

Evaluación de una prueba estadística para detectar endogeneidad en modelos Bid de uso de suelo

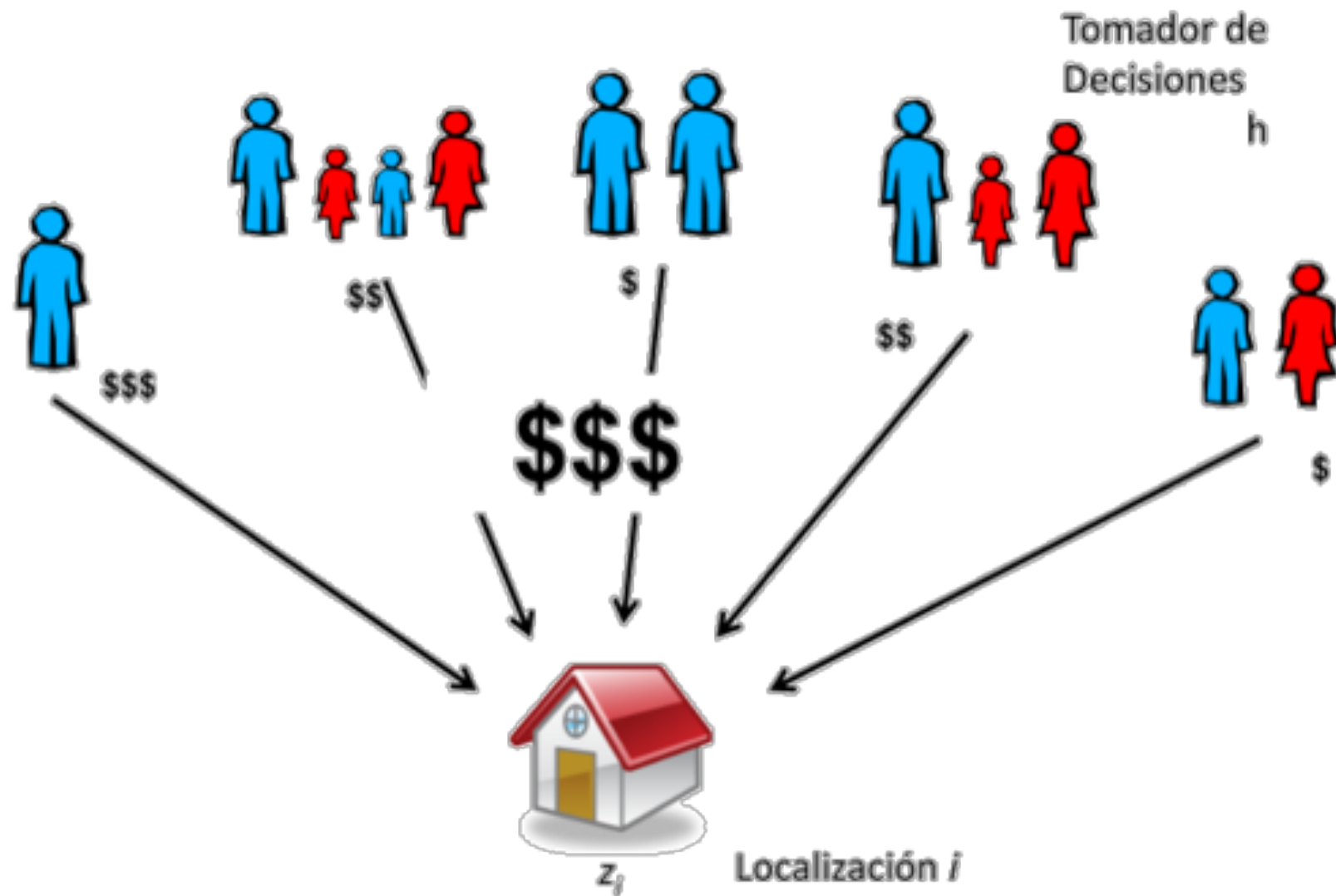
José Soto-Vaquero
CIS-T

Ángelo Guevara Cue
Universidad de Chile

Agenda

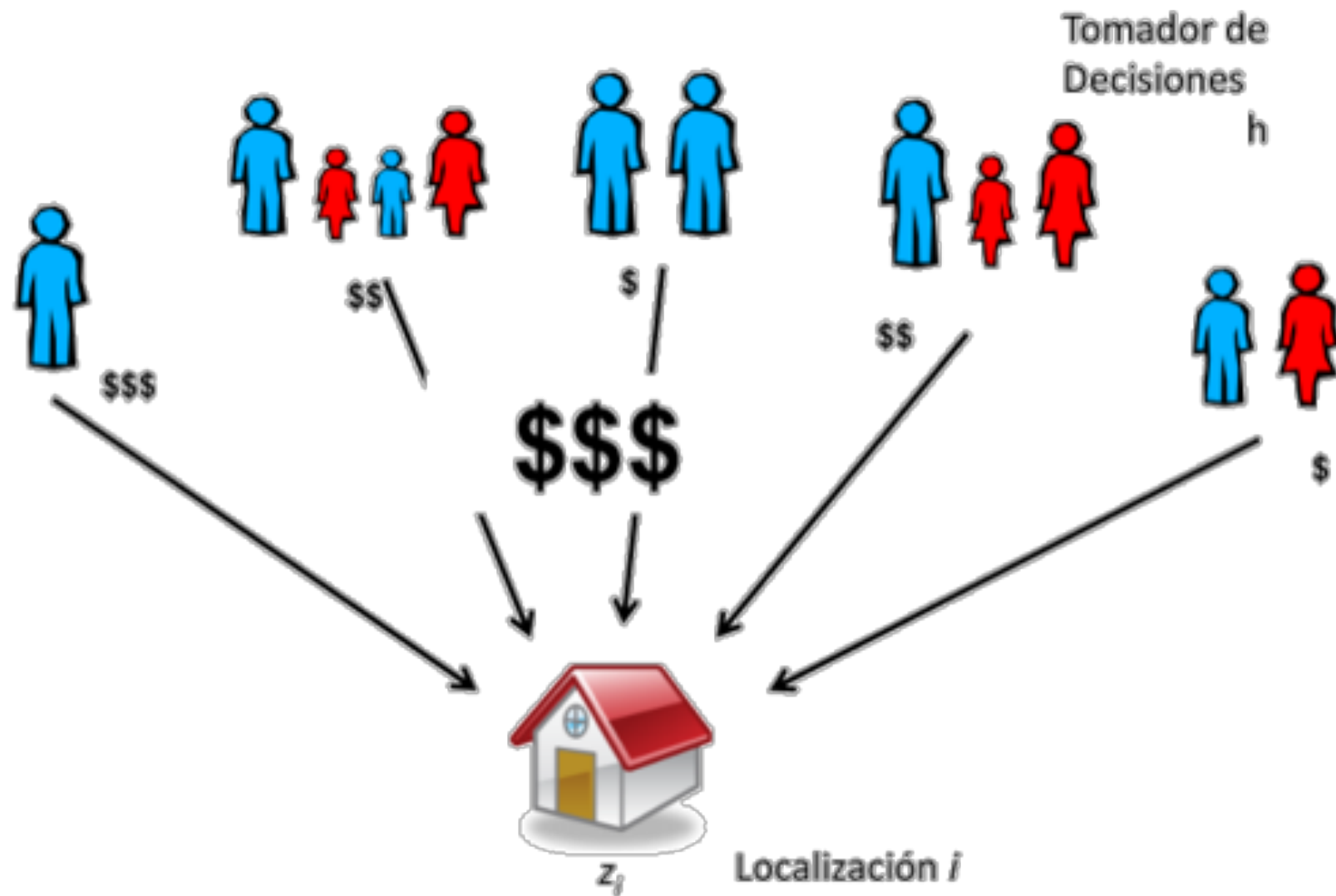
- Motivación y marco teórico
- Test de Hausman-McFadden
- Estudio del test de Hausman mediante simulación de Monte Carlo
- Resultados y conclusiones principales

Modelo Bid



Hurtubia (2014)

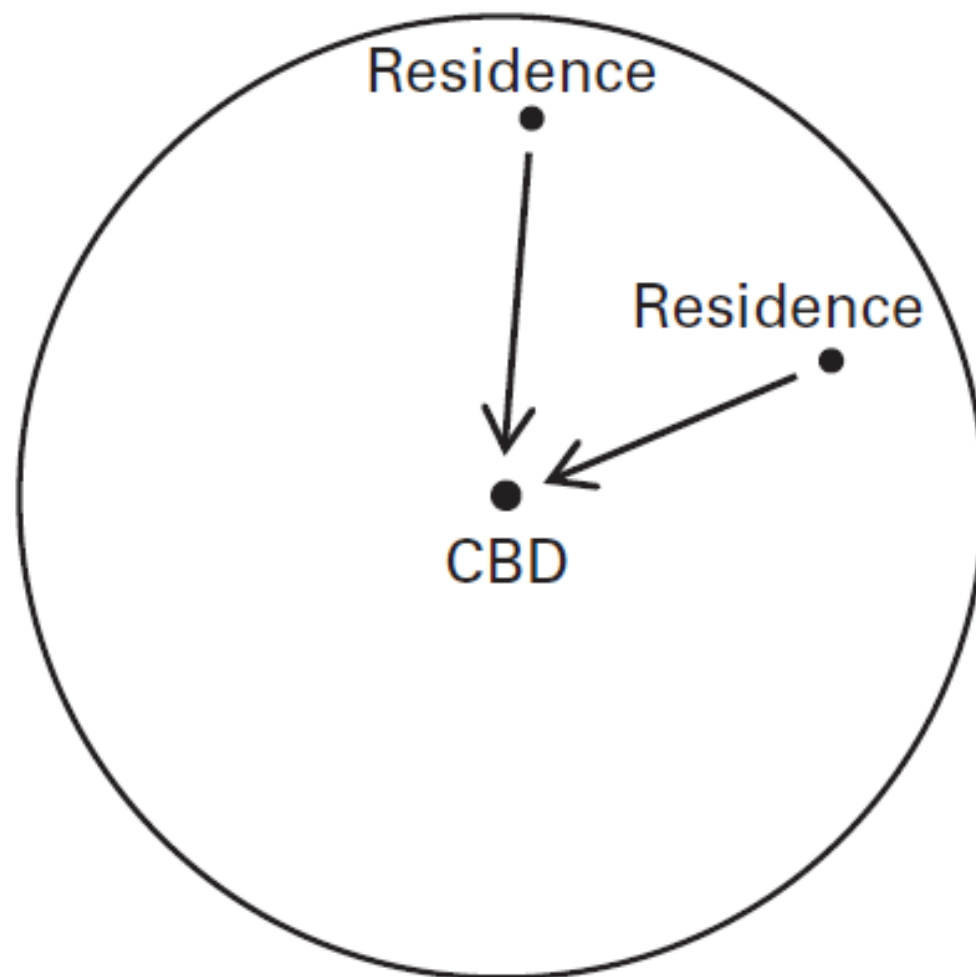
Modelo Bid



Hurtubia (2014)

$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \beta_{ruido,h} \cdot ruido_i$$

Modelo Bid



$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \beta_{ruido,h} \cdot ruido_i$$

Modelo Bid

Función de postura:

$$\tilde{B}_{hi} = B_{hi} + \varepsilon_{hi}$$

Componente aleatoria
(Elementos no medibles)

Componente sistemática

h: hogar
i: vivienda

$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \beta_{ruido,h} \cdot ruido_i$$

Endogeneidad en Modelo Bid

Por omisión de variables

$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \beta_{ruido,h} \cdot ruido_i$$

Endogeneidad en Modelo Bid

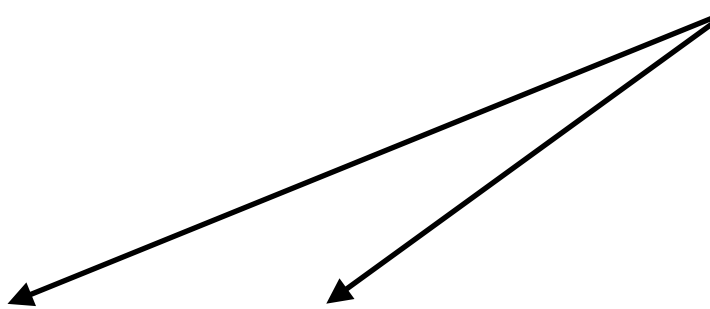
Por omisión de variables

$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \cancel{\beta_{ruido,h} \cdot ruido_i}$$

Endogeneidad en Modelo Bid

Por omisión de variables

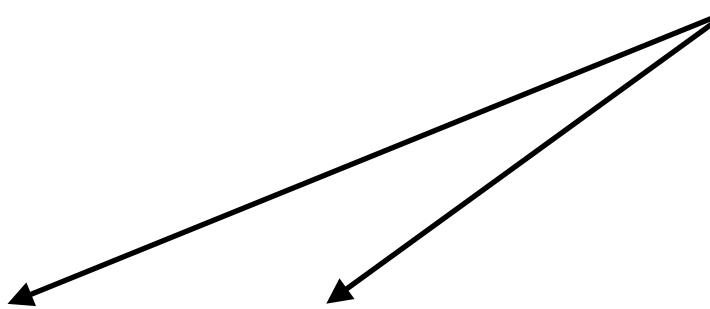
$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \cancel{\beta_{ruido,h} \cdot ruido_i}$$


$$\tilde{B}_{hi} = B_{hi} + \varepsilon_{hi}$$

Endogeneidad en Modelo Bid

Por omisión de variables

$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \cancel{\beta_{ruido,h} \cdot ruido_i}$$


$$\tilde{B}_{hi} = B_{hi} + \varepsilon_{hi}$$

¿Consecuencia?

Endogeneidad $\Rightarrow$ ¡Estimadores inconsistentes!

¿Cómo puedo detectar
endogeneidad en
modelos Bid?



Idea

Utilizar un test tipo

Hausman

Test de Hausman

Conceptos básicos

- Si existen dos estimadores **consistentes**, que varían **eficiencia**, ambos debieran ser similares
- Si difieren, es porque existe un problema de especificación => **endogeneidad**
- Modelo Bid puede **estimarse de dos maneras: Ellikson y Lerman&Kern**

Estimación del modelo Bid

Estimación del modelo Bid

Estimador de Ellickson (1981)

- Error distribuye Gumbel.
- Probabilidad: $P(h | i) = \mathbb{P}(\tilde{B}_{hi} \geq \tilde{B}_{h'i}; \forall h' \neq h; h, h' \in H)$

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

- Verosimilitud:

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

Estimación del modelo Bid

Estimador de Lerman y Kern (1983)

- Conocemos precio del remate (P_i^*) y podemos identificar la escala (μ): **Más eficiente**

- Probabilidad:

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

- Verosimilitud:

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

Comparar estimadores

Comparar estimadores

¿Posible?

Comparar estimadores

- No es directo $\Rightarrow$ Problemas de **identificación**
- Ellickson: $\widehat{\mu(ASC_h - ASC_{h'})}$, $\widehat{\mu(\beta_{sup,h} - \beta_{sup,h'})}$, $\widehat{\mu(\beta_{dist,h} - \beta_{dist,h'})}$
- Lerman y Kern: $\widehat{\mu}$, $\widehat{ASC_h}$, $\widehat{\beta_{sup,h}}$, $\widehat{\beta_{dist,h}}$
- Estimadores solo son comparables a nivel de **diferencia entre estratos de hogares**

Comparar estimadores

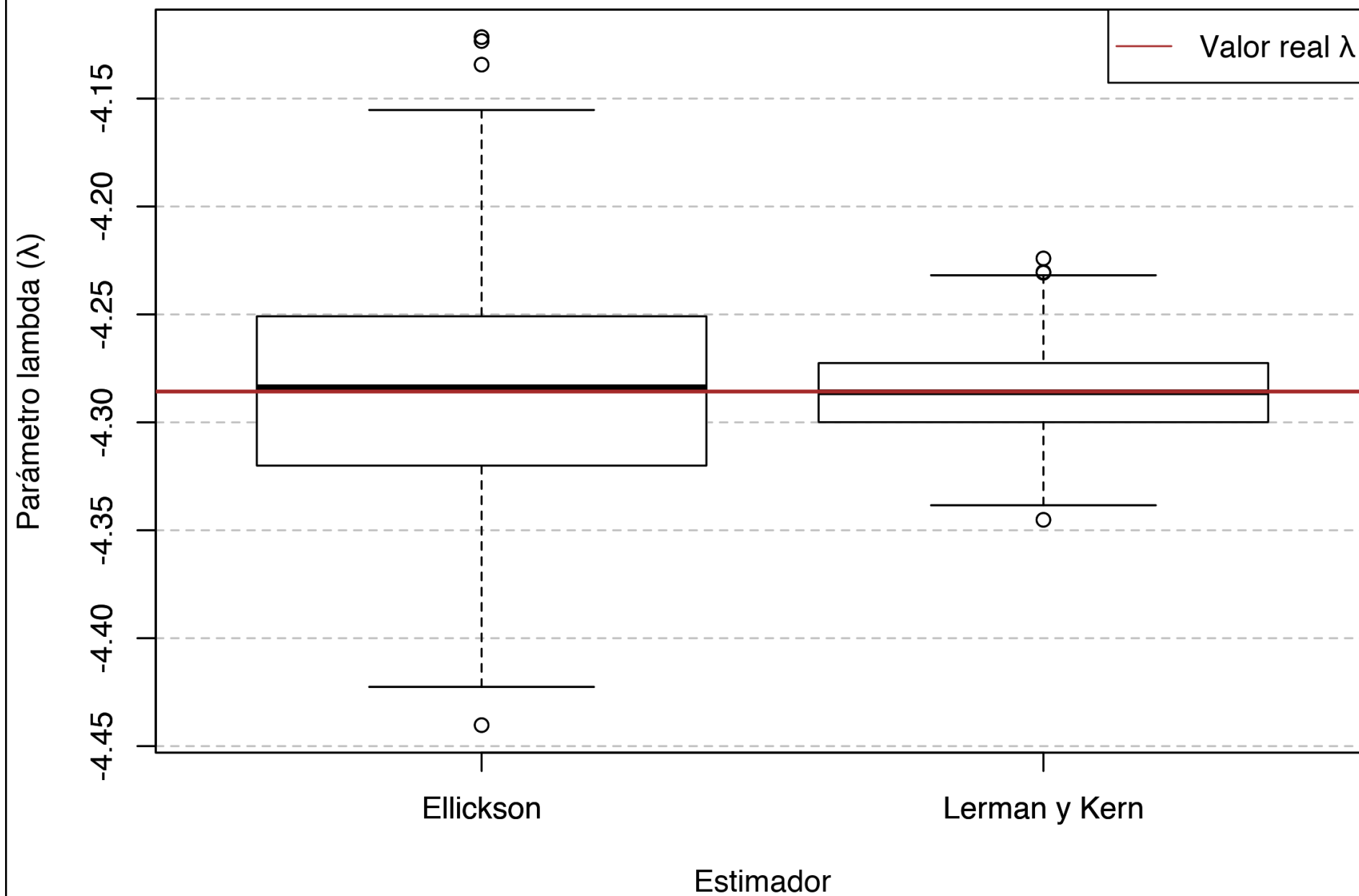
- Estadístico de comparación:

$$\hat{\lambda} = \frac{\beta_{\text{dist,alto}} - \beta_{\text{dist,bajo}}}{\beta_{\text{sup,alto}} - \beta_{\text{sup,bajo}}}$$

$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \beta_{ruido,h} \cdot ruido_i$$

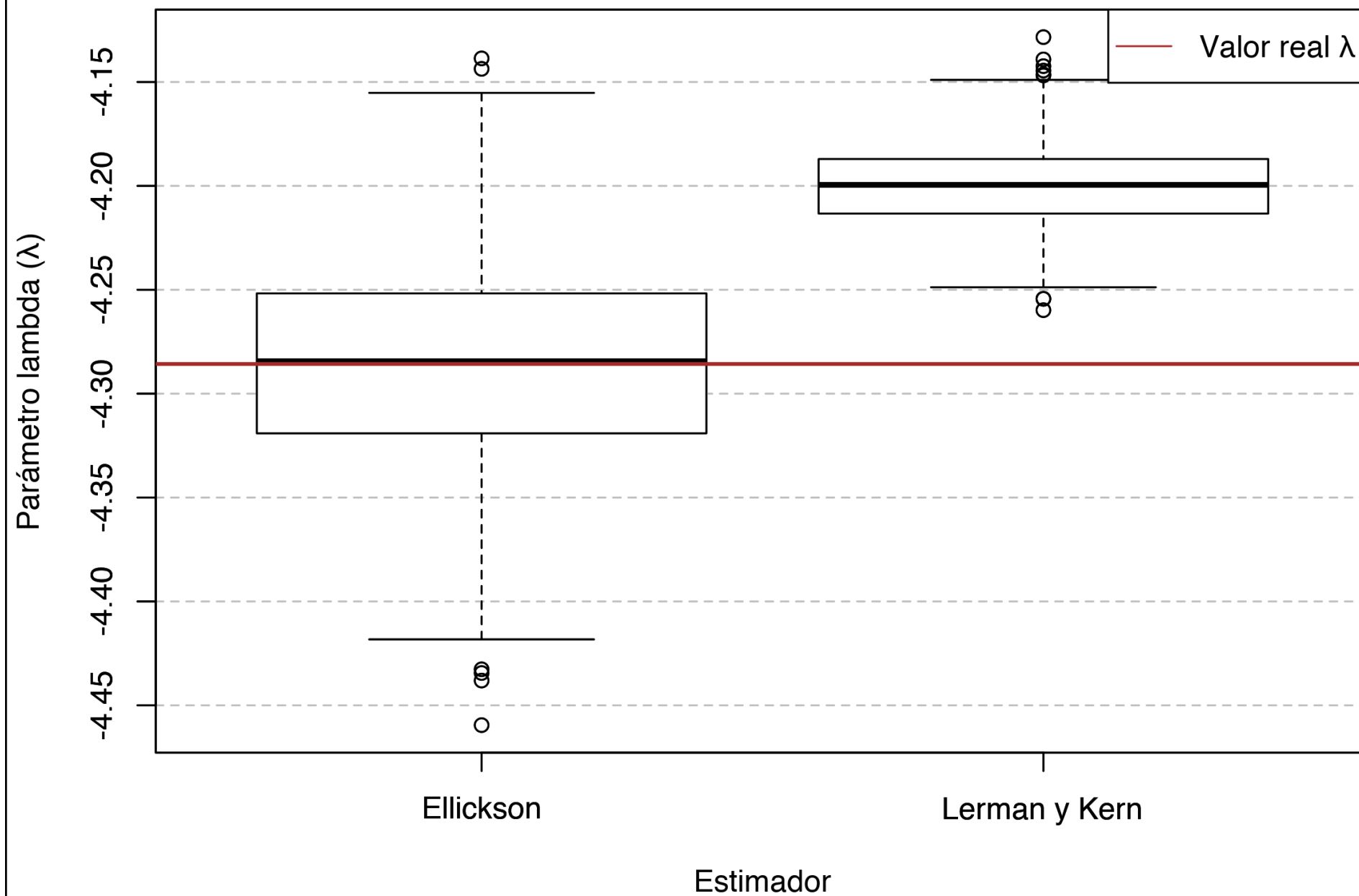
- Captura la diferencia de percepción entre estratos
- Al ser una razón entre parámetros, se elimina la escala
- Si no hay endogeneidad, el estadístico es consistente para ambos estimadores, por lo debiera ser muy similar

Simulación de Monte Carlo Estimación de parámetro λ — Caso sin endogeneidad



N=2.000, 1.000 realizaciones, error estándar < 1%

Simulación de Monte Carlo Estimación de parámetro λ — Caso endógeno



N=2.000, 1.000 realizaciones, error estándar < 1%

Test de Hausman-McFadden

- Hipótesis nula:

$$H_0 : \lim_{\mathbb{P}}(\widehat{\lambda}_E) - \lim_{\mathbb{P}}(\widehat{\lambda}_{LK}) = 0$$

- Estadístico H:

$$H = \frac{|\widehat{\lambda}_E - \widehat{\lambda}_{LK}|}{\sqrt{Var(\widehat{\lambda}_E) - Var(\widehat{\lambda}_{LK})}} \sim t$$

- Nivel de confianza del **95%**:
 - Rechazo H_0 cuando $H > 1,96$

Estudio del test

Estudio del test

- Size distortion

Estudio del test

- Size distortion
- Poder empírico

Errores

Error Tipo I (falso positivo)



Error Tipo II (falso negativo)



Adaptado de Ellis (2010). The Essential Guide To Effect Sizes. Cambridge University Press

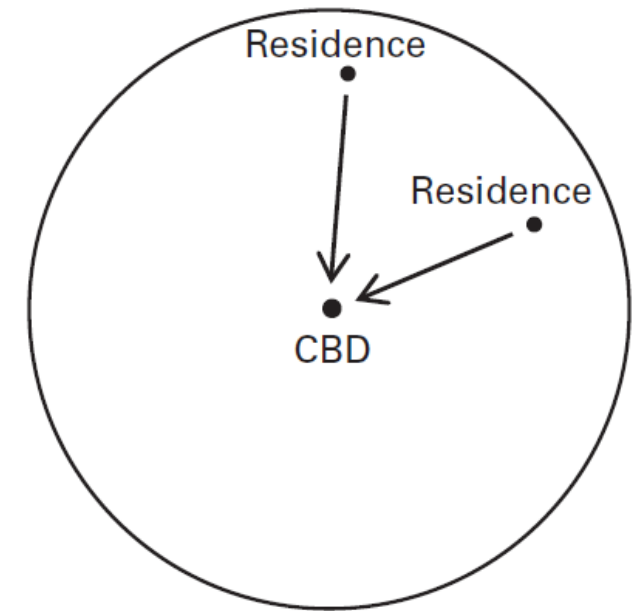
Indicadores

- **Size distortion:** Comparar probabilidad empírica de cometer un error tipo I con valor nominal asignado (5%).
 - Es deseable que valor empírico se asemeje al nominal y se determina el tamaño muestral (**N**) donde esto ocurre.
- **Poder estadístico:** Probabilidad de rechazar la hipótesis nula cuando la hipótesis alternativa es verdadera (Error tipo II) para un tamaño de efecto dado.

$$\mathbb{P}(H_0 \text{ es rechazado} | H_0 \text{ es falso})$$

Simulación de Monte Carlo

- Dos estratos: Alto y Bajo.
- Asignación por **ciudad abierta**.
 - Infinitos oferentes por estrato.
- Función de postura:



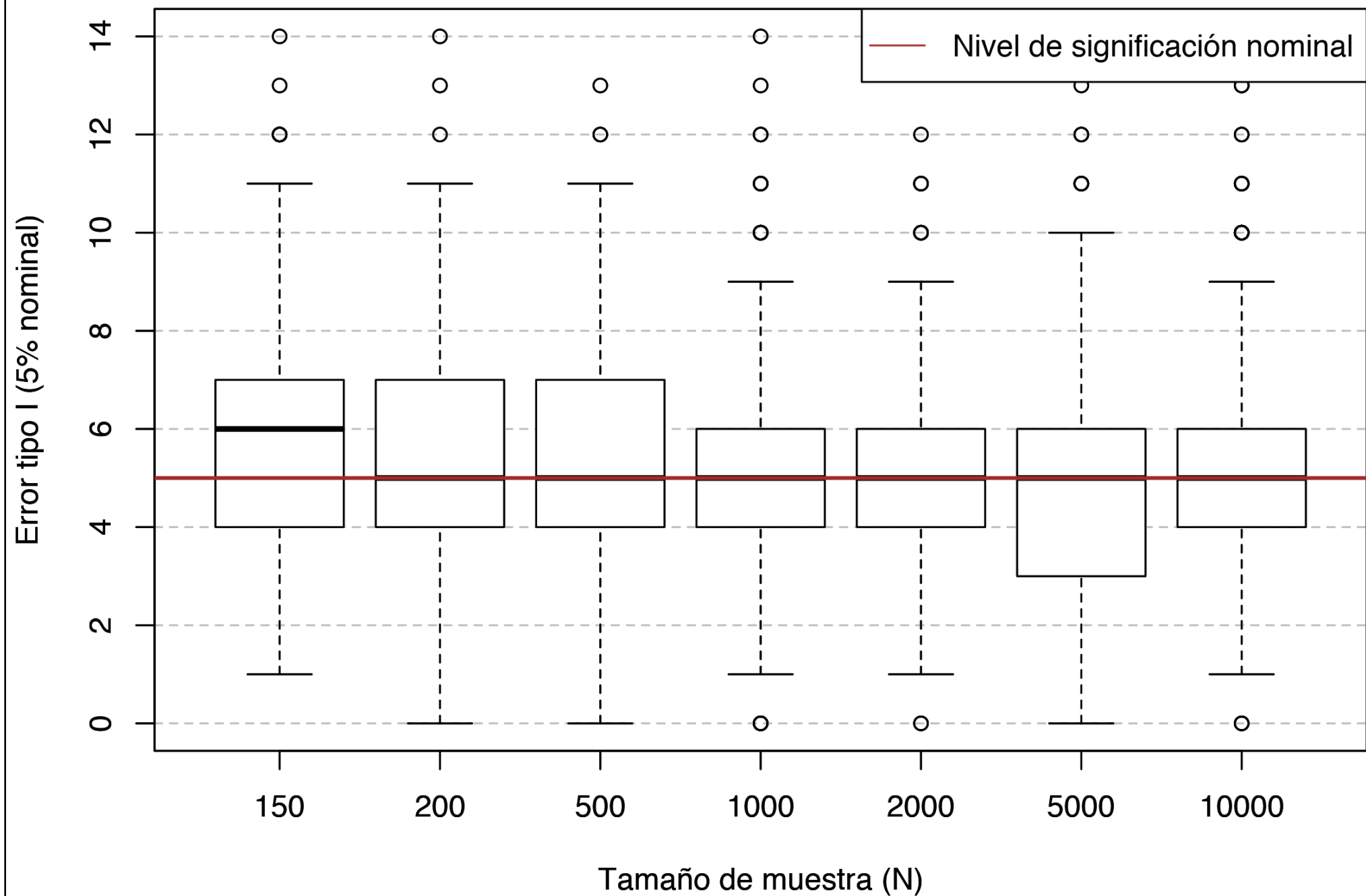
$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \beta_{ruido,h} \cdot ruido_i$$

- Remate **determina** distribución de hogares.

Resultados

Size distortion

Test de Hausman-McFadden modificado Análisis por «size distortion»



N=2.000, $\alpha=5\%$, 50.000 realizaciones, error nominal < 2%

Resultados

Poder empírico

Escenario A

Beta ruido igual por estrato

$$\beta = \beta_{\text{ruido,alto}} = \beta_{\text{ruido,bajo}} \leq 0$$

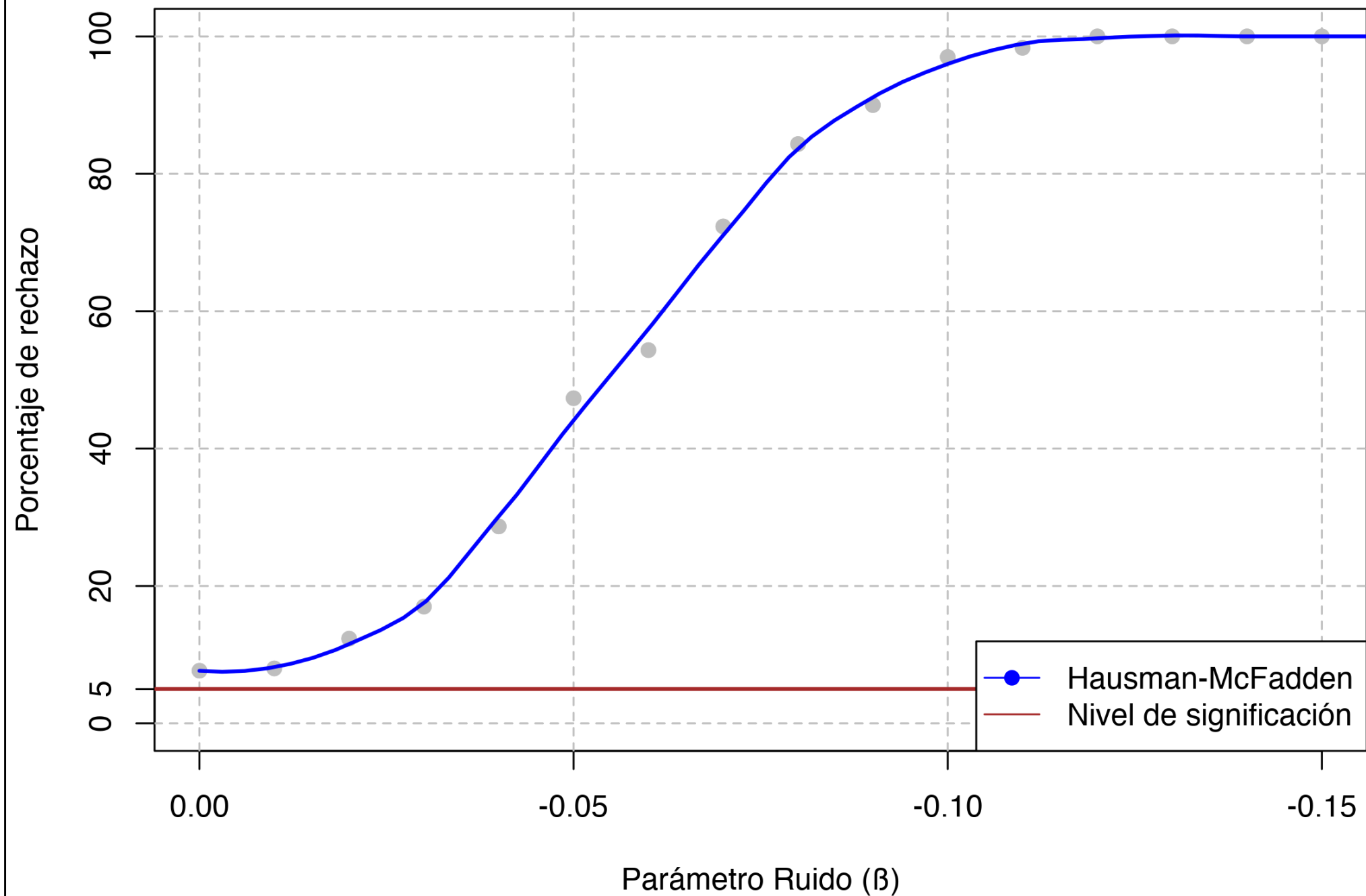
- $\beta = 0 \Rightarrow$ No se observa endogeneidad.

- $\beta \neq 0 \Rightarrow$

Altera la postura final pero no
ganador del remate

Ellickson es consistente
Lerman y Kern endógeno

Poder empírico por test
Escenario A: Betas iguales por estrato



N=2.000, $\alpha=5\%$, 10.000 realizaciones, error nominal < 5%

Escenario B

Beta ruido distinto por estrato

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

$$\beta_{\text{ruido,bajo}} = 0$$

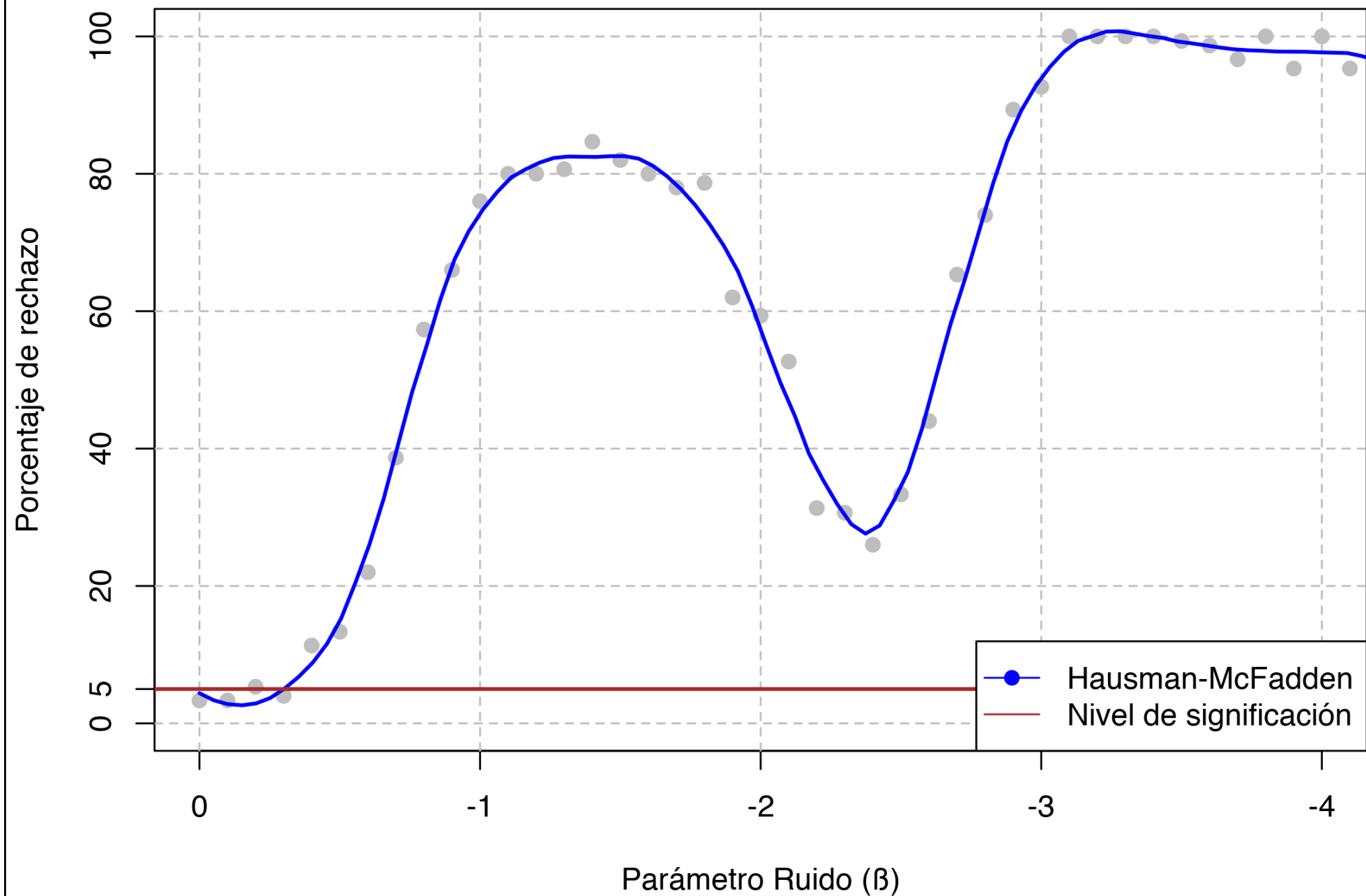
- $\beta = 0 \Rightarrow$ No se observa endogeneidad.

- $\beta \neq 0 \Rightarrow$

Cambia distribución por estrato
socioeconómico

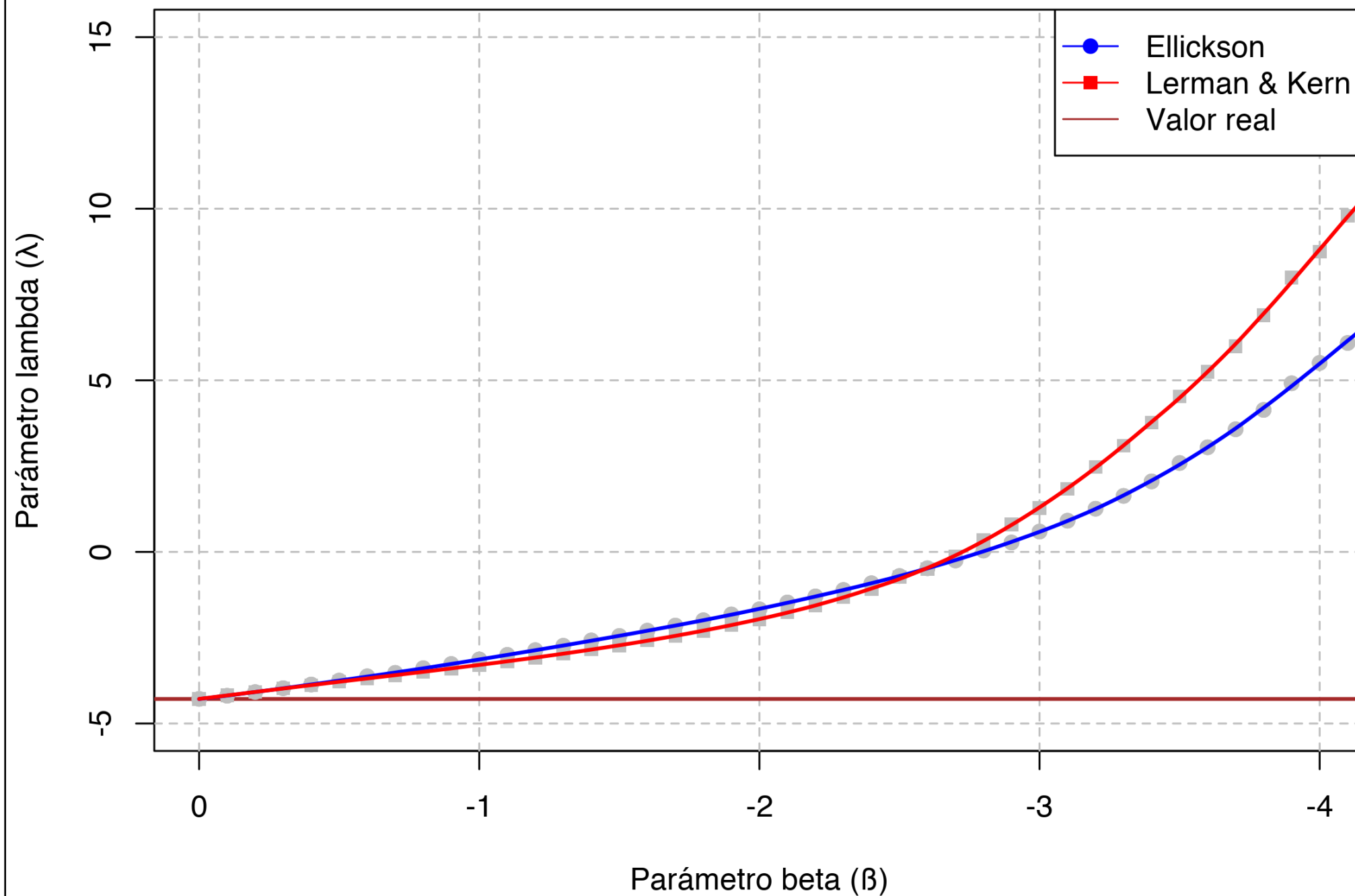
Tanto Ellickson como Lerman y Kern
se ven afectados

Poder empírico por test
Escenario B-II: Betas distintos por estrato



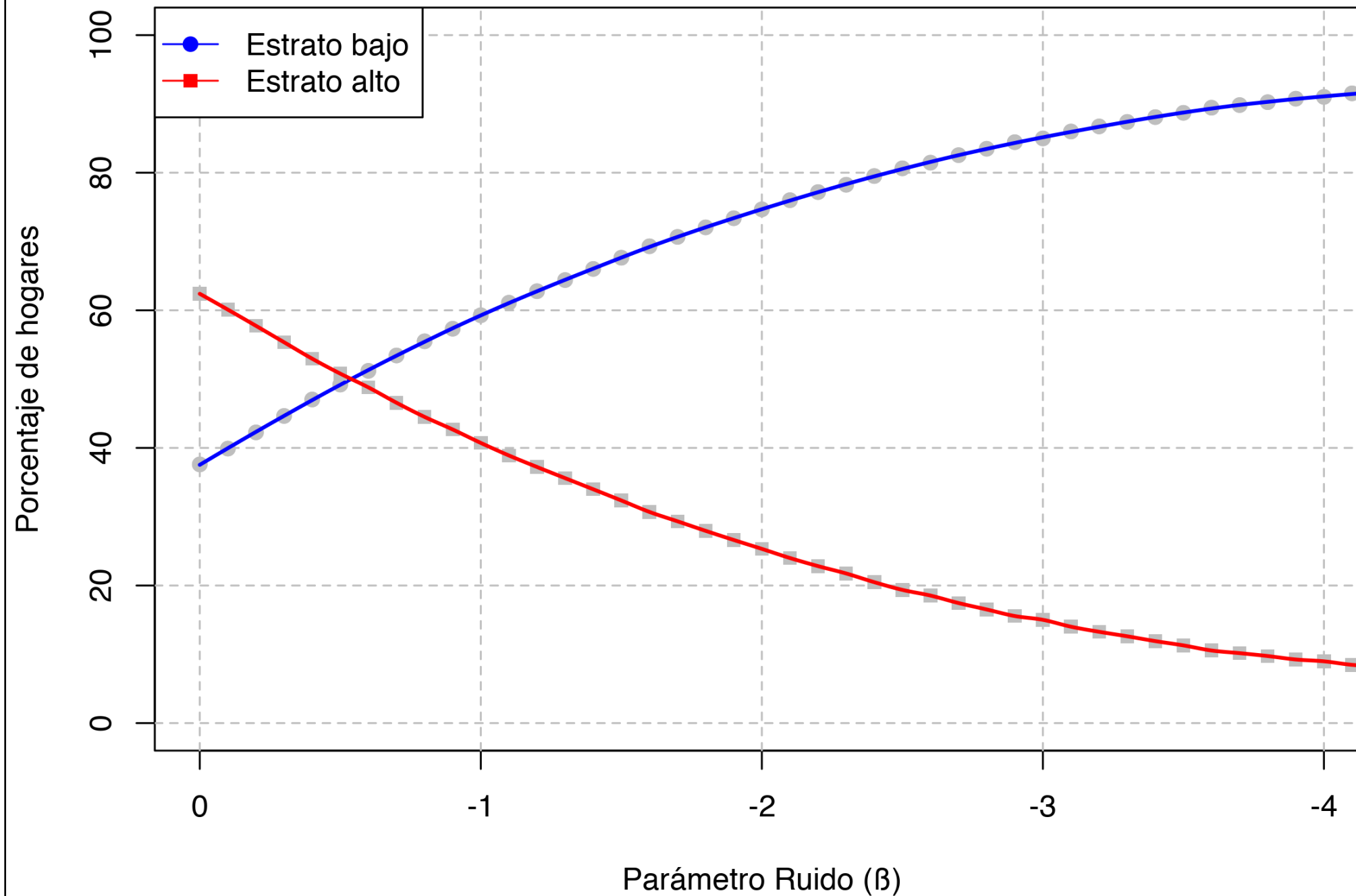
N=2.000, $\alpha=5\%$, 10.000 realizaciones, error nominal < 5%

Test de Hausman-McFadden modificado
Valor promedio de λ en Escenario B: Betas distintos por estrato



N=2.000, $\alpha=5\%$, 10.000 realizaciones, error nominal $< 5\%$

Distribución de hogares por estrato Escenario B: Betas distintos por estrato



N=2.000, $\alpha=5\%$, 10.000 realizaciones, error nominal < 5%

Síntesis y conclusión

- Se propuso y analizó test para detectar endogeneidad en modelos de postura en **ciudad abierta**
 - Size distortion está controlado
 - Mejor detección con var. omitida igual entre estratos
- Se estudió también el caso **ciudad cerrada**, mostrando conclusiones similares
- Propuesto: evaluación del test con datos reales

Evaluación de una prueba estadística para detectar endogeneidad en modelos Bid de uso de suelo

José Soto-Vaquero
CIS-T

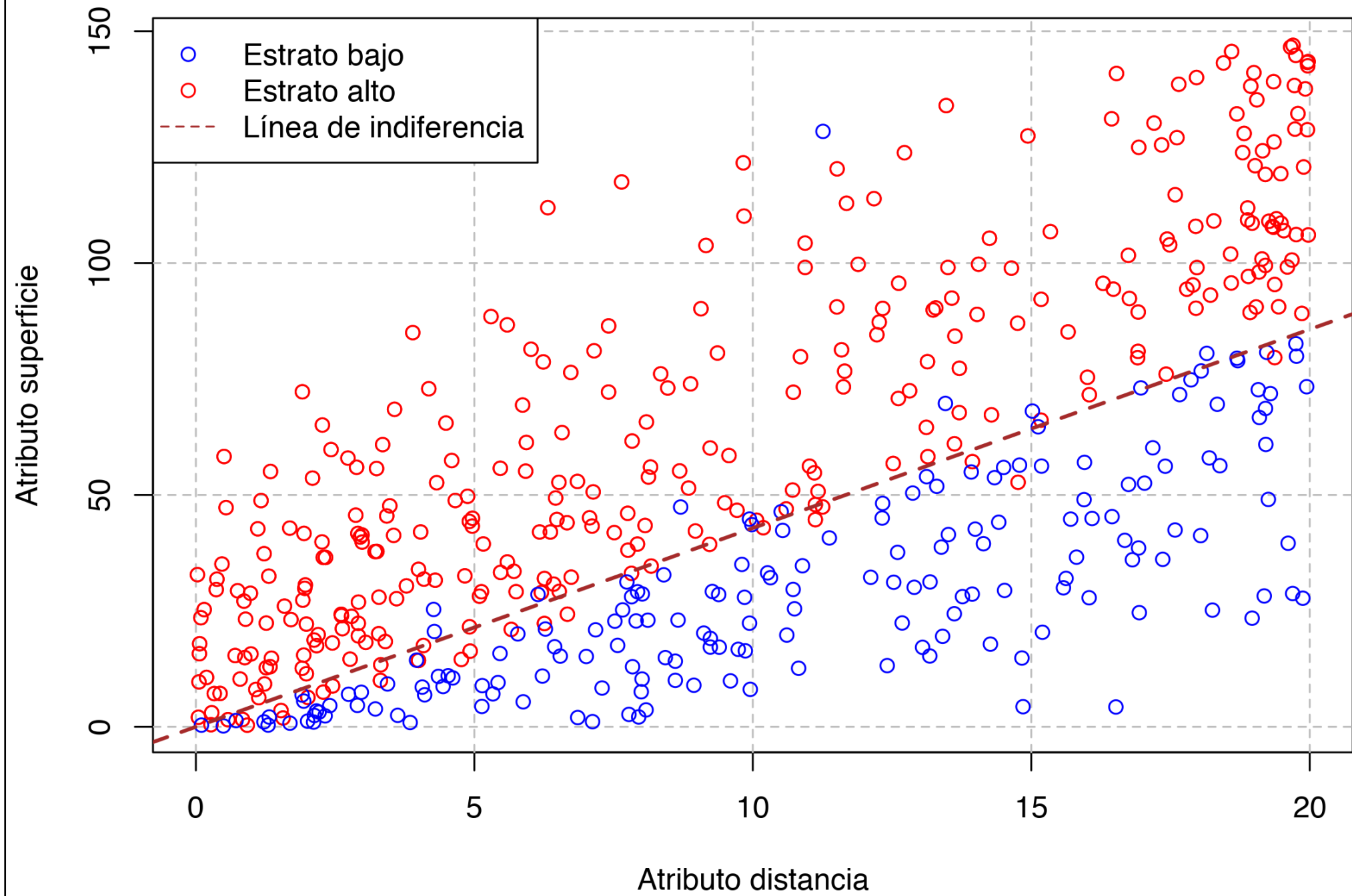
Ángelo Guevara Cue
Universidad de Chile

Anexos

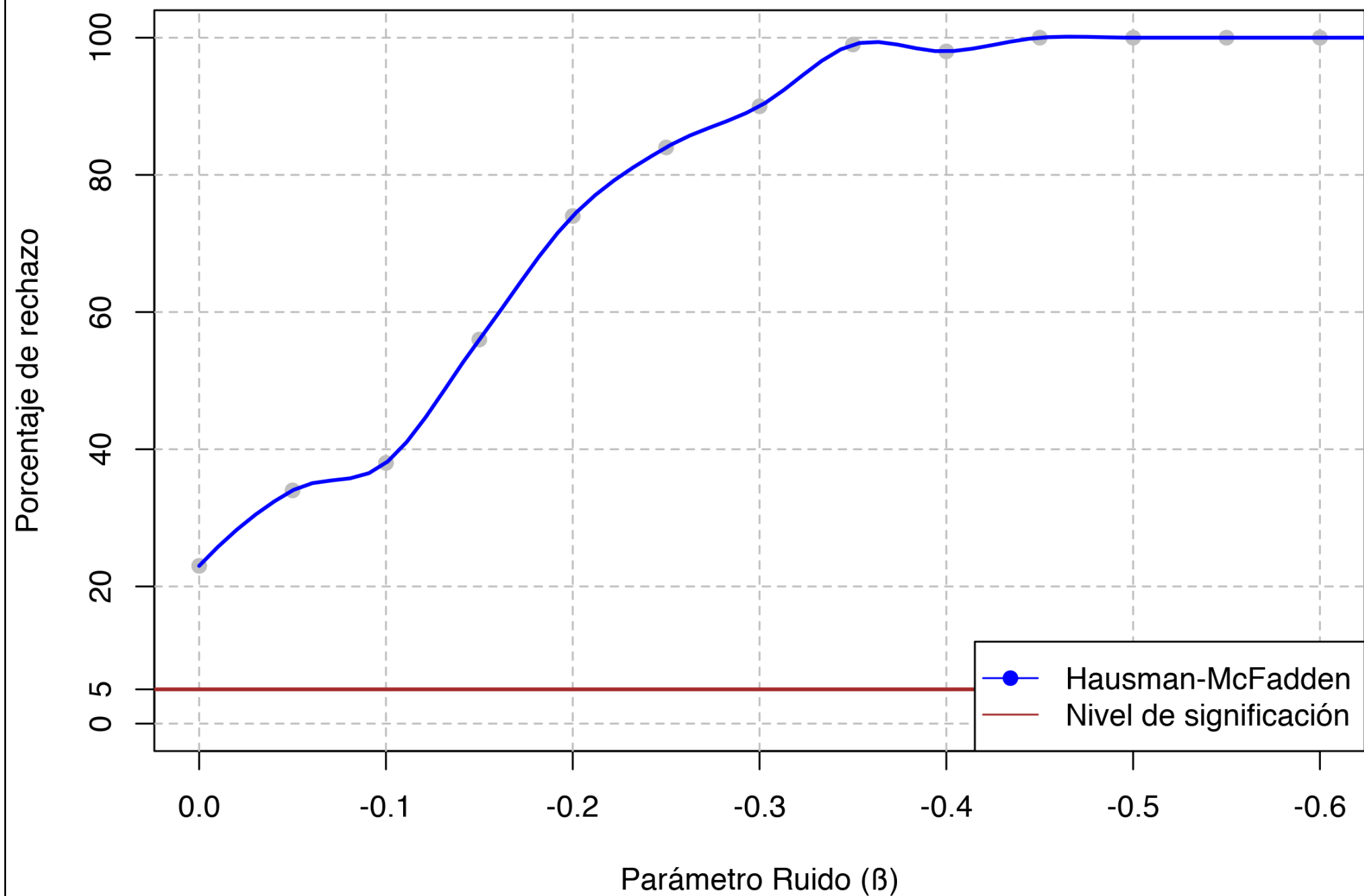
Ciudad cerrada

Distribución de hogares por estrato

Orden de asignación aleatorio

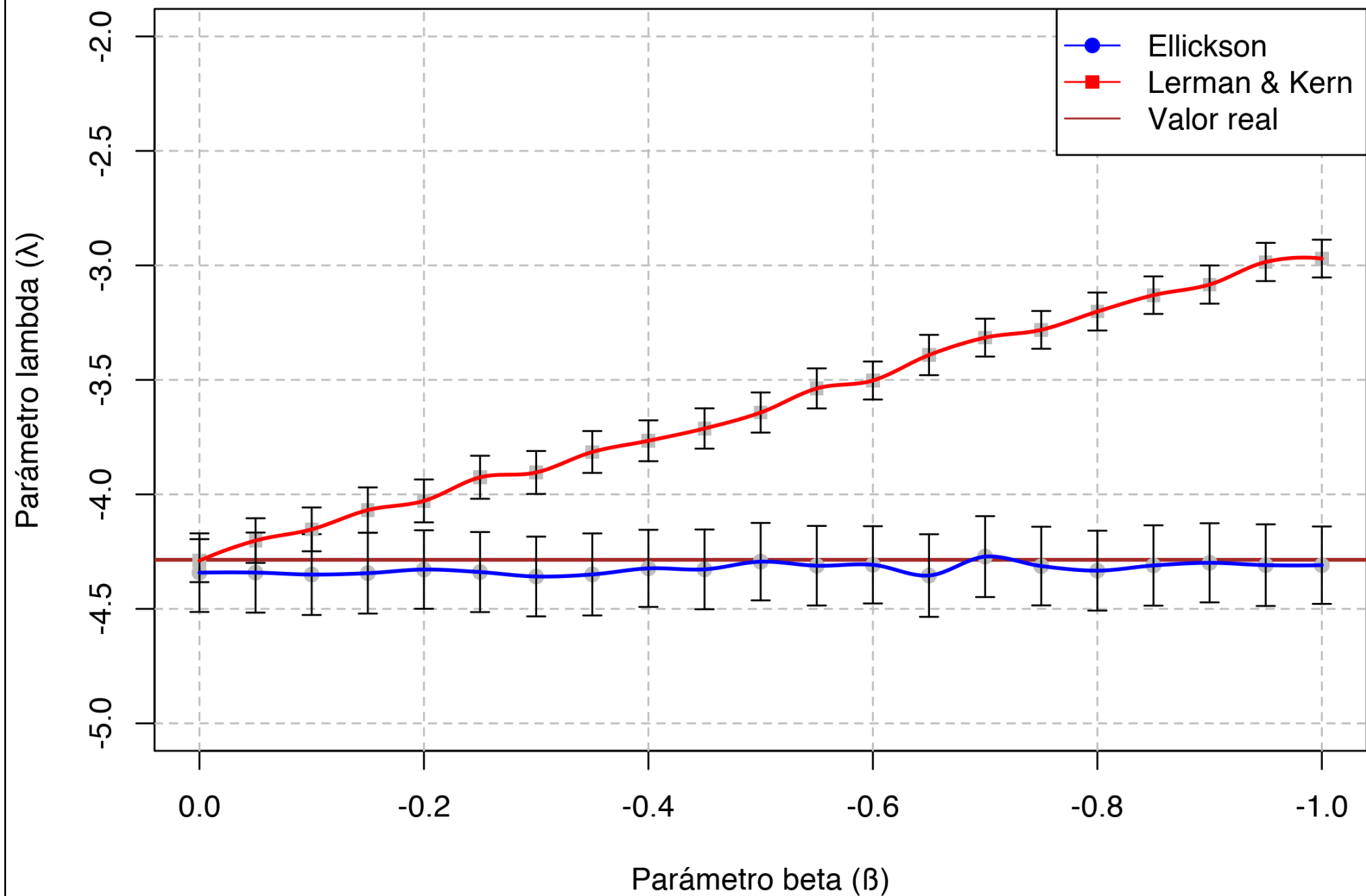


Poder empírico por test
Escenario A: Betas iguales por estrato — Ciudad Cerrada



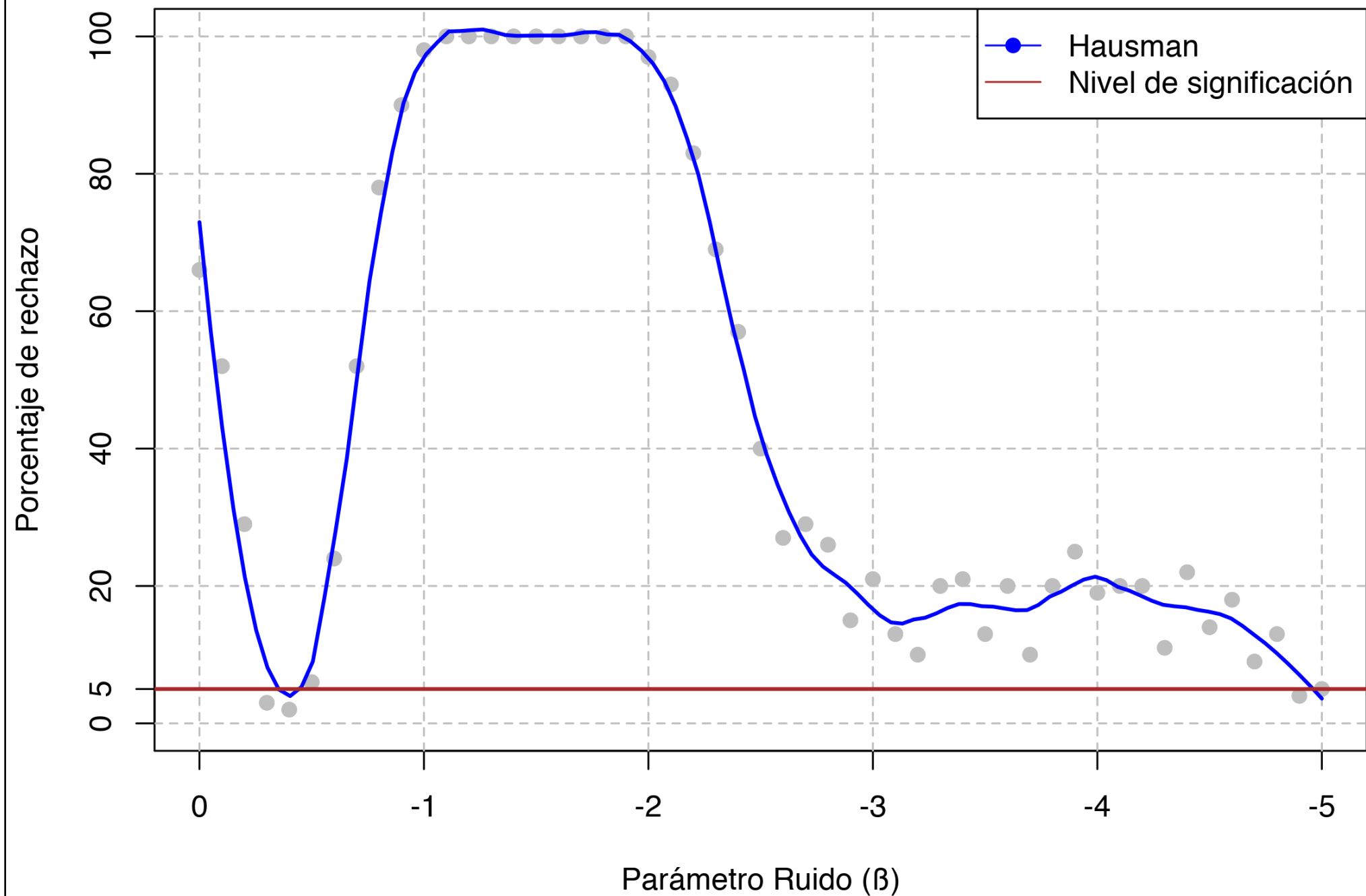
N=500. $\alpha=5\%$. 10.000 realizaciones. Error nominal < 5%.

Test de Hausman-McFadden modificado **Valor promedio de λ en Escenario A: Betas iguales por estrato**



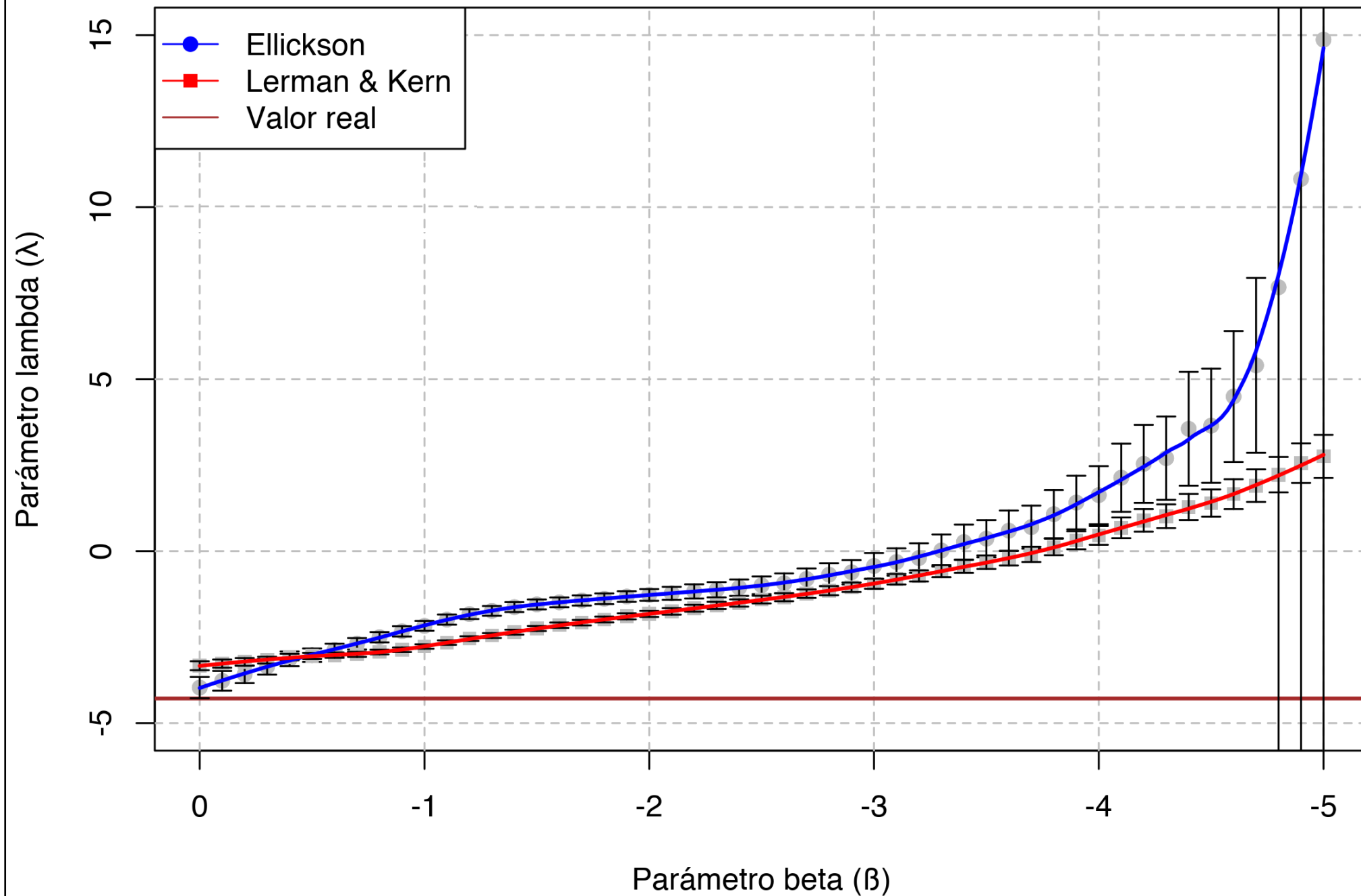
N=500. $\alpha=5\%$. 10.000 realizaciones. Error nominal < 5%.

Poder empírico por test **Escenario B: Betas distintos por estrato — Ciudad Cerrada**



N=500. $\alpha=5\%$. 10.000 realizaciones. Error nominal < 5%.

Test de Hausman-McFadden modificado **Valor promedio de λ en Escenario B: Betas distintos por estrato**



N=500. $\alpha=5\%$. 10.000 realizaciones. Error nominal < 5%.

Monte Carlo

Procedimiento

- Definir tamaño del efecto. (Escenarios)
- Simular elecciones utilizando una muestra aleatoria y la valoración de cada atributo; **conocidos de antemano.**

$$B_{hi} = ASC_h + \beta_{sup,h} \cdot sup_i + \beta_{dist,h} \cdot dist_i + \cancel{\beta_{ruido,h} \cdot ruido_i}$$

- Estimar modelos Bid con atributo omitido.
- Comparar estimadores utilizando test estadísticos.

Experimento de Monte Carlo

- Parámetros de elección

Hogar	ASC	β_{sup}	β_{dist}
Estrato Bajo	0	0,4	-0,5
Estrato Alto	0	0,55	-2

Atributos

- **Superficie:** Normal truncada

- $a=0$; $b=100$; $\mu=0$; $\sigma=75$.

- **Distancia:** Uniforme.

- $a=0$; $b=20$.

- **Ruido:** Uniforme.

- $a=0$; $b=10$.

- **Variable instrumental** (CF): Normal.

- $\mu=10$; $\sigma=5$, $\sigma_{\text{ruido, VI}} = 0,7$; $\sigma_{\text{VI, distancia}} = 0$

Correlaciones

A priori

$$\begin{bmatrix} 1 & 0,8 & 0 & 0 \\ & 1 & -0,9 & 0,5 \\ & & 1 & 0 \\ & & & 1 \end{bmatrix}$$

A posteriori [copula]

$$\begin{bmatrix} 1 & 0,62 & -0,06 & 0,04 \\ & 1 & -0,7 & 0,4 \\ & & 1 & -0,03 \\ & & & 1 \end{bmatrix}$$

Test de Funciones de Control

Definiciones

- Modelo inicial (reducido):

$$B_{hi} = ASC_h + \beta_{sup} sup_{hi} + \beta_{dist} dist_{hi} + \varepsilon_{hi} \implies L^{RED}$$

- Uso de variable instrumental:

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

- Modelo final (completo):

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

Test de Funciones de Control

- Hipótesis nula:

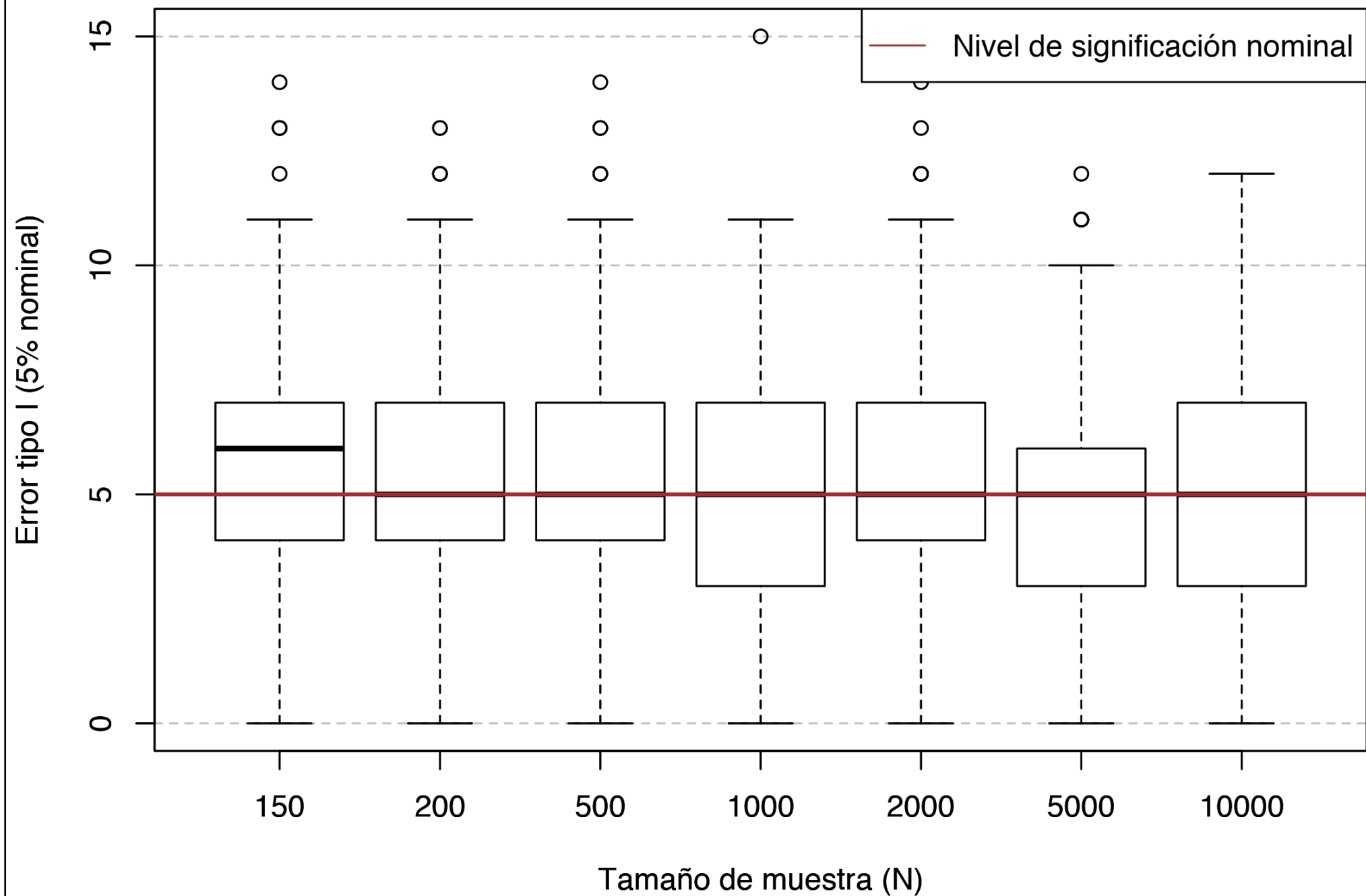
$$H_0 : \beta_\delta = 0$$

- Test de razón de verosimilitud:

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

- Nivel de confianza del 95%:
 - Dos grados de libertad.
 - Rechaza H_0 cuando $H > 5,99$.

Test de Funciones de Control **Análisis por «size distortion»**



N=2000. $\alpha=5\%$. 50.000 realizaciones. Error estándar < 2%.