1. Modelo Lineal Base

Consideren el modelo de regresión usual:

$$\textcircled{1} \quad \underset{1\times 1}{y_i} = \underset{1\times K}{x_i'} \underset{K\times 1}{\beta} + \underset{1\times 1}{u_i} \quad , \quad E[\underset{1\times 1}{u_i} \underset{K\times 1}{x_i}] = 0$$

(no hay correlación entre cada variable dentro del vector x_i y el término de error u_i).

En forma matricial

$$y = X\beta + u,$$

donde $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}$, $X = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ \underset{1\times K}{x_n'} \end{bmatrix}_{n\times K}$.

(*) Noten que $E[u_i|x_i] = 0$ es un supuesto MUCHO más fuerte porque significa $E[g(x_i)u_i] = 0$. Es decir, u_i no se correlaciona con ninguna función de x_i.

1.1. Identificación

Suponemos

(i) $E x_i u_i = 0$

(ii) $E x_i x_i'$ tiene rango completo. Es decir, $\text{rank}(E x_i x_i') = K$.

Entonces partimos de (i):

$$E x_i u_i = 0$$

$$\Rightarrow E x_i (y_i - x_i'\beta) = 0$$

$$\Rightarrow \underbrace{E x_i y_i - E x_i x_i' \beta = 0}$$

Si $E x_i x_i'$ es invertible (lo es, dado por (ii)), entonces

$$\beta = (E x_i x_i')^{-1} E x_i y_i \quad \textcircled{2}$$

Además, podemos estimar reemplazando esperanzas por promedios muestrales. A esto se le llama sample analogue. Tenemos:

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i y_i$$

$$\underset{\text{reemplazo } \textcircled{1}}{=} \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i u_i \quad \textcircled{3}$$

1.2 Insesgadez de OLS

Partimos de (3)

$$\hat{\beta} = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i u_i$$

Tomando esperanza condicionada

$$E[\hat{\beta} | X] = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i \underbrace{E[u_i | X]}$$

Solo es 0 si $E[u_i | X] = 0$.
Si solo asumimos $E[u_i x_i] = 0$
NO es insesgado.

Importantemente, OLS es sesgado en muestras finitas, excepto cuando ponemos un supuesto de mean independence $E[u_i | X_i] = 0$. Es decir, cuando $E[Y_i | X_i] = x_i' \beta$, que significa que estamos especificando correctamente el modelo.

2. Teoría Asintótica

- **Teorema 1**: (Continuous Mapping) Sea $\{X_n\} \in \mathbb{R}^d$ tal que $X_n \xrightarrow[d]{} X$. Además, sea $g(\cdot) : \mathbb{R}^d \to \mathbb{R}^k$ que es continuo en un set G tal que $P\{X \in G\} = 1$. Entonces

$$g(X_n) \xrightarrow[d]{} g(X) \quad \text{cuando } n \to \infty.$$

(*) Ejemplo, $g\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = a + b \Rightarrow X_{1n} + X_{2n} \xrightarrow[d]{} X_1 + X_2$

- **Teorema 2**: (Slutsky) Suponemos que hay $\{X_n\}$ tal que $X_n \xrightarrow[p]{} X$ y que $g(\cdot)$ nuevamente es una función continua en un set tal que $P\{X \in G\} = 1$. Entonces,

$$h(X_n) \xrightarrow[p]{} h(X).$$

- **Teorema 3**: (Ley Débil Grandes Números WLLN, iid) Sea $\{X_n\}$ una secuencia de iid vectores aleatorios tal que $E\|X_i\| < \infty$. Entonces

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[p]{} EX_i$$

- **Teorema 4**: (Límite Central CLT, iid) Sea $\{X_n\}$ una secuencia de iid vectores aleatorios tal que $E\|X_i\|^2 < \infty$ y $Var(X_i) = \Sigma$, donde Σ es no singular. Entonces,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - EX_i) \xrightarrow[d]{} N(0, \Sigma)$$

(*) $Var(X_i) = E[(X_i - EX_i)(X_i - EX_i)'] = E[X_i X_i' - (EX_i) X_i']$
$\leq E\|X_i\|^2 + [E\|X_i\|]^2 = O(1)$ si $E\|X_i\|^2 = O(1)$.

3. Propiedades del Estimador $\hat{\beta}$

3.1. Consistencia

Def.- Decimos que $\hat{\beta}$ es consistente si $\hat{\beta} \xrightarrow[p]{} \beta$.

Para mostrar esta propiedad damos los siguientes supuestos.

(i) Data $\{Y_i, X_i'\}$ es iid

(ii) $E \underset{k\times 1}{X_i} \underset{1\times 1}{u_i} = 0$

(iii) $E X_i X_i'$ tiene rango completo k

Si (i)-(iii) se cumplen, entonces

$$\hat{\beta}_n = \beta + o_p(1)$$

subíndice n
para ser explícitos que
cambia con n.

proof.-

$$\hat{\beta}_n = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} X_i u_i$$

Por el Teorema 3 (WLLN) tenemos $\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \xrightarrow[p]{} E X_i X_i'$ y $\frac{1}{n} \sum_{i=1}^{n} X_i u_i = E X_i u_i = 0$.

Además $g(z) = z^{-1}$ es continua siempre que $z \neq 0$. Entonces, por el Teorema 2 (Slutsky)

$$\left[\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right]^{-1} \xrightarrow[p]{} [E X_i X_i']^{-1}.$$

Entonces, escribimos

$$\hat{\beta}_n = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} X_i u_i$$

$$\underset{\text{WLLN}}{=} \beta + (E X_i X_i' + o_p(1))^{-1} (\underbrace{E X_i u_i}_{=0} + o_p(1))$$

$$\underset{\text{Slutsky}}{=} \beta + ((E X_i X_i')^{-1} + o_p(1)) (0 + o_p(1))$$

$$= \beta + (O(1) + o_p(1)) \, o_p(1)$$

$E X_i X_i' > 0$
$\Rightarrow (E X_i X_i')^{-1} = O(1)$

$$= \beta + o_p(1). \quad \blacksquare$$

3.2. Normalidad Asintótica

Para mostrar ello, primero daremos explícitamente los supuestos

- (i) Data $\{y_i, x_i'\}$ es iid
- (ii) $E x_i u_i = 0$
- (iii) $E x_i x_i'$ tiene rango completo k
- (iv) $E x_{ij}^4 < \infty$ para $j = 1, \dots, k$.
- (v) $E u_i^4 < \infty$
- (vi) $Var(u_i x_i)$ es positiva definida

Lema: Si (ii), (iv) y (v) se cumplen, entonces

$$Var(u_i x_i) = O(1).$$

proof:

$$\underset{k\times k}{Var(x_i u_i)} = E[\underset{k\times 1}{(u_i x_i - E u_i x_i)}\ \underset{1\times k}{(u_i x_i - E u_i x_i)'}]$$

$$\overset{\text{por (ii)}}{=} E[\underset{1\times 1}{u_i}\ x_i\ \underset{1\times 1}{u_i}\ x_i']$$

$$= E[u_i^2\ x_i x_i']$$

$$\overset{\text{Cauchy Schwarz}}{\leq} \{E|u_i|^4\ E\|x_i x_i'\|^2\}^{1/2}$$

$$\overset{\text{Frobenius norm}}{=} \{E|u_i|^4\ E[tr(x_i x_i'\ x_i x_i')]\}^{1/2}$$

$$\overset{\text{Propiedad de } tr(\cdot)}{=} \{E|u_i|^4\ E[tr(\underbrace{x_i'x_i}_{1\times 1}\ \underbrace{x_i'x_i}_{1\times 1})]\}^{1/2}$$

$$= \{E|u_i|^4\ E[x_i'x_i]^2\}^{1/2}$$

$$\overset{\text{Definición de norma } \|x_i\|^2 = x_i'x_i}{=} \{E|u_i|^4\ E\|x_i\|^4\}^{1/2}$$

$$= \{E|u_i|^4\ E[\sum_{j=1}^{k} x_{ij}^2]^2\}^{1/2}$$

$$\leq \{E|u_i|^4 .\ E[k \max_{1\leq j\leq k} x_{ij}^2]^2$$

$$\leq \{E|u_i|^4\ k^2 . \max_{1\leq j\leq k} E[x_{ij}^4]\}^{1/2}$$

$$\overset{\text{por (iv) (v)}}{=} O(1) .\ O(1)$$

$$= O(1). \quad \blacksquare$$

Ahora, si se cumplen (i) - (vi), entonces

$$\sqrt{n}\,(\hat{\beta}_n - \beta) \xrightarrow[d]{} \mathcal{N}\Big(0, \underbrace{(E\,X_iX_i')^{-1}\,Var(X_iu_i)\,(E\,X_iX_i')^{-1}}_{V := \text{ varianza asintótica}}\Big)$$

⊛ Noten que esto implica $\hat{\beta}_n = \beta + \underline{O_p(n^{-1/2})}$.

↳ esto es el ruido estadístico.

proof:

Empecemos notando lo siguiente dado por (ii)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_iu_i - \underbrace{E\,X_iu_i}_{\text{porque } = 0})$$

Además, por el Lema anterior y por (i), (vi) cumple con los requisitos del CLT. Entonces,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i \xrightarrow[d]{} \mathcal{N}(0, Var(X_iu_i))$$

⊛ Es decir, $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i = O_p(1)$.

Ahora, procedemos a escribir

$$\sqrt{n}\,(\hat{\beta}_n - \beta) = \left(\frac{1}{n}\sum_{i=1}^n X_iX_i'\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i$$

$$\overset{\text{WLLN + slutsky}}{=} \left((E\,X_iX_i')^{-1} + o_p(1)\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i$$

$$= (E\,X_iX_i')^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i + o_p(1)\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i}$$

$$= (E\,X_iX_i')^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i + o_p(1)\,O_p(1)$$

$$= (E\,X_iX_i')^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n X_iu_i + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \underbrace{(E\,X_iX_i')^{-1}X_iu_i}_{\xi_i} + o_p(1)$$

El CLT requiere que $Var(\xi_i) = E[\xi_i \xi_i']$ sea finita y definida positiva.
En primer lugar es finita porque

$$\begin{aligned} Var(\xi_i) &= E(x_i x_i')^{-1}\, Var(x_i u_i)\, E(x_i x_i')^{-1} \\ &= O(1)\,.\,O(1)\,.\,O(1) \\ &= O(1), \end{aligned}$$

dado por $E x_i x_i'$ de rango completo.

Además, $E x_i x_i' = O(1)$ porque $E\|x_i x_i'\| \leq E\|x_i\|^2 = O(1)$. Por lo tanto su inversa $(E x_i x_i')^{-1}$ no puede ser singular. Esto, junto con (vi) aseguran que $Var(\xi_i)$ cumpla con ser positiva definida.

Entonces,

$$\sqrt{n}(\hat{\beta}_n - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underbrace{(E x_i x_i')^{-1} x_i u_i}_{\xi_i} + o_p(1)$$

$$\xrightarrow{d} N\Big(0, \underbrace{E(x_i x_i')^{-1}\, Var(x_i u_i)\, E(x_i x_i')^{-1}}_{V \,:=\, \text{varianza asintótica.}}\Big). \quad \blacksquare$$

3.3. Estimación de Varianza Asintótica

Denotamos

$$\hat{M}_n = \frac{1}{n}\sum_{i=1}^{n} x_i x_i', \quad \hat{\Omega}_n = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 x_i x_i', \quad \hat{u}_i = y_i - x_i'\hat{\beta}_n.$$

Proponemos estimar V usando

$$\hat{V}_n = \hat{M}_n^{-1}\, \hat{\Omega}_n\, \hat{M}_n^{-1}.$$

Para probar su consistencia, observemos lo siguiente

$$\begin{aligned} \Omega_n &= \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 x_i x_i' \\ &= \frac{1}{n}\sum_{i=1}^{n} [(y_i - x_i'\hat{\beta}_n \pm x_i'\beta)^2 x_i x_i'] \\ &= \frac{1}{n}\sum_{i=1}^{n} [(u_i - x_i'(\hat{\beta}_n - \beta))^2 x_i x_i'] \end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n} u_i^2 x_i x_i' + \underbrace{\frac{1}{n}\sum_{i=1}^{n} [x_i'(\hat{\beta}_n - \beta)]^2 x_i x_i'} - \underbrace{\frac{2}{n}\sum_{i=1}^{n} [x_i'(\hat{\beta}_n - \beta)u_i] x_i x_i'}$$

$$= \Omega + o_p(1) + R_{1n} + R_{2n}$$

Si mostramos que R_{1n} y R_{2n} son $o_p(1)$ entonces nuestro estimador $\hat{\Omega}_n$ es consistente.

Empezamos por R_{1n} :

- $$\begin{aligned}
\| R_{1n} \| &\leq \frac{1}{n}\sum_{i=1}^{n} | x_i'(\hat{\beta}_n - \beta)|^2 \underbrace{\| x_i x_i' \|} \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \| x_i \|^2 \| \hat{\beta}_n - \beta \|^2 \| x_i \|^2 \\
&\leq \underbrace{\| \hat{\beta}_n - \beta \|^2} \underbrace{\frac{1}{n}\sum_{i=1}^{n} \| x_i \|^4} \\
&= \| o_p(1) \|^2 \left(E \| x_i \|^4 + o_p(1) \right) \\
\text{Slutsky} &= o_p(1) \left(E \| x_i \|^4 + o_p(1) \right) \\
\text{por (iv)} &= o_p(1) \left(O(1) + o_p(1) \right) \\
&= o_p(1) .
\end{aligned}$$

- $$\begin{aligned}
\| R_{2n} \| &\leq \frac{1}{n}\sum_{i=1}^{n} \| x_i'(\hat{\beta}_n - \beta) u_i \; x_i x_i' \| \\
&\leq \| \hat{\beta}_n - \beta \| \frac{1}{n}\sum_{i=1}^{n} |u_i| \| x_i \| . \| x_i \|^2 \\
&= o_p(1) \left(E |u_i| \| x_i \| \| x_i \|^2 + o_p(1) \right) \\
\text{Cauchy Schwarz} &\leq o_p(1) \left(\left(E |u_i|^2 \| x_i \|^2 \right)^{1/2} \left(E \| x_i \|^4 \right)^{1/2} + o_p(1) \right) \\
\text{Cauchy Schwarz} &\leq o_p(1) \left[\left(E |u_i|^4 \; E \| x_i \|^4 \right)^{1/4} \left(E \| x_i \|^4 \right)^{1/2} + o_p(1) \right] \\
\text{por (iv),(v)} &= o_p(1) \left[O(1) . O(1) . O(1) + o_p(1) \right] \\
&= o_p(1) .
\end{aligned}$$

Con esto hemos mostrado

$$\hat{\Omega}_n = \Omega + o_p(1) .$$

Finalmente,

$$\hat{V}_n = M_n^{-1} \hat{\Omega}_n \hat{M}_n^{-1}$$

$$= \left[(E X_i X_i')^{-1} + o_p(1) \right] \left[\Omega + o_p(1) \right] \left[(E X_i X_i')^{-1} + o_p(1) \right]$$

Por $(E X_i X_i')^{-1}$ y Ω siendo $O(1)$

$$= (E X_i X_i')^{-1} \Omega (E X_i X_i')^{-1} + o_p(1).$$

$$= V + o_p(1). \quad \blacksquare$$

4. Modelo Lineal Particionado

Consideremos el modelo

$$y_i = \beta_1 \underbrace{X_{1i}}_{1\times 1} + X_{2i}' \underbrace{\beta_2}_{k_2 \times 1} + u_i \quad , \text{ donde } \begin{array}{l} X_{1i} \to \text{var. de interés (escalar) y} \\ X_{2i} \to \text{vars de control (vector).} \end{array}$$

Además, tenemos que las variables son todas exógenas:

$$\underbrace{E\, X_i u_i}_{k\times 1} = \begin{pmatrix} E\, X_{1i} u_{i \; 1\times 1} \\ E\, X_{2i} u_{i \; k_2\times 1} \end{pmatrix} = 0_{k\times 1} \quad , \quad \text{donde } k = 1 + k_2 .$$

Noten que mientras $E\, X_i X_i'$ sea de rango completo k (no multicolinearidad perfecta) la identificación de $\beta = (\beta_1 \; \beta_2')'$ no ha cambiado.

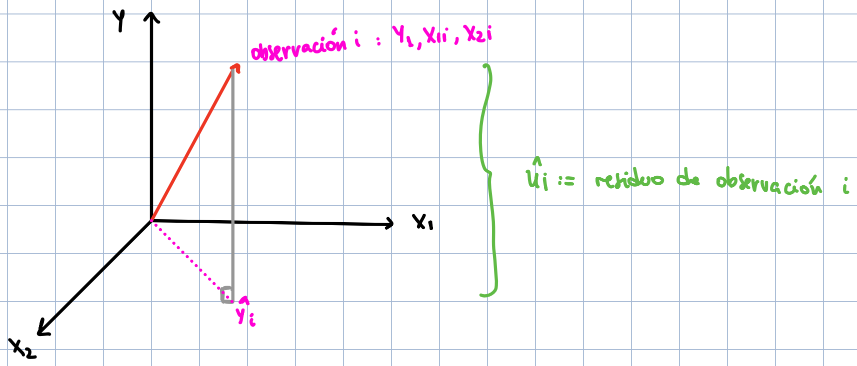
4.1. Proyecciones

Sea $X_{n\times k}$ una matriz de rango completo k, podemos definir la proyección

$$\underbrace{P_X}_{n\times n} = \underbrace{X}_{n\times k} \underbrace{(X'X)^{-1}}_{k\times k} \underbrace{X'}_{k\times n}$$

Lo que hace esta matriz es crear una predicción de una variable $Y_{n\times 1}$ a través de OLS:

$$\hat{Y} = P_X Y = X \underbrace{(X'X)^{-1} X'Y}_{\hat{\beta}_n}$$



Tenemos que P_X proyecta vectores de tamaño $n \times 1$ a un subespacio

$$S(X) = \{ y \in \mathbb{R}^n : \quad y = X b \, , \, b \in \mathbb{R}^k \}.$$

Además, mostraremos que es una proyección ortogonal.

Considere la diferencia

$$Y - P_X Y = \underbrace{(I_n - P_X)} Y$$
$$= M_X Y .$$

Noten que

$$M_X Y = Y - \hat{Y} = Y - X\hat{\beta} = \hat{U} \, . \longrightarrow \text{genera residuos de OLS.}$$

Por ello, decimos que es una proyección ortogonal

$$Y = P_X Y + M_X Y = \hat{Y} + \hat{U} \, ,$$

donde

$$\hat{Y}'\hat{U} = Y' X(X'X) \underbrace{X'\hat{U}}_{=0} = 0 \, .$$

(*) Es decir, OLS descompone Y en dos vectores ortogonales entre sí.

<u>Proposición</u>.-

(a) P_X y M_X son simétricas : $P_X = P_X'$, $M_X = M_X'$.

(b) $P_X X = X$ $\rightarrow$ X ya pertenece a ese subespacio

(c) $M_X X = 0$ $\rightarrow$ Consecuencia de lo anterior

(d) P_X y M_X son ortogonales : $M_X P_X = 0$ y $P_X M_X = 0$

(e) P_X y M_X son idempotentes : $P_X P_X = P_X$ y $M_X M_X = M_X$

(f) $\text{rank}(P_X) = k$ y $\text{rank}(M_X) = n - k$.

Proposición.- La suma de residuos al cuadrado SSR no puede decrecer a medida que aumentamos regresores.

Proof:

Consideren $X = (X_1 \quad \underset{\text{regresores extra}}{X_2})$. Tenemos

$$P_X = X(X'X)^{-1}X' \longrightarrow M_X = I_n - P_X,$$
$$P_1 = X_1(X_1'X_1)^{-1}X_1 \longrightarrow M_1 = I_n - P_1.$$

Además, ya que X_1 es parte de X se cumple que $P_X X_1 = X_1$.
Por lo tanto,

- $$\begin{aligned} P_X P_1 &= \underbrace{P_X X_1}(X_1'X_1)^{-1}X_1 \\ &= X_1(X_1'X_1)^{-1}X_1 \\ &= P_1. \end{aligned}$$

- $$\begin{aligned} M_X M_1 &= (I_n - P_X)(I_n - P_1) \\ &= I_n - P_1 - P_X + \underbrace{P_X P_1} \\ &= I_n - P_1 - P_X + P_1 \\ &= I_n - P_X \\ &= M_X. \end{aligned}$$

Definiremos ahora

$$\hat{U}_X = M_X Y,$$
$$\hat{U}_{X_1} = M_1 Y,$$

y escribimos

$$\begin{aligned} 0 &\leq (\hat{U}_X - \hat{U}_{X_1})'(\hat{U}_X - \hat{U}_{X_1}) \longrightarrow \|\hat{U}_X - \hat{U}_{X_1}\|^2 \geq 0 \\ &= \hat{U}_X'\hat{U}_X + \hat{U}_{X_1}'\hat{U}_{X_1} - 2\underbrace{\hat{U}_X'\hat{U}_{X_1}} \\ &= \hat{U}_X'\hat{U}_X + \hat{U}_{X_1}'\hat{U}_{X_1} - 2Y'\underbrace{M_X M_1}Y \\ &= \hat{U}_X'\hat{U}_X + \hat{U}_{X_1}'\hat{U}_{X_1} - 2Y'M_X Y \\ &= \hat{U}_X'\hat{U}_X + \hat{U}_{X_1}'\hat{U}_{X_1} - 2\hat{U}_X'\hat{U}_X \\ &= \hat{U}_{X_1}'\hat{U}_{X_1} - \hat{U}_X'\hat{U}_X \end{aligned}$$

Por lo tanto,

$$\underbrace{\hat{U}_{X_1}'\hat{U}_{X_1}}_{\substack{\text{Suma de} \\ \text{residuos} \\ \text{con } X_1}} \geq \underbrace{\hat{U}_X'\hat{U}_X}_{\substack{\text{Suma de} \\ \text{residuos} \\ \text{con } X_1, X_2}}. \quad \blacksquare$$

4.2. Regresión Particionada (Cont.)

Sea $X = (X_1 \quad X_2)$, tenemos:

$$X'X = \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}_{K \times K}$$

Además, $\hat{\beta}_n$ satisface la condición de primer orden

$$X'X \; \hat{\beta}_n = X'Y$$

$$\Rightarrow \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}_{K \times K} \begin{pmatrix} \hat{\beta}_{1n} \\ \hat{\beta}_{2n} \end{pmatrix} = \begin{pmatrix} X_1' \\ X_2' \end{pmatrix}_{K \times n} Y$$

$$\Rightarrow \begin{array}{ll} X_1'X_1 \; \hat{\beta}_{1n} + X_1'X_2 \; \hat{\beta}_{2n} = X_1'Y & (1) \\ X_2'X_1 \; \hat{\beta}_{1n} + X_2'X_2 \; \hat{\beta}_{2n} = X_2'Y & (2) \end{array}$$

Si $X_2'X_2$ es invertible, podemos despejar de (2):

$$\hat{\beta}_{2n} = (X_2'X_2)^{-1} X_2'Y - (X_2'X_2)^{-1} X_2'X_1 \hat{\beta}_{1n} \quad (3)$$

Reemplazando (3) en (1), nuevamente suponiendo $X_1'X_1$ es invertible:

$$X_1'X_1 \; \hat{\beta}_{1n} + X_1'X_2 \left[(X_2'X_2)^{-1} X_2'Y - (X_2'X_2)^{-1} X_2'X_1 \hat{\beta}_{1n} \right] = X_1'Y$$

$$\Rightarrow \left[X_1'X_1 - X_1'X_2 (X_2'X_2)^{-1} X_2'X_1 \right] \hat{\beta}_{1n} = X_1'Y - X_1'X_2 (X_2'X_2)^{-1} X_2'Y$$

$$\Rightarrow X_1' \left(I_n - X_2 (X_2'X_2)^{-1} X_2' \right) X_1 \; \hat{\beta}_{1n} = X_1' \left(I_n - X_2 (X_2'X_2)^{-1} X_2' \right) Y$$

$$\Rightarrow X_1' M_2 X_1 \hat{\beta}_{1n} = X_1' M_2 Y$$

$$\Rightarrow \hat{\beta}_{1n} = (X_1' M_2 X_1)^{-1} X_1' M_2 Y.$$

Además, noten que $\hat{\beta}_{1n} = (\tilde{X}_1' \tilde{X}_1)^{-1} \tilde{X}_1' \tilde{Y}$, donde $\tilde{X}_1 = M_2 X_1$ e $\tilde{Y} = M_2 Y$.

(*) Es decir, equivale a hacer una regresión de X_1 sobre X_2 y de Y sobre X_2 para luego hacer una regresión entre ambos residuos. Este resultado se conoce como el Teorema de Frisch-Waugh-Lovell.

(*) Podemos pensar en $\tilde{X}_{1i} = X_{1i} - X_{2i}'\hat{\gamma}$, donde $\hat{\gamma} = (X_2'X_2)^{-1} X_2'X_1$.

4.2.1. Consistencia

Definimos $\Sigma_{A,B} := E\, A_i B_i'$, donde A_i, B_i son vectores aleatorios.
Tenemos

$$\hat{\beta}_{1n} = (X_1' M_2 X_1)^{-1} X_1' M_2 Y$$

$$= \beta_1 + (X_1' M_2 X_1)^{-1} X_1' M_2 U \; ,$$

donde

- $$\frac{X_1' M_2 U}{n} = X_1' (I_n - X_2 (X_2' X_2)^{-1} X_2') U / n$$
$$= X_1' U / n - \underbrace{X_1' X_2}_{n} \underbrace{(X_2' X_2)}_{n}^{-1} X_2' U / n$$
$$= \sum_{i=1}^{n} x_i u_i / n - \sum_{i=1} \frac{x_{1i} x_{2i}'}{n} \left(\sum_{i=1}^{n} \frac{x_{2i} x_{2i}'}{n} \right)^{-1} \sum_{i=1}^{n} \frac{x_{2i} u_i}{n}$$
$$\overset{\text{WLLN}}{=} \underbrace{\Sigma_{x_1,u}}_{=0} - \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \underbrace{\Sigma_{x_2,u}}_{=0} + o_p(1)$$
$$= o_p(1)$$

- $$\frac{X_1' M_2 X_1}{n} = \frac{X_1' X_1}{n} - \frac{X_1' P_2 X_1}{n}$$
$$= \underbrace{\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \Sigma_{x_2,x_1}}_{\neq 0 \text{ (no singular)}} + o_p(1)$$

Solo se cancela si $x_1 = x_2$, lo cual no tendría sentido. Como no es singular, su inversa es $O(1)$

Entonces,

$$\hat{\beta}_{1n} = \beta_1 + (X_1' M_2 X_1)^{-1} X_1' M_2 U$$
$$= \beta_1 + \left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \Sigma_{x_2,x_1} + o_p(1) \right)^{-1} o_p(1)$$
$$= \beta_1 + (O(1) + o_p(1))\, o_p(1)$$
$$= \beta_1 + o_p(1) \, .$$

¿Qué pasa si el vector de controles X_{2i} fuera endógeno? A esto le llamamos Bad Controls.

Supongamos que ahora X_{2i} es endógeno : $E\, X_{2i} u_i \neq 0$.

$$\hat{\beta}_{1n} = (X_1' M_2 X_1)^{-1} X_1' M_2 Y$$

$$= \beta_1 + (X_1' M_2 X_1)^{-1} X_1' M_2 u,$$

donde

- $$\frac{X_1' M_2 u}{n} = X_1' (I_n - X_2 (X_2' X_2)^{-1} X_2') u /n$$
$$= X_1' u/n - \frac{X_1' X_2}{n} \left(\frac{X_2' X_2}{n}\right)^{-1} X_2' u /n$$
$$= \sum_{i=1}^{n} x_i u_i /n - \sum_{i=1} \frac{x_{1i} x_{2i}'}{n} \left(\sum_{i=1}^{n} \frac{x_{2i} x_{2i}'}{n} \right)^{-1} \sum_{i=1}^{n} \frac{x_{2i} u_i}{n}$$
$$\overset{WLLN}{=} \underbrace{\Sigma_{x_1,u}}_{=0} - \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \underbrace{\Sigma_{x_2,u}}_{\text{Ahora } \neq 0 \text{ (endógeno)}} + o_p(1)$$
$$= o_p(1)$$

- $$\frac{X_1' M_2 X_1}{n} = \frac{X_1' X_1}{n} - \frac{X_1' P_2 X_1}{n}$$
$$= \underbrace{\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \Sigma_{x_2,x_1}}_{\neq 0 \text{ (no singular)}} + o_p(1)$$

Solo se cancela si $x_1 = x_2$, lo cual no tendría sentido. Como no es singular, su inversa es $O(1)$

Entonces,

$$\hat{\beta}_{1n} = \beta_1 + \left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \Sigma_{x_2,x_1} + o_p(1) \right)^{-1} \left(- \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \Sigma_{x_2,u} + o_p(1) \right)$$

$$= \beta_1 - \left[\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \Sigma_{x_2,x_1} \right]^{-1} \Sigma_{x_1,x_2} \Sigma_{x_2,x_2}^{-1} \Sigma_{x_2,u} + o_p(1).$$

$$\not\xrightarrow{p} \beta_1,$$

a menos que $\Sigma_{x_1,x_2} := E\, x_{1i} x_{2i}' = 0$, que significa que X_{1i} y X_{2i} no están correlacionados.

4.2.2 Normalidad Asintótica

$$\hat{\beta}_{1n} = \beta_1 + \left(\frac{X_1' M_2 X_1}{n}\right)^{-1} \frac{X_1' M_2 u}{n}$$

$$\Rightarrow \sqrt{n}\,(\hat{\beta}_{1n} - \beta_1) = \left(\frac{X_1' M_2 X_1}{n}\right)^{-1} \frac{X_1' M_2 u}{\sqrt{n}}$$

$$= \left[\left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1} + o_p(1)\right]\left[\frac{X_1'u}{\sqrt{n}} - \frac{X_1'X_2}{n}\left(\frac{X_2'X_2}{n}\right)^{-1}\frac{X_2'u}{\sqrt{n}}\right]$$

$$= \underbrace{\left[\left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1}_{K_1\times K_1} \;\vdots\; -\left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1}\Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}{}_{K_1\times K_2}\right]}_{Q\,:\; K_1\times(K_1+K_2) \,=\, K_1\times K}\begin{bmatrix} \frac{1}{\sqrt{n}}\sum^n x_{1i}u_i \\ \frac{1}{\sqrt{n}}\sum^n x_{2i}u_i\end{bmatrix} + o_p(1)$$

Noten que nada nos impide seguir usando el CLT

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum^n x_{1i}u_i \\ \frac{1}{\sqrt{n}}\sum^n x_{2i}u_i\end{pmatrix} \xrightarrow[d]{} N\left(0, \underbrace{\begin{pmatrix} E\,u_i^2 x_{1i}x_{1i}'{}_{K_1\times K_1} & E\,u_i^2 x_{1i}x_{2i}'{}_{K_1\times K_2} \\ E\,u_i^2 x_{2i}x_{1i}'{}_{K_2\times K_1} & E\,u_i^2 x_{2i}x_{2i}'{}_{K_2\times K_2}\end{pmatrix}}_{\Omega\,:\;K\times K}\right),$$

siempre que la data sea iid y todos los bloques de esa varianza sean finitos $O(1)$.

Podemos mostrar que

$$\sqrt{n}\,(\hat{\beta}_{1n} - \beta_1) \xrightarrow[d]{} N(0, Q\,\Omega\,Q'),$$

donde

$$Q\Omega Q = \left[\left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1}_{K_1\times K_1} E\,u_i^2 x_{1i}x_{1i}'{}_{K_1\times K_1} \;\vdots\; -\left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1}\Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}{}_{K_1\times K_2}\, E\,u_i^2 x_{2i}x_{2i}'{}_{K_2\times K_2}\right]Q'$$

$$= \left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1}_{K_1\times K_1} E\,u_i^2 x_{1i}x_{1i}' \left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1}_{K_1\times K_1} +$$

$$\left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1}\Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}{}_{K_1\times K_2}\, E\,u_i^2 x_{2i}x_{2i}'{}_{K_2\times K_2}\; \Sigma_{x_2,x_2}^{-1}{}_{K_2\times K_2}\,\Sigma_{x_1,x_2}'{}_{K_2\times K_1}\left(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}\right)^{-1}_{K_1\times K_1}.$$

En el caso homoscedástico $E[u_i^2 | x_{1i}, x_{2i}] = \sigma^2$. Por tanto

- $E u_i^2 x_{1i} x_{1i}' \underset{\text{L.I.E.}}{=} E[E[u_i^2|x_i] x_{1i} x_{1i}'] = \sigma^2 E x_{1i} x_{1i}'$

- $E u_i^2 x_{2i} x_{2i}' \underset{\text{L.I.E.}}{=} E[E[u_i^2|x_i] x_{2i} x_{2i}'] = \sigma^2 E x_{2i} x_{2i}'$

Por lo tanto, $Q\Omega Q$ se reduce a

$$Q\Omega Q = (\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1})^{-1}_{K_1 \times K_1} \sigma^2 [\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}] (\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1})^{-1}_{K_1 \times K_1}$$

$$= \sigma^2 \underbrace{(\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1})}_{\text{Varianza de } X_1 \text{ una vez que hemos quitado la porción explicada por } X_2.}{}^{-1}_{K_1 \times K_1}$$

(*) A medida que añadimos regresores a X_2 hacemos dos cosas:

$\downarrow \sigma$: varianza de Y no explicada

$\downarrow (\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1})$: varianza residual de X_1 no explicada por X_2

Si aumentamos indiscriminadamente regresores es probable que la segunda fuerza termine dominando y causara errores estándar altos $\widehat{\text{var}}(\beta_1) \uparrow$. Esto causará valores t bajos y p-values altos (no significancia estadística).

5. Inferencia

Por simplicidad consideremos

$$Y_i = \underbrace{x_{1i}}_{\substack{1\times 1 \\ \text{Regresor de interés}}} \beta_1 + \underbrace{x_{2i}'}_{\substack{1\times k_2 \\ \text{Controles}}} \beta_2 + u_i \,, \quad E\, u_i x_{1i} = 0, \quad E\, u_i x_{2i} = 0 \,.$$

Como ya determinamos anteriormente

$$\hat{\beta}_1 = \frac{X_1' M_2 Y}{\underbrace{X_1' M_2 X_1}_{\text{Ahora esto es } 1\times 1}} = \beta_1 + \frac{X_1' M_2 u}{X_1' M_2 X_1}$$

Además, hemos mostrado el siguiente resultado:

$$\underbrace{\sqrt{n}\,(\hat{\beta} - \beta)}_{\sqrt{n}\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix}} \xrightarrow[d]{} \underbrace{N(0, V)}_{N\left(\begin{pmatrix} 0_{1\times 1} \\ 0_{k_2\times 1} \end{pmatrix}, \begin{pmatrix} V_{11} & 0 \\ 0 & V_2 \end{pmatrix}\right)} \,.$$

Es decir,

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow[d]{} N(0, V_{11}) \,. \quad (1)$$

Para simplificar el álgebra nos centraremos en el caso homoscedástico $E[u_i^2 | x_i] = \sigma^2$.
En ese caso:

$$V_{11} = \frac{\sigma^2}{\Sigma_{x_1,x_1} - \Sigma_{x_1,x_2}\Sigma_{x_2,x_2}^{-1}\Sigma_{x_2,x_1}} \,,$$

su estimador

$$\hat{V}_{11} = \frac{\hat{\sigma}^2}{\frac{X_1' M_2 X_1}{n}} \,, \quad \text{donde} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (Y_i - x_i'\hat{\beta})^2 \,.$$

Noten, además que (1) implica $\hat{\beta}_1 \sim N(\beta_1, V_{11}/n)$. $\sqrt{\frac{\hat{V}_{11}}{n}} := SE(\hat{\beta}_1)$ error estándar.

Por lo tanto podemos plantear el siguiente intervalo de confianza al nivel $1-\alpha$:

$$CI_{1-\alpha} = \left[\hat{\beta}_1 - z_{1-\alpha/2}\sqrt{\frac{\hat{V}_{11}}{n}} \,,\ \hat{\beta}_1 + z_{1-\alpha/2}\sqrt{\frac{\hat{V}_{11}}{n}} \right] .$$

Consideren el test

usaré subíndice 0 para referir al valor testeable.

$$H_0 : \beta_1 = \beta_{1,0} ,$$
$$H_1 : \beta_1 \neq \beta_{1,0} ,$$

donde rechazamos H_0 si

$$\hat{\beta}_1 \notin CI_{1-\alpha} .$$

Es decir, la región crítica es el complemento del $CI_{1-\alpha}$. Esto es equivalente a rechazar si

$$\beta_{1,0} < \hat{\beta}_1 - z_{1-\alpha/2} \sqrt{\frac{\hat{V}_{11}}{n}}$$

o

$$\beta_{1,0} > \hat{\beta}_1 - z_{1-\alpha/2} \sqrt{\frac{\hat{V}_{11}}{n}} .$$

Esto puede re-escribirse como algo que probablemente recuerden

$$T_n := \left| \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{V}_{11}/n}} \right| > z_{1-\alpha/2} ,$$

test estadístico

que es lo que conocemos como two sided test, porque si se cumple H_1 el valor del verdadero β_1 puede ser menor o mayor a $\beta_{1,0}$.

5.1. Test de Tamaño α

Decimos que el test es (asintóticamente) de tamaño α si

$$\lim_{n\to\infty} P(\text{Rechazar } H_0 \mid H_0 \text{ es verdadero}) \to \alpha .$$

En este caso tenemos que

$$P(\text{Rechazar } H_0 \mid H_0 \text{ es verdadero}) = P\left(\left| \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{V}_{11}/n}} \right| > z_{1-\alpha/2} \;\middle|\; \beta_1 = \beta_{1,0} \right)$$
$$= P\left(\left| \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{V}_{11}/n}} \right| > z_{1-\alpha/2} \right)$$

Tomando límite tenemos

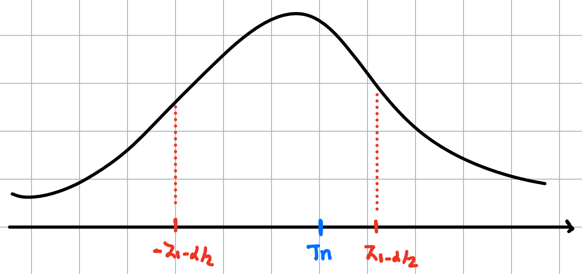
$$\lim_{n\to\infty} P\left(\left| \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{V}_{11}/n}} \right| > z_{1-\alpha/2} \right) = \alpha .$$

5.2 P-values

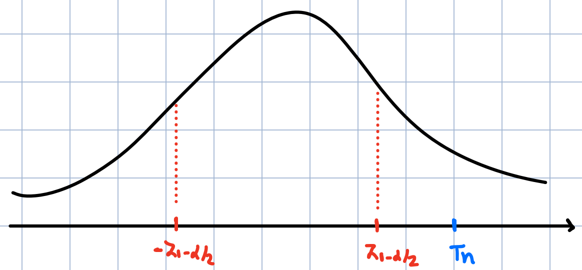
Recuerden que rechazamos si:

$$T_n := \left| \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{V}_{11}/n}} \right| > z_{1-\alpha/2}$$

test estadístico

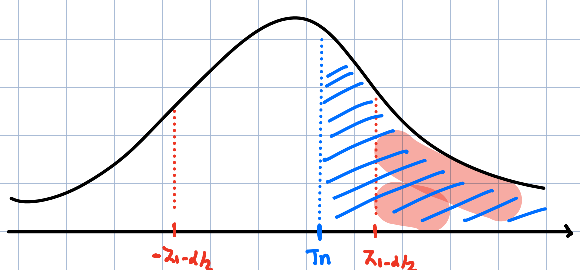


Si T_n cae debajo de $z_{1-\alpha/2}$ NO podemos rechazar.

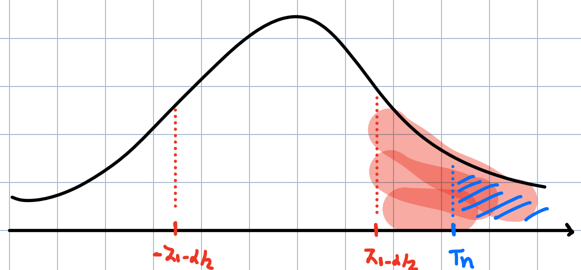


Si T_n cae encima de $z_{1-\alpha/2}$ SÍ podemos rechazar.

Así como comparamos T_n contra $z_{1-\alpha/2}$ es EQUIVALENTE a comparar el área a la derecha de T_n con el área a la derecha de $z_{1-\alpha/2}$.



Si el área por encima de T_n es mayor al área por encima de $z_{1-\alpha/2}$ no podemos rechazar.



Si el área por encima de T_n es menor al área por encima de $z_{1-\alpha/2}$ no podemos rechazar.

De hecho, podemos calcular estos números. Rechazamos si

Area encima de T_n < Area encima de T_n

$$1 - \Phi(T_n) < 1 - \Phi(z_{1-\alpha/2}) = \alpha/2$$

, donde $\Phi(\cdot)$ es el CDF de $N(0,1)$.

Podemos multiplicar ambos lados por 2 y no afectar la desigualdad:

$$\underbrace{2(1 - \Phi(T_n))}_{\text{A este número le llamamos p-value.}} < \alpha .$$

En otras palabras, el p-value es una función del estadístico T_n así que es un estadístico : NO es la probabilidad de que H_0 es cierta.

De aquí viene nuestra regla usual de rechazar con $\alpha = 0.05$ si:

p-val < 0.05 o, equivalentemente, $T_n := \left| \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{V}_{11}/n}} \right| > 1.96$.

5.3. Poder Asintótico del Test

Cuando H_0 es falsa, queremos que nuestro test rechace correctamente. A esto le llamamos poder estadístico.

$$P(\text{rechazar } H_0 \mid \beta_1 \neq \beta_{1,0}) = P\left(\left| \frac{\hat{\beta}_1 - \beta_{1,0} \pm \beta_1}{\sqrt{\hat{V}_{11}/n}} \right| > z_{1-\alpha/2} \right)$$

$$= P\left(\left| \underbrace{\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{V}_{11}/n}}}_{\xrightarrow{d} N(0,1)} + \underbrace{\frac{\beta_1 - \beta_{1,0}}{\sqrt{\hat{V}_{11}/n}}}_{\xrightarrow{p} \pm\infty} \right| > z_{1-\alpha/2} \right)$$

Entonces,

$$\lim_{n\to\infty} P(\text{rechazar } H_0 \mid \beta_1 \neq \beta_{1,0}) = 1 .$$

* El test promete que con una muestra grande siempre rechazaremos si H_0 es falsa.