



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

EVALUACIÓN DE UNA PRUEBA ESTADÍSTICA PARA DETECTAR
ENDOGENEIDAD EN MODELOS BID DE USO DE SUELO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL

JOSÉ RICARDO SOTO VAQUERO

PROFESOR GUÍA:
ÁNGELO GUEVARA CUE

MIEMBROS DE LA COMISIÓN:
FRANCISCO MARTÍNEZ CONCHA
RICARDO HURTUBIA GONZÁLEZ

Este trabajo ha sido parcialmente financiado por el proyecto Fondecyt No. 1150590

SANTIAGO DE CHILE
2019

EVALUACIÓN DE UNA PRUEBA ESTADÍSTICA PARA DETECTAR ENDOGENEIDAD EN MODELOS BID DE USO DE SUELO

Los modelos de uso de suelo permiten analizar la interacción entre los sistemas de transporte y el mercado inmobiliario. Esto los vuelve una herramienta de gran importancia para la toma de decisiones que concluye en proyectos como una línea de metro, un edificio comercial o un nuevo plano regulador. Considerando el alcance y consecuencias que tienen esta clase de proyectos en la sociedad, resulta fundamental que los modelos propuestos sean consistentes y permitan realizar pronósticos que reflejen la realidad en escenarios futuros.

Los modelos de uso de suelo actuales son desarrollados bajo un enfoque Bid de máxima postura aleatoria, que podrían sufrir de endogeneidad debido, entre otras cosas, a la omisión de atributos, difíciles de medir pero muy influyentes en lo observado. La detección y corrección de endogeneidad ha sido estudiada ampliamente en el contexto de elecciones discretas (Louviere et al., 2005; Guevara, 2015), pero no en el caso particular del modelo Bid.

Se han desarrollado múltiples metodologías para estimar modelos Bid (Ellickson, 1981; Lerman y Kern, 1983) que varían en términos de eficiencia. En consideración de lo anterior, esta memoria propone el desarrollo de una prueba de hipótesis basada en la prueba de especificación de Hausman y McFadden (1984) que logre detectar endogeneidad en modelos Bid. La principal ventaja que presenta el test propuesto frente a otras opciones es el no requerir de supuestos adicionales ni del uso de variables instrumentales para ser aplicado.

Para estudiar el test propuesto se desarrolla un experimento de Monte Carlo donde se evalúan los indicadores de *size distortion* y *poder empírico* del test, los que son contrastados con el método de funciones de control. Este último se propone como alternativa pese al inconveniente de exigir el uso de variables instrumentales. El experimento simula una ciudad monocéntrica compuesta por hogares de dos estratos socioeconómicos donde se realiza un remate de viviendas a partir de un conjunto de atributos dado, logrando un equilibrio de mercado controlando la distribución de hogares por estrato y suponiendo un precio pagado igual a la máxima postura observada. Luego, se procede a estimar el modelo omitiendo un atributo de la simulación, que deriva en una fuente de endogeneidad que permite evaluar el test.

La simulación concluye que la prueba desarrollada es capaz de detectar endogeneidad bajo ciertas condiciones. No obstante, existen casos donde el test puede fallar. En particular, cuando la valoración del atributo omitido es diferente para cada estrato se observan caídas en el poder empírico del test. Por otra parte, la comparación realizada con funciones de control muestra que este último posee mayor poder, mas el uso de variables instrumentales sigue representando un inconveniente que el test propuesto logra subsanar.

Para José Miguel.

Agradecimientos

En primer lugar deseo dar las gracias a quienes han puesto su granito de arena para llegar a ser quien soy: Mis padres Berlena y Ricardo, mis hermanas Jennifer y Ángela, mis abuelos, tíos, primos y sobrinos, especialmente a Nicolás quien acaba de cumplir su primer año. Agradezco también a todos aquellos que me apoyaron durante mi formación escolar, de quienes estaré eternamente agradecido. Una mención especial a Luis Álvarez y Daniel Wladdimiro, cuya amistad ha perdurado hasta hoy en día.

Quiero agradecer también a quienes me acompañaron durante mi paso por la universidad: A los chicos de la sección 2 en mi etapa mechona; a la Pandilla: Aurora, Emilio, Coni, Mireya, Camilo, Suafle, Karina, Pablo y Pelo, con quienes he compartido grandes experiencias; y a las personas con quienes compartí estando ya en Ingeniería de Transporte: Esteban, Benja, Javi, Raúl, Vera, Nico, Maca, André, Isa, Hans, Ali, Kim y tantos más.

Finalmente quiero agradecer a los profesores y funcionarios de Ingeniería de Transporte. A Sergio Jara por acercarme al área, a Marcela Munizaga por responder mis dudas vocacionales y en especial a mi profesor guía Ángelo Guevara, quien fue un soporte fundamental para que este trabajo llegara a buen puerto. No puedo dejar de mencionar a Mónica Chávez y Raúl Espinoza, personas clave para que todo funcione como se debe.

Tabla de Contenido

1. Introducción	1
1.1. Presentación y Motivación	1
1.2. Objetivos	2
1.3. Metodología	2
1.4. Estructura de la memoria	4
2. Marco Teórico	5
2.1. Presentación	5
2.2. Modelos de uso de suelo	5
2.2.1. Antecedentes	5
2.2.2. Formulación y estimación del modelo Bid	6
2.3. Contraste de hipótesis	7
2.3.1. Conceptos básicos	8
2.3.2. Test de especificación de Hausman	9
2.3.3. Test de especificación de Hausman-MacFadden	10
2.4. Endogeneidad en modelos de elección discreta	11
3. Desarrollo del test para detectar endogeneidad en modelo Bid de posturas	15
3.1. Presentación	15
3.2. Test de especificación de Hausman-McFadden modificado	15
3.2.1. Formulación del test	16
3.2.2. Normalización de la escala en modelos Bid	16
3.2.3. Evaluación del test	20
3.3. Test de Funciones de Control	20
3.3.1. Consideraciones previas	20
3.3.2. Formulación del test	21
3.3.3. Evaluación del test	22
3.4. Conclusiones del capítulo	22
4. Evaluación del test mediante simulación de Monte Carlo	23
4.1. Presentación	23
4.2. Formulación de la simulación	24
4.2.1. Definición de la muestra de viviendas	24
4.2.2. Definición de la muestra de hogares	25
4.2.3. Cálculo de posturas	25
4.2.4. Generación de datos	25

4.2.5.	Asignación de viviendas	28
4.2.6.	Estimación de modelos y escenarios de endogeneidad	31
4.3.	Evaluación de la distorsión de tamaño o <i>size distortion</i>	32
4.3.1.	Definición del escenario simulado	33
4.3.2.	Resultados obtenidos	33
4.4.	Evaluación del poder empírico del test	35
4.4.1.	Procedimiento general de evaluación	36
4.4.2.	Resultados Escenario A: Beta ruido igual por estrato	37
4.4.3.	Resultados Escenario B: Beta ruido distinto por estrato	39
4.5.	Extensión de la simulación a ciudad cerrada	45
4.5.1.	Modificaciones metodológicas	45
4.5.2.	Resultados de estimación del poder empírico	49
5.	Síntesis, Conclusiones y Recomendaciones	57
5.1.	Síntesis de resultados	57
5.2.	Conclusiones sobre el test desarrollado	58
5.3.	Investigación posterior	58
	Bibliografía	59

Índice de Tablas

3.1. Simulación parámetro lambda – Resumen de resultados – Caso sin endogeneidad	18
3.2. Simulación parámetro lambda – Resumen de resultados – Caso endógeno . . .	19
4.1. Distribución de probabilidad en los atributos propuestos	27
4.2. Matriz de correlación entre atributos	27
4.3. Parámetros por estrato	28
4.4. Evaluación de <i>size distortion</i>	34
4.5. Resumen de resultados – Estimación de poder empírico – Escenario A	37
4.6. Resumen de resultados – Estimación de poder empírico – Escenario B	39
4.7. Parámetros por estrato – Escenario B-II	42
4.8. Resumen de resultados – Estimación de poder empírico – Escenario B-II . .	43
4.9. Resumen de resultados – Estimación de poder empírico – Escenario A – Caso ciudad cerrada	50
4.10. Resumen de resultados – Estimación de poder empírico – Escenario B – Caso ciudad cerrada	54

Índice de Ilustraciones

3.1. Simulación parámetro lambda – Caso sin endogeneidad	18
3.2. Simulación parámetro lambda – Caso endógeno	19
4.1. Densidad de probabilidad para atributos muestreados	26
4.2. Análisis de dispersión para muestra de viviendas	28
4.3. Distribución de hogares por estrato – Ciudad Abierta – Escenario sin endogeneidad	30
4.4. Análisis por <i>size distortion</i> – Test de Hausman-McFadden modificado	35
4.5. Análisis por <i>size distortion</i> – Test de Funciones de Control	36
4.6. Poder empírico – Escenario A	38
4.7. Evolución del parámetro λ – Escenario A	38
4.8. Poder empírico – Escenario B	40
4.9. Evolución del parámetro λ – Escenario B	40
4.10. Distribución de hogares por estrato – Escenario B	41
4.11. Poder empírico – Escenario B-II	42
4.12. Evolución del parámetro λ – Escenario B	44
4.13. Distribución de hogares por estrato – Escenario B	44
4.14. Distribución de hogares por estrato – Ciudad Cerrada – Orden de asignación aleatorio	46
4.15. Distribución de hogares por estrato – Ciudad Cerrada – Eliminación secuencial al mejor postor	47
4.16. Distribución de hogares por estrato – Ciudad Cerrada – Eliminación secuencial equilibrada	48
4.17. Poder empírico – Escenario A – Caso Ciudad Cerrada	51
4.18. Evolución del parámetro λ – Escenario A – Ciudad Cerrada	51
4.19. Distribución de hogares por estrato simulada en ciudad cerrada con endogeneidad	52
4.20. Poder empírico – Escenario B – Caso Ciudad Cerrada	53
4.21. Evolución del parámetro λ – Escenario B – Ciudad Cerrada	53

Capítulo 1

Introducción

1.1. Presentación y Motivación

En Ingeniería de Transporte se habla tradicionalmente de un modelo «de cuatro etapas»; a saber, Generación, Distribución, Partición Modal y Asignación. Estas fueron planteadas con el fin de modelar escenarios donde interactúan tres sistemas: Transporte, Actividades y Flujos en red. Más adelante se incorporó al modelo una quinta etapa, denominada «Uso de Suelo», que establece un nexo entre los sistemas de transporte y el mercado inmobiliario. Esto resulta de gran interés, ya que modelar el mercado inmobiliario permite analizar el impacto de proyectos indirectamente vinculados con transporte tales como políticas de vivienda, planos reguladores, generación de subcentros en una ciudad o mejoras de accesibilidad, entre otros.

Para cuantificar el impacto del uso de suelo fueron desarrollados modelos basados en elecciones discretas, siendo en la actualidad una referencia para la toma de decisiones en proyectos de mediana y gran envergadura. Chile ha aportado en esta área desarrollando modelos específicos, como MUSSA (Martinez, 1996; Martínez y Donoso, 2010), utilizado para analizar el mercado inmobiliario de la ciudad de Santiago y que ha sido incorporado como una componente del software comercial llamado Cube (Martínez, 2011).

Considerando la relevancia de las decisiones tomadas con el apoyo de estos modelos, resulta clave garantizar una estimación consistente de estos, evitando obtener resultados contraintuitivos fruto de endogeneidad, la que puede deberse a múltiples factores como los son un conjunto de consideración inadecuado o la omisión de un atributo relevante, entre otros. Cabe destacar que a la fecha no existen trabajos que aborden el problema de detectar endogeneidad bajo el contexto de modelos Bid de posturas aleatorias.

Esto motiva el análisis del modelo Bid de posturas aleatorias, llevado a cabo mediante el estudio de un enfoque alternativo para detectar endogeneidad debido a la omisión de atributos en la función de postura. Adicionalmente, el modelo utilizado tiene la característica de poseer dos estimadores para la función de postura, Ellickson (1981) y Lerman y Kern (1983), ambos consistentes y cuya diferencia es observada al comparar su eficiencia.

Utilizando los anterior se propone desarrollar una prueba estadística de especificación basada en el trabajo de Hausman y McFadden (1984), que permita detectar endogeneidad en el modelo Bid de posturas aleatorias mediante la comparación de estimadores para la función de postura que varían en términos de eficiencia.

1.2. Objetivos

El objetivo principal de esta memoria es desarrollar un **test de hipótesis** que permita detectar endogeneidad en el modelo Bid de posturas aleatorias. Se propone verificar su correcto funcionamiento desarrollando las siguientes actividades.

1. **Formulación del test.** En esta etapa se busca estudiar la literatura de modelos de posturas aleatorias y pruebas de especificación para proponer el test de hipótesis que será evaluado posteriormente. Se discuten los supuestos requeridos por el test y la metodología utilizada para su evaluación.
2. **Desarrollo de la simulación de Monte Carlo de una ciudad.** En esta etapa se plantea la simulación del comportamiento de una ciudad monocéntrica, lo que permitirá generar muestras de datos que permitan estimar el modelo Bid y evaluar el test propuesto.
3. **Desarrollo de escenarios de endogeneidad.** Se buscar discutir condiciones de la simulación que permitan causar endogeneidad mediante la omisión de atributos. Así, el test propuesto podrá ser aplicado, estableciendo un parámetro que cuantifique el nivel de endogeneidad presente.
4. **Evaluación de desempeño del test utilizando simulación de Monte Carlo y comparación con otros test de hipótesis.** En esta etapa se propone la evaluación del test desarrollado usando dos indicadores: *Size distortion* y *Poder estadístico*. El primer indicador permite comparar el error por falso positivo nominal con el obtenido de forma empírica, mientras el segundo permite medir la capacidad del test de rechazar la hipótesis nula (H_0) cuando esta es falsa. Los resultados serán comparados con otro test utilizado para detectar endogeneidad y con ello se buscará reconocer las fortalezas y debilidades del test propuesto.

1.3. Metodología

Para llevar a cabo el primer objetivo se realiza una revisión bibliográfica que aborda tres temas: Modelos de remates y su estimación (Ellickson, 1981; Lerman y Kern, 1983), Pruebas estadísticas de especificación aplicadas en elecciones discretas (Hausman, 1978; Hausman y McFadden, 1984) y el estudio de la endogeneidad en modelos de localización (Louviere et al., 2005; Guevara, 2015). El principal desafío que presenta esta etapa es obtener un estimador del modelo Bid normalizado por su escala, el que permitirá comparar de forma directa las múltiples formas de estimación existentes.

El segundo objetivo se lleva a cabo realizando muestreos aleatorios utilizando el software R (R Core Team, 2018). Esto permite generar muestras de viviendas con tres atributos: Superficie, distancia al centro de la ciudad y ruido ambiente. Adicionalmente, se establece una muestra de hogares estratificada en dos niveles de ingreso, alto y bajo, los que difieren en la valoración de atributos para cada vivienda. Hogares y viviendas son asignados mediante la simulación de un remate, donde las posturas de cada hogar son explicadas tanto por efectos sistemáticos, producto de la valoración de atributos individuales, como aleatorios, los que son simulados mediante variables aleatorias de tipo Gumbel con escala unitaria y parámetro de localización nulo. El remate descrito, además de generar una asignación, define dos elementos adicionales: el precio de cada vivienda, igual a la máxima postura observada y la composición socioeconómica de la muestra, cuyo comportamiento es análogo a un modelo de ciudad abierta (Brueckner, 1987).

Para cumplir con el tercer objetivo se propone un escenario donde el investigador estima un modelo de remates, ignorando los efectos del ruido que presenta la muestra. Así, la valoración de este atributo en la etapa de simulación permitirá proponer escenarios donde existe endogeneidad debido a la omisión del mismo y así el grado de endogeneidad estará determinado por la valoración del ruido y el nivel de correlación que presente con el atributo distancia. Así, cuando la influencia de este sea nula se espera que el test propuesto no detecte endogeneidad más allá de un error aceptable, mientras que cuando exista influencia del ruido en la decisión, el test será capaz de detectarla y traducirá en rechazos a la hipótesis nula de no endogeneidad.

Para la última etapa se propone calcular dos indicadores que caracterizan el error de un test de hipótesis:

- **Size distortion:** Permite estudiar la presencia de falsos positivos cuando se cambia el tamaño de muestra.
- **Poder empírico:** Permite estudiar la cantidad de falsos negativos para un nivel de endogeneidad dado.

Para cumplir lo anterior se establece un escenario y nivel de endogeneidad dados, donde se simula una cierta cantidad de aplicaciones del test, donde cambia la muestras de hogares. Cada aplicación se traduce en una aceptación o rechazo de la hipótesis nula; siendo la proporción de rechazos sobre el total representará el indicador evaluado y permitirá cuantificar su efectividad en cada caso. Lo anterior se lleva a cabo partiendo del caso donde no existe endogeneidad y luego se repite aumentando el nivel de endogeneidad hasta que se alcanza una tasa de rechazo del 100%.

Finalmente, para establecer un punto de comparación con la prueba desarrollada se desarrolla un segundo test de hipótesis, el que está basado en el enfoque de funciones de control (Heckman, 1977; Villas-Boas y Winer, 1999). Esta prueba presenta una serie de hipótesis adicionales debido al uso de variables instrumentales, las que en la práctica no son fáciles de obtener. El test se evalúa siguiendo la metodología anteriormente descrita y los resultados son contrastados con los obtenidos por el test propuesto.

1.4. Estructura de la memoria

Esta memoria tiene una estructura de cinco capítulos, siendo el primero la presente introducción. El segundo capítulo aborda la revisión bibliográfica de temas afines al trabajo y establece un marco teórico para la comprensión del mismo. El capítulo tercero presenta el test desarrollado, explicando las hipótesis involucradas y la forma en que se evalúa dicho test. También se presenta la prueba de funciones de control, que servirá para comparaciones posteriores. El cuarto capítulo desarrolla y presenta los resultados de un experimento de Monte Carlo desarrollado para evaluar los tests de hipótesis, presentando como resultado final el análisis de size distortion y del poder empírico de cada test. Finalmente el capítulo quinto realiza una síntesis del trabajo, señalando las conclusiones más relevantes, recomendaciones metodológicas y otras líneas de investigación que pueden abordarse a partir de este trabajo.

Capítulo 2

Marco Teórico

2.1. Presentación

Este capítulo expone el marco teórico de la memoria y realiza una revisión del contexto y la literatura relacionada con el problema que se busca abordar. El capítulo se divide en tres secciones, las que abordan los tópicos principales que fueron investigados. El primero corresponde a modelos de uso de suelo bajo el enfoque de remates (*bid-auction*), donde se aborda el modelo Bid de posturas aleatorias, los supuestos que involucra y las metodologías de estimación utilizadas. El segundo punto refiere a pruebas estadísticas de especificación, donde se realiza una revisión de conceptos y se explican los trabajos en que se basa la prueba desarrollada en esta memoria. El último tema corresponde a una revisión de las metodologías existentes para la detección de endogeneidad en el contexto de modelos de elección discreta y particularmente en el modelo Bid.

2.2. Modelos de uso de suelo

2.2.1. Antecedentes

Los primeros modelos de uso de suelo corresponden a modelos agregados que consideran una estructura de ciudad monocéntrica (Alonso et al., 1964; Mills, 1967; Muth, 1969; Brueckner, 1987). En este sentido, el trabajo de Alonso et al. (1964) plantea un modelo donde logra estimar las curvas de disposición a pagar percibidas por un individuo al localizarse a distintas distancias del centro de una ciudad manteniendo un nivel de utilidad.

Desde la perspectiva de elecciones discretas, McFadden (1978) propone un modelo (*choice*) donde el tomador de decisiones (hogar) elige una vivienda a partir de un conjunto de opciones dado y sus respectivos atributos, donde el precio es dado. Considerando efectos aleatorios de valor extremo se deriva un modelo de elección discreta Logit Multinomial.

El modelo *choice* presenta como principal crítica el hecho de no abordar cómo se obtienen los precios utilizados en la asignación, dejando que otros modelos (e.g. Rosen, 1974; Anas, 1982) aborden dicho punto. Considerando lo anterior, Ellickson (1981) propone un modelo alternativo basado en la simulación de un remate inmobiliario (*bid-auction*), siendo pionero en aquello. Así, la asignación es realizada al mejor postor mediante una función de postura (B_i), que sirve como aproximación de la disposición a pagar por el inmueble y permite estimar los precios pagados por cada hogar, de forma similar a un modelo de precios hedónicos.

El modelo Bid de Ellickson (1981), al igual que un modelo Logit convencional, trata la función de postura como una variable latente, cuya componente sistemática es estimada hasta una escala en el propio modelo. Lerman y Kern (1983) argumentan que la escala de dicha función puede ser determinada conociendo el precio final pagado por una vivienda. Incorporando esta información proponen un nuevo estimador del modelo Bid, de mayor eficiencia y con la posibilidad de incorporar un factor de escala.

2.2.2. Formulación y estimación del modelo Bid

El modelo de Ellickson (1981) parte de una función de postura (*bid*) B_{hi} , definida de antemano, que corresponde a la oferta realizada por un hogar de tipo $h \in H$ por la vivienda $i \in V$. Esta función representa la componente sistemática de postura para cada hogar.

Para incorporar efectos aleatorios no asociados a la función de postura se incorpora una componente aleatoria ε_{hi} de distribución Gumbel con parámetros de localización 0 y escala 1. De esta forma queda definida la función \tilde{B}_{hi} , que representa la postura aleatoria observada del hogar h por la vivienda i :

$$\tilde{B}_{hi} = B_{hi} + \varepsilon_{hi} \quad (2.1)$$

Ahora se procede a definir la probabilidad $P(h|i)$ de un hogar h de ser el ganador del remate de una vivienda i particular, como:

$$P(h|i) = \mathbb{P}(\tilde{B}_{hi} \geq \tilde{B}_{h'i}; \forall h' \neq h; h, h' \in H) \quad (2.2)$$

Es directo observar que la distribución de $P(h|i)$ depende de la distribución de ε_{hi} . En este caso, al distribuir Gumbel y suponiendo independencia idénticamente distribuida entre todas las variables aleatorias ε_{hi} , es posible calcular una expresión cerrada para $P(h|i)$:

$$P(h|i) = \frac{\exp(B_{hi})}{\sum_{h' \in H} \exp(B_{h'i})} \quad (2.3)$$

la cual corresponde a la probabilidad de elección de un modelo Logit Multinomial ordinario.

Para estimar la función de postura de Ellickson se recurre a un estimador de máxima verosimilitud, representado por la siguiente función:

$$\mathcal{L} = \prod_{i \in V} \left(\prod_{h \in H} (P_{h|i})^{y_{hi}} \right) \quad (2.4)$$

donde y_{hi} es un indicador binario que toma valor uno si la vivienda i fue asignada al hogar h , en caso contrario es cero.

El estimador máximo verosímil de Ellickson no considera el precio pagado por la vivienda. En consideración, otros autores (e.g. Lerman y Kern, 1983; Dubin y McFadden, 1984; McMillen, 1997; Hurtubia y Bierlaire, 2014) han propuesto enfoques alternativos para estimar modelos de posturas incorporando el precio de los bienes rematados. A continuación se presenta el desarrollo de uno de estos.

La publicación de Lerman y Kern (1983) propone una nueva expresión para calcular la probabilidad $P(h|i)$, la que incorpora el precio final pagado por la vivienda P_i^* :

$$P(h|i) = \mathbb{P}(\tilde{B}_{hi} = P_i^* \text{ y } \tilde{B}_{h'i} \leq P_i^*; \forall h' \neq h; h, h' \in H) \quad (2.5)$$

De esta expresión se calcula un nuevo estimador para B_{hi} , también de máxima verosimilitud, que incorpora los efectos del precio ofertado.

$$\mathcal{L} = \prod_{i \in V} \left(-\mu \exp(-\mu(P_i^* - B_{hi})) \prod_{h' \in H} \exp(-\exp(-\mu(P_i^* - B_{h'i}))) \right)^{y_{hi}} \quad (2.6)$$

$$\ell = -|V| \ln(\mu) - \sum_{i \in V} y_{hi} \left((\mu(P_i^* - B_{hi})) - \sum_{h' \in H} (\exp(-\mu(P_i^* - B_{h'i}))) \right) \quad (2.7)$$

El nuevo estimador tiene la ventaja de poder determinar la escala del modelo de forma directa, ya que puede ser identificado de forma independiente en la ecuación de log-verosimilitud (2.7). Además, debido a que posee más información, tendrá un beneficio extra en términos de eficiencia, lo que redundará en estimadores de menor varianza.

2.3. Contraste de hipótesis

Las pruebas estadísticas de especificación son un elemento tratado por la literatura econométrica para validar la consistencia de un estimador aprovechando casos en que dos estimadores consistentes pueden diferir en el nivel de eficiencia. En primer lugar se realizará una revisión de conceptos básicos sobre la materia, para proceder posteriormente con la revisión de pruebas de especificación y su aplicación en elecciones discretas.

2.3.1. Conceptos básicos

Un **test de hipótesis** es un mecanismo de validación de una propiedad poblacional a partir de una muestra particular. Para su desarrollo se debe establecer una hipótesis nula (H_0), la que se pretende validar; una alternativa (H_1) que permita efectuar su contraste y un nivel de confianza $(1 - \alpha)$, con $0 < \alpha < 1$, que dependerá de los objetivos definidos por cada usuario.

Estadístico muestral

Es una medida cuantitativa computada a partir de una muestra aleatoria, la cual permite conocer características de un modelo estadístico. Los ejemplos clásicos son la media y varianza muestral, mientras que en contraste de hipótesis son conocidos estadísticos como el t de Student y el estadístico F de Fisher.

El uso de estadísticos es fundamental para el desarrollo de pruebas de hipótesis, ya que disponen de una distribución conocida y se define una **región de rechazo** a partir de ella. Así, la prueba consiste en calcular la probabilidad que tiene el estadístico de hallarse en la región de rechazo cuando (H_0) es cierta.

Errores de tipo I y II

Este concepto se asocia a dos problemas conocidos de la inferencia estadística: rechazar hipótesis verdaderas y validar aquellas falsas. En el primer caso se rechaza (H_0) cuando la hipótesis era cierta, lo que se conoce como **error de tipo I o falso positivo**. En el segundo, por el contrario, se acepta (H_0) cuando es falsa, lo que se define como **error de tipo II o falso negativo**.

Se utiliza α para definir la probabilidad de cometer un error de tipo I, mientras que β representa la probabilidad de cometer un error de tipo II. Ambos valores son independientes, aunque puede prestarse para confusiones. Ellis (2010) señala que realizar un control fuerte del nivel de significación, minimizando errores de tipo I, redundará en mayores probabilidades de cometer un error tipo II, lo que se verá manifestado en un menor poder del test.

Nivel de significación, nivel de confianza e intervalo de confianza

El nivel de significación α , como ya se mencionó, corresponde a la probabilidad de cometer un error de tipo I. Este no debe ser confundido con el nivel de confianza $(1 - \alpha)$, que es la probabilidad de no cometer un error tipo I.

Ambos términos están ligados con el concepto de **intervalo de confianza**, que es un rango de valores donde la probabilidad de que el intervalo generado de esa manera contenga el valor verdadero es igual al nivel de confianza.

Si bien es deseable tener el mayor nivel de confianza posible, se utiliza por convención un nivel del 95 por ciento (Fisher, 1925). Lo anterior no resulta trivial, puesto que controlar α incrementa la probabilidad de cometer errores tipo II (Ellis, 2010).

Poder estadístico y tamaño del efecto

Se define **poder o potencia estadística** como la probabilidad de rechazar la hipótesis nula cuando esta es falsa. En otras palabras, corresponde a la probabilidad de no cometer un error de tipo II.

$$\text{Poder estadístico} = \mathbb{P}(\text{Rechazar } H_0 | H_0 \text{ es falso}) \quad (2.8)$$

Cabe mencionar que si bien el poder de una prueba de hipótesis puede ser calculado de forma teórica, también puede ser estimado de forma empírica, lo que facilita los cálculos sin comprometer los resultados obtenidos.

El poder de un test debe ser calculado para un **tamaño del efecto** dado, el cual corresponde a la magnitud que controla el efecto estudiado por el test, como puede ser el nivel de endogeneidad de una muestra.

Size distortion

Se dice que un test posee *size distortion*, para un experimento de tamaño N dado, cuando la probabilidad de cometer un error tipo I difiere de su valor nominal, es decir del nivel de significación elegido.

La tasa de rechazos observada cuando se cumple la hipótesis nula define la categoría de esta distorsión. Si es inferior a α se habla de distorsión «conservadora» y cuando es superior se habla de distorsión «liberal». El primer caso es visto como aceptable, puesto que corresponde a una menor tasa de falsos positivos que lo esperado; mientras que el último resulta indeseado ya que conlleva a lo contrario del primer caso.

2.3.2. Test de especificación de Hausman

El trabajo de Hausman (1978) es utilizado como base para las pruebas de especificación posteriores. El trabajo presentado prueba que bajo la hipótesis nula de que el modelo esté bien especificado, el vector de diferencias entre dos estimadores consistentes que varían en su eficiencia es asintóticamente nulo. Este resultado se utiliza para definir un test de hipótesis para un modelo lineal:

$$y = X\beta + \varepsilon \quad (2.9)$$

Los estimadores de este modelo quedan definidos por la especificación de ε , que debe cumplir con los siguientes supuestos:

1. **Ortogonalidad:** ε debe tener esperanza condicional nula, para un X dado.
2. **Esfericidad:** La matriz de covarianzas debe ser diagonal, presentando homocedasticidad y no autocorrelación.

Si se verifica el supuesto de ortogonalidad es directo señalar que se cumple la hipótesis nula y por ende el modelo está bien especificado. Así, para probar dicho supuesto se estima el siguiente modelo:

$$y = X\beta + \tilde{X}\alpha + \nu \quad (2.10)$$

donde \tilde{X} es una transformación de X . Así, se define como hipótesis nula $H_0 : \alpha = 0$, ya que en caso contrario ($H_1 : \alpha \neq 0$) el supuesto de ortogonalidad dejaría de ser válido. Así, comparando ambos escenarios (H_0 y H_1) se busca comparar el valor *beta* estimado, resultando en los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$.

Para comparar ambos estimadores, se elabora un estadístico $\hat{q} = \hat{\beta}_1 - \hat{\beta}_2$ el cual es asintóticamente nulo cuando el modelo está bien especificado. Para probar lo anterior se construye el siguiente estadístico:

$$m = T\hat{q}^T \hat{V}(\hat{q})^{-1} \hat{q} \sim \chi_K^2 \quad (2.11)$$

donde T es el número de observaciones (temporales) usadas para la estimación, K es la dimensión de β y $\hat{V}(\hat{q})$ es un estimador consistente de varianza de \hat{q}

Este estimador permite probar la hipótesis $q = 0$, concluyendo la aceptación o rechazo de la hipótesis nula inicial.

El principal aporte de este trabajo en el contexto de econometría, fue permitir la elaboración de hipótesis nula no paramétricas (Godfrey, 1991) y realizar este cálculo únicamente con las varianzas del estimador \hat{q} . Este trabajo fue extendido al campo de las elecciones discretas por Hausman y McFadden (1984), como se presenta en la siguiente sección.

2.3.3. Test de especificación de Hausman-MacFadden

La publicación de Hausman y McFadden (1984) extiende el Test de especificación de Hausman como base para elaborar una prueba de hipótesis válida en modelos de elección discreta; realizando una aplicación particular para verificar la validez del supuesto de Independencia de Alternativas Irrelevantes (IIA) para el modelo Logit Multinomial. El test propuesto hereda la principal ventaja del test original, que es el solo requerir una matriz de covarianzas para la diferencia entre parámetros.

La validez del supuesto IIA, definida como hipótesis nula, implica que la razón de probabilidad entre dos alternativas es independiente de los atributos que presente una tercera alternativa. Con esto en mente, la validación del supuesto requiere comparar los estimadores obtenidos por dos modelos: Un modelo con un conjunto de consideración C dado y una versión restringida con un conjunto de consideración ($A \subset C$). Así, se construye un indicador θ de dimensión $|A|$ con los valores estimados para calcular la razón de probabilidad observada en cada escenario. Estos resultados son contrastados con el estadístico de Hausman-McFadden:

$$T = N(\theta_A - \theta_C)^T Q^{-1}(\theta_A - \theta_C) \sim \chi_K^2 \quad (2.12)$$

$$Q = V_A - V_C \quad (2.13)$$

donde N es el tamaño de la muestra y V_i representa la matriz de covarianzas de la cada escenario. Cabe destacar que los grados de libertad de χ^2 dependen directamente del rango de la matriz Q , por lo que $K = \text{rango}(Q)$.

El test es utilizado en la presente memoria como base para la prueba de hipótesis propuesta en modelos de localización. A diferencia del caso desarrollado por Hausman y McFadden (1984), en la presente memoria se utiliza el mismo conjunto de consideración y se comparan estimadores pertenecientes a dos implementaciones del mismo modelo (Ellickson y Lerman & Kern).

2.4. Endogeneidad en modelos de elección discreta

En elecciones discretas se habla de **endogeneidad** cuando alguna variable explicativa presenta correlación con el término de error del modelo asociado. Entre las posibles causas de esto se pueden nombrar: variables omitidas, errores de medición, determinación simultánea o auto selección. El fenómeno de la endogeneidad se presenta de forma casi inevitable en múltiples casos de estudio y redundante en estimadores inconsistentes de los parámetros del modelo, invalidando cualquier tipo de análisis asociado a este.

Los modelos de localización son casos susceptibles de sufrir algún tipo de endogeneidad, generalmente debido a la omisión de atributos. En un escenario real, los hogares participantes contrastan una serie de atributos, tanto de la vivienda como del propio hogar, para tomar una decisión: Ubicación, tamaño, número de habitaciones y baños, antigüedad del inmueble, accesibilidad, entre tantos otros. Al intentar modelar lo anterior, el modelador suele ignorar aquellos atributos que considera de menor importancia, principalmente por la dificultad para obtener ciertos datos o en su defecto, la inexistencia de estos. Si la hipótesis del modelador resulta errada, el efecto de la variable omitida quedará capturado por la variable con que presenta mayor correlación, estimando incorrectamente su impacto y causando endogeneidad. Un ejemplo práctico de esto se presenta en el trabajo de Guevara (2005), donde se realiza un análisis utilizando datos de la Encuesta Origen Destino 2001 de la ciudad de Santiago.

La literatura ha desarrollado numerosos métodos para corrección de endogeneidad aplicado en modelos de elección discreta. No obstante, se observa que dichos métodos presentan limitaciones en el rango de aplicación de cada método y los supuestos involucrados en ellos. Un ejemplo de ello es el uso de variables instrumentales, las cuales permiten estimar el impacto de una variable mediante otra correlacionada, pero cuya existencia es incierta. A continuación se hará una revisión de los enfoques clásicos utilizados en corrección de endogeneidad. Para ello se reportan los trabajos de Louviere et al. (2005) y Guevara (2015), señalando los enfoques de mayor utilización y aquellos más recientes.

Enfoque BLP (Berry, 1994; Berry et al., 1995)

Es un método de corrección que utiliza regresiones lineales y variables instrumentales cuando existe endogeneidad sobre la muestra de tomadores de decisión, como es el caso de mercados geográficamente distanciados que son analizados en conjunto. De lo anterior se desprende que el rango de aplicación queda limitado a casos donde se observa endogeneidad en forma agrupada.

Funciones de Control (Heckman, 1977; Villas-Boas y Winer, 1999)

Este método permite detectar y corregir endogeneidad mediante el uso de variables instrumentales. Estas deben estar correlacionadas con la variable endógena y al mismo tiempo, ser independientes del error que presente el modelo. Para aplicar el método se realiza una regresión lineal de la variable endógena en la variable instrumental, incluyendo todas las variables exógenas del modelo, permitiendo identificar los efectos por omisión de atributos y así corregir endogeneidad mediante la inclusión de los residuos obtenidos del modelo lineal.

Este método tiene su principal crítica en la necesidad de obtener instrumentos válidos. Bound et al. (1995) señala que el uso de instrumentos débiles en un modelo lineal puede llevar a resultados indeseados, por lo que resulta de interés encontrar metodologías que no requieran su uso.

Solución de múltiples indicadores (MIS) (Guevara y Polanco, 2016)

Se propone el uso de instrumentos alternativos, que en teoría sean más fáciles de medir, con el objeto de caracterizar la variable latente causante de endogeneidad. Su aplicación es realizada en dos etapas: La primera consiste en incluir uno de los instrumentos en la función a estimar, trasladando los efectos de endogeneidad a esta nueva variable. Luego se corrige esta variable haciendo uso del enfoque de funciones de control aprovechando que se dispone de un segundo instrumento.

Conclusiones

Esta sección ha presentado los enfoques principales utilizados en elecciones discretas. No obstante, la aplicación de estos a modelos de localización ha sido limitada y requiere de supuestos sofisticados. El siguiente capítulo presenta una forma para detectar endogeneidad aplicable a estos modelos que reduce notablemente la cantidad de supuestos realizados, esto al basarse en una prueba de especificación como es el test de Hausman-McFadden.

Capítulo 3

Desarrollo del test para detectar endogeneidad en modelo Bid de posturas

3.1. Presentación

Este capítulo presenta el desarrollo teórico de lo que es propuesto como principal aporte de la memoria, es decir, el desarrollo de una prueba de hipótesis que permita detectar endogeneidad en el contexto del modelo Bid de posturas aleatorias. Para ello se adapta el trabajo de Hausman y McFadden (1984), buscando aprovechar la formulación realizada para evaluar dos estimadores del modelo Bid, presentados en Ellickson (1981) y Lerman y Kern (1983).

La primera sección presenta la formulación del test de Hausman-McFadden modificado, los supuestos involucrados y la forma en que se evalúa el test. La segunda sección propone como enfoque alternativo el uso de funciones de control con el fin de establecer una base que permita comparar el desempeño de ambos test.

3.2. Test de especificación de Hausman-McFadden modificado

El test propuesto corresponde a una adaptación de lo mostrado por Hausman y McFadden (1984). De la revisión bibliográfica realizada se recuerda que la formulación original busca validar el supuesto de Independencia de Alternativas Irrelevantes (IIA) mediante la comparación de dos estimadores de máxima verosimilitud para un modelo logit multinomial, alterando el conjunto de consideración utilizado por cada estimador.

La modificación, propuesta en el contexto del modelo Bid de remates, consiste en comparar dos estimadores que varían por la cantidad de información utilizada: Ellickson y Lerman y Kern. Los estimadores propuestos difieren en términos de eficiencia, sin necesidad de cambiar el conjunto de consideración, por lo que el test sigue siendo aplicable.

3.2.1. Formulación del test

Considere un modelo de remates Bid con un vector de parámetros θ que se desea estimar y la existencia de dos estimadores: $\hat{\theta}$ y $\tilde{\theta}$, que corresponden a lo obtenido por Ellickson (1981) y Lerman y Kern (1983). Ambos logran estimar de forma consistente un modelo de remates tipo *bid-auction*, por lo que ambos estimadores deben entregar el mismo resultado cuando son normalizados por su escala. Este último punto no resulta trivial, por lo que se propondrá y discutirá una metodología que permita normalizar la escala de ambos estimadores durante la aplicación del test, obteniendo un vector de parámetros λ representativo del vector θ inicial.

De lo anterior se propone la hipótesis nula, que corresponde al escenario donde ambos estimadores son iguales y por ende se presume que no existen fuentes causantes de endogeneidad.

$$H_0: \tilde{\lambda} = \hat{\lambda} \quad (3.1)$$

Para verificar la hipótesis de consistencia (Hausman y McFadden, 1984) debe cumplirse bajo (H_0) que ambos estimadores convergen en probabilidad al valor real:

$$\tilde{\lambda} \xrightarrow{\mathbb{P}} \hat{\lambda} \quad (3.2)$$

Lo anterior se prueba aprovechando el hecho de que la diferencia entre vectores puede ser explicada a partir de la diferencia entre matrices de varianza-covarianza de cada estimador (Hausman, 1978), sin mediar la correlación entre estimadores.

3.2.2. Normalización de la escala en modelos Bid

El desafío que plantea esta etapa es obtener un parámetro normalizado que logre identificar el impacto de cada estimador. Para ello, se adelantó en la parte anterior que es posible obtener un vector de parámetros normalizados λ para cada estimador del modelo Bid, permitiendo que sean comparados utilizando la prueba de Hausman-McFadden.

Antes de proceder a definir λ como tal, se estudiarán los parámetros identificados por los estimadores del modelo Bid mediante un ejemplo. Para ello, considere el siguiente modelo ejemplo a estimar, formulado para un hogar de estrato $h \in H$ y una vivienda $i \in V$:

$$\tilde{B}_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \varepsilon_{hi} \quad (3.3)$$

Las características de la vivienda i no dependen del hogar, por lo que las diferencias entre estratos de hogares quedarán capturadas por: ASC_h , $\beta_{sup,h}$ y $\beta_{dist,h}$, obteniendo, por ejemplo, un mínimo de seis parámetros a estimar para el caso de una muestra segmentada en dos estratos socioeconómicos.

Considere la expresión de probabilidad para logit multinomial de Ellickson (1981) (Ecuación 2.3). Dividiendo el numerador y denominador por $\exp(\mu B_{hi})$ y separando la suma del

denominador se logra obtener la siguiente expresión:

$$P(h|i) = \frac{1}{1 + \sum_{h' \in H, h' \neq h} \exp(\mu(B_{hi} - B_{h'i}))} \quad (3.4)$$

Se nota que dicha expresión solo captura las diferencias observadas entre estratos ($B_{hi} - B_{h'i}$). Al desarrollar este último término se obtiene:

$$\mu(B_{hi} - B_{h'i}) = \mu(ASC_h - ASC_{h'}) + \mu(\beta_{sup,h} - \beta_{sup,h'}) \cdot sup_i + \mu(\beta_{dist,h} - \beta_{dist,h'}) \cdot dist_i \quad (3.5)$$

Así, Ellickson estima los valores de $\mu(ASC_h - ASC_{h'})$, $\mu(\beta_{sup,h} - \beta_{sup,h'})$ y $\mu(\beta_{dist,h} - \beta_{dist,h'})$, por lo que la escala debe ser normalizada cuando se comparan valores estimados por el propio modelo.

Lerman y Kern, por otra parte, permite estimar de forma independiente la escala del modelo, como fue constatado en la revisión bibliográfica mediante el análisis del término de log-verosimilitud.

$$\ell = -|V| \ln(\mu) - \sum_{i \in V} y_{hi} \left((\mu(P_i^* - B_{hi})) - \sum_{h' \in H} (\exp(-\mu(P_i^* - B_{h'i}))) \right) \quad (3.6)$$

De la expresión (3.6) es posible identificar dos estimadores: $\hat{\mu}$ y $\mu(P_i^* - B_{h'i})$. Además, el último término puede ser desarrollado utilizando (3.3) y, puesto que P_i^* es conocido, se logra identificar de forma particular todos los términos de B_{hi} : ASC_h , $\beta_{sup,h}$ y $\beta_{dist,h} \forall h \in H$. Finalmente, con los valores estimados es posible reconstruir los estimadores utilizados en Ellickson: $\mu(ASC_h - ASC_{h'})$, $\mu(\beta_{sup,h} - \beta_{sup,h'})$ y $\mu(\beta_{dist,h} - \beta_{dist,h'})$, que permiten construir el parámetro λ .

Se concluye que cada estimador entrega la diferencia observada entre estratos socioeconómicos ($h \in H$) para ASC_h , $\beta_{sup,h}$ y $\beta_{dist,h}$. Queda pendiente normalizar el efecto de la escala μ en el modelo de Ellickson, para lo que se propone estimar la razón entre dos parámetros de la diferencia observada entre estratos.

$$\lambda = \frac{\beta_{dist,alto} - \beta_{dist,bajo}}{\beta_{sup,alto} - \beta_{sup,bajo}} \quad (3.7)$$

La expresión anterior muestra, a modo de ejemplo, la razón de la diferencia entre los atributos $\beta_{sup,h}$ y $\beta_{dist,h}$ presentes en la ecuación 3.3. No obstante, es posible plantear este parámetro con cualquier otra combinación de parámetros. La dimensión de λ depende directamente de la segmentación del conjunto H , siendo equivalente a $|H| - 1$, que es el número de estratos definidos menos uno, el cual se usa para identificar la diferencia entre estratos.

Para estimar el término λ basta con definir $\hat{\lambda}$ a partir de los valores estimados en Ellickson y Lerman y Kern:

$$\hat{\lambda} = \frac{\mu(\beta_{dist,h} - \beta_{dist,h'})}{\mu(\beta_{sup,h} - \beta_{sup,h'})} \quad (3.8)$$

donde el término μ se cancela y con ello recupera el valor λ buscado.

El estimador propuesto puede ser calculado en cualquiera de los escenarios planteados, permite detectar las diferencias observadas entre estratos y normaliza los efectos producto de la escala. Así, cuando el modelo es consistente **tanto la estimación de Ellickson como de Lerman y Kern arrojarán el mismo valor de λ** y reflejarán la variación de eficiencia mediante el análisis de su varianza.

Para ilustrar la validez de las aseveraciones anteriores se realizó una simulación de Monte Carlo, donde se calculó λ para cada estimador y se comparó este con el valor real para el escenario donde no existe endogeneidad ($\beta_{ruido,alto} = \beta_{ruido,bajo} = 0$). Se realizaron 1000 simulaciones, obteniendo un error estándar menor al 1%. Los detalles sobre la formulación de esta simulación son presentados en el capítulo cuarto.

Tabla 3.1: Simulación parámetro lambda – Resumen de resultados para el caso sin endogeneidad

Estimador	Valor estimado	Desviación estándar
Ellickson	-4,28	0,05
Lerman & Kern	-4,29	0,02
Valor real	-4,286	

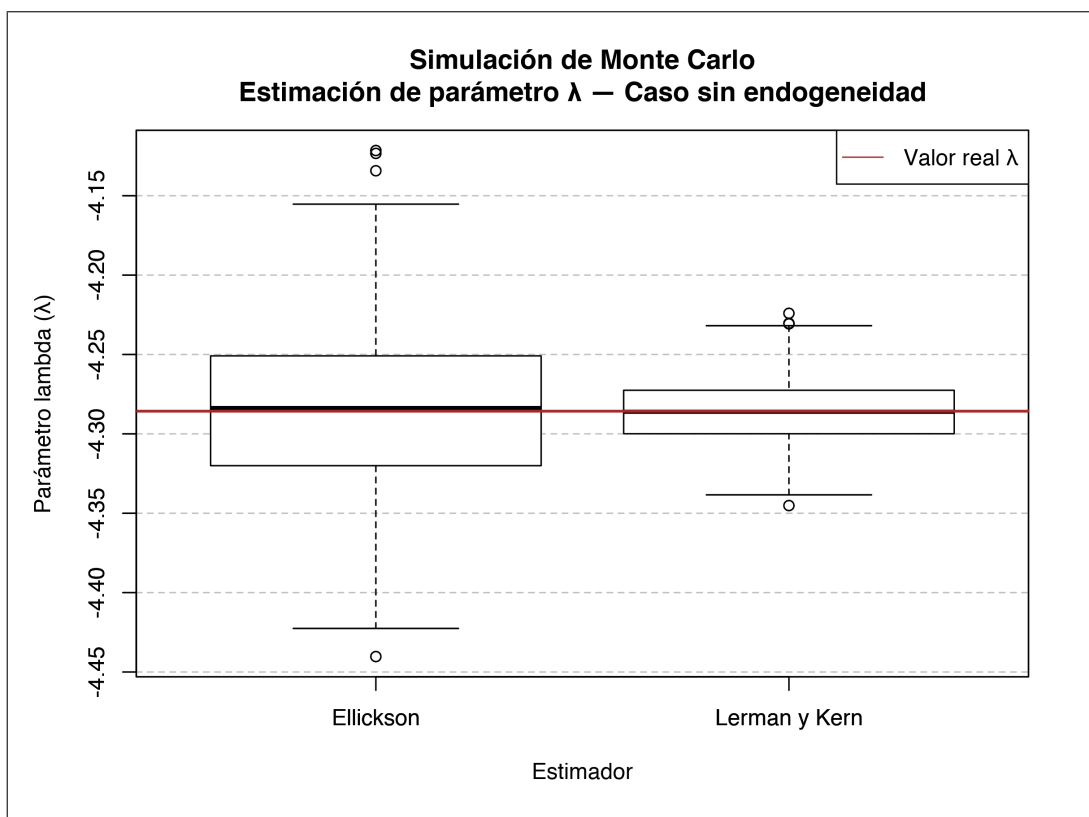


Figura 3.1: Simulación parámetro lambda – Caso sin endogeneidad. La línea marrón representa el valor real de λ sobre el gráfico de caja cuando no hay endogeneidad.

La tabla 3.1 resume los resultados obtenidos en esta simulación, destacando que la estimación de Lerman y Kern muestra mayor precisión. El gráfico de caja (Figura 3.1) muestra que la mediana de los datos obtenidos por cada estimador es idéntica al parámetro real, validando la consistencia del valor estimado. También se recalca que la varianza obtenida a partir del estimador de Lerman y Kern es menor, lo que es consistente con las hipótesis planteadas sobre la eficiencia de cada estimador.

Finalmente, con el fin de ilustrar el efecto de un escenario de endogeneidad sobre este parámetro, se supone la existencia de un atributo no considerado en el proceso de estimación (e.g. ruido), el cual presenta correlación con otro atributo de la función (3.3) y por ende causa correlación entre esta función y su término de error. El atributo omitido reduce uniformemente las posturas los hogares de la muestra, de modo que las decisiones de localización no cambian, pero si cambia el precio final pagado. Recordando que Ellickson no utiliza precios observados para estimar (3.3), mas si Lerman y Kern, se espera el primero siga siendo consistente en términos de λ y que Lerman y Kern se vea perjudicado por esta información.

Tabla 3.2: Simulación parámetro lambda – Resumen de resultados para el caso endógeno

Estimador	Valor estimado	Desviación estándar
Ellickson	-4,28	0,05
Lerman & Kern	-4,2	0,02
Valor real	-4,286	

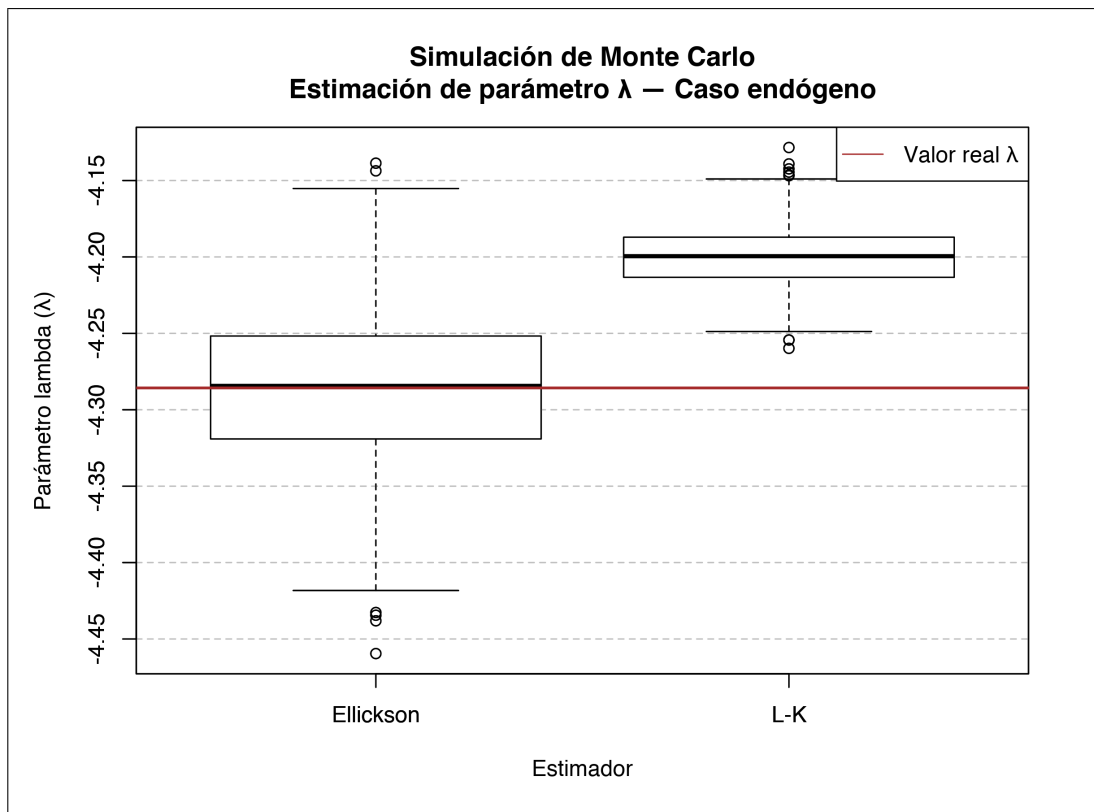


Figura 3.2: Simulación parámetro lambda – Caso endógeno. La línea marrón representa el valor real de λ sobre el gráfico de caja cuando no hay endogeneidad.

La tabla 3.2 muestra una clara diferencia en el valor estimador por Lerman y Kern, confirmando la hipótesis inicial. No obstante, la eficiencia del estimador se mantiene constante. La figura 3.2 ilustra la alteración observada en el estimador de Lerman y Kern, mientras que Ellickson no presenta cambios. Esto muestra que el parámetro λ es capaz de detectar diferencias de especificación si los datos utilizados por uno de los estimadores se encuentran alterados. Este escenario será estudiado en mayor detalle en el capítulo cuarto, referente a la simulación de Monte Carlo, la evaluación de *size distortion* y poder empírico del test.

3.2.3. Evaluación del test

Para evaluar la prueba de Hausman-McFadden modificada se define el estadístico H (Ecuación 3.9), que corresponde a la versión unidimensional del test original propuesto en Hausman y McFadden (1984), por lo que es válido cuando la dimensión de $\tilde{\theta}$ y $\hat{\theta}$ es igual a 1.

$$H = \frac{(\tilde{\lambda} - \hat{\lambda})}{\sqrt{Var_{\tilde{\lambda}} - Var_{\hat{\lambda}}}} \sim t \quad (3.9)$$

Para aceptar o rechazar la hipótesis nula se compara el valor obtenido de H con el valor crítico observado para los grados de libertad deseados. En esta implementación se tiene **un grado de libertad**, por lo que se establece el valor crítico de rechazo en $H = 1,96$ para un nivel de confianza del 95 %.

3.3. Test de Funciones de Control

Con el objetivo de establecer una base de comparación para el test de Hausman-McFadden modificado se propone la aplicación de una prueba alternativa, basada en un enfoque probado como son las Funciones de Control (Heckman, 1977; Villas-Boas y Winer, 1999). Esta prueba requiere una serie de supuestos no contemplados en el test desarrollado, por lo que se procede a describir los fundamentos del método y la prueba de hipótesis para detección de endogeneidad asociada.

3.3.1. Consideraciones previas

El enfoque de funciones de control tiene por objetivo detectar y corregir endogeneidad que presenta un modelo particular valiéndose de la eventual existencia de variables instrumentales (Villas-Boas y Winer, 1999). Dichas variables deben estar correlacionadas con aquella que presente endogeneidad en el modelo y ser independientes del término de error presente.

La principal dificultad que subyace a este enfoque es la capacidad de encontrar instrumentos válidos. Esto, que es sencillo de obtener en un escenario de simulación como el que propone esta memoria, resulta no trivial en casos reales, donde se suele buscar variables que expliquen atributos de carácter cualitativo tales como confort y seguridad. El uso de instrumentos inadecuados puede afectar significativamente los resultados obtenidos (Bound et al., 1995), por lo que resulta de interés buscar métodos que, como el propuesto en esta memoria, no requieran su uso.

3.3.2. Formulación del test

Considere un modelo de posturas aleatoria que busca definir una función B_{hi} a partir de dos variables: $dist_{hi}$, sup_{hi} y una constante ASC_h . Se sospecha que $dist$ se encuentra sesgado por endogeneidad y se propone un instrumento I_{hi} para corregirla.

$$B_{hi} = ASC_h + \beta_{sup}sup_{hi} + \beta_{dist}dist_{hi} + \varepsilon_{hi} \quad (3.10)$$

Para capturar la información causante de endogeneidad se realiza una regresión lineal de la variable endógena $dist$ a partir del instrumento I y otras variables exógenas (sup). De la regresión se recuperan los residuos δ del modelo.

$$dist = \alpha_{sup}sup_{hi} + \alpha_I I_{hi} + \delta_{hi} \implies \hat{\delta}_{hi} \quad (3.11)$$

Se espera que los residuos del modelo lineal logren capturar los efectos de endogeneidad presentes en la formulación inicial (Heckman, 1977), sin verse afectado por otras causas al incluirse también las variables exógenas. Luego, se puede usar esta información en el modelo inicial y estudiar su significancia estadística.

$$B_{hi} = \beta_{sup}sup_{hi} + \beta_{dist}dist_{hi} + \beta_{\delta}\hat{\delta}_{hi} + \varepsilon_{hi} \quad (3.12)$$

En caso de no existir endogeneidad se espera que β_{δ} sea insignificante, mientras que en caso contrario una prueba de razón de verosimilitud permita afirmar que es no nula. Con esto queda definida la siguiente hipótesis nula (H_0).

$$H_0: \beta_{\delta} = 0 \quad (3.13)$$

3.3.3. Evaluación del test

Para validar o refutar la hipótesis presentada se aplica un **Test de Razón de Verosimilitud** que compara las especificaciones señaladas anteriormente.

$$B_{hi} = \beta_{sup}sup_{hi} + \beta_{dist}dist_{hi} + \beta_{\delta}\hat{\delta} + \varepsilon_{hi} \implies L^{COM} \quad (3.14)$$

$$B_{hi} = \beta_{sup}sup_{hi} + \beta_{dist}dist_{hi} + \varepsilon_{hi} \implies L^{RED} \quad (3.15)$$

La ecuación (3.14) identifica la especificación completa (COM) y (3.15) la especificación reducida (RED). De la estimación de cada modelo se recupera su verosimilitud: L^{COM} y L^{RED} , necesaria para aplicar el test. Finalmente se evalúa el estadístico CF , descrito por la ecuación (3.16).

$$CF = -2(L^{RED} - L^{COM}) \sim \chi_d^2 \quad (3.16)$$

El estadístico CF distribuye chi cuadrado con d grados de libertad, el cual depende de la cantidad de variables adicionales que presenta la especificación completa. En el escenario de simulación propuesto se tiene $d = 2$ ya que la cantidad de variables depende del número de estratos socioeconómicos utilizados. Así, el valor crítico de CF donde la hipótesis nula se rechaza corresponde a $CF = 5,99$.

3.4. Conclusiones del capítulo

Este capítulo presentó el test de Hausman-McFadden modificado, explicando las modificaciones realizadas para su aplicación en el contexto del modelo Bid de posturas aleatorias. En términos de su aplicación, el uso de un parámetro normalizado (λ) resulta crucial, puesto que de otra forma los resultados obtenidos por cada estimador no serían comparables. Finalmente se enunció el test de funciones de control aplicado al mismo modelo de posturas, destacando la hipótesis adicional sobre variables instrumentales, la cual dificulta su aplicación en escenarios reales.

Queda pendiente realizar un análisis de desempeño para la prueba de hipótesis desarrollada, en que se logre estimar la capacidad que posee el test para detectar endogeneidad en varios escenarios. Esto será estudiado mediante simulaciones de Monte Carlo en el siguiente capítulo.

Capítulo 4

Evaluación del test mediante simulación de Monte Carlo

4.1. Presentación

El presente capítulo busca presentar la simulación de Monte Carlo realizada para evaluar la prueba de especificación de Hausman-McFadden modificada y comparar su desempeño con la prueba de Funciones de Control.

El experimento a desarrollar busca replicar las condiciones para producir endogeneidad en el precio final pagado por una vivienda. Se define una función de postura para la simulación con tres atributos, donde dos de ellos se hallan correlacionados. Así, la omisión de uno de estos atributos en una eventual estimación por parte de un investigador afectará el valor estimado de aquel que no fue omitido, causando endogeneidad y cumpliendo el objetivo propuesto. Con esto en consideración, el experimento queda descrito por dos etapas:

1. **Simulación:** Etapa donde se replica un remate inmobiliario considerando una lista de atributos dada. Se obtiene como resultado una muestra de datos que permite estimar modelos de remates.
2. **Estimación:** Se procede a estimar el modelo Bid de posturas aleatorias **omitiendo una variable** usada durante la simulación. Se tendrá como consecuencia directa que el modelo estimado pueda verse afectado por endogeneidad. Con esto se procede a la aplicación de las pruebas de hipótesis que se desean evaluar.

El resultado final del experimento es una realización de cada test evaluado. Mediante la repetición de este se calcularán dos indicadores de desempeño para los test evaluados:

1. **Size distortion:** Que corresponde a la estimación del error tipo I observado cuando se cumple la hipótesis de no endogeneidad.
2. **Poder empírico:** El que se obtiene planteando posibles escenarios de endogeneidad, los que son controlados mediante un parámetro β denominado tamaño del efecto. Con ello se calcula la probabilidad de no cometer un error tipo II para un valor β dado.

La primera sección de este capítulo presenta la metodología y datos empleados para realizar la simulación, describiendo también los escenarios donde la prueba de hipótesis será evaluada. La segunda sección presenta los resultados del análisis por *size distortion*. En la tercera sección se presentan los resultados de la estimación el poder empírico del test para los escenarios propuestos. Finalmente, en la última sección se propone levantar el supuesto de ciudad abierta planteado inicialmente, mostrando los cambios realizados en el proceso de asignación y su impacto en los resultados, en términos del poder empírico estimado.

4.2. Formulación de la simulación

Esta sección describe los pasos necesarios para replicar el proceso de remate inmobiliario desarrollado en esta memoria, lo que permite obtener muestras de datos que permitan evaluar el test propuesto. Así, se describe la muestra de viviendas y hogares, la función de postura asociada y la metodología de asignación utilizada. Para finalizar se describe el proceso de estimación del modelo, esta vez desde el punto de vista de un investigador que omite una variable relevante del modelo, con lo que se proponen dos escenarios que susceptibles a endogeneidad.

4.2.1. Definición de la muestra de viviendas

En esta etapa se definen los atributos característicos de cada vivienda, buscando condiciones que permitan replicar endogeneidad causada por variables omitidas. Con esto en consideración, se propone generar una muestra de viviendas caracterizada por tres atributos, donde uno de ellos sea causante de endogeneidad dada su omisión.

1. **Superficie:** Este cuantifica el tamaño físico del inmueble. Se considera un indicador de calidad de vida y por ende muestra una alta disposición de parte de los hogares a pagar por mejorar este atributo.
2. **Distancia al CBD**¹: Este atributo es característico de los modelos de ciudad monocéntrica, puesto que caracteriza la accesibilidad del inmueble elegido al lugar de trabajo. Así, se observará una relación inversa entre este atributo y la disposición a pagar por la vivienda.
3. **Ruido ambiente:** La elección de este atributo se fundamenta en la dificultad que presenta para su medición en cada vivienda de la muestra, lo que motivaría su omisión por parte de un investigador en una aplicación real.

Por otra parte, es válido suponer que los centros urbanos presentan mayores niveles de ruido que la periferia, por lo que existiría correlación entre este atributo y la distancia al centro. Así, la omisión de este atributo resultaría en una eventual fuente de endogeneidad, debido a una sobrestimación de la disposición a pagar por localizarse más cerca del centro.

¹Del inglés *Center Business District*, hace referencia al centro de la ciudad. Visto en Alonso et al. (1964).

4.2.2. Definición de la muestra de hogares

Esta etapa busca definir los hogares que participarán del remate de viviendas, replicando las condiciones que se pueden observar en la realidad. Cada hogar participante ($h \in H$) posee una disposición a pagar marginal por atributo observado $k \in K$ que denominamos $\beta_{k,h}$. Dado que la muestra de hogares puede ser tan grande como la de viviendas, resulta poco práctico proponer un parámetro para cada hogar. Por lo tanto, se propone que la muestra de hogares sea segmentada por estratos.

Así, se divide la muestra en dos categorías de hogar: Estrato alto y estrato bajo. Estos agrupan la preferencia del total de hogares, facilitando la estimación de $\beta_{k,h}$ en el modelo. El estrato alto presentará mayor disposición a pagar por localizarse en viviendas de mayor tamaño, cercanas al centro y con bajo nivel de ruido; mientras que el estrato bajo tenderá a pagar menos por dichos atributos.

4.2.3. Cálculo de posturas

En esta etapa se calcula la disposición a pagar de cada hogar participante del remate por cada vivienda. Se busca incorporar efectos aleatorios que capturen aquellos elementos no observados por el investigador, además de las variables superficie, distancia y ruido.

Se propone una función de postura aleatoria \tilde{B}_{hi} , correspondiente a lo ofertado por un hogar de estrato $h \in H$ para adquirir una vivienda $i \in V$.

$$\tilde{B}_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \beta_{ruido,h} \cdot ruido_i + \varepsilon_{hi} \quad (4.1)$$

La función de postura propuesta es una adaptación de Martínez y Henríquez (2007). El primer término (ASC_h) agrupa los efectos del ingreso y utilidad percibidas por el hogar, mientras que los siguientes términos representan la valoración de atributos individuales de la vivienda rematada. El término ε_{hi} representa la componente aleatoria de la postura, que por conveniencia se asumió distribuye Gumbel con parámetro de localización 0 y escala 1, siendo independiente e idénticamente distribuido sobre hogares y viviendas.

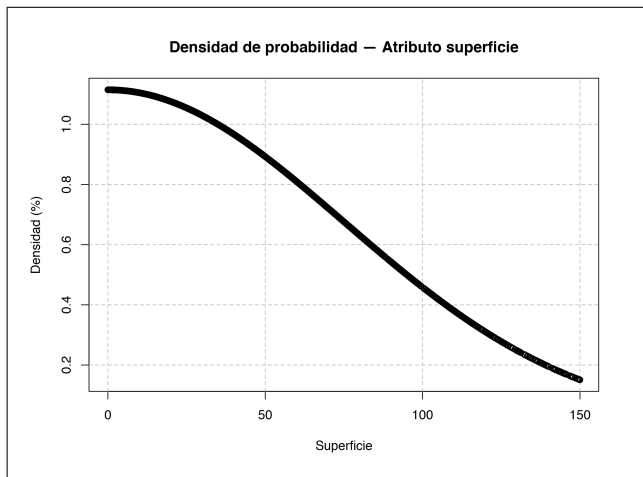
4.2.4. Generación de datos

Una vez definidas las etapas anteriores, se procede a generar la muestra de datos que será usada para evaluar las pruebas de hipótesis en la simulación. Estos serán generados mediante una muestra aleatoria, para lo cual es necesario definir las distribuciones de probabilidad a utilizar, el nivel de correlación deseado entre atributos y los parámetros de cada atributo por estrato socioeconómico.

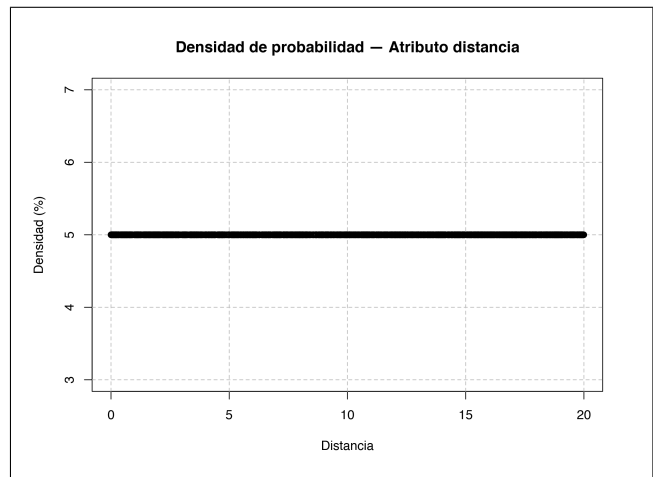
Cada simulación genera una lista de $N = 2000$ observaciones, que representan un conjunto de viviendas y su asignación a cada hogar. Adicionalmente y producto de los requisitos impuestos por el método de funciones de control ya discutidos en la sección 3.3, se genera una variable instrumental que será utilizada en la aplicación del test de Funciones de Control.

Distribución de probabilidad

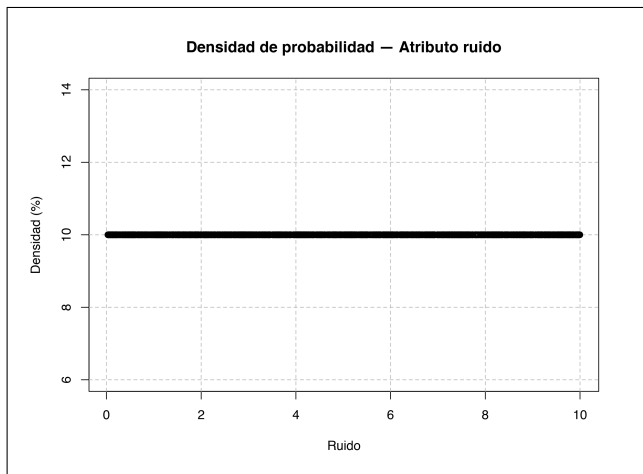
La distribución y parámetros característicos de cada atributo se presentan en la tabla 4.1. La figura 4.1 ilustra la distribución de los atributos generados para una iteración arbitraria.



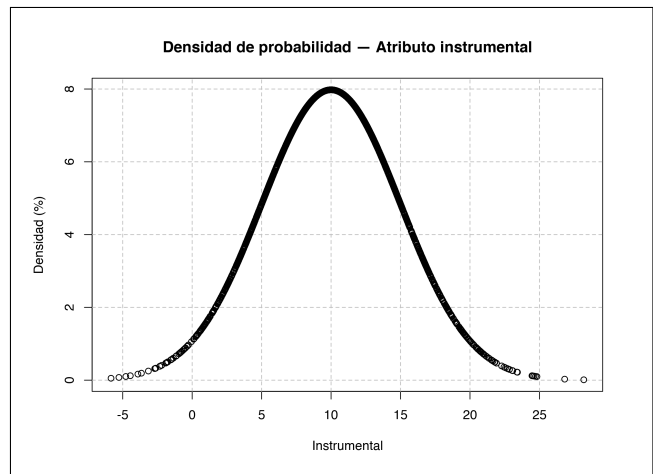
(a) Superficie



(b) Distancia



(c) Ruido



(d) Instrumental

Figura 4.1: Función densidad de probabilidad observada para los atributos muestreados.

Tabla 4.1: Distribución de probabilidad en los atributos propuestos

Atributo	Distribución	Parámetros
1. Superficie	Normal truncada	$a=0, b=100, \mu=0, \sigma=75$
2. Distancia al CBD	Uniforme	$a=0, b=20$
3. Ruido	Uniforme	$a=0, b=10$
4. Instrumental	Normal	$\mu=10, \sigma=5$

Correlación entre atributos

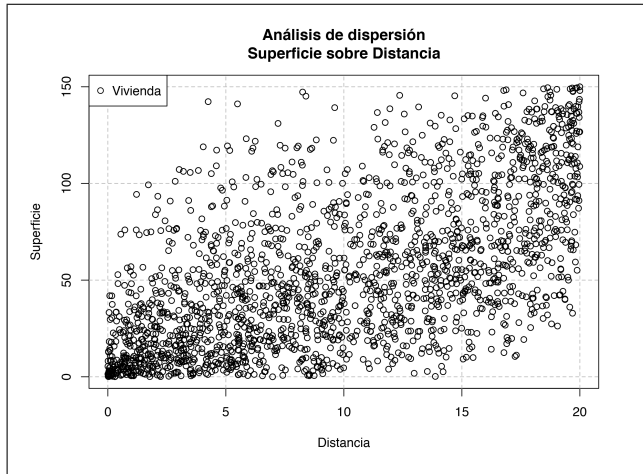
Con el fin de poder controlar la correlación entre atributos de cada vivienda, se utilizó el paquete de R «copula» (Hofert et al., 2017), el cual permite generar vectores aleatorios multivariados usando una matriz de correlaciones definida de antemano. La tabla 4.2 muestra la matriz de correlaciones propuesta para generar los datos y la calculada a partir de una iteración arbitraria.

Tabla 4.2: Matriz de correlación entre atributos

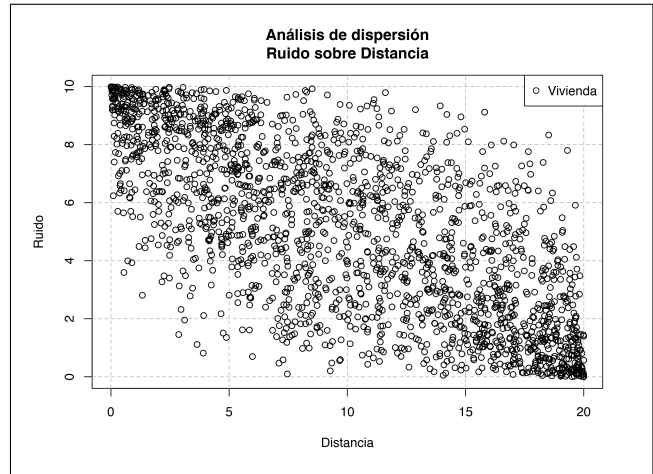
Atributo	1	2	3	4
Propuesto				
1. Superficie	—	0,8	0	0
2. Distancia al CBD		—	-0,9	0,5
3. Ruido			—	0
4. Instrumental				—
Observado				
1. Superficie	—	0,62	-0,06	0,04
2. Distancia al CBD		—	-0,7	0,4
3. Ruido			—	-0,03
4. Instrumental				—

El principal análisis derivado de esta tabla se encuentra en las diferencias observadas entre la matriz propuesta y la observada luego de generar los datos. En esta última se nota que las correlaciones tienen una magnitud menor que la deseada, lo que puede atribuirse a las limitaciones del paquete compatibilizando múltiples variables con distintas distribuciones de probabilidad. No obstante, el resultado obtenido es suficiente para los objetivos buscados en la simulación al presentar signos adecuados y magnitudes de correlación aceptables.

La correlación entre superficie y distancia al CBD (Figura 4.2a) busca replicar lo observado en un ciudad, donde viviendas de mayor tamaño se localizan a mayor distancia del centro. Por otra parte, la correlación entre distancia y ruido ambiente (Figura 4.2b) está fundamentada en la alta concentración de puestos de trabajo que posee el CBD, lo que repercute en una mayor actividad en dicha zona. Finalmente, la correlación que presenta el atributo instrumental fue propuesta para cumplir los supuestos que requiere la aplicación de la prueba de funciones de control. En este sentido se tiene un atributo que logra explicar bien la variable latente, por lo que se espera un resultado satisfactorio de esta prueba.



(a) Superficie sobre Distancia



(b) Ruido sobre Distancia

Figura 4.2: Análisis de dispersión para muestra de viviendas.

Parámetros de la función de postura

La tabla 4.3 muestra los parámetros utilizados para evaluar la función de postura (Ecuación 4.1). Cabe destacar que el atributo ruido no queda definido en este punto ya que depende del escenario de endogeneidad que sea evaluado.

Tabla 4.3: Parámetros por estrato

Parámetros β por estrato	Bajo	Alto
0. Constante (ASC)	0	0
1. Superficie	0,40	0,75
2. Distancia al CBD	-0,5	-2
3. Ruido	Variable (ver 4.2.6)	

Un punto a destacar de la tabla anterior es la definición de constantes específicas por estrato como nulas. Si bien se reconoce que su omisión afecta la asignación de viviendas en como se reflejan elementos como el ingreso y utilidad percibida; aquello no influye en los resultados de este estudio, de modo que los conclusiones obtenidas siguen siendo válidas. Así, por simplicidad se optó por su omisión.

4.2.5. Asignación de viviendas

En esta etapa se procede a comparar las posturas calculadas para cada estrato y con ello se aplica una regla de decisión que determina el hogar ganador del remate. Como resultado adicional se obtendrá el precio final de la vivienda, definido igual a la máxima postura observada en el remate.

Para llevar a cabo lo anterior es necesario definir la cantidad de hogares oferentes en cada estrato, puesto que el resultado del remate dependerá de ello. Existen dos enfoques clásicos para ciudades monocéntricas, los que serán descritos y discutidos a continuación.

Ciudad abierta

Este escenario de asignación se fundamenta en la posibilidad de migrar desde y hacia otras ciudades «sin costo adicional» (Brueckner, 1987), de modo que residentes externos pueden localizarse en esta ciudad. En términos económicos existe un nivel de utilidad exógeno al modelo, el cual permite determinar las condiciones de equilibrio para esta ciudad.

La simulación de un remate para ciudad abierta puede realizarse de forma agregada por estrato, lo que supone la existencia de infinitos oferentes pertenecientes a cada uno. Así, en cada remate compite un hogar representativo de cada estrato y se determina el estrato ganador de cada vivienda. Este enfoque redundante en que la distribución de la muestra de hogares simulada queda determinada por el propio remate, permitiendo que esta distribución cambie debido a las características de las viviendas y la disposición a pagar de cada estrato.

Lo explicado anteriormente puede observarse en la realidad cuando se analiza una localidad particular o comuna dentro de un área urbana, como puede ser el Gran Santiago, por lo que una muestra de hogares obtenida con este enfoque puede replicar las características propias de un sector de la ciudad. También, en términos de tiempo, los análisis de carácter estratégico tienden a considerar ciudades abiertas para modelar dado que poseen un enfoque de largo plazo.

Ciudad cerrada

Este escenario, contrario a una ciudad abierta, está cerrado a migraciones, por lo que la cantidad de hogares y su distribución socioeconómica es conocida de antemano. En términos económicos, la utilidad percibida por los hogares queda determinada por el equilibrio de mercado, dada la existencia de población cautiva. Este enfoque resulta práctico en escenarios donde el efecto de migraciones tenga una relevancia menor o nula, como puede ocurrir cuando se analiza una ciudad en términos tácticos o de corto plazo, o si se estudia el comportamiento de hogares al adquirir vivienda dentro de un condominio.

La aplicación de este enfoque a un remate inmobiliario requiere supuestos más sofisticados que el caso de ciudad abierta. En primer lugar, es necesario caracterizar de forma individual a cada hogar participante del remate, esto ya que la cantidad de participantes es conocida antes de la realización de cada remate. Luego, debido a que existe la misma cantidad de hogares y viviendas, es claro que el orden de realización del remate repercutirá en el resultado de la asignación cuando se realiza de forma secuencial. Este punto puede ser abordado mediante modelos de equilibrio como MUSSA (Martinez, 1996) o utilizando métodos de microsimulación como los ILUTE (Miller et al., 2004) o Urbansim (Waddell et al., 2003). No obstante, esto último supera los objetivos alcances del trabajo realizado y por ello será tratado de forma simplificada cuando se evalúe el test bajo este enfoque de asignación.

Conclusiones

Del análisis de metodologías de asignación se concluye que el enfoque de ciudad abierta logra cumplir con los objetivos buscados en esta etapa, pese a la limitación de no poder definir la distribución de hogares por estrato deseada. La figura 4.3 ilustra los resultados de una asignación arbitraria para el escenario sin endogeneidad, destacando la presencia de una línea de indiferencia que es equivalente a la ecuación (4.1).

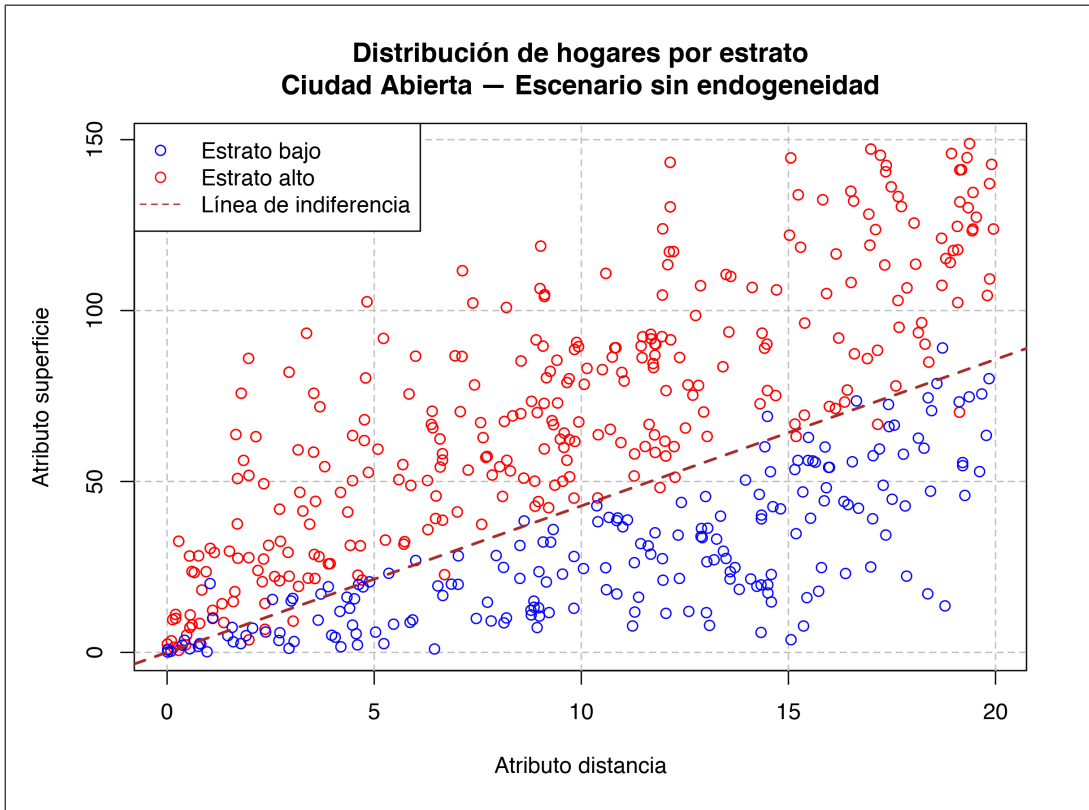


Figura 4.3: Distribución de hogares por estrato simulada mediante ciudad abierta para el caso sin endogeneidad. Destaca la presencia de una línea de indiferencia que divide en términos generales a la muestra de hogares.

Por otra parte, el enfoque de ciudad cerrada, que resulta interesante de estudiar por los supuestos que levanta, presenta desafíos adicionales que están fuera del alcance de este trabajo, por lo que se propone realizar una revisión simplificada para evaluar su comportamiento. Otra diferencia a destacar sobre el enfoque de ciudad abierta radica en que la simulación de este escenario no garantiza una condición de equilibrio de precios, por lo que estos pueden variar puntualmente respecto al enfoque de ciudad abierta.

Finalmente, como resultado del proceso de asignación se obtiene una muestra de datos que permita estimar el modelo Bid de posturas aleatorias. La estimación del modelo permitirá evaluar los test de hipótesis propuestos en esta memoria.

4.2.6. Estimación de modelos y escenarios de endogeneidad

Como paso final y utilizando la muestra de datos generada a partir de la simulación, se procede a estimar los modelos de Ellickson y Lerman y Kern. Este paso supone que el investigador ignora los efectos del ruido para estimar el modelo y utiliza la expresión propuesta en (4.2). Así, se espera que el nuevo término de error $\tilde{\varepsilon}_{hi}$ se vea afectado por la omisión del término de ruido, estableciendo una correlación con el parámetro asociado a distancia.

$$\tilde{B}_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \underbrace{\beta_{dist,h} \cdot dist_i + \tilde{\varepsilon}_{hi}}_{\beta_{ruido,h} \cdot ruido_i + \varepsilon_{hi}} \quad (4.2)$$

Debido a la correlación presente entre los atributos de distancia y ruido, se espera que los efectos de endogeneidad sean observados principalmente en el parámetro de distancia. El efecto de la omisión del atributo ruido dependerá de la influencia que este tenga en la elección final de cada hogar. Para controlar dicha influencia se propone hacer variar el parámetro asociado al ruido durante la simulación, partiendo desde un escenario base donde no existe influencia y por ende se espera que el modelo estimado no presente endogeneidad.

A continuación se describen los dos escenarios estudiados para estimar el poder de cada test, señalando los pros, contras y consecuencias esperadas en cada caso.

Escenario A: Beta ruido igual por estrato

El primer escenario propuesto utiliza el atributo omitido para modificar el precio final pagado por cada inmueble de forma uniforme sobre los hogares de la muestra, siendo independiente del estrato al que pertenezcan. Se utiliza la variable β_{ruido} para cuantificar el tamaño del efecto que permite estudiar el test de hipótesis. Un cambio en el valor de β_{ruido} hará que los precios finales por la vivienda sean menores que lo esperado si dicho efecto no se considera. Este escenario es extrapolable a otras situaciones, como la aplicación de subsidios sobre la propiedad que no dependan del tipo de hogar que se adjudica el remate.

Para simular este escenario se hace variar el parámetro β_{ruido} de ambos estratos en igual magnitud, como se muestra en 4.3. Es importante notar el valor de β propuesto siempre será negativo, con el fin que represente un perjuicio para el hogar, reduciendo las posturas por cada vivienda.

$$\beta = \beta_{ruido,alto} = \beta_{ruido,bajo} \leq 0 \quad (4.3)$$

La condición inicial $\beta = 0$ deriva en que no exista diferencias entre la simulación y el modelo real estimado ($\tilde{\varepsilon}_{hi} = \varepsilon_{hi}$), de modo que bajo esta condición no existe endogeneidad en la función de postura por omisión de variables.

A medida que se incrementa la magnitud del parámetro β , se espera que las posturas de cada estrato cambien en igual magnitud. Es por esto que, pese a la existencia de endogeneidad, la asignación de hogares en ciudad abierta no se ve afectada. Por otro lado, los cambios en la postura final provocan endogeneidad en el precio ofertado, alterando los resultados obtenidos por el estimador de Lerman y Kern al sobrestimar los efectos del atributo distancia. El modelo de Ellickson no se ve afectado por lo anterior ya que este solo utiliza la elección de vivienda y atributos de cada hogar para realizar la estimación, las que no cambian en este escenario.

Escenario B: Beta ruido distinto por estrato

El segundo escenario propuesto corresponde a una variación del primer caso, donde el cambio en el precio final pagado por el inmueble solo es percibido por uno de los estratos. Se decide en esta instancia que solo los hogares de estrato alto reduzcan su disposición a pagar mediante el parámetro β_{ruido} . El análisis realizado resulta válido también para el escenario inverso, es decir que el cambio solo sea percibido por hogares de estrato bajo, y podría deberse a motivos ajenos al ruido como puede ser un subsidio para la adquisición de viviendas.

La simulación de este escenario queda descrita en las expresiones (4.4) y (4.5), donde β representa el tamaño del efecto que será considerado al evaluar los test. Nuevamente los valores de β serán negativos por los motivos ya descritos en el escenario anterior.

$$\beta = \beta_{ruido,alto} \leq 0 \tag{4.4}$$

$$\beta_{ruido,bajo} = 0 \tag{4.5}$$

La condición inicial $\beta = 0$ provoca que, al igual que el Escenario A, no exista endogeneidad por variable omitida, lo que está condicionado también a que $\beta_{ruido,bajo}$ sea nulo pues en caso contrario se observará lo mismo que en el primer escenario. Fijar el valor de $\beta_{ruido,bajo}$ en cero también implica que solo se aprecian cambios en la posturas de hogares pertenecientes al estrato alto, reduciendo su disposición a pagar. Así, a medida que la magnitud de β aumenta, **se observará un cambio en la asignación de hogares por estrato bajo en ciudad abierta**, donde aquellos de estrato bajo lograrán competir en remates de viviendas ante posturas cada vez menores por parte de hogares del estrato alto. Esto continuará hasta alcanzar un escenario degenerado, donde todas las viviendas de la ciudad serán adjudicadas a hogares de estrato bajo.

4.3. Evaluación de la distorsión de tamaño o *size distortion*

La presente sección busca evaluar la presencia de *size distortion* en las pruebas evaluadas. Esto requiere simular una determinada cantidad de realizaciones del test para un tamaño de muestra (N) dado, calcular las veces que el test arroja un error tipo I y comparar la tasa de

error empírico con la nominal. De la revisión bibliográfica se recuerda que error tipo I o falso positivo es cuando la hipótesis nula (H_0) se cumple y el test de hipótesis rechaza H_0 , por lo que en el escenario simulado siempre debe cumplirse H_0 .

Del experimento anterior se espera que entregue una tasa de error tipo I en torno al nivel de significación definido α (Ellis, 2010). Si la tasa de rechazo supera α se concluye que existe distorsión de carácter «liberal», mientras que si es menor se habla de la existencia de una distorsión de carácter «conservador». Se considera que un tamaño muestral N es *acceptable* cuando no se observan distorsiones de carácter liberal.

4.3.1. Definición del escenario simulado

Considere la función de postura definida para la simulación (4.1) y la definida para la estimación (4.2). El término $\beta_{ruido,h}$ sirve como parámetro para desencadenar endogeneidad en el modelo estimado, por lo que estableciendo $\beta_{ruido,h} = 0 \forall h \in H$ se tendrá un escenario que siempre cumple la hipótesis nula.

Posteriormente se define una serie de valores de N a ser evaluados: 150, 200, 500, 1.000, 2.000, 5.000 y 10.000. El tamaño mínimo fue fijado en 150 ya que se observaron múltiples fallas en la estimación para valores menores. Por otra parte, el tamaño máximo estudiado es cinco veces mayor que la muestra utilizada finalmente para la estimación del poder empírico ($N = 2000$), ya que se busca estudiar si existe alguna mejora apreciable producto de ello.

Para estimar el error tipo I empírico en cada escenario se simulan 50.000 aplicaciones del test, agrupadas en 500 muestras de 100 realizaciones del test. Esto tiene por objetivo garantizar que el error de muestreo² sea bajo, logrando valores inferiores al 2%. Para cada muestra se calcula la cantidad de rechazos del test, que en este escenario representan falsos positivos.

4.3.2. Resultados obtenidos

A partir del conjunto de simulaciones realizadas se elabora la Tabla 4.4 que resume los resultados de la simulación. Un análisis en detalle para cada test se presenta a continuación.

Test de Hausman-McFadden modificado

La figura 4.4 ilustra mediante un gráfico de caja el rango de valores observados para el error tipo I en la prueba de Hausman-McFadden modificada, señalándose con una línea marrón el nivel de significación elegido para el test.

²Este valor se calcula dividiendo la desviación estándar sobre la raíz del número de valores muestreados (500). La diferencia porcentual es calculada sobre la media de los datos.

Tabla 4.4: Evaluación de *size distortion*. Nivel de significación nominal del 5 %.

Test	Hausman-McFadden		Funciones de Control	
	Media	Desviación estándar	Media	Desviación estándar
150	6	2,4	6	2,4
200	6	2,2	6	2,3
500	5	2,3	5	2,4
1000	5	2,3	5	2,2
2000	5	2	5	2,3
5000	5	2,2	5	2,2
10000	5	2,2	5	2,3

Error tipo I nominal (α) del 5 %. 50.000 repeticiones.

Nivel de precisión (error estándar) del 2 %.

De los resultados obtenidos se concluye que la prueba de Hausman posee una media de error igual o menor al 5 % a partir de $N = 500$. Se observa además que a partir de $N = 200$ se tiene una mediana igual al 5 %, es decir que la cantidad de casos conservadores detectados igual a la cantidad de casos liberales observados. También se puede notar la menor varianza observada se logra en $N = 2000$, por lo que no se justifica el uso de valores mayores.

Test de Funciones de Control

La figura 4.5 muestra mediante un gráfico de caja los valores observados para el error tipo I en el test de Funciones de Control.

Este caso presenta leves diferencias sobre lo observado para el test anterior. Destaca el hecho de que la varianza permanezca uniforme a medida que aumenta N , por lo que se espera obtener resultados razonables a partir de $N = 500$.

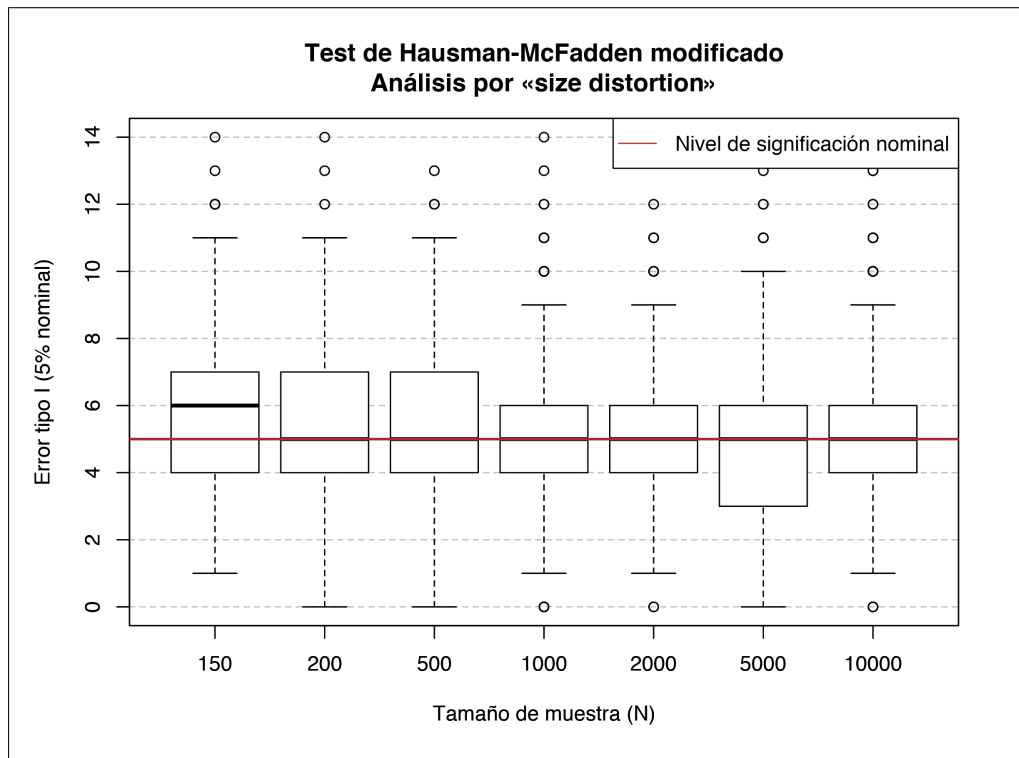


Figura 4.4: Análisis por *size distortion* para la prueba de Hausman-McFadden modificada. La línea marrón indica el error tipo I nominal de la prueba realizada.

Conclusiones

Observando los resultados se concluye que ninguno de los test presentará *size distortion* si se escoge un valor de N lo suficientemente alto, obteniendo resultados aceptables a partir de $N = 500$ en ambos test. Para el caso de una simulación esto debe ser contrastado con los recursos computacionales requeridos, puesto que un tamaño mayor de N incrementa el tiempo requerido para obtener resultados. Tomando en cuenta lo anterior, el tamaño muestral utilizado para estimar el poder del test ($N = 2000$) cumple con ser un equilibrio entre el uso de recursos y la obtención de resultados razonables.

4.4. Evaluación del poder empírico del test

Esta sección presenta los resultados de la estimación del poder estadístico del Test de Hausman-McFadden modificado y el de Funciones de Control bajo los escenarios ya descritos en la sección 4.2.6.

Para llevar a cabo esta evaluación es necesario simular una determinada cantidad de realizaciones del test para un tamaño del efecto dado, que corresponde al parámetro β definido en cada escenario. Del total de simulaciones realizadas se calcula la cantidad de rechazos obtenidos, los que corresponden por construcción a errores de tipo II a excepción del caso $\beta = 0$ que ya fue estudiado. Es útil recordar de la revisión bibliográfica que un error de tipo

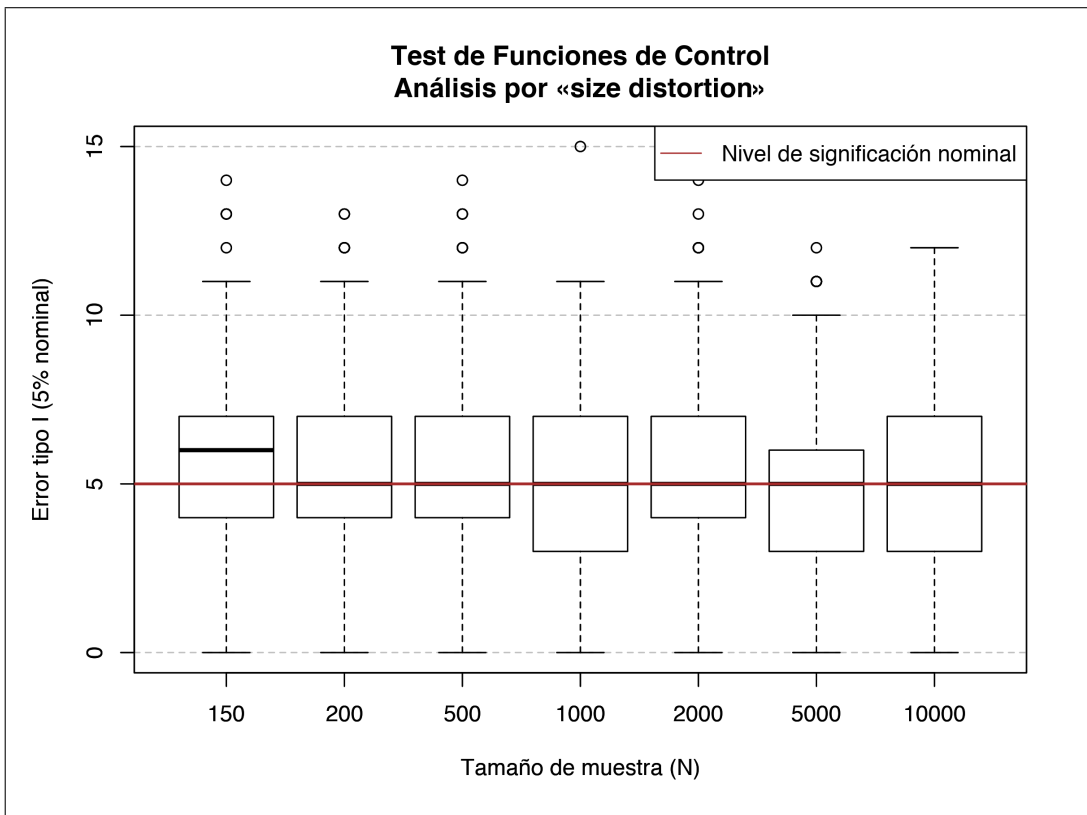


Figura 4.5: Análisis por *size distortion* para la prueba con Funciones de Control. La línea marrón indica el error tipo I nominal de la prueba realizada.

II o falso negativo corresponde a la aceptación de la hipótesis nula (H_0) cuando esta es falsa. También, dada la construcción de escenario realizada en 4.2.6 se sabe que H_0 es falsa cuando $\beta \neq 0$.

4.4.1. Procedimiento general de evaluación

Considere las funciones de postura definidas para la simulación (4.1) y para la estimación del modelo (4.2), donde el término $\beta_{ruido,h}$ es el que determina si el caso simulado presenta endogeneidad. Es directo señalar que si existe h tal que $\beta_{ruido,h} \neq 0$ se observará endogeneidad y por ende no se cumplirá la hipótesis nula.

Para estimar el error tipo II para un tamaño del efecto β dado se simulan 100 aplicaciones del test y se cuenta la cantidad de rechazos de H_0 , obteniendo así el porcentaje de rechazo del test que corresponde al poder del test para dicho valor de β .

4.4.2. Resultados Escenario A: Beta ruido igual por estrato

Recordemos que el escenario A queda definido por el tamaño del efecto $\beta = \beta_{\text{ruido,bajo}} = \beta_{\text{ruido,alto}}$, por lo que ambos estratos de hogares se ven afectados por una variación de precios, sin alterar la distribución de hogares producto de hallarse en una ciudad abierta. Esto afecta exclusivamente al estimador de Lerman y Kern, por lo que se espera que el test de Hausman-McFadden detecte exitosamente este escenario.

La tabla 4.5 presenta un resumen de los resultados obtenidos para el escenario A. La figura 4.6 muestra gráficamente la evolución del poder de las pruebas de Hausman-McFadden y Funciones de Control en este escenario a medida que cambia el tamaño del efecto β .

Tabla 4.5: Resumen de resultados – Estimación de poder empírico – Escenario A

β	Tasa de rechazo por test		Valor λ por estimador		Desviación estándar	
	Hausman-McFadden	Funciones de Control	Ellickson	Lerman y Kern	Ellickson	Lerman y Kern
0	8	5	-4,29	-4,29	0,051	0,020
0,1	8	14	-4,29	-4,27	0,051	0,020
0,2	12	34	-4,28	-4,25	0,051	0,020
0,3	17	72	-4,28	-4,24	0,051	0,020
0,4	29	93	-4,28	-4,22	0,051	0,020
0,5	47	98	-4,28	-4,20	0,051	0,020
0,6	54	100	-4,29	-4,18	0,051	0,020
0,7	72	100	-4,28	-4,16	0,051	0,020
0,8	84	100	-4,28	-4,15	0,051	0,020
0,9	90	100	-4,29	-4,13	0,051	0,021
1	97	100	-4,28	-4,11	0,051	0,021
1,1	98	100	-4,29	-4,09	0,051	0,021
1,2	100	100	-4,29	-4,07	0,051	0,021
1,3	100	100	-4,29	-4,05	0,052	0,021
1,4	100	100	-4,29	-4,03	0,051	0,021
1,5	100	100	-4,28	-4,01	0,051	0,021

N=2000, 10.000 repeticiones, Nivel de significación (α) del 5% y precisión de 5%.

La prueba de Funciones de Control alcanza un rechazo del 100% con $\beta = 0,6$, mientras que Hausman-McFadden obtiene este nivel de rechazo con $\beta = 1,2$, lo que sugiere que la prueba de Funciones de Control posee mayor poder. Sin embargo, este resultado debe ser tomado con cautela, puesto que la simulación fue realizada suponiendo la existencia de un instrumento válido, lo que no puede extrapolarse a la realidad. Es en este punto donde el test de Hausman-McFadden resulta de interés, puesto que no requiere el uso de variables instrumentales para su aplicación.

Un segundo análisis, particular del test de Hausman-McFadden, corresponde al estudio de la media de valores λ observados en cada simulación. La figura 4.7 ilustra la evolución de λ a medida que el impacto del ruido aumenta.

Se logra constatar que el comportamiento del estimador de Lerman y Kern se ve afectado por el cambio en β , mientras que el estimador de Ellickson permanece torno al valor real, validando las hipótesis planteadas cuando se desarrolló el escenario.

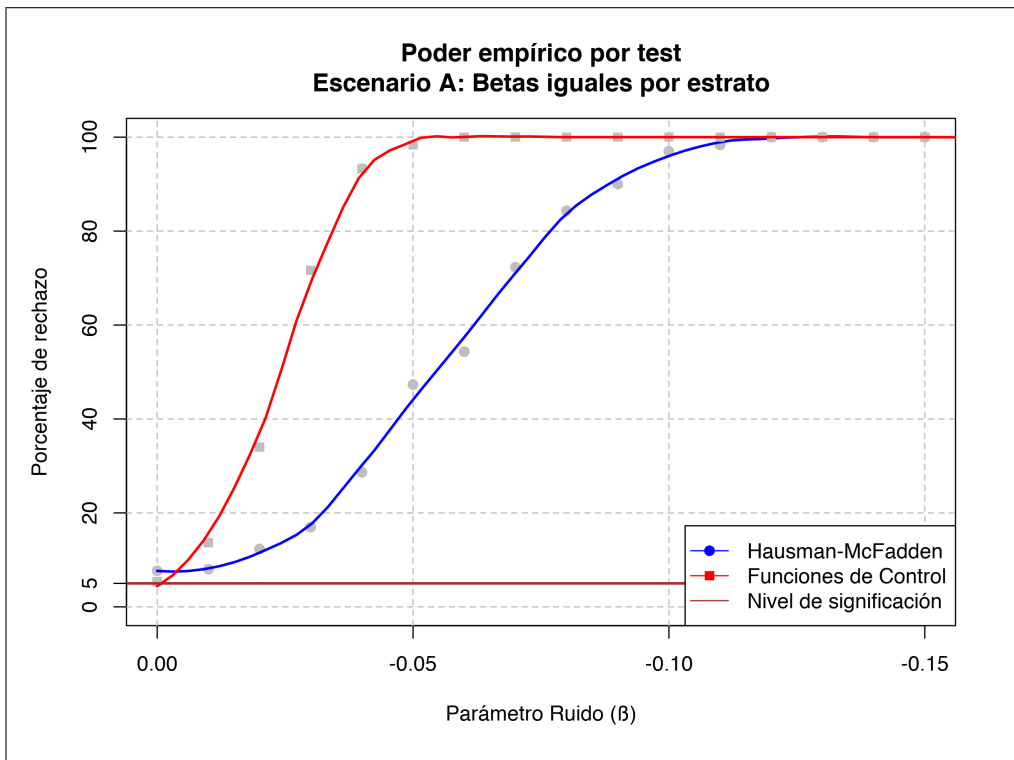


Figura 4.6: Poder empírico observado en Escenario A: Beta ruido igual por estrato. La línea marrón indica el nivel de significación

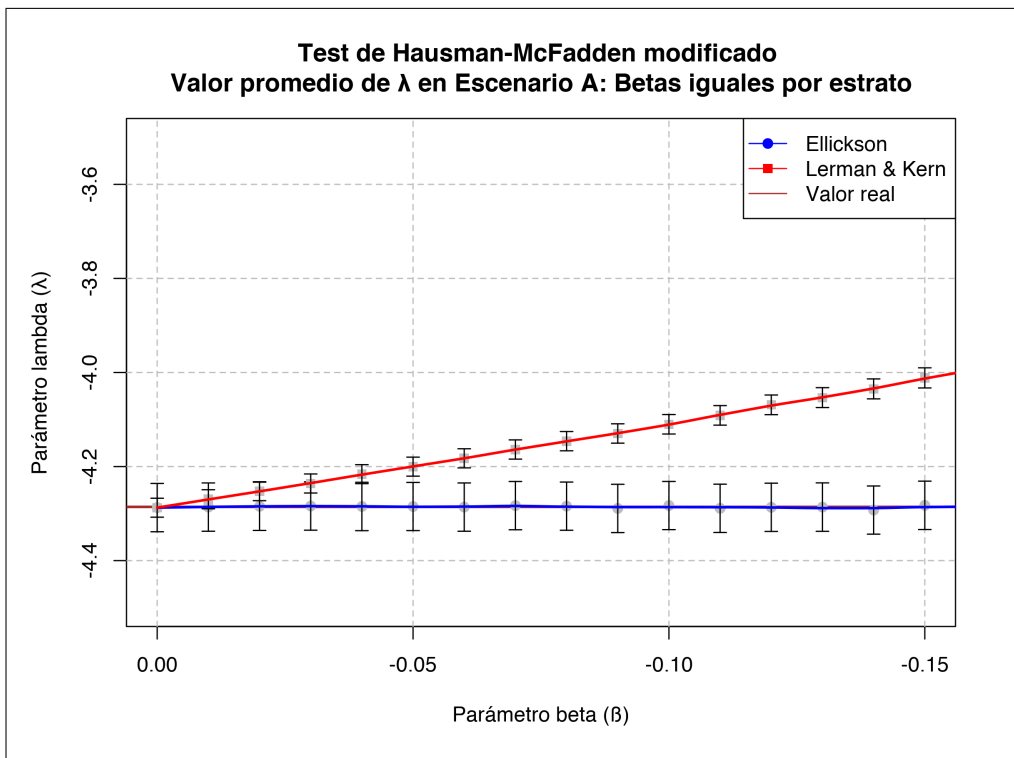


Figura 4.7: Evolución observada del parámetro λ en el Escenario A: Beta ruido igual por estrato. La línea marrón representa el valor verdadero.

4.4.3. Resultados Escenario B: Beta ruido distinto por estrato

En la sección 4.2.6 se definió el escenario B a partir del parámetro $\beta = \beta_{\text{ruido,alto}}$, que es utilizado como tamaño del efecto para determinar el poder de cada test. Por otra parte, se estableció que $\beta_{\text{ruido,bajo}}$ tenga valor nulo para estudiar los efectos causados cuando un solo estrato ve alterada su disposición a pagar. Con lo anterior y recordando que la simulación es realizada sobre una ciudad abierta, este escenario altera la composición de la muestra de hogares cuando varía el tamaño del efecto β , además del efecto en el precio que ya fue estudiado en el escenario anterior. Los resultados y consecuencias de ello son discutidas a continuación.

La tabla 4.6 resume los resultados obtenidos para el escenario B. La figura 4.8 ilustra el poder de las pruebas de Hausman-McFadden y Funciones de Control para este escenario en términos de cobertura empírica alcanzada cuando cambia el tamaño del efecto β .

Tabla 4.6: Resumen de resultados – Estimación de poder empírico – Escenario B

β	Tasa de rechazo por test		Valor λ por estimador		Desviación estándar		Distribución de hogares	
	Hausman-McFadden	Funciones de Control	Ellickson	Lerman y Kern	Ellickson	Lerman y Kern	E. Bajo	E. Alto
0	5	7	-4,28	-4,28	0,052	0,020	38 %	62 %
0,1	6	99	-4,19	-4,18	0,052	0,020	40 %	60 %
0,2	5	100	-4,09	-4,08	0,053	0,020	42 %	58 %
0,3	6	100	-3,97	-3,98	0,056	0,021	45 %	55 %
0,4	7	100	-3,86	-3,88	0,059	0,022	47 %	53 %
0,5	7	100	-3,75	-3,78	0,062	0,023	49 %	51 %
0,6	16	100	-3,62	-3,69	0,066	0,024	51 %	49 %
0,7	16	100	-3,52	-3,59	0,069	0,025	53 %	47 %
0,8	36	100	-3,39	-3,49	0,072	0,026	55 %	45 %
0,9	47	100	-3,27	-3,40	0,077	0,028	57 %	43 %
1	59	100	-3,13	-3,29	0,081	0,030	59 %	41 %
1,1	68	100	-3,00	-3,19	0,086	0,032	61 %	39 %
1,2	76	100	-2,86	-3,08	0,091	0,035	63 %	37 %
1,3	72	100	-2,73	-2,97	0,097	0,037	64 %	36 %
1,4	80	100	-2,58	-2,84	0,102	0,040	66 %	34 %
1,5	85	100	-2,45	-2,73	0,108	0,043	68 %	32 %
1,6	79	100	-2,30	-2,57	0,116	0,047	69 %	31 %
1,7	81	100	-2,15	-2,45	0,122	0,051	71 %	29 %
1,8	71	100	-1,99	-2,30	0,129	0,056	72 %	28 %
1,9	66	100	-1,82	-2,12	0,137	0,061	73 %	27 %
2	64	100	-1,67	-1,96	0,145	0,067	75 %	25 %
2,1	58	100	-1,47	-1,76	0,156	0,075	76 %	24 %
2,2	49	100	-1,30	-1,57	0,164	0,084	77 %	23 %
2,3	35	100	-1,11	-1,32	0,175	0,093	78 %	22 %
2,4	33	100	-0,91	-1,08	0,187	0,105	80 %	20 %
2,5	22	100	-0,70	-0,73	0,199	0,119	81 %	19 %
2,6	23	100	-0,48	-0,50	0,213	0,137	81 %	19 %
2,7	28	100	-0,25	-0,14	0,226	0,153	83 %	17 %
2,8	45	100	0,04	0,35	0,247	0,177	83 %	17 %
2,9	67	100	0,27	0,80	0,266	0,202	84 %	16 %
3	83	100	0,59	1,28	0,285	0,228	85 %	15 %
3,1	91	100	0,91	1,84	0,309	0,263	86 %	14 %
3,2	99	100	1,26	2,48	0,344	0,309	87 %	13 %
3,3	99	100	1,63	3,10	0,371	0,345	87 %	13 %
3,4	100	100	2,05	3,77	0,407	0,400	88 %	12 %
3,5	100	100	2,59	4,53	0,453	0,449	89 %	11 %
3,6	100	100	3,05	5,25	0,500	0,515	89 %	11 %
3,7	99	100	3,57	6,00	0,564	0,580	90 %	10 %
3,8	100	100	4,14	6,89	0,637	0,665	90 %	10 %
3,9	100	100	4,91	8,00	0,756	0,789	91 %	9 %
4	100	100	5,51	8,74	0,850	0,875	91 %	9 %

N=2000, 10.000 repeticiones, Nivel de significación (α) del 5 % y precisión de 5 %.

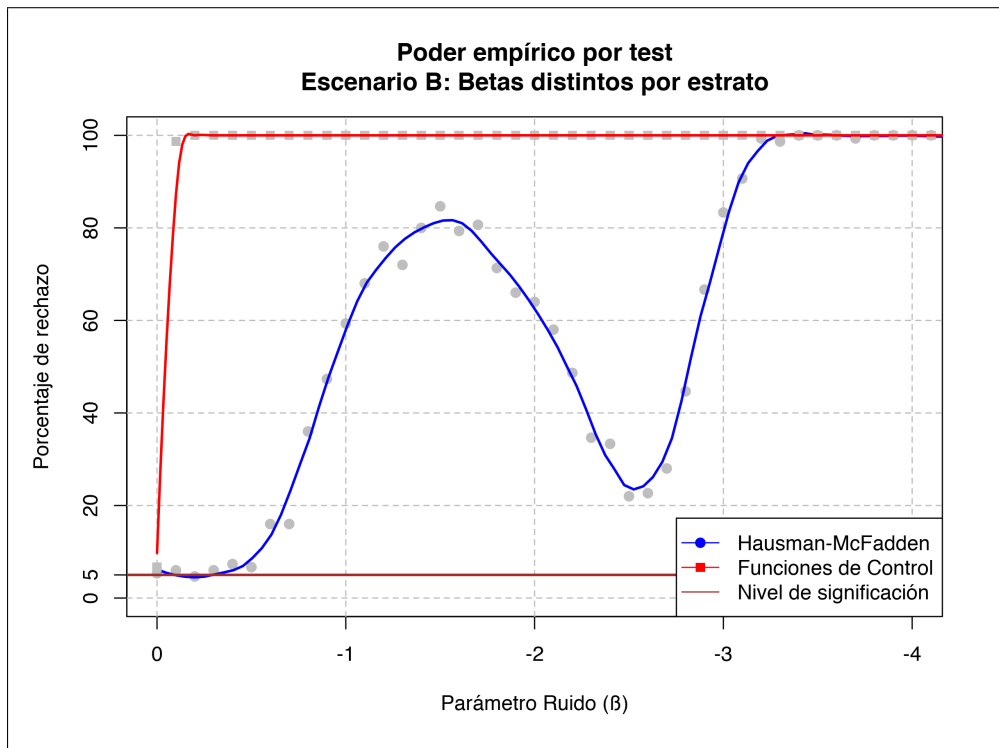


Figura 4.8: Poder empírico observado en Escenario B: Beta ruido distinto por estrato. La línea marrón indica el nivel de significación.

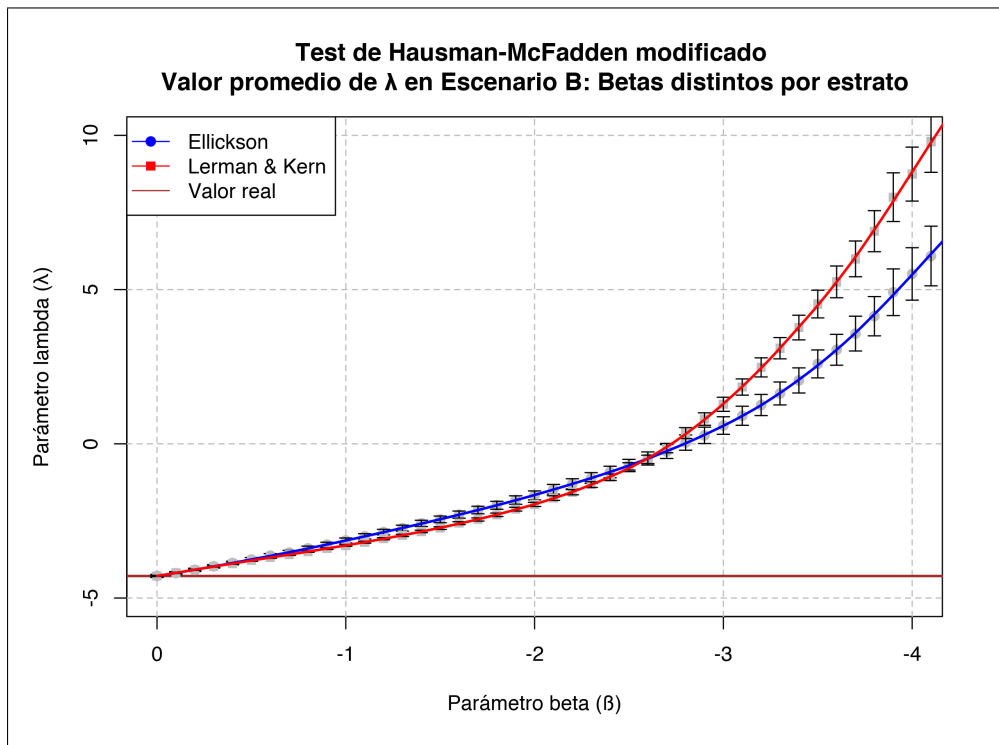


Figura 4.9: Evolución observada del parámetro λ en el Escenario B: Beta ruido distinto por estrato. La línea marrón representa el valor verdadero.

La prueba de funciones de control alcanza un 100 % de cobertura rápidamente, mientras que la prueba de Hausman-McFadden se muestra menos sensible ante cambios de β e incluso registra una bajada en el valor observado de cobertura empírica antes de alcanzar el 100 %. Para explicar dicha caída se realizó un análisis sobre el comportamiento del parámetro λ , presentado en la figura 4.9.

Del análisis de la figura 4.9 se puede concluir que el fenómeno tiene su origen en el comportamiento de los valores observados, donde destaca la intersección de los valores de λ para cada estimador en el caso $\beta = 2,6$, lo que coincide con la caída de cobertura empírica en dicho punto. Las razones de la existencia de dicho punto pueden hallarse en el impacto que tienen los cambios de la distribución de hogares (Figura 4.10), el cual pierde a medida que varía β y se estabiliza la partición de hogares. Así, el estimador de Ellickson puede verse inicialmente más afectado por β , dominando el efecto de la distribución de hogares; hasta que el crecimiento del estimador de Lerman y Kern se acelera y se alcanza un escenario donde domina el efecto del precio pagado.

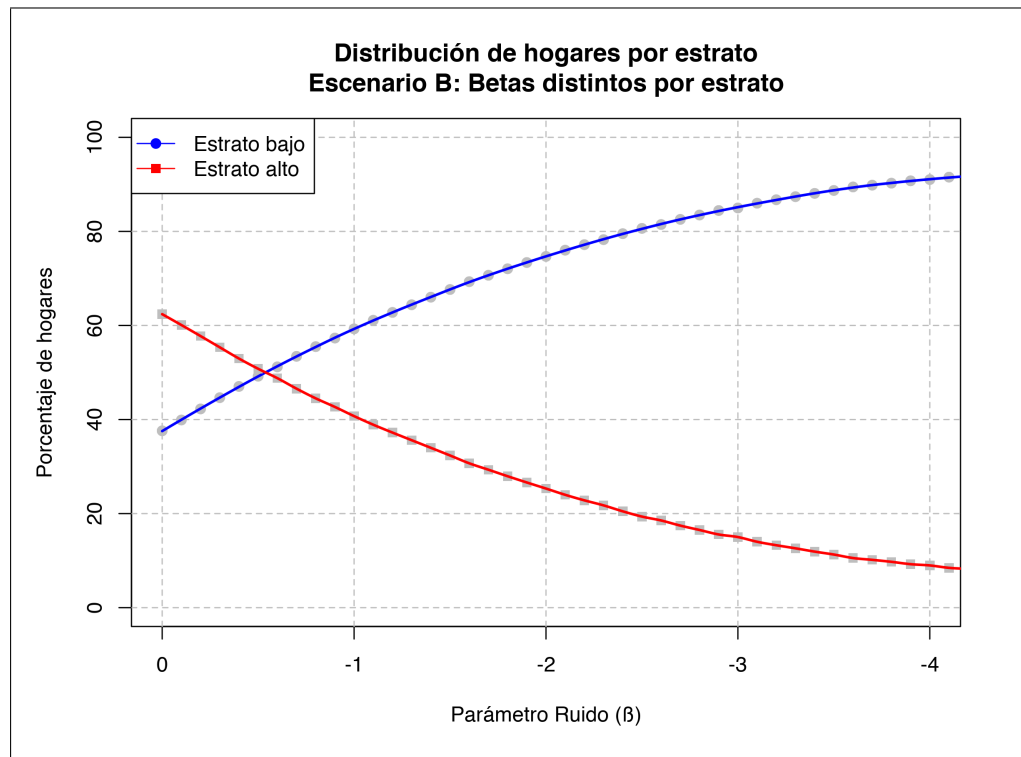


Figura 4.10: Distribución de hogares por estrato en el Escenario B: Beta ruido distinto por estrato. La cantidad de hogares de estrato alto se muestra decreciente conforme aumenta β .

El análisis de la figura 4.10 revela que la distribución de hogares por estrato se invierte a medida que cambia β , llegando a un punto donde la estimación de parámetros para uno de los estratos será imposible, lo que redundará en un escenario degenerado. Como primera extensión del caso estudiado se propone cambiar las constantes de ingreso para establecer una hipótesis nula donde la distribución de hogares es igualitaria. El escenario antes descrito será denominado B-II y se utilizarán los valores de la tabla 4.7.

Tabla 4.7: Parámetros por estrato – Escenario B-II

Parámetros β por estrato	Bajo	Alto
0. Constante (ASC)	0	-4
1. Superficie	0,40	0,75
2. Distancia al CBD	-0,5	-2
3. Ruido	0	Variable

La tabla 4.8 presenta un resumen de los resultados obtenidos en este nuevo escenario, donde solo se evalúa la prueba de Hausman-McFadden modificada. La figura 4.11 presenta el poder empírico reportado por este nuevo escenario, mientras que las figuras 4.12 y 4.13 presentan la evolución del parámetro λ y la distribución de hogares respectivamente. Este escenario permite concluir que el hecho de que se invierta la partición de hogares en el escenario anterior no es relevante al momento de evaluar sus resultados, puesto que no se reportan diferencias apreciables con comparación al Escenario B original.

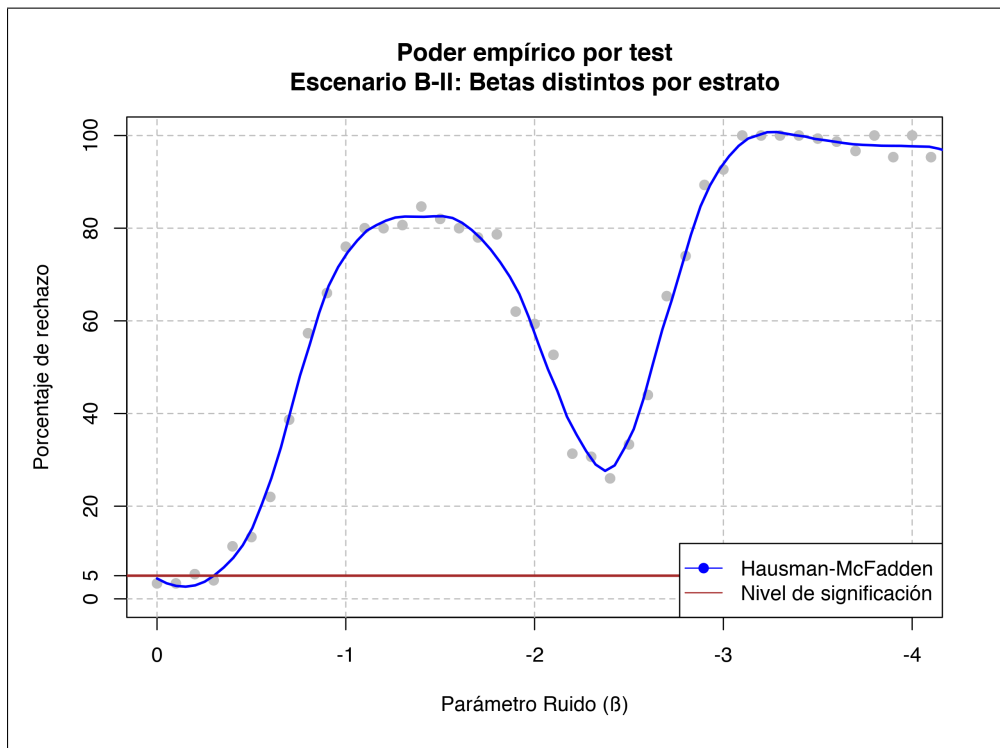


Figura 4.11: Poder empírico observado en Escenario B-II: Beta ruido distinto por estrato. La línea marrón indica el nivel de significación.

Para concluir esta sección, se propone estudiar el supuesto de ciudad abierta utilizado durante la simulación. Para ello la sección siguiente abordará el enfoque de ciudad cerrada y discutirá si este mejora o empeora el poder del test de Hausman-McFadden.

Tabla 4.8: Resumen de resultados – Estimación de poder empírico – Escenario B-II

β	Tasa de rechazo	Valor λ por estimador		Desviación estándar		Distribución de hogares	
	Hausman-McFadden	Ellickson	Lerman y Kern	Ellickson	Lerman y Kern	E. Bajo	E. Alto
0	3	-4,28	-4,29	0,055	0,018	54 %	46 %
0,1	3	-4,18	-4,18	0,057	0,019	56 %	44 %
0,2	5	-4,07	-4,08	0,060	0,020	58 %	42 %
0,3	4	-3,96	-3,99	0,063	0,021	60 %	40 %
0,4	11	-3,85	-3,89	0,066	0,022	62 %	38 %
0,5	13	-3,73	-3,79	0,068	0,022	63 %	37 %
0,6	22	-3,62	-3,71	0,071	0,024	65 %	35 %
0,7	39	-3,49	-3,61	0,076	0,025	67 %	33 %
0,8	57	-3,36	-3,51	0,080	0,027	68 %	32 %
0,9	66	-3,24	-3,43	0,085	0,028	70 %	30 %
1	76	-3,09	-3,31	0,089	0,031	71 %	29 %
1,1	80	-2,97	-3,22	0,095	0,033	73 %	27 %
1,2	80	-2,84	-3,10	0,101	0,036	74 %	26 %
1,3	81	-2,70	-2,99	0,107	0,038	76 %	24 %
1,4	85	-2,55	-2,88	0,114	0,042	77 %	23 %
1,5	82	-2,41	-2,75	0,121	0,045	78 %	22 %
1,6	80	-2,24	-2,60	0,128	0,050	79 %	21 %
1,7	78	-2,08	-2,43	0,138	0,055	80 %	20 %
1,8	79	-1,91	-2,28	0,146	0,062	81 %	19 %
1,9	62	-1,73	-2,07	0,156	0,069	83 %	17 %
2	59	-1,55	-1,88	0,167	0,079	84 %	16 %
2,1	53	-1,35	-1,64	0,180	0,091	85 %	15 %
2,2	31	-1,19	-1,38	0,190	0,105	85 %	15 %
2,3	31	-0,97	-1,14	0,208	0,121	86 %	14 %
2,4	26	-0,72	-0,74	0,223	0,138	87 %	13 %
2,5	33	-0,49	-0,37	0,239	0,160	88 %	12 %
2,6	44	-0,19	0,12	0,260	0,187	89 %	11 %
2,7	65	0,12	0,67	0,282	0,216	89 %	11 %
2,8	74	0,39	1,09	0,305	0,245	90 %	10 %
2,9	89	0,80	1,72	0,336	0,290	91 %	9 %
3	93	1,19	2,37	0,372	0,340	91 %	9 %
3,1	100	1,70	3,26	0,423	0,411	92 %	8 %
3,2	100	2,03	3,73	0,457	0,456	92 %	8 %
3,3	100	2,74	4,74	0,541	0,548	93 %	7 %
3,4	100	3,24	5,51	0,621	0,638	93 %	7 %
3,5	99	3,90	6,41	0,719	0,738	94 %	6 %
3,6	99	4,63	7,46	0,863	0,840	94 %	6 %
3,7	97	5,31	8,29	1,006	0,985	94 %	6 %
3,8	100	5,97	9,30	1,149	1,101	94 %	6 %
3,9	95	6,82	10,48	1,360	1,286	95 %	5 %
4	100	7,61	11,74	1,544	1,455	95 %	5 %

N=2000, 10.000 repeticiones, Nivel de significación (α) del 5% y precisión de 5%.

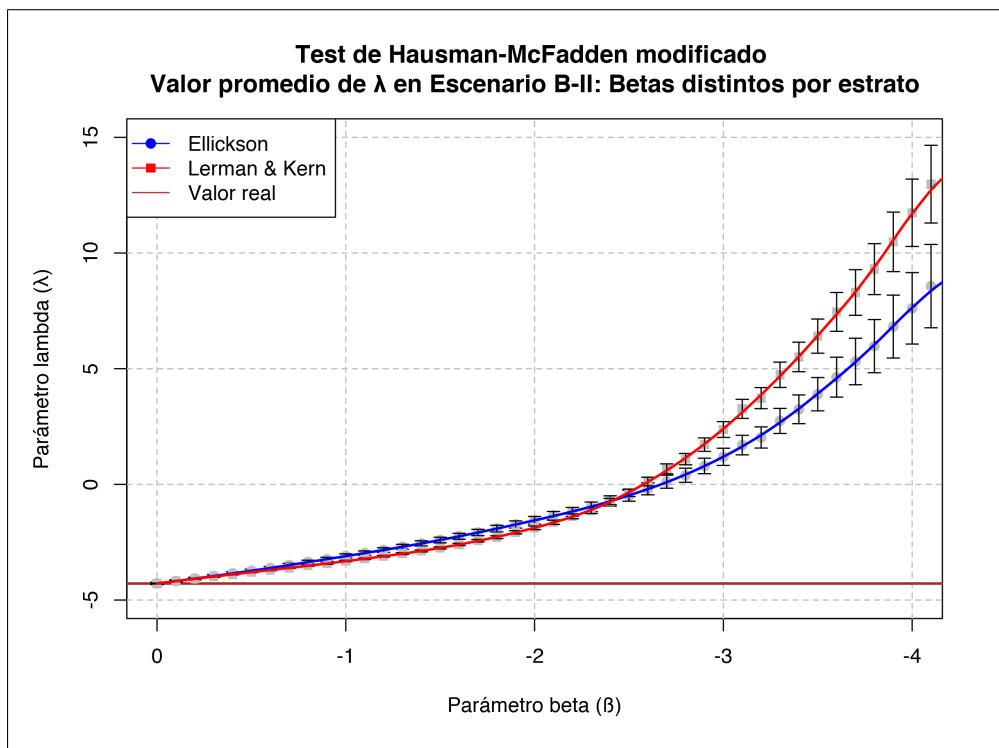


Figura 4.12: Evolución observada del parámetro λ en el Escenario B-II: Beta ruido distinto por estrato. La línea marrón representa el valor verdadero.

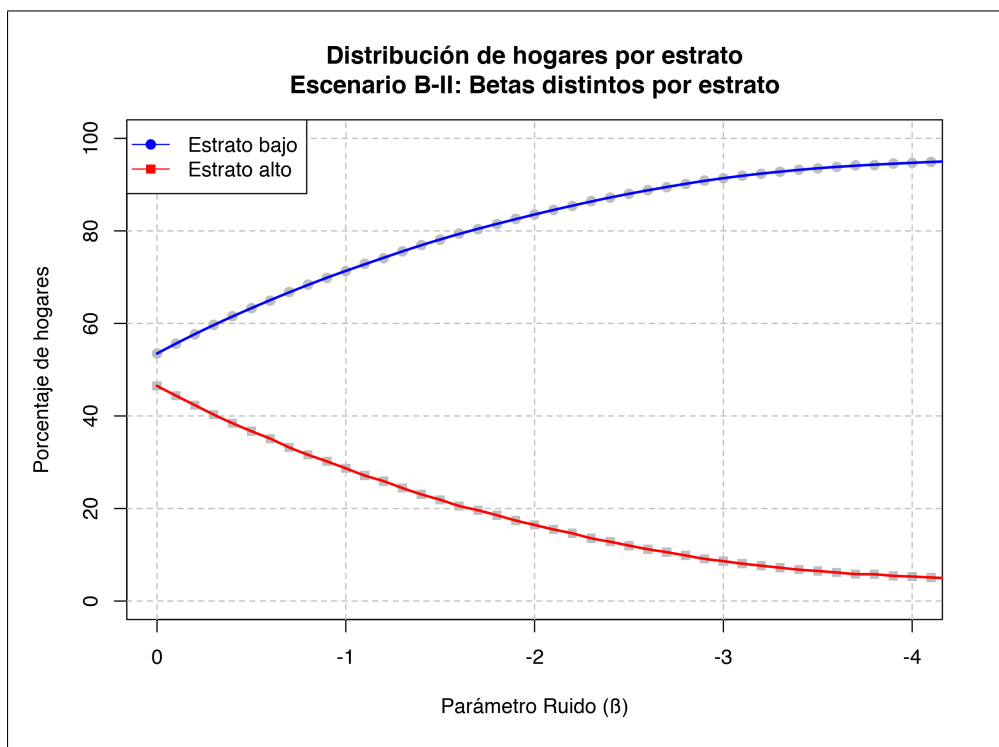


Figura 4.13: Distribución de hogares por estrato en el Escenario B-II: Beta ruido distinto por estrato. En este escenario no se invierten la partición de hogares.

4.5. Extensión de la simulación a ciudad cerrada

Los resultados obtenidos para la estimación del poder en el escenario B motivan la discusión de algunos supuestos realizados para realizar la simulación. En esta línea, durante la definición de la metodología (sección 4.2.5) se discutieron dos posibles enfoques de asignación:

1. **Ciudad abierta:** En ella se supone la existencia de una cantidad de hogares cuasi-infinita que oferta de forma agrupada por estrato sobre un número limitado de viviendas. Esto resulta en un enfoque agregado para el remate, donde la competencia es a nivel de estrato socioeconómico y el resultado del remate define también la distribución de hogares de la muestra.
2. **Ciudad cerrada:** En esta la distribución de hogares es definida de antemano, de modo que los hogares pueden ser modelados como agentes independientes. Esto resulta importante, puesto que ahora el orden en que se realiza el remate resulta relevante y existen múltiples efectos sobre la interacción entre agentes no son capturados por la simulación del remate.

En primera instancia se optó por ciudad abierta, ya que simplifica en gran parte el proceso de asignación de viviendas. El caso de ciudad cerrada, por otro lado, presenta dificultades adicionales, cuyo alcance no espera ser cubierto para este trabajo. La revisión de antecedentes presentada en Rocco (2014) señala que se requiere una representación detallada de la interacción entre compradores y vendedores, la cual considere expectativas y percepciones del mercado, concluyendo en un precio final a pagar (Filatova et al., 2009). La misma revisión también se da cuenta sobre múltiples modelos propuestos, donde cada uno de ellos ofrece un nuevo enfoque para describir la asignación de hogares y viviendas.

Así, en consideración de lo presentado y los alcances de este trabajo, se opta por realizar un procedimiento simplificado de remate, donde se hará énfasis en el **orden de asignación de las viviendas** durante el remate. Para ello se discute sobre tres posibles formas de asignación y se presentan los resultados de evaluar nuevamente el poder empírico de los test bajo este nuevo enfoque.

4.5.1. Modificaciones metodológicas

En primer lugar se estableció que la distribución inicial de hogares por estrato fuera igual a la identificada para el caso de ciudad abierta cuando se cumple la hipótesis nula de no endogeneidad. Con esto en mente se estableció una muestra de hogares donde el 60% de la población fuera de estrato alto y el restante fuera de estrato bajo.

Como ya fue mencionado, el orden en que se asignan las viviendas cobra especial importancia bajo un escenario de ciudad cerrada, ya que un orden arbitrario no entregará necesariamente una asignación representativa del mercado inmobiliario, lo que provocaría una fuente de endogeneidad no deseada por el investigador. Se procede a discutir tres formas de asignación que pretenden abordar este punto: completamente aleatoria, eliminación al mejor postor y eliminación secuencial.

Asignación aleatoria

En este caso se procede a elegir en forma aleatoria el orden en que las viviendas pertenecientes a la muestra salen a remate y son asignadas. El objetivo de presentar este caso es identificar las discrepancias con la realidad que genera este tipo de asignación.

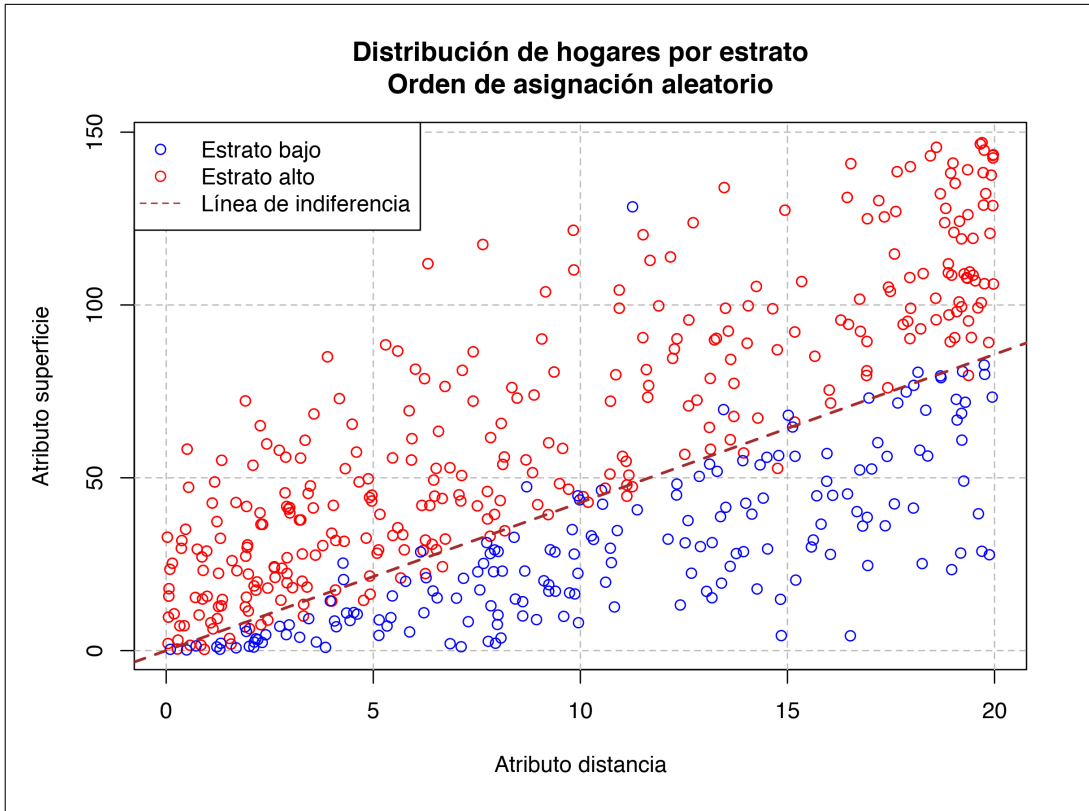


Figura 4.14: Distribución de hogares por estrato simulada en una ciudad cerrada donde el orden de asignación de las viviendas es aleatorio.

La figura 4.14 presenta la distribución de hogares asignados considerando los atributos de superficie y distancia. Se observa una asignación regular para la mayoría de las viviendas, con discrepancias puntuales en cada asignación. Esta discrepancia ocurre porque, al momento de asignar aquellas viviendas, todos los hogares de estrato alto ya han sido asignados. Pese a lo anterior, los resultados observados parecen razonables y además se logra observar una línea de indiferencia entre estratos, caracterizada por el parámetro λ de la simulación.

Eliminación secuencial al mejor postor

Este método alternativo realiza una asignación secuencial basada en las posturas observadas sobre cada hogar y vivienda. Los pasos para realizar esta asignación se describen a continuación:

1. Se genera una matriz de posturas donde todos los hogares de la muestra realizan una oferta por cada vivienda disponible.
2. Se identifica el par vivienda-hogar que registra el valor máximo global de la matriz de posturas, correspondiente a una asignación.
3. Una vez registrada la asignación, se elimina la información de este hogar y vivienda, obteniendo una nueva matriz de posturas.
4. Se repite el paso 2 utilizando la nueva matriz, aquello ocurre hasta que todos los pares hogar-vivienda son asignados.

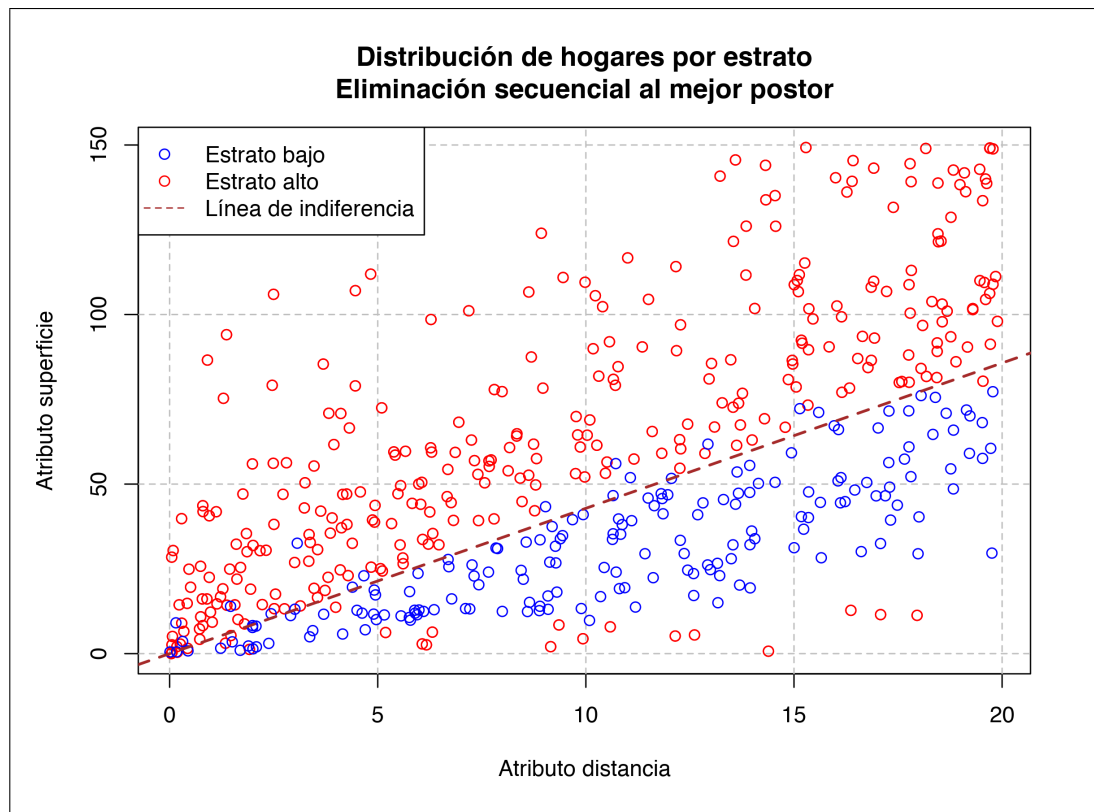


Figura 4.15: Distribución de hogares por estrato simulada en una ciudad cerrada donde la asignación se realiza mediante eliminación secuencial al mejor postor.

La figura 4.15 presenta la distribución obtenida bajo este método. Llama la atención la presencia de hogares de estrato alto en la zona inferior de la gráfica, contrario a toda lógica. Dicha anomalía tiene que ver con el hecho de asignar por mejor postura, ya que al analizar las posturas observadas para las viviendas de la zona inferior se constata que los hogares de estrato bajo realizan mejores posturas. Sin embargo, cuando se realiza la asignación de estas viviendas ocurre que la totalidad de hogares pertenecientes al estrato bajo han sido asignados, por lo que solo compiten hogares del estrato alto. Lo anterior no es representativo de la realidad dado que no se alcanza una situación de equilibrio y por tanto ocurrirían inconsistencias en la estimación de parámetros. En consideración de lo anterior este enfoque es descartado.

Eliminación secuencial equilibrada

Buscando solucionar la anomalía descrita en la parte anterior, se modifica el orden de asignación buscando que las viviendas de bajo valor sean asignadas de forma temprana. La asignación realizada queda descrita de la siguiente forma:

1. Se genera una matriz de posturas donde todos los hogares de la muestra realizan una oferta por cada vivienda disponible.
2. Se computa la suma acumulada de posturas por vivienda, donde un mayor valor redonda en una mayor postura por parte de los hogares pertenecientes a cada estrato.
3. Se ordenan las viviendas de mayor a menor, utilizando el valor calculado en el paso anterior.
4. Se elige la primera vivienda de la lista y se identifica su mejor postura, realizando la asignación correspondiente y marcando hogar y vivienda como asignadas.
5. Se elige la última vivienda de la lista y se repite lo descrito en el paso anterior.
6. Se repiten los pasos 3 y 4 hasta que ya no queden viviendas ni hogares por asignar.

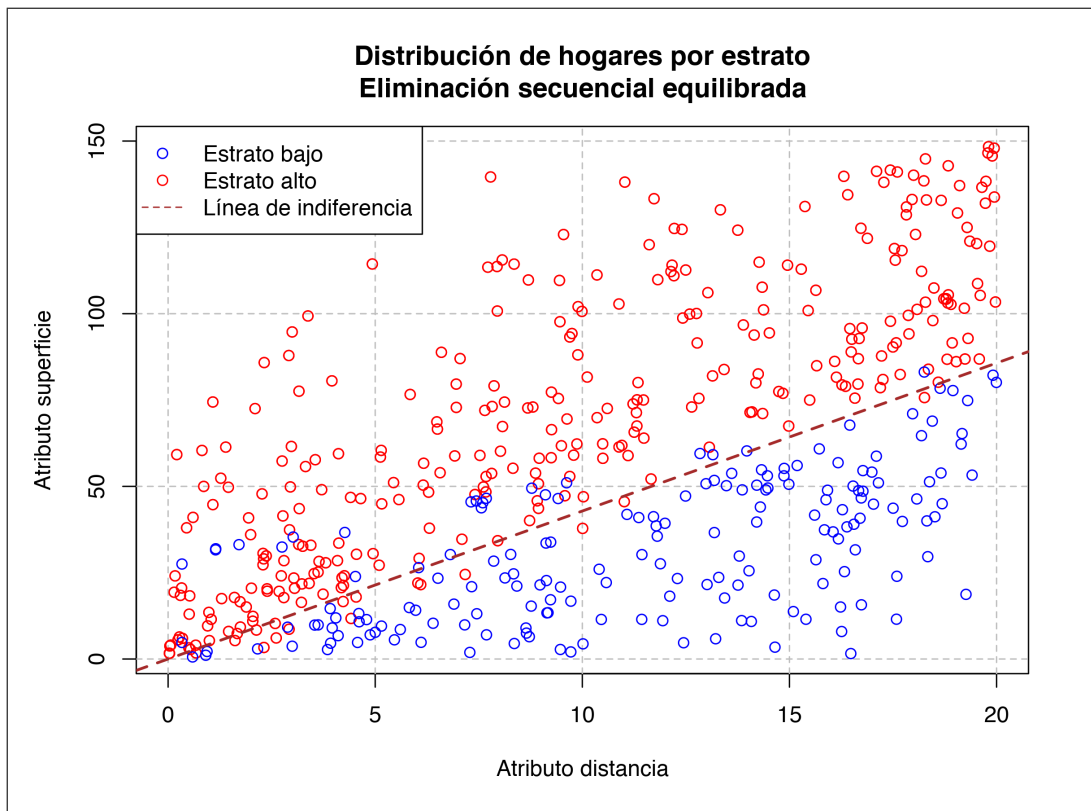


Figura 4.16: Distribución de hogares por estrato simulada en una ciudad cerrada donde la asignación se realiza mediante eliminación secuencial equilibrada.

El resultado de aplicar dicha metodología es presentado en la figura 4.16 y muestra un resultado más realista que el caso anterior. La justificación de esto se encuentra en que las últimas viviendas en ser asignadas son aquellas donde ambos estratos realmente compiten por su adjudicación, al tener posturas similares donde es más influyente el término de error que la componente sistemática de la postura.

Conclusiones

El análisis realizado sugiere que tanto el enfoque aleatorio como el secuencial equilibrado logran replicar condiciones de equilibrio aceptables para evaluar el test de hipótesis desarrollado; siendo importante recalcar que se trata de un escenario simplificado.

De lo anterior se desprende que es posible usar cualquier enfoque, de modo que se opta por el caso aleatorio por su facilidad inherente para reproducir el experimento planteado.

4.5.2. Resultados de estimación del poder empírico

Escenario A: Beta ruido igual por estrato

La tabla 4.9 resume los resultados obtenidos de la simulación. La figura 4.17 ilustra la evolución del poder empírico a medida que cambia el tamaño del efecto β establecido para el escenario.

El análisis de los resultados obtenidos para el test de Hausman-McFadden sugiere en primera instancia que la prueba se ve afectada por *size distortion* de carácter liberal, es decir que la tasa de rechazo cuando se cumple la hipótesis nula ($\beta = 0$) supera con creces el nivel de significación nominal del 5%. No obstante, también se observa que el test es capaz de detectar endogeneidad, alcanzando una cobertura empírica del 100% a partir de $\beta = 0,45$.

El análisis en detalle del parámetro λ estimado para este escenario (Figura 4.18) muestra que el parámetro se comporta de forma similar al caso de ciudad abierta (Figura 4.7). Sin embargo, se observa un leve sesgo en el valor estimado de λ para el estimador de Ellickson, incluyendo el caso donde la hipótesis nula es válida.

Tabla 4.9: Resumen de resultados – Estimación de poder empírico – Escenario A – Caso ciudad cerrada

β	Tasa de rechazo		Valor λ por estimador		Desviación estándar	
	Hausman-McFadden		Ellickson	Lerman y Kern	Ellickson	Lerman y Kern
0	23		-4,34	-4,29	0,17	0,09
0,05	34		-4,34	-4,20	0,17	0,10
0,1	38		-4,35	-4,15	0,18	0,10
0,15	56		-4,34	-4,07	0,18	0,10
0,2	74		-4,33	-4,03	0,17	0,09
0,25	84		-4,34	-3,93	0,17	0,09
0,3	90		-4,36	-3,90	0,17	0,09
0,35	99		-4,35	-3,81	0,18	0,09
0,4	98		-4,32	-3,77	0,17	0,09
0,45	100		-4,33	-3,71	0,17	0,09
0,5	100		-4,29	-3,64	0,17	0,09
0,55	100		-4,31	-3,54	0,17	0,09
0,6	100		-4,31	-3,50	0,17	0,08
0,65	100		-4,35	-3,39	0,18	0,09
0,7	100		-4,27	-3,32	0,18	0,08
0,75	100		-4,31	-3,28	0,17	0,08
0,8	100		-4,33	-3,20	0,17	0,08
0,85	100		-4,31	-3,13	0,18	0,08
0,9	100		-4,30	-3,08	0,17	0,08
0,95	100		-4,31	-2,99	0,18	0,08
1	100		-4,31	-2,97	0,17	0,08

N=500, 10.000 repeticiones, Nivel de significación (α) del 5% y precisión de 5%.

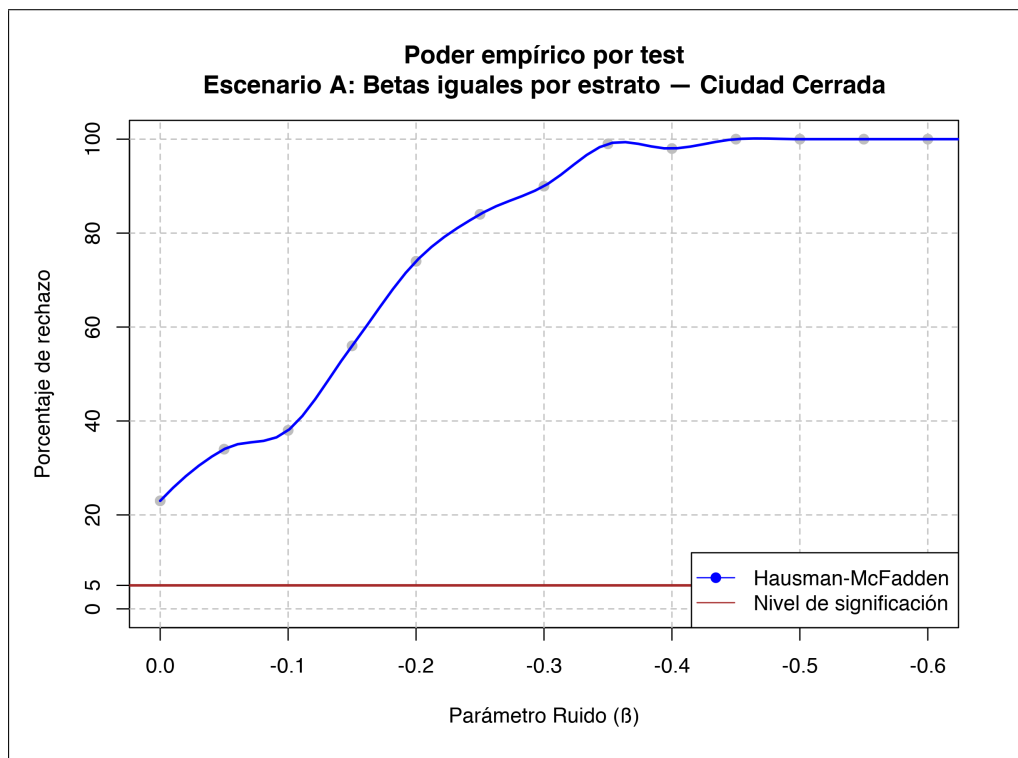


Figura 4.17: Poder empírico observado en Escenario A cuando se realiza la asignación por Ciudad Cerrada. La línea marrón indica el nivel de significación

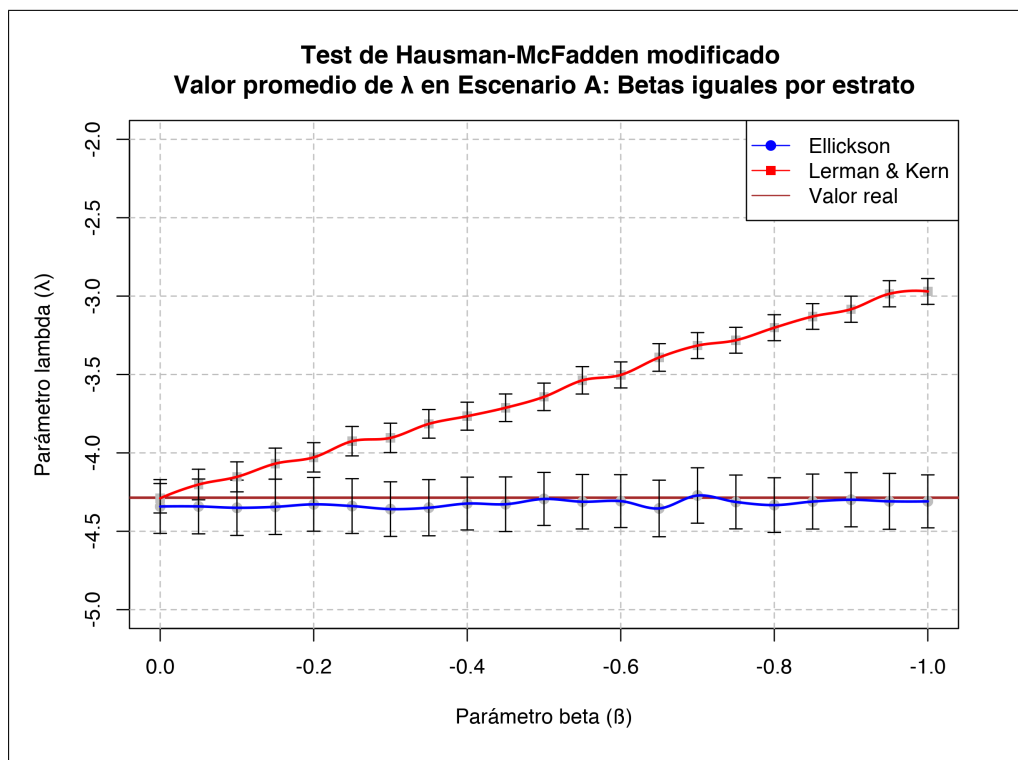


Figura 4.18: Evolución observada del parámetro λ en el Escenario A para el caso de Ciudad Cerrada. La línea marrón representa el valor verdadero.

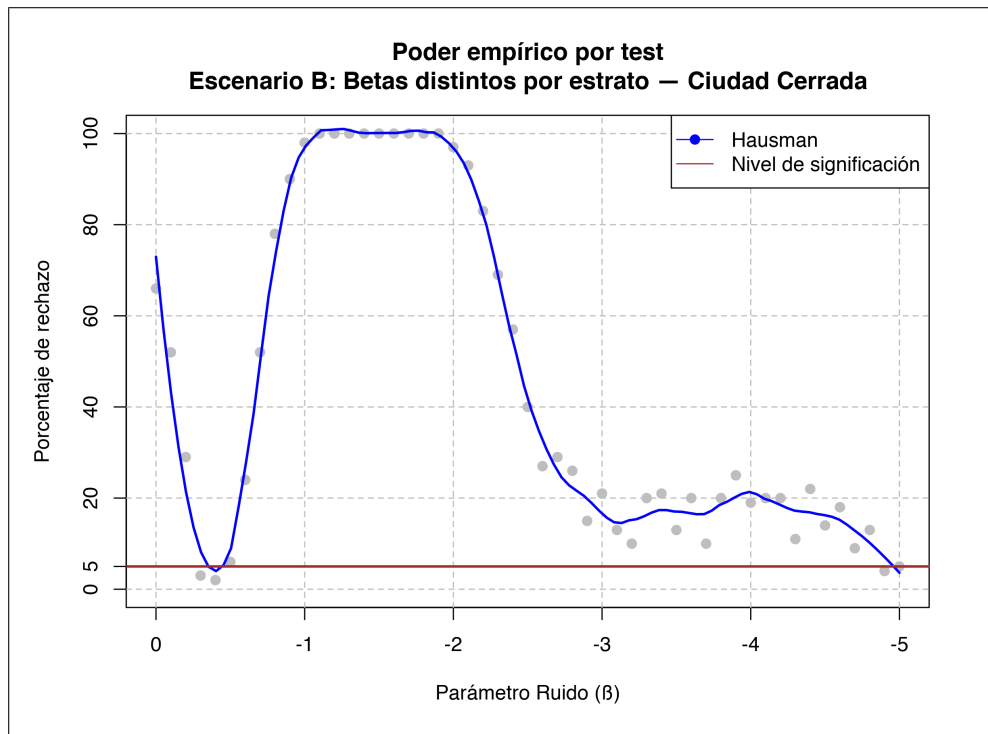


Figura 4.20: Poder empírico observado en Escenario B cuando se realiza la asignación por Ciudad Cerrada. La línea marrón indica el nivel de significación

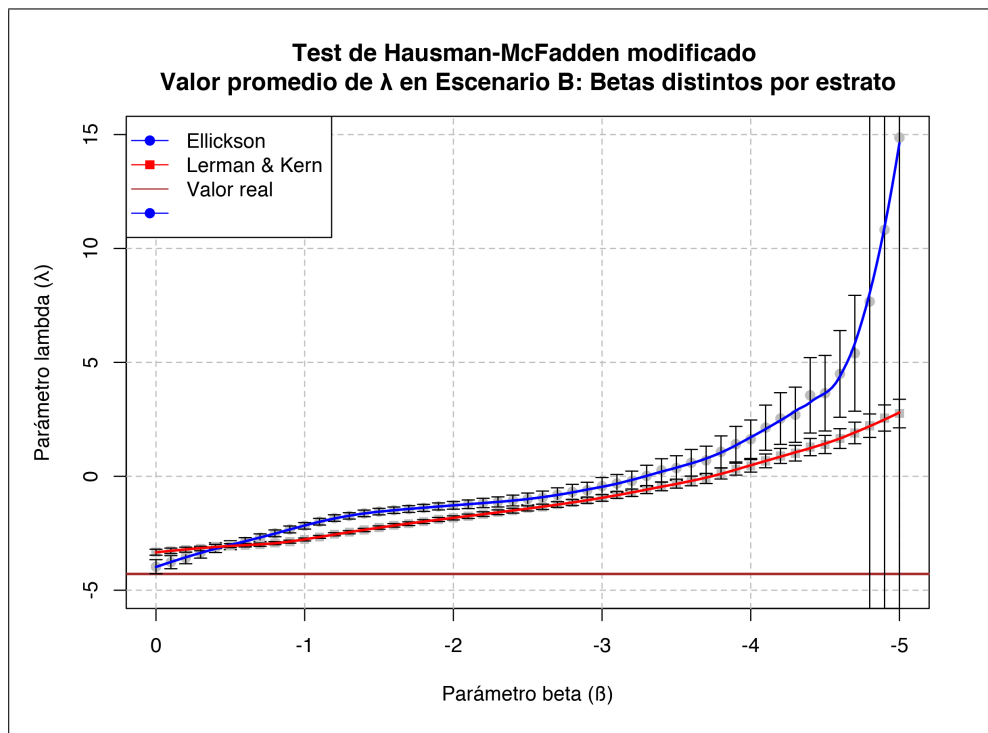


Figura 4.21: Evolución observada del parámetro λ en el Escenario B para el caso de Ciudad Cerrada. La línea marrón representa el valor verdadero.

Tabla 4.10: Resumen de resultados – Estimación de poder empírico – Escenario B – Caso ciudad cerrada

Beta	Tasa de rechazo		Valor lambda por estimador		Desviación estándar	
	Hausman-McFadden	Ellickson	Lerman y Kern	Ellickson	Lerman y Kern	
0	66	-3,97	-3,33	0,31		0,13
0,1	52	-3,77	-3,27	0,29		0,12
0,2	29	-3,58	-3,22	0,26		0,11
0,3	3	-3,36	-3,17	0,23		0,10
0,4	2	-3,13	-3,08	0,21		0,09
0,5	6	-3,03	-3,04	0,20		0,09
0,6	24	-2,87	-3,02	0,18		0,08
0,7	52	-2,69	-3,00	0,17		0,07
0,8	78	-2,50	-2,93	0,15		0,07
0,9	90	-2,33	-2,87	0,15		0,07
1	98	-2,17	-2,77	0,14		0,06
1,1	100	-1,99	-2,66	0,15		0,07
1,2	100	-1,83	-2,54	0,14		0,07
1,3	100	-1,74	-2,45	0,14		0,07
1,4	100	-1,63	-2,36	0,14		0,07
1,5	100	-1,55	-2,25	0,14		0,07
1,6	100	-1,49	-2,15	0,14		0,08
1,7	100	-1,44	-2,08	0,15		0,08
1,8	100	-1,39	-1,99	0,16		0,08
1,9	100	-1,32	-1,90	0,15		0,09
2	97	-1,27	-1,83	0,17		0,09
2,1	93	-1,23	-1,75	0,19		0,09
2,2	83	-1,17	-1,66	0,21		0,10
2,3	69	-1,12	-1,58	0,22		0,10
2,4	57	-1,06	-1,51	0,24		0,11
2,5	40	-0,99	-1,41	0,26		0,11
2,6	27	-0,94	-1,34	0,29		0,12
2,7	29	-0,81	-1,24	0,30		0,12
2,8	26	-0,69	-1,15	0,33		0,13
2,9	15	-0,61	-1,05	0,35		0,14
3	21	-0,43	-0,95	0,38		0,15
3,1	13	-0,32	-0,82	0,40		0,15
3,2	10	-0,20	-0,72	0,43		0,16
3,3	20	0,02	-0,58	0,47		0,18
3,4	21	0,26	-0,45	0,51		0,19
3,5	13	0,36	-0,32	0,54		0,20
3,6	20	0,59	-0,20	0,59		0,21
3,7	10	0,70	-0,09	0,63		0,22
3,8	20	1,07	0,12	0,70		0,24
3,9	25	1,41	0,31	0,78		0,26
4	19	1,63	0,46	0,84		0,27
4,1	20	2,13	0,68	0,99		0,30
4,2	20	2,54	0,89	1,13		0,33
4,3	11	2,70	1,01	1,21		0,34
4,4	22	3,55	1,28	1,66		0,38
4,5	14	3,65	1,40	1,66		0,40
4,6	18	4,49	1,66	1,90		0,43
4,7	9	5,40	1,90	2,54		0,47
4,8	13	7,67	2,22	18,01		0,51
4,9	4	10,82	2,56	81,94		0,57
5	5	14,87	2,75	227,79		0,63

N=500, 10.000 repeticiones, Nivel de significación (α) del 5% y precisión de 5%.

Conclusiones

El análisis de escenarios bajo un contexto de ciudad cerrada muestra resultados peores en términos de indicadores (*Size distortion*, Poder estadístico) que su contraparte simulada con ciudad abierta, lo que se atribuye directamente al hecho de fijar la cantidad de hogares pertenecientes a cada estrato. Pese a realizar cambios en la metodología de asignación con el fin de minimizar el impacto de dicho cambio, los resultados confirman que esta no se logra controlar del todo, en especial considerando el test se ve afectado por *size distortion* en ambos escenarios.

Capítulo 5

Síntesis, Conclusiones y Recomendaciones

5.1. Síntesis de resultados

El trabajo realizado en esta memoria de título propuso una prueba de hipótesis aplicable a modelos de localización basados en máxima postura aleatoria. Dicha prueba está basada en el test de especificación de Hausman-McFadden, una prueba de tipo χ^2 que permite comparar estimadores utilizando únicamente el valor estimado y su varianza correspondiente, sin requerir supuestos adicionales como el uso de variables instrumentales. Para evaluar la prueba desarrollada se simuló una ciudad monocéntrica de tipo abierta y se plantearon múltiples escenarios a partir de una especificación afecta a endogeneidad de la función de postura, causada por la omisión de variables. Para contrastar los resultados se aplicó el enfoque de funciones de control a los mismos escenarios para poder evaluar un test de hipótesis alternativo, siempre teniendo en cuenta que el uso de variables instrumentales complejiza su aplicación.

Los experimentos realizados usando simulación de Monte Carlo permitieron cuantificar indicadores de *calidad* para las pruebas hipótesis: *Size distortion* y Poder estadístico. Bajo un escenario de ciudad abierta se logró garantizar la robustez del test de Hausman-McFadden; donde no se observa *size distortion* para un tamaño de muestra suficiente y obteniendo valores aceptables para el poder del test en el primer escenario evaluado (betas iguales), mas no en el segundo (betas distintos) donde los cambios en la distribución de hogares jugaron un rol importante. Por otra parte, la extensión del análisis a un contexto de ciudad cerrada permitió observar las principales limitaciones de este test de hipótesis, donde un incremento en la varianza de los datos perjudicó notablemente la efectividad del test estudiado, manifestado en un bajo poder estadístico.

5.2. Conclusiones sobre el test desarrollado

La primera conclusión que se puede obtener es que la prueba desarrollada logra detectar el efecto de endogeneidad en la función de postura de forma eficiente, como muestran los resultados de la simulación realizada bajo un contexto de *ciudad abierta*. No obstante, es clave mencionar que aquello no significa que la prueba sea efectiva en cualquier caso. Los resultados del escenario B, donde se causa endogeneidad mediante un parámetro β omitido y distinto para cada estrato, es prueba de aquello. Por otro lado, el uso de un enfoque de asignación por *ciudad cerrada* no muestra mejoras en términos de poder y plantea supuestos adicionales sobre la interacción entre hogares cuando se pretende realizar una simulación. Queda pendiente evaluar la efectividad del test en escenarios reales mediante el uso de datos de panel, que es donde se espera que la prueba planteada sea de utilidad.

En términos de la simulación efectuada, es importante recordar que el trabajo realizado utilizó una muestra de hogares estratificada por ingreso, donde no existen diferencias entre hogares del mismo estrato salvo por la componente aleatoria de la función de postura. Así, la inclusión de atributos propios del hogar oferente queda propuesta y puede ser realizada, por ejemplo, mediante la especificación de Martínez y Henríquez (2007), donde se puede incorporar el efecto del consumo de bienes del hogar con tal de mejorar la estimación del precio real de las viviendas.

Como último punto, se destaca que el uso de muestras de datos que se ajusten a la realidad en términos de magnitudes, varianzas y correlaciones es clave para garantizar resultados verosímiles; donde la omisión de atributos relevantes puede afectar significativamente los resultados de la estimación de modelos tal como se logró observar.

5.3. Investigación posterior

Como consecuencia de lo señalado anteriormente, la continuación natural de este trabajo consiste en una aplicación del Test de hipótesis propuesto en un escenario con datos reales, con el fin de evaluar el comportamiento del test en términos de la detección de endogeneidad.

Continuando la línea de análisis mediante simulaciones de Monte Carlo, el abanico de posibilidades es amplio: Se puede proponer el estudio de especificaciones alternativas, ampliar el análisis a una muestra más segmentada mediante la inclusión de más estratos en la muestra de hogares, estudiar formas alternativas de implementar el test de Hausman-McFadden o estudiar el comportamiento del test en otros escenarios susceptibles a endogeneidad.

Un escenario que se sugiere estudiar es aquel donde la correlación entre atributos omitidos y observados sea variable. Partiendo por el caso sin correlación, se espera que en dicho caso la variable omitida sea absorbida totalmente por el término de error y a medida que se aumenta la correlación entre atributos se observe endogeneidad producto de la variable omitida.

Bibliografía

- Alonso, W. et al. (1964). Location and land use. Harvard University Press Cambridge, MA.
- Anas, A. (1982). Residential Location Markets and Urban Transportation. Economic Theory, Econometrics and Policy Analysis With Discrete Choice Models. No. Monograph.
- Berry, S., J. Levinsohn y A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.
- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.
- Bound, J., D. A. Jaeger y R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association* 90 (430), 443–450.
- Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the Muth-Mills model. *Handbook of regional and urban economics* 2, 821–845.
- Dubin, J. A. y D. L. McFadden (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, 345–362.
- Ellickson, B. (1981). An alternative test of the hedonic theory of housing markets. *Journal of Urban Economics* 9 (1), 56–79.
- Ellis, P. D. (2010). The essential guide to effect sizes: an introduction to statistical power, meta-analysis, and the interpretation of research results. Cambridge University Press.
- Filatova, T., D. Parker y A. Van der Veen (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.
- Fisher, R. A. (1925). Theory of statistical estimation. En: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 22. Cambridge University Press, pp. 700–725.
- Godfrey, L. G. (1991). Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches. No. 16. Cambridge University Press.
- Guevara, A. (2005). Addressing endogeneity in residential location models. Tesis de Magíster, Massachusetts Institute of Technology.

- Guevara, C. A. (2015). Critical assessment of five methods to correct for endogeneity in discrete-choice models. *Transportation Research Part A: Policy and Practice* 82, 240–254.
- Guevara, C. A. y D. Polanco (2016). Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution. *Transportmetrica A: Transport Science* 12 (5), 458–478.
- Hausman, J. y D. McFadden (1984). Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society*, 1219–1240.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251–1271.
- Heckman, J. J. (1977). Dummy endogenous variables in a simultaneous equation system.
- Hofert, M., I. Kojadinovic, M. Maechler y J. Yan (2017). *copula: Multivariate Dependence with Copulas*. R package version 0.999-18.
URL <https://CRAN.R-project.org/package=copula>
- Hurtubia, R. y M. Bierlaire (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. y C. R. Kern (1983). Hedonic theory, bid rents, and willingness-to-pay: Some extensions of ellickson’s results. *Journal of Urban Economics* 13 (3), 358–363.
- Louviere, J., K. Train, M. Ben-Akiva, C. Bhat, D. Brownstone, T. A. Cameron, R. T. Carson, J. Deshazo, D. Fiebig, W. Greene et al. (2005). Recent progress on endogeneity in choice modeling. *Marketing Letters* 16 (3-4), 255–265.
- Martinez, F. (1996). MUSSA: land use model for Santiago city. *Transportation Research Record: Journal of the Transportation Research Board* (1552), 126–134.
- Martínez, F. (2011). *Discover Cube Land: Economic-based Land Use Forecasting Tool*.
URL <http://citilabs-website-resources.s3.amazonaws.com/resources/DiscoverCubeLand.pdf>
- Martínez, F. y P. Donoso (2010). The MUSSA II land use auction equilibrium model. En: *Residential Location Choice*. Springer, pp. 99–113.
- Martínez, F. J. y R. Henríquez (2007). A random bidding and supply land use equilibrium model. *Transportation Research Part B: Methodological* 41 (6), 632–651.
- McFadden, D. L. (1978). Modelling the choice of residential location. En: A. Karlqvist, F. S. y J. Weibull (Eds.), *Spatial Interaction Theory and Residential Location*. North Holland, Amsterdam, The Netherlands, pp. 75–96.
- McMillen, D. P. (1997). Multiple regime bid-rent function estimation. *Journal of Urban Economics* 41 (2), 301–319.

- Miller, E. J., J. D. Hunt, J. E. Abraham y P. A. Salvini (2004). Microsimulating urban systems. *Computers, environment and urban systems* 28 (1-2), 9–44.
- Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *The American Economic Review* 57 (2), 197–210.
- Muth, R. (1969). *Cities and housing: The spatial patterns of urban residential land use*. University of Chicago, Chicago 4, 114–123.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL <https://www.R-project.org/>
- Rocco, V. (2014). *Microsimulación de transacciones en el mercado inmobiliario*. Tesis de Ph.D., Universidad de Chile.
- Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy* 82 (1), 34–55.
- Villas-Boas, J. M. y R. S. Winer (1999). Endogeneity in brand choice models. *Management science* 45 (10), 1324–1338.
- Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke y G. Ulfarsson (2003). Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and spatial economics* 3 (1), 43–67.