TEMA 3:
PROPIEDADES DEL ESTIMADOR MCO

S. Álvarez, A. Beyaert, M. Camacho, M. González, A. Quesada
Departamento de Métodos Cuantitativos para la Economía y la Empresa
UNIVERSIDAD DE
MURCIA

UNIVERSIDAD DE MURCIA

Lo que estudiaremos en este tema:

1. El valor esperado del estimador MCO
1.1 Supuestos necesarios para la insesgadez
1.2 Errores de especificación e insesgadez del estimador MCO

2. La varianza del estimador MCO y su estimación
2.1 La varianza del estimador MCO
2.2 Multicolinealidad fuerte y varianzas MCO
2.3 El estimador de la varianza del estimador MCO
2.4 Varianzas en modelos mal especificados

3. Optimalidad del estimador MCO
3.1 Supuestos necesarios
3.2 El Teorema de Gauss-Markov

4. Consistencia del estimador MCO
4.1 Introducción
4.2 Supuestos necesarios
4.3 Teorema

Bibliografía básica: Wooldridge, 2006, cap. 3 y 5

- Propiedad de insesgadez:

$$E\left(\hat{\beta}_j / x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}\right) = \beta_j,\ j = 0,1,\cdots,k$$

- $E\left(\hat{\beta}_j\right)$ es una medida de **posición** del estimador

- El estimador MCO de los coeficientes poblacionales $\beta_0, \beta_1, \cdots, \beta_k$ es **insesgado** si se cumplen **ciertos supuestos:**

  – RLM.1 Modelo lineal en los parámetros

  – RLM.2 Muestreo aleatorio

  – RLM.3 Media condicionada nula: $\rightarrow$ $\begin{cases} E(\varepsilon_i \mid x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}) = E(\varepsilon_i) \\ E(\varepsilon_i) = 0 \qquad \forall i = 1, \ldots, N \end{cases}$

  – RLM.4 No multicolinealidad perfecta

(“RLM”=Regresión Lineal Múltiple)

- RLM.1 Modelo lineal en los parámetros

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

Supuesto bastante flexible: permite usar funciones de las variables de interés

Ejemplo:

$$\log(\text{salario}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \exp er + \beta_3 \exp er^2 + \varepsilon$$

¿Cómo se interpretan los coeficientes de este ejemplo?

UNIVERSIDAD DE MURCIA

- RLM.2 Muestreo aleatorio

Los datos necesarios para estimar el modelo se han obtenido por muestreo aleatorio

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i \qquad i = 1, \cdots, N$$

Por tanto, los datos y el error de la observación "i" y de la observación "h" son independientes

$$\Rightarrow \operatorname{cov}(\varepsilon_i, \varepsilon_h) = 0 \quad \forall i \neq h \text{ con } i, h = 1, \cdots, N$$

UNIVERSIDAD DE MURCIA

- RLM.3 Media condicionada nula

El valor esperado del error, condicionado a las explicativas, es igual al valor esperado no condicionado, y vale cero

$$\begin{cases} E(\varepsilon_i | x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}) = E(\varepsilon_i) & \forall i = 1, \cdots, N \quad \text{(RLM.3a)} \\ E(\varepsilon_i) = 0 & \forall i = 1, \cdots, N \quad \text{(RLM.3b)} \end{cases}$$

- $\text{(RLM.3a)} \equiv$ **exogeneidad estricta**

  $\Rightarrow \varepsilon_i$ incorrelacionado con las N observaciones de cada una de las explicativas

- $\text{(RLM.3b)} \equiv$ **media nula**

- **exogeneidad débil** :

$$E(\varepsilon_i | x_{1i}, x_{2i}, \cdots, x_{ki}) = E(\varepsilon_i) \ \forall i \text{ con } i = 1, \cdots, N$$

$\Rightarrow \varepsilon_i$ incorrelacionado con la observación i-ésima de cada una de las explicativas

- Con muestreo aleatorio, exogeneidad débil $\Rightarrow$ exogeneidad estricta

Por tanto:
- RLM.2
- Exogeneidad débil
- RLM.3b

$\Rightarrow$ RLM.3

- Si no hay exogeneidad débil, por lo menos una explicativa, x_{ji} , está correlacionada con ε_i , y esta variable explicativa es **endógena.**

- RLM.4 No multicolinealidad exacta

  En la muestra ninguna de las variables explicativas es constante, y no existen relaciones *lineales* *exactas* entre las explicativas

- se refiere *sólo* a las variables explicativas en la muestra.
- No excluye cierta correlación -*no perfecta*- entre las explicativas

¿Puede cumplirse el supuesto RLM.4 en este ejemplo?

$$\log(salario) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \varepsilon$$

- Si hay multicolinealidad exacta, **no es posible obtener el estimador MCO** porque **no está bien definido**
- ¿Por qué?

Ejemplo: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} r_{1i} y_i}{\sum_{i=1}^{N} r_{1i}^2}$$

donde r_{1i} = residuo de regresar x_1 sobre x_2, y un término constante
es la parte de x_1 no relacionada con x_2

pero si $x_{1i} = kx_{2i}$ (multicolinealidad exacta): $r_{1i} = 0 \forall i$

- En general, la multicolinealidad exacta se debe a un error del analista

UNIVERSIDAD DE MURCIA

- **Teorema de insesgadez**

  Bajo los supuestos RLM.1 a RLM.4,

$$E\left(\hat{\beta}_j / x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}\right) = \beta_j, \; j = 0, 1, \cdots, k$$

  para cualquiera de los valores del coeficiente poblacional.

- Los estimadores MCO son estimadores insesgados de los coeficientes poblacionales
- No se puede hablar de la insesgadez de una estimación, sino del estimador.
- Demostración: Wooldridge p. 54 para modelo de regresión simple

Nota: Todas las esperanzas y varianzas están condicionadas a los valores muestrales de las variables explicativas, pero para simplificar notación en las ecuaciones del resto del tema no aparece dicha condición, salvo cuando sea útil.

- Dos tipos frecuentes de error de especificación:

  - **Inclusión de irrelevantes:**

    una (o más) variable(s) explicativa(s) incluida(s) no tiene(n) ningún efecto sobre la variable dependiente del modelo => modelo sobreespecificado

  - **Omisión de relevantes**:

    se omite una (o más) variable(s) explicativa(s) que sí tiene(n) efecto sobre la variable dependiente del modelo => modelo subespecificado

- Tienen efectos distintos sobre la insesgadez

UNIVERSIDAD DE MURCIA

- Efecto de la inclusión de irrelevantes:

modelo verdadero: $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$, cumple RLM.1-RLM.4

modelo planteado: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad \Rightarrow \beta_2 = 0$

modelo estimado: $\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{1i} + \tilde{\beta}_2 x_{2i}$

- Se siguen cumpliendo RLM.1-RLM.4 $\Rightarrow E\left(\tilde{\beta}_j\right) = \beta_j \quad \forall j = 0,1,2 \quad \forall \beta_j$

$$\Rightarrow \begin{cases} E\left(\tilde{\beta}_0\right) = \beta_0 \\ E\left(\tilde{\beta}_1\right) = \beta_1 \\ E\left(\tilde{\beta}_2\right) = \beta_2 = 0 \end{cases}$$

la inclusión de irrelevantes no afecta a la insesgadez

UNIVERSIDAD DE **MURCIA**

- <u>Efecto de la omisión de relevantes:</u>

modelo verdadero: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$, $\beta_2 \neq 0$, cumple RLM.1-RLM.4

modelo planteado: $y_i = \beta_0 + \beta_1 x_{1i} + \nu_i \Rightarrow \nu_i = \beta_2 x_{2i} + \varepsilon_i$, puede fallar RLM.3

modelo estimado: $\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{1i}$

- Usando el modelo *verdadero*, $E\left(\hat{\beta}_1\right) = \beta_1$

- Pero con el modelo mal especificado:

$$E(\tilde{\beta}_1) = \beta_1 + \boxed{\beta_2 \frac{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sum_{i=1}^{N}(x_{1i} - \bar{x}_1)^2}} \rightarrow$$

Sesgo de mala especificación, función de:
- β_2
- cov. muestral entre x_1 y x_2 (si no es cero, falla RLM.3)

UNIVERSIDAD DE MURCIA

- ¿Por qué?

$$\tilde{\beta}_1 = \frac{\sum (x_{1i}-\bar{x}_1)(y_i-\bar{y})}{\sum (x_{1i}-\bar{x}_1)^2} = \frac{\sum (x_{1i}-\bar{x}_1)\left(\beta_1(x_{1i}-\bar{x}_1)+\beta_2(x_{2i}-\bar{x}_2)+(\varepsilon_i-\bar{\varepsilon})\right)}{\sum (x_{1i}-\bar{x}_1)^2}$$

$$=\beta_1+\beta_2\frac{\sum (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sum (x_{1i}-\bar{x}_1)^2}+\frac{\sum (x_{1i}-\bar{x}_1)(\varepsilon_i-\bar{\varepsilon})}{\sum (x_{1i}-\bar{x}_1)^2}$$

$$E\left(\tilde{\beta}_1\right)=E\left[\beta_1+\beta_2\frac{\sum (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sum (x_{1i}-\bar{x}_1)^2}+\frac{\sum (x_{1i}-\bar{x}_1)(\varepsilon_i-\bar{\varepsilon})}{\sum (x_{1i}-\bar{x}_1)^2}\right]=$$

$$=\beta_1+\beta_2\frac{\sum (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sum (x_{1i}-\bar{x}_1)^2}$$

- No sólo es importante el signo del sesgo sino también el tamaño del mismo.

- Ejemplo: salario en función del nivel de estudios y de la habilidad

Modelo verdadero (estimado): $\hat{sal}_i = \hat{\beta}_0 + \hat{\beta}_1 ne_i + \hat{\beta}_2 habil_i$

Modelo planteado (estimado): $\tilde{sal}_i = \tilde{\beta}_0 + \tilde{\beta}_1 ne_i$

$$\Rightarrow \quad Sesgo\left(\tilde{\beta}_1\right) = \beta_2 \frac{\sum (ne_i - \overline{ne})(habil_i - \overline{habil})}{\sum (ne_i - \overline{ne})^2}$$

- Es probable que $Cov(ne, habil) > 0$ y también $\beta_1 > 0, \beta_2 > 0$
- Por tanto:
  - o Sobreestimación de β_1
  - o Sobrevaloración del impacto de ne sobre sal

UNIVERSIDAD DE MURCIA

- Necesitamos una medida de la **<u>dispersión</u>** de $\hat{\beta}_j$ alrededor del valor central $E\left(\hat{\beta}_j\right)$: $V\left(\hat{\beta}_j\right)$

- $V\left(\hat{\beta}_j\right)$ depende de las características del modelo.

- Característica más simple: <u>Supuesto RLM.5: homoscedasticidad</u>

$$V(\varepsilon_i | x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}) = V(\varepsilon_i) = \sigma^2 \;\; \forall i = 1, \cdots, N$$

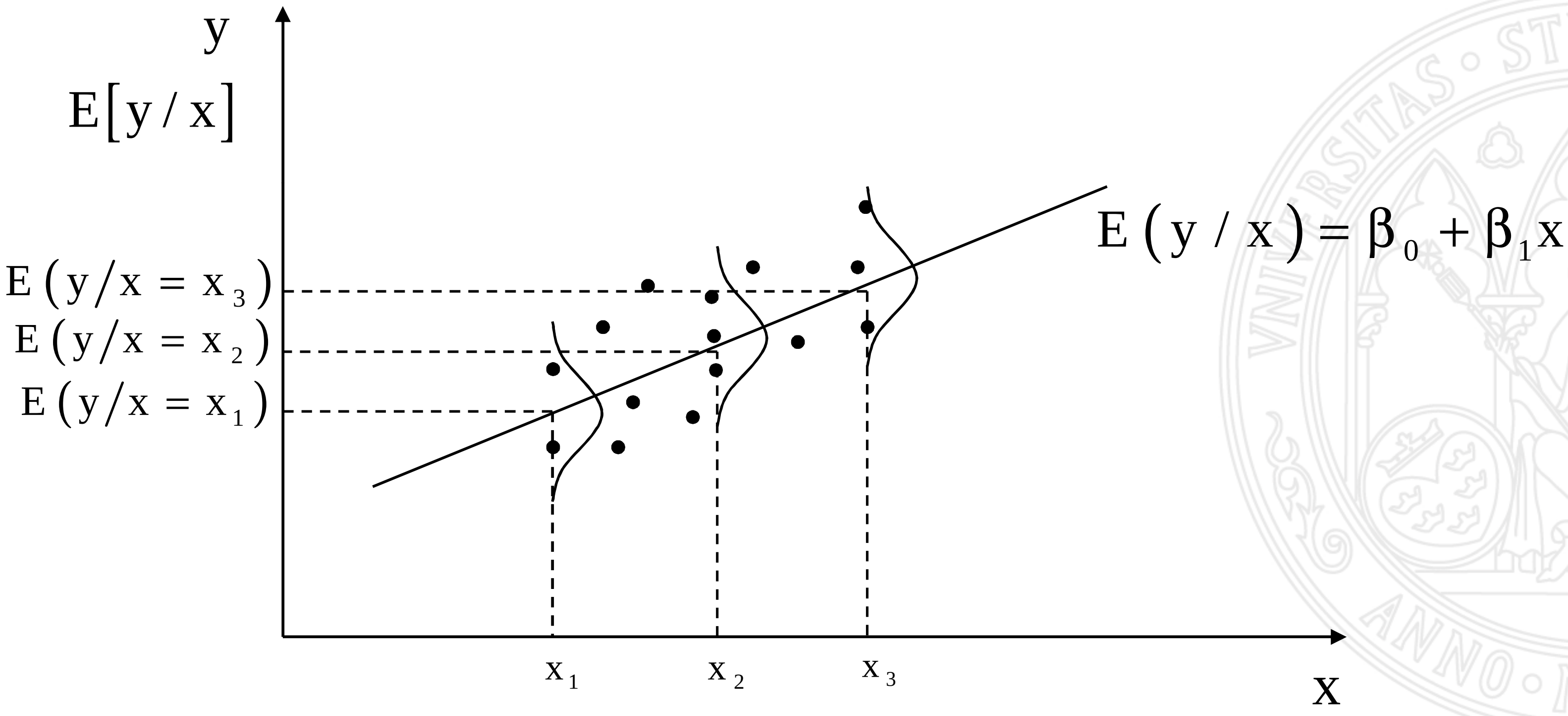
$$\Rightarrow V(y_i | x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}) = V(\varepsilon_i | x_{11}, x_{21}, \cdots, x_{(k-1)N}, x_{kN}) = \sigma^2 \;\; \forall i$$

La varianza del error (y por tanto de y) es independiente de los valores de las explicativas, y además es constante.

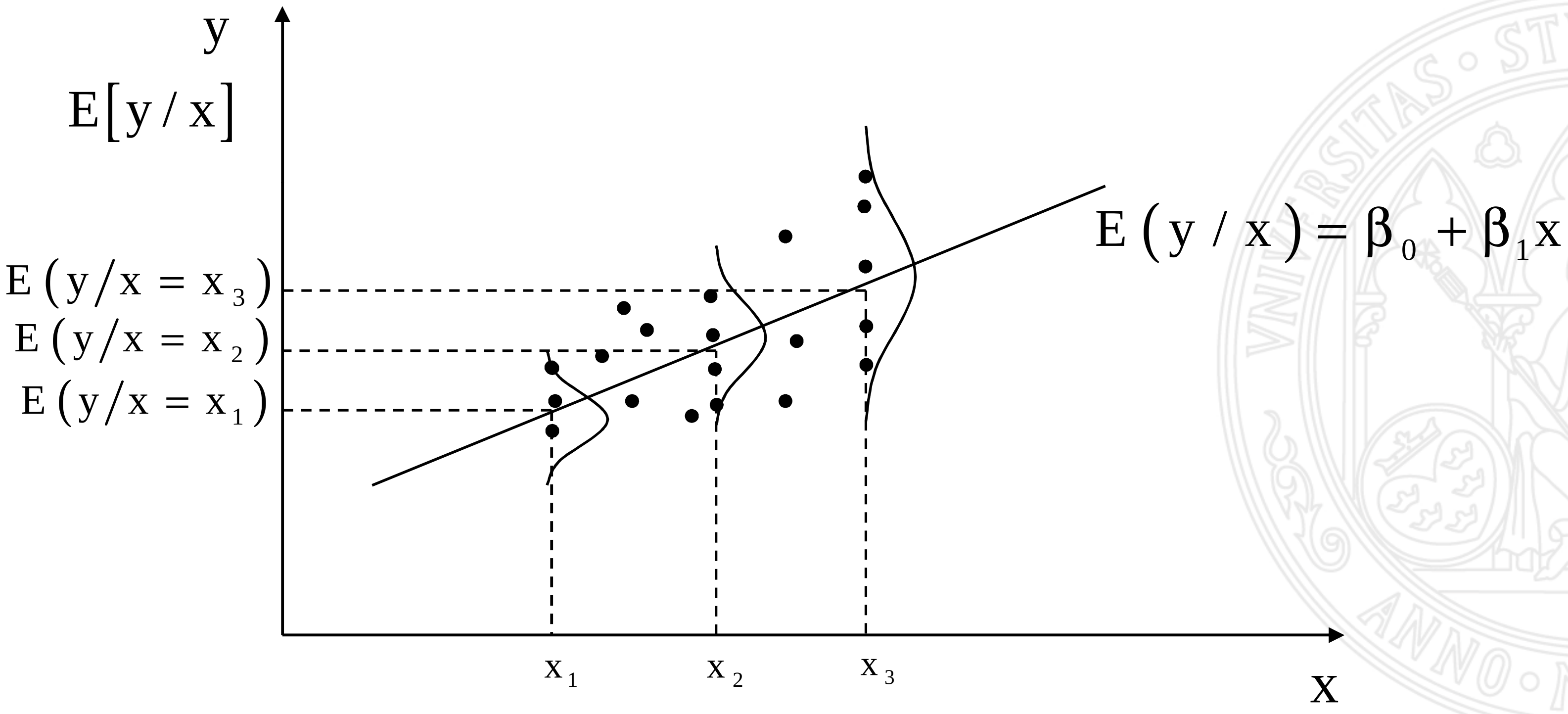
➡ homoscedasticidad $\equiv$ “la varianza del error es constante “

UNIVERSIDAD DE MURCIA

- Homoscedasticidad:



- Heteroscedasticidad:



UNIVERSIDAD DE MURCIA

- **<u>Teorema de las varianzas muestrales de los estimadores MCO de los coeficientes de los regresores:</u>**

Bajo RLM.1 a RLM.5, condicionando a los valores muestrales de las variables explicativas,

$$V\left(\hat{\beta}_j\right)=\frac{\sigma^2}{STC_j(1-R_j^2)} \quad j=1,2,\cdots,k$$

$STC_j=\sum_{i=1}^{N}\left(x_{ji}-\bar{x}_j\right)^2$, proporcional a la varianza muestral de x_j

$R_j^2=$ R-cuadrado de la regresión de x_j sobre el resto de explicativas y un término constante

UNIVERSIDAD DE MURCIA

$$V\left(\hat{\beta}_j\right) = \frac{\sigma^2}{STC_j(1-R_j^2)} \quad j=1,2,\cdots,k$$

Tres componentes:

- <u>varianza del error σ^2</u>: a mayor σ^2, mayor $V(\hat{\beta}_j)$

  σ^2 grande, muestra más dispersa $\Rightarrow$ estimación menos precisa de las pendientes

- <u>variación muestral de x_j</u>: a mayor STC_j, menor $V(\hat{\beta}_j)$

  x_j varía mucho $\Rightarrow$ más fácil capturar con precisión el efecto ceteris paribus de sus variaciones sobre y

- <u>relación de x_j con las demás explicativas</u>: a menor R_j^2, menor $V(\hat{\beta}_j)$

  x_j muy relacionado con el resto de explicativas $\Rightarrow$ más difícil capturar con precisión el efecto ceteris paribus de sus variaciones sobre y

- **Multicolinealidad fuerte** (o multicolinealidad)

  existe una correlación alta (no exacta) entre dos o más variables explicativas

- No viola ninguno de los supuestos RLM.1 a RLM.5 ☺

- Pero genera valores altos de varios R_j^2 => varias $V(\hat{\beta}_j)$ (muy) altas

fuente de **(mucha) imprecisión en las estimaciones** ☹

- ¿cómo detectarla? ¿soluciones posibles?

UNIVERSIDAD DE MURCIA

- **Para detectarla:**
  - Si es entre 2 variables explicativas
    - Coeficiente de correlación lineal
    - Diagrama de dispersión
  - Si es entre más de dos variables explicativas
    - los R_j^2 j=1,2,…k
- **Para solucionarla:**
  - Recoger más datos (aumentar N)
  - Si no es posible, reducirla eliminando alguna variable menos relevante (problema: ¿cuál es "menos relevante"? ¿reducimos multicolinealidad y dispersión a costa de introducir sesgo? (Ver tema4)

$$V(\hat{\beta}_j) = \frac{\sigma^2}{\boxed{STC_j(1-R_j^2)}}$$

→ desconocido

→ calculable

- Necesitamos un estimador insesgado de $\sigma^2 = E(\varepsilon_i^2)$

- Bajo RLM.1 - RLM.5 (demostración: Wooldridge p.62 para modelo de regresión simple)

$$s^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-K} \quad \text{con } E\left(s^2\right) = \sigma^2$$

➡ estimador insesgado de $V(\hat{\beta}_j)$: $\boxed{\hat{V}(\hat{\beta}_j) = \frac{s^2}{STC_j(1-R_j^2)}}$

UNIVERSIDAD DE MURCIA

- Inclusión de irrelevantes:

modelo verdadero: $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$, cumple RLM.1-RLM.5

modelo planteado: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad \Rightarrow \beta_2 = 0$

modelo estimado: $\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{1i} + \tilde{\beta}_2 x_{2i}$

$$V(\hat{\beta}_1) = \frac{\sigma^2}{STC_1} \quad \text{en mod. verdadero}$$

$$\text{pero } V(\tilde{\beta}_1) = \frac{\sigma^2}{STC_1(1-R_1^2)} \quad \text{en mod. estimado}$$

➡ las varianzas aumentan

➡ se pierde precisión en la estimación

- Omisión de relevantes:

modelo verdadero: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \; , \beta_2 \neq 0,$ cumple RLM.1-RLM.5

modelo planteado: $y_i = \beta_0 + \beta_1 x_{1i} + \nu_i \quad \Rightarrow \; \nu_i = \beta_2 x_{2i} + \varepsilon_i$ pero se ignora

modelo estimado: $\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{1i}$

$$\hat{V}(\hat{\beta}_1) = \frac{s^2}{STC_1(1 - R_1^2)}$$

pero $\hat{V}(\tilde{\beta}_1) = \dfrac{s_\nu^2}{STC_1}$

➡ las varianzas estimadas podrían reducirse

➡ mayor precisión aparente en la estimación

UNIVERSIDAD DE MURCIA

- Hasta ahora, hemos visto:
  - RLM.1-RLM.4: necesarios para la insesgadez de MCO
  - RLM.1-RLM.5: necesarios para la fórmula de $V(\hat{\beta}_j)$

- RLM.1-RLM.5 = “supuestos de Gauss-Markov”

- Garantizan la **optimalidad** del estimador MCO.

  Recordemos: “estimador óptimo” $\equiv$ estimador de mínima varianza dentro de una clase concreta

- **<u>Teorema de Gauss-Markov</u>**:

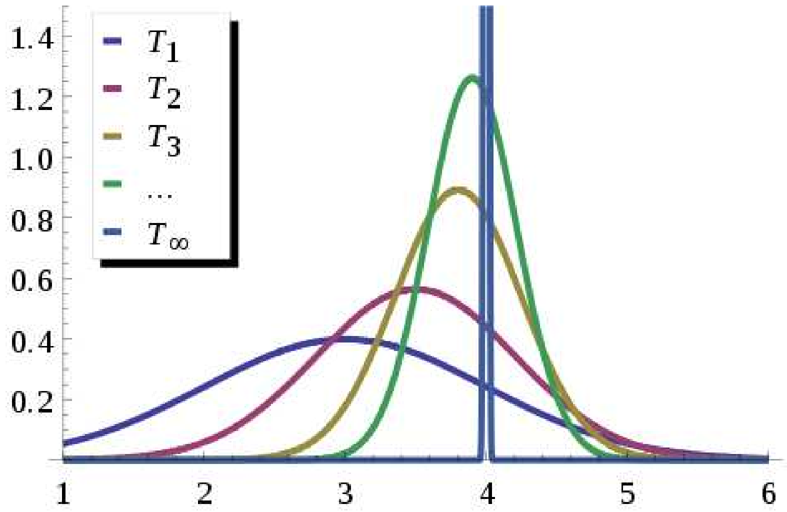
  Bajo los supuestos RLM.1-RLM.5, $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$ son los **estimadores lineales insesgados óptimos** (**ELIO**) de $\beta_0, \beta_1, \cdots, \beta_k$, respectivamente.

- Interpretación:

  - RLM.1-RLM.5 garantizan que el MCO es el estimador más preciso de entre todos los estimadores lineales e insesgados.
  - Si falla un supuesto de RLM.1 a RLM.4 (p.ej. RLM.3 que suele fallar), deja de ser insesgado.
  - Si falla RLM.5, existe otro estimador más preciso (ver Tema 6)

- No siempre se cumplen RLM.1-RLM.4 que aseguran insesgadez del MCO (*propiedad de muestra finita*)

- Pero un estimador debe ser por lo menos **consistente:** a mayor tamaño muestral, menos probabilidad de que el estimador se aleje del valor poblacional (*propiedad asintótica*)



Tamaño muestral : $T_1 < T_2 < \ldots < T_\infty$

Valor poblacional: 4

UNIVERSIDAD DE MURCIA

- Bajo RLM.1-RLM.4, MCO es insesgado y **consistente**:

$$E(\hat{\beta}_j) = \beta_j \quad \text{y también} \quad \underset{N\to\infty}{p\lim}(\hat{\beta}_j) = \beta_j$$

(dem. Consistencia Wooldridge p. 183)

- Puede fallar la exogeneidad estricta. Entonces, MCO está sesgado.
- Pero con supuestos más débiles, garantizamos por lo menos la consistencia.

- **RLM.3'**:

$$\begin{cases} E(\varepsilon_i \mid x_{1i'}, x_{2i}, \cdots, x_{ki}) = E(\varepsilon_i) & \forall i = 1, \cdots, i, \cdots, N \qquad \text{(RLM.3a')} \\ E(\varepsilon_i) = 0 & i = 1, \cdots, N \qquad \text{(RLM.3b)} \end{cases}$$

RLM.3a' : exogeneidad débil

- **<u>Teorema de consistencia del estimador MCO:</u>**

  Bajo los supuestos RLM.1, RLM.3' y RLM.4,

$$\underset{N\to\infty}{\text{plim}}(\hat{\beta}_j) = \beta_j$$

- No hace falta RLM.2
- La validez práctica de este resultado requiere un tamaño muestral grande

- <u>Nota</u>: con sólo exogeneidad débil, el estimador MCO está sesgado y deja de ser ELIO, y las fórmulas de las varianzas utilizadas ya no sirven.

Lo que hemos aprendido:

UNIVERSIDAD DE **MURCIA**

1. Propiedades de muestra finita del MCO:

- Supuestos mínimos que garantizan la insesgadez del MCO

  Linealidad en los parámeros, muestreo aleatorio, exogeneidad estricta, no multicolinealidad exacta.

- Efectos de la mala especificación sobre la insegadez:
  - *Incluir una irrelevante no sesga la estimación MCO*
  - *Omitir una relevante sesga la estimación, y el sesgo es mayor cuanto más relacionada esté la omitida con las incluidas*

Lo que hemos aprendido:

UNIVERSIDAD DE MURCIA

- Varianza del estimador MCO, bajo homoscedasticidad y los supuestos de insesgadez
- Cómo estimar esta varianza
- Cómo afecta la multicolinealidad fuerte a la precisión de las estimaciones
- Cómo la inclusión de irrelevantes y la omisión de relevantes afectan a la precisión de la estimación
- Los supuestos para que MCO sea óptimo (ELIO)

2. Propiedad asintótica del MCO:

- Unos supuestos menos exigentes, para que MCO sea al menos consistente: *exogeneidad débil*