify and estimate. However, the CMNL model is a heuristic that is based on convenient assumptions about the functional form of the utility function. This is why the CMNL model can at most be considered as an approximation to Manski's model. The next section evaluates the quality of this approximation.

3.3 Comparison of CMNL with Manski's model

This section compares the CMNL model with Manski's model. For this, we first present a simple example where we analyze the difference between the choice probabilities obtained using both models. Second, we estimate the CMNL model and Manski's model over synthetic data and compare the results. For notational simplicity, we subsequently omit the index n for the decision maker.

3.3.1 Simple example

Consider a logit model with only 2 alternatives, where alternative 1 is always considered ($\phi_1 = 1$) and alternative 2 has probability ϕ_2 of being considered by the decision maker. Figure 3.1 shows the structure of Manski's framework if we consider every possible combination of alternatives as a choice set. This simple situation corresponds to a case where the decision maker is captive to alternative 1 with probability $1-\phi_2$ (see also the captivity logit model proposed by Gaudry and Dagenais (1979)).

The CMNL model defines the probability of choosing alternative 1 as

$$P(1) = \frac{e^{V_1}}{e^{V_1}+e^{V_2+\ln\phi_2}}. \tag{3.13}$$

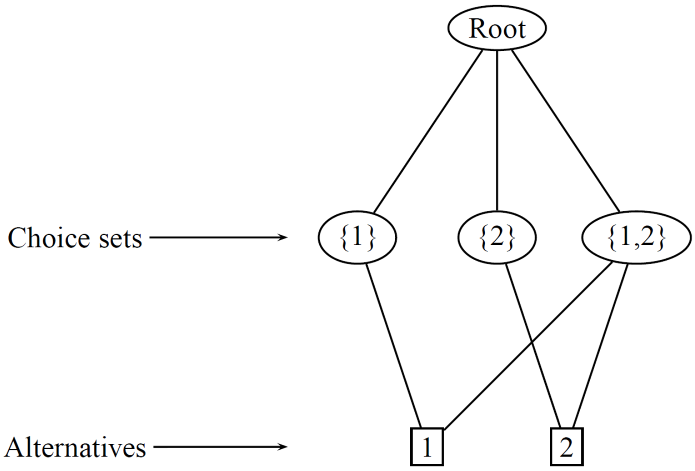


Figure 3.1: Example of a model in Manski's framework

Manski's model (3.1) defines the probability of choosing alternative 1 as

$$P(1) = P(\{1\})\frac{e^{V_1}}{e^{V_1}} + P(\{1,2\})\frac{e^{V_1}}{e^{V_1} + e^{V_2}} \tag{3.14}$$

where $P(\{1\})$ is the probability of considering the choice set composed only of alternative 1 and $P(\{1,2\})$ is the probability of considering the choice set containing both alternatives. According to (3.2), the choice set probabilities are

$$P(\{1\}) = \frac{\phi_1(1-\phi_2)}{1-(1-\phi_1)(1-\phi_2)} = 1-\phi_2 \tag{3.15}$$

and

$$P(\{1,2\}) \quad = \quad \frac{\phi_1\phi_2}{1-(1-\phi_1)(1-\phi_2)} = \phi_2. \tag{3.16}$$

The probability of considering choice set {2} is zero because alternative 1 is always available.

Therefore, 3.14 becomes

$$P(1) = (1-\phi_2) + \phi_2 \frac{e^{V_1}}{e^{V_1} + e^{V_2}} \tag{3.17}$$

In the deterministic limit ($\phi_2 = 0$ or $\phi_2 = 1$), both models are equivalent. However, this is not the case anymore when ϕ_2 takes values between zero and one. The resulting choice probabilities are shown in Figure 3.2, assuming the same utility level $V_1 = V_2$ for both alternatives.

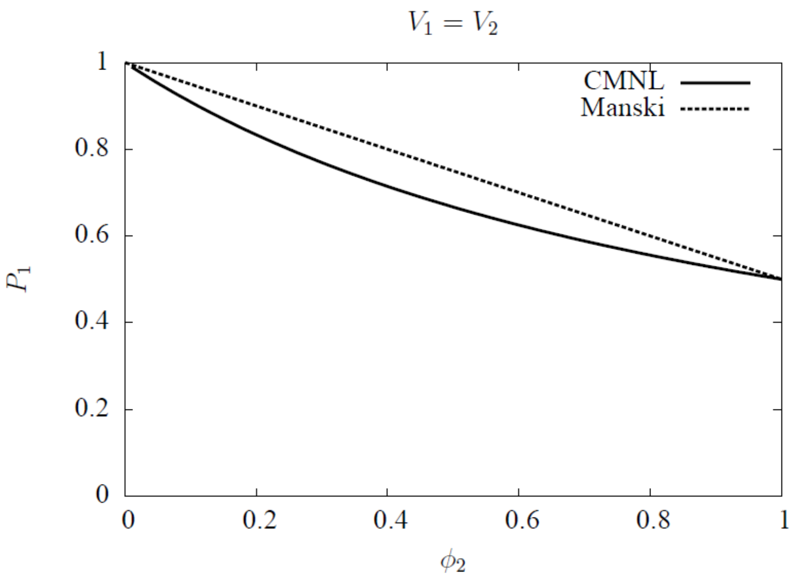


Figure 3.2: Choice probability of alternative 1 ($V_1 = V_2$)

This figure shows that the CMNL is a good approximation of Manski's model only when ϕ_2 is close to either zero or one, but it underestimates the probability of alternative 1 elsewhere. If the utility for alternative 1 is larger than the utility for alternative 2 (Figure 3.3), the approximation improves. This makes sense since the more an alternative is dominated, the less important it is to know if it really belongs to the choice set.

However, as the utility of alternative 1 becomes smaller and smaller compared to the utility of alternative 2, the CMNL becomes a poorer and poorer approximation of Manski's model for intermediate ϕ_2 values, which is demonstrated in Figures 3.4 and 3.5.

These results can be interpreted as an unwanted compensatory effect in the CMNL model. The availability constraint is enforced by modifying the utility of the constrained alternative. However, when the utility of this alternative is high, it compensates the penalty. This means that the use of the CMNL model as an efficient choice set generation mechanism requires the

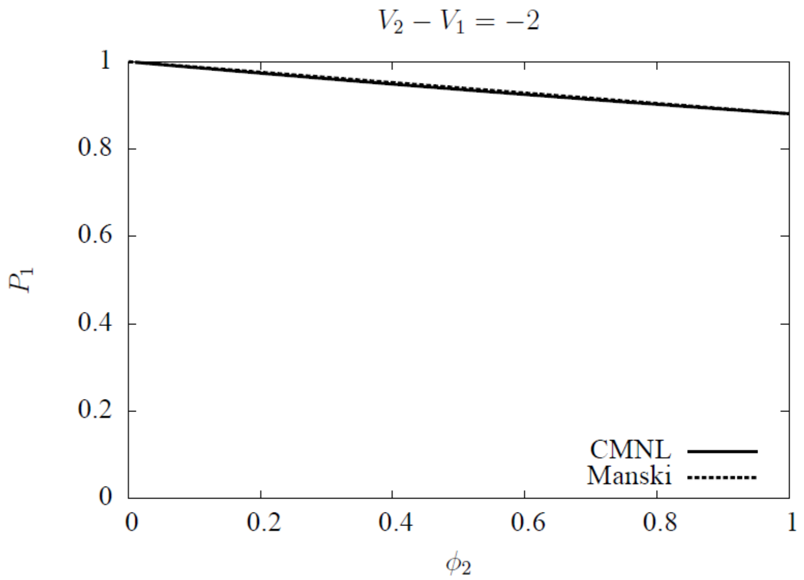


Figure 3.3: Probability of alternative 1 ($V_1 > V_2$)

assumption that the consideration probability for an alternative grows with its utility, meaning that the choice set depends only on the preferences of the individual. But alternatives with a high utility may be discarded in the presence of constraints such as budget or physical constraints. In the context of repetitive choices over a long period the individual may try to change her initial constraints in order to make the high-utility alternative available (for example, if the train produces high utility, a user may consider moving his residence closer to the train station), but in an instantaneous or short-term decision this may not be possible. This motivates to analyze the performance of the CMNL on synthetic data, which is shown in the next section.

3.3.2 Synthetic data

This section describes a series of controlled experiments where some of the data is synthetically generated. We start from a real stated preference data set that was collected for the analysis of a hypothetical high speed train in Switzerland (Bierlaire et al., 2001). The alternatives are:

1. Driving a car (CAR)
2. Regular train (TRAIN)
3. Swissmetro, the future high speed train (SM)

From this data set, which consists of 5607 observations, we use the attributes of the alternatives and simulate synthetic choices based on a postulated "true" model: a logit model with linear-

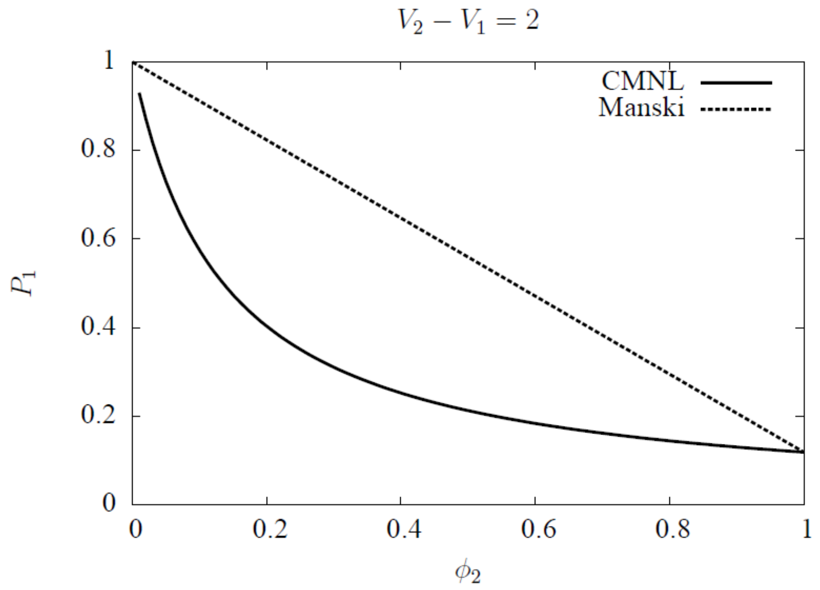


Figure 3.4: Choice probability of alternative 1 ($V_1 < V_2$)

in-parameters utility functions. The specification table as well as the "true" values of the parameters are reported in Table 3.1. The values have been obtained by estimating the model on real choices, and by rounding the estimates.

It is assumed that the TRAIN and the SM alternatives are always considered, whereas the consideration of the CAR alternative depends on the travel time according to

$$\phi_{\text{CAR}} = \frac{1}{1 + \exp(\omega(TT_{\text{CAR}}/60 - a))}, \tag{3.18}$$

which states that the probability of considering CAR as an available alternative decreases with the travel time TT_{CAR}, in minutes, and that this probability is 0.5 when the availability threshold a, in hours, is reached.

This implies that, depending on the availability of the CAR alternative, there are two possible choice sets: the full choice set and the choice set containing only the TRAIN and the SM alternative. The random constraints approach (Ben-Akiva and Boccara, 1995) defines the probability of each choice set as follows:

$$\begin{aligned} P(\{\text{TRAIN, SM}\}) &= \frac{\phi_{\text{TRAIN}}\phi_{\text{SM}}(1-\phi_{\text{CAR}})}{1-(1-\phi_{\text{CAR}})(1-\phi_{\text{TRAIN}})(1-\phi_{\text{SM}})} \\ &= 1-\phi_{\text{CAR}} \end{aligned} \tag{3.19}$$

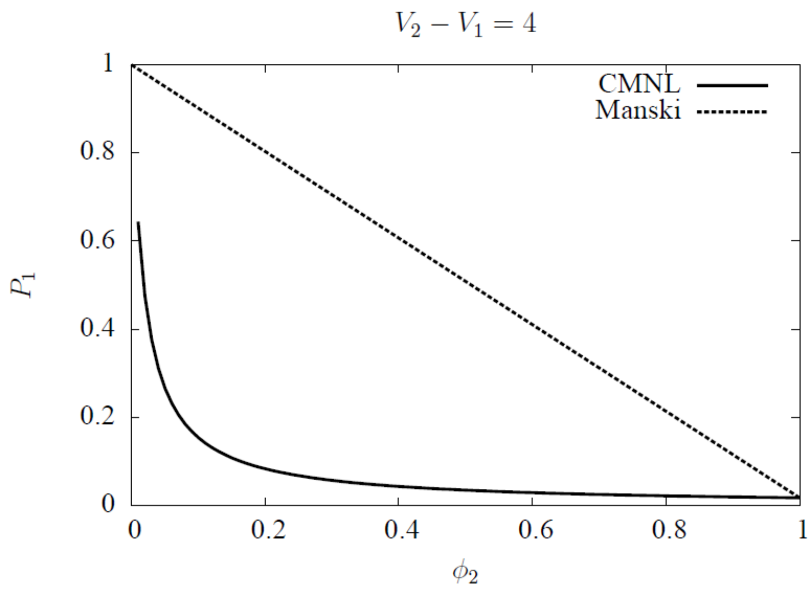


Figure 3.5: Choice probability of alternative 1 ($V_1 < V_2$)

and, accordingly,

$$P(\{\text{CAR, TRAIN, SM}\}) = \phi_{\text{CAR}}. \tag{3.20}$$

The synthetic choices are generated by (i) simulating a choice set for each decision maker according to (3.19) and (3.20), and (ii) simulating a choice for each decision maker using the "true" model specified in Table 3.1.

100 choice data sets are simulated for each value of ω. These values generate constraints with different levels of uncertainty. Figure 3.6 shows the shape of these constraint functions.

Table 3.1: Parameter descriptions and values

Parameter	Value	Car	Train	Swissmetro
ASC_{CAR}	0.3	1	0	0
ASC_{SM}	0.4	0	0	1
β_{cost}	-0.001	Cost (CHF)	Cost (CHF)	Cost (CHF)
β_{tt}	-0.001	In veh. travel time (minutes)	In veh. travel time (minutes)	In veh. travel time (minutes)
β_{he}	-0.005	0	Headway (minutes)	Headway (minutes)
a	3	Consideration threshold of car (hours)		
ω	1,2,3,5,10	Consideration dispersion of car		

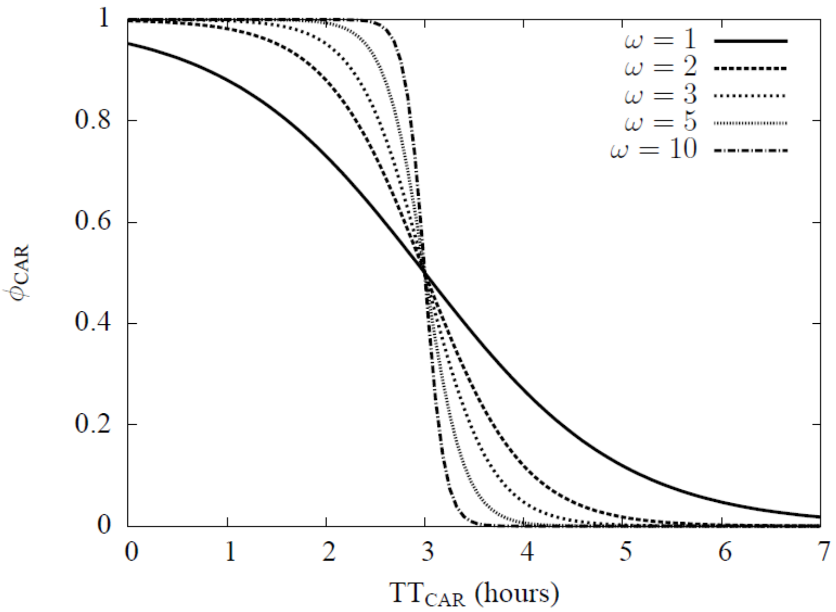


Figure 3.6: Shape of the constraint for different values of ω

Estimation results for both the Manski and the CMNL model are given in Tables 3.2 and 3.3. For each parameter β, the average value $\bar{\beta}$ and the standard error σ over 100 simulations are computed. In the tables, both $\bar{\beta}$ and the t-statistic $(\bar{\beta} - \beta)/\sigma$ are reported, the latter value being used to test if the estimated value is significantly different from the true one. Note that, since the tested hypothesis is that the average estimated value is equal to the "true" one, a low value of the t-statistic indicates that the estimate is not significantly different from the real parameter.

The estimates of Manski's model are unbiased. We cannot reject the hypothesis that the true value of any parameters is equal to the postulated value, at 95% level. Several estimates of the CMNL model are biased (marked with *), the hypothesis that the true value of the parameter is equal to the postulated value being rejected at the 95% level. The quality of the CMNL estimates improves with decreasing dispersion (increasing ω). This is consistent with the findings of Section 3.3.1.

Figures 3.7 and 3.8 shows the t-statistics for the cost and travel time parameter over different ω values for Manski's model and the CMNL model.

The quality of the estimates is constant across different values of ω for Manski's model. The quality of the CMNL estimates increases with ω, and their t-statistics reach acceptable values when the constraint function becomes steep.

Table 3.2: Estimation results for Manski's model

real ω value		1		2		3		5		10	
parameter	real value	estimate	t-test	estimate	t-test	estimate	t-test	estimate	t-test	estimate	t-test
ASC_{CAR}	0.3	0.304	0.027	0.288	0.113	0.300	0.010	0.301	0.012	0.314	0.184
ASC_{SM}	0.4	0.396	0.044	0.399	0.010	0.405	0.053	0.401	0.017	0.410	0.151
β_{cost}	-0.01	-0.010	0.283	-0.010	0.001	-0.010	0.179	-0.010	0.052	-0.010	0.012
β_{he}	-0.005	-0.005	0.241	-0.005	0.010	-0.005	0.048	-0.005	0.082	-0.005	0.078
β_{time}	-0.01	-0.01	0.074	-0.010	0.050	-0.010	0.049	-0.010	0.003	-0.010	0.001
a	3	2.963	0.019	3.008	0.118	3.000	0.100	2.998	0.081	3.002	0.101
ω	see top	1.003	0.028	2.014	0.079	3.066	0.210	5.095	0.170	10.523	0.353

Table 3.3: Estimation results for CMNL model

real ω value		1		2		3		5		10	
parameter	real value	estimate	t-test	estimate	t-test	estimate	t-test	estimate	t-test	estimate	t-test
ASC_{CAR}	0.3	0.503	0.950	0.421	1.153	0.406	1.365	0.380	0.988	0.326	0.313
ASC_{SM}	0.4	0.565	2.013 *	0.550	2.375 *	0.536	1.804	0.506	1.485	0.463	0.872
β_{cost}	-0.01	-0.008	4.825 *	-0.008	3.580 *	-0.009	2.309 *	-0.009	1.182	-0.010	0.613
β_{he}	-0.005	-0.005	0.202	-0.005	0.151	-0.005	0.071	-0.005	0.120	-0.005	0.090
β_{time}	-0.01	-0.007	3.929 *	-0.008	3.645 *	-0.008	2.813 *	-0.009	2.316 *	-0.009	1.523
a	3	2.186	1.753	2.656	3.073 *	2.773	3.762 *	-2.869	3.305 *	2.948	1.864
ω	see top	1.043	0.239	2.094	0.403	3.118	0.431	5.238	0.424	12.146	3.149 *

(* indicates a biased parameter)

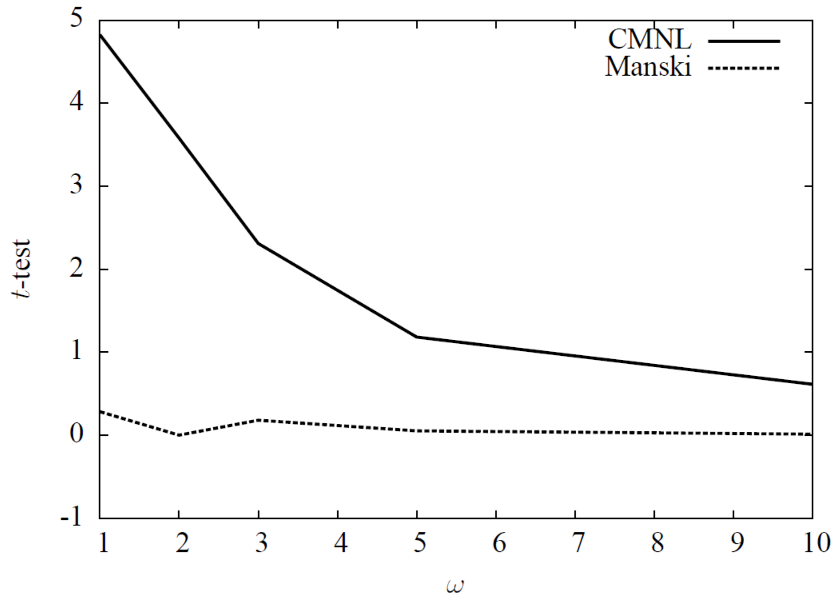


Figure 3.7: t-statistics for the cost parameter over ω

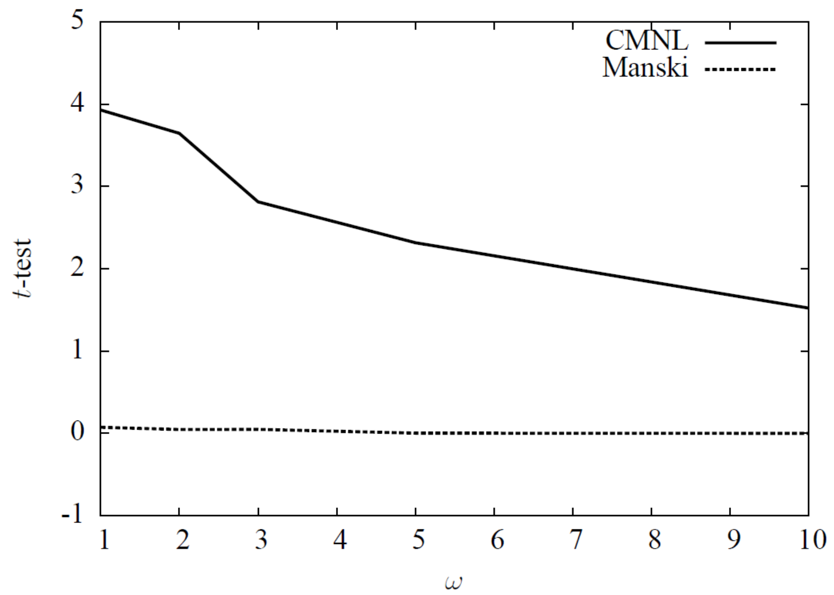


Figure 3.8: t-statistics for the time parameter over ω

3.4 Conclusions and further work

We have shown on simple examples that the Constrained Multinomial Logit (CMNL) model is not adequate to model the choice set generation process consistently with Manski's framework, except when the scale parameter of the considerantion probability is large and, therefore, approximates a discrete step function. Consequently, the CMNL model should be considered as a model on its own, derived from semi-compensatory assumptions as described by Martínez et al. (2009), but not as a way to capture the choice set generation process. Its complexity is linear with the number of alternatives, while Manski's framework exhibits an exponential complexity.

Model estimation with the CMNL, the IAP and other similar approaches, although feasible for larger choice sets, is still a difficult task even for not so large ones, especially when more than one threshold affects the availability of the alternatives (Martinez et al., 2008). This difficulty will be increased if the choice set is large.

The evidence presented by Frejinger et al. (2009) suggests that it is better to consider the universal choice set, sample from it and properly correct for sampling. These arguments allow to conclude that, until further analysis is performed, heuristic approaches like the CMNL should be used carefully in location choice problems

Further research should attempt to validate the CMNL in forecast scenarios of location choice for a real case study. Also, the derivation of a good approximation of Manski's model with the complexity of the CMNL would be particularly useful to handle models with a large number of alternatives.

4 Estimation of bid-auction functions

In land use models, forecasting location choice requires forecasting real estate prices and vice versa. It is widely accepted (Alonso, 1964; Fujita, 1989; Fujita et al., 1999; Glaeser, 2008) that location choice depends heavily on the prices of locations while, simultaneously, real estate prices are determined in part by the location preferences of agents. This endogenous interdependence makes the modeling of prices and location a particularly complex task.

Location choice and real estate prices have been traditionally modeled under two different main assumptions regarding the way the market operates: the choice approach and the bid-auction approach (See chapter 2).

In the field of urban economics, the bid-auction model has been used mostly as an alternative to hedonic models for the estimation of prices and marginal willingness to pay for attributes of real estate goods. The original model proposed by Ellickson (1981) considered an Extreme Value distribution of the willingness to pay that each agent has for a particular location. This generates a Logit model, conditional on the location, that can be estimated via maximum likelihood. The estimation process assumes that every located agent was the best bidder for the location. However, because of the under-determined nature of the Logit model, Ellickson's approach does not allow to find absolute estimates of the willingness to pay. It is only able to estimate relative rents and relative willingness to pay for groups of homogeneous agents.

Improving on Ellickson's work, Lerman and Kern (1983) proposed a method that maximizes the likelihood of an agent being the best bidder for his observed location while, simultaneously, maximizing the likelihood of his bid being equal to the observed transaction price. This method solves the original problem of under-determination in Ellickson's approach, generating absolute estimates of rents or prices and the associated willingness to pay for the location attributes. However, implementing Lerman and Kern's approach requires information that, in general, is not easy to collect: the actual price or rent paid for a particular real estate good, together with the attributes of the best bidder and the location or good itself . Moreover, as in the case of Ellickson's approach, the method imposes a simplification of the bid function, aggregating agents into homogeneous groups of bidders and estimating a single, linear in parameters, bid function for each of them. This simplifying assumption seems to have been originally motivated by computational constraints due to the need to enumerate the groups of bidders and should be easily ignored by current practice.

The simultaneous location choice and price estimation method of Lerman and Kern has been applied, among others, by Gross (1988), Gross et al. (1990), Gin and Sonstelie (1992), McMillen (1997) and Chattopadhyay (1998) to estimate bid-rent function in several urban case studies. The literature shows that, in general, the bid-rent generates more useful results than hedonic price models, thanks to the possibility of estimating the willingness to pay of different groups of agents and, therefore, providing information about consumer behavior. Despite this, the bid-auction approach has not been extensively applied due to a more complex estimation process than standard hedonic models and the already mentioned expensive data requirements. Moreover, the emphasis has been put in estimation of prices and marginal willingness to pay, giving little attention to the location choice distribution and with scarce validation of the resulting model when forecasting prices or locations. Muto (2006) analyzed location choice results when using Lerman and Kern's method, finding significant and systematic deviations in the results when compared with observed location distributions for the city of Tokyo. This result suggests that, while Lerman and Kern improve over Ellickson's model by estimating absolute rents, it does so at the cost of worse location forecast capabilities.

The bid-auction approach is particularly attractive for location choice modeling since it provides an explicit explanation of the market clearing process that generates the transaction prices (or rents in the case of the rental market) of real estate. This has motivated the development of several land use models that base their location choice process on the bid-auction approach. Examples of this are RURBAN (Miyamoto and Kitazume, 1989), MUSSA (Martínez, 1996), IRPUD (Wegener, 2008) and, to some extent, UrbanSim (Waddell et al., 2003). In these models, the bid-auction approach has been applied with a focus on modeling the spatial distribution of agents (households and firms) in a city, most of the times using Ellickson's approach to find the relative willingness to pay of different households for the attributes of a location. In these models, if done, the adjustment of the bid functions to absolute levels is achieved in the context of a market clearing process, separated from the original estimation.

Besides the theoretical appealing, the bid-auction approach is attractive for location choice modeling from an econometric point of view, because it does not have the price endogeneity problems usually found when using the choice approach. Price endogeneity occurs because the price is highly correlated with unobserved attributes of the location, therefore complicating the estimation of parameters. In the worst case, if descriptive attributes of the location are omitted, price endogeneity may lead to wrong estimates of the price elasticity and proper estimation will require the use of correcting mechanisms like the Control Function method (Guevara and Ben-Akiva, 2006). Because the price of the location does not enter the bid function as a variable, the bid-auction approach does not present price endogeneity issues.

The relevance and advantages of the bid-auction approach motivates the search for bid-rent estimation methods that allow for consistent estimation of both location choice (maximum bid probabilities) and price distributions without the need of individual level price data. At the same time it is interesting to explore the possibility of estimating bid rent models where the bidding agents don't have to be aggregated in homogeneous groups or regimes and where bid functions are not constrained to be linear in parameters. This chapter proposes a method for the estimation of bid functions that maximizes the likelihood of the observed maximum bids while simultaneously adjusting the bid levels to observed average prices or price indicators. The main assumption behind the proposed method is that, as observed many times in practice, real estate goods are traded in auctions that don't take place explicitly. This implies that the outcome of the auction (the expected maximum bid) is a latent construct that can not be observed but is, however, structurally related to the transaction price. This assumption implies that the potential bid of all agents affects the final price of a real estate good, regardless if they are active in the market (currently looking for a location) or not.

The structure of the proposed model is inspired by the Generalized Random Utility Model (Walker and Ben-Akiva, 2002) and defines structural relationships for two latent variables: the bid and the auction price with the corresponding measurement relationships that relate them to observed choices (or best bidders) and observed prices.

The chapter is organized as follows: Section 4.1 reviews the literature on estimation of bid-rent function and analyzes the advantages and drawbacks of the different existing methods. Section 4.2 describes the method proposed in this paper and Section 4.3 describes a case study where the method is implemented, validated and compared with other methods. Finally, Section 4.4 concludes the paper and identifies future lines of research.

4.1 Estimation of bid rent functions

The first work on estimation of bid rent functions was developed by Ellickson (1981) who introduced stochasticity in the bid function specification and proposed for the first time the conditional probability of a household being the best bidder for a location (2.10). The original formulation by Ellickson considers a linear in parameters bid function and is estimated via maximization of the following likelihood function:

$$\mathscr{L} = \prod_{i \in S} \left(\prod_{h \in C_i} (P(h|i))^{y_{hi}} \right) \tag{4.1}$$

where y_{hi} is a binary indicator that assumes the value of one if household h is observed to be located in dwelling i and zero otherwise. The term $P(h|i)$ corresponds to the best bidder probability of (2.10).

Ellickson's method had as main objective the estimation of the willingness to pay for housing attributes by different agents, as an alternative to the hedonic rent model originally proposed by Rosen (1974). However, Ellickson's method only allows to estimate relative parameters because the scale parameter (μ) cannot be identified and, as depicted in (2.13), rent estimates are known only up to an undefined constant.

A method accounting for observed prices in the estimation to adjust the bids level was first proposed by Lerman and Kern (1983), as a direct extension of Ellickson's model. The method is based on estimating the joint probability of a household being the best bidder for a particular location and of that particular bid being equal to the observed transaction price or land rent (R_i). As a probability, this event can be expressed as:

$$P(h|i) = \text{Prob}\left\{B_{hi} + \varepsilon_h = R_i \text{ and } B_{hi} + \varepsilon_h > B_{h'i} + \varepsilon_{h'}, \forall h' \neq h\right\} \tag{4.2}$$

Lerman and Kern's approach considers that the land rent has exactly the same value of the maximum bid. If the error terms are Extreme Value distributed, the probability of (4.2) can be written as:

$$P(h|i) = f(R_i - B_{hi}) \prod_{h' \neq h} F(R_i - B_{h'i}) \tag{4.3}$$

with the density (f) and cumulative distribution (F) functions given by:

$$f(\varepsilon) = \mu \exp\left(-\mu\varepsilon\right) \exp\left(-\exp\left(-\mu\varepsilon\right)\right) \tag{4.4}$$

and

$$F(\varepsilon) = \exp\left(-\exp\left(-\mu\varepsilon\right)\right) \tag{4.5}$$

Therefore the likelihood function that needs to be maximized in order to estimate the parameters of $B_{hi}$is:

$$\mathscr{L} = \prod_{i=1}^{S} \left(-\mu \exp\left(-\mu\left(R_i - B_{hi}\right)\right) \prod_{h'=1}^{H} \exp\left(-\exp\left(-\mu\left(R_i - B_{h'i}\right)\right)\right)\right)^{y_{hi}} \tag{4.6}$$

where H is the total number of households participating in the auction and S is the total number of dwellings in the market. The term $y_{hi}$is a binary indicator that assumes the value of one if household h is observed to be located in dwelling i and zero otherwise. According to Lerman and Kern, the parameters of (4.6) can only be consistently estimated if the bid function is linear in parameters.

Lerman and Kern's method has been applied to estimate the real estate rents and the different agent's willingness to pay for particular attributes of housing units in several instances. For example, Gross (1988) applied the model on the city of Bogota, Colombia, finding that the bid-rent approach performs better than hedonic models when forecasting rents and marginal willingness to pay. Gross et al. (1990) and Gin and Sonstelie (1992) applied the model to the cities of Philadelphia and Baton Rouge (Louisiana) respectively, finding reasonable rent estimates. Chattopadhyay (1998) applied the model to the city of Chicago, finding that the rent estimates do not differ much from those of a hedonic model, but have the advantage of providing estimates of the willingness to pay for different groups of agents. Muto (2006) expands Lerman and Kern's model by incorporating an instrumental variable in the estimation and estimates the model for the city of Tokyo, obtaining reasonable results for rent forecasting but a significant bias for location choice. In all the applications found in the literature agents are grouped in homogeneous groups, therefore considering h as a type of agent instead of an individual household or firm. The estimation is done over a sample of locations for which detailed information on the attributes and individual transaction price is available.

An alternative way of estimating bid-rent function can be derived from the two stage estimation procedure originally proposed by Lee (1982) and adapted by Dubin and McFadden (1984) for the particular case of electric appliances and energy consumption. In this method a choice model is estimated in a first stage, obtaining parameters for the endogenous price function that are adjusted to observed prices in a second stage. In the particular case of bid-rent functions, the choice model is the maximum bidder probability described by (2.10) and the adjustment of the bid-rent function is done through the estimation of an hedonic price model where, besides the bid function itself, an instrumental variable is used as an explanatory element. The instrumental variable is obtained via regression of the price against attributes of the location that appear to be correlated with the price but not correlated with the error term in the agent's bid function. The two-stage model has been applied to the bid-rent problem and compared to Lerman and Kern's approach by McMillen (1997). Results show significant differences between the estimates of both approaches and suggests that Lerman and Kern's approach generates distorted results when implemented over data with selection bias problems. As in the bid-rent approach, the two-stage approach requires the aggregation of agents into a restricted number of homogeneous agents.

Similar methods for simultaneous estimation of price functions and location of agents can be found in the literature on locational sorting models (Tiebout, 1956). Based on the choice approach and generally focused on the estimation of willingness to pay and hedonic prices, these models also relate individual location choices with aggregate outcomes of the location choice of all agents in equilibrium (Epple, 1987) and spatial agglomeration phenomena (Bayer and Timmins, 2005). A specific example of this line of research in the context of spatial choice is the work of Bayer et al. (2007), where household preferences for school is analyzed under the effect of unobserved (and endogenous) neighborhood characteristics.

The literature on bid-rent function estimation has been focused on reproducing rent or price levels more than the agent's spatial distribution. One exception to this is the work by Muto (2006), where the location choice model obtained using Lerman and Kern's approach is compared with the original choice model using Ellickson's approach, finding a systematic difference between them. This results suggest that the particular solution proposed by Lerman and Kern allows to adjust bid levels to observed prices but with a cost in terms of the location-forecasting capability of the model.

4.2 Latent variable approach for bid rent function estimation

We propose a new approach for the estimation of the bid-rent function. We assume that real estate goods are traded in auctions, but that these auctions never take place explicitly. This means that the potential bid of all agents is latent and determines the price of the good, but only in relative terms. We call the outcome (or expected maximum bid) of this latent auction the "latent auction price". To adjust the latent auction price to the level of real prices it must be related to price indicators through a measurement relationship. For this we propose a model formulation based on the latent variable approach for discrete choice Walker and Ben-Akiva (2002); Walker and Li (2007), allowing for simultaneous estimation of the parameters of the bid function and of the price model.

Figure 4.1 shows the structure of the proposed model. Boxes represent observable data, like the attributes of households and locations, transaction prices and observed locations. Circles represent unobservable variables (or latent constructs) like the willingness to pay (bid) and the latent auction price. The dashed lines represent measurement relationships and the continuous lines describe structural relationships.

The proposed model is different from Lerman and Kern's model because it does not impose the bid of the located household to be equal to the observed price but, instead, imposes a linear relation between the latent auction price and a price indicator. An advantage of this approach is the fact that the price indicator (although it would be preferable) does not have to be the actual price of the transaction but, instead, it can be a much simpler and coarse proxy of price, like the zonal average price or rent by type of location.

The Bid function is related to the attributes through the structural equation that defines its functional form: $B_{hi} = f(x_h, z_i, \beta)$. Simultaneously, the measurement relationship between the Bid and the observed location is defined by the choice (best bidder) probability defined by equation (2.10). The structural relation of the latent auction price with the observed attributes of the location and the agents is given by the expected maximum bid, which is

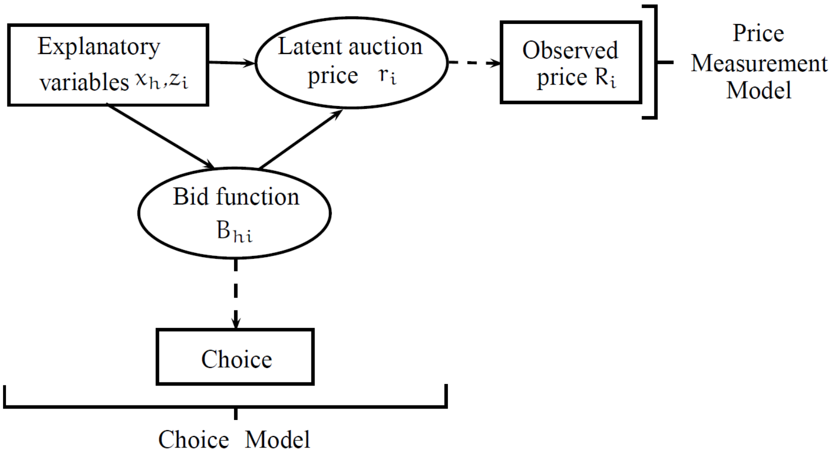


Figure 4.1: Latent auction model structure

defined by the logsum expression of (2.13). A new measurement relationship is considered in this formulation, assuming there is a linear relation between the latent auction price (r_i) and the observed prices (R_i), expressed as the following equation:

$$R_i = a + \gamma r_i + \eta. \tag{4.7}$$

Assuming a normal distribution for the error term η, a probability density function $f(R_i|r_i)$ with mean zero can be defined for the measurement relation of (4.7) as follows:

$$f(R_i|r_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{R_i - a - \gamma r_i}{2\sigma^2}\right) \tag{4.8}$$

The estimation of the proposed model can be done through traditional maximum likelihood but, in this case, the likelihood function is the product of the choice probability and the density function for the price for all observations:

$$\mathcal{L} = \prod_{i\in S}\left(\prod_{h\in C_i}\big(P(h|i)\cdot f(R_i|r_i)\big)\right)^{y_{hi}} \tag{4.9}$$

where $y_{hi} = 1$ if household h is the best bidder for location i and zero otherwise. In the context of the previous equation, S represents the set of available observations for estimation and

C_i is the set of households that participate in the auction for i. If no set generation model is available, it is reasonable to assume that all households participate in all auctions, therefore making $C_i = H$ for all i.

The outcome of the maximization of (4.9) will be the set of parameters (β) for the bid function (B_{hi}) and the a, γ and σ parameters of the density function for the price. However, in application, only the choice probability determines the best bidding household, therefore making the location probabilities independent of the price parameters. The measurement equation (4.7) can be used to estimate the expected prices as a function of the latent auction price.

4.3 Brussels case study

The model is estimated for the residential market of the city of Brussels. Data was collected from three main sources: the 2001 Belgium National Census the 2000 Brussels Land Registry Record and a travel survey to household performed in year 2000 (MOBEL, Hubert and Toint (2002)). The study area considers an extended metropolitan region, including 151 communes that contain a total of 4945 zones, denoted by the index i. Dwelling alternatives are classified in 4 types (isolated, semi-isolated and attached houses and apartments), denoted by the index v. Data adds to a total of 1274701 residential units or location alternatives, characterized by their average physical and land use attributes by type of dwelling and zone (vi), which are calculated from the Census and the Land Registry. The area of study contains a total of 1267998 households, therefore having an aggregate vacancy rate (supply surplus) of 0.5%. The estimation is done over a sample of 1007 observations of located households from the travel survey. After testing several different specifications, the linear-in-parameters specification described in Table 4.1 was considered for the bid function B_{hvi}, which can be interpreted as the willingness to pay of household h for a dwelling of type v in zone i.

Table 4.1: Bid function specification

parameter	variables
β_{surf}	surface$_{vi}$ (m^2) × log(size$_h$) (number of people)
β_{sup}	high_educ$_i$ (%) × high_educ$_h$ (number of people)
β_{house}	is_house$_{vi}$ (dummy) × size$_h$ (number of people)
$\beta_{\text{mid\_inc}}$	avg_income$_i$ (Euros) × high_income$_h$ (dummy)
$\beta_{\text{high\_inc}}$	avg_income$_i$ (Euros) × mid_income$_h$ (dummy)
β_{trans0}	PT_accessibility$_i$ (facilities/km^2) × 0_cars$_h$ (dummy)
β_{trans2}	PT_accessibility$_i$ (facilities/km^2) × 2_cars$_h$ (dummy)
β_{comm}	commerce$_i$ (jobs/m^2) × log(size$_h$) (number of people)
β_{office}	office$_i$ (jobs/m^2) × workers$_h$ (number of people)
β_{green}	green$_i$ (parks/m^2) × children$_h$ (number of people)

The variable surface$_{vi}$ is the average surface of a residential unit of type v in zone i and it is interacted with the number of individuals in the household. The building types consider three types of house (fully-detached, semi-detached and attached) and apartments. The percentage of people in a zone with a university degree (high_educ$_i$) is interacted with the number of individuals in the household that have a degree as well. The average income by zone (avg_income$_i$) was calculated from tax declarations and it is interacted with a dummy that indicates if household h is of high income level (more than 3099 Euros per month) or of mid income level (between 1860 and 3098 Euros per month). The public transport accessibility variable (PT_accessibility$_i$) was calculated as the density of public transport facilities within a zone and it is interacted with a dummy variable than indicates if the household has no car or if it has two or more cars.

Price data is available as average by commune (i') and for a simplified classification of dwelling types that aggregates them into houses and apartments (v'). The measurement equation for prices is defined following (4.7) and using the explicit definition of the maximum expected bid given by (2.13). Instead of price we use the natural logarithm of the price, to capture the diminishing marginal utility of housing attributes (DiPasquale and Wheaton, 1996). The resulting expression is similar to a log-log regression for price, a convenient specification due to its good performance for price forecasting when data describing the dwelling is not complete (Cropper et al., 1988).

$$\ln(R_{v'i'}) = a + \gamma \cdot \ln \sum_h \exp(\mu B_{hvi}) \tag{4.10}$$

For the estimation process, the scale parameter μ of the bid probability (2.10) is assumed to be one.

4.3.1 Estimation results

The model was first estimated for Ellickson's specification in order to get the best possible maximum bid model. Once good estimates were obtained the model was re-estimated with the approach proposed in Section 4.2, but keeping the same specification for the bid function, defined by Table 4.1. The estimation in both cases was done using an extended version of the software package BIOGEME (Bierlaire, 2003; Bierlaire and Fetiarison, 2009); results are shown in Table 4.2, where the columns on the left show the results using Ellickson's approach while the columns on the right show the results obtained when using the method proposed in this paper, from now on called "Latent Auction" model.

For Ellickson's model all parameters are significant with a 95% confidence. The signs of the parameters show that the willingness to pay increases with the surface of the dwelling and the size of the household, and that households with members having university degrees prefer to locate in neighborhoods with a high presence of people with a similar education level. Something similar happens with households of mid and high income level, who have a higher

Table 4.2: Estimation results for Brussels

	Ellickson			Latent Auction		
Parameter	**Value**	**Std err**	**t-test**	**Value**	**Std err**	**t-test**
β_{surf}	0.00636	0.00261	2.43	0.000311	0.000225	1.38*
$\beta_{\text{mid\_inc}}$	0.0439	0.0111	3.94	-0.00317	0.00717	-0.44*
$\beta_{\text{high\_inc}}$	0.0574	0.0153	3.76	0.0161	0.00998	1.61*
β_{sup}	0.403	0.108	3.73	0.728	0.0739	9.84
β_{trans0}	0.408	0.136	3.00	0.599	0.0849	7.06
β_{trans2}	-0.532	0.153	-3.48	-0.31	0.0791	-3.91
β_{house}	0.461	0.0615	7.5	0.0563	0.00702	8.03
β_{comm}	-1.34	0.278	-4.83	-0.0366	0.031	-1.18*
β_{green}	-0.349	0.0717	-4.86	0.136	0.0201	6.74
β_{office}	-0.295	0.0931	-3.16	0.0896	0.0413	2.17
a	-	-	-	-16.4	3.23	-5.08
γ	-	-	-	1.92	0.229	8.39
σ	-	-	-	-1.92	0.0225	-85.48
Final Log-Likelihood	-7011.03			-6387.76 (-7091.13**)		

*parameters not significant at the 95% level

** log-likelihood considering only the choice probabilities

willingness to pay for location on zones with high average income. Households without a car give a positive value to the presence of public transport facilities while households with more than one car prefer to locate in regions with low accessibility for public transport. An interesting result is the effect of the presence of commerce, public green areas and office space, with a negative parameter for all of them and decreasing with the size of the household or the number of workers, depending on the case. These negative estimates were originally interpreted as households preferring to locate in peripheral areas of the city, where the density of commerce, public areas and offices is lower. However, this conclusion is challenged by the results obtained when using the Latent Auction model, as it will be shown next.

When estimating the Latent Auction model some of the parameters become insignificant and some change their sign. For example the surface of the dwelling, the presence of commerce and the average income of the zone have a less relevant effect, with parameters that are significant with less than a 95% confidence. Other estimates like β_{green} and β_{office}, that were originally negative, came out positive in the estimation with the Latent Auction model. The change in the sign of the estimates can be explained as an endogeneity effect in the Standard logit formulation that happens due to the lack of price information. The data for estimation shows that bigger households prefer to locate in the outskirts of the urban area, this is likely to be due to lower prices for bigger dwellings in these regions where, incidentally, the presence of public green areas and offices is low. When the price indicator is considered, the estimation generates positive parameters for green areas and offices because, as expected, these attributes are likely to increase the average price in a neighborhood. This result suggest that, by accounting for price indicators, the Latent Auction model is able to generate more realistic estimates.

Table 4.3: Estimation results for Brussels

	Ellickson			L&K		
Parameter	**Value**	**Std err**	**t-test**	**Value**	**Std err**	**t-test**
β_{surf}	0.00636	0.00261	2.43	-0.00136	0.000855	-1.59*
$\beta_{\text{mid\_inc}}$	0.0439	0.0111	3.94	0.0194	0.00608	3.19
$\beta_{\text{high\_inc}}$	0.0574	0.0153	3.76	0.0474	0.00796	5.95
β_{sup}	0.403	0.108	3.73	0.416	0.0669	6.22
β_{trans0}	0.408	0.136	3.00	-1.01	0.0716	-14.1
β_{trans2}	-0.532	0.153	-3.48	-0.226	0.0887	-2.54
β_{house}	0.461	0.0615	7.5	0.0167	0.0182	0.92*
β_{comm}	-1.34	0.278	-4.83	-0.768	0.0977	-7.85
β_{green}	-0.349	0.0717	-4.86	0.286	0.0367	7.78
β_{office}	-0.295	0.0931	-3.16	-0.767	0.0533	-14.38
μ	1	-	-	1.66	0.0173	95.74
Final Log-Likelihood	-7011.03			-7569.645 (-11813.1**)		

*parameters not significant at the 95% level

** log-likelihood considering only the choice probabilities

For comparison purposes, the same specification of Table 4.1 is estimated using Lerman and Kern's method, therefore maximizing the likelihood function of (4.6). Results for this method are shown in the second column of Table 4.3 (L&K). The original estimates obtained with Ellickson's method are shown in the first column.

Some of the results obtained with the Lerman and Kern method are counter intuitive. For example the parameter for the unit surface becomes negative indicating a higher value (and preference) for smaller dwellings. Same thing happens with the parameter for presence of public transport for household with no car. Regarding the likelihood ratio test for location choice, L&K's method is clearly dominated by both Ellickson's and the method proposed in this paper, however, it generates relatively good rent estimates as it is shown next.

4.3.2 Model likelihood and fit analysis

It is not straightforward to evaluate and compare the likelihood of each model; the different expressions for the likelihood functions make the direct comparison of final log-likelihoods unfair. The final log-likelihood, calculated as the logarithm of sum of the probabilities of the chosen alternatives, is a comparable indicator because it considers the same specification for the bid function in both models. This statistic suggests that the standard logit fits better than the Latent Auction model and that both models are significantly better than Lerman and Kern's approach. However, this is only valid for the data used in estimation and an expected result because the standard logit models attempts to fit only to this data set, while the models using with a price indicator attempts to fit simultaneously an additional set of observations.

Regarding the price model, the fit of the estimated prices is a good indicator of the quality of each model. Figure 4.2 shows the difference between estimated and observed average prices per commune and dwelling type for the estimation data set. Each column in the boxplot graphic shows results for a different model; the box indicates the value of the two quartiles of observations that are closer to the reference value, the extremes of the column indicate the value of the biggest positive and negative error. Since both the relative and absolute differences are relevant, both statistics are shown, in the upper and lower plot respectively.

Both the Latent Auction and Lerman and Kern's method perform reasonably well. The method proposed in this paper generates estimates that are in 75% of the cases deviated less than 1% from the observed prices with a maximum deviation of 4%. Lerman and Kern also performs well, with 75 percent of the estimates deviated less than 4% and a maximum deviation of 6%. In both cases, some deviation is reasonable because the estimated prices are calculated for a wider classification of dwelling types and for a much finer basic spatial unit than those of the observed average prices .

As expected, Ellickson's method does not perform well in this regard, systematically overestimating the prices. However, it seems to be the best models regarding estimation of the spatial distribution of agents. Because of this, the result analysis so far does not allow to identify which model is performing better in general and further validation is required.

4.3.3 Validation

Validation is performed by simulating the location distribution for all the locations in the city with each model, and comparing the results with observed statistics. For this, all the real estate supply is generated from the census data and households are assigned following the different maximum bid distributions obtained with each method. The analysis is performed for three variables: prices, number of individuals in the household and number of individual with university degree. Results are shown in Figures 4.3, 4.4 and 4.5 as the difference at the commune level of the forecast variables against their observed value.

The difference between price forecast and observed average price is shown in Figure 4.3. Results show that, when applying the models to a different data set, the Latent Auction approach is superior to Lerman and Kern, where a systematic overestimation occurs. This is probably due to the intensive data requirements of L&K, which are not met by the relatively poor nature of the available data.

Figure 4.4 shows the results for total number of people (the sum of the number of individuals per household), aggregated by commune, against the official population statistics coming from the 2001 Belgium National Census. The Latent Auction model tends to underestimate the population at the commune level with 50% of the communes having a deviation smaller than 7%. Ellickson's model tend to overestimate the population, with a slightly higher deviation while Lerman and Kern's model systematically underestimates this variable.

Figure 4.5 shows the difference between the forecast of people with university degree by

Figure 4.2: Estimation fit (natural log of price)

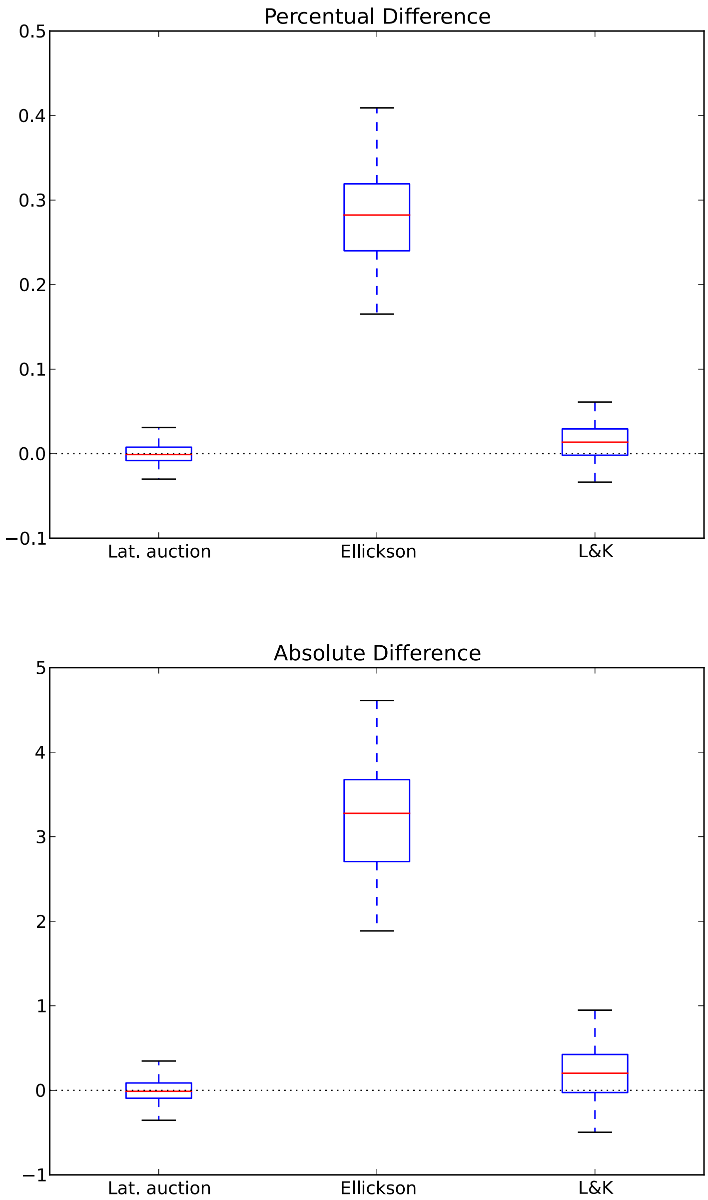
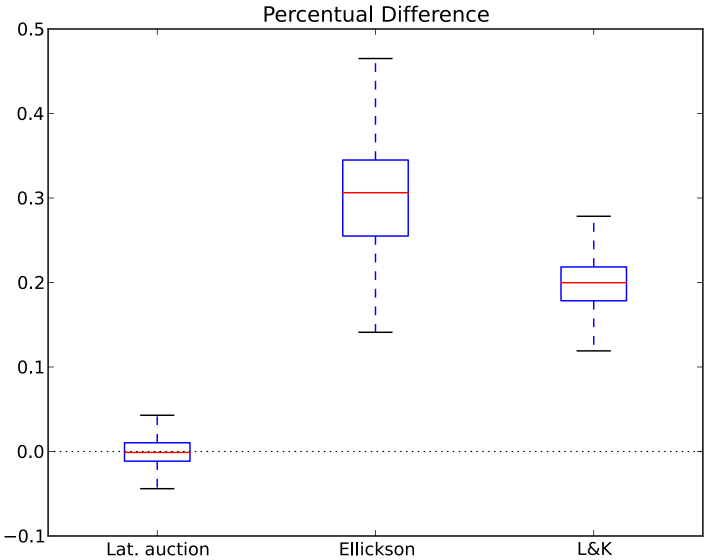


Figure 4.3: Error in forecast: natural log of price



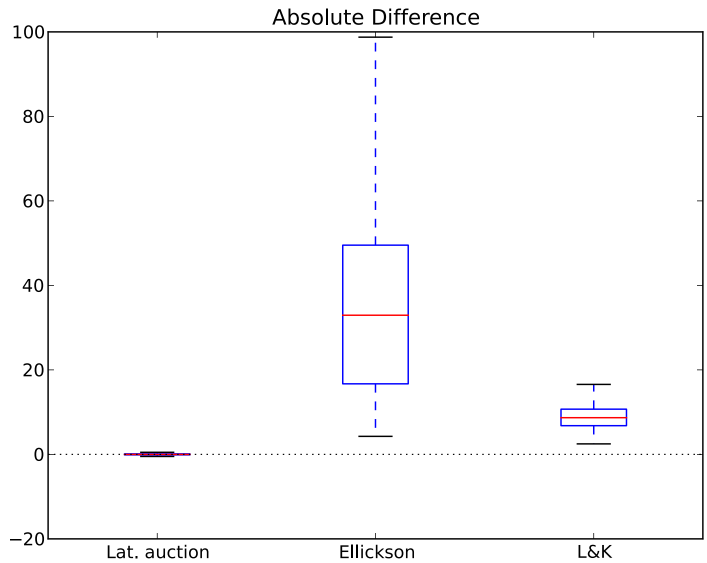
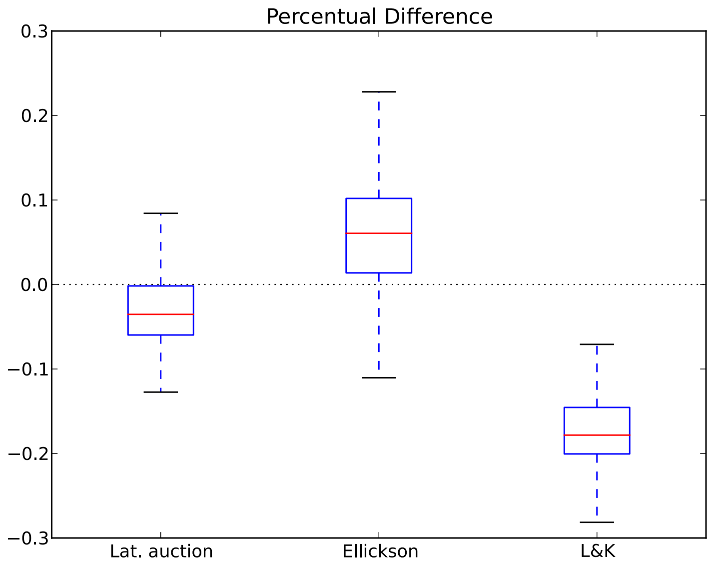


Figure 4.4: Error in forecast : number of people by commune



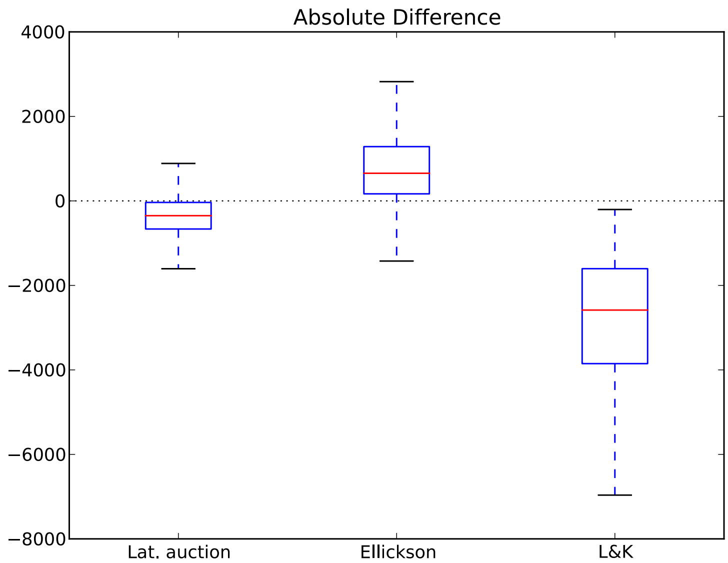
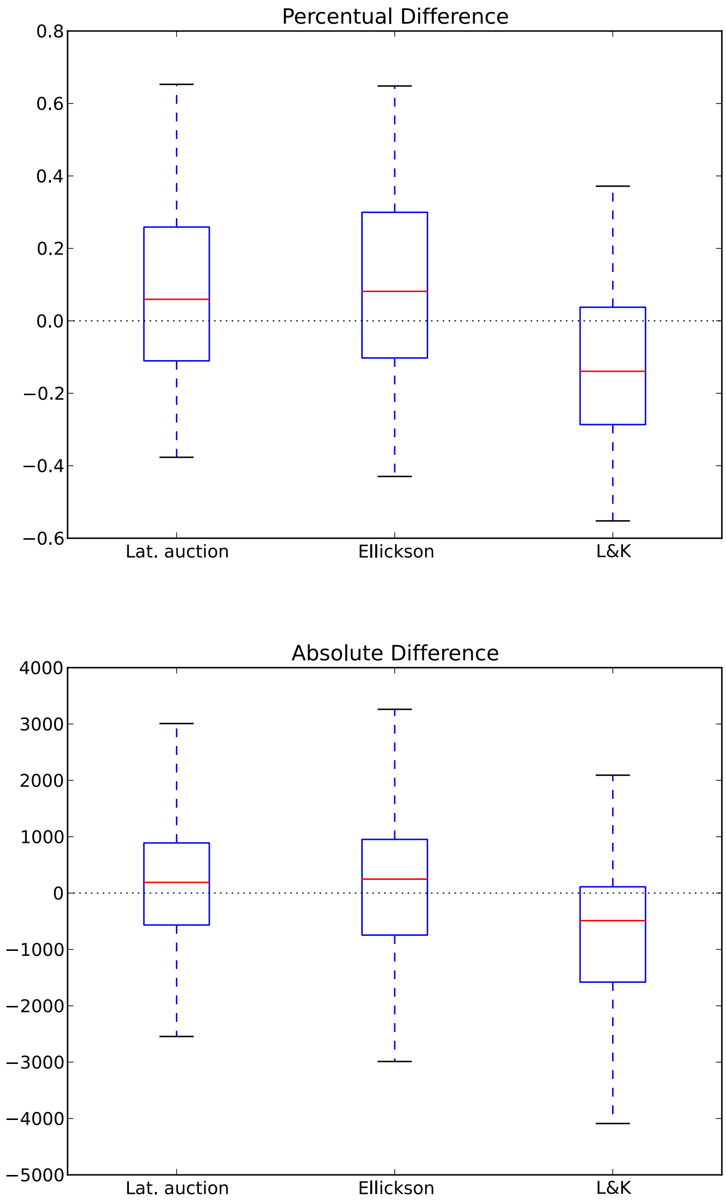


Figure 4.5: Error in forecast: Number of people with university degree by commune



commune against the official statistic from the Census. In this case both the Latent Auction and Ellickson's model perform relatively well, with a tendency to overestimate the variable and with 50% of the communes having a deviation not larger than 25% from the observed value. Lerman and Kern tends to underestimate this variable. It's worth noticing that, at the absolute level, the Latent Auction model outperforms the forecast of the other models

4.4 Conclusions

An estimation method for bid-rent functions that accounts for observed locations and price indicators is proposed. Results show that including a measurement equation for the expected auction price and the observed prices in the log-likelihood maximization process allows to obtain better estimates of the bid function parameters. The proposed model is able to forecast, with a reasonable error, the location choice distribution of agents in the city while, simultaneously, adjusts the bids to the price indicators. Because of this, the Latent Auction model outperforms Lerman and Kern's model, since the later adjusts well the bid-rent level but deviates significantly from the observed spatial distribution of agents. Moreover, when applied in forecasting, Lerman and Kern is not able to adjust to the price indicators.

The proposed model has the advantage of not requiring detailed data about real estate goods and prices. This is because the measurement relationship for prices is defined as a function of the average expectation of the maximum bid, considering all potential bidders. Therefore it does not require to know the set of bidders or the actual value of the winning bid. In fact, for the case study, only average values were available for both dwelling attributes and prices. This makes the method easier to implement when data is scarce or of aggregate nature.

The differences observed between forecast and observed prices is expected and explained by the aggregate nature of the price indicator. A more disaggregated indicator should allow for a better estimation and, consequently, a better fit. However, it is not clear if under this conditions the model outperforms the approach proposed by Lerman and Kern. Therefore, performance of the model when detailed data is available should be analyzed. For this, information on the values of all bids participating in an auction should be collected from places where real estate auctions take place explicitly, like Scotland or Australia (Quan, 1994).

Further research should also investigate the relevance of choice set formation phenomena (identification of the active bidders in each auction) and the use of more sophisticated (non-linear) structural relationships between the latent auction price and the observed price indicators.

5 Market clearing method for agent-based location choice models

Interventions on urban systems, such as real estate developments, modifications to the transport system and changes in urban policy, are usually costly to implement and, therefore, require models to forecast and evaluate their performance and effects in other elements of the system. Land use and transportation models are tools used for this purpose. Among these, microsimulation models are becoming more relevant and attractive due to the possibility of representing individual agents and their complex interactions in a simple, yet robust and flexible, way. Moreover, agent based microsimulation can easily account for the dynamics in the system, something that is hard to achieve in equilibrium models.

Location choice is one of the most important processes to model in a land land use model. Together with real estate prices, it has been traditionally modeled under two different paradigms: the choice approach and the bid-auction approach. Under the choice paradigm, households select the location that maximizes their utility and behave as price takers, with prices being determined exogenously to the location choice process. The bid-auction approach assumes that real estate goods are traded in an auction market, where the best bid for a particular location determines both the located household and the price of the dwelling.

The real estate market has two unique characteristics: first, all goods are quasi-unique since, because of their spatial nature, all locations are different and, at the same time, each good must be assigned to one buyer only. Second, demand is inelastic because all agents need to be located somewhere. This means that the market needs to clear because several agents are competing for a finite number of locations and this competition will have an effect on both the prices of goods and the spatial distribution of agents. Market clearing is traditionally achieved in aggregate models by solving an equilibrium: finding the state of the system where all agents can't improve their situation. Finding an equilibrium requires solving a fixed point problem and some strong assumptions like perfect information, perfect match between demand and supply or allowing the relocation of all agents in the system. The market clearing problem has been less visited in the case of agent-based, microsimulation problems and it's either ignored or solved at an individual level, assuming that agents are price takers and therefore ignoring market effects or simulating several iterations of individual auctions and therefore involving a lot of computational resources.

This chapter proposes a method to model location choice and real estate prices simultaneously in a microsimulation context. The method is based on the bid-auction approach and understands both location and prices as a function of the households' preferences. The proposed approach does not require solving for equilibrium, but estimates the maximum bid in each period by simulating the underlying auction process. Given exogenous supply levels, households adjust their preferences (and their willingness to pay) as a reaction to the (observed) market conditions. This adjustment goes in the direction of an equilibrium (although it does not reach it) and produces prices that are higher when goods are scarce and lower when goods are abundant. The approach has the advantage of not being excessively expensive in terms of computational time, thanks to the estimation of the price as an expectation of the maximum bid over all agents instead of simulating individual transactions.

The chapter is organized as follows: Section 5.1 explains the market clearing mechanisms used in equilibrium models and agent-based model. Section 5.2 proposes a market clearing mechanism for a bid auction location choice model. Section 5.3 proposes a general framework for land use simulation that embeds the previously proposed market clearing mechanism. Section 5.4 describes the estimation and implementation of the proposed framework for the city of Brussels and shows validation results. Finally, Section 5.5 concludes the paper and identifies possible further research.

5.1 Market clearing

Real estate markets are particular and different from other markets because of the scarce nature of the traded goods and the inelastic demand for them. Given their spatial nature, real estate goods have a quasi-unique nature. Two locations are basically different because they can't use the same space, and their distance to other points of space and perception of amenities/externalities will be different. This is true even for multi-family, high-rise housing: two apartments can be located in the exact same building but they will have, for example, different views and access to sunlight or be in different floors. Besides the fact that only one consumer can use a location, the main implication of the quasi-unique nature of real estate goods is the fact that demand for them will be differentiated: the preference and willingness to pay of one consumer for a particular location will also be quasi-unique. Besides this, housing is a basic need and, therefore, demand for it is essentially inelastic: a household can not afford not to locate anywhere because it has to live somewhere. These particular characteristics generate a lot of interactions and friction between agents in the real estate market.

The location choice problem can be described, in simple terms, as matching agents with locations in a way that avoids conflicts. That is, establishing a unique relationship between each agent and each location, avoiding situations with more than one agent per location and avoiding having agents located in more than one place. This can be described through the joint probability of a match between agent h and location i ($P(h,i)$). According to Bayes theorem, this probability can be decomposed as follows:

$$P(h,i) = P(h|i)P(i) = P(i|h)P(h) \tag{5.1}$$

Each of the terms of equation (5.1) have an interpretation that is consistent with either the Bid-auction (maximum bid) or the Choice (maximum utility) approach. The term $P(h|i)$ represents the probability of agent h being the best bidder for a location i (2.10) while $P(i|h)$ represents the probability of location i being selected by agent h. This allows to interpret the term $P(h)$ as the probability of a household being able to locate somewhere and and $P(i)$ as the probability of a location being occupied by someone. Note that, if supply equals demand and no constraints apply, they are both equal to one.

Market clearing is the process of interaction between supply and demand that assigns goods to consumers (or vice versa). It can be understood, in simple terms, as matching agents with locations in a way that avoids conflict. That is, establishing a unique relationship between each agent and each location, avoiding situations with more than one agent per location or with agents located in more than one place. In a market context conflicts are solved through prices that are adjusted in order to discriminate between agents or goods.

From a modeling point of view, market clearing is about finding the joint probability $P(h,i)$ that describes the conflict-free association between agents and locations in an aggregate way. If the market clearing problem is analyzed in disaggregate terms, it can be described as finding a set of conflict free agent-location relationships that will define the joint probability $P(h,i)$ for all agents and locations. In both cases, market clearing will require a price adjustment process.

Depending on the modeling paradigm (Bid-auction or Choice) the market clearing process adjusts prices in different ways. If the real estate market is understood as a choice market, consumers (agents) are assumed to be price takers, selecting the location that provides the maximum utility given a price defined by the provider (for example, the owner or seller of the location). If two or more agents select the same location, the seller can discriminate between them by increasing the price marginally until only one agent is willing to buy. The seller does this because it behaves as a profit maximizer and will try to sell at the maximum possible price, regardless of who gets the location. In the contrary case, if no agent chooses a particular location, the seller can decide to lower its price until it becomes attractive for at least one agent or, if the price reaches a certain minimum selling price threshold, it can decide to withdraw the location from the market (meaning that he is willing to wait for better conditions, with the associated lost of profit).

Under the Bid-auction paradigm it is assumed that agents bid their willingness to pay for each location, with each seller selecting the highest bidder. If an agent is the best bidder for more than one location it can discriminate between them by marginally (and homogeneously) reducing his bid until he remains the best bidder in only one. The bidder does this because it behaves as a utility (or consumer surplus) maximizer and it is, given his willingness to pay, indifferent between locations so it will try to pay the smallest possible amount. If an agent does not win any auction it will increase his bid until it wins at least one auction, if the agent runs out of available budget and still can't find a location it becomes "homeless". It is important to notice that an adjustment of the willingness to pay does not imply a change in the preferences of agents, but rather a re-assessment and adjustment of their expected utility level.

In a real scenario, the previously mentioned adjustments happen for many locations and many agents, some of them simultaneously and some of them sequentially, which makes the problem a very complex one. The following sections review how market clearing has been approached in the context of aggregate and disaggregate location choice models.

5.1.1 Market clearing in aggregate (equilibrium) models

In economic theory, market clearing is generally understood as the process of adjustment of prices that equilibrates demand and supply at some aggregate level. Equilibrium models assume that the market clears through the adjustment of prices until every agent (or group of agents) has achieved a state where unilateral decision can't improve their perceived utility. This means that there is a combination of location choices and equilibrium prices where every agent achieves maximum utility (Fujita, 1989). Achievement of this estate requires the simultaneous adjustment of all locations and prices, which means that perfect information of agents need to be assumed. Traditionally, equilibrium-based land use model find equilibrium by adjusting these variables until every agent is located somewhere and every location is selected by someone. This is equivalent to set the probabilities $P(h)$ and $P(i)$ of (5.1) to be equal to one for every agent and every location respectively.

In the case of a choice approach to location choice this translates into finding prices such that supply equilibrates to demand or, more specifically, into finding the vector of rents or prices (r) that solves the following system of equations:

$$\sum_h H_h P\left(i|h, r_i, z_i(\overline{h})\right) = S_i \quad \forall i \tag{5.2}$$

We explicitly introduce the price of the location (r_i) and the socioeconimic attributes of the location (z_i)that will depend on the location choice of the rest of the households $(\overline{h})$. This is done in order to show the decision variables and how the location preferences of one agent depend on the location of other agents. Because of the aggregate nature of the problem, indexes h and i indicate a group or cluster of agents and locations respectively. The variables H_h and S_i represent the number of households in group h and the supply in location

i respectively. The system of equations is a fixed point problem due to the interdependence of the location decisions and to the fact that prices will depend on socioeconomic attributes of the location, that are (also) defined by the location choices. Equation (5.2) can be interpreted as the seller of each good finding the prices that will ensure all supply (S_i) will be sold.

Most land use equilibrium models use the choice approach and, therefore, solve a problem similar to the one described by (5.2). Examples of this are TRANUS (De La Barra, 1980; De La Barra et al., 1984; De La Barra, 1989), MEPLAN (Echenique et al., 1990) and RELU-TRANS (Anas and Liu, 2007).

Similarly, in bid-auction based equilibrium models,clearing the market involves finding the vector of bids (B_{hi}) that solve the following system of equations:

$$\sum_i S_i P\left(h|i, B_{hi}, z_i(\overline{h})\right) = H_h \quad \forall h \tag{5.3}$$

where the probability of being the best bidder depends on the willingness to pay for the location (B_{hi}) and the location distribution of the rest of the agents ($\overline{h}$). The analogy with the system of equations for the choice problem (5.2) is evident although not absolute. An important difference is the fact that, the decision variable (B_{hi}) is partly affected by the location probability of other agents. However, the main difference with the choice approach from (5.2) is the fact that the decision variable in this case has two dimensions (h and i), although it can be adjusted only in h . The interpretation of this is as follows: each agent is indifferent about the location he gets, as long as he is able to pay his willingness to pay for it, which is described by his bid function. If the relative preferences for each location remain the same, the only adjustable factor is the expected utility the agent will get from locating somewhere. This justifies understanding the bid function of an agent as having two components (see equation 2.8 in Chapter 2): one "hedonic" component (b_{hi}) describing his willingness to pay for the attributes of each location and one "utility" component (b_h) indicating the change required in the bid level to solve (5.3):

$$B_{hi} = b_{hi} + b_h \tag{5.4}$$

Therefore, equation (5.3) can be interpreted as each cluster of bidders finding the utility level and corresponding bid adjustment (b_h) than will ensure their location somewhere. Some examples of equilibrium models using an equilibrium-based bid-auction approach are RURBAN (Miyamoto and Kitazume, 1989) and MUSSA (Martinez, 1992).

If prices are the outcome of an auction process and the market clears, the distribution of households across locations obtained through solving (5.2) will be the same as the distribution obtained from solving (5.3) when supply and demand are equilibrated (Wheaton, 1977; Martinez, 1992). This also means that the previous equilibrium conditions of (5.2) and (5.3) can be achieved only when an absolute equality between supply (the number of location alternatives) and demand (the number of households) holds, meaning that:

$$\sum_h \sum_i P(i,h) = H = S \tag{5.5}$$

with H the total number of households and S the total number of locations.

The equilibrium approach to model the real estate market is useful for its elegance and simplicity. The analytical nature of the involved equations and adjustment processes allows for explicit and stochastic-variation-free analysis of the market trends. However, the assumption of supply being equal to demand is un-realistic and does not allow to account for scenarios where supply (or demand) surplus might be possible. The required adjustment of all locations and prices (or bids) is also unrealistic since it requires to assume perfect information for all agents while, in reality, only partial information is available to decision makers in both the demand and supply sides of any urban system. Because of this, changes in the city generally happen with parsimony and slow (sometimes marginal) adjustments, rendering the dynamics of the real estate market hard to model using an aggregate equilibrium approach. Finally, the complexity of the problem and fixed point nature of the systems of equation defined by (5.2) and (5.3) forces to consider groups of agents and real estate goods instead of individual agents and goods. Because of these limitations, other approaches to market clearing location choice modeling have been proposed in the literature.

5.1.2 Market clearing in aggregate (dynamic and disequilibrium) models.

The lack of dynamics and the forced equilibration between supply and demand in equilibrium motivated the development of different approaches to model the real estate market clearing problem. For example Horowitz (1986) proposed a model based on the bid auction approach, but lifting several of the assumptions typically taken under the equilibrium paradigm. Horowitz assumed that the bidding process for real estate goods happens sequentially instead of simultaneously and that decision makers don't have perfect information of the market. Because of this, sellers may not necessarily sell to the highest bidder, either because they accept an earlier, lower bid or because the potential best bidder never participates in the auction of the good. Horowitz also constrained the bids for households by introducing a minimum selling price defined by the developer or seller of the real estate. The model was estimated, generating better results than the standard bid-auction model and thus indicating that, by introducing more realistic behavioral assumptions for agents, it can describe the market in a better way. However, Horowitz's model requires to include the asking price as an explanatory variable and does not propose methods to estimate it. The model also ignores the dynamic effects of demand or supply surplus in the market clearing process being, in this regard, equivalent to equilibrium models.

Introduction of dynamics is relevant for location choice modeling because of the non-instantaneous nature of supply generation and location change in a urban system. Cities change constantly and parsimoniously, the urban system reacts to interventions at different speeds and the outcome of processes depends in many cases of the path taken. Many land use models introduce dynamics by modeling periods of time and accounting for time lags and feedback effects (mostly excess of supply or surplus and transport system performance) that make decision in one period dependent on previous periods. Examples of this are DELTA (Hunt and Simmonds, 1993; Simmonds, 1999), PECAS (Hunt and Abraham, 2003) and IRPUD (Wegener, 2008). A relevant characteristic of dynamic models is that they model the relocation of only a fraction of the agents in each period, therefore not solving an equilibrium for all agents and locations. These models deal with market clearing by adjusting prices in a similar way as equilibrium models but without solving a full, long-term, equilibrium, which requires some simplifying assumptions like periodic adjustment of rents to reduce supply surplus or reduction of the available supply to match aggregate demand.

Very few dynamic models rely absolutely on equilibrium to clear the market since it requires the simultaneous solution of several cross-sectional and inter-temporal equilibrium problems. An example of this is the model proposed by Anas and Arnott (1991), where real estate developers are assumed to have perfect foresight regarding future prices and consumers have perfect information of all markets. The model solves the equilibrium for all periods at the same time, with producers matching supply in each an every one of them in order to find a vector of equilibrium prices. Although mathematically and economically robust, the strong assumptions required by this model imply limitations that render it difficult to implement in a real case study

Martínez and Hurtubia (2006) proposed a dynamic model for the real estate market that finds equilibrium while accepting differences between supply and demand and lack of perfect information. The model introduces constraints to the bidding behavior of agents by setting thresholds that can be related with (but not limited to) budget constraints (Martínez et al., 2009). The supply is also constrained by minimum selling prices that are determined by the construction or development cost of a microeconomic supply model that generates new supply with a temporal lag, therefore not necessarily satisfying demand. The model finds equilibrium prices in each period and determines which fraction of the demand or supply will not be utilized or located respectively. The model does not require to assume perfect foresight for suppliers and allows to consider different levels of myopia for both consumers and developers but it requires a large amount of detailed (and often private) data to be calibrated and it is, therefore, not easy to implement in a real case study.

All the models mentioned in this section work at an aggregate level to represent agents, which implies a limitation in terms of modeling heterogeneity within groups of agents. Detailed dynamic microsimulation models, accounting for individual agents, require a different approach for market clearing, as discussed in the next section.

5.1.3 Market clearing in disaggregate models

Agent-based location choice models assign individual agents (households or firms) to specific locations (dwellings or buildings) and usually do this through simulation. The price formation process is therefore harder to model since it is not explicitly linked to the relation between supply and demand at an aggregate level.

For example UrbanSim (Waddell, 2002; Waddell et al., 2003) assumes that real estate price are exogenous to the market clearing process and that agents behave as price takers, selecting the location that provides maximum utility. Prices in UrbanSim are computed in each simulation period through a hedonic price model estimated in the base year, therefore making prices dependent only on the attributes of the location. This implies that the equilibrium or market conditions of the base year are assumed to remain constant in time. In cases of conflict like, for example, two agents choosing the same location, UrbanSim assigns it using a *"first-come, first-served"* approach (Waddell, 2010). The idea behind using such an approach is that neither sellers or buyers have perfect information on the market and sellers will minimize risks by giving the location to the first buyer at the asking price. This approach is realistic because it does not require strong assumptions regarding demand/supply equilibrium or perfect information and, at the same time, is easy to implement in a agent-based framework. However, the clearing process completely ignores market effects, generating the same transaction prices regardless of the level of supply/demand surplus or socioeconomic composition of the demand.

An alternative way to deal with the market clearing is the disaggregate approach proposed in models like ILUTE (Salvini and Miller, 2005) where each transaction (or matching between agent and location) is microsimulated by modeling the interaction between sellers and buyers who negotiate based on their willingness to pay and reservation prices respectively. In simplified terms, the approach attempts to solve a simplified version of the systems of equations of (5.2) and (5.3) at an individual level, meaning that there are as many equations to solve as goods to trade and several iterations are required in order to reach a solution. The simulation of auctions has the advantage of being flexible and, therefore, applicable when modeling different markets, as proposed by Abraham and Hunt (2005). This approach, described in detail and applied to model the residential real estate market of Toronto in Farooq and Miller (2012), does not require strong assumptions like the perfect satisfaction of demand or perfect information for decision makers, but is extremely expensive in computational terms. It also depends heavily in the choice set formation process that defines which households consider which locations to negotiate prices, which is hard to model because the choice sets of decision makers are usually unidentifiable, both in size an characteristics.

Ignoring market effects may lead to mis-estimation of prices because they will only depend on location attributes while, at the same time, it requires to oversimplify the conflict solution to a lottery. On the other hand, simulating auctions individually is computationally expensive and may produce an overestimation of prices, as suggested by the validation results described in Farooq and Miller (2012). This motivates the search of alternative ways to clear the market that are compatible with an agent based approach but retaining the stability and economic foundation of equilibrium-based market clearing approaches.

5.2 A quasi-equilibrium approach to market clearing

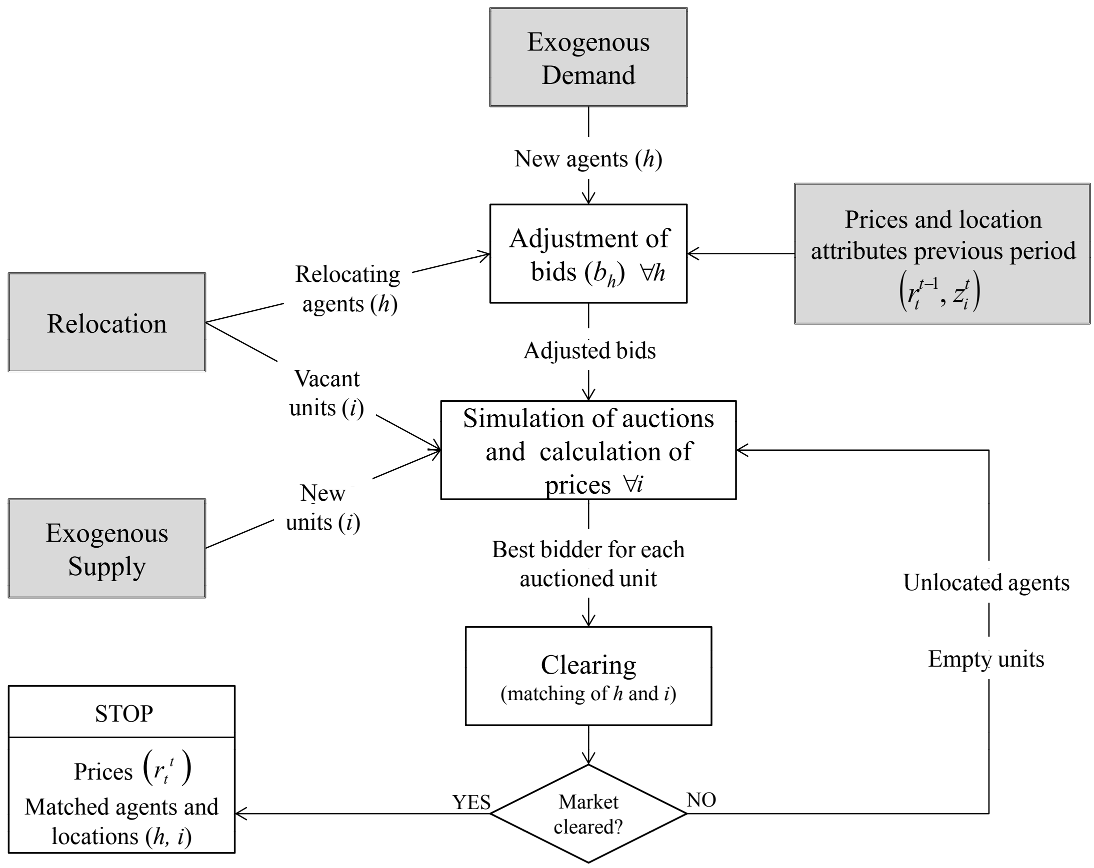
A model for real estate market clearing is proposed. The model takes into account the equilibrium forces that take place in market interactions like the adjustment of expectations of agents and the corresponding adjustment of behavior. However, the model does not attempt to solve an equilibrium and, instead proposes a disaggregate adjustment process with some outcomes, like the price, being the result of an aggregate market interaction.

Our proposed model has the following assumptions

- Interactions between agents take place in a discrete period framework. A period can be any amount of time large enough to account for a change in the levels of supply and demand.
- In each period, a group of agents enters the market looking for a location. Real estate supply for the same period is determined independently and does not necessarily satisfy demand perfectly.
- Real estate goods are transacted in auctions. All active agents (those looking for a location) are potential bidders for all locations .
- Agents do not have access to perfect information on the conditions of the real estate market, they can only infer them from the prices they observe in previous periods. Before bidding for locations agents adjust their expectations according to this information.
- Sellers use prices in previous period as a reference price.
- Auctions take place simultaneously, the best bidder gets the location. Prices are computed as the expected maximum bid.
- If an agent is the best bidder for more than one location, it chooses the one that provides maximum consumer surplus. Vacant locations and unlocated agents participate in new auctions until the market clears

Figure 5.1 describes the auction and market clearing sequence in the proposed model. The relocating agents and vacant units are exogenous to the process. The relocating agents adjust their willingness to pay after observing prices in the previous period and bid for each location.

Figure 5.1: Algorithm for market clearing



The auctions determine who is the best bidder for each location and the clearing mechanism assigns agents to locations. If due to conflicts there are unlocated agents or empty units a new set of auctions takes place. The process is repeated until all agents are located or all locations are occupied. The main components of the algorithm are described next.

5.2.1 Adjustment of bids

Following (2.8) and (5.4), we assume that the deterministic part of the bid function is composed of two elements, therefore, for a particular period t:

$$B_{hi}^t = b_h^t + b_{hi}(z_i^t, \beta) \tag{5.6}$$

where b_h^t is the adjustment component that relates the bid with the utility level of the household and b_{hi} is the hedonic part of the bid expressing the value a household h gives to the attributes (z_i^t) of a location i through a set of parameters β. The functional form of (5.6) implies the assumption of a quasi-linear underpinning utility function which allows the additive decomposition and simplifies the interpretation of each element (Martínez and Henríquez, 2007). We assume the preferences of households remain constant in time, therefore the value of the hedonic part of a bid for a particular pair (b_{hi}) will remain constant in time unless there is a change in the attributes of the location (z_i^t).

The adjustment of b_h follows the logic of households changing their expectations given what they observe in the market and, therefore, increasing or decreasing their bids depending on their perceived chance of winning an auction. We define the perceived probability (q) that agent h has of winning the auction for location i in period t following the bid probability equation of (2.14), but considering that only the prices of the previous period are observable:

$$q^t(h|i) = \exp\left(B_{hi}^t - r_i^{t-1}\right) \tag{5.7}$$

For simplicity we assume the scale parameter (μ) to be equal to one. The expected probability (perceived by h) of winning any auction is the sum over all available supply (S^t) of the perceived winning probability:

$$q^t(h) = \sum_{i \in S^t} q^t(h|i) \tag{5.8}$$

Because demand for location is inelastic and agents can use only one location at the time, they try to make this probability to be equal to one. Therefore, by replacing (5.6) and (5.7) in (5.8) we get the following description of the bid adjustment that accounts for the expectations of the agent, given the market conditions described by r^{t-1}:

$$\sum_{i \in S^t} q^t(h|i) = \sum_{i \in S^t} \exp\left(b_h^t + b_{hi}(z_i^t, \beta) - r_i^{t-1}\right) = 1 \tag{5.9}$$

Clearing b_h^t we get:

$$b_h^t = -\ln\left(\sum_{i \in S^t} \exp\left(b_{hi}(z_i^t, \beta) - r_i^{t-1}\right)\right) \tag{5.10}$$

The adjustment of b_h is, to some extent, similar to adjusting or recalibrating the alternative specific constant of a logit model, in order to capture unobserved factors that describe the market conditions of the forecast scenario (Train, 2009).

It is important to notice that if supply (S^t) is large, the sum of (5.10) will tend the be large as well. This ensures that, all things being equal, larger supply will always generate smaller values for b_h. This is consistent with the expected lower overall bids that should occur when supply is abundant. In the opposite case, if supply is scarce, b_h will have bigger values. The effect of demand levels on bid adjustment is indirectly introduced through the price variable (see Appendix A). In general, supply surplus will generate low values of b_h while a demand surplus scenario will trigger increases in the value of b_h.

Equation (5.9) follows the same logic as equation (5.3), in the sense that it attempts to ensure location for every agent. However, the current equation is not solved for equilibrium because the attributes of the locations (z_i^t) and the observed prices will not change within the period t. This means that equation (5.10) does not represent a fixed point problem and, therefore, can be easily evaluated for each agent h in each period t. However, the solution of (5.10) will not ensure the location of agent h when the auctions take place.

5.2.2 Simulation of auctions and calculation of prices

After the bid adjustments have been calculated by each agent, the auctions take place with the following equation describing the probability of agent h winning auction i:

$$P^t(h|i) = \frac{\exp\left(b_h^t + b_{hi}(z_i^t, \beta)\right)}{\sum\limits_{g \in H^t} \exp\left(b_g^t + b_{gi}(z_i^t, \beta)\right)} \tag{5.11}$$

A simulation is performed generating an auction for each location i, where the best bidder will be chosen following the cumulative probability distribution defined by (5.11). Prices are the expected maximum bid of each auction considering the bids of all agents looking for a location (H^t) and the reference or asking price set by the seller. We model the asking price as the potential price the location will achieve if auctioned between all located households ($\overline{H^t}$). This is equivalent to computing the equilibrium price for each dwelling using the logsum expression of (2.13) but considering that new (re-locating) agents bid their adjusted willingness to pay.

$$r_i^t = \ln\left(\sum_{g \in \overline{H^t}} \exp\left(b_{gi}(z_i^t)\right) + \sum_{h \in H^t} \exp\left(b_h^t + b_{hi}(z_i^t)\right)\right) \tag{5.12}$$

The inclusion of the asking price is equivalent to accounting for the potential bid of other actors in the market and it is consistent with the definition of equilibrium prices of (2.13). This generates a stable price dynamic, since the prices will not depend only on the bids of active agents but also on that of all potential bidders. The prices will react to scenarios of supply or demand surplus thanks to the inclusion of the bid adjustments (b_h).

5.2.3 Clearing

Since all auctions are simulated simultaneously it is possible to find agents that are best bidders for more than one location and agents that could not win any auction. The clearing process sorts agents and locations by solving conflicts and determines which agents and locations will go through a new sequence of auctions.

If an agent is the best bidder in more than one auction it will choose the location that provides maximum consumer surplus (CS), defined as the difference between the willingness to pay for the location minus the (equilibrium) price defined in (5.12):

$$CS_{hi}^t = B_{hi}^t - r_i^t \tag{5.13}$$

If desired, a probabilistic approach can be applied here too, by computing the probability of choosing a particular location as:

$$P(i|h) = \frac{\exp(B_{hi}^t - r_i^t)}{\sum\limits_{j \in S(h)} \exp(B_{hj}^t - r_j^t)} \tag{5.14}$$

where $S(h)$ is the set of locations where agent h was the best bidder. The probability of (5.14) can be used to generate a choice using simulation. Because all active agents bid simultaneously for all locations, the order in which winning bidders are drawn does not affect the outcome. After the selection is made, the agent and the location are taken out of the pool. After repeating this process for all locations, some of them will be empty because they were discarded by a winning agent and some agents will remain unlocated because they weren't best bidders in any auction. This set of empty locations and unlocated agents enters a new auctioning sequence where the clearing process is repeated until all conflicts are solved (when either all agents are located or all locations occupied). Bids could be re-adjusted during the new auction sequences, however in this formulation we assume that the willingness to pay of agents will remain fixed during the matching process.

This mechanism has the advantage of simulating market clearing in a realistic way. The bid adjustment has the effect of avoiding under or over-bidding of households, therefore the best bidder for a particular location is also likely to be an agent that perceives a high utility in that particular location. This is consistent with economic theory and similar to the situation observed in equilibrium models, although the relation between best bid and maximum utility is not absolute, due to the asymmetries of information and temporal lag in attribute perception.

The adjustment of the bids before entering the auctions can also be understood as a (simplified) way to model strategic behavior by households. Because of the simultaneous bid-adjustment for all households, each of them is less likely to overbid and end up winning auctions for unattractive locations or, in the opposite case to underbid and end up loosing systematically in all auctions. A more explicit model for strategic behavior could allow households to participate in different sequential auctions and re-adjust their bids in each iteration as a function of the observed winning prices. However, there is now evidence on how the auctioning and expectation adjustment process takes place in practice. The proposed method aims at being an operational model for these processes without incurring in excessive complexity.

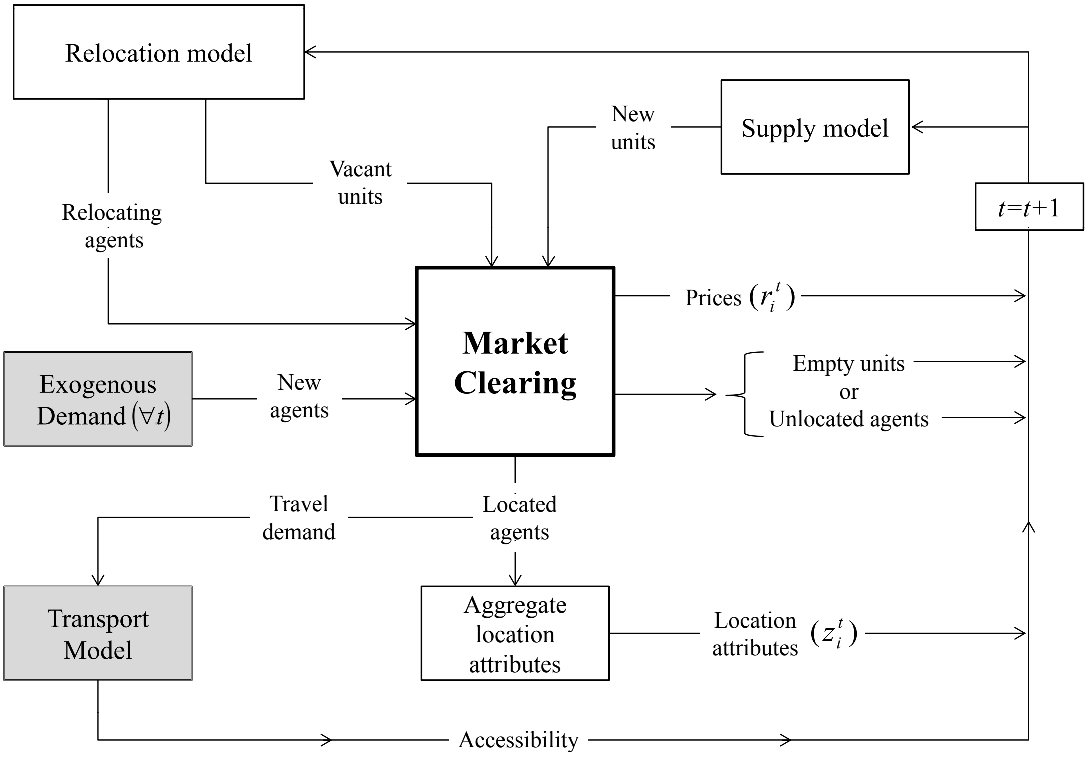
5.3 General framework for land use modeling

In order to operationalize our proposed methodology we insert the market clearing method described in the previous section within a broader framework for land use modeling that considers a series of modules accounting for each of the main models: Demand, Supply, Transport, Relocation and Market Clearing. Figure 5.2 shows the different modules and how they interact in one simulation period. In this implementation sub-models are simplified and follow observed distributions instead of having a behavioral approach, with the exception of the residential market clearing (and location choice) model. However, the framework supports any behavioral assumption for each of the sub-models and could be improved in further implementations.

We consider an exogenous demand module that, for each simulation period, generates a random set of new agents that enter the system, following control totals defined by official population statistics coming from Census projections. Relocating agents are drawn randomly in each period, following an exogenous (and fixed) relocation rate for the whole region. If there are unlocated agents from the previous period they are added to the relocating agents set. The new agents, together with those agents that are relocating, conform the demand. The uniform sampling protocol for both models generates a set of new and relocating households with the same distribution of socioeconomic attributes of the observed population.

Supply is determined by a model that, in each period, generates enough supply to satisfy the total demand, with a spatial and building type distribution that follows that of the supply in the base year. This is a strong simplifying assumption but generates a supply distribution which is close enough to that observed in reality for the purposes of this experimental implementation.

Figure 5.2: General modeling framework



The supply model also keeps track of unoccupied supply and makes it available for the next period.

Demand and supply interact in the market clearing model where the mechanism described in Section 5.2 takes place and auctions are simulated. The results is a set of sorted agents and locations together with prices, updates attributes of the locations and a set of empty units or unlocated agents (depending on the market conditions). These elements are the main input of the simulation in the next period. Because in this particular implementation the supply model is set to satisfy total demand plus a structural (relatively small) vacancy, only empty units are possible.

The transport model is exogenous and provides accessibility measures that characterize the transport system in each location or zone. In this particular implementation, the results of a MATSim (Rieser et al., 2007) simulation for the base year are used and kept constant for all periods. The accessibility is computed as the logsum of the travel (dis)utility from the zone of origin to every possible zone of destination)

For simplicity (and due to the scarce data for job location) the current implementation simulates the dynamics for the residential real estate market only. Therefore, the non-residential attributes of the locations (number and distributions of jobs by type and zone) are computed for the base year and kept constant in the simulation periods.

If data allows, all the simplifying assumptions previously mentioned can be easily relaxed and replaced by behavioral models. For this implementation the focus is placed on the behavioral models for household location and residential market clearing.

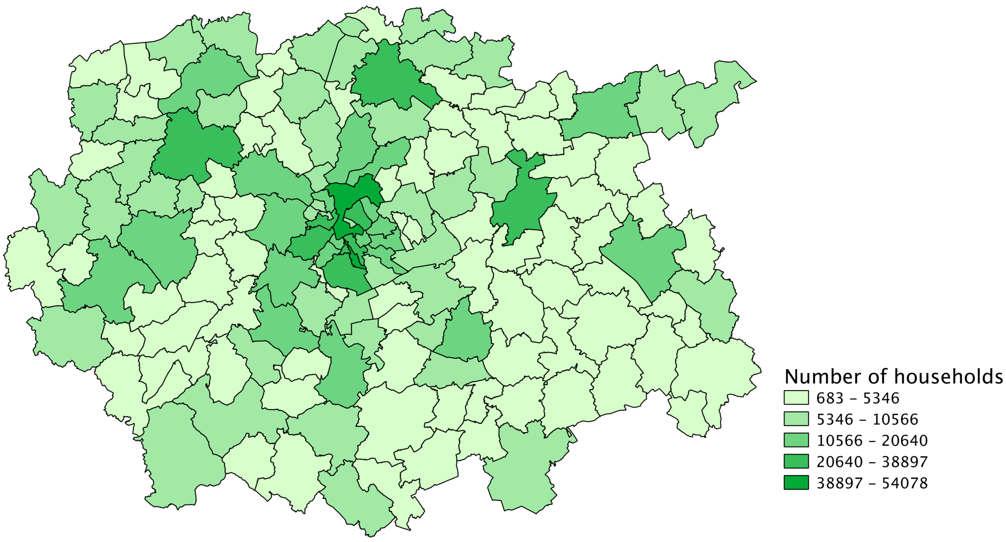
In summary, the current implementation of the proposed integrated land use and transport simulation frameworks will only model residential dynamics. Although reducing the complexity and scope of the simulation, this setting will allow to test and evaluate the market clearing model as a proof of concept, avoiding confusion with the effects that could come from the non residential market dynamics.

5.4 Case study

The proposed model is implemented for the city of Brussels, where data has been collected in the context of the European research project SustainCity[1]. The main data sources are the 2001 Population Census and a travel survey (MOBEL, Hubert and Toint (2002)) containing information about location preferences and socioeconomics for a sample of households. Information about average income per zone and average transaction prices of dwellings was also collected from the Belgian Statistical Office (StatBel). The study area considers an extended region around Brussels (see Appendix B), including 151 communes (c) that contain a total of 4945 zones (i), covering an important part of the Dutch speaking region to the north

[1] www.sustaincity.org

Figure 5.3: Number of households by commune, 2001



(Flanders) and the French speaking region to the south (Wallonia). Dwelling alternatives (v) are classified in 3 types of houses (fully detached, semi-detached and attached) and 1 type of multi-family units (apartments). Each zone is assumed to have one representative building by type (with a variable number of units in each of them), generating a total of 19780 possible location types (combinations of zones and building types).

The area of study contains a total of 1213169 households. Figure 5.3 shows the distribution of households across the communes in the area of study for the base year. Central communes (the city of Brussels) concentrate the larger amount of located households and are, at the same time, the most dense communes. Besides Brussels, the main urban agglomerations inside the area of study are Leuven (east of Brussels), Mechelen (north) and Aalst (west, north-west). Outer communes are less dense with the less populated communes located south east and south-west of Brussels city.

Given data availability the modeling period-length is a year. For the base year a synthetic population is generated (for details of the process see Farooq et al. (2011) and Farooq et al. (2012)) where individual households are described in term of their socioeconomic attributes and their location (building type and zone). For the following modeling periods, control totals coming from official estimations of population size are used to generate new households from a sample of the synthetic population. Households are characterized by their size, income level, number of children, number of workers and education level of its members. Table 5.1 describes the values for each attribute level .

The marginal distributions of attributes for the synthetic population are consistent with observed distributions coming from the census and other data sources. This consistency allows to estimate models over census data that can be implemented over the synthetic

Table 5.1: Household attributes

Attribute	levels
Income level of the household (inc_h)	1 (0-1859 Euros)
	2 (745-1859 Euros)
	2 (1860-3099 Euros)
	4 (3100-4958 Euros)
	5 (>4959 Euros)
Household size (hh_size_h)	1,2,3,4,5+
Number of children (children_h)	0,1,2+
Number of workers (workers_h)	0,1,2+
Number of cars (cars_h)	0,1,2,3+
Number of people with university degree (univ_h)	0,1,2+

population. Moreover, the generation of a synthetic population generates variables that have more information than observed marginal distributions. For example, the synthetic population provides information about income distribution within a zone that can be used instead of the average income per zone that was used in the estimation process described in Chapter 4. The synthetic population also allows to run a transport simulation for Brussels using the agent-based transport simulator MATSim. This motivates the re-estimation of the location choice (maximum-bid) model using the new available information.

5.4.1 Bid-auction model estimation

A residential auction-based location choice model is estimated using the double equation method described in Chapter 4, that adjusts the choice probabilities to the observed locations of households while simultaneously reproducing observed prices as a function of the expected maximum bid (see also Hurtubia et al. 2011 and Hurtubia and Bierlaire 2012). A total of 1346 observations from the MOBEL survey are used in the estimation. Table 5.2 describes the specification of the linear-in-parameters bid function that was finally estimated while Table 5.3 shows the estimation results, obtained with the statistical software BIOGEME (Bierlaire, 2003; Bierlaire and Fetiarison, 2009). The specification used for this model is an adaptation of the model described in Section 4.3 to the data available for simulation.

All parameters have the expected signs. The scale parameter μ has been normalized to one. Socioeconomic agglomeration effects are explained by the positive value of $\beta_{\text{high-inc}}$ and the negative value of $\beta_{\text{low-inc}}$, meaning that middle and high income prefer locations with a higher income distribution while high income households decrease their willingness to pay for a location when low income households are located in a zone. Presence of shopping, services and education increase the willingness to pay for a location while the presence of industry has a negative effect for high income households. Car accessibility has a positive effect for households with one or more cars while access to public transport attracts households with no car. Households with two or more cars have a low willingness to pay for locations with

Table 5.2: Bid function specification

Parameter	spatial attribute	×	household (hh) attribute
ASC_2	-		income level 2 constant (745-1859 Euros)
ASC_3	-		income level 3 constant (1860-3099 Euros)
ASC_4	-		income level 4constant (3100-4958 Euros)
ASC_5	-		income level 5 constant (>4959 Euros)
β_{house}	dummy for houses (types 1,2 or 3)	×	dummy for hh_size$_h$ > 2 and inc$_h$ > 2
$\beta_{\text{apartment}}$	dummy for apartment (type 4)	×	dummy for hh_size$_h$ > 2 and inc$_h$ > 2
β_{surface}	surface of dwelling v in zone i (m^2)	×	logarithm of hh_size$_h$
$\beta_{\text{high-inc}}$	% of hh's of income level 4 and 5 in commune c	×	dummy for income inc$_h$ > 2
$\beta_{\text{low-inc}}$	% of hh's of income level 1 and 2 in commune c	×	dummy for income inc$_h$ > 3
$\beta_{\text{education}}$	density of education jobs in commune c	×	dummy for univ$_h$ > 0
β_{industry}	% of industry jobs in commune c	×	dummy for inc$_h$ > 3
β_{service}	% of service (office and hotel) jobs in zone i	×	dummy for workers$_h$ > 0
β_{shopping}	density of retail jobs in zone i	×	dummy for income inc$_h$ > 2
β_{pubtrans}	public transport acces$_i$ (facilities/km^2)	×	dummy for cars$_h$ = 0
$\beta_{\text{pubtrans2}}$	public transport acces$_i$ (facilities/km^2)	×	dummy for cars$_h$ > 1
$\beta_{\text{car-access}}$	car accessibility in zone i (MATSim)	×	dummy for cars$_h$ > 0

Table 5.3: Estimation results

Parameter	Value	Std error	t-test
ASC_2	-0.171	0.083	-2.07
ASC_3	-0.461	0.113	-4.1
ASC_4	2.05	0.374	5.47
ASC_5	2.19	0.385	5.68
β_{house}	-0.128	0.0472	-2.7
$\beta_{\text{apartment}}$	-0.702	0.181	-3.88
β_{surface}	0.002	0.001	2.6
$\beta_{\text{high-inc}}$	3.97	1.24	3.21
$\beta_{\text{low-inc}}$	-3.94	0.701	-5.62
$\beta_{\text{education}}$	0.356	0.127	2.8
β_{industry}	-0.562	0.25	-2.25
β_{service}	0.046	0.020	2.31
β_{shopping}	0.040	0.018	2.24
β_{pubtrans}	0.257	0.094	2.72
$\beta_{\text{pubtrans2}}$	-0.249	0.101	-2.46
$\beta_{\text{car-access}}$	0.007	0.004	1.9*
α	-8.98	5.82	-1.54*
γ	1.46	0.421	3.46
σ	-1.93	0.022	-89.42

* Parameter not significant at the 95% level

high access to public transport, probably due to the street priority of the former over private modes.

The parameters α, γ and σ are coefficients for the latent auction price model that computes prices p_{vi} as a function of the expected maximum bid for each real estate unit, following:

$$\ln(p_{vi}) = \alpha + \gamma \cdot r_{vi} \tag{5.15}$$

where r_{vi} is the logsum of all potential bids for a unit of type v located in zone i as described by equation (2.13). The price model is applied at the zone level (i) but, given the available data, it is estimated only at the commune level, therefore replacing p_{vi} for p_{vc} in estimation mode.

5.4.2 Simulation results

Simulations are run for a period of 8 years, from 2001 to 2008. The reason to select this period is the availability of validation data regarding transaction prices for residential units and population at the commune level. Results are analyzed in the following sections for each of these variables and for the spatial distribution of income, which is analyzed on its relation to real estate prices.

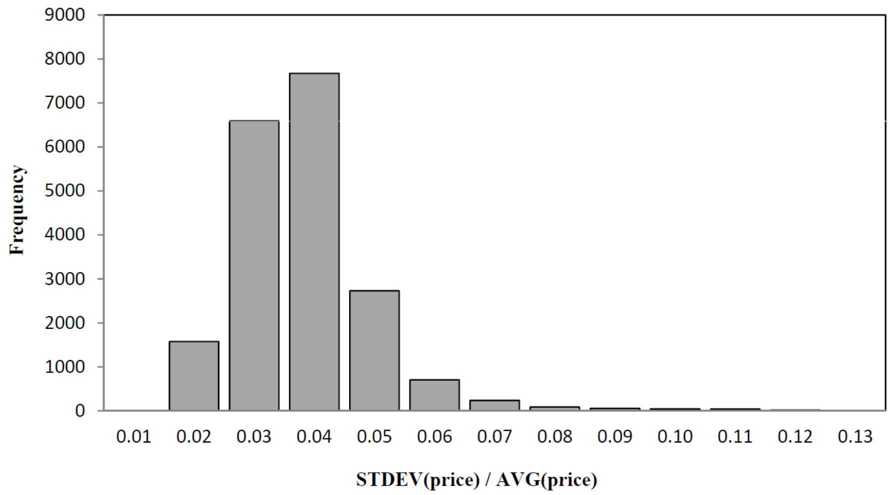
Additionally, the stability of the simulation is tested by running 100 full simulations with different random seeds and computing the standard deviation of the simulated prices in the last simulation year. Figure 5.4 shows the distribution of the standard deviation over the average value of the price for all 19781 combinations of building types and zones. Most values have a remarkably low standard deviation, with more than 94% of prices presenting a standard deviation of less than 5% of the average value. This confirms the stability of the simulation and indicates that results are not path dependent.

Price dynamics

Figures 5.5 and 5.6 show the evolution of prices at the zone level for each of the simulation years. The values are shown as the difference between the simulated prices and the observed average prices in absolute and relative terms respectively. The boxplot graphic describes the span for each quartile of the values with the central box containing the 50% of values that are closer to the median error. Since observed prices are also affected by elements that are not considered by the model presented in this thesis, like inflation or interest rate effects, we normalize all prices to the average of the observed price in the base year. This allows to analyze the relative variation of prices.

The simulation begins with a small average underestimation of prices that turns into an average overestimation for later periods. This is explained by the fact that simulation prices

Figure 5.4: Standard deviation of prices in 2008 (100 simulations)



will always increase due to the increase of population and its postive effect on the expected maximum (see equation 5.12). The quality of the forecast for prices is relatively stable across time, with the increase in the extreme values of year 2005 explained by a large increase of the population for that particular year. The relatively large errors are explained by the fact that the plot shows the difference between average price by zone with the observed average price by commune. Therefore there is natural heterogeneity within the zones inside communes

In terms of quality of the price forecast a more aggregate analysis is shown Figures 5.7 and 5.8, comparing commune-level forecasted (simulated) prices and observed prices for years 2001 and 2008 respectively.

Prices in 2001 are concentrated in the range between 70000 and 140000 Euros. However in year 2008 a group of communes goes beyond the 150000 Euros threshold with the simulated prices following this trend, although systematically underestimating the magnitude of the price increase. The higher prices appear in the communes of Woluwe-Saint-Pierre, Woluwe-Saint-Lambert, Lasne and Ukkel, some of the richest communes in Belgium, measured by the average taxable income of their inhabitants. The sistematic underestimation of prices in 2008 for these communes may be due to the fact that the uniform increase in supply mis-predicts the number of new units by type, possible under predicting new supply of the more expensive types. Additionally, the log-normal functional form of the regression price measurement (5.15) may have an effect in the predicted values because, although reducing variance in the estimation process, it may tend to under-predict larger values.

It is important to notice that the model was estimated over a reduced number of observations (1346) where some of the communes are not included. Therefore, despite results having a fit that is far from good, the trend-following pattern is an indicator of the good quality of the

Figure 5.5: Error in price forecast, 2001 - 2008

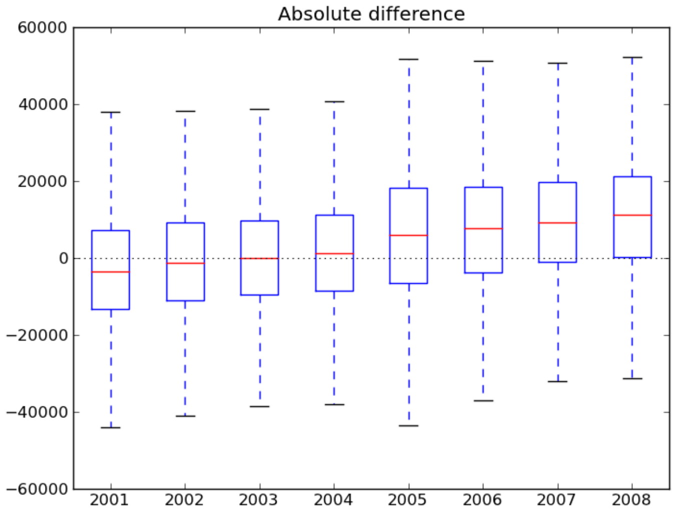


Figure 5.6: Error in price forecast, 2001 - 2008 (%)

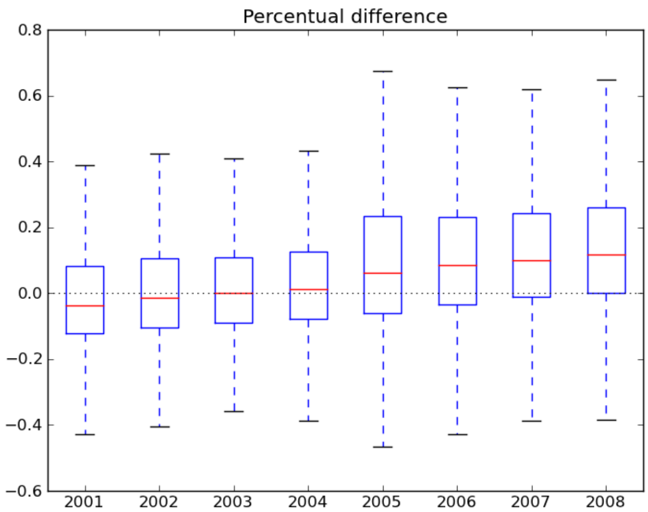


Figure 5.7: Forecasted vs observed price by commune (2001)

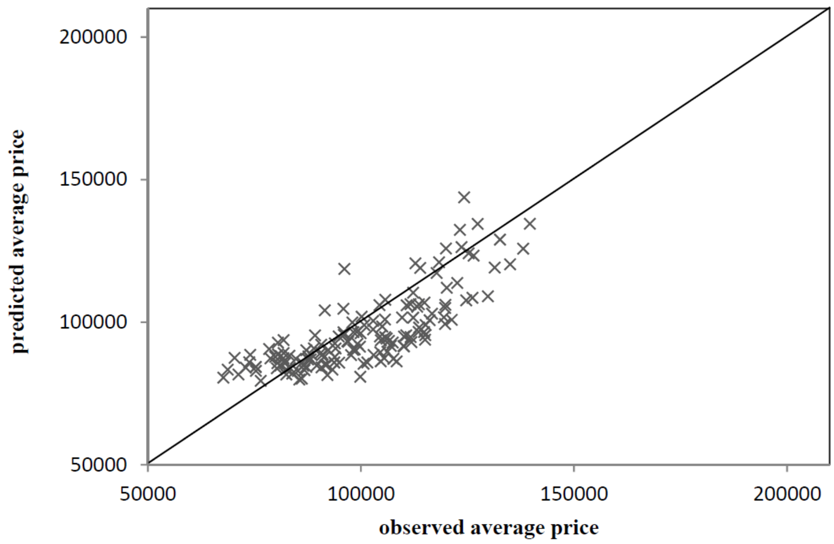
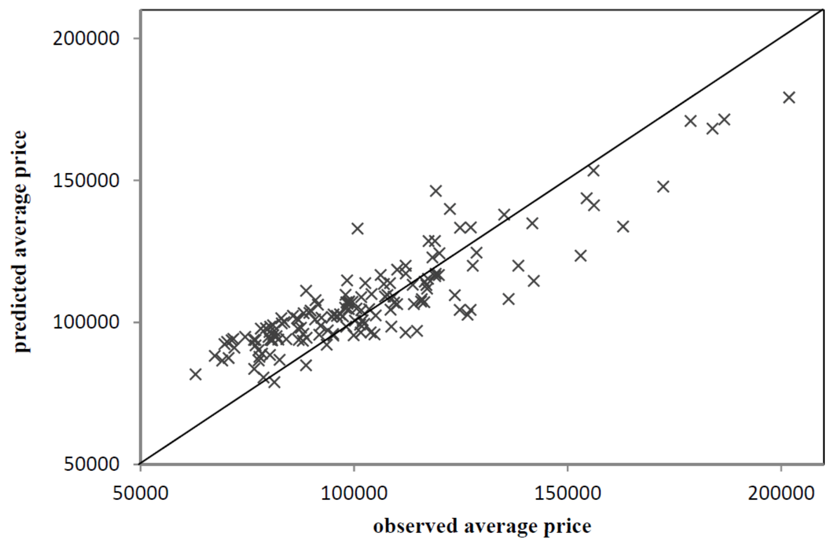


Figure 5.8: Forecasted vs observed price by commune (2008)

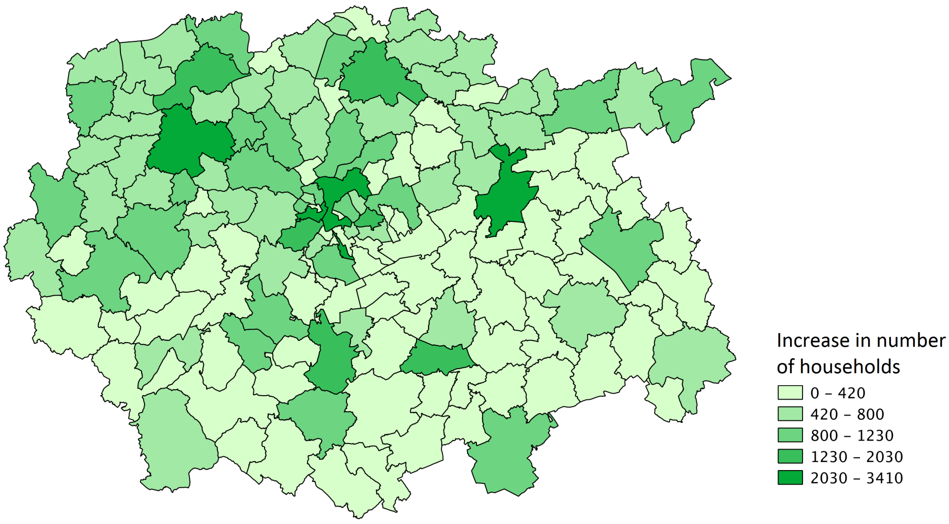


approach.

Population distribution

Figure 5.9 shows the increase in the number of households by commune. The location of new households follows the spatial distribution of the new supply. Since the generation of new supply follows the observed distribution, most new households are located in communes that originally presented high density, specially in the urban areas. The simulation does not take into account land use regulations or development constraints and results could be clearly improved by doing so. However, results are consistent with the observed trend of increase in rural areas of the Flanders regions.

Figure 5.9: Increase in number of households by commune, 2001-2008

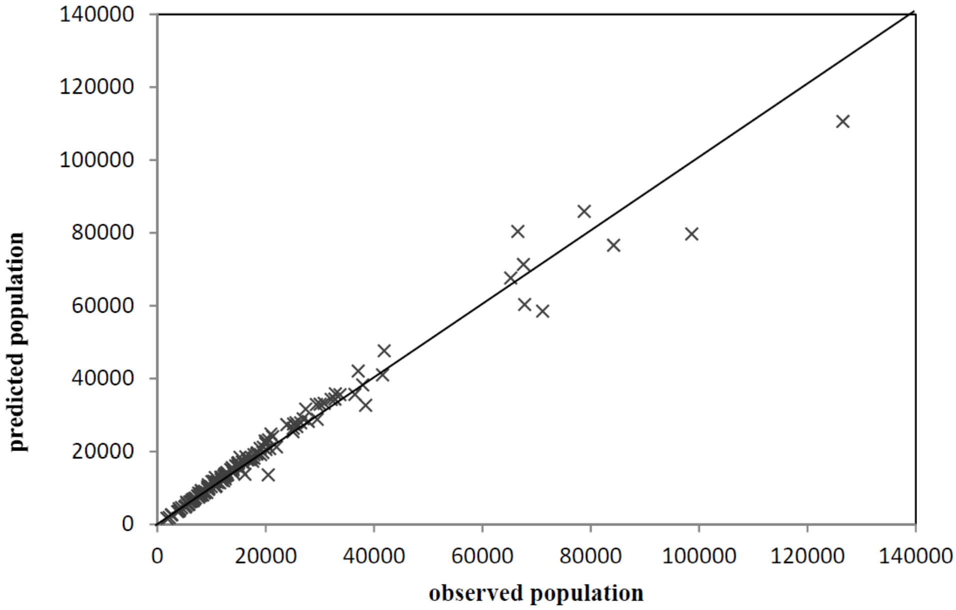


Since the simulation locates households that can have different sizes, the quality of the forecast number of people by commune is an indicator of the capacity of the model to forecast the spatial distribution of agents according to socioeconomic attributes. Figure 5.10 shows the comparison between simulated and observed data for the number of people by commune in 2008. The simulation predicts with very good fit for smaller communes and underestimates the population in 13% for the largest commune (Brussels). However, this underestimation is in part explained by the original underestimation of the number of people in this commune (by approximately 10%) in the synthetic population for the base year.

Income distribution

There is no available data for validation regarding the number of households by income level and zone in 2008, but information on the average income by commune is available from tax declarations. Predicted average income per commune can be roughly estimated from the

Figure 5.10: Forecasted vs observed population by commune (2008)



simulation results using the observed average income (in Euros) per income level from the MOBEL database. The relevant variable to analyze is the variation in the average income by commune, as a proxy of change in the income distribution. Figure 5.11 shows a comparison between observed and predicted variation in the average income by commune. Although the simulation tend to overestimate the increase in the average income, it is clear that results follow the trend observed in reality, predicting large and positive variations for the communes with the greatest increase in observed income.

The correct prediction of the trend in change of the income distribution explains the quality of the price forecast results shown in Section 5.4.2. Figure 5.12 shows the difference between 2001 and 2008 of the ratio between the increase in the number of rich households (with income higher than 3100 Euros) and the increase in the number of poor households (with income lower than 1860 Euros). A darker color indicates a predicted increase in the proportion of rich households in the commune while a lighter color indicates an increase in the proportion of poor households. Figure 5.13 shows the relative increase of the average real estate prices by commune.

There is a clear pattern of relative increase in income in communes where a relative increase in price also takes place, as seen in figure 5.14. This is due to the fact that the willingness to pay for a dwelling increases with a high income in the location and decreases with the presence of low income households (see Tables 5.2 and 5.3).

Figure 5.11: Observed vs simulated variation in average income per commune

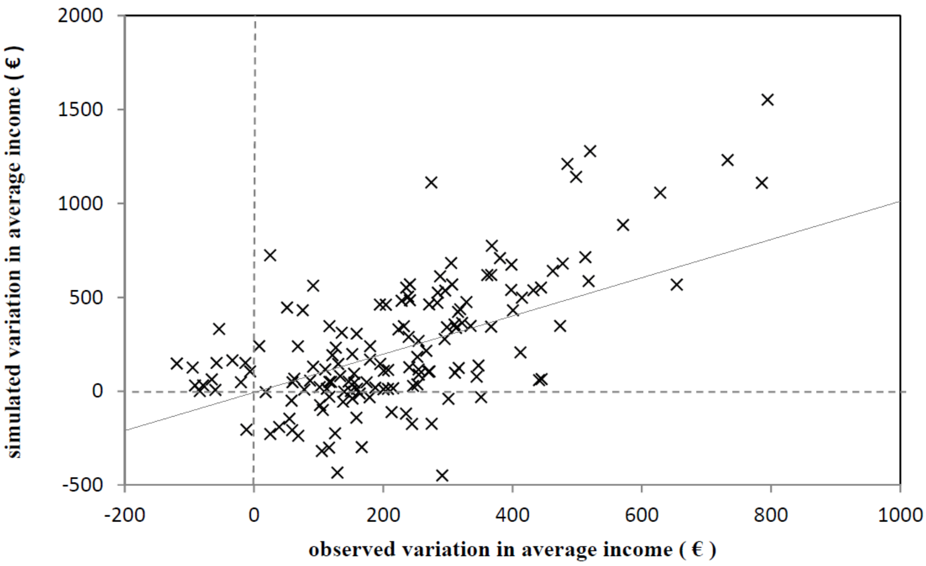
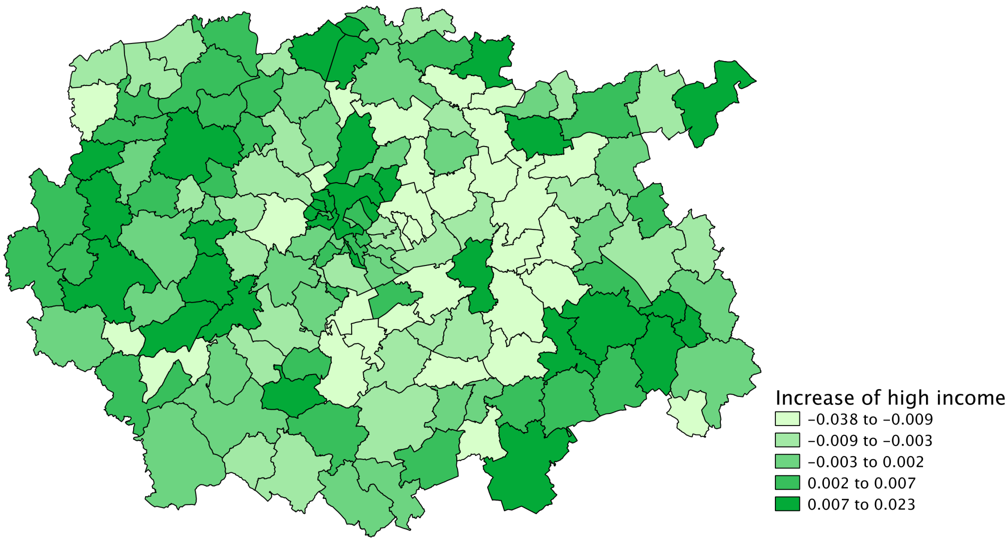


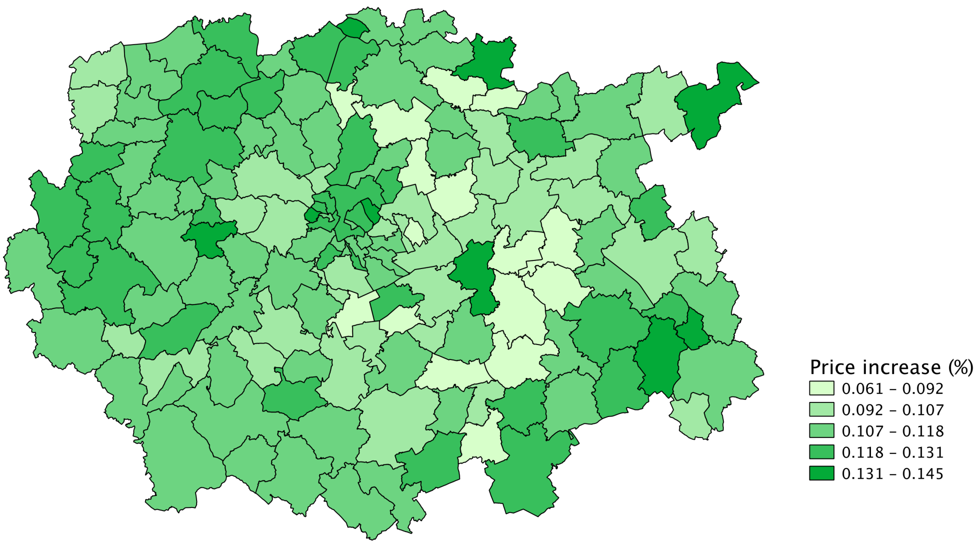
Figure 5.12: Predicted variation in the rich/poor ratio per commune, 2001-2008



5.5 Conclusions

A method for market clearing in agent-based location choice simulations is proposed. The method is based in a bid-auction approach and assumes that agents adjust their perceived expected utility by observing market prices before entering the simulation. The adjustment translates in a correction of each agent bid levels as they attempt to ensure their location. Actions for each real estate good are simulated. Prices are computed as the expected maximum

Figure 5.13: Predicted increase in price (%) by commune, 2001-2008



bid of all agents in the market.

The method is feasible to be implemented in agent-based simulations because it does not require to solve an equilibrium. Although, through the adjustment of the bids, it follows the direction of an equilibrium, without reaching it. Results are stable across simulations with different random seeds, suggesting that the process is not path dependent.

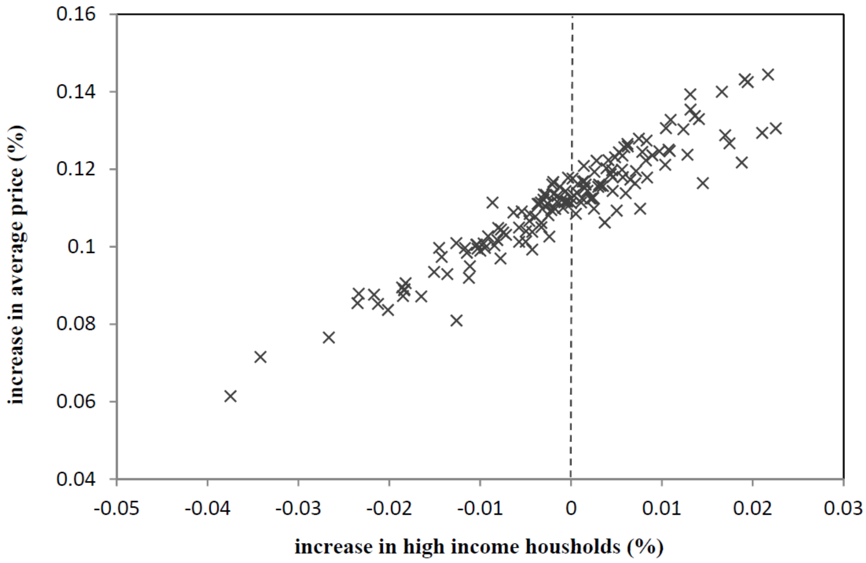
The market clearing method is embedded in a larger land use simulation framework and is applied to a real case study in the city of Brussels considering the year 2001 as base year. Simulation results for year 2008 are compared with observed data. The proposed model is able to forecast trends in price increase and change in the income distribution by commune that are consistent with reality. Some heteroscedasticity is observed in the spatial distribution of agents, with different variances being observed for different sectors or communes (see Figures 5.11 and 5.14). This suggest different behavior for different segments of agents or zones that could be modeled with different sets of equations and parameters.

The model implementation described in this chapter considers many simplifications for the supply-generation and non-residential location components of the general framework. The amount of error that can be explained by these simplifying assumptions is not clear and should be analyzed in future work.

The functional form of the price measurement equation (5.15) seems to have a biasing effect when used in price forecasting. Future work should consider exploring different functional forms and distributions to relate observed and predicted prices.

This framework can be applied to other markets or choice situations where expectations

Figure 5.14: Predicted price vs income increase by commune, 2001-2008



and competition between decision makers play an important role. An example of a possible extension is the labor market, where individuals compete for different jobs by "bidding" their experience and skills (their CV) and their salary expectations, with the job being assigned to the best "bidder". Another example is the "marriage market", where matching between individuals take place in a way that could also be modeled as a sequential (probably trial and error) bidding process. In general, any market where the assignment mechanism is an auction could be modeled with the proposed approach or extensions of it.

6 Comprehensive integrated Transport and Land Use model for Brussels

Integrated Land use and transport simulation is a complex task that requires to model several subsystem and markets and their corresponding interactions within a the city. Moreover, if the simulation is done at the agent level and with disaggregate spatial resolution, the computational complexity is increased due to the size of the problem.

UrbanSim (Waddell, 2002; Waddell et al., 2003) has become one of the most widely used platforms for land use microsimulation. It has been under development since the late 1990's and it is available as open source software. UrbanSim has a modular structure, with sub-models that can be estimated and run independently and uses a dynamic disequilibrium approach, with all sub-models interacting in a period-based temporal framework.

UrbanSim has been implemented in many cities with diverse results. Most of the application have been implemented by the developers team, including the cities of Eugene, Oregon; Salt Lake City, Utah; San Francisco, California; Seattle, Washington and San Antonio, California. Most of these implementations are reported in the literature as producing successful results in terms of accuracy of estimations and market-trend predictions (Waddell et al., 2007a, Waddell et al., 2007b, Waddell et al., 2007c). There is also an increasing number of applications implemented by researchers, urban planning offices or public authorities outside the UrbanSim core developers team. These include, among others, the case studies of Paris, France (De Palma et al., 2005); Phoenix, Arizona (Joshi et al., 2006); Volusia County, Florida (Zhao and Chung, 2006); Zurich, Switzerland (Buergle et al., 2005; Loechl et al., 2007); Lausanne, Switzerland (Patterson and Hurtubia, 2008; Patterson and Bierlaire, 2010); Lyon, France (Kryvobokov et al., 2008); Brussels, Belgium (Patterson et al., 2010); Rome, Italy (Zio et al., 2010) and Seoul, Korea (Joo et al., 2011; Hassan and Jun, 2011).

The literature reports fewer successful applications in the cases implemented by teams outside of the UrbanSim developers group, the most frequently reported problems are lack of appropriate data for the required disaggregation level, excessive data processing time, partial estimation or implementation of sub-models and inconsistency of results with observed urban trends. Many of the applications are prototype or proof-of-concept models with simplified datasets and assumptions. Some of the applications report an intensive use of human resources,in some cases excessive considering the quality of the obtained results (Nguyen-Luong, 2008).

There is few literature reporting with detail the amount of effort an application of UrbanSim is expected to require. In general, implementation of a fully fledged UrbanSim model can be very expensive in terms of data collection, time and effort for data processing, familiarization with the software platform, model estimation and implementation of the simulations. Accounting for the required implementation effort is fundamental for the potential user before deciding which tool, or even which modeling approach, fits best to the objectives of a project and, maybe even more relevant, available resources.

This chapter reports the efforts required, and results obtained, for an application of UrbanSim to the city of Brussels. The model was implemented in the context of the *SustainCity* project[1], financed by the European Research Commission with the objective of adjusting the UrbanSim modeling platform to the context of European cities, focusing on evaluation of sustainability-oriented transport and land use policies. The project considered a simultaneous application of UrbanSim to the cities of Zurich and Paris.

The chapter is organized as follows: Section 6.1 contains an overview of the data, models and mechanisms involved in UrbanSim. Section 6.2 describes in general terms the steps that are needed to implement a complete UrbanSim application. Section 6.3 describes the data, estimated models and obtained results for the Brussels case study. Section 6.4 offers a critical analysis of the UrbanSim application in terms of implementation difficulties, resources required and quality of the results. Finally, Section 6.5 concludes the chapter.

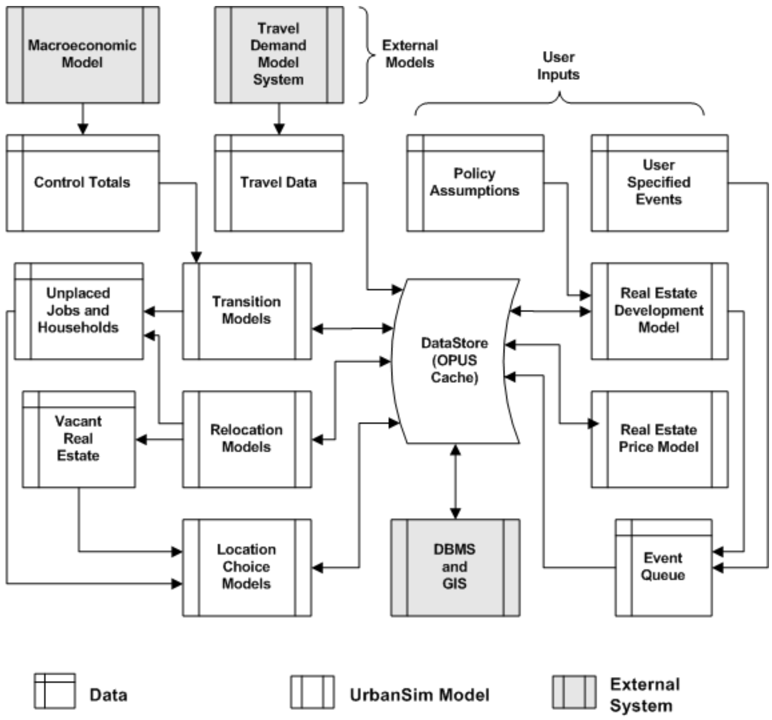
6.1 UrbanSim

UrbanSim consists of a series of models, some representing the choice of particular agents in a diverse decision scenarios and some describing market dynamics and other processes that take place in the urban context. The models interact in a simulation and data management tool (Open Platform for Urban Simulation: OPUS) that keeps tract of the agents location and real estate evolution through time. Figure 6.1 shows the data structure of UrbanSim and the involved models, that can be classified in:

- **Transition models** that describe the evolution of the demand agents (households and jobs) for each simulation period.
- **Development models** that describe the generation of new real estate supply (residential and non-residential buildings)
- **Relocation models** where the agents' decision of moving from their current location is

[1] www.sustaincity.org

Figure 6.1: UrbanSim data and model structure



Source: UrbanSim Users Guide and Reference Manual (The UrbanSim Project, 2011)

simulated

- **Location Choice models** describing the spatial allocation of new and relocating agents across the different alternatives existing in the city or the spatial distribution of new supply previously defined by the development models.
- **Price models** that compute the prices of real estate good in each period, based on the attributes of the buildings and their locations.

The models interact in each simulation period, generating a new state of the system that is used as a starting point for the simulation in the next period. Of all the models involved in UrbanSim the most relevant ones are probably the Location Choice models, because they describe the preferences of agents in terms of spatial distribution, therefore giving shape to the city and defining its dynamics and socioeconomic landscape. All location choice models in UrbanSim are based on a "choice" multinomial logit approach, where decision makers (agents) choose from a sample of available alternatives (buildings or locations) selecting the one that provides maximum utility given its attributes and price. Market clearing is treated with a "first come first served" approach (Waddell, 2010) meaning that, when two agents select the same location, conflict is solved by randomly selecting one of them.

Real estate price models are also of high relevance because they described the market value of the traded goods. In UrbanSim, real estate prices are modeled using a hedonic regression of property value per surface unit on attributes of the building and its environment and market-level vacancy rates (Waddell et al., 2003), following:

$$\ln\left(p_{vit}\right) = \alpha + \delta\left(\frac{Q_v^s - Q_{vt}^c}{Q_v^s}\right) + \beta X_{vit} \tag{6.1}$$

where $\ln(p_{vit})$ is the natural logarithm of price of land per surface unit for development type v at location i at time t, Q_{vt}^c is the current vacancy rate at time t, Q_v^s is the long-term structural vacancy rate, X_{vit} is a vector of building and location attributes, and α, δ and β are estimated parameters.

In each period, new households and firms are generated by the Transition models. Simultaneously, new supply is generated by the Development models and distributed within locations in the city. Relocating and new agents enter the market and choose their location following the distribution defined by the location Choice models. At the end of each period, prices are computed and all location and building attributes are updated to enter as the main input to the next period simulation.

One of the main characteristics of UrbanSim is the independent estimation process for each of the involved sub-models. This is a practical advantage that simplifies the implementation of the model but also implies strong assumptions about the behavior of agents and the interdependence of the decision processes that take place in the city. The modular structure and open source nature of the code makes feasible to customize UrbanSim for several different circumstances and conditions, although this may require advance knowledge of the software.

6.2 Implementing UrbanSim

Implementing UrbanSim requires to follow a series of steps to generate data, prepare the software, estimate and simulate. This section describes the process for the particular case study of Brussels, but it can easily apply to any implementation with similar characteristics. Most of these steps have to be taken sequentially (in the shown order), unless otherwise indicated. The data requirements analyzed in this section correspond to those required by a zone version of UrbanSim coupled with MATSim (Rieser et al., 2007) transport simulations.

6.2.1 Data collection

UrbanSim requires several datasets to run, the number and complexity of them varying depending on the type of application. The fundamental datasets that are required to estimate and run an UrbanSim application are those describing the basic demand agents (households, jobs) and the basic supply units (buildings). For an extensive list of agents and attributes that may be considered for land use modeling see Hurtubia, Gallay and Bierlaire (2010).

Households must be described at the individual level and exhaustively (all households in the area of study). The basic attributes that should be considered for household characterization are size (number of people), income (or income level), number of cars and residential location (building type and zone). However, if available, other basic attributes that could be considered are: the age of the head of the household, number of workers, number of children, race and education level just to mention a few. This data is usually found in censal databases but in many cases is not available at the individual level, therefore requiring synthesis of the household population. Population Synthesis procedures like Iterative Proportional Fitting (Beckman et al., 1996) require control total and aggregate statistics coming from census data and a sample of individual households with detailed attribute information. For a review of different possible population synthesis methods see Müller and Axhausen (2010).

Jobs must be described in terms of economic activity type and location. This information is usually available at an aggregate spatial level from statistical authorities and can be easily disaggregated into smaller zones and buildings following some distribution rule. UrbanSim also requires to identify when jobs are home based or not and need information regarding the average surface consumed by each type of job.

Buildings should be described in terms of type and location (zone), land area used, number of housing units and their average surface, non-residential surface and price. Additionally, UrbanSim requires data about residential and non-residential developable capacity (coming from zoning regulations) and historical data of previously built real estate developments.

Besides the main datasets additional information describing the dynamics of aggregate supply and demand is required. Control totals for households (if possible by categories of attributes) and jobs by type are necessary for all the simulation years. UrbanSim also requires data about structural real estate vacancy by building type (Q_v^s) and relocation rates for households and jobs by type. Recent developments incorporate a set of demographic models to UrbanSim in order to simulate the evolution of population within the platform. However, the inclusion of an internal demographic model is extremely data hungry and was not implemented in the reported application.

Regarding travel data, if an activity based travel demand model (like MATSim in the case of this application) is to be connected to UrbanSim, additional data about individual persons is required. Persons data must describe, at a minimum level, the residential location and work location. If a different transport model is available, accessibility data for the base year must.

Finally, additional data describing (non-dynamic) attributes of the location can be collected. For example the presence of green areas or large sport facilities like stadiums, the distance to the CBD, distance to transportation facilities (airports, train stations) or other specific landmarks.

6.2.2 Data processing

Data must be processed and organized in tables with a specific format required by UrbanSim. The fundamental UrbanSim tables are those related with the basic demand and supply elements (households, jobs and buildings), these tables concentrate most of the data and require the biggest processing effort.

Synthesis of missing data

Since the level of detail and disaggregation of required data is high, in most cases it won't be directly available from the original data source, either because it was not collected at that level or due to privacy issues. This triggers the need to simulate data, a process that must be done in a way that ensures consistency between all agents and locations.

For example, the already mentioned synthetic population of households must be consistent not only with the real population in term of socioeconomic attributes, but also in terms of location. The households of the synthetic population must be located in one particular residential unit and a unit can't be used by more than one household. This requires to simulate the location distribution of all households.

Another example is the distribution of jobs across buildings. In the reported application job data was available only at the commune level. Distribution within (smaller) zones and buildings was done through Monte Carlo simulation, following the observed distribution of available non-residential surface by zone and building type.

Adapting data to the platform

Some of the modeling assumptions of UrbanSim generate specific requirement regarding the structure of the data. For example, land use regulations and location constraints must be translated into developable capacities by zone and included in the buildings table as available non-residential surface to build and number of possible residential units to build. Similarly, the capacity of non-residential building, in terms of the number of jobs they can hold, is defined through the average surface used by each job type in each zone. All these data adaptation require assumptions regarding the average use of land and built space by the agents and the way this varies across different locations.

Additional data adaptations are those required by the software due to inherited model structure. This is translated mostly into additional identifiers for agent or supply types and, sometimes, variables that are required by default. An example of this is the need to include the age of the head of the household in the households table, even if this variable is not involved in any modeling or simulation process. In this particular application, a random age was assigned to each household.

Other datasets have a predefined data structure that may require adaptation. For example, travel data must come in the form of a zone to zone impedance (cost or travel time) matrix and UrbanSim is not prepared to deal with a different structure for this data.

Scenario generation

Sometimes the simulation scenarios that will be run consider specific conditions on supply, demand or land use regulations. Every pre-set border condition must be explicitly introduced in the corresponding UrbanSim tables that will be considered for scenarios other than the baseline. For example, a particular increase of households in some specific socioeconomic class must be explicitly included in the control totals table or pre-defined (probably large) real estate developments must be included in a new buildings table.

6.2.3 Software preparation

After installing the OPUS modeling platform, the analyst needs to familiarize with the software and set up the project before estimation or simulation is possible. Of all the processes described next, familiarization can be performed during (or even before) the data collection and processing steps.

Software familiarization

As any software, UrbanSim requires the user to become familiar with the interface and the coding language used to define the utility function specifications. This can be easily done following the tutorials available online and by running the example projects provided with the UrbanSim installation package: Eugene (gridcell), Seattle (parcel) and San Antonio (zone). These projects are previously tested and run smoothly, they can be used to practice model estimation and simulation with the latest stable version settings.

Project set up

Setting up a project requires to define all the initial settings and relations between sub-models. Every sub-model has a large set of parameters that configure them in terms of sample size for estimation and simulation, tables to consider, dependencies with other models, filtering of alternatives, participating agents, etc. The project setting can be done either by using a template and defining all parameters from scratch or by adapting a previously existing model. The last option is certainly simpler but it implies inheriting structure and setting from the previous model that can interfere in the estimation and running processes. Additionally, a database can be set up to manage data, UrbanSim supports MySQL and PostgreSQL as platforms for this. For technical details of project and database set up see Gallay (2010).

Database and software debugging

Getting UrbanSim to run is not straight forward.It is very likely that, after initial preparation, many details in the data definitions or structure are not exactly complying with the software requirements. A trial and error process is required before estimation and simulation is possible.

Because of its open source nature, the code behind UrbanSim is constantly evolving and, therefore, troubleshooting can be hard. The error messages can be sometimes cryptic and it's not always easy to identify the source of error because of the multilayer software structure. In most cases, help from the developers is required when implementing a stable version of UrbanSim. If modifications to the model structure or simulation algorithms are to be introduced the developer version should be used, in which case software debugging will be also necessary.

6.2.4 Model Estimation and scenario simulation

After setting up the data and the software, model estimation should be straight forward using the OPUS language (based on python) for model specification definition. After all sub-models have been estimated, simulations can be run and, ideally, a validation comparing observed data with forecast results obtained for a baseline scenario should be performed. However it is likely that both the model estimation and simulation processes will trigger the need to collect and process more data or re-define the way the system is simulated, and this will probably require further software adaptation and debugging . This makes likely that all four steps described in this section are re-visited to some extent simultaneously during the final implementation of the model. This is important to take into account when scheduling the application project and defining the amount of resources to use in each step.

6.3 Brussels case study

UrbanSim is applied to an extended region around the city of Brussels. The application is done in the context of the European Commission financed project SustainCity and is done in parallel with applications for the cities of Paris and Zurich (Schirmer et al., 2011).

The main data sources used for this application were the 2001 Belgium Population Census and the Belgium Land Registry (a cadastre of real estate goods). Both datasets were obtained at an aggregate level from the Belgian Statistical Authority (SPF - Economie[2]). Aggregate data regarding employment by activity type and commune was collected from the ONSS and INASTI databases, from the same source. Additionally, individual level data for households and persons was obtained from the travel survey MOBEL (Hubert and Toint, 2002), performed in the area of study during 2002.

[2] http://economie.fgov.be/en/

Table 6.1: Building types

type id	description	category
1	isolated houses	residential
2	semi-attached houses	residential
3	attached houses	residential
4	apartments buildings	residential
5	agricultural	non-residential
6	quarrying, mining	non-residential
7	industry	non-residential
8	office (private sector, including bank)	non-residential
9	shops, retail	non-residential
10	hotels, bars, restaurant	non-residential
11	government and public service	non-residential
12	education	non-residential
13	health	non-residential
14	leisure activities	non-residential

The area of study consists of 151 communes covering a large area around the Brussels Capital Region. For a list of the communes conforming the area of study see Appendix B. Each commune is subdivided in zones, defined by the Census statistical sector, adding up to a total of 4945 zones in the area of study. Each zone contains a representative building for each type that is described by aggregate attributes (number of residential units, average surface per unit, total non-residential surface, etc.). Four types of residential buildings and ten types of non-residential buildings are considered, as shown in Table 6.1.

Accessibility measures are obtained from an adaptation of the agent-based transport microsimulator MATSim to the OPUS platform (Nicolai and Nagel, 2010; Nicolai et al., 2011).

Jobs, known at the commune level by type from the original data source, are distributed across zones and buildings following the observed surface by building type. By definition, the economic activity sectors of each job category is consistent with the non-residential building types of Table 6.1.

Individual households were generated combining Census aggregate data with individual household level data from MOBEL using a synthetic population generation process (Farooq et al., 2011, 2012) that ensures consistency with observed marginal statistics by zone. Income are characterized by income level, size, number of children, education level, number of workers and number of cars categories. The levels of each category are described in table 6.2.

6.3.1 Estimation results

All models were estimated with the OPUS/UrbanSim software platform. Several different specifications were tried, given available data and following what the literature and urban

Table 6.2: Household attributes

Attribute	levels
Income level of the household (inc_h)	1 (0-1859 Euros) 2 (745-1859 Euros) 3 (1860-3099 Euros) 4 (3100-4958 Euros) 5 (>4959 Euros)
Household size (hh_size$_h$)	1,2,3,4,5+
Number of children (children$_h$)	0,1,2+
Number of workers (workers$_h$)	0,1,2+
Number of cars (cars$_h$)	0,1,2,3+
Number of people with university degree (univ$_h$)	0,1,2+

economic theory suggests as explanatory variables for each of the modeled phenomena. Final specifications were selected following estimate-significance and theoretical-consistency criteria.

For all choice models a linear-in-parameters utility function specification like the following was chosen:

$$V_i = \sum_k \beta_k x_i^k \tag{6.2}$$

where β_k is the k-th parameter to estimate and x_i^k is the k-th attribute of alternative i. For some models, like the household location choice model (see Table 6.3), x_i may be replaced by $x_{in} = x_i \cdot x_n$, describing an interaction between an attribute of the alternative i and a characteristic of the decision maker n.

The households location choice models estimation results are presented in Table 6.4. All parameters are statistically significant and have the expected signs. Price has a negative effect in the utility for all households but is stronger for households of mid and low income which is reasonable due to their smaller available income. The presence of high income households attracts other households of high income but makes locations less attractive for low income households. This is consistent with the expected social agglomeration and segregation effects usually observed in residential location. The presence of retail (reflected in the number of shopping jobs) increases the attractiveness of a location while the presence of offices has the opposite effect. Public transport accessibility has a positive effect for households with no cars while car accessibility increases the utility of car-owning households. The two spatial alternative-specific constants account for unobserved attributes that makes central locations more attractive. One of the constants is active when the location is inside the Brussels Capital Region (see Appendix B for a definition of the involved communes) while the other refers to

Table 6.3: Household location choice model specification

Parameter	location attribute	×	household attribute
ASC_{BCR}	constant for locations inside Brussels Capital Region	×	-
$ASC_{Brussels}$	constant for locations inside Brussels commune	×	-
β_{price_low}	price of dwelling type v in zone i (2001 Euros)	×	dummy for $inc_h < 3$
$\beta_{price\text{-}mid}$	price of dwelling type v in zone i (2001 Euros)	×	dummy for $inc_h = 3$
$\beta_{price\text{-}high}$	price of dwelling type v in zone i (2001 Euros)	×	dummy for $inc_h > 3$
β_{educ}	% of university degree holders in i	×	-
$\beta_{income\text{-}low}$	% of households with high income (>3) in i	×	dummy for $inc_h < 3$
$\beta_{income\text{-}high}$	% of households with high income (>3) in i	×	dummy for $inc_h > 3$
β_{office}	density of office jobs in commune c (jobs/HA)	×	dummy for $workers_h > 0$
$\beta_{shopping}$	density of retail jobs in commune c (jobs/HA)	×	dummy for $workers_h > 0$
β_{CBD}	distance to Business Central District	×	dummy for $workers_h > 0$
β_{rail}	dummy for urban rail presence within 1km	×	dummy for $cars_h = 0$
$\beta_{car\text{-}access}$	car accessibility in zone i (MATSim)	×	dummy for $cars_h > 0$

Table 6.4: Household location choice model estimation results

Parameter	Coeff	stdev	t-test
ASC_{BCR}	0.838	0.016	53.52
$ASC_{Brussels}$	0.144	0.032	4.48
β_{price_low}	-1.020	0.035	-29.28
$\beta_{price\text{-}mid}$	-0.987	0.040	-24.42
$\beta_{price\text{-}high}$	-0.955	0.084	-11.36
β_{educ}	0.039	0.001	39.10
$\beta_{income\text{-}low}$	-0.111	0.007	-15.73
$\beta_{income\text{-}high}$	0.201	0.013	14.92
β_{office}	-0.013	0.002	-6.60
$\beta_{shopping}$	0.060	0.013	4.46
β_{CBD}	-0.144	0.012	-11.69
β_{rail}	0.380	0.022	17.14
$\beta_{car\text{-}access}$	0.039	0.004	10.97
log-likelihood	-105329		

the specific commune of Brussels.

The employment location choice model is subdivided in eight sub-models, one for each type of economic activity. Table 6.5 shows the estimation results for each sub-model. Jobs in the agricultural and mining sectors are not considered for modeling purposes.

Industry jobs indicate the presence of manufacturing industry which is usually characterized by the generation of negative externalities and is attracted by the presence of jobs of the same type within a zone, as explained by the positive parameter for jobs of the same time. This is consistent with the observed agglomeration of industry, usually due to economies of scale and by the strong regulations over location of manufacturing firms. It also explains the negative effect of job density and positive effect of available surface since industrial activity usually requires large plots of land with fewer employees per surface unit than other types of activities.

Office jobs and retail jobs are of particular relevance because they appear in the utility function of the household location choice model. Office jobs prefer to locate in zones with agglomeration economies and therefore favor density of jobs of the same type. Population and job density also have a positive effect in the utility for office jobs, probably because office jobs are service providers and prefer to locate near potential clients. Retail jobs also benefit for agglomeration economies and therefore the presence of jobs of the same type and of jobs in general have a positive effect in their location preferences. The communal population density has a negative effect on retail jobs, probably because, despite having major retail agglomerated in central communes, shopping facilities are present in all communes, even those with low population density, this is consistent with the positive parameter for presence of high income households who tend to locate in communes with lower density.

In general, all jobs are attracted by the presence of jobs in the same economic field, because of agglomeration economies and have a positive parameter for the presence of residential surface and total job density. This is overall consistent with the observed behavior and location of non-residential activities.

Estimation results for the residential development Location Choice model are presented in Table 6.6. The behavioral assumption in UrbanSim is that developers try to maximize their profit and therefore will locate new developments in zones with high prices and attributes that will attract prospective households. The models are estimated over data for individual real estate developments that took place in the ten year period previous to the base year and, therefore, are not representative of all existing supply in the city. Because of endogeneity problems, price is not always feasible to be included and proxy variables that describe the quality of the locations are used instead. In general houses are built in zones with presence of green areas while apartments tend to be built in (probably more urbanized) zones with less green areas. All types of residential development tend to agglomerate and therefore have a positive parameter for the logarithm of the number of buildings of the same type. The distance to the CBD has an interesting effect in the location of new buildings: while isolated and semi-detached houses prefer to be built near the CBD, smaller (attached) houses and

Table 6.5: Employment location choice model

Industry (n=13699)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
jobs_ind_per_surf	Density of jobs in industry sector in commune	0.0757	0.0028	27.23
ln_jobs_den_zone	Logarithm of jobs density in zone	-0.0557	0.0082	-6.82
ln_non_res_sqft	Logarithm of non residential surface	1.2585	0.0102	123.18
AIC=29436	***Log-likelihood***=-14715			
Office (n=14875)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
jobs_off_den_com	Density of jobs in private sector (office) in commune	0.0202	0.0031	6.54
ln_jobs_den_zone	Logarithm of jobs density in zone	0.6542	0.0090	72.52
pop_den_com	Population density in commune	0.1285	0.0094	-13.66
ln_non_res_sqft	Logarithm of non residential surface	0.5200	0.0072	72.10
AIC=49186	***Log-likelihood***=-24587			
Retail (n=3821)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
per_high_inc	Percentage of households in high income scale (>3) in commune	0.0508	0.0067	7.53
jobs_ret_den_com	Density of jobs in retail sector in commune	0.2829	0.0346	8.19
ln_jobs_den_zone	Logarithm of jobs density in zone	0.0719	0.0143	5.04
pop_den_com	Population density in commune	-0.1348	0.0320	-4.21
ln_non_res_sqft	Logarithm of non residential surface	0.8932	0.0177	50.51
AIC=13945	***Log-likelihood***=-6967			
Hotels/Bar/Restaurants (n=1950)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
jobs_hbr_den_com	Density of jobs in hotels/bar/restaurants in commune	0.1738	0.0222	7.81
ln_jobs_den_zone	Logarithm of jobs density in zone	0.3371	0.0158	21.27
ln_dist_cbd	Logarithmic distance of zone from CBD	-0.2153	0.0352	-6.11
pop_den_com	Population density in commune	-0.1989	0.0229	-8.69
ln_non_res_sqft	Logarithm of non residential surface	0.3376	0.0144	23.53
AIC=10649	***Log-likelihood***=-5319			
Government and public service (n=8359)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
jobs_off_den_com	Density of jobs in private ofiices in commune	0.0204	0.0028	7.27
ln_jobs_den_zone	Logarithm of jobs density in zone	0.7735	0.0128	60.21
pop_den_com	Population density in commune	-0.0777	0.0124	-6.27
ln_non_res_sqft	Logarithm of non residential surface	0.5073	0.0115	44.00
AIC=23607	***Log-likelihood***=-11798			
Education (n=3908)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
jobs_edu_den_com	Density of jobs in education sector in commune	0.2813	0.0160	17.60
ln_job_den_zone	Logarithm of jobs density in zone	0.1844	0.0167	11.03
ln_dist_cbd	Logarithmic distance of zone from CBD	0.2006	0.0309	6.50
pop_den_com	Population density in commune	-0.1352	0.0201	-6.72
ln_non_res_sqft	Logarithm of non residential surface	0.8213	0.0176	46.62
AIC=13176	***Log-likelihood***=-6583			
Health (n=5252)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
per_high_inc	Percentage of households in high income scale (>3) in commune	0.0570	0.0057	9.99
jobs_health_den_com	Density of jobs in health sector in commune	0.1601	0.0103	15.53
ln_job_den_zone	Logarithm of jobs density in zone	0.4404	0.0123	35.86
ln_dist_cbd	Logarithmic distance of zone from CBD	0.3730	0.0236	15.78
pop_den_com	Population density in commune	-0.0687	0.0197	-3.49
ln_non_res_sqft	Lof of non residential surface	0.4908	0.0119	41.29
AIC=23054	***Log-likelihood***=-11521			
Leisure activities (n=1354)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
per_high_inc	Percentage of households in high income scale (>3) in commune	0.0825	0.0125	6.60
jobs_leis_den_com	Density of jobs in leisure sector in commune	0.3673	0.0184	19.98
ln_dist_cbd	Logarithmic distance of zone from CBD	-0.2198	0.0448	-4.91
pop_den_com	Population density in commune	0.0956	0.0300	3.19
ln_non_res_sqft	Lof of non residential surface	0.6439	0.0241	26.67
AIC=4967	***Log-likelihood***=-2479			

apartments appear with a positive parameter for distance. This can be explained by the fact that land tends to be more expensive near the CBD while smaller (and cheaper) supply will tend to locate in zones with a lower land costs.

The estimation results for the location choice model of non-residential real estate developments are shown in Table 6.7. New non-residential supply tends to locate in places that already show agglomeration and with high concentration of other activities in general. Locations with good car and public transport accessibility tend to be attractive for the location of new developments.

Finally, results for the real estate price model are shown in Table 6.8. Two models are estimated: one for houses and one for apartments. The house price model has a better fit than the apartment price model, which has a very low r-squared indicator. This is due to the estimation data, where only communal average of transactions for these two types are available, but a much larger number of observations is available for houses (due to the existence of three types of house). There is no available observation of non-residential real estate prices and, therefore, no model is estimated for this market.

The price of houses and apartments is positively affected by the presence of green areas and by socioeconomic attributes of the zone or neighborhood, like the presence of households with a high education level or high income. In general, the presence of non-residential activities has a negative effect on the price, with the notable exception of education that makes locations more attractive, probably due to a higher access to schools. In general, real estate supply will have higher prices in location that are nearer to the Central Business District.

The real estate price model presented here is mostly based on location (neighborhood or commune) attributes with no building-specific attributes. This is due to the fact that the only available data characterizing building is surface. UrbanSim, by default, computes prices by surface unit and therefore inclusion of surface is not possible in the hedonic model specification. Despite this problem and the fact that the literature shows that prices are largely explained by attributes of the buildings (see for example Loechl and Axhausen (2010)), the presented models are still able to capture land use effects that should be relevant for the modeling purposes.

6.3.2 Simulation results

For validation purposes, the simulation is run from 2001 to 2008 considering a baseline (business as usual) scenario. Figure 6.2 shows that new supply, and consequently, households, are located in the major urban areas around Brussels, Leuven, Mechelen and Aalst. This is consistent with the observed trend. UrbanSim predicts a very small variation (a maximum difference of 1,4%) in the income distribution within each commune for the simulation period. Figure 6.3 shows the difference between the rich to poor income ratio of 2008 and 2001 by commune. Higher values indicate a relative increase of rich households in the commune,

Table 6.6: Residential development project location choice model

Detached (n=6112)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
car_accessibility	Car accessibility in zone	0.3066	0.0117	26.16
green_score	Green area score	0.6424	0.1136	5.66
per_high_inc	Percentage of high income (>3) households in commune	0.0230	0.0061	3.79
jobs_ret_den_c	Density of jobs in retail sector in commune	0.3461	0.0346	10.01
ln_detached_house_units	Logarithm of number of detached house units	0.6364	0.0120	53.25
ln_dist_cbd	Logarithmic distance of zone from CBD	-0.2821	0.0314	-8.97
pop_den_com	Population density in commune	-0.4340	0.0321	-13.50
AIC=42873				
Log-likelihood=-21429				
Semi-detached (n=14007)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
car_accessibility	Car accessibility in zone	0.1406	0.0074	18.96
green_score	Green area score	1.1384	0.0889	12.81
per_high_inc	Percentage of high income (>3) households in commune	-0.0298	0.0055	-5.38
jobs_ret_den_c	Density of jobs in retail sector in commune	0.0919	0.0424	2.17
ln_semi-detached_units_price	Logarithm of price of semi-detached houses	-0.5416	0.0905	-5.98
ln_semi-detached_house_units	Logarithm of number of semi-detached house units	1.0423	0.0089	117.48
ln_dist_cbd	Logarithmic distance of zone from CBD	-0.3289	0.0271	-12.14
pop_den_com	Population density in commune	-0.5547	0.0486	-11.41
AIC=88913				
Log-likelihood=-44449				
Attached (n=59558)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
car_accessibility	Car accessibility in zone	-0.0112	0.0028	-4.02
green_score	Green area score	1.4831	0.0448	33.10
per_high_inc	Percentage of high income (>3) households in commune	-0.0096	0.0020	-4.91
jobs_ret_den_c	Density of jobs in retail sector in commune	-0.1433	0.0293	-4.88
ln_attached_units_price	Logarithm price of attached houses	0.8180	0.0366	22.37
ln_attached_house_units	Logarithm of number of attached house units	0.2008	0.0031	65.50
ln_dist_cbd	Logarithmic distance of zone from CBD	0.1797	0.0145	12.43
pop_den_com	Population density in commune	-1.0104	0.0444	-22.78
AIC=458860				
Log-likelihood=-229422				
Apartments (n=5119)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
car_accessibility	Car accessibility in zone	0.0508	0.0121	4.20
green_score	Green area score	-1.5951	0.1294	-12.32
per_high_inc	Percentage of high income (>3) households in commune	-0.0184	0.0074	-2.47
ln_apart_units_price	Logarithm price of apartements	0.7437	0.0883	8.42
ln_apart_house_units	Logarithm of number of apartment units	1.0340	0.0123	84.04
ln_dist_cbd	Logarithmic distance of zone from CBD	0.0026	0.0347	0.08
pop_den_com	Population density in commune	0.1622	0.0284	5.71
AIC=24299				
Log-likelihood=-12143				

Table 6.7: Non-residential development project location choice model

Industry (n=2770)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
car_accessibility	Car accessibility in zone	0.0556	0.0126	4.42
jobs_ind_den_c	Density of jobs in industrial sector in commune	-0.1108	0.0118	-9.43
ln_dist_cbd	Logarithmic distance of zone from CBD	0.3665	0.0535	6.85
pop_den_com	Population density in commune	0.7032	0.0585	12.03
ln_jobs_in_zone	Logarithm of total number of jobs in zone	0.4034	0.0126	32.04
ln_pop_in_zone	Logarithm of total number of population in zone	0.2214	0.0141	15.70
AIC=21472	***Log-likelihood***=-10730			
Office (private sector) (n=767)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
car_accessibility	Car accessibility in zone	0.0566	0.0277	2.04
jobs_off_den_c	Density of jobs in private sector in commune	0.0229	0.0076	3.03
pop_den_com	Population density in commune	-0.2611	0.0540	-4.83
ln_jobs_in_zone	Logarithm of total number of jobs in zone	1.1792	0.0338	34.86
ln_pop_in_zone	Logarithm of total number of population in zone	-0.1722	0.0283	-6.08
AIC=3947	***Log-likelihood***=-1968			
Shops (n=1466)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
car_accessibility	Car accessibility in zone	0.0565	0.0171	3.30
jobs_ret_den_c	Density of jobs in retail sector in commune	-0.1082	0.0583	-1.86
ln_dist_cbd	Logarithmic distance of zone from CBD	0.6581	0.0649	10.15
pubtrans_score	Public transport score	1.3846	0.1929	7.18
ln_jobs_in_zone	Logarithm of total number of jobs in zone	0.4895	0.0222	22.01
ln_pop_in_zone	Logarithm of total number of population in zone	0.3880	0.0313	12.39
AIC=10662	***Log-likelihood***=-5325			
Hotels, bar, restaurants (n=107)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
jobs_hbr_den_c	Density of jobs in hotels/bar/restaurants in commune	0.1952	0.1023	1.91
pop_den_com	Population density in commune	-0.3668	0.1343	-2.73
ln_jobs_in_zone	Logarithm of total number of jobs in zone	0.7706	0.0733	10.52
AIC=724	***Log-likelihood***=-359			
Government and public service (n=264)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
ln_dist_cbd	Logarithmic distance of zone from CBD	0.4491	0.0817	5.50
pubtrans_score	Public transport score	2.0769	0.4621	4.49
ln_jobs_in_zone	Logarithm of total number of jobs in zone	0.7448	0.0521	14.29
ln_pop_in_zone	Logarithm of total number of population in zone	0.0915	0.0509	1.80
AIC=1850	***Log-likelihood***=-921			
Education (n=140)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
ln_jobs_in_zone	Logarithm of total number of jobs in zone	0.3639	0.0588	6.19
ln_pop_in_zone	Logarithm of total number of population in zone	0.3647	0.0602	6.06
AIC=1072	***Log-likelihood***=-534			
Health (n=225)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
green_score	Green area score	1.4191	0.5441	2.61
ln_jobs_in_zone	Logarithm of total number of jobs in zone	0.3942	0.0665	5.93
ln_pop_in_zone	Logarithm of total number of population in zone	0.5946	0.0697	8.53
AIC=1669	***Log-likelihood***=-830			
Leisure activities (n=970)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
car_accessibility	Car accessibility in zone	-0.0871	0.0196	-4.44
jobs_lei_den_c	Density of jobs in leisure sector in commune	3.4914	0.3962	8.81
ln_jobs_in_zone	Logarithm of total number of jobs in zone	0.2433	0.0212	11.49
AIC=7721	***Log-likelihood***=-3857			

Table 6.8: Real estate price model

Houses (n=14835)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
constant	-	12.1215	0.0214	565.43
edu_high_per	Percentage of university degree holders in zone	0.0003	0.0001	2.43
green_score	Green area score	0.1508	0.0066	22.67
per_high_inc	Percentage of high income (>3) households in commune	0.0339	0.0004	88.91
jobs_edu_den_c	Density of jobs in education sector in commune	0.0114	0.0017	6.90
jobs_ind_per_z	Percentage of jobs in industry sector in zone	-0.0248	0.0085	-2.90
jobs_ret_den_c	Density of jobs in retail sector in commune	-0.0188	0.0035	-5.43
ln_dist_cbd	Logarithmic distance of zone from CBD	-0.1065	0.0019	-54.60
pop_den_com	Population density in commune	0.0258	0.0023	11.44
R^2=0.65				

Apartments (n=4945)				
Variable	**Interpretation**	**Coefficient**	**SE**	**t-values**
constant	-	11.1558	0.0576	193.80
green_score	Green area score	0.0445	0.0167	2.66
per_high_inc	Percentage of high income (>3) households in commune	0.0312	0.0009	34.01
jobs_edu_den_c	Density of jobs in education sector in commune	0.0369	0.0042	8.73
jobs_ind_den_c	Density of jobs in industry sector in commune	-0.0057	0.0013	-4.41
jobs_ret_den_c	Density of jobs in retail sector in commune	-0.0197	0.0079	-2.50
ln_dist_cbd	Logarithmic distance of zone from CBD	-0.0137	0.0052	-2.64
R^2=0.31				

which follows the same distribution as the location of households.

The evolution of commune average prices for the 2001-2008 period is shown in Figure 6.4. Prices also variate little with respect to the base year prices, having a maximum increase of 4% that takes place in the central communes. Figure 6.5 shows the difference between the average prices by commune predicted by UrbanSim and the observed transaction averages for 2008. While the model predicts prices with reasonable accuracy for communes with low prices in average it clearly under-predicts for the communes where prices are high. This is apparently due to the strong inertia of the hedonic price model (because it is indifferent to market conditions like supply or demand surplus) and the little variation in income distribution, already shown in Figure 6.3, that clearly underestimates the relative growth of high income households in the communes where prices increase in the 2001-2008 period. The fit of the predicted prices against observed ones for 2008 is $R^2 = 0.69$.

6.4 Analysis of used resources and obtained results

The application of the modeling tools is next analized in terms of the trade-off between involved resources and the quality of the obtained results.

Figure 6.2: Predicted increase in the number of households (2001-2008)

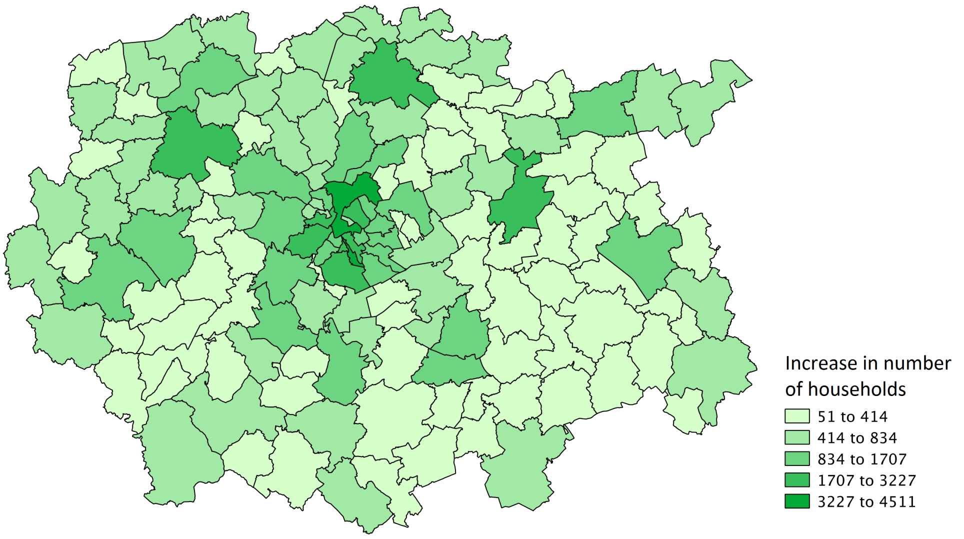


Figure 6.3: Variation in the rich/poor ratio by commune (2001-2008)

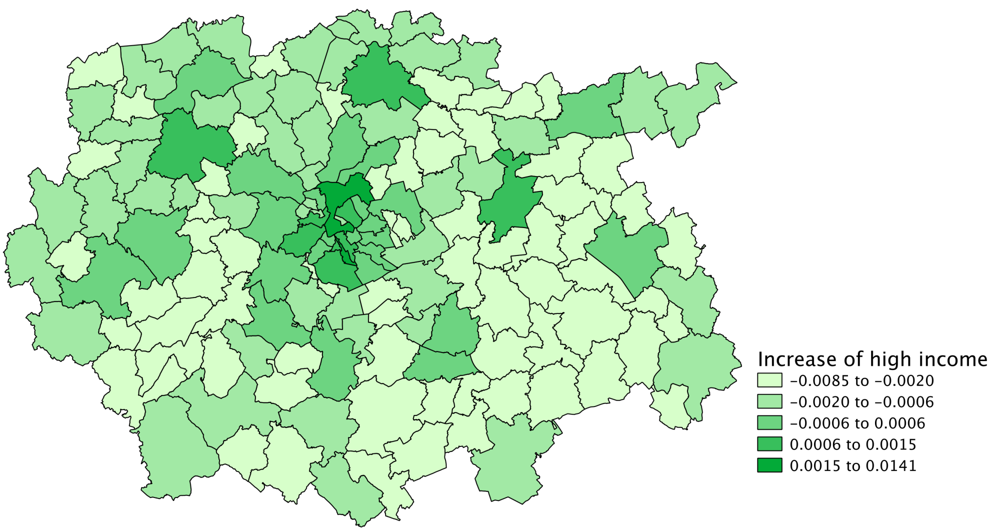


Figure 6.4: Predicted evolution of prices (2001-2008)

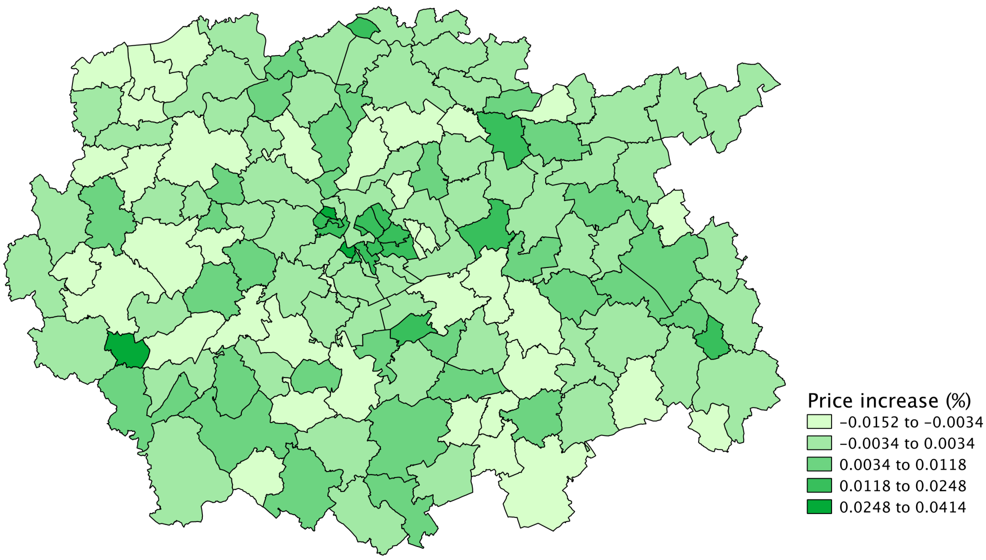
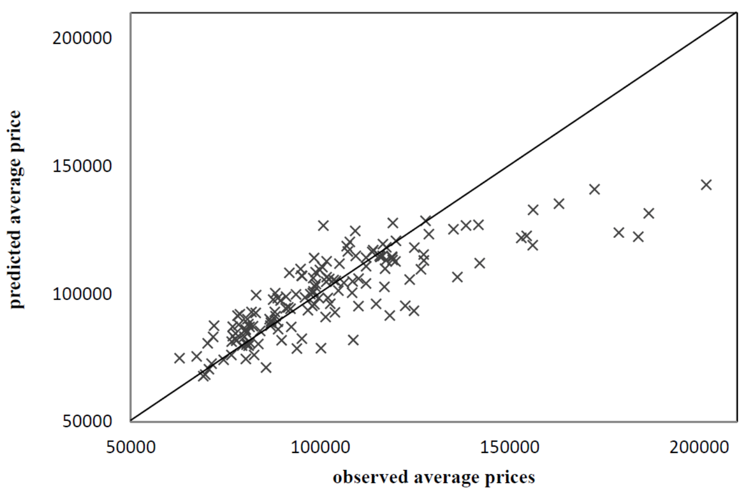


Figure 6.5: Predicted vs observed average prices by commune (2008)



6.4.1 Amount of resources used

The development of the model for Brussels, following the steps described in Section 6.2, began in march 2010 with the initial data collection efforts. A simultaneous process of data cleaning and preparation took place for a period of approximately 1.5 years. The implementation of the data into the platform, model estimation and simulation preparations took approximately one year until results presented here were obtained. The process was undertaken by several different PhD students who worked full time on each of the implementation steps in a more or less sequential fashion. This amounts to a total of approximately 2.5 years or 30 persons-month of work. It's important to notice that, since most of the processes can't be performed in parallel, the amount of time required for the full implementation can't be reduced significantly by increasing the amount of people working simultaneously on it.

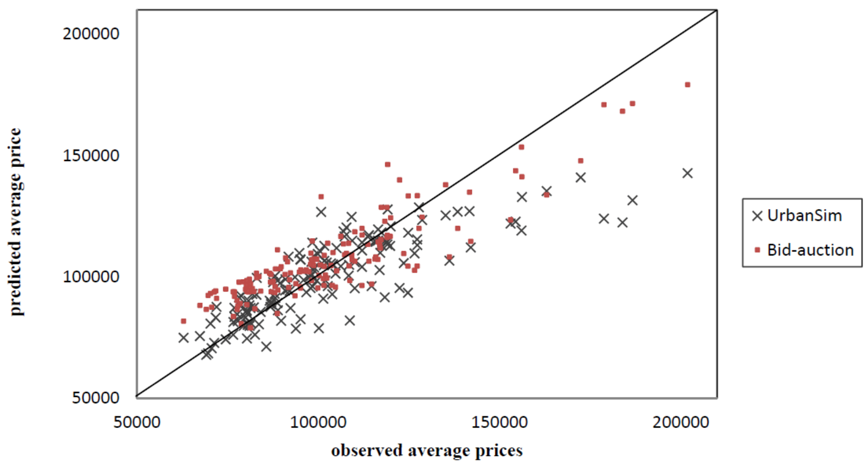
The time and effort required for this application is considerably less that those reported in other applications. For example Nguyen-Luong (2008) reports an effort of four years for a team of four people working full time in a Paris application. Experience is clearly a relevant factor in the amount of time required to implement UrbanSim and implementing a prototype model before the actual application can help to reduce the amount of development time. Patterson and Bierlaire (2010) report a 5 person-month effort to implement a prototype model, with extremely simplified data and models. It is important to mention that, for the application described here, some of the involved persons already had previous experience implementing prototype UrbanSim models and Land Use models in general.

6.4.2 Quality of results

Results obtained with UrbanSim seem reasonable, although inertia in income distribution and prices seems to be strong, with little variation in an eight year simulation period. Because of this, prices are underestimated for communes where a large price increase took place between 2001 and 2008. Figure 6.6 shows the difference between UrbanSim and the model presented on Chapter 5 (Bid-auction model) when forecasting prices for 2008. The Bid-auction model (red dots) has a superior fit to observed average price ($R^2 = 0.79$) and is able to follow the trend of "expensive" communes quite well. It is important to mention that the model presented on chapter 5 uses the same data as the current UrbanSim application and required a much smaller amount of time for implementation (3 person-months against 12 for UrbanSim implementation only). Moreover, the model of Chapter 5 accounts only for residential location dynamics, a disadvantage against the more comprehensive approach of UrbanSim.

The are two possible explanations for the price results of Figure 6.6. First, the use of an hedonic model for prices does not allow to capture the effect of market conditions like supply and demand surplus. The model specification of (6.1) attempts to account for market conditions by including a ratio between simulated and structural vacancy rates. However, the economic justification of the inclusion of this ratio is not clear.

Figure 6.6: Comparison of predicted prices (2008)



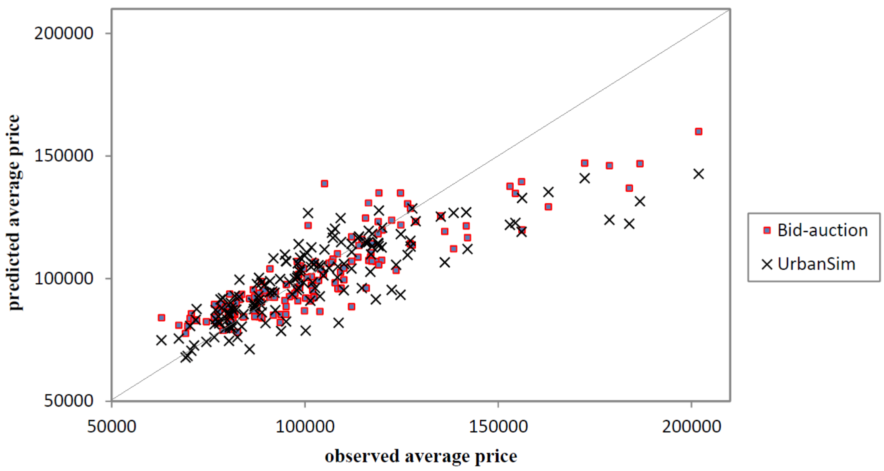
A second possible explanation comes from the fact that UrbanSim solves simultaneous location conflicts by random assignment. This may prevent the model from properly predicting socioeconomic agglomeration that happen when the market can discriminate through price increase. Since prices depend on the income distribution in the surrounding of the location, underestimation of these effects will produce an underestimation of prices.

Figure 6.7 shows the predicted prices for 2008 when the hedonic price model of Table 6.8 is evaluated with the explanatory variables predicted by the model of Chapter 5 (red dots) and compares it with the UrbanSim results. The fit of the predicted prices to observed average prices is is $R^2 = 0.75$, being higher than the fit obtained when using the explanatory variables generated by UrbanSim ($R^2 = 0.69$). This is in part explained by price increase trend in the expensive communes being better captured by the explanatory variables of the bid auction-mode with market clearing. However, the price model itself (hedonic or bid-auction) also seems to have a significant role in the goodness of fit of the predicted prices.

6.4.3 Software and modeling problems

Despite its rather simple structure of independent sub-models, UrbanSim is complex to implement because, basically, it deals with a complex problem. Many details in the way models are estimated and simulations are run can be taken into consideration: sample size, involved agents and datasets, model dependency, definition and sampling of alternatives, data filters, agent group clustering and data storage options, just to name a few, can be defined by the user. For example, the Household Location Choice Model of the application described here has more than 50 setting options, with a similar number for other sub-models. Modifying the settings of a model can easily generate errors that are hard to solve without help from the

Figure 6.7: Hedonic price model evaluated with bid-auction explanatory variables (2008)



developers team and, therefore, customizing the model to the analyst's needs can be hard.

Because of its open source nature, the software is constantly and quickly evolving and, therefore, requires more than basic user knowledge from the analyst if the latest version is to be used or if some aspect of the simulation needs to be adapted to particular conditions of the case study. If modifications beyond basic definitions are required like, for example, a change in the behavioral assumptions or type of model for any of the sub-models, the user will certainly need an "almost-developer" level of knowledge or, if possible, intensive and extensive help from the developer team members. Because of this difficulty, the user is likely to end up adapting a pre-existing model to his own case study, meaning that the new application will have to deal with inherited data structures or model settings.

Also because of the constant evolution of the software, documentation is usually not up to date. This complicates an overall understanding of the model, sometimes forcing the user to use a "black box" approach when working with it.

When comparing results with those obtained with the model presented in Chapter 5 it is possible to see that the more-comprehensive scope of UrbanSim (accounting for every agent and system) does not ensure better results. Accounting for all markets, agents and their interactions is clearly an advantage and should be the objective of any land use modeling system but doing this through the use of simplifying assumptions may harm the quality of the results. Moreover, the complexity (and extension) of the system to model implies a trade-off between the level of detail that can be achieved and cost in terms of data collection and processing (Wegener, 2011). Comprehensive and detailed models allow to introduce more phenomena, more heterogeneity between agents and to represent more interactions at the individual level. However, this also means that more complex models have more interacting components, rendering difficult to ensure the quality and consistency of results.

6.5 Conclusions

This chapter reports an application of the land use modeling platform UrbanSim to the city of Brussels and its surroundings. The implementation efforts, in term of data collection, data processing, software preparation and model estimation are also reported. Implementing UrbanSim for Brussels took approximately 30 persons-month or 2.5 years until a complete simulation was possible.

Most difficulties during implementation of the model are related to data processing and software preparation. UrbanSim requires very detailed data in a specific format that may not be directly available from existing data. Simulation of agent-level data is necessary while data-adaptation, many times using simplifying assumptions, is required to fit the format requirements. Software preparation is a slow and complex task that requires intensive support from the developers team.

Simulation results are reasonably good, although real estate price variation and socioeconomic agglomeration of households is underestimated. This is possibly due to the hedonic approach for price modeling and the strong simplifying assumptions for solving conflicts when clearing the market. Because it is a more comprehensive model, accounting for dynamics and evolution of more agents and markets, UrbanSim generates residential market results (prices and agent distribution) that are not as good as those obtained with the model described in Chapter 5. These results suggest that is preferable to model fewer key processes (or markets) of the land use system than to develop a comprehensive model, accounting for several processes but considering simplifying assumptions for the market clearing mechanism. Since the (re)location of agents is the main driver of urban dynamics, the modeling effort should be focused there.

The assumption that every spatial choice follows a disaggregate utility (or profit) maximizing criteria, given current attributes of locations, may be a wrong assumption. This is especially relevant for models where decision makers can behave strategically, like the real estate development location choice model or the job (firm) location choice model.

The comparative advantage of using a disaggregate approach seems to be counterweighted by the fact that available data is not enough to properly estimate all the required models. A disaggregate approach allows to simulate very detailed behavior by individual agents but it must be supported by, also detailed, empirical data. Therefore, there is a trade-off between comprehensiveness, level of disaggregation and implementation cost that should be carefully analyzed before applying this type of model to real case studies where, in general, detailed data is difficult to collect.

7 Conclusions

This thesis contributes in several methodological and practical aspects of location choice models for land use simulation.

A first conclusion, from **Chapter 2**, is the theoretical appeal and practical convenience of using the bid-auction approach to model location choice models. The possibility of having a framework that explains simultaneously the formation of prices and the location preferences of agents permits to develop more robust models. Many of the model proposals of this thesis are based on the use of this approach.

Chapter 3 reviews and analyzes methods for choice set formation. Explicit choice set formation is not easy to implement in problems with a large number of alternatives, like location choice, and different approaches are therefore required for this. However, the existing heuristics for implicit choice set formation (based on alternative elimination) are unable to reproduce the results of the explicit method and therefore, they should be carefully analyzed before considering them for choice set formation. Following this conclusion, the location choice models presented in this thesis do not attempt to model choice set formation.

Chapter 4 proposes a method to simultaneously estimate a location choice and a price model, under the bid-auction paradigm and using a latent variable framework. The proposed method is less data-hungry than already existing methodologies for the same purpose because it relates aggregate or coarse indicators of prices to the expectation of the maximum bid for the auction of each real estate good, instead of requiring detailed transaction prices for each observation of a located household. The method, named "Latent Auction", is applied to the city of Brussels and compared to other models. Results show a significant improvement in estimation results when using the Latent Auction approach, reproducing both observed average prices and spatial distribution of agents with reasonably small errors.

Chapter 5 proposes a market clearing method for agent-based location choice microsimulation. The method considers that agents adjust their expectations by observing prices from previous periods and modifying their willingness to pay based on them. Baseline bid functions are estimated with the method proposed in Chapter 4. Market is cleared by selecting the best bidder in auctions that are performed for every available location, but computing transaction prices as the expected maximum bid. This allows to incorporate the potential bid of all agents in the simulation, therefore making prices dependent on the market and not just the active bidders. If an agent is the best bidder in more than one location, it selects the one that provides maximum consumer surplus, given the market price. The method is embedded in a larger and more comprehensive framework for land use simulation and is applied to the city of Brussels. Simulations are run for the 2001-2008 period and results are compared with observed data. Results show that the proposed framework is capable of predicting trends in price increase and changes in income distribution. The proposed method has the advantage of being feasible to implement in an agent-level modeling context without the need to solve an equilibrium, but still considering market effects.

Chapter 6 describes the application of a fully-fledged UrbanSim model to the city of Brussels. Estimation for sub-models describing the behavior of various agents and markets is performed, including Household Location Choice, Job Location Choice, Residential and Non-residential Real Estate Development Location Choice and a Real Estate Price Model. Simulations are also run for the 2001-2008 period. Results are not able to reproduce observed average prices for year 2008, probably due to a strong inertia in the simulations that keep income distributions practically constant across time. Despite accounting for more agents and market, results obtained with UrbanSim are not as good as those obtained with the model proposed in Chapter 5 and represented a much higher cost in terms of implementation effort. The Brussels model took 1.5 years for data collection and processing, mostly due to the need to adapt the existing data to the platform requirements. Implementation, estimation and simulation of the UrbanSim model required an additional year, mostly due to software debugging. The model from Chapter 5 took three months for estimation, implementation and simulation.

7.1 Future research

Beyond further application and validation of the methodologies proposed in Chapters 4 and 5 to different case studies, future work can be organized in four possible lines of research:

7.1.1 Behavioral aspects of location choice

Location choice models should be improved in order to account for behavioral aspects that nowadays are neglected or barely taken into account in land use models. Life cycles and lifestyles should have an influence on the moving and location decisions of the households. For example, the high relevance of this type of decision makes reasonable to assume some households will plan ahead of their actions and make more informed and rational decisions, while other households may have a more "spontaneous" behavior. Some households may have a lifestyle that makes them naturally attracted to urban areas while others, with similar socioeconomic characteristics, may have a suburban lifestyle. This type of phenomena can be modeled through the use of latent variables and latent class models (see for example Walker and Li (2007)), accounting for characteristics of the households that go beyond their pure socioeconomic attributes, like their attitudes or perceptions. However, the implementation of this type of model to the particular problem of location choice requires a particular treatment, due to the long term nature of the modeled decisions and the fact that households are not single minded entities, but a bundle of several decision makers that negotiate and compete between them (de Palma et al., 2011). For example, residential location choice can be considered as a decision made jointly with the location choice of activities performed by different members of the household, like working or studying (or conditional on them).

Latent class models are especially interesting for policy making, since they allow to identify target segments of the population and therefore make possible the design of more effective policies. Additionally, latent class models have a structure that naturally fits in the life cycle framework for household behavior modeling and, therefore, can explain how the preferences of households evolve in time.

Another aspect to research is how the expectations of the households are built over their previous experience. For example a household may reduce their choice set of locations to those similar or spatially close to their current one or may build the expected utility when looking for a new location as a function of the utility perceived in the current or previous locations. Reviewing approaches that go beyond utility maximization may be useful to account for this type of phenomena. For example, prospective theory (Kahneman and Tversky, 1979) can be used to account for the way perceived risk and expectations affect the decisions of a household in terms of moving from its current location, the set of alternatives it will consider and the final location that will be chosen. Introducing modeling approaches and ideas from other fields and reconciling them with the discrete choice framework for location choice is an open and challenging research topic.

Development of better location choice models will require specific data collection efforts, developing and executing survey specially designed to measure and model the previously mentioned aspects. Collection of historical data accounting for previous locations, re-location triggers and location of main activities (like work or study) and qualitative data (like future prospects or life expectations) should be considered for future research in this field. Finally, new potential data sources, like social networks data, can be used to improve the behavioral dimension of the models (Carrasco et al., 2008), by measuring the role of peer pressure or social activity patterns in the choice of residential location. Moreover, this data could be use to develop better (short term) activity-based travel demand models that can be explicitly linked to the (long term) location choice model.

7.1.2 Disaggregation and comprehensiveness trade-off

One of conclusions of Chapter 6 is that comprehensive and disaggregate models are not necessarily better than models where the detail is focused on few (but relevant) processes.

This opens the question on the level of aggregation at which each element of an urban system should be modeled. While household location choice seems to benefit naturally from a very disaggregate, agent-based approach, it is possible that other processes involved in the dynamics of the city can be better modeled with higher levels of aggregation. For example, evidence shows that in many cases real estate developers behave as an oligopoly, coordinating their actions and profiting from economies of scale in the construction process. In this case it seems reasonable to explore more aggregated modeling approaches for the generation of new supply. Similarly, firms are likely to behave strategically and are strongly influenced by agglomeration economies. The way firms distribute their jobs in space may be better explained by aggregate microeconomic approaches like, for example, the New Economic Geography (Fujita et al., 1999).

There is a trade-off between the disaggregation level and the comprehensiveness of a model. The bigger the number of phenomena or processes taken into account by the model, the higher the cost of using a disaggregate approach, because of data collection and processing efforts and because of more difficult tractability in terms of interaction between the different components of the model. Additionally, the use of several different models accounting for the decisions of different agents may create an error-propagation problem that has not been carefully analyzed. Identifying the most appropriate aggregation level and modeling approach for each component of a fully fledged land use model will be a topic for future research.

7.1.3 Market clearing for agent-based models with disaggregate time

This thesis proposed a method for real estate market clearing considering that agents adjust their expectations and interact simultaneously in a relatively long period of time. This assumption, although convenient for computational and tractability purposes, is unrealistic and imposes an arbitrary time frame to represent the dynamics of the real estate market. Different aspects of a city's dynamics may happen at different speeds and a more flexible time framework may be required. For example, different agents may have different time windows for location searching or construction times may be different for different types of buildings.

An alternative approach to model the real estate market could be to assume a more disaggregate time framework (for example using days as the basic period) where agents become active as they start looking for a new location and enter the market one by one to sequentially bid for several locations, leaving the market when an auction is won . The adjustment of utility levels by agents looking for a location can be a function of the time spent on the task, adjusting their expectations after each unsuccessful bid. In a similar way, asking prices may also be affected by the time a real estate good has (unsuccessfully) been offered in the market, with the asking price being adjusted after receiving a bid that didn't meet the minimum price threshold. This would allow to model the market clearing mechanism at an individual level while accounting explicitly for the process that pushes prices up or down.

Estimation of this kind of model is a complex task since it requires to estimate both bid functions and asking prices, which need to be treated as latent variables since observed transaction prices register only the values at which bids and asking prices are matched, after the clearing mechanism. An interesting possibility to explore is to expand the methodology proposed in Chapter 4 by introducing both bids and asking prices as latent variables. Calibration of a model like this is also an interesting problem since the bidding frequency will have an effect in the market prices. Finding the right bid frequency for each agent and the right protocol for bid and asking price adjustment is a fundamental research task. Finally, from an implementation point of view, such an approach will require a strong computational framework to deal with all the individual interactions.

7.1.4 Integrated transport and land use

A fundamental aspect that requires further investigation is the connection between the land use and transport sides of the model. Like in most operational models, the transport and land use components are linked through accessibility measures and travel demand vectors that have an aggregate nature. Real integration between land use and transport could be achieved if the travel behavior of individual agents is directly linked to the activity opportunities described by the land use system. An activity-based travel model can be directly linked at an individual level with the agents (households and persons) of the residential location choice model, describing the destination choice and tour-structure as a function of the spatial distribution other agents. This type of interaction is already being explored in projects like SustainCity [1] and ILUTE (Salvini and Miller, 2005). A model like this is certainly difficult to implement and expensive in computational terms since it will require to fully track agents in all the location and travel decisions they make. Despite this, the potential benefit from implementing this type of model outweighs the costs, because it can explicitly account for the effect of the urban environment in the travel patterns of individuals, while simultaneously generating a richer and disaggregate description of travel demand. Moreover, the possibility of incorporating social network data in this kind of models is extremely interesting. For example, as it may be the case in location choice, the travel pattern of an individual can be affected by the location of the members of his social network.

The final purpose of any land use model is to forecast the future evolution of cities and to evaluate policy scenarios. In order to achieve this, a comprehensive model, incorporating the contributions of this thesis and the potential improvements mentioned in this chapter, should be implemented for a real city. A possible case study for this is the city of Santiago, Chile, for which abundant data is available and where several previously implemented land use and transport models can be used as benchmark.

[1] www.sustaincity.org

A Appendix: Possible values of the bid adjustment

The bid adjustment is defined by:

$$b_h^t = -\ln\left(\sum_{i \in S^t} \exp\left(b_{hi}^t - r_i^{t-1}\right)\right) \tag{A.1}$$

meaning that b_h^t will decrease with the size of the available supply (S_i^t) and it will be positive if $\sum_{i \in S^t} \exp\left(b_{hi}^t - r_i^{t-1}\right) \leq 1$ and negative otherwise.

The effect of the demand size in the adjutstment term b_h is introduced through the price of the goods. The rent or price of location i in $t-1$ is defined by the following equation:

$$r_i^{t-1} = \ln\left(\sum_{g \in H^{t-1}} \exp(b_g^{t-1} + b_{gi}^{t-1})\right) \tag{A.2}$$

Therefore, the argument of the sum in expression A.1 can be re-written as:

$$b_{hi}^t - r_i^{t-1} = b_{hi}^t - \ln\left(\sum_{g \in H^{t-1}} \exp(b_g^{t-1} + b_{gi}^{t-1})\right) \tag{A.3}$$

re-arranging the terms:

$$b_{hi}^{t} - r_{i}^{t-1} = b_{hi}^{t} - \ln\left(\exp(b_{h}^{t-1} + b_{hi}^{t-1}) + \sum_{\substack{g \in H^{t-1} \\ g \neq h}} \exp(b_{g}^{t-1} + b_{gi}^{t-1})\right) \tag{A.4}$$

The well know "addition identity" for logarithms indicates that $\ln(a+b) = \ln(a) + \ln\left(1 + \frac{b}{a}\right)$, therefore:

$$b_{hi}^{t} - r_{i}^{t-1} = b_{hi} - \ln\left(\exp(b_{h}^{t-1} + b_{hi}^{t-1})\right) - \ln\left(1 + \frac{\sum_{\substack{g \in H^{t-1} \\ g \neq h}} \exp(b_{g}^{t-1} + b_{gi}^{t-1})}{\exp(b_{hi}^{t-1})}\right) \tag{A.5}$$

Therefore

$$b_{hi}^{t} - r_{i}^{t-1} = b_{hi} - b_{hi}^{t-1} - b_{h}^{t-1} - \ln\left(1 + \frac{\sum_{\substack{g \in H^{t-1} \\ g \neq h}} \exp(b_{g}^{t-1} + b_{gi}^{t-1})}{\exp(b_{hi}^{t-1})}\right) \tag{A.6}$$

The argument of the natural logarithm in equation (A.6) will be always larger than 1 and, everthing else kept constant, will increase with the number of bidding agents in the previous period (H^{t-1}). This means that b_h increases with the number of active agents participting in the market, but with a lag of one period. Further analysis of the values of the bid adjustment can be found in Hurtubia et al. (2010)

B Appendix: Definition of the area of study

Table B.1: Communes in the Sustaincity study area for Brussels

INS code	name	BCR?	INS code	name	BCR?
11005	BOOM	-	21016	UKKEL	-
11037	RUMST	-	21017	WATERMAAL-BOSVOORDE	-
12005	BONHEIDEN	-	21018	SINT-LAMBRECHTS-WOLUWE	-
12009	DUFFEL	-	21019	SINT-PIETERS-WOLUWE	-
12025	MECHELEN	-	23002	ASSE	-
12029	PUTTE	-	23003	BEERSEL	-
12030	PUURS	-	23009	BEVER	-
12034	SINT-AMANDS	-	23016	DILBEEK	-
12035	SINT-KATELIJNE-WAVER	-	23023	GALMAARDEN	-
12040	WILLEBROEK	-	23024	GOOIK	-
21001	ANDERLECHT	-	23025	GRIMBERGEN	-
21002	OUDERGEM	-	23027	HALLE	-
21003	SINT-AGATHA-BERCHEM	-	23032	HERNE	-
21004	BRUSSEL	yes	23033	HOEILAART	-
21005	ETTERBEEK	yes	23038	KAMPENHOUT	-
21006	EVERE	-	23039	KAPELLE-OP-DEN-BOS	-
21007	VORST	-	23044	LIEDEKERKE	-
21008	GANSHOREN	-	23045	LONDERZEEL	-
21009	ELSENE	yes	23047	MACHELEN	-
21010	JETTE	-	23050	MEISE	-
21011	KOEKELBERG	-	23052	MERCHTEM	-
21012	SINT-JANS-MOLENBEEK	-	23060	OPWIJK	-
21013	SINT-GILLIS	yes	23062	OVERIJSE	-
21014	SINT-JOOST-TEN-NODE	yes	23064	PEPINGEN	-
21015	SCHAARBEEK	yes	23077	SINT-PIETERS-LEEUW	-

Table B.2: Communes in the Sustaincity study area for Brussels

INS code	name	BCR?	INS code	name	BCR?
23081	STEENOKKERZEEL	-	24045	HULDENBERG	-
23086	TERNAT	-	24048	KEERBERGEN	-
23088	VILVOORDE	-	24055	KORTENBERG	-
23094	ZAVENTEM	-	24059	LANDEN	-
23096	ZEMST	-	24062	LEUVEN	-
23097	ROOSDAAL	-	24066	LUBBEEK	-
23098	DROGENBOS	-	24086	OUD-HEVERLEE	-
23099	KRAAINEM	-	24094	ROTSELAAR	-
23100	LINKEBEEK	-	24104	TERVUREN	-
23101	SINT-GENESIUS-RODE	-	24107	TIENEN	-
23102	WEMMEL	-	24109	TREMELO	-
23103	WEZEMBEEK-OPPEM	-	24133	LINTER	-
23104	LENNIK	-	24134	SCHERPENHEUVEL-ZICHEM	-
23105	AFFLIGEM	-	24135	TIELT-WINGE	-
24001	AARSCHOT	-	24137	GLABBEEK	-
24007	BEGIJNENDIJK	-	25005	BEAUVECHAIN	-
24009	BERTEM	-	25014	BRAINE-L'ALLEUD	-
24011	BIERBEEK	-	25015	BRAINE-LE-CHATEAU	-
24014	BOORTMEERBEEK	-	25018	CHAUMONT-GISTOUX	-
24016	BOUTERSEM	-	25023	COURT-SAINT-ETIENNE	-
24020	DIEST	-	25031	GENAPPE	-
24033	HAACHT	-	25037	GREZ-DOICEAU	-
24038	HERENT	-	25043	INCOURT	-
24041	HOEGAARDEN	-	25044	ITTRE	-
24043	HOLSBEEK	-	25048	JODOIGNE	-

Table B.3: Communes in the Sustaincity study area for Brussels

INS code	name	BCR?	INS code	name	BCR?
25050	LA HULPE	-	41082	ERPE-MERE	-
25068	MONT-SAINT-GUIBERT	-	42003	BERLARE	-
25072	NIVELLES	-	42004	BUGGENHOUT	-
25084	PERWEZ	-	42006	DENDERMONDE	-
25091	RIXENSART	-	42008	HAMME	-
25105	TUBIZE	-	42010	LAARNE	-
25107	VILLERS-LA-VILLE	-	42011	LEBBEKE	-
25110	WATERLOO	-	42025	WETTEREN	-
25112	WAVRE	-	42026	WICHELEN	-
25117	CHASTRE	-	42028	ZELE	-
25118	HELECINE	-	45059	BRAKEL	-
25119	LASNE	-	45063	LIERDE	-
25120	ORP-JAUCHE	-	52055	PONT-A-CELLES	-
25121	OTTIGNIES-LOUVAIN-LA-NEUVE	-	52063	SENEFFE	-
25122	RAMILLIES	-	52075	LES BONS VILLERS	-
25123	REBECQ	-	55004	BRAINE-LE-COMTE	-
25124	WALHAIN	-	55010	ENGHIEN	-
41002	AALST	-	55023	LESSINES	-
41011	DENDERLEEUW	-	55039	SILLY	-
41018	GERAARDSBERGEN	-	55040	SOIGNIES	-
41024	HAALTERT	-	55050	ECAUSSINNES	-
41027	HERZELE	-	64034	HANNUT	-
41034	LEDE	-	64047	LINCENT	-
41048	NINOVE	-	64075	WASSEIGES	-
41063	SINT-LIEVENS-HOUTEM	-	92142	GEMBLOUX	
41081	ZOTTEGEM	-			

C Appendix: Notation summary

n : index for (individual) decision maker

h : index for household

i : index for alternative (a zone, region in the location choice problem)

c : index for a commune

v : index for type of building or dwelling

t : index for time period

V_{in}^t : utility perceived by decision maker n for selecting alternative i in period t (in the case of residential location choice $n = h$)

B_{hi}^t : bid or willingness to pay of household h for location alternative i (if a type of dwelling is identifiable $i = vi$)

b_h^t : household-specific component of the bid function (related to the expected utility level)

b_{hi}^t : hedonic component of the bid function (willingness to pay for attributes of the dwelling/location)

$P^t(h|i)$: probability of a household h being the best bidder for location i in period t

$P^t(i|h)$: probability of a household h choosing alternative i in period t

H^t : number of households in period t

S^t : supply size (number of residential units, dwellings or buildings) in period t

S_{vi}^t : supply size of type v units in zone i in period t

r_{vi}^t : expected maximum bid for a type v unit in zone i in period t

R^t_{vi} : observed price indicator (e.g average price) for a type v unit in zone i

p^t_{vi} : price for unit type v in zone i

$q^t(h|i)$: perceived probability of being the best bidder for alternative i by household h

Bibliography

Abraham, J. and Hunt, J. (2005). Dynamic submodel integration using an offer-accept discrete event simulation, *Networks and Spatial Economics* **5**(2): 129–146.

Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*, Harvard University Press, Cambridge, Massachusetts.

Anas, A. (1981). The estimation of multinomial logit models of joint location and travel mode choice from aggregated data, *Journal of Regional Science* **21**(2): 223–242.

Anas, A. (1982). *Residential location markets and urban transportation: Economic theory, econometrics, and policy analysis with discrete choice models*, Academic Press, London.

Anas, A. and Arnott, R. (1991). Dynamic housing market equilibrium with taste heterogeneity, idiosyncratic perfect foresight, and stock conversions, *Journal of Housing Economics* **1**(1): 2–32.

Anas, A. and Liu, Y. (2007). A regional economy, land use, and transportation model (relu-tran): Formulation, algorithm design, and testing, *Journal of Regional Science* **47**(3): 415–455.

Bayer, P., Ferreira, F. and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods, *Journal of Political Economy* **115**(4): pp. 588–638.

Bayer, P. and Timmins, C. (2005). On the equilibrium properties of locational sorting models, *Journal of Urban Economics* **57**(3): 462 – 477.

Beckman, R. J., Baggerly, K. A. and McKay, M. D. (1996). Creating synthetic baseline populations, *Transportation Research Part A: Policy and Practice* **30**(6): 415 – 429.

Ben-Akiva, M. E. and Boccara, B. (1995). Discrete choice models with latent choice sets, *International Journal of Research in Marketing* **12**: 9–24.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Ma.

Bhat, C. R. and Guo, J. (2004). A mixed spatially correlated logit model: formulation and application to residential choice modeling, *Transportation Research Part B: Methodological* **38**(2): 147 – 168.

Bierlaire, M. (2003). Biogeme: a free package for the estimation of discrete choice models, *Proceedings of the Swiss Transport Research Conference*, Ascona, Switzerland.

Bierlaire, M., Axhausen, K. and Abay, G. (2001). Acceptance of modal innovation: the case of the Swissmetro, *Proceedings of the 1st Swiss Transportation Research Conference*, Ascona, Switzerland. www.strc.ch.

Bierlaire, M. and Fetiarison, M. (2009). Estimation of discrete choice models: extending biogeme, *Proceedings of the 9th Swiss Transport Research Conference*, Ascona, Switzerland.

Bierlaire, M., Hurtubia, R. and Flötteröd, G. (2009). A comparative analysis of implicit and explicit methods to model choice set generation, *Proceedings of the 9th Swiss Transport Research Conference*, Ascona, Switzerland.

Bierlaire, M., Hurtubia, R. and Flötteröd, G. (2010). Analysis of the implicit choice set generation using a constrained multinomial logit model, *Transportation Research Record* **2175**: 92–97.

Buergle, M., Loechl, M., Waldner, U. and Axhausen, K. W. (2005). Land use and transport simulation: applying urbansim in the greater zurich area, *Proceedings of the 9th Computers in Urban Planning and Urban Management Conference (CUPUM)*, London, UK.

Carrasco, J. A., Hogan, B., Wellman, B. and Miller, E. J. (2008). Collecting social network data to study social activity-travel behavior: an egocentric approach, *Environment and Planning B: Planning and Design* **35**(6): pp. 961–980.

Cascetta, E., Pagliara, F. and Axhausen, K. (2007). The use of dominance variables in choice set generation, *Proceedings of the 11th World Conference on Transport Research*, University of California at Berkeley.

Cascetta, E. and Papola, A. (2001). Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand, *Transportation Research Part C: Emerging Technologies* **9**: 249–263.

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

Cropper, M. L., Deck, L. B. and McConnell, K. E. (1988). On the choice of funtional form for hedonic price functions, *The Review of Economics and Statistics* **70**(4): pp. 668–675.

De La Barra, T. (1980). Integrating micro-economic models with spatial interaction theory., *Transactions, Martin Centre for Architectural and Urban Studies, University of Cambridge; 4* pp. 231–274.

De La Barra, T. (1989). *Integrated land use and transport modelling: decision chains and hierarchies*, Cambridge University Press.

De La Barra, T., Perez, B. and Vera, N. (1984). Tranus-j: putting large models into small computers., *Environment and Planning B* **11**(1): 87–101.

De Palma, A., Motamedi, K., Picard, N. and Nguyen-Luong, D. (2005). An integrated land use - transportation model for the Paris area, *ERSA conference papers ersa05p421*, European Regional Science Association.

de Palma, A., Picard, N. and Ziegelmeyer, A. (2011). Individual and couple decision behavior under risk: evidence on the dynamics of power balance, *Theory and Decision* **70**: 45–64.

Dieckmann, A., Dippold, K. and H., D. (2009). Compensatory versus noncompensatory models for predicting consumer preferences, *Judgment and Decision Making* **4**: 200–213.

DiPasquale, D. and Wheaton, W. (1996). *Urban Economics and Real-Estate Markets*, Prentice Hall, Englewood Cliffs, New Jersey.

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

Dubin, J. A. and McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption, *Econometrica* **52**(2): pp. 345–362.

Echenique, M. H., Flowerdew, A. D. J., Hunt, J. D., Mayo, T. R., Skidmore, I. J. and Simmonds, D. C. (1990). The meplan models of bilbao, leeds and dortmund, *Transport Reviews* **10**(4): 309–322.

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

Elrod, T., Johnson, R. D. and White, J. (2004). A new integrated model of noncompensatory and compensatory decision strategies, *Organizational Behavior and Human Decision Processes* **95**(1): 1 – 19.

Epple, D. (1987). Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products, *Journal of Political Economy* **95**(1): pp. 59–80.

Farooq, B., Bierlaire, M. and Flötteröd, G. (2011). Simulation based population synthesis, *Presentation at the Seventh workshop on discrete choice models, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne.*

Farooq, B., Bierlaire, M. and Flötteröd, G. (2012). A new approach to synthesize heterogeneous agents and their associations for urban microsimulations, *Presentation at the 2nd Workshop on Urban Dynamics*, Termas de Chillán, Chile.

Farooq, B. and Miller, E. J. (2012). Integrated land use and transportation: A dynamic disequilibrium based microsimulation framework for built space markets, *Transportation Research Part A: Policy and Practice* **46**(7): 1030 – 1053.

Frejinger, E., Bierlaire, M. and Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling, *Transportation Research Part B: Methodological* **43**(10): 984–994.

Fujita, M. (1989). *Urban Economic Theory: Land Use and City Size*, Cambridge University Press, Cambridge, Massachussets.

Fujita, M., Krugman, P. and Venables, A. (1999). *The Spatial Economy: Cities, Regions and International Trade*, MIT Press, Cambridge, Massachussets.

Gallay, O. (2010). Starting with urbansim: On the creation of an introductory project, *Technical report*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Gaudry, M. and Dagenais, M. (1979). The dogit model, *Transportation Research Part B: Methodological* **13**: 105–111.

Gilbride, T. J. and Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules, *Marketing Science* **23**(3): 391–406.
URL: *http://www.jstor.org/stable/30036705*

Gin, A. and Sonstelie, J. (1992). The streetcar and residential location in nineteenth century philadelphia, *Journal of Urban Economics* **32**(1): 92 – 107.

Glaeser, E. L. (2008). *Cities, Agglomeration, and Spatial Equilibrium*, Oxford University Press.

Gross, D. J. (1988). Estimating willingness to pay for housing characteristics: An application of the ellickson bid-rent model, *Journal of Urban Economics* **24**(1): 95 – 112.

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

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

Hassan, M. M. and Jun, C. (2011). A microscale simulation of land prices, household and employment growth using urbansim in yongsan-gu, seoul, *International Journal of Urban Sciences* **15**(3): 201–213.

Hauser, J., Ding, M. and Gaskin, S. (2009). Non-compensatory (and compensatory) models of consideration-set decisions, *Proceedings of the Sawtooth Software Conference*, Delray Beach, FL.

Horowitz, J. L. (1986). Bidding models of housing markets, *Journal of Urban Economics* **20**(2): 168 – 190.

Hubert, J. P. and Toint, P. L. (2002). La mobilité quotidienne des belges, *Mobilité et Transports* **1**.

Hunt, J. and Abraham, J. (2003). Design and application of the PECAS land use modelling system, *Proceedings of the 8th Computers in Urban Planning and Urban Management Conference (CUPUM)*, Sendai, Japan.

Hunt, J. D., Kriger, D. S. and Miller, E. J. (2005). Current operational urban land-use and transport modelling frameworks: A review, *Transport Reviews* **25**(3): 329–376.

Hunt, J. D. and Simmonds, D. C. (1993). Theory and application of an integrated land-use and transport modelling framework, *Environment and Planning B: Planning and Design* **20**(2): 221–244.

Hurtubia, R. and Bierlaire, M. (2011). Bid rent model for simultaneous determination of location and rent in land use microsimulations, *Proceedings of the Swiss Transport Research Conference*, Ascona, Switzerland.

Hurtubia, R. and Bierlaire, M. (2012). Estimation of bid functions for location choice and price modeling with a latent variable approach, *Technical Report TRANSP-OR 120206*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Hurtubia, R., Bierlaire, M. and Martinez, F. (2012). Dynamic microsimulation of location choices with a quasi-equilibrium auction approach, *Proceedings of the 12th Swiss Transport Research Conference*, Ascona, Switzerland.

Hurtubia, R., Flötteröd, G., Martínez, F. and Bierlaire, M. (2010). Comparative analysis of hedonic rents and maximum bids in a land-use simulation context, *Swiss Transport Research Conference 2010*.

Hurtubia, R., Gallay, O. and Bierlaire, M. (2010). Attributes of households, locations and real-estate markets for land use modeling (sustaincity deliverable), *Technical Report TRANSP-OR 100419*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Hurtubia, R., Martínez, F. and Bierlaire, M. (2011). Bid auction model for simultaneous determination of location and rent in land use microsimulation models, *Proceedings of the XV Congreso Chileno de Ingeniería de Transporte*.

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

Joo, Y., Hassan, M. M. and Jun, C. (2011). An application of the urbansim land price model in yongsan-gu, seoul, korea, *International Journal of Urban Sciences* **15**(1): 15–24.

Joshi, H., Guhathakurta, S., Konjevod, G., Crittenden, J. and Li, K. (2006). Simulating the effect of light rail on urban growth in phoenix: An application of the urbansim modeling environment, *Journal of Urban Technology* **13**(2): 91–111.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica* **47**(2): pp. 263–292.

Kryvobokov, M., Ovtracht, N. and Thiebaut, V. (2008). Analysis and prediction of household location choice in grand lyon with urban land use simulation tool urbansim, *Proceedings of the International Conference on Territoral Intelligence*, Salerno, Italy.

Lee, L.-F. (1982). Some approaches to the correction of selectivity bias, *The Review of Economic Studies* **49**(3): pp. 355–372.

Lerman, S. R. (1977). Location, housing, automobile ownership, and mode to work: a joint choice model, *Transportation Research Record* **610**: 6–11.

Lerman, S. R. and Kern, C. R. (1983). Hedonic theory, bid rents, and willingness-to-pay: Some extensions of ellickson's results, *Journal of Urban Economics* **13**(3): 358 – 363.

Loechl, M. and Axhausen, K. W. (2010). Modelling Hedonic Residential Rents for Land Use and Transport Simulation While Considering Spatial Effects, *Journal of Transport and Land Use* **3**(2).

Loechl, M., Buergle, M. and Axhausen, K. (2007). Implementierung des integrierten flaechennutzungsmodells urbansim fuer den grossraum zuerich, *DISP - The Planning Review* **168**(1): 13–25.

Malhotra, N. K. (1982). Information load and consumer decision making, *Journal of Consumer Research* **8**(4): 419–430.

Manski, C. (1977). The structure of random utility models, *Theory and Decision* **8**: 229–254.

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

Martínez, F. (1996). Mussa: Land use model for santiago city, *Transportation Research Record: Journal of the Transportation Research Board* **1552**(1): 126–134.

Martinez, F. (2000). Towards a land use and transport interaction framework, *in* D. Hensher and K. Button (eds), *Handbooks in Transport - Hanbook I: Transport Modelling*, Elsevier Science, pp. 145–164.

Martínez, F., Aguila, F. and Hurtubia, R. (2009). The constrained multinomial logit: A semi-compensatory choice model, *Transportation Research Part B* **43**(3): 365 – 377.

Martinez, F., Cascetta, E., Bierlaire, M. and Axhausen, K. W. (2008). An application of the constrained multinomial logit(cmnl) for modelling dominated choice alternatives, *Proceedings of the 8th Swiss Transport Research Conference*, Ascona, Switzerland.

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

Martínez, F. and Hurtubia, R. (2006). Dynamic model for the simulation of equilibrium states in the land use market, *Networks and spatial economics* **6**: 55 – 73.

McFadden, D. (1978). Modeling the choice of residential location, *in* A. Karlqvist (ed.), *Spatial Interaction Theory and Residential Location*, North-Holland, Amsterdam, pp. 75–96.

McMillen, D. P. (1997). Multiple regime bid-rent function estimation, *Journal of Urban Economics* **41**(2): 301 – 319.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information, *Psychological Review* **63**: 81–97.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area, *The American Economic Review* **57**(2): 197–210.

Miyamoto, A. and Kitazume, K. (1989). A land-use model based on random utility/rent-bidding analysis (rurban), *Transport Policy, Management and Technology - Towards 2001, Selected Proceedings of the Fifth World Conference on Transport Research, Yokohama.*

Müller, K. and Axhausen, K. W. (2010). Population synthesis for microsimulation: State of the art, *Proceedings of the 10th Swiss Transport Research Conference*, Ascona, Switzerland.

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

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

Nguyen-Luong, D. (2008). An integrated land use -transport model for the Paris region (SIMAURIF): Ten lessons learned after four years of development, *Technical report*, Institut d'Aménagement et d'Urbanisme de la Région d'Ile-de-France (IAURIF).

Nicolai, T. and Nagel, K. (2010). Coupling matsim and urbansim: Software design issues, *Technical report*, SustainCity Working Paper.
URL: *http://www.sustaincity.org*

Nicolai, T., Wang, L., Nagel, K. and Waddell, P. (2011). Coupling an urban simulation model with a travel model – a first sensitivity test, *Technical report*, SustainCity Working Paper.
URL: *http://www.sustaincity.org*

Pagliara, F. and Wilson, A. (2010). The state-of-the-art in building residential location models, *in* F. Pagliara, J. Preston and D. Simmonds (eds), *Residential Location Choice*, Advances in Spatial Science, Springer Berlin Heidelberg, pp. 1–20.

Patterson, Z. and Bierlaire, M. (2010). Development of prototype urbansim models, *Environment and Planning B* **37**(2): 344–366.

Patterson, Z. and Hurtubia, R. (2008). A prototype integrated transportation land-use model for the lausanne region, *Technical report*, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Patterson, Z., Kryvobokov, M., Marchal, F. and Bierlaire, M. (2010). Disaggregate models with aggregate data: Two urbansim applications, *Journal of Transport and Land Use* **3**(2): 5–37.

Quan, D. C. (1994). Real estate auctions: A survey of theory and practice, *The Journal of Real Estate Finance and Economics* **9**: 23–49.

Quigley, J. M. (1976). *Housing Demand in the Short Run: An Analysis of Polytomous Choice*, National Bureau of Economic Research, pp. 76–102.

Rieser, M., Nagel, K., Beuck, U., Balmer, M. and Rümenapp, J. (2007). Agent-oriented coupling of activity-based demand generation with multiagent traffic simulation, *Transportation Research Record* **2021**: 10–17.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition, *The Journal of Political Economy* **82**(1): 34 – 55.

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

Schirmer, P., Zöllig, C., Müller, K., Bodenmann, B. and Axhausen, K. (2011). The zurich case study of urbansim, *Ersa conference papers*, European Regional Science Association.

Shocker, A., Ben-Akiva, M., Boccara, B. and Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions, *Marketing Letters* **2**: 181–197.

Simmonds, D. C. (1999). The design of the DELTA land-use modelling package, *Environment and Planning B: Planning and Design* **26**(6): 665 – 684.

Swait, J. (2001a). Choice set generation within the generalized extreme value family of discrete choice models, *Transportation Research Part B: Methodological* **35**(7): 643–666.

Swait, J. (2001b). A non-compensatory choice model incorporating attribute cutoffs, *Transportation Research Part B: Methodological* **35**(10): 903–928.

Swait, J. and Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation, *Transportation Research Part B: Methodological* **21**(2).

The UrbanSim Project (2011). Opus: The open platoform for urban simulation and urbansim version 4.3 - users guide and reference manual, *Technical report*, University of California Berkeley and University of Washington, Berkeley, California.

Tiebout, C. M. (1956). A pure theory of local expenditures, *Journal of Political Economy* **64**(5): pp. 416–424.

Train, K. (2009). *Discrete Choice Methods with Simulation*, Cambridge University Press.

Tversky, A. (1972). Elimination by aspects: a theory of choice, *Psychological Review* **79**: 281–299.

United Nations (2011). *Population Distribution, Urbanization, Internal Migration and Development: An International Perspective*, United Nations, Department of Economic and Social Affairs, Population Division.

von Thünen, J. (1826). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationaloekonomie*, Hamburg.

Waddell, P. (1992). A multinomial logit model of race and urban structure, *Urban Geography* **13**: 127–141.

Waddell, P. (2002). UrbanSim: Modeling Urban Development for Land-Use, *Transportation and Environmental Planning: Journal of the American Planning Association* **68**(3): 297–314.

Waddell, P. (2010). Modeling residential location in urbansim, *in* F. Pagliara, J. Preston and D. Simmonds (eds), *Residential Location Choice, Models and Applications*, Advances in Spatial Science, Springer Berlin Heidelberg, pp. 165–180.

Waddell, P., Bhat, C., Eluru, N., Wang, L. and Pendyala, R. (2007). Modeling the interdependence in household residence and workplace choices, *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, Transportation Research Board, Washington, DC.

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

Waddell, P., Ulfarsson, G. F., Franklin, J. P. and Lobb, J. (2007). Incorporating land use in metropolitan transportation planning, *Transportation Research Part A: Policy and Practice* **41**: 382–410.

Waddell, P., Wang, L. and Charlton, B. (2007). Integration of a parcel-level land use model and an activity-based travel model, *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, Transportation Research Board, Washington, DC.

Walker, J. and Ben-Akiva, M. (2002). Generalized random utility model, *Mathematical Social Sciences* **43**(3): 303 – 343.

Walker, J. L. and Li, J. (2007). Latent Lifestyle Preferences and Household Location Decisions, *Journal of Geographical Sytems* **9**: 77–101.

Wegener, M. (2004). Overview of land use transport models, *in* D. Hensher (ed.), *Handbook of transport geography and spatial systems*, Elsevier Science, p. 672.

Wegener, M. (2008). The irpud model: Overview, *Technical report*, Spiekermann & Wegener, Urban and Regional Research, http://irpud.raumplanung.unidortmund.de/irpud/pro/ mod/-mod.htm.

Wegener, M. (2011). From macro to micro, how much micro is too much?, *Transport Reviews* **31**(2): 161–177.

Wheaton, W. C. (1977). A bid rent approach to housing demand, *Journal of Urban Economics* **4**(2): 200 – 217.

Zhao, F. and Chung, S. (2006). A study of alternative land use forecasting models - final report, *Technical Report BD015-10*, Florida Department of Transportation, Tallahassee, Florida.

Zio, S. D., Montanari, A. and Staniscia, B. (2010). Simulation of urban development in the city of rome: Framework, methodology, and problem solving, *Journal of Transport and Land Use* **3**(2).

Ricardo Hurtubia González

Chemin de Champ Soleil 14
1012 - Lausanne, Switzerland
Phone: 41-21-6939329
Fax: 41-21-6938060
ricardo.hurtubia@epfl.ch
rhurtubi@gmail.com

PERSONAL INFORMATION

Birth date:	April 24, 1978
Nationality:	Chilean
Marital Status:	Married (one daughter)
Web:	http://sites.google.com/site/ricardohurtubia/

TEACHING EXPERIENCE

Master thesis supervision (Civil Engineering):

- 2012 - "*Mode choice models for the city of Nice using psychometric indicators and latent variables*". My Hang Nguyen, École Polytechnique Fédérale de Lausanne, Switzerland.
- 2011 - "*Urban development effects of the construction of a metro line*". Aurelién Odobert, École Polytechnique Fédérale de Lausanne, Switzerland and Concordia University, Montreal, Canada.
- 2010 - "*Analyse des relations entre la forme urbaine, l'accessibilité aux transports en commun et la mobilité quotidienne des habitants de la ville de Québec, Canada*". Laetitia Bettex, École Polytechnique Fédérale de Lausanne and McGill University, Montréal, Canada.
- 2009 - "*Combining heterogeneous sensor information for the activity tracking of smart phone users*". Willem Himpe, Katholieke Universiteit Leuven and École Polytechnique Fédérale de Lausanne.

Aug. 2008 – Jun. 2012	Supervisor for semester projects on land use and transportation modeling for students of civil engineering and mathematics (12 projects in total). École Polytechnique Fédérale de Lausanne.
Sep. 2010 – Jan. 2011	Teaching assistant of Optimization (Introduction à l'optimisation différentiable). École Polytechnique Fédérale de Lausanne.
Feb. 2010 and 2011	Teaching assistant of Discrete Choice analysis: Predicting Demand and Market Shares (http://transp-or.epfl.ch/dca). École Polytechnique Fédérale de Lausanne.
Sep. 2008 - Dec. 2009	Teaching assistant of Modeling of Transport and Energy Systems. École Polytechnique Fédérale de Lausanne.
Jul. 2002 – Dec. 2004	Teaching assistant of Transport Systems Analysis. Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
Mar. 1999 – Jun. 1999	Teaching assistant of Introduction to Industrial Engineering. Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

PROFESSIONAL EXPERIENCE

Jun. 2008 to present	**Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne**. Research and teaching assistant.
Jan. 2007 – May. 2008	**CIS Consultores**. Project Engineer.
Nov. 2004 – Dec. 2006	**Laboratorio de Modelación del Transporte y el Uso del Suelo (LABTUS), Universidad de Chile**. Project Engineer.

EDUCATION

2012	**PhD in Mathematics** – École Polytechnique Fédérale de Lausanne. Under the supervision of Prof. Michel Bierlaire.
2006	**Master of Science in Transport Engineering** – Universidad de Chile. Thesis supervised by Prof. Francisco Martinez.
2006	**Industrial Engineer** – Universidad de Chile. Thesis supervised by Prof. Francisco Martinez.
2002	**Bachelor of Sciences in Engineering** – Universidad de Chile

PUBLICATIONS – ISI JOURNALS

Bierlaire, M., Hurtubia, R., and Flötteröd, G. (2010) An analysis of the implicit choice set generation using the Constrained Multinomial Logit model, Transportation Research Record, 2175:92-97.

Martínez F., Aguila F. and Hurtubia R. (2009) The Constrained Multinomial Logit Model: A Semi-Compensatory Choice Model. Transportation Research Part B: Methodological, 43(3):365-377.

Martínez, F. and Hurtubia, R. (2006) Dynamic model for the simulation of equilibrium states in the land use market. Networks and Spatial Economics, 6, pp. 55-73.

PUBLICATIONS – SELECTED PEER-REVIEWED CONFERENCE PROCEEDINGS

Hurtubia, R., Bierlaire, M., and Martinez, F. (2012). Dynamic microsimulation of location choices with a quasi-equilibrium auction approach. Proceedings of the 12th Swiss Transport Research Conference. Ascona, Switzerland

Hurtubia, R., Martinez, F., and Bierlaire, M. (2011). Bid auction model for simultaneous determination of location and rent in land use microsimulation models. Proceedings of the XV Congreso Chileno de Ingeniería de Transporte October 3 - 6, 2011.

Hurtubia, R., Flötteröd, G., and Bierlaire, M. (2010) Estimation techniques for MEV models with sampling of alternatives. Proceedings of the European Transport Conference. Glasgow, Scotland, October 11 - 13, 2010.

Hurtubia, R., Martinez, F., Flötteröd, G., and Bierlaire, M. (2010) Comparative analysis of hedonic rents and maximum bids in a land-use simulation context. Proceedings of the 10th Swiss Transport Research Conference. Ascona, Switzerland, September 1 - 3, 2010.

Hurtubia, R., Atasoy, B., Glerum, A., Curchod, A., and Bierlaire, M. (2010) Considering latent attitudes in mode choice: The case of Switzerland. Proceedings of the World Conference on Transport Research. Lisbon, Portugal, July 11-15, 2010.

Hurtubia, R., Flötteröd, G., and Bierlaire, M. (2009) Inferring the activities of smartphone users from context measurements using Bayesian inference and random utility models. Proceedings of the European Transport Conference. Leeuwenhorst, The Netherlands, October 5 - 7, 2009.

SPEAKER IN SEMINARS AND CONFERENCES

2012

- 2nd Workshop on Urban Dynamics - URBANICS II. Termas de Chillán, Chile
- 12th Swiss Transport Research Conference, Ascona, Switzerland.

2011

- XV Congreso Chileno de Ingeniería de Transporte, Santiago, Chile.
- 51st European Congress of the Regional Science Association International, Barcelona, Spain.
- Seventh Workshop on Discrete Choice Models, EPFL, Lausanne, Switzerland.
- 11th Swiss Transport Research Conference, Ascona, Switzerland.

2010
- Seminario especial del curso Modelos Avanzados de Demanda, Universidad de Chile.
- European Transport Conference 2010, Glasgow, Scotland.
- 10th Swiss Transport Research Conference, Ascona, Switzerland.
- Sixth Workshop on Discrete Choice Models, EPFL, Lausanne, Switzerland.
- Integrated Modeling and Simulation Workshop (MIT Portugal), Lisbon, Portugal
- World Conference on Transport Research, Lisbon, Portugal
- First Workshop on Urban Dynamics (URBANICS), Marbella, Chile.

2009
- European Transport Conference 2009, Leeuwenhorst, The Netherlands.
- 9th Swiss Transport Research Conference, Ascona, Switzerland.
- Fifth Workshop on discrete choice models. EPFL, Lausanne, Switzerland.
- Fourth Kuhmo-Nectar Conference on Transport and Urban Economics, Copenhagen, Denmark.

2008
- 8th Swiss Transport Research Conference, October 2008, Monte Veritá, Ascona, Switzerland.
- European Transport Conference 2008, Leeuwenhorst, The Netherlands.
- Fourth workshop on discrete choice models, EPFL, Lausanne, Switzerland.

2004
- XIII Panamerican Conference of Traffic and Transportation Engineering, Albany, NY, USA.

RESEARCH PROJECTS

Aug. 2010 to present — **Sustainable Urban Patterns (SUPat)**. Project manager. Sponsor: *Swiss National Science Foundation*. The project aims at developing new techniques for the evaluation of urban projects. This considers the combination of expert's knowledge (architects and urban planners) with quantitative land use and transport microsimulation models such as UrbanSim and MATSim. The project will be implemented for the city of Zurich.

Jan. 2010 to present — **Micro-simulation for the prospective of sustainable cities in Europe (SustainCity)**. Project manager. Sponsor: *European Commission*. The project developes a new UrbanSim-based modeling platform for three European cities (Brussels, Paris and Zurich). The objective is to create a tool for urban scale policy-evaluation, with a focus on environmental, social and economical sustainability. In charge of the model for the city of Brussels.

Jan. 2007 – May. 2008 — **Modelo integrado de uso de suelo y transporte: Formulación y desarrollo**. Research assistant. *Fondecyt regular – 2006 – 1060788*. Development of a new version of the Santiago Land Use Model (MUSSA II)

Jul. 2002 – Oct. 2004 — **Mejoramiento de modelación de ciudades y extensión a ciudades de diversas características**. Student (Master Thesis). *Fondecyt regular – 2001 – 1010422*. Master thesis under the supervision of Prof. Francisco Martínez.

AWARDS AND HONOURS

2009 — Ranked third in the “TOP 25 hottest articles” of Transportation Research Part B (Martínez, Aguila and Hurtubia, 2009)

2009 — BecasChile scholarship for PhD studies

2006 — Graduated with highest honours, Master of Science in Transport Engineering, Universidad de Chile
Graduated with highest honours, Professional Degree in Industrial Engineering, Universidad de Chile

---

SKILLS

Computing: Python, Gauss, Matlab, SQL, Biogeme.

Languages: Spanish (native language), English (advanced) and French (intermediate)
