



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA CIVIL

MICROSIMULACIÓN DE TRANSACCIONES EN EL MERCADO INMOBILIARIO

TESIS PARA OPTAR AL GRADO DE DOCTOR EN SISTEMAS DE INGENIERÍA

VÍCTOR ANTONIO ROCCO CASTILLO

PROFESOR GUÍA:
FRANCISCO MARTÍNEZ CONCHA

PROFESOR CO-GUÍA:
RICARDO HURTUBIA GONZÁLEZ

MIEMBROS DE LA COMISIÓN:
MOSHE BEN-AKIVA
ROBERTO COMINETTI COTTI-COMETTI
NICOLAS FIGUEROA GONZÁLEZ

Este trabajo ha sido parcialmente financiado por Conicyt

SANTIAGO DE CHILE
SEPTIEMBRE 2014

Resumen

Los modelos de uso de suelo surgen históricamente como modelos agregados en población, espacio y tipo de bien inmueble (departamento, casa, oficina, etc.). Destacan los modelos de renta y elección discreta, ambos ámpliamente usados en la literatura para estudiar las ciudades, obteniendo precios para grupos de viviendas en condiciones de equilibrio en el marco de la ciudad. En modelos posteriores de microsimulación tales niveles de agregación, así como la agregación a un nivel temporal en equilibrio estático, han sido levantados implementando modelos donde cada agente y localización se asigna de manera detallada, sin embargo el precio de las viviendas se obtiene normalmente mediante funciones hedónicas que proyectan un corte temporal observado sin edongenizar la dinámica propia del mercado. Otros trabajos buscan representar la formación de precio a nivel de interacción comprador-vendedor (y no a nivel zonal agregado), sin embargo en este enfoque la formación del precio de las viviendas sólo se ha estudiado empleando modelos de búsqueda para el vendedor (Salant, 1991, Yavaş, 1992, Merlo et al., 2013), que requieren supuestos muy simplificadorios para el comportamiento de los compradores, y no corresponden al resultado de una interacción. Diversos autores reconocen que estos temas han sido sub estudiados tanto teórica cómo empíricamente (Filatova et al., 2009).

Por otra parte, los modelos de microsimulación del sistema de uso de suelo se han transformado en una herramienta cada vez más usada para predecir los estados futuros de las ciudades, evaluar distintas políticas y sus efectos sobre la ciudad. Se observa además, la inexistencia de un modelo de microsimulación con un enfoque microeconómico consistente con la teoría de elección discreta, que considere los aspectos relevantes de los modelos de búsqueda tanto del comprador como del vendedor, representando la construcción de expectativas y la dinámica del mercado de manera detallada y con la capacidad de reproducir y predecir tanto precios de viviendas como localizaciones de agentes.

La investigación propone un modelo que busca representar la interacción multi-agente, entre un comprador que busca localizarse y un oferente, representando explícitamente la búsqueda y el proceso de negociación. En este proceso los agentes construyen expectativas que se ajustan dinámicamente al estado de la negociación. De esta manera el modelo pretende integrar los procesos de negociación a nivel de agentes y donde ambos, comprador y vendedor, fijan precios en base a sus expectativas de lo cual emerge la dinámica del mercado.

Para su implementación en la ciudad de Singapur se desarrolla una técnica de imputación de datos, basada en imputación multiple tipo "Hot-Deck", se desarrolla también una técnica ad-hoc de calibración para el modelo basada en maximización de la verosimilitud y se realizan pruebas sobre datos obtenidos de transacciones inmobiliarias de Singapur.

Este trabajo forma parte del desarrollo de una plataforma de microsimulación que integra fenómenos de corto, mediano y largo plazo desarrollada en conjunto por el Instituto Tecnológico de Massachusetts (MIT), la Universidad Nacional de Singapur (NUS) y la Universidad de Chile.

Dedicado a los que amo, mi esposa y toda mi familia.

Agradecimientos

En primer lugar quiero dar gracias a mis padres quienes me han entregado de manera incondicional todo su amor, cariño y paciencia. Ellos son quienes me han motivado a perseguir metas ambiciosas, me alentaron y confiaron siempre en mis capacidades. También a mis hermanos: Paz Andrea, Matías, Claudina y mi tía Rosita.

Quiero agradecer a mis amigos de la universidad en particular a Pablo, Alejandro, Rodrigo y María-Paz. Agradezco también el apoyo de todos los nuevos amigos que conocí en estos años del doctorado.

Quiero agradecerle a mi esposa Constanza por acompañarme siempre y apoyarme en la etapa final del doctorado. Por tus consejos, tu cariño y tu amor.

Finalmente, quiero agradecer el consejo y apoyo de los profesores Ricardo Hurtubia, Moshe Ben-Akiva, Angelo Guevara y en particular a Francisco Martínez por introducirme al mundo de la investigación.

Tabla de Contenido

1. Introducción a los artículos	1
1.1. Motivacion y objetivos	1
1.1.1. Antecedentes	1
1.1.2. Objetivos	3
1.1.3. Estructura de la tesis	3
1.2. Revisión Bibliográfica	4
1.2.1. Elección Discreta y Utilidad Aleatoria	4
1.2.2. Modelos de Localización basados en Agentes	10
1.2.3. Sobre la disponibilidad de Datos	12
1.2.4. Modelos de Búsqueda para el Mercado Inmobiliario	12
Bibliografía	14
2. Primer Artículo	20
2.1. Introduction	21
2.2. Classical Formulation of the Bid Rent	24
2.3. Proposed Model	26
2.3.1. Bidding Behavior	26
2.3.2. Probability of the household h to submit a bid of size P_s	29
2.3.3. Probability of Selecting the Unit	29
2.3.4. Probability of Being Accepted	31
2.4. Likelihood Estimation	31
2.5. Numerical Estimation	32
2.5.1. Data Description	32
2.5.2. Asking Prices	33
2.5.3. Model Specification	33
2.6. Results	35
2.7. Conclusions	38
Bibliography	40
3. Segundo Artículo	43
3.1. Introduction	44
3.2. Microsimulation models: brief literature review	46
3.3. The proposed framework	47
3.3.1. A Sellers Model	50
3.3.2. The buyers model	51
3.4. Conclusions and Future Work	52

Bibliography	54
4. Tercer Artículo	57
4.1. Introduction	58
4.2. The Model	60
4.3. Hot Deck Multiple Imputation Algorithm	61
4.4. Proof of concept	63
4.4.1. First test: Synthetic Data	63
4.4.2. Second Test: Brussels' Real Data	66
4.5. Application to Singapore data	68
4.6. Summary and Conclusion	71
Bibliography	71

Índice de tablas

2.1. Willingnes to pay variables	34
2.2. Log-Likelihood vs number of interactions T	36
2.3. Results of the New Model with T=2 (Log-Likelihood: -3819.78)	37
2.4. Results of estimation with Lerman and Kern Model (Log-Likelihood: -10032.96)	37
4.1. Simulated Parameters Estimation	65
4.2. Brussels Interacting Attributes	66
4.3. Lerman and Kern's Model Estimates - Results without imputation	67
4.4. Results obtained with imputation	67
4.5. Hits Survey	69
4.6. REALIS Survey	69
4.7. Singapore Lerman and Kern's model Estimation	70

Índice de figuras

3.2. Evolution of seller's expectation and buyer's bid	48
3.1. Diagram of the Microsimulation Model	49
4.1. Levels of X_1 and X_2 in the linear city	64
4.2. Zoning Structure	64

Capítulo 1

Introducción a los artículos

1.1. Motivacion y objetivos

1.1.1. Antecedentes

Los primeros modelos de uso de suelo corresponden modelos agregados que consideran estructuras de ciudades monocéntricas. Dentro de los modelos clásicos se encuentran los que surgen a partir del trabajo de Alonso (1964), quien plantea un modelo donde se calculan curvas de precio que un individuo estaría dispuesto a pagar a distintas distancias de un centro de una ciudad manteniendo un nivel de utilidad. Posteriormente se desarrollan dos enfoques de modelos basados en la teoría de la elección discreta (McFadden et al., 1978), uno bajo el supuesto de individuos que eligen su localización mediante máxima utilidad y equilibrio de precios Walrasiano (Anas, 1982) llamado *choice* y otro bajo el supuesto de localización mediante máxima postura (Ellickson, 1981) llamado *bid-auction*. En el primero se asume que el tomador de decisión observa los atributos de la localización, entre ellos el precio, el cual surge de un equilibrio Walrasiano, y elige localizarse maximizando su utilidad y pagando el precio exógeno. En el segundo modelo se asume que la vivienda será asignada mediante un remate al mejor postor, definiendo por este mecanismo el precio de venta o de arriendo según corresponda. Ambos modelos corresponden a enfoques de equilibrio estático del mercado inmobiliario. Horowitz (1986) presenta un primer modelo que considera la interacción entre comprador y vendedor mediante un precio de lista (o *asking price*) de la vivienda, sin embargo lo planteado por Horowitz no representa una dinámica explícitamente.

De manera creciente los modelos de interacción entre transporte y uso de suelo tienden a ser más desagregados, llegando finalmente a modelos de microsimulación basados en agentes como son los casos de ILUTE (Miller et al., 2004) y Urbansim (Waddell et al., 2003). En ambos modelos los precios son generados a nivel agregado considerando regresiones hedónicas que no dependen de la demanda de la vivienda en el mercado ni de las características propias del vendedor. La comparación entre usar modelos hedónicos y remate realizada por Hurtubia et al. (2010) concluye que ambos modelos entregan resultados diferentes y los modelos hedónicos

no garantizan la representación de la heterogeneidad en las preferencias de los hogares, ni cambios en la distribución del ingreso.

Otros modelos como ILUMAS (Moeckel et al., 2002) describen la re-localización de los hogares como un proceso de búsqueda (de los compradores) en los cuales los hogares evalúan viviendas una a una y aceptan y pagan un precio de lista sólo en caso de incrementar significativamente su utilidad. Sin embargo, no describe cómo los precios observados son resultado de transacciones entre un comprador y un vendedor, en su lugar se emplea una función agregada del número de localizaciones disponibles en el área.

El modelamiento apropiado del terreno y los mercados inmobiliarios basados en agentes, requiere una representación detallada de la interacción entre los vendedores y compradores, considerando en ambos casos las expectativas y percepciones del mercado. Esta interacción es la que debe dar origen al precio de las viviendas como destacan Filatova et al. (2009).

En un trabajo pionero Salant (1991), formula y resuelve la estrategia óptima para un vendedor neutro al riesgo mediante programación dinámica. Además, muestra que la solución óptima generalmente involucra un decrecimiento monótono de los precios de lista. Sin embargo carece de especificaciones sobre el comportamiento del comprador, quién sólo acepta el precio pedido por el vendedor con una cierta probabilidad, similar a los modelos de elección tipo *choice*. Una extensión de este modelo es presentado por Yavaş (1992), donde considera la interacción entre comprador y vendedor como un proceso de búsqueda, donde al coincidir ambos reparten de manera proporcional sus excedentes. Horowitz (1992) realiza un primer intento de estimación empírica del problema del vendedor, considerando un modelo dinámico en un horizonte fijo, a diferencia de Salant quien considera un horizonte infinito de venta. Otros modelos recogen los trabajos anteriores bajo un enfoque de microsimulación, como por ejemplo el trabajo de Merlo et al. (2013), quienes desarrollan un modelo que incorpora algunos puntos básicos de Salant, sin modelar en detalle el comportamiento del comprador.

Considerando además la interacción del comprador, en el trabajo de Devisch et al. (2005) se describe la transacción como el resultado de una negociación entre un vendedor y un comprador. Este modelo considera el hecho de que los agentes aprenden de los eventos del mercado y emplean esa información en las ofertas siguientes en un proceso bilateral, considerando un enfoque basado en agentes. Todos estos trabajos han sido desarrollados sin mayor aplicación a datos observados de transacciones.

Con el aumento del uso de las herramientas de microsimulación para modelos de uso de suelo, se plantea el desafío de considerar la interacción entre los agentes del sistema en un nivel detallado, explotando los beneficios de contar con la diversidad de sus características individuales y el impacto sobre el mercado. Si bien existen aproximaciones para obtener estos resultados, carecen de soporte microeconómico que permita incorporar algunas decisiones de los individuos, las que se consideran usualmente como exógenas. En este trabajo se busca desarrollar un mecanismo endógeno que represente la construcción de expectativas para modelar la interacción entre un vendedor y un comprador de un bien inmobiliario,

considerando aspectos como el precio de lista de la vivienda, percepción sobre los potenciales compradores en cada instante y un mecanismo de asignación.

1.1.2. Objetivos

El objetivo principal es desarrollar un modelo negociación entre comprador y vendedor de un bien inmobiliario, considerando el enfoque microeconómico para una elección de bienes discretos, y la teoría de remates, además del proceso dinámico de una negociación. Además, implementar el modelo computacionalmente y generar una técnica de calibración y estimación de parámetros. Finalmente se busca probar su aplicación sobre datos de transacciones obtenidos de la ciudad de Singapur.

Los objetivos específicos son:

- a) Desarrollar un modelo para un hogar que considere cambiar su localización o ingresar al mercado inmobiliario. Modelar el proceso de elección de la residencia entre las opciones disponibles en el mercado y la construcción de la postura (bid) que dependa dinámicamente de sus expectativas de lograr éxito en la transacción, en función del precio publicado de venta, su excedente y consideraciones sobre el comportamiento de los otros jugadores.
- b) Desarrollar un modelo dinámico para el vendedor de una propiedad, que genere el precio de venta a publicar mediante la construcción de expectativas de ventas, las que dependerán de consideraciones sobre los atributos y precios de las viviendas en su misma zona, considerando además el potencial que presenta el precio como un atractor negativo de ofertas. El modelo debe incluir un mecanismo para la aceptación de ofertas donde un vendedor pueda elegir una oferta sobre su expectativa o precio de reserva, o esperar un periodo definido de tiempo antes de realizar la elección.
- c) Definir una técnica de estimación y calibración basada en metodologías de calibración de microsimulación, que permita recuperar la dinámica de la ciudad y los valores de precios observados. Se realizará una calibración basada en máxima verosimilitud para las funciones del modelo, mientras que la dinámica se calibrará estableciendo una función objetivo y resolviendo un problema de optimización.
- d) Desarrollar técnicas de imputación de datos para combinar bases de datos parciales de oferta y demanda residencial

1.1.3. Estructura de la tesis

Esta tesis consiste en dos partes, en la primera se realiza una breve revisión bibliográfica, donde se revisa la literatura correspondiente a modelos de elección discreta, modelos de uso de suelo y modelos de localización basados en agentes en términos generales.

En una segunda parte se adjuntan tres artículos científicos que resuelven las inquietudes planteadas en esta tesis. Un primer artículo describe el comportamiento de un hogar que busca relocalizarse, para ello se desarrolla un modelo de ofertas dinámicas basado en la teoría

de rentas. En un segundo artículo se presenta la dinámica para un modelo de microsimulación, la cual toma el modelo para un comprador desarrollado en un primer artículo e incorpora un modelo de búsqueda para un vendedor, describiendo una interacción dinámica. Un tercer artículo es incluido para mostrar la aplicación de una técnica de imputación de datos que permite realizar las estimaciones de los modelos clásicos de localización y el modelo propuesto para el comprador.

1.2. Revisión Bibliográfica

1.2.1. Elección Discreta y Utilidad Aleatoria

La teoría de elección discreta corresponde a una de las principales herramientas empleadas en la actualidad para la modelación y predicción de las elecciones que realizará un individuo o agente entre un grupo de posibles alternativas mutuamente excluyentes, las cuales son de conocimiento del individuo.

El conjunto de elección, *choice set* debe presentar tres principales características. Deben ser mutuamente excluyentes desde la perspectiva de la decisión del individuo. El conjunto de elección debe ser exhaustivo, es decir todas las posibles alternativas deben estar representadas y en tercer lugar el conjunto de alternativas debe ser finito.

La derivación de las probabilidades de elección normalmente se basan en los supuestos de maximización de la utilidad del tomador de decisión. Estos conceptos fueron derivados originalmente por Thurstone (1927) en termino de estímulos psicológicos. El trabajo de Marschak (1950) interpretó los estímulos como utilidades y derivó las probabilidades como una maximización de la utilidad. Estos modelos derivados siguiendo el enfoque de Marschack, son usualmente llamados modelos de utilidad aleatoria (*Random utility models*)

Un individuo h realizará una elección de un conjunto de I alternativas disponibles. h obtendrá un nivel de utilidad (o ganancia) de cada alternativa de tamaño U_{ih} para la alternativa $\forall i \in I$. Esta utilidad es comúnmente asumida de conocimiento para el agente h pero no para el modelador. El individuo h elige de las alternativas solamente aquella que le entrega mayor utilidad, es decir, escoge la alternativa i tal que $U_{ih} > U_{kh} \forall k \neq i \in I$.

Desde la perspectiva del modelador, resulta imposible conocer la utilidad del agente h directamente. El modelador solo tendrá a la vista algunos atributos de las alternativas tal como las observa h llamados x_{hi} , también observará algunas características o atributos de h , llamados s_h , para todas las alternativas $i \in I$. El modelador entonces especificará una función que relacione los factores observados con los atributos del individuo h , llamando $\hat{U}_{hi} = \hat{U}(x_{hi}, s_h)$ para todas las alternativas $i \in I$. Esta función \hat{U}_{hi} es comúnmente llamada utilidad representativa.

La utilidad del individuo se descompone en la utilidad representativa mas un término que

captura los factores que afectan a la utilidad y que no son incluidos en \hat{U}_{hi} , de esta forma se puede escribir $U_{hi} = \hat{U}_{hi} + \varepsilon_{hi}$.

Debido a que el modelador no conoce ε_{hi} para todas las alternativas i , es que el error se trata normalmente como un término aleatorio. El vector que contiene todas las diferencias entre las utilidades reales y observadas $\varepsilon_h = (\varepsilon_{h1}, \varepsilon_{h2}, \dots, \varepsilon_{hI})$ tiene una densidad $f(\varepsilon_h)$. Las distintas especificaciones dan origen a distintas formas de modelos de elección discreta. De manera general, la probabilidad de que un individuo h escoja una alternativa i esta dada por:

$$P_{hi} = \Pr(\hat{U}_{hi} + \varepsilon_{hi} > \hat{U}_{ki} + \varepsilon_{ki} \quad \forall k \neq i \in I) \quad (1.1)$$

Considerando la función indicadora $\mathbb{I}()$ que toma el valor de 1 cuando el término en paréntesis es verdad y 0 en caso contrario, la probabilidad P_{hi} se puede escribir de la siguiente forma

$$P_{hi} = \Pr(\hat{U}_{hi} + \varepsilon_{hi} > \hat{U}_{ki} + \varepsilon_{ki} \quad \forall k \neq i \in I) \quad (1.2)$$

$$= \Pr(\hat{U}_{hi} - \hat{U}_{ki} > \varepsilon_{ki} - \varepsilon_{hi} \quad \forall k \neq i \in I) \quad (1.3)$$

$$= \int_{\varepsilon} \mathbb{I}(\varepsilon_{ki} - \varepsilon_{hi} < \hat{U}_{hi} - \hat{U}_{ki} \quad \forall k \neq i \in I) f(\varepsilon_h) d\varepsilon_h \quad (1.4)$$

Esta ultima expresión corresponde a una integral multidimensional sobre la densidad de las partes no observadas de las utilidades. La expresión anterior solo toma valores cerrados para ciertas especificaciones de densidad. Por ejemplo si f es una normal multivariada se obtiene el modelo Probit, si los errores son independientes e idénticamente distribuidos (i.i.d.) siguiendo una distribución de valor extremo se obtiene el clásico modelo Logit.

McFadden et al. (1978) realiza una extensión econométrica del modelo, al considerar que una población de individuos realiza la misma elección sobre un conjunto de alternativas y luego determina la fracción de la población que escogerá una alternativa sobre la otra.

Muchos aspectos del proceso de decisión afectan la especificación y estimación de los modelos de elección discreta. Principalmente debido a que solamente las diferencias de utilidades importan y que la escala de la utilidad es arbitraria.

El efecto de las diferencias entre utilidades puede ejemplificarse suponiendo que a todas las alternativas se les agrega un termino constante finito positivo, las elecciones permanecerán inalteradas. El individuo elegirá la misma alternativa que antes de agregar la constante.

El nivel de utilidad tampoco afecta a los individuos pues la probabilidad de elección depende solamente de la diferencia entre utilidades y no del valor absoluto de esta diferencia.

Las implicancias de estos aspectos recaen en que solamente podrán ser estimados parámetros que capturen efectivas diferencias entre las alternativas. Por ejemplo en le caso de las constantes específicas para cada alternativa. En términos generales se podrá estimar $I - 1$ constantes específicas cuando se tengan I alternativas. Las variables socio-demográficas que ingresan a la utilidad y que presenten poco variación sobre las alternativas solo podrán entrar al modelo si están especificados en formas en que generen diferencias en la utilidad sobre las alternativas.

En los casos en que las variables socio-demográficas, tales como el ingreso u otras, quieran ser incorporadas a la función de utilidad representativa, estas podrán ingresar interactuando con otras variables. Por ejemplo, el ingreso muchas veces ingresa dividiendo al costo de una alternativa. En el desarrollo de este trabajo, la inclusión de variables socio-demográficas será empleado para la estimación de modelos de localización.

En el caso del nivel de utilidad, debido a que la escala esta directamente relacionada con la varianza, cuando la utilidad es amplificada por un factor positivo λ la varianza aumenta en λ^2 . En modelos logit, usualmente el parámetro de escala μ es fijado en 1.

Cuando los errores ε_{hi} distribuyen i.i.d. Gumbel, con media 0 y parámetro de escala μ la probabilidad de elección de la alternativa i por un individuo h toma la siguiente forma:

$$P_{hi} = \frac{\exp(\mu \hat{U}_{hi})}{\sum_{j \in A_h} \exp(\mu \hat{U}_{hj})} \quad (1.5)$$

Aplicación de los modelos de Localización

En el contexto de localización urbana, los modelos de elección son descritos como los procesos en los que los agentes urbanos (hogares, firmas) se asignan con las localizaciones. La naturaleza discreta del proceso de decisión hace que los modelos de utilidad aleatoria sean una herramienta apropiada para realizar estimaciones y predecir las localizaciones. Domencich and McFadden (1975) fueron los primeros en proponer los modelos de elección discreta para comprender la localización residencial, como una manera de predecir la disponibilidad de los diferentes modos de transporte en el contexto de la demanda de viaje.

Dos enfoques clásicos existen en la modelación de la localización mediante elección discreta. Un enfoque llamado Choice y otro llamado Bid. Estos enfoque surgen de la apreciación del problema desde la perspectiva de la oferta residencial o la demanda. Bajo el modelo Choice los agentes seleccionan la localización que maximiza su utilidad, bajo el supuesto de que ellos se comportan como tomadores de precio. En el enfoque Bid-Auction, se asume que las viviendas son comerciadas en un mercado de remates, donde la localización y la renta se definen mediante el mejor postor.

A continuación se describen ambos enfoques de manera detallada:

Enfoque Choice

El problema de la localización puede ser expresado como el problema micro-económico clásico, donde un consumidor escoge la localización que le reporta máxima utilidad sujeto a una restricción de ingreso.

$$\begin{aligned} \text{máx } U(x, z_i) \\ px + r_i \leq I \end{aligned} \quad (1.6)$$

En este problema el consumidor maximiza su utilidad al escoger un vector de bienes continuos x y una localización discreta i , la cual es descrita mediante un set de atributos z_i . En la restricción de presupuesto se considera que el total gastado en bienes de consumo a precio p mas el precio de la vivienda seleccionada r_i debe ser menor o igual que el total de ingreso disponible al consumidor I . La resolución de este problema en el vector de bienes continuos x y considerando activa la restricción de ingresos nos permite reescribir el problema (Rosen, 1974).

$$\max_i V(p, I - r_i, z_i) \quad (1.7)$$

En este caso V represente la utilidad indirecta condicional en la localización i . Al considerar que los distintos individuos tienen diversas preferencias y que las alternativas de localización tienen atributos no observables podemos escribir la utilidad como $V_{hi} + \varepsilon_i$, donde ε corresponde a un término de error estocástico.

Siguiendo la formulación de los modelos de elección discreta generales, y considerando que el vector de errores se compone de variables aleatorias i.i.d. de Valor Extremo, con media 0 y parámetro de escala μ , la probabilidad de que un hogar h escoja una localización i que pertenece a un conjunto de alternativas S esta dada por (McFadden et al., 1978)

$$P(i|h, S) = \frac{\exp(\mu V_{ih})}{\sum_{j \in S} \exp(\mu V_{jh})} \quad (1.8)$$

En este enfoque de elección, el precio de la vivienda r_i se encuentra explícito en la función de utilidad. Esto genera que un supuesto del modelo sea que los individuos h sean tomadores de precio. Estos precios r_i se pueden obtener como una función de atributos de la localización i , siguiendo el marco teórico para precios hedónicos propuesto por Rosen (1974) a través del resultado de la interacción entre oferta y demanda (Anas, 1982) Una desventaja de este enfoque es la gran correlación que podría existir entre el precio de la localización y los atributos de la misma, lo cual podría llevar a problemas de endogeneidad. Esto generaría parámetros que no sean significativos o incluso que no sean positivos. Un tratamiento para este problema de endogeneidad es estudiado por Guevara and Ben-Akiva (2006).

Enfoque Bid-Auction

El mercado inmobiliario fue modelado por primera vez como un mercado de remate por primera vez en el trabajo de Alonso 64, en el cual los hogares ofrecen su disposición a pagar por una vivienda i la cual es asignada al mejor postor. Este proceso define de manera simultánea la localización del agente y el precio a pagar, entendido como la postura mas alta del remate.

Desde el punto de vista económico, la disposición a pagar puede ser derivada desde el problema del consumidor planteado para el caso choice.

Consideremos el problema planteado 1.6 que entrega como resultado una utilidad indirecta condicional 1.7, al fijar un nivel de utilidad condicional V_0 e invertir en el precio de la vivienda r_i , de esta forma se obtiene:

$$r_i = I - V^{-1}(V_0, p, z_i) \quad (1.9)$$

Bajo el enfoque del remate, este precio se puede interpretar como la disposición a pagar para

una localización particular i condicional en el nivel de utilidad V_0 . (Jara-Díaz and Martínez, 1999)

De esta forma, la función de postura B_{hi} para el individuo h en la localización i puede expresarse como

$$B_{hi} = I_h - V^{-1}(V_{0h}, p, z_i) \quad (1.10)$$

La postura B_{hi} se puede entender como el máximo valor que un individuo h puede pagar por una localización particular i y obtener un nivel de utilidad V_{0h} .

Ellickson (1981) propone una postura que considere la heterogeneidad en preferencias dentro de grupos de hogares mediante un termino aleatorio ε_h , considerando que el término de error sigue una distribución aleatoria i.i.d. de Valor Extremo, entonces la probabilidad de ser el mejor postor se puede expresar como un modelo Logit.

$$P(h|i, H) = \frac{\exp(\mu B_{hi})}{\sum_{k \in H} \exp(\mu B_{ki})} \quad (1.11)$$

Debemos notar que las dos expresiones choice (ecuación 1.8) y bid-auction (ecuación 1.11) son similares, pero representan supuestos totalmente diferentes. La equivalencia entre ambos modelos es estudiada por Martínez (1992), quien muestra que ambos enfoques son equivalentes bajo supuestos de mercados competitivos.

A continuación se describen como se realizan las estimaciones de estos modelos.

Estimación de modelos

En su modelo Ellickson (1981) formula un modelo Logit, cuya estimación supone que cada agente se localiza en el lugar donde es el mejor postor, pero como es común en los modelos Logit, los parámetros son estimables salvo el parámetro de escala. Este problema de identificación no permite a la estimación de Ellickson obtener estimaciones de valores absolutas de las posturas, salvo obtener estimaciones relativas entre ellas. El trabajo de Ellickson es el primer trabajo en plantear la estimación de la función de postura de los agentes. Para ello Ellickson plantea un modelo lineal donde la estimación es realizada mediante maximización de la verosimilitud, la cual esta planteada en la siguiente ecuación:

$$= \prod_{i \in S} \prod_{h \in C_i} P(h|i)^{y_{hi}} \quad (1.12)$$

Donde y_{hi} toma el valor de 1 si se observa al hogar h localizado en la vivienda i . $P(h|i)$ se encuentra definido en la ecuación (1.11)

En un trabajo posterior, Lerman and Kern (1983) plantean incorporar la información del precio de transacción observado en la probabilidad. De esta forma, llamando P_i al precio observado de la vivienda i :

$$P(h|i) = \Pr[B_{hi} = P_i + \varepsilon_h > B_{ki} + \varepsilon_k \quad \forall k \in H] \quad (1.13)$$

Luego si los errores son i.i.d. y siguen una distribución Gumbel,

$$P(h|i) = f(B_{hi} = P_s) \prod_{k \neq h} F(B_{ki}) \quad (1.14)$$

Donde:

$$f(x) = \mu \exp(-\mu x) \exp(\exp(-\mu x)) \quad (1.15)$$

$$F(x) = \exp(-\exp(\mu)) \quad (1.16)$$

Luego la verosimilitud queda representada por

$$L = \prod_{i \in S} \left(-\mu \exp(-\mu(P_i - B_{hi})) \prod_{k \neq h \in H} \exp(-\exp(-\mu(P_i - B_{ki}))) \right)_{hi}^y \quad (1.17)$$

Donde H corresponde al conjunto total de hogares en el remate, S es el total de viviendas en el mercado. Este modelo presenta la propiedad, al igual que el modelo de Ellickson, de ser estimables de manera consistente, pues ambas funciones de verosimilitud son convexas.

Variaciones sobre el modelo de Lerman and Kern (1983) pueden encontrarse en la literatura, por ejemplo Hurtubia and Bierlaire (2014) realiza una estimación del modelo considerando que el remate nunca se realiza, de esta manera no obliga al remate a ser igual a la máxima postura. (Muto, 2006) levanta el supuesto de errores i.i.d. en el modelo para definir errores diferentes entre viviendas de uso residencial y de uso comercial.

Comentarios

Los modelos de uso de suelos basados en la teoría de elección discreta son ampliamente utilizados actualmente (Wegener, 2004, Hunt et al., 2005). Sin embargo, ambas representaciones entregan informaciones diferentes de como ocurre el equilibrio de mercado.

El enfoque bid-auction permite obtener el precio de la vivienda como el valor esperado de la variable B_{hi} , de esta forma :

$$P_i = E[\text{máx } B_{ki}, k \in H] \quad (1.18)$$

Dado que los errores son i.i.d. Gumbel, podemos escribir :

$$P_i = \frac{1}{\mu} \ln \left(\sum_{g \in H} \exp(\mu B_{gi}) \right) + \frac{\mu}{\gamma} \quad (1.19)$$

Con γ igual a la constante de Euler.

Al definir el precio de la vivienda mediante la interacción de todos los oferentes, se engloba la dinámica del mercado y la formación de precio de manera explícita.

El modelo bid-choice de Martínez ha dado origen a diversos modelos tales como MUSSA (Martinez, 1996, Martínez and Henríquez, 2007)

1.2.2. Modelos de Localización basados en Agentes

El deseo de representar y explorar de mejor manera los sistemas económicos ha llevado al desarrollo y aplicación de modelos basados en agentes, los cuales han visto un gran incremento en el pasar del tiempo (Wegener, 2004).

A pesar de su rápido crecimiento, los modelos basados en agentes son relativamente nuevos, su aplicación se puede observar en áreas que incluyen mercados financieros, mercados de emisión de contaminaciones, entre otros. El concepto de equilibrio es central para la gran mayoría de los modelos económicos, sin embargo los modelos normalmente se encuentran en adaptación dinámica y pueden permanecer fuera de un estado de equilibrio Arthur (2006). Empleando modelos de mercado basados en agentes la dinámica hacia un equilibrio puede ser modelado.

Los modelos de equilibrio muchas veces recogen el enfoque de un agente representativo, como es el caso de los modelos de Ellickson y Lerman-Kern presentados antes. En estos modelos el comportamiento de un agente (u hogar) representa el comportamiento de un grupo heterogéneo de hogares en la realidad, pero homogéneos en atributos socio-económicos, como por ejemplo en nivel de ingreso. Los modelos de equilibrio excluyen la mayor parte de la interacción agente-agente, lo cual ignora el efecto que las negociaciones podrían tener durante la formación del precio de un bien.

Mientras que la mayoría de los modelos de microsimulación se basan en elección discreta Hunt et al. (2005), Horowitz (1986) presenta un modelo a nivel de agente, en el que las ofertas son aceptadas o rechazadas en la medida en la que son recibidas. En este modelo, los compradores no son capaces de juntar todas ofertas que llegarán antes de tomar una decisión. Por lo anterior, el vendedor no necesariamente aceptará la oferta más alta. Los compradores potenciales de las viviendas llegan a estas ignorando las otras ofertas realizadas sobre la vivienda. La ignorancia de este proceso lleva a los compradores a realizar ofertas mas altas o bajas que las que realizarían con información completa.

En el modelo de Horowitz, un hogar realizará una compra tratando de maximizar su utilidad esperada, donde u_c representa su nivel de utilidad condicional actual y $H(b)$ corresponde a la probabilidad que un comprador percibe de localizarse

$$E(u) = H(b)u(b, z_i) + (1 - H(b))u_c \quad (1.20)$$

Horowitz emplea la siguiente forma funcional para la distribución $H(b)$

$$H(b) = \exp\left(\frac{b - P_a}{\alpha}\right) \text{ if } b \leq P_a \quad (1.21)$$

Donde P_a corresponde al precio de lista o *asking price* de la vivienda. Podemos notar la semejanza en la función de probabilidad $H(b)$ empleada por Horowitz con la forma funcional obtenida si el precio de lista correspondiera a la máxima utilidad esperada, es decir Si:

$$P_a = \frac{1}{\mu} \ln \left(\sum_{g \in H} \exp(\mu B_{gi}) \right) \quad (1.22)$$

Con B_{gi} correspondiente a la postura de los hogares $g \in H$, donde H es el conjunto de todos los compradores.

Entonces (1.21) corresponde a la ecuación $P(h|i)$ que se encuentra en la ecuación (1.11)

Finalmente Horowitz estima una función de postura dada por la siguiente relación:

$$B_{hi}^* = \begin{cases} w_c - \alpha & \text{if } B_{hi}^* \leq P_a \\ P_a & \text{otherwise} \end{cases} \quad (1.23)$$

Donde C_0 corresponde a un costo fijo y C_1 a un costo variable por realizar la venta. $W_c(z_i)$ corresponde a la disposición a pagar para mantener un nivel de utilidad u_c evaluada en los atributos de la vivienda i , z_i .

Como se puede ver, el modelo de Horowitz corresponde a un primer acercamiento en la reproducción de una secuencia dinámica de ofertas, sin embargo la función de postura corresponde a un modelo estático, a pesar de que la función de probabilidad H incorpora el precio de lista P_a , su efecto final en la postura no se encuentra representado.

Los trabajos de Parker and Filatova (2008), Farooq et al. (2013), Parker et al. (2003) destacan las ventajas de realizar modelos basados en agentes para modelar el uso de suelo. Resaltando las capacidades de enfocar de manera heterogénea a los agentes, y capturando interacciones agente-agente que además afectan las decisiones de los demás mediante externalidades generadas por su localización.

En modelos operativos, UrbanSim (Walker, 2001) corresponde a un modelo de simulación conformado por submodelos que interactúan entre si. Urbansim emplea un enfoque de desequilibrio dinámico, donde los ajustes se producen en distintas escalas temporales. A diferencia de la teoría bid-auction, relaja el supuesto de equilibrio, mediante la consideración de precios temporalmente exógenos para el proceso de localización. Los compradores u hogares se asumen como tomadores de precio. Luego, los precios de transacción son determinados mediante una función hedónica, la cual se actualiza con información de las condiciones anteriores del mercado. Esta función hedónica también toma en consideración las características del espacio construido, así como también condiciones del barrio y accesibilidades. En cada periodo de tiempo, los hogares, firmas y agentes son evaluados. La localización de los agentes en el mercado es el resultado de la aplicación de un modelo Logit Multinomial.

El mecanismo de localización de UrbanSim genera una localización evaluada anualmente, en lugar de ser el resultado de decisiones de movilidad encadenadas. Este modelo no representa un precio de las viviendas como el resultado endógeno de la localización.

El trabajo de Rosenfield et al. (2013) describe un modelo de localización de hogares para ser empleado en el model ILUTE Miller et al. (2004). En el cual se describen las posturas en función de la disposición a pagar para las viviendas considerando el efecto del precio de lista de la vivienda. En este modelo todo hogar h tiene un set de viviendas disponibles C_h y cada vivienda activa tiene un set de compradores B_i . El mecanismo empleado para la localización consiste en que los compradores determinan su utilidad no-precio para cada vivienda en su conjunto C_h , esta utilidad es evaluada en función de los atributos de la vivienda i . El error ε_{hi} es simulado mediante una distribución de valor extremo. Luego los compradores determinan el nivel de utilidad mas alto que pueden alcanzar para todos los hogares en su conjunto, salvo su actual vivienda i El precio de lista P_{aj} refleja la valoración en el mercado para cada vivienda.

1.2.3. Sobre la disponibilidad de Datos

Un problema comúnmente encontrado en la estimación de modelos de elección discreta, en particular ante modelos que requieren tanta información esta dada por la disponibilidad de datos. Por ejemplo, Picard and Antoniou (2011), destaca que el desarrollo de modelos más desagregados requiere datos mas detallados. En el contexto de un proyecto, Picard destaca que los datos recolectados para Bruselas, París y Zurich necesitaron ser procesados, igualados y conectados, antes de ser utilizados.

En algunos casos los precios son un secreto de los desarrolladores de viviendas, quienes no desean dar a conocer los detalles de las transacciones inmobiliarias (Kong et al., 2007), una salida posible es utilizar programas que capturen información de las características de las viviendas de los portales o información pública, sin embargo este tipo de información no describirá la vivienda que habita esta localización.

1.2.4. Modelos de Búsqueda para el Mercado Inmobiliario

Los modelos de búsqueda para el mercado inmobiliario pueden considerarse desde el problema clásico de parada optima para un vendedor. El problema llamado *house selling problem* consiste en que una secuencia de ofertas $b_{i=1..N}$ son ofrecidas a un vendedor, las ofertas deben ser aceptadas o rechazadas por el vendedor, quien debe decidir el momento optimo para aceptar la oferta. La solución común para este problema consiste en encontrar el valor esperado de la venta V^* y luego aceptar la primera oferta que se encuentre sobre este valor Si las ofertas llegan siguiendo una distribución F entonces el problema se encuentra resolviendo.

Donde c corresponde al costo de esperar. Feinberg and Johnson (1977) muestran resultados para distribuciones normal y cuadrada. Siguiendo un enfoque similar, Haurin (1988) muestra que una vivienda que presente características atípicas recibirá ofertas con mayor varianza, luego estará un mayor tiempo en el mercado. En el modelo de Salant (1991) se estudia el comportamiento del vendedor considerando el precio de lista o *asking price* de una vivienda. Para ello, Salant considera que en un periodo de tiempo a lo mas un comprador visitara a un

vendedor, la probabilidad por periodo que un comprador llegue dependerá si el comprador use o no una agencia para vender su vivienda. Si un comprador ofrece un precio mayor que el esperado por el vendedor ocurre la transacción al asking price fijado por el vendedor. Si no existe ninguna oferta, la vivienda permanece sin ser vendida hasta el próximo periodo. El modelo considera un número fijo de T periodos, al cabo del cual el vendedor decide no vender su vivienda y obtiene un valor fijo por ella. Este modelo corresponde al primero que considera la determinación endógena de secuencias de precios de lista para realizar la transacción.

Horowitz (1992) presenta un modelo donde estudia el rol del *asking price* y como este se relaciona con un precio de reserva en el mercado inmobiliario. Horowitz considera un modelo estacionario donde el precio aceptado se encuentra entre el precio de reserva y el asking price fijado por el vendedor. Para encontrar el precio de reserva y el precio de lista, el vendedor maximiza una función que representa el valor esperado de la venta.

Ettema (2011) modela a un vendedor mediante la maximización del precio esperado de venta. Para ello determinan que el precio esperado de venta, asumiendo que para cada precio que se fije existe una probabilidad conocida de vender la vivienda. En caso de no poder realizar la venta, se espera vender a un precio reducido en un factor α exógeno. La novedad del modelo es que permite actualizar las probabilidades esperadas de manera endógena.

Otros modelos que emplean la búsqueda, como por ejemplo el modelo de Merlo et al. (2013) o Albrecht et al. (2014) describen la interacción entre un vendedor y los compradores mediante el empleo de un modelo de búsqueda donde el vendedor determina el precio de venta para maximizar su retorno esperado de la transacción. En todos los modelos mencionados antes, el comportamiento del comprador es modelado suponiendo que las ofertas llegan con valores que distribuyen de manera aleatoria.

Comentarios

De la revisión bibliográfica podemos notar que gran parte de los modelos basados en agentes, salvo contadas ocasiones tienden a simplificar la generación de las ofertas; las cuales generalmente son obtenidas de una distribución, o bien presentan comportamientos poco realistas, como por ejemplo en Carrillo (2012) donde el comprador luego de realizar una oferta, realiza una contra-oferta correspondiente al precio de reserva del vendedor.

Esta simplificación es enfrentada en el primer artículo, mediante un desarrollo que recoge la formulación de Horowitz (1986) para desarrollar un modelo que incorpora supuestos microeconómicos y genera una secuencia de ofertas que permitan simular el comportamiento del agente.

El rol de precio de lista, o *asking price*, como herramienta para indicar y orientar el precio de una vivienda ha sido sub-utilizado en los modelos de microsimulación, en esta tesis se desarrolla un modelo de interacción comprador vendedor donde el precio de venta corresponde al resultado del ajuste de expectativas por parte de ambos.

Los resultados de este trabajo se encuentran sujetos a la disponibilidad de datos. Para superar la falta de datos, en el primer artículo, ante la falta de disponibilidad de precios de

lista o *asking prices* para realizar la estimación del modelo, se considera a la variable no observada como una variable latente del modelo, permitiendo estimar de manera consistente los parámetros restantes del modelo. En el tercer artículo se presenta la aplicación de una técnica de imputación múltiple para la realizar la estimación de un modelo bid-auction. El cual permite emplear datos de dos bases de datos separadas para la estimación de un modelo.

Bibliografía

- Albrecht, J., Gautier, P. A., and Vroman, S. (2014). Directed search in the housing market. *Econometrica*, 82(2):705–30.
- Alonso, W. (1964). Location and land use. toward a general theory of land rent. *Location and land use. Toward a general theory of land rent*.
- Anas, A. (1982). Residential location markets and urban transportation: Economic theory, econometrics, and policy analysis with discrete choice models.
- Andridge, R. R. and Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, 78(1):40–64.
- Arthur, W. B. (2006). Out-of-equilibrium economics and agent-based modeling. *Handbook of computational economics*, 2:1551–1564.
- Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Börsch-Supan, A., Brownstone, D., Bunch, D. S., et al. (2002). Hybrid choice models: Progress and challenges. *Marketing Letters*, 13(3):163–175.
- Carrillo, P. E. (2012). An empirical stationary equilibrium search model of the housing market*. *International Economic Review*, 53(1):203–234.
- Chingcuanco, F. and Miller, E. J. (2013). Demographic microsimulation model for integrated land use, transportation, and environment model system. In *Transportation Research Board 92nd Annual Meeting*, number 13-3445.
- Cranmer, S. J. and Gill, J. (2013). We have to be discrete about this: A non-parametric imputation technique for missing categorical data. *British Journal of Political Science*, 43(02):425–449.
- Curlette, W. L. (2006). A framework for research studies: Mixed methods through combining bayesian statistics and qualitative research in individual psychology. *The Journal of Individual Psychology*.
- Devisch, O., Arentze, T., Borgers, A., and Timmermans, H. (2005). An agent-based model of residential choice dynamics in non-stationary housing markets. In *Paper CUPUM Conference, London*.
- DiPasquale, D. and Wheaton, W. C. (1996). *Urban economics and real estate markets*,

volume 23. Prentice Hall Englewood Cliffs, NJ.

- Domencich, T. A. and McFadden, D. (1975). Urban travel demand-a behavioral analysis. Technical report.
- Ellickson, B. (1981). An alternative test of the hedonic theory of housing markets. *Journal of Urban Economics*, 9(1):56–79.
- Ettema, D. (2011). A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. *Computers, environment and urban systems*, 35(1):1–11.
- Farooq, B. and Miller, E. J. (2012). Towards 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.
- Farooq, B., Miller, E. J., Chingcuanco, F., and Giroux-Cook, M. (2013). Microsimulation framework for urban price-taker markets. *Journal of Transport and Land Use*, 6(1):41–51.
- Fay, R. E. (1996). Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association*, 91(434):490–498.
- Feinberg, R. M. and Johnson, W. R. (1977). The superiority of sequential search: a calculation. *Southern Economic Journal*, pages 1594–1598.
- Feldman, O., Simmonds, D., Ballas, D., Clarke, G., Gibson, P., Jin, J., and Stillwell, J. (2005). A spatial microsimulation approach to land-use modelling. In *CUPUM conference*, volume 29.
- Filatova, T., Parker, D., and van der Veen, A. (2009). Agent-based urban land markets: agent’s pricing behavior, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation*, 12(1):3.
- Frazier, T. and Alfons, A. (2012). Generating a close-to-reality synthetic population of ghana. *FBE Research Report KBI_1206*, pages 1–15.
- 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(1):60–66.
- Haurin, D. (1988). The duration of marketing time of residential housing. *Real Estate Economics*, 16(4):396–410.

- Horowitz, J. L. (1986). Bidding models of housing markets. *Journal of urban economics*, 20(2):168–190.
- Horowitz, J. L. (1992). The role of the list price in housing markets: Theory and an econometric model. *Journal of Applied Econometrics*, 7(2):115–129.
- Horton, N. J. and Lipsitz, S. R. (2001). Multiple imputation in practice: comparison of software packages for regression models with missing variables. *The American Statistician*, 55(3):244–254.
- Hunt, J. D., Kriger, D. S., and Miller, E. J. (2005). Current operational urban land-use–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*, 20:221–221.
- Hurtubia, R. and Bierlaire, M. (2014). Estimation of bid functions for location choice and price modeling with a latent variable approach. *Networks and Spatial Economics*, 14(1):47–65.
- Hurtubia, R., Flötteröd, G., Martinez, F., and Bierlaire, M. (2010). Comparative analysis of hedonic rents and maximum bids in a land-use simulation context. In *Swiss Transport Research Conference*.
- 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.
- Kirman, A. P. (1992). Whom or what does the representative individual represent? *The Journal of Economic Perspectives*, pages 117–136.
- Kong, F., Yin, H., and Nakagoshi, N. (2007). Using gis and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in jinan city, china. *Landscape and Urban Planning*, 79(3):240–252.
- 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.
- Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica: Journal of the Econometric Society*, pages 111–141.
- Martinez, F. (1996). Mussa: land use model for santiago city. *Transportation Research Record: Journal of the Transportation Research Board*, 1552(1):126–134.
- Martínez, F. and Donoso, P. (2011). A quasi-equilibrium logit model for residential suburbanization. In *Computers in Urban Planning and Urban Management 2011 Conference*.
- Martinez, F. J. (1992). The bid-choice land-use model: an integrated economic framework. *Environment and Planning A*, 24(6):871–885.

- Martínez, F. J. and Henríquez, R. (2007). A random bidding and supply land use equilibrium model. *Transportation Research Part B: Methodological*, 41(6):632–651.
- McFadden, D. et al. (1978). *Modelling the choice of residential location*. Institute of Transportation Studies, University of California.
- McMillen, D. P. (1997). Multiple regime bid-rent function estimation. *Journal of Urban Economics*, 41(2):301 – 319.
- Merlo, A., Ortalo-Magné, F., and Rust, J. (2008). The home selling problem: Theory and evidence. In *Cowles Foundation Summer Conference*.
- Merlo, A., Ortalo-Magné, F., and Rust, J. (2013). The home selling problem: Theory and evidence. Technical report, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- Miller, E. J., Douglas Hunt, J., Abraham, J. E., and Salvini, P. A. (2004). Microsimulating urban systems. *Computers, Environment and Urban Systems*, 28(1):9–44.
- Moeckel, R., Schürmann, C., and Wegener, M. (2002). Microsimulation of urban land use. In *42nd European Congress of the Regional Science Association, Dortmund*, pages 27–31.
- Muto, S. (2006). Estimation of the bid rent function with the usage decision model. *Journal of Urban Economics*, 60(1):33 – 49.
- Parker, D. C. and Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems*, 32(6):454–463.
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., and Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the association of American Geographers*, 93(2):314–337.
- Picard, N. and Antoniou, C. (2011). Econometric guidance for developing urbansim models. first lessons from the sustainability project. In *Proceedings of the 51st European Congress of the Regional Science Association International, Barcelona, Spain*.
- Rocco, V. A., Hurtubia, R., and Martinez, F. J. (2014). A dynamic bidding model for the housing market. *Unpublished*.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, pages 34–55.
- Rosenfield, A., Chingcuanco, F., and Miller, E. J. (2013). Agent-based housing market microsimulation for integrated land use, transportation, environment model system. *Procedia Computer Science*, 19:841–846.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.

- Salant, S. W. (1991). For sale by owner: When to use a broker and how to price the house. *The Journal of Real Estate Finance and Economics*, 4(2):157–173.
- Salvini, P. and Miller, E. J. (2005). Ilute: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, 5(2):217–234.
- Steimetz, S. S. and Brownstone, D. (2005). Estimating commuters’ value of time with noisy data: a multiple imputation approach. *Transportation Research Part B: Methodological*, 39(10):865–889.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.
- Timmermans, H. (2003). The saga of integrated land use-transport modeling: how many more dreams before we wake up? In *Keynote paper, Moving through nets: The Physical and social dimension of travel, 10th International Conference on Travel Behaviour Research, Lucerna, www.ivt.baug.ethz.ch/allgemein/pdf/timmermans.pdf*.
- Toomet, O. and Henningsen, M. A. (2013). Package maxlik.
- Waddell, P. (2009). Parcel-level microsimulation of land use and transportation: The walking scale of urban sustainability. In *12th International Conference on Travel Behavior Research*.
- 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.
- Walker, J. L. (2001). *Extended discrete choice models: integrated framework, flexible error structures, and latent variables*. PhD thesis, Massachusetts Institute of Technology.
- Wegener, M. (2004). Overview of land-use transport models. *Handbook of transport geography and spatial systems*, 5:127–146.
- Yavaş, A. (1992). A simple search and bargaining model of real estate markets. *Real estate economics*, 20(4):533–548.

Capítulo 2

Primer Artículo

A DYNAMIC BIDDING MODEL FOR MICROSIMULATION

Abstract

An increasing number of microsimulation tools for land use prediction has been developed over the last decades, most of them rely on an aggregated models in population, space and type of dwelling. Predicting prices of dwellings and location of households by considering equilibrium or dynamic equilibrium conditions for the city.

Agent based models has been developed for representing activities and travel decisions, but there are few detailed model of the behavior that represent the behavior of a search and bidding for a microsimulation land use model. Most of the microsimulation models still uses hedonic functions, without representing the outcome of the interaction of a seller and a buyer.

This work describes a dynamic bidding model that represents the behavior of a household buyer at agent level. The model incorporates the changes in expectation and dynamic adjustment made over the bargaining process to relocate. The models is developed to simulate a bargaining process on a microsimulation model, with the interaction of a dynamic selling model. The model is estimated by maximizing the likelihood of the observed transaction prices, estimations results shows a good fit for the model.

2.1. Introduction

Integrated land use and transportation models have been used to model and forecast the potential impact of policies or urban interventions such as regulation changes, real state development, modifications in the transport system among others, a task increasingly relevant in modern cities. These models represent the interaction between land use, transport and accessibility, where travel demand is a function of the spatial distribution of agents and activities. Modeling the location of the agents distributed across the space is one of the main challenges of this models.

The firsts location models were aggregated in the description and forecasting of a city, starting with the work of Alonso (1964), who developed a monocentric city model where price functions describe the willingness to pay of a household as a function of the distance to the CBD. With the development of the discrete choice theory by McFadden et al. (1978) two approaches emerges. The first approach, known as "choice" considers the household selecting the location by looking at the attributes of the dwelling including the price, and then selecting the one that maximizes his utility. Prices are estimated from a Walrasian equilibrium in the housing market (Anas, 1982) .

A second approach, called "bid-auction", is obtained considering that dwellings are traded in an auction market, and assigned to the best bidder, thus defining the final price of the dwelling (Ellickson, 1981). An extension to that model for the sake of parameter estimation

follows when observed transaction price is incorporated as information of the trade, introduced by Lerman and Kern (1983) allows the estimation of the prices directly from the model.

Several other location models are set in an aggregated basis, and assume market equilibrium. Models such as MEPLAN or TRANUS (Hunt and Simmonds, 1993), follows a choice-walrasian approach or MUSSA (Martinez, 1996)) follows a bid-auction approach. A complete revision on these models can be found in Wegener (2004). In recent models, the equilibrium assumption has been combined and partially replaced by microsimulation. See for example models as ILUTE (Salvini and Miller, 2005) or Urbansim Waddell et al. (2003) among other (see also Farooq et al. (2013), Chingcuanco and Miller (2013), Feldman et al. (2005)).

In equilibrium models agents are grouped into homogeneous clusters to make equilibrium problem treatable, but this is subject to criticism. Kirman (1992) describes the limitations in such models that employs a representative agent to model homogeneous groups, for example the expected reaction against a change in a policy, from a representative individual may not be the same reaction of the aggregated reactions of individuals inside that group. Also, representing the behavior of a group by one individual does not incorporate interaction among agents, and the sum of collective behavior may not be well represented.

Equilibrium aggregated models imposes limitations by the demanding assumption of complete information and perfect foresight. Capturing the effects of specific rules for an specific agent type, can be difficult or even unfeasible with aggregated equilibrium models. Finally the equilibrium approach fails to represent the behavior adjustment of individuals, and when multiple equilibrium exists, models usually fails to explain what led to the chosen equilibrium, for a more detailed discussion see Arthur (2006).

Microsimulation approach presents the advantage of allowing the models to work at a more detailed agent level, avoiding the aggregation problems and allowing the possibility of setting specific rules for a type of agent on specific situations. They also offer the capacity of capturing the decisions of individual agents and their interaction with others in the system, something in this level of detail impossible to accomplish with equilibrium models.

Most of microeconomic models based on search theory present a detailed model for the seller and the representation of the bidder usually is very simplistic. For example in the work of Albrecht et al. (2014), the seller behavior is very detailed, while the buyers bids are calculated as draws from a random distribution. Ettema (2011) presents a multi-agent model to micro-simulate the housing market. The work presented assumes one to one interaction, and the relocation process for the household is made by considering the expected utility. The buyer household behaves as a price taker without a bidding model that makes offers/counteroffers. In this approach the buyer household will be allocated if he accepts a list price, that is set by the seller.

Other group of microsimulation models are based on partial equilibrium and discrete choice, for example in Urbansim (Waddell et al., 2003) the location of agents is based on discrete choice theory and rents follows hedonic theory. Other works such as Farooq and Miller (2012) a microsimulation-disequilibrium approach where price formation market clearing, the price of dwelling is calculated to clear the market. These kind of models don't represent the bargaining on the market.

Few works have been developed to simulate the specific behavior of the buyers, for example in the work by Horowitz (1986) buyers submit one offer and the seller can of accept or reject. A more complete model is developed by Devisch et al. (2005), where buyers build bids considering the lifetime utility and a perceived utility of being accepted.

To study the behavior of the buyer, the asking prices and the information they gave to the seller has been incorporated in different models. For example Horowitz (1986) proposes to consider the individual decision of an agent, which incorporates the information of the asking price of the dwelling and the expectations of being the best bidder to calculate the bid price function. In most of search models (Salant, 1991, Merlo et al., 2013, Ettema, 2011), the asking price only has influence on the distributions of the bids. In a recent agent based micro-simulation model developed to work with ILUTE model (Miller et al., 2004), Rosenfield et al. (2013) incorporates the asking price serving two main proposes, narrowing down the set of bidders and acting as a benchmark for calculating the utility.

To develop a detailed representation of the buyer household following the bid-auction approach, Parker and Filatova (2008) present two open questions: How the standard equilibrium approach of Alonso (1964) can be translated to account for heterogeneous agent in spatial environment? and given a theoretical willingness to pay function, how the bid price are set for every agent? In this work we move towards answering those questions. We present a dynamic bidding model to be used in microsimulation agent based models, it considers the prediction of bids that a buyer is willing to submit in order to change its current location. The model is developed to an agent level, where each household has a different willingness to pay for every dwelling on the market. Our approach is similar to Horowitz (1986) model and incorporates some of the dynamic behavior proposed by Devisch et al. (2005), Parker and Filatova (2008). In this model the size of new bids is endogenously calculated by the buyer based on his believes about the market and the seller willingness to accept a price.

The model is derived in a generic formulation, considering the expected utility of the buyer and taking into account the current utility level, using the asking price as information of the likelihood it has to succeed with the bid. In our model, as in Ellickson (1981) and Lerman and Kern (1983) models, we focus on obtaining a joint probability of observing a transaction of a household h in a location l at a given price P_s : $\Pr(P_s, h, l)$. This probability consists on three events: the probability of the household h of selecting the dwelling l to make an offer $\Pr_{select}(h, l)$, the probability of the household h to submit a bid of size P_s : $\Pr_{bid}(P_s, h, l)$, and the probability of being accepted by the seller at price P_s , $Q(P_s)$. We put our focus on building the first two probabilities, leaving the behavior of the seller as an exogenous distribution.

An assumption for the derivation of the model is that the seller is willing to accept a bid of the size of the asking price. In this work we put our focus into the development of the bidding procedure, the available choice set of dwelling units is taken as exogenous. The proposed model differs from Horowitz (1986) formulation because it introduces dynamics into the problem, allowing the calculation of a sequence of bids to represent the response of a bidder in a bargaining.

The bidding model is estimated by using real transaction data for the city of Brussels. We employ a simple but realistic model to represent the behavior of the buyer. The estimation of

the parameters is made using maximum likelihood, to achieve this, assumption about the dynamic are made: 3 probabilities are calculated, the probability of the household of selecting the dwelling unit on the market, the probability of the household of a bid of size b and the probability of the seller to accept. The estimation shows that the models presents a good adjustment to the observed transaction price. The estimates of the model are compared against real data.

This work is structured as follows: In section 2.2 we introduce the classical formulation of the bid-rent models, specifically describing Ellickson and Lerman and Kern model, in section 3 the new model is presented. Section 4 describes the econometric model to estimate the parameters of the proposed model. Finally in section 5 an application using real data is presented, section 6 concludes.

2.2. Classical Formulation of the Bid Rent

To introduce the micro-economical formulation of the buyer problem, lets consider a quasi linear utility $u_h(x, z_l)$ of a household $h \in H$ located in a unit $l \in L$, where H is the set of all the household agents in the market, L is the set of available locations for the buyer, x represents a vector of continuous consumption goods and z_l is a set of attributes of the dwelling l . The quasi linear utility is linear in one continuous good x_0 .

The utility can be written as:

$$u_h(x, z_l) = \alpha_h x_0 + g_h(x_1, \dots, x_n, z_l) \quad (2.1)$$

the household faces an income constrain given by:

$$\sum_{k=0}^n p_k x_k + r_l \leq I_h \quad (2.2)$$

The income constrain imposes that the total amount spent in goods, exogenously priced at p_k for each of the k components of x , plus the price payed for the selected location must be smaller or equal to the household income I_h .

The household is assumed to maximize his utility by choosing an optimal x^* and a location l^* paying r_l^* .

Solving the optimization problem on the continuous goods, we obtain the conditional utility function $V_h(p, I_h - r_l, z_l) = u_h(x^*, z_l)$ that explicitly considers the value of the house selection l given a vector of characteristics z_l and a price for the location of r_l .

$$V_h(r, z_l) = \lambda_h(I_h - r_l) + g_h(x^*, z_l) \quad (2.3)$$

where λ_h is the Lagrange multiplier of the income constrain, and represents the marginal utility of the income.

Given a fixed utility level V_{hc} that represents the current utility level for a household, the value r_l is interpreted as the willingness to pay for the location l that yield a utility V_{hc} . Thus, the willingness to pay $w_h(z_l, V_{hc})$ of the household h for the location l , conditional on the utility level V_{hc} is given by:

$$w_h(z_l, V_{hc}) = I_h - \frac{V_{hc} - g_h(x^*, z_l)}{\lambda_h} \quad (2.4)$$

The willingness to pay is understood as the maximum rent or price a household can pay for a dwelling of attributes z_l conditional on attaining a utility level V_{hc} .

By this interpretation, if a household h pays $w_h(z_l)$ for any unit i he will obtain the same utility level than bidding $w_h(z_l)$ for other dwelling l , the utility level corresponds to V_{hc} represents and indifferent condition. In particular, if the household bids less than their willingness to pay he obtains a greater utility level. This fact is important for the design of a bargaining strategy of the buyer, because a bid of size b sets the utility level in $V_h(b, z_l)$ and $V_h(b, z_l) - V_{hc}$ defines a surplus; this property will be used in the derivation of the new model.

The econometric model developed by Ellickson (1981) (bid-auction) assumes an stochastic bid price function $w_{hl} = W_{hl}(z_l) + \varepsilon$, with $W_{hl}(z_l)$ the deterministic component of the bid function, representing the systematic part of the willingness to pay, that depends on the vector of attributes of the dwelling l . ε represents a stochastic term that takes into account differences in tastes among the household group.

In this setting, the probability of the dwelling l of being occupied by the household h is the probability household h being the best bidder for the dwelling:

$$\Pr(h, l) = \Pr[w_{hl} \geq w_{h'l} \quad \forall h' \in H] \quad (2.5)$$

where H is the set of all available households.

Ellickson's model assumes an independent and identically distributed Gumbel extreme value term ε for all the dwellings in the market, generating a Logit model from the equation (2.5). The probability of a household h to select a dwelling unit l to bid is always equal to one. Finally in this Logit model formulation, only relative estimates of the parameters can be obtained.

To overcome this limitation Lerman and Kern (1983) incorporates observed transaction prices to the probability of being the best bidder. The method proposed by Lerman and Kern is based on estimating the joint probability of a household being the best bidder for a particular location and of that the particular bid is equal to the observed transaction price.

Therefore, the probability of observing a sale price P_s for a dwelling l is the probability of the household h to be the best bidder with a bid of size P_s :

$$\Pr(P_s, l, h) = \Pr[w_{hl} = P_s \text{ and } W_{hl} + \varepsilon_h \geq W_{h'l} + \varepsilon_{h'} \quad \forall h' \in H] \quad (2.6)$$

Then a location l with attributes z_l is sold to a household h at a sale price P_s with probability given by:

$$\Pr(P_s, l, h) = f_\varepsilon(P_s - W_{hc}(z_l)) \prod_{r \in H, r \neq h} F_\varepsilon[P_s - W_{rc}(z_l)] \quad (2.7)$$

where H is the set of all households on the market.

In Lerman and Kern's model we can distinguish the probability of the buyer submitting a bid of size P_s given by $\Pr_{bid}(P_s, h, l) = f_\varepsilon(P_s - W_{hc}(z_l))$ and the probability of the seller accepting a bid corresponds to $Q(P_s) = \prod_{r \in H, r \neq h} F_\varepsilon[P_s - W_{rc}(z_l)]$. In Ellickson's and Lerman and Kern's models all households in the market bid for all available dwellings in the set S . The probability of observing a bid of the household h for a dwelling i is always equal to one ($\Pr_{select}(h, l) = 1$). This representation of the consumer behavior is unrealistic, because it is unlikely that a household h evaluate and bid all the dwelling units in the market; however it avoids representing a complex search process.

This model has been applied and reported in several studies for example: Gross (1988), Muto (2006), Hurtubia and Bierlaire (2014)

2.3. Proposed Model

We aim to generate a microsimulation model for a buyer who is actively looking for a dwelling unit in the housing market. Since we try to replicate a realistic behavior, assumptions such as the buyer household submitting bids for all the dwelling units on the market are omitted, instead we propose to generate a probability based on the expected surplus when bidding for all the units. Other assumptions has to do with the definition of the sale, for example how the seller accepts or rejects a bid.

2.3.1. Bidding Behavior

In the proposed dynamic model our first assumption is that the household h has an estimate of his probability of being accepted in dwelling l with a bid of size b^t given by $H_{hl}(b^t)$ at a time t . The length of the time interval t is assumed to be very short to mimic the dynamic of a bargaining process. Probability H_{hl} may depend on both dwelling and household attributes but to keep the notation simple, we omit the sub-indexes h and l .

The problem of a household is to select the size of the bid to submit in order to maximize his expected utility. Outcomes of the bidding process consists of winning with the bid offered in time t , or losing an bid again in time $t + 1$, or quit the market and maintain the current utility level.

Calling $E(V(b^t))$ to the expectation of the household's utility evaluating the unit with attributes z_l at the time t and $V(b, z_l)$ to the conditional utility described in equation (2.3) with V_c the current conditional utility level, we can write:

$$E(V(b^t)) = H(b^t)V(b^t, z_l) + (1 - H(b^t))\beta \max\{V(b^{t+1}, z_l); V_c\} \quad (2.8)$$

This formulation assumes a bidder who believes that if he fails bidding at time t , that is, the bid is not accepted by the seller, he will bid again the next period and expect to win a discounted utility $\beta V(b^{t+1}, z_l)$, or remain with the current utility V_c

To solve the preceding equation, we consider two cases. First let's consider that the current utility level is greater than the expected utility for the next period, then $V_c \geq V(b^{t+1}, z_l)$, and the equation (3.7) led us to the one period case which replicates the result of Horowitz (1986):

$$E(V(b^t)) = H(b^t)V(b^t, z_l) + (1 - H(b^t))V_c \quad (2.9)$$

The second case is when $V_c < V(b^{t+1}, z_l)$, that is, the buyer expects to increase his utility bidding again. Then equation (3.7) takes the following form:

$$E(V(b^t)) = H(b^t)V(b^t, z_l) + (1 - H(b^t))\beta V(b^{t+1}, z_l) \quad (2.10)$$

The problem of the household can be written as:

$$\max_b E(V(b^t)) \quad (2.11)$$

Calling b_{opt}^t the solution of the problem (2.11), taking the first order conditions and the quasi linear utility definition (2.3), we obtain the solution of the maximization problem.

Equation (2.9), can be written as:

$$E(V(b^t)) = H(b^t) (V(b^t, z_l) - V_c) + V_c \quad (2.12)$$

Taking the first order condition ¹ $E(V(b^t))' = 0$:

$$0 = H(b^t)' (V(b^t, z_l) - V_c) + H(b^t)V(b^t, z_l)' \quad (2.13)$$

By using equation (2.3):

$$V(b^t, z_l)' = -\lambda \quad (2.14)$$

And from equation (2.4), since the household bid can be interpreted as the willingness to pay, conditional on the utility $V(b^t, z_l)$:

$$b^t = w(z_l, V(b^t, z_l)) = I - \frac{V(b^t, z_l)}{\lambda} - \frac{g(x^*, z_l)}{\lambda} \quad (2.15)$$

and at the current utility level:

$$w(z_l, V_c) = I - \frac{V_c}{\lambda} - \frac{g(x^*, z_l)}{\lambda} \quad (2.16)$$

¹Where $H'(b)$, $E(V(b^t))'$, $V(b^t, z_l)'$ stands for the derivative of $H(b)$, $E(V(b^t))$, $V(b^t, z_l)$ respectively

Then:

$$V(b^t, z_l) - V_c = -\lambda(b^t - w(z_l, V_c)) \quad (2.17)$$

Replacing (2.14) and (2.17) in (2.13) we obtain:

$$0 = -H'(b^t)\lambda(b^t - w(z_l, V_c)) - \lambda H(b^t) \quad (2.18)$$

then, given that $\lambda > 0$ is the marginal income utility, we obtain b^t :

$$b^t = w(z_l, V_c) - \frac{H(b_{opt}^t)}{H'(b_{opt}^t)} \quad (2.19)$$

This result is equivalent to the obtained by Horowitz (1986). The value $w(z_l, V_c)$ corresponds to the willingness to pay to achieve the utility level V_c on a dwelling l , and represent the amount to bid to achieve the same utility as in the current utility.

A household is indifferent in winning an auction with a bid of size $w(z_l, V_c)$ or staying in the current location.

For the second case, $V_c < V(b^{t+1}, z_l)$, the utilities are derived from the application of the first order condition:

$$b_{opt}^t = \frac{g^t(x^*, z_l) - \beta g^{t+1}(x^*, z_l)}{\lambda} + (I_t - \beta I_{t+1}) + \beta b_{opt}^{t+1} - \frac{H(b_{opt}^t)}{H'(b_{opt}^t)} \quad (2.20)$$

Here assuming that the bargain duration is short enough to have the income level I and tastes constants that conforms in the equation g ; A period t is for example a couple of weeks. Also notice that β , the discount factor, will be close and assumed equal to 1. Finally we obtain :

$$b_{opt}^t = \begin{cases} b_{opt}^{t+1} - \frac{H(b_{opt}^t)}{H'(b_{opt}^t)} & \text{if } u(b_{opt}^{t+1}, z_l) > u_c \\ w_c - \frac{H(b_{opt}^t)}{H'(b_{opt}^t)} & \text{otherwise} \end{cases} \quad (2.21)$$

Equation (2.21) defines the behavior of a buyer when facing the problem of determining the optimal size of a bid. This model defines an increasing sequence of bids for a particular household, because $H(b)$ takes positive values and $H(b)^'$ is positive since a increase in the bid size increases the expected probability of winning. Therefore, the bid at time $t + 1$ is always greater than the bid on t .

An opposite effect of increasing bids, expected utility decreases (see equation 2.3), down to the lower bound V_c , the current utility.

If buyer defines a maximum number of periods for the bargaining T , we can rewrite (2.21) as following:

$$b_{opt}^t = w_c(z) - (T + 1 - k) \frac{H(b_{opt}^t)}{H'(b_{opt}^t)} \quad k = 1 \dots T \quad (2.22)$$

with k the bargain step of number of counter offers.

2.3.2. Probability of the household h to submit a bid of size P_s

As in Ellickson model, we introduce an stochastic term ε associated with the unobserved part of the willingness to pay function, given in equation (2.4), where ε is a random variable with mean 0 and variance σ^2 .

We consider that the error ε distributes with density function g , so we can write the probability of observing the bid at time t being equal to the sale price P_s :

$$\Pr_{bid}^{[b^t = P_s]} = g \left(P_s - W_{hc}(z) + (T + 1 - k) \frac{H_h(P_s)}{H'_h(P_s)} \right) \quad k = 1 \dots T \quad (2.23)$$

Depending on the probability function H_{hl} selected, the problem may define a fixed point for each bid.

So far, this bidding model can be used to estimate the size of the optimal bid on sequence of offers with fixed number of steps. The formulation is written in a general way, allowing flexibility by incorporating assumptions in the functional form of $H_{hl}()$, subject to the assumption of the cuasilinear utility.

As it was discussed in the previous section, asking prices gives information to the buyer about the seller's reservation price. This information feeds the subjective perception of the household's likelihood of being the winner. The model has been so far formulated considering a generic value for H , so in the next section we develop the model for an explicit form of the probability, that constrains the bids to be at most the asking price.

2.3.3. Probability of Selecting the Unit

In the proposed model the buyer household will select a single unit to submit a bid and enter in a bargain process, the selected unit will be the available dwelling that maximizes his expected surplus. To find the appropriate dwelling unit, the household may evaluate all the dwelling units available on the market to make a selection. Since that is unlikely to happen in real life because of the very large cost of collecting information in the large number of units, we assume that expected surplus is a consequence of the attributes of the dwelling units in the market, and those attributes are randomly distributed among the available houses, defining a probability of selecting the household of maximum surplus. A similar approach can be found in Horowitz (1986) and Carrillo (2012)

Calling the expected surplus $S_{hl}(b^t)$, the distribution of the probability of selecting the maximum will follow an extreme value distribution. Let $\Pr_{select}(h, l, b^t)$ be the probability of

a household h of selecting dwelling $l \in L$ with a bid of size b^t , where L represents all the units currently on the market, with cardinal of L large, then.

$$\Pr_{select}(h, l, b^t) = \Pr(\max_{k \in L} S_{hk} \leq S_{hl}) \quad (2.24)$$

The expected surplus is calculated as the difference between expected utility at the location and current utility level V_{hc} of a household h given a bid b^t , assumed known by the household. Then:

$$S_{hl}(b^t) = E(V_h(b^t)) - V_{hc} \quad (2.25)$$

From (2.12) we can calculate $E(V_h(b_l^t)) - V_{hc}$,

$$E(V_h(b_l^t)) - V_{hc} = H_h(b_l^t)(V_h(b_l^t, z_l) - V_{hc}) \quad (2.26)$$

and since from (2.17) we know the value $V_h(b_l^t, z_l) - V_{hc} = -\lambda_h(b^t - w_h(z_l, V_c))$, we obtain:

$$E(V_h(b_l^t)) - V_{hc} = \lambda_h H_h(b_l^t)(w_h(z_l, V_c) - b^t) \quad (2.27)$$

this equation shows that the expected surplus is the difference between the willingness to pay and the bid submitted by the perceived probability of winning, the λ translates the monetary term into utility.

Following our extreme value distribution assumption:

$$\Pr_{select}(h, l, b^t) = \exp\left(-\exp\left(-\frac{S_{hl}(b^t) - a_h}{b_h}\right)\right) \quad (2.28)$$

Where a_h and b_h are parameters of the Gumbel distribution to be estimated, which are specific for each household; a_h is the location parameter representing the maximum expected surplus for the set L and b_h is inversely proportional to the variance.

Using (2.27) and (2.28):

$$\Pr_{select}(h, l, b^t) = \exp\left(-\exp\left(-\frac{\lambda_h H_h(b_l^t)(w_h(z_l, V_c) - b^t) - a_h}{b_h}\right)\right) \quad (2.29)$$

Corresponds to the probability of a household h selects a dwelling unit i , conditional on the bid of size b^t . In the case of the observed transaction price P_s the expression (2.29) takes the following form:

$$\Pr_{select}(h, l, P_s) = \exp\left(-\exp\left(-\frac{\lambda_h H_h(P_s)(w_h(z_l, V_c) - P_s) - a_h}{b_h}\right)\right) \quad (2.30)$$

2.3.4. Probability of Being Accepted

In the proposed model, the behavior of the seller is not modeled. It has a probability of accepting the bid b . For the offer to be accepted, the bid value should be greater than the reservation price of the seller. We assume that there is a certain relation between this reservation price and the asking price proposed. Then, the probability of being accepted at a price P is going to be called $Q(P)$ and it will be an implicit function of the asking price P_a . $Q(b|P_a)$, where $b \geq 0$ and $Q(b|P_a) = 1 \quad \forall b \geq P_a$

Finally the proposed model is formulated as it follows as the interaction of the probabilities of selecting the unit, offering a bid equal to the observed transaction price P_s and being accepted by the offer

2.4. Likelihood Estimation

To develop an estimable econometric model, we define the probability of observing a sale in terms of the events likely to happen to observe the transaction. Our approach to estimate the parameters of the dynamic model is the use of maximum likelihood method.

We use the three probabilities described in the model for the transaction to be observed, the first one is the probability of the household has of selecting the dwelling to bid P_{select} , this probability is defined in terms of the expected surplus that the buyer thinks could get (see equation (2.29)). A second probability is given by the probability of observing a bid with size equal to the transaction price P_{bid} , similar to in Lerman and Kern's model, but in this case bids follows a dynamic bargaining process. Finally there is a third probability corresponding to the likelihood of being accepted by the seller Q . That last probability represents the behavior of the seller, who is not modeled in this work. Since the probabilities has been defined in previous section, the derivation of the joint probability and the likelihood is straightforward.

The probability of observing a transaction depends on the number of attempts made by the household; On each attempt the seller will reject or accept the offered bid. For example, if the bids are b_1 , b_2 and b_3 for attempts from 1 to 3, the probability of accepting a bid of size one is given by $Q(b_1)$ and the probability of rejecting the bid is given by $1 - Q(b_1)$, to accept a second bid b_2 it has to reject bid b_1 and accept bid b_2 then the probability is given by $Q(b_2)(1 - Q(b_1))$, the probability of accepting the bid of size b_3 is $Q(b_3)(1 - Q(b_2))(1 - Q(b_1))$. In this case, since the transaction price is observed, but the number of attempts is unknown, $b_1 = P_s$ if there is one bid, $b_2 = P_s$ if there are two offers, and $b_3 = P_s$ if there are three offers. Then the probability of having up to three offers is given by : $Q(P_s)$ for one offer, $(1 - Q(P_s))Q(P_s)$ for two offers, and $(1 - Q(P_s))^2Q(P_s)$ for three offers. With that in mind, the probability of observing a transaction at price P_s for the dwelling l is given by:

$$\Pr(P_s, l, h) = \sum_{k=1}^T \prod_{j=1}^{k-1} \Pr_{select}(h, l, P_s)(1 - Q_l(P_s))^j Q_l(P_s) \Pr_{bid}(P_s, h, l) \quad (2.31)$$

Substituting the probabilities with the values we obtain:

$$\Pr(P_s, l, h) = \Pr_{select}(h, l, P_s) \sum_{k=1}^T \prod_{j=1}^{k-1} (1 - Q_l(P_s))^j Q_l(P_s) g \left(P_s - W_{hc}(z) + (T + 1 - k) \frac{H_h(P_s)}{H'_h(P_s)} \right) \quad (2.32)$$

Where

$$\Pr_{select}(h, l, P_s) = e^{-e^{-\frac{\lambda_h H_h(P_s)(w_h(z_l, V_c) - P_s) - a_h}{b_h}}}$$

. Then the likelihood of the model could be written as:

$$L = \prod_{l \in L} \prod_{h \in H} \Pr(P_s, h, l)^{y_{hl}} \quad (2.33)$$

where y_{hl} indicates if the household h is observed located in dwelling l . The model estimates the sets of parameters of $\Pr(P_s, l, h)$ that maximizes the likelihood of reproducing the set of observations $Y = \{y_{hl}\}_{l \in S, h \in H}$

2.5. Numerical Estimation

2.5.1. Data Description

The model is estimated for the residential market of the city of Brussels. Data contains information of the current household and dwelling, as zonal and commune attributes, with 4945 zones and 151 communes. The dwelling alternatives are divided into 4 types: detached, semi-detached, back-to-back houses and apartments.

Data adds to a total of 1,274,701 residential units or location alternatives, characterized by their average physical and land use attributes, by type of dwelling and zone, which are calculated from the Census and the Land Registry of Belgium. The area of study contains a total of 1,267,998 households, therefore having an aggregate vacancy rate (supply surplus) of 0.5 percent.

Price data is available as the average by commune and for a simplified classification of dwelling types that aggregates them into houses and apartments. The database does not contain information about asking prices, this is modeled by assuming it as a latent variable; this process is explained in the next section.

2.5.2. Asking Prices

Since the database does not contain information about the asking prices, we perform our estimations by considering a latent variable approach. Latent variables are employed in behavioral sciences to incorporate into the model the effect of phenomena that cannot be directly measured. The phenomena are usually called latent constructs, based on the hypothesis that the effects of such latent constructs are measurable by using variables called indicators; See Walker (2001) for a complete description of the latent variables. One application of this latent variable approach to location choice models can be seen in the work of Hurtubia and Bierlaire (2014), where a latent approach of the transaction price is employed to estimate a Lerman and Kern type of model.

In this case, the asking price is modeled as a latent construct where the indicators are attributes of the dwelling l . The distribution of the latent variables defined by a structural model $f(P_a|Z)$ where Z stands for the vector of attributes of the zone and dwelling. In this case the latent asking price takes the following functional form:

$$P_{al} = h(Z_l) + \omega_l \quad (2.34)$$

with $h(Z_l)$ a linear “hedonic” function.

Since for the willingness to pay function the attributes are depending both on the household and on the attributes of the dwellings, the variables of the asking price indicator function are not correlated with bids.

The idea behind the model is to consider the probability of observing a transaction price as the following probability conditional on the asking price (P_a).

$$\Pr(P_s, h, l) = \int \Pr(P_s, h, l | P_{al} = h(Z_i) + \omega_l) f(\omega_l) d\omega_l \quad (2.35)$$

For the indicator function $h(Z_i)$ we use a linear form function with a random error ω_l independent and identically distributed for each dwelling unit, normally distributed with mean zero and variance σ_1^2 . This independence is achieved since our probability of selection does not depend on other dwelling unit’s surplus.

2.5.3. Model Specification

Willingness to pay

The willingness to pay function is specified considering 3 constants: one for each group of income: low, medium and medium high. To take into account the difference of the willingness to pay for space when the size of the household varies, floor space of the dwelling is interacted with dummy variables for households of size 1, 2 and more than 3 persons. To capture

segregation effects, the percentage of households of level low in a zone is interacted with a dummy variable for incomes high, and to capture positive externalities of the high incomes households, the percentage of households of high income in a zone interacting with a dummy variable for medium income group. Finally accessibility measures of car and public transport are interacted with a dummy variable indicating the car availability, to capture attractiveness of the different transportation modes.

The willingness to pay function used for the proof of concept consists on the interaction of houses and dwelling attributes, the variables estimated are the following:

Table 2.1: Willingnes to pay variables

Name	Household attribute	Dwelling Attribute
ASC1, ASC2, ASC3	Constant for incomes, low, - medium, high	-
bfloor1,2,3+	Dummy if household is of size 1,2,3 or more	floor space of the dwelling
blowinc	Dummy: mid-high income	% of low Income households in the zone
bhighinc	Dummy: mid-low income	% of high income households in the zone
bcar	Dummy: Has car	Car accessibility
btrans	Dummy: Not has Car	Public transport accessibil- ity

The dummies variables: mid-high income and mid-low income take the value of one when the household belongs to the respective income groups. The variable *bfloor1*, *bfloor2* and *bfloor3* interacts with dummies taking value of one when the household size is one, two or three respectively and interacts with the floor space of the dwelling unit. The dummies has car and not has car indicates if the household has access to car, these variables are interacted with car accessibility which is calculated with the usage of the MATSim, and public transport accessibility which is calculated as the number of facilities in the commune per square kilometer.

Latent Price estimation

We assumed a linear function on the dwelling attributes for the asking price function (2.34), following (Rosen, 1974). The specification was made by considering different combinations of attributes available in the dataset, the one that reported best results were the following: *floor5* is a variable multiplying the floor space of the dwelling interacting with the percentage of high income households in the zone, capturing higher values of the floor space in zones where rich people lives. *floor1* interacts the size of the dwelling with the percentage of people of low income in the zone. A last variable *DD_shop_c* is the density of shopping places in the commune, this parameter captures the valuation of shopping activities for the selected dwelling.

Distributions Specifications

To have an estimable econometric model for the probability of the household $H_h()$ we assume a logistic distribution truncated on the asking price P_a for $H(b)$, therefore $H(b) = 1$ for all $b \geq P_a$. This distribution yields a treatable model. This distribution generates a fixed point problem (see (3.8)) for each household on a specific dwelling.

$$H_h(b) = \frac{\int_a^b g(u)du}{F(b) - F(a)} \quad (2.36)$$

with :

$$g_{\nu_1}(u) = \begin{cases} \frac{e^{-\frac{u-P_a}{\nu_1}}}{\nu_1 \left(1 + e^{-\frac{u-P_a}{\nu_1}}\right)^2} & \text{if } 0 \leq u \leq P_a \\ 0 & \text{else} \end{cases} \quad (2.37)$$

where $F(x)$ and $g(x)$ corresponds to a Logistic cumulative distribution with scale parameter P_a

$$F(x)_{\nu_1} = \frac{1}{1 + e^{-\frac{x-P_a}{\nu_1}}} \quad (2.38)$$

the scale parameter s is related to the variance of the distribution given by $\nu_1^2 \pi^2 / 3$

For the calculation of the probability of being accepted $Q(P)$ we also assume a logistic distribution truncated on the asking price P_a with location parameter equal to the asking price and scale parameter ν_2 to make the numerical estimation.

$$Q_i(b) = \frac{\int_a^b g_{\nu_2}(u)du}{F_{\nu_2}(b) - F_{\nu_2}(a)} \quad (2.39)$$

For the stochastic term of the willingness to pay function ε we assume a normal distributed error term with mean equal to zero, and variance σ_2^2 . Therefore the equation (2.23) takes the following form:

$$\Pr_{bid}^{[b^t = P_s, k]} = \phi \left(P_s - W_c(z) + (T + 1 - k) \frac{\int_0^{b^t} g(u)du}{g(b^t)} \right) \quad k = 1 \dots T \quad (2.40)$$

The estimation was made in the software R and the maxLik package (Toomet and Henningsen, 2013), the maximization of the likelihood was made employing the BHHH algorithm, on each iteration the a numerical integration for each observation was made. Results are shown in next section.

2.6. Results

The estimation results of the model are found in Table 2.3. Parameters on the willingness to pay and latent price specification gives statistically significant parameters and signs of the

parameters are the expected. The parameters of the distribution employed also gave good results, σ_1 corresponds to the variance of the normally distributed error term for the bid function. a and b parameters correspond to the parameters of the probability of selecting the location to submit a bid, for estimation proposes only one parameter for a and b was estimated, although they depend on each individual household h believes. ν_1 corresponds to the scale parameter of the truncated logistic distribution representing the probability H , ν_2 represents the scale parameter of the logistic distribution of accepting a bid Q , both parameters are related with the variance of the distribution.

Values for the constants $ASC1$, $ASC2$, $ASC3$ shows a “u-shape” , this non monotonic effect could be explained because of the specification of the willingness to pay function in which the effect of low income percentage and high income percentage is not included as base reference.

The variables $bfloor1$, $bfloor2$ and $bfloor3$ shows similar values, therefore the effect of the size of household varies little. $blow_inc$ shows a negative value, indicating a negative effect of the percentage of low income households in the zone, an opposite effect is shown by the variable $bhigh_inc$, showing a larger and positive value that indicates the possitive effect of the precense of high income households in the area. The $baccess$ and $btrans0$ variables present signs as expected.

The results obtained for the variables $lfloor1$ and $lfloor5$ of the asking price latent variable show that the valuation of floor space is larger in zones with higher percentage of households of high income.

Since the value of T was not estimated in the model, we estimated with values of $T = 1, 2, 3$ and 4

The results that gave the lowest likelihood was obtained with $T = 2$ (see Table 2.6), that is, up to one counteroffer from the bidder. We notice that in case that $T = 1$ the model takes a form very similar to Horowitz model, since only allows one bid to be submitted.

Table 2.2: Log-Likelihood vs number of interactions T

T value	Log-Likelihood
1	-4161.848
2	-3819.778
3	-3834.722
4	-4153.334

To compare results with classical Lerman and Kern’s model we notice that the same willingness to pay function, results are in Table 2.4. The log-likelihood obtained shows that the new model adjust better to the data.

Table 2.3: Results of the New Model with T=2 (Log-Likelihood: -3819.78)

Parameter	Estimate	Std. Error	t value
σ_1	1.833263	0.03549	51.6564
ASC1	3.176146	0.40344	7.8727
ASC2	2.474697	0.360887	6.8573
ASC3	5.669976	0.85248	6.6512
bfloor1	0.664801	0.034819	19.0931
bfloor2	0.652699	0.031778	20.5391
bfloor3	0.644491	0.030072	21.4314
blow_inc	-3.924604	1.382709	-2.8383
bhigh_inc	11.812181	1.318284	8.9603
baccess	0.336727	0.035916	9.3754
btrans0	2.559115	0.323567	7.9091
ν_1	0.181099	0.108439	1.6701
ν_2	9.423204	0.392472	24.0099
a	-5.528982	0.420153	-13.1594
b	14.158511	1.264808	11.1942
lfloor5	22.468655	1.035373	21.701
lfloor1	13.341703	0.509394	26.1913
lshop	5.845362	1.248552	4.6817
σ_2	5.323914	0.344777	15.4416

Table 2.4: Results of estimation with Lerman and Kern Model (Log-Likelihood: -10032.96)

Parameter	Estimate	Std. Error	t value
μ	0.600412	0.015881	37.8061
ASC1	-6.342477	0.393705	-16.1097
ASC2	-7.265766	0.367743	-19.7577
ASC3	-2.496591	0.748946	-3.3335
bfloor1	0.661242	0.020057	32.9682
bfloor2	0.669068	0.019980	33.4864
bfloor3	0.627053	0.022900	27.3828
blow_inc	-7.567690	1.304356	-5.8019
bhigh_inc	9.197043	0.750224	12.2591
baccess	0.321170	0.035547	9.0351
btrans0	2.467306	0.320836	7.6902

2.7. Conclusions

In this work, we developed a bidding model that represents the behavior of a household buyer at individual level. The calculation of the buyer's behavior on dynamically setting bids is endogenous to the model and depends on the expectation of the household. This model has economic support on buyer's behavior as maximizes his expectation considering the possibility of bidding again if he loses. The model is presented in a generic way, and it is flexible to incorporate different behavioral assumptions by selecting the specification of the perceived probability of whether the bid will be accepted H . In the model, the role of the asking price is incorporated in the household's perceived probability of winning the property.

A maximum likelihood approach was employed for the estimation of the model. The dynamic of the bargaining process is captured by using a probability function to describe the behavior of the seller. To overcome the absence of the asking price in the data, a latent variable approach was incorporated. The assumption of i.i.d. normally distributed error term allows the estimation by using traditional numeric methods for integration, methods which are available in most of estimation software.

The selection of a logistic function for the probability H , resembles the shape of the expectation probability suggested by Devisch et al. (2005). The selection of a truncated logistic function generates a fixed point problem, to be solved for every household at every stage t to determine the bid's size (see equation (3.8)). Other forms of H could avoid the fixed point problem, for example $H(b) = \exp((b - P_a)/\gamma)$, whit γ a constant, which corresponds to Horowitz probability. The convergence of the fixed point problem remain to be analyzed to characterize zones where the convergence is assured.

We used the size of the dwelling unit both for the willingness to pay and the latent price. To avoid correlation of parameters we employed interacting variables, in one case interacting the variable with the size of the household and in the other with the percentage of households of different income groups in the communes.

The log-likelihood difference between the proposed model and Lerman and Kern indicates that the latter model does not replicate to the data as well as the proposed model, similar results where found in Horowitz work. In Lerman and Kern's model estimation, a similar "u-shape" form for the constants was found, floor area and accessibility estimated parameters have similar magnitude than in the new model estimation.

The proposed model extends previous work by Ellickson and Lerman and Kern's model in line with Horowitz agent-based approach, and extends the latter by introducing a dynamic process of a number of counter offers, in doing that we assume that number ex-ante. i.e. the buyer decides how many offers will be ready to submit before quitting.

The model could be improved by incorporating the buyer's learning effect on the probability. Indeed, notice than in the estimated model, H function was considered constant for all the

periods, but this simplification of the model that allows a feasible estimation can be modified by considering for example a Bayesian update of the perception; as for example in Ettema (2011), Curlette (2006).

Bibliography

- Albrecht, J., Gautier, P. A., and Vroman, S. (2014). Directed search in the housing market. *Econometrica*, 82(2):705–30.
- Alonso, W. (1964). Location and land use. toward a general theory of land rent. *Location and land use. Toward a general theory of land rent*.
- Anas, A. (1982). Residential location markets and urban transportation: Economic theory, econometrics, and policy analysis with discrete choice models.
- Arthur, W. B. (2006). Out-of-equilibrium economics and agent-based modeling. *Handbook of computational economics*, 2:1551–1564.
- Carrillo, P. E. (2012). An empirical stationary equilibrium search model of the housing market*. *International Economic Review*, 53(1):203–234.
- Chingcuanco, F. and Miller, E. J. (2013). Demographic microsimulation model for integrated land use, transportation, and environment model system. In *Transportation Research Board 92nd Annual Meeting*, number 13-3445.
- Curlette, W. L. (2006). A framework for research studies: Mixed methods through combining bayesian statistics and qualitative research in individual psychology. *The Journal of Individual Psychology*.
- Devisch, O., Arentze, T., Borgers, A., and Timmermans, H. (2005). An agent-based model of residential choice dynamics in non-stationary housing markets. In *Paper CUPUM Conference, London*.
- Ellickson, B. (1981). An alternative test of the hedonic theory of housing markets. *Journal of Urban Economics*, 9(1):56–79.
- Ettema, D. (2011). A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. *Computers, environment and urban systems*, 35(1):1–11.
- Farooq, B. and Miller, E. J. (2012). Towards 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.
- Farooq, B., Miller, E. J., Chingcuanco, F., and Giroux-Cook, M. (2013). Microsimulation framework for urban price-taker markets. *Journal of Transport and Land Use*, 6(1):41–51.

- Feldman, O., Simmonds, D., Ballas, D., Clarke, G., Gibson, P., Jin, J., and Stillwell, J. (2005). A spatial microsimulation approach to land-use modelling. In *CUPUM conference*, volume 29.
- 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.
- Horowitz, J. L. (1986). Bidding models of housing markets. *Journal of urban economics*, 20(2):168–190.
- Hunt, J. D. and Simmonds, D. C. (1993). Theory and application of an integrated land use and transport modelling framework. *Environment and Planning B*, 20:221–221.
- Hurtubia, R. and Bierlaire, M. (2014). Estimation of bid functions for location choice and price modeling with a latent variable approach. *Networks and Spatial Economics*, 14(1):47–65.
- Kirman, A. P. (1992). Whom or what does the representative individual represent? *The Journal of Economic Perspectives*, pages 117–136.
- 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.
- Martinez, F. (1996). Mussa: land use model for santiago city. *Transportation Research Record: Journal of the Transportation Research Board*, 1552(1):126–134.
- McFadden, D. et al. (1978). *Modelling the choice of residential location*. Institute of Transportation Studies, University of California.
- Merlo, A., Ortalo-Magné, F., and Rust, J. (2013). The home selling problem: Theory and evidence. Technical report, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- Miller, E. J., Douglas Hunt, J., Abraham, J. E., and Salvini, P. A. (2004). Microsimulating urban systems. *Computers, Environment and Urban Systems*, 28(1):9–44.
- Muto, S. (2006). Estimation of the bid rent function with the usage decision model. *Journal of Urban Economics*, 60(1):33 – 49.
- Parker, D. C. and Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems*, 32(6):454–463.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, pages 34–55.
- Rosenfield, A., Chingcuanco, F., and Miller, E. J. (2013). Agent-based housing market microsimulation for integrated land use, transportation, environment model system. *Procedia Computer Science*, 19:841–846.

- Salant, S. W. (1991). For sale by owner: When to use a broker and how to price the house. *The Journal of Real Estate Finance and Economics*, 4(2):157–173.
- Salvini, P. and Miller, E. J. (2005). Ilute: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, 5(2):217–234.
- Toomet, O. and Henningsen, M. A. (2013). Package maxlik.
- 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.
- Walker, J. L. (2001). *Extended discrete choice models: integrated framework, flexible error structures, and latent variables*. PhD thesis, Massachusetts Institute of Technology.
- Wegener, M. (2004). Overview of land-use transport models. *Handbook of transport geography and spatial systems*, 5:127–146.

Capítulo 3

Segundo Artículo

MODEL FOR MICROSIMULATION

Abstract

In this work develop a microsimulation model of the housing market, where the interaction between a seller and a buyer is simulated based on the expectations mutually induced in the bargaining procedure. The seller generates expectations for the property and publishes an asking price for the property. The bidder observes the attributes of the dwelling and the asking price and decides whether to submit a bid considering the likelihood of winning. This entails a bargaining dynamic with endogenously changing asking prices and bids. Our proposed framework is designed to be implemented in the SimMobility Long Term simulation software of Singapore.

3.1. Introduction

Land use model are traditionally applied for forecasting future changes in land use, socioeconomic and demographics of the city, which is performed by models usually based on economic theories and social behavior assumptions. Their use is relevant since allows the decision maker to forecast the effects of changes in the city, anticipating potential effects of public policies or infrastructure investments, such as land use changes that impact on transportation demand.

Historically the development of large and computable land use models started with the earlier work of Lowry 1964, based on the gravity theory; followed by spatial input-output model a, such as PECAS, all of this based on equilibrium assumptions. With the arise of the discrete choice theory of McFadden et al. (1978) models started to rely on the multinomial logit model to forecast residential choice, firms' location choice and the various destination choices underlying trip behavior, as for example the Santiago Land Use model MUSSA from Martinez (1996). Traditionally, land use changes and urban development have been modelled using aggregate models, based on zonal information (Timmermans, 2003). With the increase of computational power, simulation capabilities and behavioral theory, microsimulation models have become more prominent in land use models. (Wegener, 2004)

Location choice is one of the key process of the land use models. The capability to determine the prices of housing market and the location of agents is usually described by some market clearing procedure. In aggregate models this is involves solving an equilibrium problem, where all agents locate in dwellings until there is no incentive to change its current location. Finding an equilibrium requires strong assumptions such as perfect information and perfect match between demand and supply. Disequilibrium or quasi-equilibrium models has been proposed to soften this requirements, but they don't represent the interaction of the agents in a proper an agent-based way. See for example Martínez and Donoso (2011).

The usage of microsimuation agent-based models allows the representation of effects that

could not be observed on equilibrium models. But at this level, the representation of the population requires to describe the household's decision to select and bid for dwellings, and the seller decision to trade the dwelling.

The market clearing problem has been less studied in the case of agent-based microsimulation models, and it's either ignored or solved at an individual level assuming that agents behave as price takers, ignoring market effects or simulating several iterations of individual auctions. Parker and Filatova (2008) highlights the need of proper representation of the bidders and sellers, based on both side expectations and perceptions of the market, and price cannot be interpreted as indicator of current demand level for housing.

In this work we present a framework to describe the microeconomic behavior of a seller and a buyer in the housing market. Our goal is to define two models of agent's behavior, one for the seller and one for the buyer, and integrate them to represent the interaction of them in a agent-based microsimulation compatible way. For representing the buyer's behavior, we took a bidding model developed by Rocco et al. (2014)¹ to incorporate a dynamic bidding behaviour that considers expectations (based on attributes and the asking price of the dwelling), as well as the utility on his current location and the perceived probability of being accepted by the seller. The representation of the seller behavior considers a search strategy, where the seller defines in each period of time an asking price and a reservation price. The seller's behavior takes into account the belief of the market and the influence of the asking price in the attractiveness of bidders to the dwelling. The transaction will happen if exist a bid larger than the reservation price of the seller.

Interaction between seller and buyer is not explicitly modeled through a two sided bargaining process, instead the bargaining is implicit in the behavior of buyers who can submit counter offers to a seller unit. In the same way sellers can accept, reject or accumulate bids in a period of time, and decide whether to stay or withdraw the housing unit from the market. In our proposed model, the information common to both agents, are housing attributes given by a vector z_i , the size of the bid b and the asking price of the property P_a .

The model is currently being implemented in the SimMobility software, a simulation platform with an integrated model of human and commercial activities, land use, transportation, environmental impacts, and energy use. First numerical simulation are currently being carried out.

This work is structured in the following way: next section describes shortly the usage of other microsimulation models and software currently available; section 3.3 describes the proposed framework and the interaction between the seller and the buyer, describing also the equations that rules the behavior of both selling and buyer agent. Section 3.3 concludes.

¹ See this in chapter 2

3.2. Microsimulation models: brief literature review

To describe the behavior of the seller we could highlight the classical optimal stopping strategy. In the classical optimal search approach of the "house selling problem", a sequence of bids unknown to the seller $\{b_1, b_2, \dots\}$ arrives, the seller should pay a price $c > 0$ every time he observes the next bid. The problem of the seller is to decide an optimal rule to stop. One solution is found by computing the expected value V^* and if any bid larger than V^* appears, then the seller should take the offer. The number of bids observed before trading is:

$$N = \min\{n \geq 1 : b_n \geq V^*\} \quad (3.1)$$

If bids arrive following a known distribution F , the optimal value can be found solving:

$$V^* = \int_{-\infty}^V V^* dF(b) + \int_{V^*}^{\infty} bF(b) dF(b) - c \quad (3.2)$$

Feinberg and Johnson (1977) uses this approach to study market behavior under assumptions of functional distribution assumptions. In the same line, a search model is proposed by Haurin (1988), where its shown that an atypical dwelling will have larger time on the market. Salant (1991), formulates the problem of the seller by using dynamic programming, which involves an initial choice of whether to use a real state agent or to sell themselves. Under this assumptions the seller must choose a list price for each period.

Another search model to describe the behavior of the seller is done by Horowitz (1992), which performs an estimation of an asking price model. In this work he considers an stationary search framework where the seller should set the reservation and the list price to maximize the expected outcome of the sale. In this stationary equilibrium model the seller enters the market, fixes an asking price and informs all the buyers that the home is for sale, and the characteristics of the house. The asking price constitutes a ceiling for the offers of the buyers and also constitutes an obligation, so if a buyer offers the asking price, the unit must be sold at that price. The seller problems consists on finding a reservation price and a asking price. To do so, they maximize the expected gain from the trade. In this model the rate of at which buyers visits a particular seller depends on the asking price. The idea of the seller selecting an asking price to attract bidders could be found in several models. For example in Salant (1991) and in Merlo et al. (2008).

Models that consider multi-period adjustment for the seller usually consist on variations of the optimal stopping rule problem. In these models the reservation price and asking price is determined endogenously by the seller. For example, Merlo et al. (2013) develops a model where the seller can exit the market and decline from selling the dwelling unit. To do that he writes a Bellman equation of the problem:

$$V_t(P_t, d_t) = \max\{W_t(P_t, \beta); \max_p [u_t(p, p_t, d_t) + E(V_{t+1}(p, p_t, d_t))]\} \quad (3.3)$$

where $V_t(P_t, d_t)$ is the expected optimal value of selling the dwelling at the beginning of time t , u_t represents a holding cost that depends on the period t and the price P_t is the asking

price on time t . The value $W_t(P_t, \beta)$ represents a withdraw utility he gets if leaves the market at time t , and $E(V_{t+1}(p, p_t, d_t))$ is the expected utility for the next period.

This representation of the behavior of the agent is consistent with the description of an agent-based microsimulation model, however, an agent based bidding model based Alonso's bid rent theory (Alonso, 1964) has not yet been introduced properly into agent based models.

Improving the representation of the housing market at an agent based level, Devisch et al. (2005) introduce a model considering the interaction of households and real estate agents. This model consists on a bilateral trade where a buying agent will only start the negotiation process if this guarantees the largest expected increase in lifetime utility. The buyer submits a bid that maximizes his expectations, and the seller has to decide to set an asking price. Ettema (2011) presents a dynamic model of household's relocation where the behavior of the buyer is not represented in an explicit bidding-counter bidding model. The seller can set an asking price to maximize the expected selling price. Filatova et al. (2009) present different ways in which the seller may be represented and introduces a dynamic approach in which the outcome of a transaction is a price that represents an explicit equilibrium between the buyer's willingness to pay and the seller's willingness to accept. In most models, instead of modeling the interaction of seller and buyer as a bargaining model with two sided incomplete information, they assume that bids are random draws from bids distribution that depends on the asking price among other dwelling attributes.

On the other hand, most of operative microsimulation models for location choice currently available in the market have representations that came from the discrete choice theory, see Wegener (2004) and Hunt et al. (2005). For example, the case of the widely used Urbansim model (Waddell et al., 2003) employs a disequilibrium approach where every household behaves as price taker, choosing their location using a multinomial logit approach. Price of dwellings is then updated by using an hedonic model. ILUMASS model (Moeckel et al., 2002) model the relocation of households similar to a search model. In this model households evaluate the dwellings one by one, and if it presents a significant improvement over their current location the dwelling is accepted. Prices setting is determined by an aggregated function of the number of vacant dwellings in the area. Rosenfield et al. (2013) presents an agent-based housing market simulation for the ILUTE model (Salvini and Miller, 2005), in his work a market clearing procedure is described, which is based on individual's willingness to pay for dwellings and considers the asking price in the utility function. The seller can accept the offers submitted or reject; if the highest bidder's offer is accepted, the dwelling is transacted a price determined by a Vickrey auction. All of these models have lack the representation of the trade process and the explicit representation of the seller - buyer interaction proper as an agent based microsimulation model.

3.3. The proposed framework

The behavior of a seller, representing a household wanting to sell a dwelling unit, is described in this section. The details of the calculations are presented in the sellers model

subsection.

A seller enters the market with the dwelling unit to be sold and determines an initial expected selling price and a lower reservation price W , also determines the maximum number of attempts to sell the property T_s . She enters the active dwelling pool of available units, and must publish an asking price. This asking price and the attributes of the dwelling to be sold are common information for all the buyers in the market. Once the asking price of the dwelling is calculated, the seller will wait for the bidders interested in the dwelling unit. On every time-step t of T_s , the seller will select a bid from the list of bids and will compare the maximum bid offered with his current reservation price; if the maximum bid of those received is larger than the reservation price, then the seller will sell the dwelling unit to the best bidder buyer. In case of bids lower than the current reservation price, the seller can set a new asking price and update the expected selling price. This process takes place in a time scale t of weeks or months. If the seller expected price reaches a minimum value, he could withdraw his dwelling unit from the market.

A buyer will enter the market following an exogenously given awakening process. Once they enter the market, they decide where to bid and a maximum number of attempts he can make to be allocated in a certain dwelling T_b . This number defines the maximum number of counter offers the buyer is willing to make. The search process is made considering all the available housing units that exist on the market at the time he enters. The selection of the unit corresponds to the maximum expected utility or maximum surplus. Rules of eligibility must be exogenously given to the model, so the buyer will look only on feasible dwellings. To select a dwelling unit where to bid, he looks at the dwellings available and the asking prices. The household will compute his willingness to pay for the dwelling unit, which will depend on his current utility level, and the surplus expected. If no houses on the market gives positive surplus, she will exit the market; otherwise the buyer submits a bid for one dwelling unit and waits the answer of the seller. In the case the seller rejects the buyer bid, he will re calculate the bid and the surplus for a maximum of T_b times, when he reaches the maximum number of attempts allowed, he will be indifferent on being the winner of the dwelling or stay in his current location. The complete dynamic of the model is represented in Figure 3.3.

For a transaction to take place the buyer must submit a bid larger than the seller's reservation price. A representation of the evolution of a particular buyer submitting bids of size b_t an a seller updating the reservation price V_t is shown in Figure 3.3. The final price is determined by the size of the bid b_t .

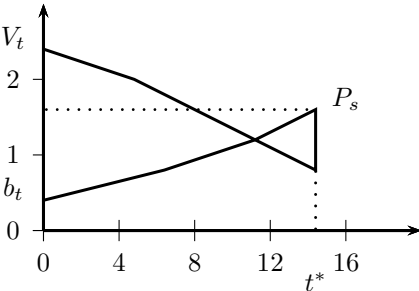


Figure 3.2: Evolution of seller's expectation and buyer's bid

Dynamic Bidding Process

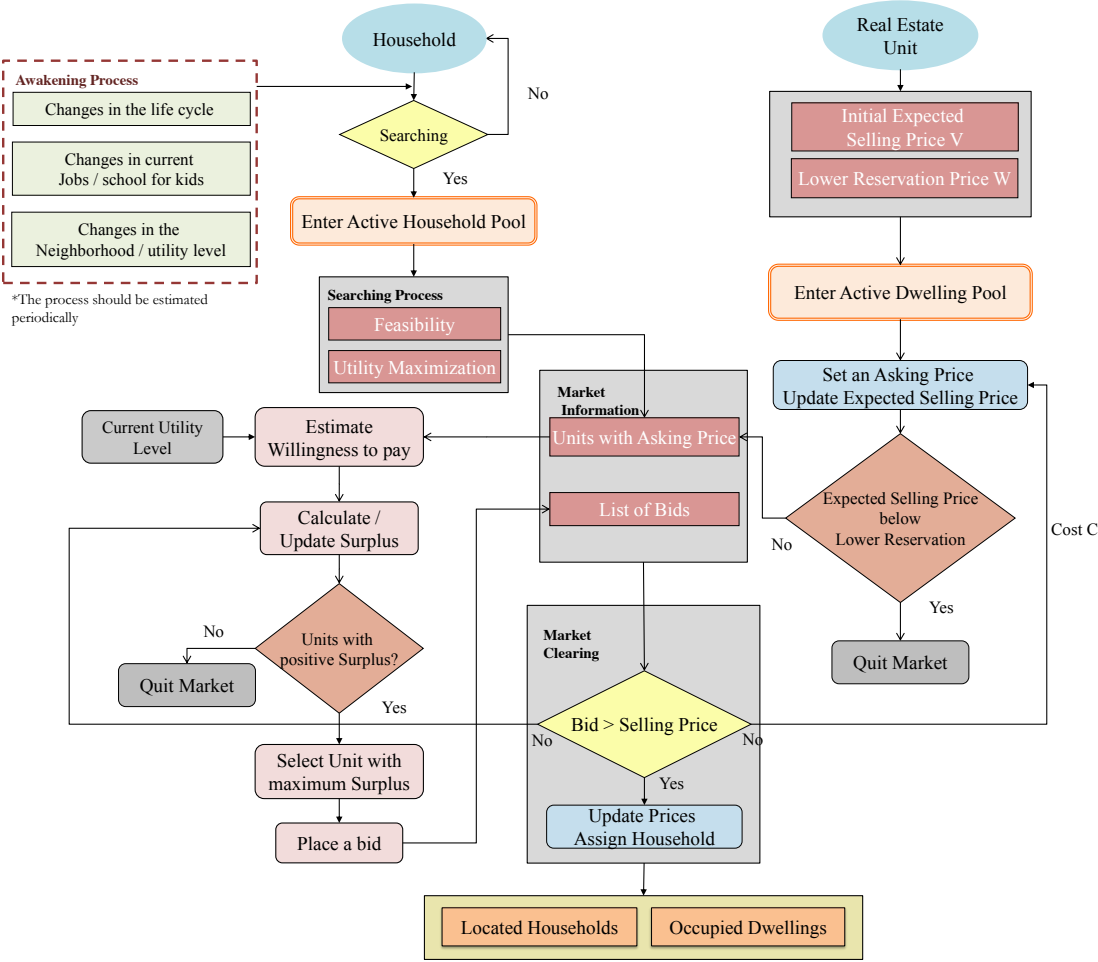


Figure 3.1: Diagram of the Microsimulation Model

3.3.1. A Sellers Model

In this model, the seller has a subjective value of the property, given by $R > 0$, which represents the lower reservation price, or the final price a seller would accept for his property. For the construction of the value $R > 0$ the seller has a perceived market value, that depends on the attributes of the houses. At each step, the seller will sets an asking price P_t and will set an expected reservation price V_t , the seller will select the largest bid of a group received on each time t . All the bids are rejected if none of them is larger than V_t , in that case, the seller will update her expected reservation price and sets a new asking price V_{t+1} and P_{t+1} , thus defining a sequence of reservation prices and asking prices. The maximum number of attempts to sell the dwelling unit is set ex-ante by the seller, this number T_s defines the strategy and the maximum time on the market the seller will be. This behavior is similar to the search process described by Merlo et al. (2013).

The behavior of the seller consists in setting the asking price that maximizes her expected utility of the trade. This originates a dynamic problem, given by:

$$V_t(P_{at}) = \max\{R, \max_{P_a} [C_t + E(V_{t+1}(P_a))]\} \quad (3.4)$$

where C_t is a cost that depends on the time on the market of the dwelling, $E(V_{t+1})$ the expected reservation price in the period $t + 1$, and V_{t+1} the reservation price in $t + 1$.

Our assumptions of the seller corresponds to her believes on the market. A first assumption is that from the sellers perspective bids will arrive following a poisson process with a rate λ , that depends on the asking price she sets.

A second assumption corresponds to the size of the bids, the seller assumes that bids will arrive following a probability density function f , and cumulative distribution F . A final assumption is that buyers wont submit offers larger than the asking price P_{at} set by the seller at time t .

With those assumptions, the expected reservation price for the seller $E(V_{t+1}(P_{at}))$ will be the expected value of bids in next period, which depends on the number of offers received (that depends on λ) and the expected size of the bids (depending on F). For the duration between t and $t + 1$, the seller will accumulate a groups of n offers, and will select the maximum. Then the expected value of V_{t+1} , $E(V_{t+1}(P_{at}))$, can be calculated as follows (see the details of the calculation in the annex):

$$E(V_{t+1}(P_{at})) = P_{at} - \int_{V_{t+1}}^{P_{at}} e^{-\lambda(1-F(b))} db \quad (3.5)$$

Replacing (3.5) in (3.4) we obtain:

$$V_t(P_{at}) = \max_{P_{at}} \left[C_t + P_{at} - \int_{V_{t+1}}^{P_{at}} e^{-\lambda(1-F(b))} db \right] \quad (3.6)$$

To solve this equation, we use T_s : the maximum time to be on the market; setting the final value of the seller could set $V_{T_s} = R$ and solving the equation backwards. This generates a sequence of asking prices and reservation price (P_{at}, V_t) for $t = 1, \dots, T_s$ to use in the proposed framework.

This sequence has been tested numerically with decreasing linear rates $\lambda = \alpha_0 - \alpha_1 P_t$ and, obtaining a decreasing sequence of reservation price and asking price.

3.3.2. The buyers model

The derivation of the model for the buyer is taken from a previous work, (Rocco et al., 2014). Briefly, a household actively looking in the market, once a dwelling unit is selected, he faces the problem of defining an optimum bid to submit. In classical formulations (see for example Alonso (1964), Ellickson (1981), Lerman and Kern (1983)) buyers bid a price that is equal to their willingness to pay for the specific location. In this model, the buyer faces a different problem, similar to the work of Horowitz (1986), where he has a perceived probability of winning the location. This assumption incorporates the fact that the asking price gives information about the reservation price of the seller.

The buyer solves the following problem:

$$\max_{b^t} E(V(b^t)) = H(b^t)V(b^t, z_l) + (1 - H(b^t)) \beta(\max\{V(b^{t+1}, z_l); V_c\}) \quad (3.7)$$

Where $H(b)$ represents the probability perceived by the buyer of the sellers acceptance of a bid of size b ; $V(b, z)$ represents the indirect utility of the buyer, l corresponds to the dwelling unit and z_l the attributes of the housing unit, β stands for a discount factor, and V_c represents the utility at his current location.

To derive the bidding behavior, we assume that the bargain duration is short enough to not have changes in the income level I or in the tastes of the utility function. This assumption fits when a period t is for example a couple of weeks; also notice that in this case β , the discount factor, will be close to 1.

By considering a quasi-linear utility and solving the optimization problem in (3.7) :

$$b_{opt}^t = w_c(z) - (T_b + 1 - k) \frac{H(b_{opt}^t)}{H'(b_{opt}^t)} \quad k = 1 \dots T_b \quad (3.8)$$

Where T_b represents the maximum number of attempts to buy the dwelling, this equation solves the problem of the buyer of determining the optimal size of a bid. Details on the calculation can be found in Rocco et al. (2014), also we need can notice than the bid sequence defined in (3.8) is an increasing bid, that conforms with the model proposed. See 3.3

3.4. Conclusions and Future Work

In this work we have proposed a framework for the microsimulation of the housing market, that represents the behavior of the buyer and the seller and their interactions in the market. A seller model based on search theory captures the behavior of a seller that can accumulate a set of bids. The model depends on the specification of the bids distribution F and the rate of arrival of the buyers λ . The behavior of the buyer is taken from a previous work, and it adapts the bid-rent theory to be applicable in an agent-based framework. The model depends on the assumption of the perceived probability of the seller of being allocated (H), the willingness to pay function and buyer's current utility level.

In our model the formation of the prices and the relocation of agents is given as the result of the interaction of agents performing as buyers and sellers. The buyer and seller model have been developed to mimic the basic behavior in a city. Although the asking price assumption, limiting the maximum value of the bids seems a strong assumption, in most of housing markets transaction prices are below the asking price as a general rule (Merlo et al. (2008)).

Preliminary results obtained so far show the applicability of this model. The SimMobility software engineers have implemented the models in the platform and tests has been carried out. The estimation of the bidding model as it was proposed in our previous work not only gives the estimated parameters for the willingness to pay, but also gives estimated asking price parameters, which can be used to determine an starting value for the iterations of the seller model.

Future work includes incorporating learning behavior into both seller and buyer's models, which can be accomplished by following different techniques, for example see Devisch et al. (2005)'s work. In simulation tests we have employed uniform distribution and constant parameters, but the model can incorporate macroeconomic information, such as unemployment rate, growth of the country among others. In the formulation, the parameter λ that represents the rate of arrival of buyers to bid for the dwelling unit could take that information into account.

Annex

Seller's Expected Utility

For the calculation of the expected utility $E(V^{t+1}(P_t))$ in (3.7), we separate the case off 0 offers with probability $\Pr(n = 0)$, from the case of more than one offer. In the former case:

$$E(V^{t+1}(P_t)) = \Pr(n = 0)V^{t+1} + E(\max(b)|n = 1..\infty) \quad (3.9)$$

Then:

$$E(V^{t+1}(P)) = V^{t+1} \Pr(n = 0) + \sum_{n=1} \int_0^P \max\{V^{t+1}, b\} \Pr(b = \max\{b\}_n) \Pr(n) db \quad (3.10)$$

If the bids arrive i.i.d. with f, F the density and cumulative distributions, then probability of having b as the maximum of n random variables is given by:

$$\Pr(b = \max\{b\}_n) = nF(b)^{n-1} f(b) \quad (3.11)$$

the probability of having n different bids follows a poisson model, then we can write:

$$\Pr(n) = \frac{e^{-\lambda} \lambda^n}{n!} \quad (3.12)$$

Then using previous equations (3.11) and (3.12) into the expectation (3.10), yields:

$$E(V^{t+1}(P)) = V^{t+1} e^{-\lambda} + \sum_{n=1} \int_0^P \max\{V^{t+1}, b\} n F(b)^{n-1} \frac{e^{-\lambda} \lambda^n}{n!} db \quad (3.13)$$

$$= V^{t+1} e^{-\lambda} + \sum_{n=1} \left(\frac{e^{-\lambda} \lambda^n}{n!} \left(\int_0^{V^{t+1}} V^{t+1} n F(b)^{n-1} db + \int_{V^{t+1}}^P b n F(b)^{n-1} db \right) \right) \quad (3.14)$$

$$= \sum_{n=0} \left(\frac{e^{-\lambda} \lambda^n}{n!} \left(V^{t+1} F(V^{t+1})^n + P - V^{t+1} F(V^{t+1})^n - \int_{V^{t+1}}^P F(b)^n db \right) \right) \quad (3.15)$$

Now we can incorporate the case when $n = 0$

$$E(V^{t+1}(P)) = P - \int_{V^{t+1}}^P \sum_{n=0} \frac{e^{-\lambda} \lambda^n}{n!} F(b)^n db \quad (3.16)$$

$$E(V^{t+1}(P)) = P - \int_{V^{t+1}}^P e^{-\lambda(1-F(b))} db \quad (3.17)$$

Another way to see this problem is to define a random variable which corresponds to the maximum between N i.i.d. random variables arriving at a Poisson rate, where N is a random value greater than 0:

$$Y = \max\{b_1, b_2, \dots, b_N\}$$

Computing the cumulative distribution of Y , $F_Y(x)$ by filtering by the rate of arrivals, and defining a process where there are no bids bigger than x yields:

$$F_Y(x) = e^{-\lambda(1-F(x))} \quad (3.18)$$

Given the distribution, we can separate the expected reservation price in two cases : no offer greater than V^{t+1} and with the maximum greater between V^{t+1} and P , then :

$$E(V^{t+1}(P)) = e^{-\lambda(1-F_Y(V^{t+1}))}V^{t+1} + \int_{V^{t+1}}^P xF_Y(x)dx \quad (3.19)$$

but

$$\int_{V^{t+1}}^P xF_Y(x)dx = P - V^{t+1}F_Y(V^{t+1}) - \int_{V^{t+1}}^P F_Y(x)dx$$

replacing in the previous equation we obtain:

$$E(V^{t+1}(P)) = P - \int_{V^{t+1}}^P F_Y(x)dx$$

Replacing F_y

$$E(V^{t+1}(P)) = P - \int_{V^{t+1}}^P e^{-\lambda(1-F(x))}dx$$

which is the same equation as 8.

Notice that if $\lambda \rightarrow +\infty$, then $E(u(P)) \rightarrow P$, and if $\lambda \rightarrow 0$, then $E(u(P)) \rightarrow V^{t+1}$

Bibliography

- Alonso, W. (1964). Location and land use. toward a general theory of land rent. *Location and land use. Toward a general theory of land rent.*
- Devisch, O., Arentze, T., Borgers, A., and Timmermans, H. (2005). An agent-based model of residential choice dynamics in non-stationary housing markets. In *Paper CUPUM Conference, London.*
- Ellickson, B. (1981). An alternative test of the hedonic theory of housing markets. *Journal of Urban Economics*, 9(1):56–79.
- Ettema, D. (2011). A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. *Computers, environment and urban systems*, 35(1):1–11.
- Feinberg, R. M. and Johnson, W. R. (1977). The superiority of sequential search: a calculation. *Southern Economic Journal*, pages 1594–1598.
- Filatova, T., Parker, D., and van der Veen, A. (2009). Agent-based urban land markets: agent’s pricing behavior, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation*, 12(1):3.
- Haurin, D. (1988). The duration of marketing time of residential housing. *Real Estate Economics*, 16(4):396–410.
- Horowitz, J. L. (1986). Bidding models of housing markets. *Journal of urban economics*, 20(2):168–190.
- Horowitz, J. L. (1992). The role of the list price in housing markets: Theory and an econometric model. *Journal of Applied Econometrics*, 7(2):115–129.
- Hunt, J. D., Kriger, D. S., and Miller, E. J. (2005). Current operational urban land-use–transport modelling frameworks: A review. *Transport Reviews*, 25(3):329–376.
- 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.
- Martinez, F. (1996). Mussa: land use model for santiago city. *Transportation Research Record: Journal of the Transportation Research Board*, 1552(1):126–134.
- Martínez, F. and Donoso, P. (2011). A quasi-equilibrium logit model for residential suburban-

- ization. In *Computers in Urban Planning and Urban Management 2011 Conference*.
- McFadden, D. et al. (1978). *Modelling the choice of residential location*. Institute of Transportation Studies, University of California.
- Merlo, A., Ortalo-Magné, F., and Rust, J. (2008). The home selling problem: Theory and evidence. In *Cowles Foundation Summer Conference*.
- Merlo, A., Ortalo-Magné, F., and Rust, J. (2013). The home selling problem: Theory and evidence. Technical report, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- Moeckel, R., Schürmann, C., and Wegener, M. (2002). Microsimulation of urban land use. In *42nd European Congress of the Regional Science Association, Dortmund*, pages 27–31.
- Parker, D. C. and Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems*, 32(6):454–463.
- Rocco, V. A., Hurtubia, R., and Martinez, F. J. (2014). A dynamic bidding model for the housing market. *Unpublished*.
- Rosenfield, A., Chingcuanco, F., and Miller, E. J. (2013). Agent-based housing market microsimulation for integrated land use, transportation, environment model system. *Procedia Computer Science*, 19:841–846.
- Salant, S. W. (1991). For sale by owner: When to use a broker and how to price the house. *The Journal of Real Estate Finance and Economics*, 4(2):157–173.
- Salvini, P. and Miller, E. J. (2005). Ilute: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, 5(2):217–234.
- Timmermans, H. (2003). The saga of integrated land use-transport modeling: how many more dreams before we wake up? In *Keynote paper, Moving through nets: The Physical and social dimension of travel, 10th International Conference on Travel Behaviour Research, Lucerna*, www.ivt.baug.ethz.ch/allgemein/pdf/timmermans.pdf.
- 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.
- Wegener, M. (2004). Overview of land-use transport models. *Handbook of transport geography and spatial systems*, 5:127–146.

Capítulo 4

Tercer Artículo

USING HOT-DECK MULTIPLE IMPUTATION TO MATCH DATABASES

Abstract

Although Census data and mobility surveys contain most of the information needed to perform the estimation of the classical location choice models, they usually are not obtained from the same source and therefore don't conform a unique match. The objective of this paper is to study the performance of the hot-deck multiple imputation method to match information from two different databases to obtain the data needed for the calibration of a Lerman and Kern Model. Our work shows that the results of the estimation with imputed and observed data give estimates that are statistically identical.

4.1. Introduction

One of the most well known models for location choice modeling was developed by Ellickson (1981) based on the discrete choice of Domencich and McFadden (1975). In Ellickson's work, dwellings are traded in an auction market, where the best bidder is assigned to the dwelling. Improving Ellickson's work, Lerman and Kern (1983) incorporate the transaction price to the likelihood of the auction, allowing better estimates of the willingness to pay function. Applications of the Lerman and Kern model has been reported by several authors (Gross, 1988, Gross et al., 1990, McMillen, 1997, Muto, 2006, Hurtubia and Bierlaire, 2014).

Estimation is a key process in building of any model. A commonly reported problem for the estimation of location choice models is the difficulty to obtain data in a detailed level to perform estimation of spatially detailed models. Data requirements for the estimation of this models typically consists on the description of both dwelling and household characteristics. In particular, the estimation of Lerman and Kern's model requires an econometric description of the willingness to pay that considers attributes of the location, the dwelling and the household. With detailed data, the aggregation into groups usually made in both Ellickson and Lerman and Kern's model could be avoided, therefore avoiding aggregation and biases on estimation and forecast. An example of a disaggregated willingness to pay function is in Hurtubia and Bierlaire (2014), where the specification considers the interaction between dwelling and household attributes, generating a functional form that estimates individual values for each particular combination of household-dwelling variables.

Typical sources of data for the estimation are census and travel surveys, but when data describing dwellings units and data characterizing households come from separate sources, the estimation of the models requires assumptions to match the data before making the estimation, usually at a zone level, see for example Picard and Antoniou (2011).

The estimation made over a single matched database does not incorporate the variance in the estimators from the different configurations that the database could take. Thus, the motivation of this work is to estimate Lerman and Kern’s model by using two separated databases, one describing household characteristics and other describing the dwellings attributes, both databases contain information of spatial location of the households and the dwelling at different zonal aggregations. Postcode, zone number, commune, or other spatial identifier allows to link the information at spatial level, but in absence of a unique match between the two databases, multiple configurations could be created. Our goal is not to obtain a new matched database, but to estimate a statistically significant model given two related databases and their multiple matches.

To perform the estimation we propose to use a multiple imputation technique usually called *hot-deck multiple imputation* to obtain estimates that considers the variance coming from multiple matches. The estimation of the model is made without considering the aggregation into groups for the households or types of dwellings, to avoid the groups formation we formulate a model interacting attributes from dwelling and households on the specification of the willingness to pay function. Hot deck imputation method is a kind of non parametric imputation method, it consists on replacing the missing value with observed values. Where the selection of the imputed value usually needs to specify a pool of donors, and then randomly select a donor to fill a missing value.

Multiple imputation is a widely known technique for filling in missing observations, and was first introduced by Rubin (1987). It consists on in filling the missing values with some plausible values obtained from the sampling of a prior distribution. In each imputation the missing values of the database are completed, allowing the usage of complete-data methods. The application of this methodology can be seen in the work of Steimetz and Brownstone (2005), where multiple imputation is used to obtain the estimation of a discrete choice model.

Imputation methods are usually employed in land use modeling for the construction of a micro data set that represents the population, for example the work of Frazier and Alfons (2012) shows the usage of distance-based imputation methods to generate a synthetic population for Ghana. While the imputation methods used in population synthesis generates disaggregated data, the estimation of the models over this generated data does not takes into account the propagation of errors of the generated data. The notion of imputed values in discrete choice models has been recognized by some authors, see for example Ben-Akiva et al. (2002), Waddell (2009).

In section II we review and describe the theoretical fundamentals of the Lerman and Kern model and the multiple imputation technique, in section III we present the application to a full database for the city of Brussels. In section IV proof-of-concept is made by using the complete database and estimating the model, then taking this data as if it were obtained from two separated sources and by using the hot-deck matching a new set of parameters is estimated. With this sets of estimated parameters we perform a Hausman test to find statistically differences in the models. We present the application of the methodology to

Singapore data, where the source of the information came from two state agencies. Finally we simulate synthetic data to test the behavior of the estimates using the imputation method.

4.2. The Model

Following the discrete choice theory of Domencich and McFadden (1975), Ellickson (1981) developed a model considering that real state goods are traded in an auction market (an approach called commonly bid-auction), where households bid their willingness to pay for a particular dwelling which is assigned to the best bidder defining the allocation of dwellings to households.

In Ellickson's work the use of a stochastic bid price function $\hat{B}_{hi} = B_{hi}(z_i) + \varepsilon$, represents the willingness to pay for a dwelling unit i of a household that belongs to a socioeconomic group h . $B_{hi}(z_i)$ is the deterministic component of the bid function, representing the observable part of the willingness to pay that depends on the vector of attributes of the dwelling type i . ε represents a stochastic term that represents differences in incomes and tastes among the households in the group.

The probability of observing a household h located in a dwelling i , as a result of the auction is defined by Ellickson's model by:

$$\Pr(h|i) = \Pr(B_{hi} + \varepsilon > B_{ki} + \varepsilon, \forall k \neq h, \quad k, h \in H) \quad (4.1)$$

Considering ε extreme value Gumbel independent and identically distributed ε , H the set containing all households groups, then the probability takes the Logit form:

$$P(h|i) = \frac{\exp(\mu B_{hi})}{\sum_{k \in H} \exp(\mu B_{ki})} \quad (4.2)$$

With Ellickson's model, only relative estimates of the willingness to pay parameters can be found, because the scale parameter μ cannot be identified.

Lerman and Kern (1983) incorporates information about the observed transaction price into Ellickson's model. They consider the probability of the winning bid to be equal to the observed transaction price R_i in the likelihood function. By doing that, the model predicts both the absolute values of willingness to pay estimates and the location of the bidder.

Then the probability of a household h being the best bidder and simultaneously offering a bid equal to the observed transaction price is given by:

$$\Pr(h|i) = \Pr(B_{hi} + \varepsilon = R_i \text{ and } B_{hi} + \varepsilon > B_{ki} + \varepsilon, \quad \forall k \neq h, \quad k, h \in H) \quad (4.3)$$

the error ε distributes Gumbel with cumulative function F and density f given by:

$$f(\varepsilon) = \mu \exp(-\mu\varepsilon) \exp(\exp(-\mu\varepsilon)) \quad (4.4)$$

$$F(\varepsilon) = \exp(-\exp(-\mu\varepsilon)) \quad (4.5)$$

Lerman and Kern's model probability of observing h located in unit i is written as:

$$\Pr(h|i) = f(R_i - B_{hi}) \prod_{k \neq h} F(R_i - B_{ki}) \quad (4.6)$$

The log-likelihood function of a sample of households of size H , with S the set of dwellings in the market, is given by:

$$l = H \ln(\mu) - \mu \sum_{i \in S} \sum_{h \in H} (R_i - B_{hi}) - \sum_{h \in H} \sum_{k \in S_h} \exp(\mu(R_k - B_{hk})) \quad (4.7)$$

And by maximizing the likelihood or the log-likelihood function the parameters of B_{hi} could be estimated. Lerman and Kern's model allows the simultaneous estimation of the parameters in B_{hi} and the scale parameter μ

Both preceding models are usually estimated via maximum likelihood, the data requirements for the estimation includes data that characterizes both households and dwellings; for Lerman and Kern estimation the price of the dwelling is also needed. Although this information may be extracted from census and mobility surveys, is usually not always available in a detailed level.

Estimations of the willingness to pay function B_{hi} consider the aggregation of the households into homogeneous groups in both models. This kind of aggregation does not capture the heterogeneity accross households in the same group. By making groups for estimation, the willingness to pay function will predict the same values for individuals belonging to the same group, which does not suits the needs of modern agent based models.

4.3. Hot Deck Multiple Imputation Algorithm

The Hot-Deck Multiple imputation algorithm consists on a non parametric imputation technique for matching missing data with observed one. The technique was first introduced by Rubin (1987) seminal work and it corresponds to a particular case of general multiple imputation methods. Asymptotic properties of the hot-deck estimator has been developed by Fay (1996). A more recent revision of the methods is in Andridge and Little (2010) where they analyze the current statistical properties of the hot deck imputation and perform a comparison between the different techniques existing for computing the multiple imputation method.

In a hot-deck multiple imputation, the missing values are replaced by values of other completely observed values. The selection of the value to replace is made by randomly picking a donor from a pool of potential donors. This pool is generated usually by the usage of some distance-based measurement. Estimation on complete data is then performed with the imputed database. This replacement procedure is repeated several times, and final values

of the estimated parameters are obtained taking the average of the estimates of all the imputation. Variance and standard errors are estimated by considering the between and within imputation variance (see below).

In this work, we match information from one database to another in every imputation, this procedure treats the information in one database as missing information on second database. This interpretation is a key assumption, because it allows the use of the methods proposed by Rubin (1987) to perform the calculation of the final estimates of the model.

Now we summarize the procedure implemented for the hot-deck imputation for the estimation of the Lerman and Kern model:

- **Creation of Donors:** For every dwelling available on the database, we select households to create the pool of donors. This is made by looking for all the households that constitute a consistent match. For example, if the dwelling and the household have the same postal code, zone or commune. Consistency also implies that households reporting living in a landed dwelling wont be assigned to apartments.
- **Selection of the match:** The selection of the household to be imputed is performed by randomly assigning one household from the pool of donors of the dwelling unit. After this imputation is done for every dwelling unit, one complete database is obtained
- **Estimation:** Once the data is filled with the imputed households, we estimate the Lerman and Kern model. This is done by maximizing the log-likelihood function described in equation ((4.7)). The vector of parameters θ_k of the k imputation is stored, also with the variance matrix U_k .
- **Final estimation:** The final values of the models with the statistic values are obtained after M imputations; the calculation is made by applying the equations from Rubin's work described next.

The value of the θ_i parameter of the willingness to pay function is calculated by taking the average among the M imputations, where θ_{im} is the i parameter estimated by maximum likelihood during the m imputation, then:

$$\theta_i = \frac{1}{M} \sum_{m=1}^M \theta_{im} \quad (4.8)$$

The variance and standard errors for each parameter considers the variability within imputation U and the between imputation variability B .

Given by the following relations:

$$U = \sum_{m=1}^M U_{im}/M \quad (4.9)$$

$$B = \sum_{m=1}^M \left(\hat{\theta}_m - \theta_M \right) \left(\hat{\theta}_m - \theta_M \right)' / (M - 1) \quad (4.10)$$

The associated total variance of θ is T , given by:

$$T = U_M + \frac{M+1}{M} B_m \quad (4.11)$$

For each scalar parameter, the standard error and significance are calculated individually by the relation given by Rubin, with a t-student distribution with v degrees of freedom given by:

$$v = (M - 1) \left(1 + \frac{MU_m}{(1 + M)B_m} \right) \quad (4.12)$$

This formulae are derived by considering the imputation as random samples of a posterior distribution, details of this derivation are available by Rubin (1987), Chapter 3.

4.4. Proof of concept

For the validation of the usage of the multiple imputation technique in the Lerman and Kern model, we first study the performance of the imputation technique over the estimation of a multinomial logit model using synthetic data. A second test consists on comparing the estimates obtained with the multiple imputation method and the Lerman and Kern's model using a complete database with real data. We perform the estimation considering a complete database without performing any imputation, then parameters estimated are compared with those obtained by applying the imputation to the same database.

4.4.1. First test: Synthetic Data

A numerical simulation was made to estimate how the imputation technique performed under controlled situations. The numerical test consists on simulating a multinomial Logit Ellickson (1981) in a linear city.

In this city two attributes are generated, X_1 and X_2 , where X_1 represents the size of the unit of dwelling, increasing the the distance to the city center and X_2 represents a measure of accessibility, that decreases with the distance to the city center. The micro-zones are distributed homogeneously along the city center. A representation of the values is shown in Figure 4.1

The order in this city is defined by the micro zone numbers, the lower the number of the zone, the closer to the CBD the dwelling. Zone, communes and macro zone follows the same order. Attributes are generated by taking an increasing (X_1) and a decreasing (x_2) function plus a random noise. Three types of spatial aggregation units were defined: micro zones with

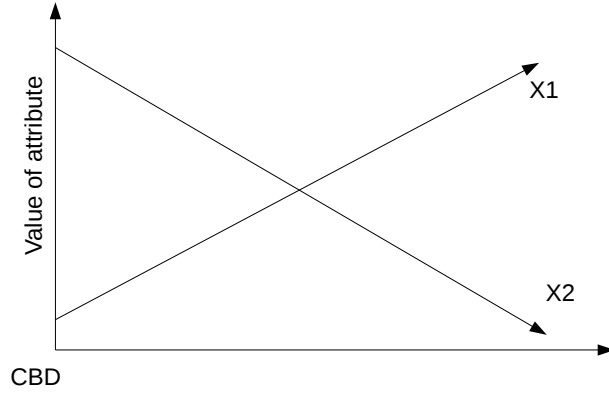


Figure 4.1: Levels of $X1$ and $X2$ in the linear city

15000 units, each one with one household allocated, 500 zones, 50 communes and 5 Macrozone (see Figure 4.2).

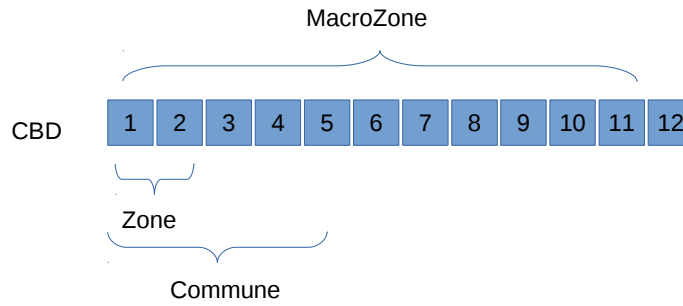


Figure 4.2: Zoning Structure

Since the estimation of the multinomial logit gives estimation for differences of utilities, our simulation was made by considering three types of agents and fixing the utility of one of them. By doing this the parameters of the other two are identifiable. Estimation results consider a number of 15 imputations. Results and parameters of the utility function employed for individuals 1 and 2 are shown in Table 4.1.

$$U_1 = \beta_1 X1 + \gamma_1 X2 \quad (4.13)$$

$$U_2 = \beta_2 X1 + \gamma_2 X2 \quad (4.14)$$

$$U_3 = \alpha \quad (4.15)$$

The value α was set to be a constant average of the utilities U_1 and U_2 to assure equivalent distribution of the selections. The assignation was made by computing U_1, U_2 and U_3 with the values of $X1$ and $X2$ and simulating the choice by drawing from the multinomial logit

probability

$$P_1 = \frac{\exp(\mu U_1)}{\exp(\mu U_1) + \exp(\mu U_2) + \exp(\mu U_3)} \quad (4.16)$$

$$P_2 = \frac{\exp(\mu U_2)}{\exp(\mu U_1) + \exp(\mu U_2) + \exp(\mu U_3)} \quad (4.17)$$

$$P_3 = \frac{\exp(\mu U_3)}{\exp(\mu U_1) + \exp(\mu U_2) + \exp(\mu U_3)} \quad (4.18)$$

taking an error term Gumbel distributed with scale parameter $\mu = 1$.

The imputation was performed by matching X_1 and X_2 values with the observed selection 1,2 or 3. The pool of donors was generated by selecting from different levels of aggregation. When the pool of donors is created with information at a microzone level, the values of X_1 and X_2 corresponds to a unique match. As the scale increases the size of the pool of donors and variance inside the group increases too. Estimation of the multinomial logit was made by maximum likelihood employing R software.

Table 4.1: Simulated Parameters Estimation

Par.	Real Value	Microzone	S.E.	Zone	S.E.	Commune	S.E.	Macrozone	S.E.
β_1	0.7	0.7032	0.0058	0.6980	0.0074	0.6992	0.0110	0.7648	0.0071
γ_1	0.5	0.4961	0.0077	0.4901	0.0126	0.4919	0.0189	0.6810	0.0096
β_2	0.6	0.5985	0.0053	0.6010	0.0112	0.6087	0.0140	0.6712	0.0084
γ_2	0.8	0.8002	0.0046	0.8043	0.0146	0.8021	0.0139	0.6717	0.0129

From the estimation results we noticed that the estimates for microzone, zone and commune are almost the same, which is expected because the parameters are correlated with the microzone number. In each zone attributes X_1 and X_2 are close to each other, therefore the randomly imputed values are close relatively close to the true X_1 and X_2 .

In the case of the commune level the model performs well, the average imputed values are almost the same real value, although data at commune level could have larger variance inside, and values for the parameters X_1 and X_2 could be very different than the true parameter.

The worst prediction is made at a Macrozone level, where the differences between the imputations are larger, the low value for the standard error could be explained because at every imputation the specification could accommodate to explain the differences, therefore the variance of in between imputation was low. This error will be larger if the functional form could not accommodate to the imputed value, for example when the functional form includes specific dummies variables.

Two conclusions arise from this test: first, the method works well, specially where the parameters are homogeneously distributed in the donor pool. Secondly, the quality of estimates decreases as the size of the donors pool increases as the variance inside the pool.

4.4.2. Second Test: Brussels' Real Data

In this case we estimated a Lerman and Kern's model with real data from the residential market of Brussels (Hurtubia and Bierlaire, 2014). Data contains information of the located household and the current dwelling, as zonal and commune attributes, with 4945 zones and 151 communes. The dwelling alternatives consists on four types: detached, semi-detached, back-to-back houses and apartments.

The specification of the willingness to pay function is inspired in the previous work of Hurtubia and Bierlaire (2014), where the interaction between dwelling and household attributes is made to account for the heterogeneity of households otherwise gruped. We consider three constants for groups of income: low, medium and medium-high. A dummy variable interacts with the households size the dwelling is a house is interacted with a dummy variable indicating if the household has more than 2 members and belongs to a level of income medium high, to capture the preferences of largest families for houses. Floor space of the dwelling is interacted with dummy variables for households of size 1, 2 and 3 or more persons, to take into account the difference of the willingness to pay for space when the size of the household varies. The percentage of households of low income in a zone interacts with a dummy variable for high incomes households to capture segregation effects. To capture spatial agglomeration and positive externalities of the high incomes households the percentage of households of high income in a zone interacting with a dummy variable indicating if the household belongs to a group of medium-high income. Finally accessibility measures of car and public transport interacts with a dummy variable indicating the car availability, to capture attractiveness of the different transportation modes.

The willingness to pay function used for the proof of concept consists on the interaction of households and dwelling attributes, the estimation was made via maximum likelihood. The natural logarithm of the price was used to capture the diminishing marginal utility of housing attributes as suggested by DiPasquale and Wheaton (1996) and Hurtubia and Bierlaire (2014)

Table 4.2: Brussels Interacting Attributes

Name	Household attribute	Dwelling Attribute
ASC1, ASC2, ASC3	Constant for incomes, low, - medium, high	
bhouse	Size of household	Dummy if is house
bsurface1,2,3+	Dummy if household is of size 1,2,3 or more	Surface of the dwelling
blowinc	Dummy if mid-high income	Percentage low income hh in the zone
bhighinc	Dummy if mid-low income	Percentage of high income hh in the zone
bcar	Dummy if has car	Car accessibility
btrans	Dummy if not has car	Public transport accessibility

Results are shown in Tables 4.3 and 4.4. The values obtained are consistent with the results reported by (Hurtubia and Bierlaire, 2014). All the parameters are statistically significant and signs are as expected for both with and without imputation.

The imputation was made by creating a pool of donors of a commune level (151 communes), for every dwelling unit we randomly draw a household within the same commune. The pool of donors varies from 100 to 200 candidates. The standard error and t-value are estimated with equations (4.11) and (4.12)

Table 4.3: Lerman and Kern's Model Estimates - Results without imputation

Parameter	Value	Standard Error	t-value
mu	5.4900	0.1098	50.0212
asc1	9.6364	0.0412	233.6930
asc2	9.5471	0.0368	259.7766
asc3	9.8621	0.0944	104.4311
bhouse	0.0938	0.0204	4.5926
bsurface1	0.0713	0.0029	24.3315
bsurface2	0.0698	0.0029	23.8416
bsurface3+	0.0669	0.0033	20.4939
blowinc	-0.4498	0.1504	-2.9895
t-value bhighinc	0.9625	0.1089	8.8365
bcar	0.0330	0.0040	8.2177
btrans	0.2650	0.0353	7.5061

Table 4.4: Results obtained with imputation

Parameter	Value	Standard Error	t-value
mu	5.4814	0.1106	49.5589
asc1	9.6699	0.0445	217.0964
asc2	9.5644	0.0373	256.5583
asc3	9.9704	0.1069	93.2364
bhouse	0.0568	0.0282	2.0151
bsurface1	0.0708	0.0031	22.7900
bsurface2	0.0701	0.0032	22.1635
bsurface3+	0.0676	0.0038	17.6235
blowinc	-0.5940	0.1723	-3.4469
bhighinc	1.0096	0.1168	8.6452
bcar	0.0285	0.0043	6.6571
btrans	0.2216	0.0373	5.9347

A chi-squared Hausman test was made to determine significant differences between two set of estimated parameters. For vector of parameters θ_0 and θ_1 , the hausman statistic is given by

$$H = (\theta_1 - \theta_0)'(V_0 - V_1)^{-1}(\theta_1 - \theta_0) \quad (4.19)$$

The value obtained of $H = 0.73749$ shows that for 12 variables used, there are not significant differences between the estimation made by the use of multiple imputation and by the complete data set. The results from this test shows that the imputation method performed well for the case of Brussels. This confirms the good results obtained with the first test.

4.5. Application to Singapore data

Finally an application of the imputed method is performed over real non matched data. This application is part of the development of an agent-based transport and land use microsimulation model for Singapore (Simmobility). In this application the hot deck multiple imputation method is used to estimate of a willingness to pay function that considers the interaction between households and dwelling attributes.

Singapore is divided into 5 areas called planning regions, created to facilitate the planification of the use and development of land. The 5 regions are Central, East, North East, North and West regions. Each Planning Region consists of several planning areas which are further divided into sub-zones. Data is available at individual household level and they have a common spatial information given by the postcode.

Data Availability

The data came from two separated sources. One dataset is obtained from the Real Estate Information System (REALIS), which is an internet based information system that depends on the Singapore's Urban Redevelopment Authority. The database contains information of the real state market, such as the specific location of the dwelling, type of sale (whether sold by developer or resale), tenure of land, unit size, transaction price, and date. It also contains information that links the transaction with an specific zonal identification (planning areas, postal codes, street names).

The other dataset is the Household Interview Travel Survey (HITS), conducted by the Singapore's Land Transport and Authority (LTA) every four to five years, and developed to have an insight on travelers' behavior. From this survey we obtain data that describes the characteristics of the households: Number of persons, Number of Children, Income group, access to car, among others.

The next tables present the data available from both surveys.

Table 4.5: Hits Survey

Name	Description
Pcode	Postal Code
Household Size	Number of members of household
H3Ethnic	Ethnicity : Malay, Chinese, Indian or Other
H5VehAvailable	Vehicle availability
H7Bike	Bike availability
Children	Number of children in the household
Income	Income group from 1 to 11

Table 4.6: REALIS Survey

Name	Description
Project Name	Name of the Project
PostCode	Postal Code
Size	Size of the dwelling in square meters
Property Type:	Condominium, Apartment, Detached House, Semi Detached House, Executive Condominium, Terrace House
Sale Type	New Sale or Resale
Price	Price of the dwelling in Singapore Dollars (SGD)
Postcode	Postal code of the dwelling
DistanceToCBD	Distance to the Central Business District

Residential property in Singapore can be classified into three distinct categories: housing and development board units, private apartments and landed property. The majority of the Singapore population lives in the housing and development board units which are within housing estates. The remaining Singaporeans live in private apartments or landed property. Immigrants have ownership restrictions on the private apartments and on landed property as government policies, while Singapore residents can own and live in any of these three types of properties.

Specification of the Willingness to Pay

We use similar specification as the used in the the Brussels case: one alternative specific constant is estimated for each income group (ASC); dwelling floor space interacting with a dummy for households of size 1, 2 and 3 or more, to account for different valuation of the need for space of the dwelling for different households sizes. A measure of accessibility is given by the distance to the CBD, wich interacts with a dummy indicating the availability of the car of the household. The estimated model is a simple specification only used to test the

performance of the imputation method.

The estimation of the parameters is performed by using the multiple imputation procedure described previously. Imputation was performed by selecting a sample size of 200 households, and matching was made at a postcode level. The model results are shown in Table 4.7, the method was performed with 15 imputations.

Table 4.7: Singapore Lerman and Kern's model Estimation

Param	theta	se
ASC1	4.8220	0.0548
ASC2	4.9019	0.0599
ASC3	4.7780	0.0804
ASC4	4.8298	0.0800
ASC5	4.8937	0.0540
ASC6	4.8395	0.0467
ASC7	4.8581	0.0438
ASC8	4.8467	0.0435
ASC9	4.8485	0.0459
ASC10	4.8643	0.0483
ASC11	4.8721	0.0410
b_area1	0.0037	0.0003
b_area2	0.0041	0.0003
b_area3	0.0045	0.0002
b_distCDB	-0.0151	0.0012
mu	3.9754	0.0958

From the results we can see that the parameters obtained have the expected sign and all of them are significant. As in the case of Brussels estimation constants vary little, this fact could be understood considering the use of the logarithm of the price, so differences in ASC constants are logarithm of differences on the willingness to pay.

Results from the application of this technique are important because allows the estimation of a bid-auction approach as an alternative to hedonic models Rosen (1974). In general terms, the bid-auction approach generates better results since it takes into account the preferences of the consumers in addition to land prices. The estimation of a bid-auction approach without the use of the multiple imputation method although is possible would likely require the estimation of a unique matched database. This unique match would require time and wouldn't take into the account the variance of the several combinations that would be possible.

4.6. Summary and Conclusion

In this work we have shown the use of a non parametric multiple imputation technique for matching databases with related fields to estimate the Lerman and Kern's model. The main advantages of this technique is to avoid making an explicit unique match between the databases, that do not take into account the variability of the different combinations that could take place. The time required to create a unique match and the possible bias in the outcome are avoided with this method. With this model is possible to obtain estimated parameters with an error that considers the effect from the uncertainty of the imputed database and between matches.

The estimation of the model was made "hands off", i.e. without making any assumption or hand-picking units, with the additional benefit of recovering statistical significant parameters. The rationale behind the multiple imputation arises from considering the set of imputed values of the households as random draws from a Bayesian posterior distribution. Data missing in the imputation method (household attributes) does not depend on the observed data, conforming with the Missing Completely at Random assumption required for the hot deck imputation.

Although the number of imputation needed for proper estimation will depend on the level of missing information, we performed a number of imputation until obtaining stable results; this criteria usually needed less than 10 imputations. A usual criteria is to stop the imputations when the results are stables (Horton and Lipsitz, 2001). Simulated data shows that if the pool is created with a scale level where the attributes are homogeneous inside of it, then the expected estimates are closely related to the real parameters.

The proof of concept performed consisting in two exercises showing the good properties of the imputation method, allowing to recover almost identical parameters with synthetic data and creating a pool of proper size. In the second excersise, with the city of Brussels, results show that the models estimated with the hot-deck multiple imputation and with the complete database are statistically the same model. This fact may be explained because of the use of categorical data: income, size of the household, availability of cars. This type of data is common in land use modeling and it has been reported that this is the case when this imputation technique perform best (see Cranmer and Gill (2013)). Since categorical data employed present a low variance from commune to commune, households inside the pool of donors have very homogeneous parameters, and as our simulated experiment for the linear city shows, this is when the hot-deck imputation performs best.

The application of this technique allows the matching of dwelling database with census data or any other database that characterizes the households for estimating bid-rent models. This gives the advantage of incorporating the consumers behavior in the estimation.

Bibliography

- Andridge, R. R. and Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, 78(1):40–64.
- Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Börsch-Supan, A., Brownstone, D., Bunch, D. S., et al. (2002). Hybrid choice models: Progress and challenges. *Marketing Letters*, 13(3):163–175.
- Cranmer, S. J. and Gill, J. (2013). We have to be discrete about this: A non-parametric imputation technique for missing categorical data. *British Journal of Political Science*, 43(02):425–449.
- DiPasquale, D. and Wheaton, W. C. (1996). *Urban economics and real estate markets*, volume 23. Prentice Hall Englewood Cliffs, NJ.
- Domencich, T. A. and McFadden, D. (1975). Urban travel demand-a behavioral analysis. Technical report.
- Ellickson, B. (1981). An alternative test of the hedonic theory of housing markets. *Journal of Urban Economics*, 9(1):56 – 79.
- Fay, R. E. (1996). Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association*, 91(434):490–498.
- Frazier, T. and Alfons, A. (2012). Generating a close-to-reality synthetic population of ghana. *FBE Research Report KBI_1206*, pages 1–15.
- 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.
- Horton, N. J. and Lipsitz, S. R. (2001). Multiple imputation in practice: comparison of software packages for regression models with missing variables. *The American Statistician*, 55(3):244–254.
- Hurtubia, R. and Bierlaire, M. (2014). Estimation of bid functions for location choice and price modeling with a latent variable approach. *Networks and Spatial Economics*, 14(1):47–65.

- 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.
- McMillen, D. P. (1997). Multiple regime bid-rent function estimation. *Journal of Urban Economics*, 41(2):301 – 319.
- Muto, S. (2006). Estimation of the bid rent function with the usage decision model. *Journal of Urban Economics*, 60(1):33 – 49.
- Picard, N. and Antoniou, C. (2011). Econometric guidance for developing urbansim models. first lessons from the sustainability project. In *Proceedings of the 51st European Congress of the Regional Science Association International, Barcelona, Spain*.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, pages 34–55.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.
- Steinmetz, S. S. and Brownstone, D. (2005). Estimating commuters' value of time with noisy data: a multiple imputation approach. *Transportation Research Part B: Methodological*, 39(10):865–889.
- Waddell, P. (2009). Parcel-level microsimulation of land use and transportation: The walking scale of urban sustainability. In *12th International Conference on Travel Behavior Research*.