

# Estimation of bid functions for location choice and price modeling with a latent variable approach

Ricardo Hurtubia \*

Michel Bierlaire \*

February 6, 2012

Report TRANSP-OR 120206  
Transport and Mobility Laboratory  
Ecole Polytechnique Fédérale de Lausanne  
transp-or.epfl.ch

## Abstract

A new approach for the estimation of bid-rent functions for location choice is proposed. The method considers that the expected maximum bid in the auction of a good is a latent variable than can be related to observed prices for similar goods. The model generates estimates for the spatial distribution of agents and prices simultaneously that are better than those obtained by estimating a maximum bid and a price model independently. The model is applied and validated for a case study. Results show that the proposed model outperforms other methods for bid rent estimation, especially when detailed data describing the real estate goods and their prices is not available.

*Key words:* location choice, bid function, auction, real estate

---

\*Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, ricardo.hurtubia@epfl.ch

# 1 Introduction

Land use models are an increasingly used tool for 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 policy. They are of particular relevance in the field of transport modeling, since travel demand is explained in a large amount by the spatial distribution of agents and activities in a region.

Modeling the location choice of the different agents that interact in a city is one of the main objectives of any land use model. The distribution and agglomeration of agents (households and firms) is one of the main sources of a wide variety of externalities such as congestion, pollution or social segregation and, simultaneously, is one of the main factors that affect the value of land and real estate goods.

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. Under the choice approach (McFadden, 1978a; Anas, 1982), agents select the location that maximizes their utility, with prices or rents being determined exogenously through a hedonic model (Rosen, 1974). 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.

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 (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, since the under determined nature of the Logit model does not allow to find absolute estimates of the willingness to pay, Ellickson's model 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 price or rent paid for a particular real estate good and its corresponding attributes. Moreover, as in the case of Ellickson, 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.

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 case studies. The literature shows that, in general, the bid-rent generates better results than hedonic price models, thanks to the possibility of estimating 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 ILUTE (Salvini and Miller, 2005). 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, the adjustment of the bid functions to absolute levels is done 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 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 paper 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 prices or average 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 (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 paper is organized as follows: Section 2 describes the bid-auction approach to location choice modeling. Section 3 reviews the literature on estimation of bid-rent function and analyzes the advantages and drawbacks of the different existing methods. Section 4 describes the method proposed in this paper and Section 5 describes a case study where the method is implemented, validated and compared with other methods. Finally, Section 6 concludes the paper and identifies future lines of research.

## 2 The bid approach to location choice

Since Alonso (1964), the real estate market has been understood as an auction market, where agents (households and firms) bid their willingness to pay for a particular good (residential unit, land, etc.) 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 be derived from the classical consumer's problem of maximum utility, given income constraints:

$$\max_{x,i} U(x, z_i) \quad (1)$$

$$\text{s.t. } px + r_i \leq I$$

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 than 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

$$\max_i V(p, I - r_i, z_i) \quad (2)$$

where  $V$  is the indirect utility function, conditional on the the location. Given the maximum utility level ( $\bar{U}$ ) a consumer can achieve, the indirect utility can be inverted in the price variable:

$$r_i = I - V^{-1}(\bar{U}, p, z_i) \quad (3)$$

Under the auction market assumption, the price or rent variable ( $r_i$ ) of (3) can be understood as the willingness to pay for a particular location (Jara-Díaz and Martínez, 1999), therefore the bid function  $B$  can be expressed as:

$$B_{hi} = I_h - V_h^{-1}(\bar{U}, p, z_i) \quad (4)$$

The bid, or bid-rent, function can be understood as the maximum rent (or price) a household can pay for a particular dwelling, while enjoying a fixed utility level  $\bar{U}$  (Fujita, 1989). In (4) the index  $h$  has been included to take into account heterogeneity in preferences within different households. Ellickson (1981) showed that the bid defined by (4) can also be written directly as a function of the location attributes ( $B_{hi}(z_i)$ ) and proposed to account for the unobserved heterogeneity in preferences across households by adding a random term,

$$\tilde{B}_{hi} = B_h(z_i) + \varepsilon_h = B_{hi} + \varepsilon_h \quad (5)$$

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:

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

If the error terms follow an Extreme Value distribution, the best bid probability can be expressed as a logit model (McFadden, 1978b):

$$P_{h/i} = \frac{\exp(\mu B_{hi})}{\sum_g \exp(\mu B_{gi})} \quad (6)$$

Under the auction market assumption, the price or rent ( $r_i$ ) of a good will be the maximum bid and it can be expressed as the following expectation:

$$r_i = E \left( \max_h (B_{hi}) \right) \quad (7)$$

The extreme value distribution assumption allows to express the expected maximum bid for a particular location as the logsum of the bids, in the same way the logsum represents the expected maximum utility in a traditional maximum utility discrete choice problem (Ben-Akiva and Lerman, 1985):

$$r_i = \frac{1}{\mu} \ln \left( \sum_g \exp(\mu B_{gi}) \right) + C \quad (8)$$

where  $C$  is an unknown constant indicating that the absolute value of the bids cannot be measured. This happens because the logit model is under-identified and, while relative bids are enough to calculate the best bidder probability of (6), they do not necessarily relate to real prices or rents .

### 3 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 (6). The original formulation by Ellickson considers a linear in parameters bid function and is estimated via maximization of the following likelihood function:

$$\mathcal{L} = \prod_{i \in S} \left( \prod_{h \in C_i} (P_{h/i})^{y_{hi}} \right)$$

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 (6).

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 ( $\mu$ ) cannot be identified and, as depicted in (8), 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:

$$P_{h/i} = \text{Prob} \{ B_{hi} + \varepsilon_h = R_i \text{ and } B_{hi} + \varepsilon_h > B_{h'i} + \varepsilon_{h'}, \forall h' \neq h \} \quad (9)$$

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 (9) can be written as:

$$P_{h/i} = f(R_i - B_{hi}) \prod_{h' \neq h} F(R_i - B_{h'i}) \quad (10)$$

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

$$f(\varepsilon) = \mu \exp(-\mu\varepsilon) \exp(-\exp(-\mu\varepsilon)) \quad (11)$$

and

$$F(\varepsilon) = \exp(-\exp(-\mu\varepsilon)) \quad (12)$$

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

$$\mathcal{L} = \prod_{i=1}^S \left( -\mu \exp(-\mu(R_i - B_{hi})) \prod_{h'=1}^H \exp(-\exp(-\mu(R_i - B_{h'i}))) \right)^{y_{hi}} \quad (13)$$

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 (13) 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 (6) 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

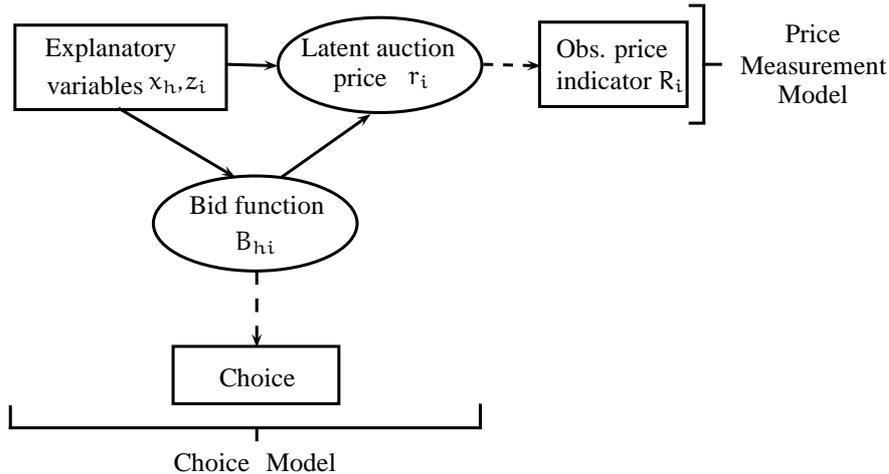


Figure 1: Model structure

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.

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 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 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 probability (6). 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 defined by the logsum expression of (8). 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:

$$R_i = \alpha + \gamma r_i + \eta. \quad (14)$$

Assuming a normal distribution for the error term  $\eta$ , a probability density function  $f(R_i|r_i)$  with mean zero can be defined for the measurement relation of (14) as follows:

$$f(R_i|r_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{R_i - \alpha - \gamma r_i}{2\sigma^2}\right) \quad (15)$$

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:

$$\mathcal{L} = \prod_{i \in S} \left( \prod_{h \in C_i} (P_{h/i} \cdot f(R_i|r_i)) \right)^{y_{hi}} \quad (16)$$

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 (16) will be the set of parameters ( $\beta$ ) for the bid function ( $B_{hi}$ ) and the  $\alpha$ ,  $\gamma$  and  $\sigma$  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 (14) can be used to estimate the expected prices as a function of the latent auction price.

## 5 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). 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 aggregated 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 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 1: Bid function specification

parameter	variables
$\beta_{surf}$	$surface_{vi}$ ( $m^2$ ) $\times$ $\log(size_h)$ (number of people)
$\beta_{sup}$	$high\_educ_i$ (%) $\times$ $high\_educ_h$ (number of people)
$\beta_{house}$	$is\_house_{vi}$ (dummy) $\times$ $size_h$ (number of people)
$\beta_{mid\_inc}$	$avg\_income_i$ (Euros) $\times$ $high\_income_h$ (dummy)
$\beta_{high\_inc}$	$avg\_income_i$ (Euros) $\times$ $mid\_income_h$ (dummy)
$\beta_{trans0}$	$PT\_accessibility_i$ (facilities/ $km^2$ ) $\times$ $0\_cars_h$ (dummy)
$\beta_{trans2}$	$PT\_accessibility_i$ (facilities/ $km^2$ ) $\times$ $2\_cars_h$ (dummy)
$\beta_{comm}$	$commerce_i$ (jobs/ $m^2$ ) $\times$ $\log(size_h)$ (number of people)
$\beta_{office}$	$office_i$ (jobs/ $m^2$ ) $\times$ $workers_h$ (number of people)
$\beta_{green}$	$green_i$ (parks/ $m^2$ ) $\times$ $children_h$ (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 that 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 (14) and using the explicit definition of the maximum expected bid given by (8). 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).

$$\ln(R_{v'i'}) = \alpha + \gamma \cdot \ln \sum_h \exp(B_{hvi}) \quad (17)$$

For the estimation process, the scale parameter  $\mu$  of the bid probability (6) is assumed to be one.

## 5.1 Estimation results

The model was first estimated for a 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, but keeping the same specification for the bid function, defined by (16). The estimation in both cases was done using an extended version of the software package BIOGEME (Bierlaire, 2003; Bierlaire and Fétiarison, 2009); results are shown in Table 2, where the first column shows the results using Ellickson's approach while the second column shows

Table 2: Estimation results for Brussels

Parameter	Ellickson			Latent Auction		
	Value	Std err	t-test	Value	Std err	t-test
$\beta_{\text{surf}}$	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*
$\beta_{\text{sup}}$	0.403	0.108	3.73	0.728	0.0739	9.84
$\beta_{\text{trans0}}$	0.408	0.136	3.00	0.599	0.0849	7.06
$\beta_{\text{trans2}}$	-0.532	0.153	-3.48	-0.31	0.0791	-3.91
$\beta_{\text{house}}$	0.461	0.0615	7.5	0.0563	0.00702	8.03
$\beta_{\text{comm}}$	-1.34	0.278	-4.83	-0.0366	0.031	-1.18*
$\beta_{\text{green}}$	-0.349	0.0717	-4.86	0.136	0.0201	6.74
$\beta_{\text{office}}$	-0.295	0.0931	-3.16	0.0896	0.0413	2.17
$\alpha$	-	-	-	-16.4	3.23	-5.08
$\gamma$	-	-	-	1.92	0.229	8.39
$\sigma$	-	-	-	-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

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 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  $\beta_{\text{green}}$  and  $\beta_{\text{off}}$ , 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.

For comparison purposes, the same specification of Table 1 is estimated using Lerman and Kern’s

Table 3: Estimation results for Brussels

Parameter	Ellickson			L&K		
	Value	Std err	t-test	Value	Std err	t-test
$\beta_{\text{surf}}$	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
$\beta_{\text{sup}}$	0.403	0.108	3.73	0.416	0.0669	6.22
$\beta_{\text{trans0}}$	0.408	0.136	3.00	-1.01	0.0716	-14.1
$\beta_{\text{trans2}}$	-0.532	0.153	-3.48	-0.226	0.0887	-2.54
$\beta_{\text{house}}$	0.461	0.0615	7.5	0.0167	0.0182	0.92*
$\beta_{\text{comm}}$	-1.34	0.278	-4.83	-0.768	0.0977	-7.85
$\beta_{\text{green}}$	-0.349	0.0717	-4.86	0.286	0.0367	7.78
$\beta_{\text{office}}$	-0.295	0.0931	-3.16	-0.767	0.0533	-14.38
$\mu$	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

method, therefore maximizing the likelihood function of (13). Results for this method are shown in the second column of Table 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.

## 5.2 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 valid 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 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 left and right hand 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

Figure 2: Estimation fit (natural log of price)

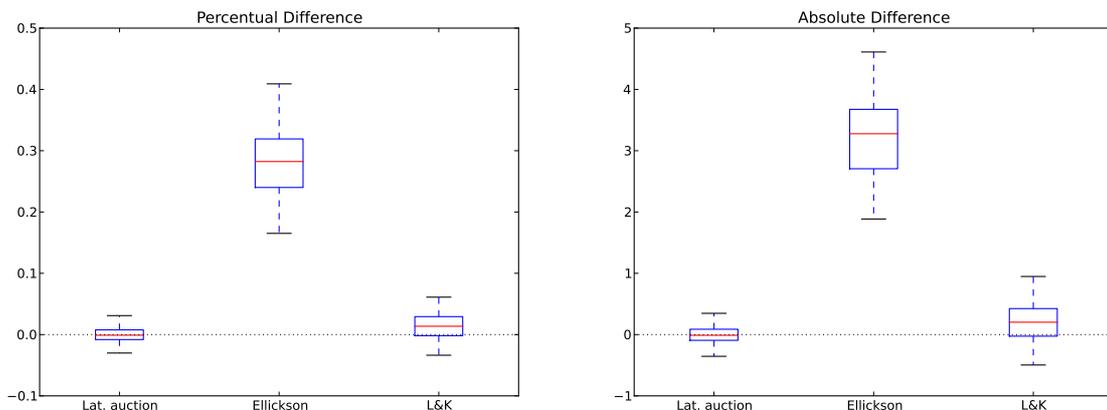
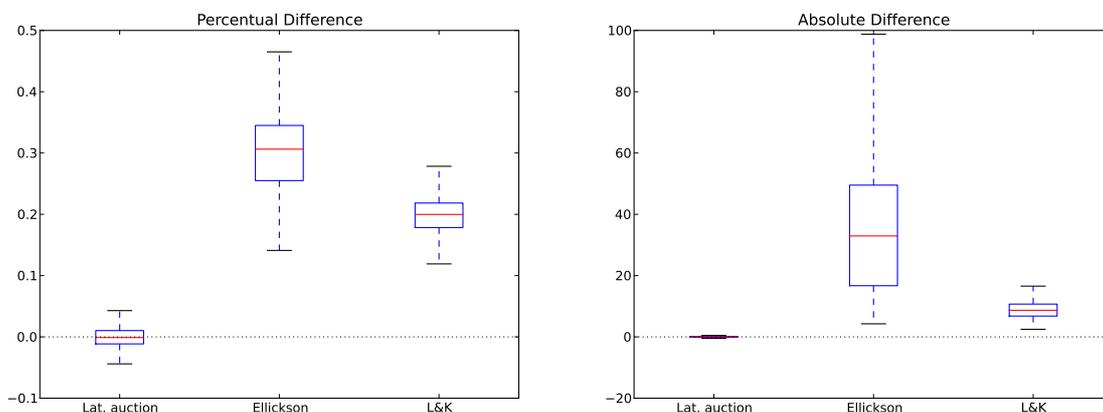


Figure 3: Error in forecast: natural log of price



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.

### 5.3 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 3,4 and 5 as the difference at the commune level of the forecast variables against their observed value.

Figure 4: Error in forecast : number of people by commune

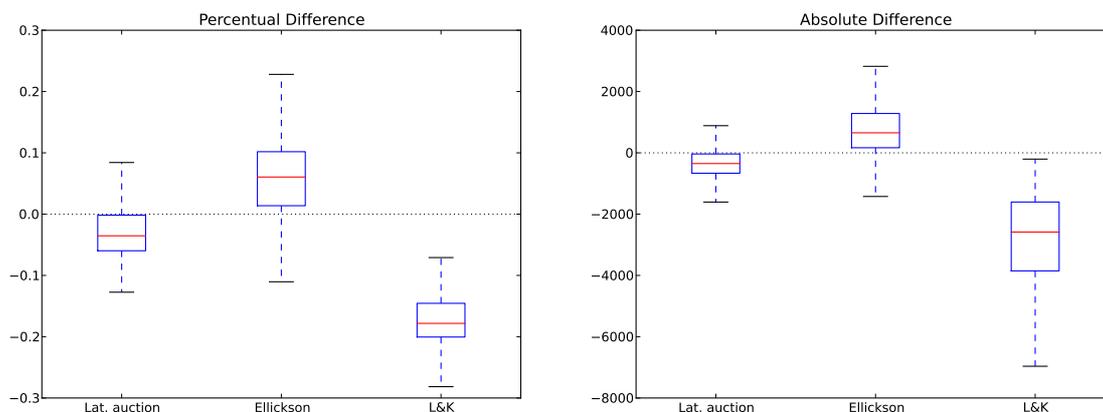
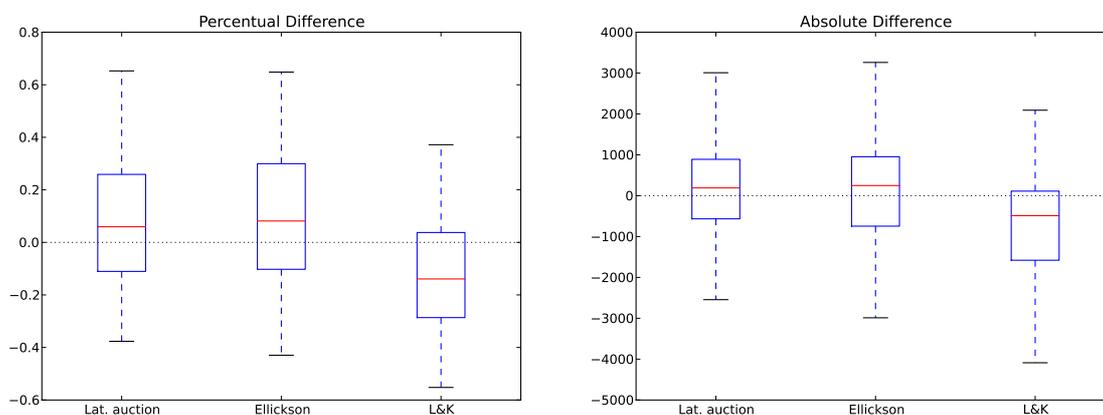


Figure 5: Error in forecast: Number of people with university degree by commune



The difference between price forecast and observed average price is shown in Figure 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 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 5 shows the difference between the forecast of people with university degree by 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

## 6 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. 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 aggregated nature.

The differences observed between forecast and observed prices is expected and explained by the aggregated nature of the price indicator. A more disaggregated indicator should allow for a better estimation and, consequently, a better fit. Further research should 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.

## References

- Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*, Harvard University Press, Cambridge, Massachusetts.
- Anas, A. (1982). *Residential location markets and urban transportation: Economic theory, econometrics, and policy analysis with discrete choice models*, Academic Press, London.
- Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Ma.
- 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. and Feterarison, M. (2009). Estimation of discrete choice models: extending biogeme, *Proceedings of the 9th Swiss Transport Research Conference*, Ascona, Switzerland.
- 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 functional form for hedonic price functions, *The Review of Economics and Statistics* **70**(4): pp. 668–675.
- DiPasquale, D. and Wheaton, W. (1996). *Urban Economics and Real-Estate Markets*, Prentice Hall, Englewood Cliffs, New Jersey.
- Dubin, J. A. and McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption, *Econometrica* **52**(2): pp. 345–362.
- Ellickson, B. (1981). An alternative test of the hedonic theory of housing markets, *Journal of Urban Economics* **9**(1): 56 – 79.

- Fujita, M. (1989). *Urban Economic Theory: Land Use and City Size*, Cambridge University Press, Cambridge, Massachusetts.
- Gin, A. and Sonstelie, J. (1992). The streetcar and residential location in nineteenth century philadelphia, *Journal of Urban Economics* **32**(1): 92 – 107.
- 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.
- 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.
- 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. 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.
- Martínez, F. (1996). Mussa: Land use model for santiago city, *Transportation Research Record: Journal of the Transportation Research Board* **1552**(1): 126–134.
- McFadden, D. (1978a). Modeling the choice of residential location, in A. Karlqvist (ed.), *Spatial Interaction Theory and Residential Location*, North-Holland, Amsterdam, pp. 75–96.
- McFadden, D. (1978b). *Modeling the Choice of Residential Location*, In: *Spatial Interaction Theory and Residential Location*, edited by A. Karlquist et al., pp. 75-96, North-Holland, Amsterdam.
- McMillen, D. P. (1997). Multiple regime bid-rent function estimation, *Journal of Urban Economics* **41**(2): 301 – 319.
- 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*.
- Muto, S. (2006). Estimation of the bid rent function with the usage decision model, *Journal of Urban Economics* **60**(1): 33 – 49.
- 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.
- 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 Systems* **9**: 77–101.
- 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>.