Teoría Asintótica

Basada en el libro *Econometrics* de Bruce E. Hansen (2019)

(capítulos 2 y 7) disponible online en
`https://www.ssc.wisc.edu/~bhansen/econometrics/`

Bibliografía complementaria: Takeshi Amemiya. *Advanced Econometrics* (1985) Harvard University Press.

Fumio Hayashi. *Econometrics* (2000), Princeton University Press.

Mejor predictor

A partir de observar los valores $\mathbf{x} = (x_0, x_1, \ldots, x_{p-1})$ queremos predecir a y. Escribimos $g(\mathbf{x})$ para cualquier predictor de y. Aquí $\mathbf{x}$ es un vector aleatorio, y es una variable aleatoria. El error de predicción es

$$y - g(\mathbf{x})$$

Una medida no estocástica de la magnitud del error de predicción es

$$\mathbb{E}\left([y - g(\mathbf{x})]^2\right) \tag{1}$$

Definimos al mejor predictor como la función $g(\mathbf{x})$ que minimiza a (1). Vimos que

$$m(\mathbf{x}) = \mathbb{E}\left(y|\mathbf{x}\right),$$

la esperanza condicional, es el mejor predictor.

La esperanza condicional como mejor predictor

Teorema A1

Si $\mathbb{E}\left(y^2\right) < \infty$, *entonces para todo predictor* $g(\mathbf{x})$ *se tiene que*

$$\mathbb{E}\left((y - g(\boldsymbol{x}))^2\right) \geq \mathbb{E}\left((y - m(\boldsymbol{x}))^2\right)$$

donde $m(\boldsymbol{x}) = \mathbb{E}(y|\boldsymbol{x})$.

Vimos

Teorema A2

Asumamos que $\mathbb{E}\left(y^2\right) < \infty$. $y = g(\mathbf{x}) + \varepsilon$ *con* $\mathbb{E}(\varepsilon|\boldsymbol{x}) = 0$ *si y sólo si* $g(\mathbf{x}) = \mathbb{E}(y|\boldsymbol{x})$.

Modelo: Esperanza Condicional Lineal (ECL)

Supuestos S.0 Asumimos que $\mathbb{E}(y|\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}$

Bajo S.0, tenemos que el modelo lineal cumple

$$y = \boldsymbol{x}^T\boldsymbol{\beta} + \varepsilon, \quad \text{donde}$$

- $\mathbb{E}(\varepsilon|\boldsymbol{x}) = 0$, y por lo tanto, $\mathbb{E}(\varepsilon) = 0$.

Si además agregamos el supuesto de que sea homoscedástico, se cumple que

- $\mathbb{E}(\varepsilon^2|\boldsymbol{x}) = \sigma^2$.

Modelo: Esperanza Condicional Lineal (ECL)

Modelo de esperanza condicional lineal homoscedástico

$$
\begin{aligned}
y &= \boldsymbol{x}^T \boldsymbol{\beta} + \varepsilon \\
\mathbb{E}(\varepsilon | \boldsymbol{x}) &= 0 \qquad \text{(exogeneidad)} \\
\mathbb{E}(\varepsilon^2 | \boldsymbol{x}) &= \sigma^2
\end{aligned}
$$

¿Qué pasa cuando $m(\boldsymbol{x})$ no es lineal? ¿O no podemos garantizar que lo sea? En general, el modelo lineal para la esperanza condicional será, a lo sumo aproximado en las aplicaciones. ¿Qué cambia cuando el modelo lineal no vale en forma exacta?

Mejor predictor lineal

Siempre podemos buscar el mejor predictor lineal de $\mathbf{x}$ que aproxime a y, es decir, buscar la función de la forma

$$g_{lin}(\mathbf{x}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_1 + \boldsymbol{\beta}_2 x_2 + \cdots + \boldsymbol{\beta}_{p-1} x_{p-1}$$

que minimice el error cuadrático medio de predicción

$$\mathbb{E}\left([y - g_{lin}(\mathbf{x})]^2 \right) = \mathbb{E}\left(\left[y - \boldsymbol{x}'\beta \right]^2 \right) = S(\boldsymbol{\beta}) \qquad (2)$$

A dicha función g_{lin} la llamaremos el **mejor predictor lineal de y dado $\mathbf{x}$**, o el **predictor óptimo lineal**, y lo notaremos por $\mathcal{P}(y|\boldsymbol{x}) = \boldsymbol{x}'\beta$ donde β minimiza el ECM de predicción, y se llama el **coeficiente lineal óptimo**, o **coeficiente de proyección lineal**.

$$\boldsymbol{\beta}_{OL} = argmin_{\boldsymbol{\beta} \in \mathbb{R}^p} S(\boldsymbol{\beta})$$

Mejor predictor lineal

¿Por qué sólo buscar entre las lineales?

- Vimos en el Teorema N9 que si $\left(y, \mathbf{x}^T\right)^T \sim N_p(\boldsymbol{\mu}, \Sigma)$, con Σ definida positiva, y escribimos

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_y \\ \boldsymbol{\mu}_{\mathbf{x}} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_y^2 & \Sigma_{y\mathbf{x}} \\ \Sigma_{\mathbf{x}y} & \Sigma_{\mathbf{xx}} \end{bmatrix}$$

  con $\boldsymbol{\mu}_{\mathbf{x}} \in \mathbb{R}^{p-1}$, $\Sigma_{\mathbf{xx}} \in \mathbb{R}^{(p-1)\times(p-1)}$ y $cov(y, \mathbf{x}) = \Sigma_{y\mathbf{x}} \in \mathbb{R}^{1\times(p-1)}$ entonces la distribución condicional de y dado $\mathbf{x}$ es normal con esperanza condicional dada por

$$\mathbb{E}\left(y|\mathbf{x}\right) = \mu_y + \Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}\right),$$

  que resulta **lineal**, y varianza condicional,

$$var\left(y|\mathbf{x}\right) = \sigma_y^2 - \Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y} \quad \text{que es constante}$$

- Porque es un problema fácil de atacar

Mejor predictor lineal

Consideremos las siguientes condiciones de regularidad

Supuestos S.1

1. $\mathbb{E}\left(y^2\right) < \infty$
2. $\mathbb{E}\left(\|\mathbf{x}\|^2\right) < \infty$
3. $Q_{\mathbf{xx}} = \mathbb{E}\left(\mathbf{xx}'\right)$ es definida positiva.

Ojo: $Q_{\mathbf{xx}}$ no es una matriz de covarianza. ¿Cómo es la primer fila de $Q_{\mathbf{xx}}$ si x_0 es constante?

Observemos que para todo $\boldsymbol{\alpha} \in \mathbb{R}^k$ no nulo se tiene que
$\boldsymbol{\alpha}' Q_{\mathbf{xx}} \boldsymbol{\alpha} = \mathbb{E}\left(\boldsymbol{\alpha}'\mathbf{xx}'\boldsymbol{\alpha}\right) = \mathbb{E}\left(\boldsymbol{\alpha}'\mathbf{x}\right)^2 \geq 0$

Luego, $Q_{\mathbf{xx}}$ es semidefinida positiva por construcción. Sólo le agregamos el supuesto de que sea definida positiva. Recordemos que las matrices definidas positivas son inversibles.

Mejor predictor lineal

Teorema A3

Bajo los Supuestos S.1,

1. *Los momentos* $\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)$ *y* $\mathbb{E}(\boldsymbol{x}y)$ *existen.*
2. *El coeficiente lineal óptimo existe, es único y es igual a* $\boldsymbol{\beta}_{OL} = \left(\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1} \mathbb{E}(\boldsymbol{x}y)$
3. *El mejor predictor lineal de* y *dado* $\boldsymbol{x}$ *es* $\mathcal{P}(y|\boldsymbol{x}) = \boldsymbol{x}'\left(\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1} \mathbb{E}(\boldsymbol{x}y)$
4. *Podemos definir el error de proyección*

$$u = y - \boldsymbol{x}'\boldsymbol{\beta}_{OL}$$

El error de proyección satisface

   1. $\mathbb{E}\left(u^2\right) < \infty$
   2. $\mathbb{E}(\boldsymbol{x}u) = \mathbf{0}$
   3. *Si* $\mathbf{x}$ *contiene una coordenada constante, entonces* $\mathbb{E}(u) = 0$

Mejor predictor lineal

Es interesante reflexionar sobre la generalidad del Teorema A3 La única restricción es que valgan los Supuestos S.1. Por lo tanto, para cualquier vector aleatorio $(y, \mathbf{x})$ con varianzas finitas podemos definir la ecuación lineal $y = x'\boldsymbol{\beta}_{OL} + u$ con las propiedades enumeradas en el Teorema A3. Las suposiciones más fuertes, como asumir la linealidad de la esperanza condicional (Modelo ECL), no son necesarias.

Modelo de Proyección Lineal

$$\boxed{\begin{aligned} y &= \boldsymbol{x}'\boldsymbol{\beta} + u \\ \mathbb{E}(\boldsymbol{x}u) &= \mathbf{0} \\ \boldsymbol{\beta} &= \left(\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1}\mathbb{E}(\boldsymbol{x}y) \end{aligned}}$$

Mejor predictor lineal: variables centradas

A veces es útil separar la constante de las demas predictoras y escribir una reparametrización del modelo:

Teorema A4

En el modelo de proyección lineal, $y = \mathbf{x}'\boldsymbol{\beta} + \beta_0 + u$*, donde* β_0 *es el intercept y* $\mathbf{x}$ *no contiene una constante, se tiene que*

$$\boldsymbol{\beta} = \text{var}(\mathbf{x})^{-1}\,\text{cov}(\mathbf{x}, y) \quad (3)$$

$$\beta_0 = \mu_y - \boldsymbol{\mu}_{\mathbf{x}}'\boldsymbol{\beta} \quad (4)$$

donde $\mu_y = \mathbb{E}(y)$ *y* $\boldsymbol{\mu}_{\mathbf{x}} = \mathbb{E}(\mathbf{x})$.

Teoría asintótica

DEFINICIÓN: (CONVERGENCIA EN PROBABILIDAD) Diremos que una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ converge en probabilidad a otra variable aleatoria X si para todo $\delta > 0$ se tiene

$$\lim_{n\to+\infty} P\left(|X_n - X| \leq \delta\right) = 1.$$

Escribiremos

$$X_n \xrightarrow{p} X.$$

Vale también la misma definición para vectores o matrices aleatorias, reemplazando el módulo por la norma de vectores o matrices.

DEFINICIÓN: (CONVERGENCIA EN DISTRIBUCIÓN) Sea $(\mathbf{x}_n)_{n\geq 1}$ una sucesión de vectores aleatorios y F una función de distribución. Diremos que la sucesión $\mathbf{x}_n$ converge en distribución a F si para todo $\mathbf{u}$ para el cual $F(\mathbf{u})$ es continua $F_{\mathbf{x}_n}(\mathbf{u}) \to F(\mathbf{u})$ cuando $n \to \infty$. Escribiremos

$$\mathbf{x}_n \xrightarrow{d} F.$$

Ley débil de los grandes números (LGN) Sean Y_i, Y variables aleatorias i.i.d. con $\mathbb{E}|Y| < \infty$, entonces, cuando $n \to \infty$ tenemos

$$\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow{p} \mathbb{E}(Y) = \mathbb{E}(Y_1) = \mathbb{E}(Y_2)$$

LGN para vectores o matrices Sean $\mathbf{y}_i$, $\mathbf{y}$ vectores aleatorios i.i.d. con $\mathbb{E}\|\mathbf{y}\| < \infty$, entonces, cuando $n \to \infty$ tenemos

$$\bar{\mathbf{y}}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{y}_i \xrightarrow{p} \mathbb{E}(\mathbf{y})$$

Vale también en el caso en el que y_i, y sean matrices aleatorias.

Aquí la norma de matrices considerada es $\|A\| = \left(\sum_{i=1}^{m} \sum_{j=1}^{k} a_{ij}^2\right)^{1/2}$, es decir, la norma del vector que se obtiene concatenando las filas de A (norma Frobenius).

TEOREMA DE FUNCIONES CONTINUAS (TFC). Sean $\{\mathbf{z}_n\}_{n\in\mathbb{N}}$ una sucesión de vectores aleatorios y $\mathbf{c}$ un vector constantes tales que tales que $\mathbf{z}_n \xrightarrow{p} \mathbf{c}$ cuando $n \to \infty$. Sea $\mathbf{g}(\cdot)$ una función a valores reales o vectoriales, continua. en $\mathbf{c}$, entonces $\mathbf{g}\left(\mathbf{z_n}\right) \xrightarrow{p} \mathbf{g}(\mathbf{c})$ cuando $n \to \infty$.

Más generalmente, si el límite en probabilidad de la sucesión, $\mathbf{c}$, no fuera constante sino un vector aleatorio, y la función $\mathbf{g}$ fuera continua en el soporte de $\mathbf{c}$, entonces el resultado también es cierto, pero no usaremos este resultado.

TEOREMA CENTRAL DEL LÍMITE PARA VECTORES(TCL) Si $\{\mathbf{y}_i\}_{i\in\mathbb{N}}$ es una sucesión de vectores aleatorios $k-$dimensionales iid, con $\mathbb{E}\left\|\mathbf{y}_i\right\|^2 < \infty$. Entonces

$$\sqrt{n}\,(\overline{\mathbf{y}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathrm{N}_k(\mathbf{0}, V)$$

cuando $n \to \infty$, donde $\boldsymbol{\mu} = \mathbb{E}(\mathbf{y})$ y $V = \mathbb{E}\left((\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'\right)$

TEOREMA DE SLUTSKY Sean $\{z_n\}_{n\in\mathbb{N}}$ y $\{c_n\}_{n\in\mathbb{N}}$ dos sucesiones de variables tales que $z_n \xrightarrow{d} z$ y $c_n \xrightarrow{p} c$, cuando $n \to \infty$, con c una constante. Entonces

1. $z_n + c_n \xrightarrow{d} z + c$
2. $z_n c_n \xrightarrow{d} zc$
3. $\frac{z_n}{c_n} \xrightarrow{d} \frac{z}{c}$ si $c \neq 0$

Herramientas de Proba V

DESIGUALDAD PARA ESPERANZAS. Para cualquier matriz aleatoria A para la que $\mathbb{E}\|A\| < \infty$ se tiene que $\|\mathbb{E}(A)\| \leq \mathbb{E}\|A\|$

DESIGUALDAD DE CAUCHY-SCHWARZ. Sean A y B matrices aleatorias de dimensión $m \times n$, entonces

$$\mathbb{E}\left\|A'B\right\| \leq \left(\mathbb{E}\left(\|A\|^2\right)\right)^{1/2}\left(\mathbb{E}\left(\|B\|^2\right)\right)^{1/2}$$

Teoría asintótica para los EMC en regresión

Toda la teoría asintótica que mostraremos vale bajo el escenario más general que el que usamos a lo largo de toda la materia, es decir, vale aun sin necesidad de suponer el modelo ECL, de esperanza condicional lineal. Sólo necesitamos poder definir el modelo de proyección lineal homoscedástico para nuestras variables. En ese escenario vale la consistencia de los estimadores de los parámetros $\boldsymbol{\beta}$ y del estimador S^2, y la normalidad asintótica de los estimadores de los $\boldsymbol{\beta}$ que nos permitirá hacer tests e IC de nivel aproximado α cuando n sea suficientemente grande **sin** requerir que la distribución de los errores del modelo sea **normal**.

Consistencia del EMC

Supuestos S.1 Asintóticos

1. $\mathbb{E}\left(y^2\right) < \infty$
2. $\mathbb{E}\left(\|\mathbf{x}\|^2\right) < \infty$
3. $Q_{\mathbf{xx}} = \mathbb{E}\left(\mathbf{xx}'\right)$ es definida positiva.
4. Las observaciones $\left(y_i, \mathbf{x}_i\right), i = 1, \ldots, n$, son iid.

Teorema A5

Bajo los Supuestos S.1 Asintóticos, la sucesión $\left(\widehat{\boldsymbol{\beta}}_n\right)_n$ con $\widehat{\boldsymbol{\beta}}_n$ el estimador de mínimos cuadrados para regresión basado en $\left(y_i, \mathbf{x}_i\right), i = 1, \ldots, n$, cuando $n \to \infty$, $\widehat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}$, donde $\boldsymbol{\beta}$ es el coeficiente de proyección lineal,

$$\boldsymbol{\beta} = \left[\mathbb{E}\left(\mathbf{xx}'\right)\right]^{-1} \mathbb{E}\left(\mathbf{x}y\right)$$

Consistencia del EMC

Primero. Observemos que el estimador de mínimos cuadrados

$$\widehat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i y_i \right) = \widehat{Q}_{\boldsymbol{xx}}^{-1} \widehat{Q}_{\boldsymbol{x}y}$$

es función de los momentos muestrales $\widehat{Q}_{\boldsymbol{xx}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$ y $\widehat{Q}_{\boldsymbol{x}y} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i y_i$

Segundo. Como los $(y_i, \boldsymbol{x}_i)$ son vectores i.i.d., resulta que cualquier función de $(y_i, \boldsymbol{x}_i)$ son iid, incluyendo $\boldsymbol{x}_i \boldsymbol{x}_i'$ y $\boldsymbol{x}_i y_i$. Estas variables también tienen esperanzas finitas, bajo los Supuestos S.1. Entonces, por la LGN para vectores, tenemos

$$\widehat{Q}_{\boldsymbol{xx}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \xrightarrow{p} \mathbb{E}\left(\boldsymbol{x} \boldsymbol{x}' \right) = Q_{\boldsymbol{xx}}$$

$$\widehat{Q}_{\boldsymbol{x}y} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i y_i \xrightarrow{p} \mathbb{E}\left(\boldsymbol{x}' y \right) = Q_{\boldsymbol{x}y}$$

Consistencia del EMC

Tercero. Observemos que

$$\widehat{\beta} = g\left(\widehat{Q}_{\mathbf{xx}}, \widehat{Q}_{\mathbf{x}y}\right)$$

donde $g(A, \mathbf{b}) = A^{-1}\mathbf{b}$. La función $g(A, \mathbf{b})$ es una función continua de A y $\mathbf{b}$ siempre que A^{-1} exista. Los supuestos S.1 garantizan que $Q_{\mathbf{xx}}^{-1}$ exista y que la función $g(A, \mathbf{b})$ sea continua en $A = Q_{\mathbf{xx}}$.

El TFC nos permite combinar estos resultados, y concluir que, cuando $n \to \infty$,

$$\widehat{\beta} = \widehat{Q}_{\mathbf{xx}}^{-1}\widehat{Q}_{\mathbf{x}y} \xrightarrow{p} Q_{\mathbf{xx}}^{-1}Q_{\mathbf{x}y} = \beta. \quad \blacksquare$$

Corolario A6

$$\widehat{Q}_{\mathbf{x}y} \xrightarrow{p} Q_{\mathbf{x}y},\ \widehat{Q}_{\mathbf{xx}} \xrightarrow{p} Q_{\mathbf{xx}},\ \widehat{Q}_{\mathbf{xx}}^{-1} \xrightarrow{p} Q_{\mathbf{xx}}^{-1},\ \widehat{Q}_{\mathbf{x}\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} x_i u_i \xrightarrow{p} Q_{\mathbf{x}u}$$

Ilustración empírica (de*Econometrics*, de B. Hansen)

La Encuesta de Población Actual (Current Population Survey: CPS) es una encuesta mensual de aproximadamente 57.000 hogares de EE. UU. realizada por la Oficina de Estadísticas Laborales. El CPS es la fuente primaria de información sobre las características de la fuerza laboral de los Estados Unidos. La encuesta cubre características demográficas, de empleo, ingreso y educación. Se pueden encontrar detalles de la misma en `www.census.gov/cps` y `dataferrett.census.gov`.

De la encuesta de marzo de 2009, extrajeron los empleados a tiempo completo durante el año anterior (excluyendo militares). Los datos están en el archivo `cps09mar.txt` y consisten en 12 variables medidas en 50.742 individuos. Las variables se describen en `cps09mar_description.pdf`. Todos los archivos están disponibles en
`http://www.ssc.wisc.edu/~bhansen/econometrics/`

Ilustración empírica (de*Econometrics*, de B. Hansen)

`education` puede pensarse como años de estudio, está codificada numéricamente del siguiente modo

- 0 Less than 1st grade
- 4 1st, 2nd, 3rd, or 4th grade
- 6 5th or 6th grade
- 8 7th or 8th grade
- 9 9th grade
- 10 10th grade
- 11 11th grade or 12th grade with no high school diploma
- 12 High school graduate, high school diploma or equivalent
- 13 Some college but no degree
- 14 Associate degree in college, including occupation/vocation programs
- 16 Bachelor's degree or equivalent (BA, AB, BS)
- 18 Master's degree (MA, MS, MENG, MED, MSW, MBA)
- 20 Professional degree or Doctorate degree (MD, DDS, DVM, LLB, JD, PHD, EDD)

earnings: total annual wage and salary earnings
hours: number of hours worked per week
week: number of weeks worked per year

Ilustración empírica (de*Econometrics*, de B. Hansen)

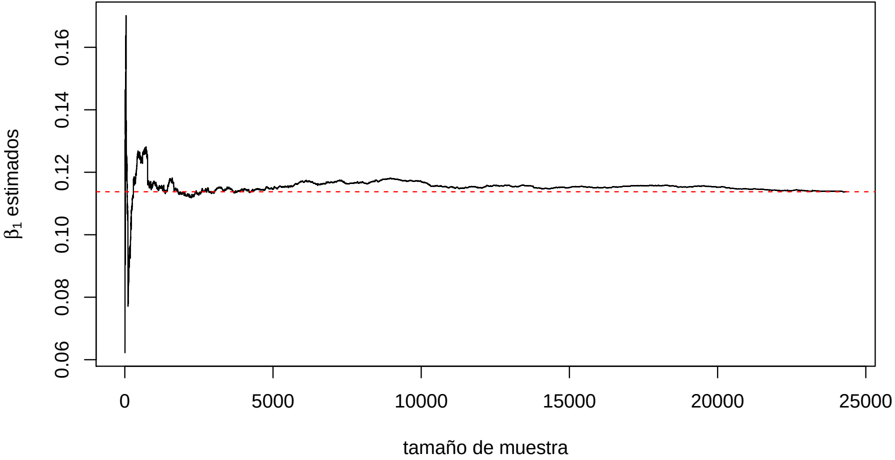
Consideremos el modelo

$$\ln(wage_i) = \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{experience}_i + \beta_3 \text{experience}_i^2 + e_i$$

donde la `experience` se define como `age-education-6` El `wage` es el total de ingresos anuales divididos por la cantidad de horas trabajadas `wage= earnings/(hours*week)` Utilizamos la muestra de 24.344 hombres blancos del CPS de marzo de 2009.

El script `cps.R` ordena aleatoriamente las observaciones y luego estima secuencialmente el modelo ajustando por mínimos cuadrados, comenzando con las primeras 5 observaciones reordenadas, y continuando hasta usar la muestra completa. La figura que sigue tiene la secuencia de estimaciones del parámetro β_1 en función del tamaño de muestra utilizado. Puede verse cómo la estimación de mínimos cuadrados cambia con el tamaño de la muestra, y a medida que el número de observaciones aumenta se estabiliza al valor estimado con la muestra completa, $\widehat{\beta_1} = 0.11376$

Ilustración empírica (de*Econometrics*, de B. Hansen)



Normalidad asintótica del EMC

Vale bajo supuestos más exigentes.

Supuestos S.2

1. $\mathbb{E}\left(y^4\right) < \infty$
2. $\mathbb{E}\left(\|\mathbf{x}\|^4\right) < \infty$
3. $Q_{\mathbf{xx}} = \mathbb{E}\left(\mathbf{xx}'\right)$ es definida positiva.
4. Las observaciones $\left(y_i, \mathbf{x}_i\right), i = 1, \ldots, n$, son iid.

Normalidad asintótica del EMC

Teorema A7

(Normalidad asintótica) Bajo los Supuestos S.2, y asumiendo homoscedasticidad de los errores,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, V_{\boldsymbol{\beta}}\right)$$

donde $V_{\boldsymbol{\beta}} = Q_{\boldsymbol{xx}}^{-1}\sigma^2, \;\; Q_{\boldsymbol{xx}} = \mathbb{E}\left(\boldsymbol{x}_1\boldsymbol{x}_1'\right)$

La matriz $V_{\boldsymbol{\beta}} = \sigma^2 Q_{\boldsymbol{xx}}^{-1}$ es la varianza de la distribución asintótica de $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

En consecuencia, se la denomina la *matriz de varianza asintótica* de $\widehat{\boldsymbol{\beta}}$.

La podemos comparar con la varianza condicional del estimador $\widehat{\boldsymbol{\beta}}$ en el modelo ECL homoscedástico, que es

$$V_{\widehat{\beta}} = \mathrm{var}(\widehat{\boldsymbol{\beta}}|\mathrm{X}) = \sigma^2\left(\mathrm{X}'\mathrm{X}\right)^{-1}$$

Hay un factor n de diferencia, ¿por qué?

Normalidad asintótica del EMC

Teorema A8

(Consistencia del estimador de la varianza del error) Bajo los Supuestos S1 Asintóticos,

1. $S^2 \xrightarrow{p} \sigma^2$ *cuando* $n \to \infty$
2. *Si los errores son homoscedásticos, entonces el estimador de la varianza de* $\widehat{\boldsymbol{\beta}}$*, que es* $\widehat{V}_{\widehat{\boldsymbol{\beta}}} = \widehat{Q}_{xx}^{-1} S^2 \xrightarrow{p} Q_{x\mathbf{x}}^{-1}\sigma^2 = V_\beta$ *cuando* $n \to \infty$*, es por lo tanto, un estimador consistente.*

Observación: El modelo de errores heteroscedástico está tratado en el libro de Hansen con bastante detalle. Ahí se proponen y estudian distintos estimadores de la covarianza de $\widehat{\boldsymbol{\beta}}$.

Tests de Hipótesis e Intervalos de Confianza para combinaciones lineales de $\boldsymbol{\beta}$

Teorema A9

Sea $\mathbf{a} \in \mathbb{R}^p$ *fijo, no nulo. Definimos* $\theta = \mathbf{a}'\boldsymbol{\beta} \in \mathbb{R}$, *y consideremos su estimador* $\widehat{\theta}_n = \mathbf{a}'\widehat{\boldsymbol{\beta}_n}$ *basado en la muestra iid* $\{(y_i, \mathbf{x}_i), i = 1, \ldots, n, \}$. *Entonces cuando* $n \to \infty$,

i.
$$T = \frac{\sqrt{n}\,(\widehat{\theta}_n - \theta)}{S\sqrt{\mathbf{a}'\left[\widehat{Q}_{\boldsymbol{xx}}\right]^{-1}\mathbf{a}}} = \frac{\sqrt{n}\,\mathbf{a}'(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})}{S\sqrt{\mathbf{a}'\left[\widehat{Q}_{\boldsymbol{xx}}\right]^{-1}\mathbf{a}}} \xrightarrow{d} N(0,1)$$

ii. *Un IC de nivel asintótico* $1-\alpha$ *para* $\theta = \mathbf{a}'\boldsymbol{\beta}$ *está dado por*
$$\mathbf{a}'\widehat{\boldsymbol{\beta}_n} \pm z_{(\alpha/2)}\frac{S}{\sqrt{n}}\sqrt{\mathbf{a}'\left[\widehat{Q}_{\boldsymbol{xx}}\right]^{-1}\mathbf{a}}$$

donde $\Phi\left(z_{(\alpha/2)}\right) = 1 - \alpha/2$.

Tests de Hipótesis e Intervalos de Confianza para combinaciones lineales de $\boldsymbol{\beta}$

Teorema A9 (cont.)

iii. *Un test de nivel asintótico α para $H_0 : \mathbf{a}^T\boldsymbol{\beta} = c$, con $c \in \mathbb{R}$ versus $H_0 : \mathbf{a}^T\boldsymbol{\beta} \neq c$ está dado por la regla de decisión que consiste en*

$$\textit{Rechazar } H_0 \textit{ cuando } |T_c| \geq z_{(\alpha/2)} \textit{ con } T_c = \frac{\sqrt{n}\left(\mathbf{a}'\widehat{\boldsymbol{\beta}} - c\right)}{S\sqrt{\mathbf{a}'\left[\widehat{Q}_{\boldsymbol{xx}}\right]^{-1}\mathbf{a}}}$$

Resulta que T dado en (i) es una expresión *asintóticamente pivotal* para obtener intervalos de confianza para $\mathbf{a}^T\boldsymbol{\beta}$

Observar que el ítem (ii) de este teorema permite construir IC de nivel asintótico para $\mathbb{E}(y|\boldsymbol{x} = \boldsymbol{x}_0)$, bajo el modelo ECL. La longitud del IC dependerá de $\boldsymbol{x}_0$.

Deuda

Probemos los items que quedaron pendientes de la demostración del Teorema A3 hecha en clase. Queremos ver que $\mathbb{E}(\boldsymbol{x}y)$ existe. Como $\mathbb{E}(\boldsymbol{x}y) \in \mathbb{R}^p$, veamos que $|\mathbb{E}(x_j y)|$ es finita, para $j = 0, \ldots, p-1$. Tenemos

$$\begin{aligned} |\mathbb{E}(x_j y)| &\leq \mathbb{E}(|x_j y|) && \text{(por la desigualdad de esperanzas)} \\ &\leq \left(\mathbb{E}\left(|x_j|^2\right)\right)^{1/2} \left(\mathbb{E}\left(|y|^2\right)\right)^{1/2} && \text{(por Cauchy–Schwarz)} \end{aligned}$$

Como $|x_j| \leq \|\mathbf{x}\|$, entonces $|x_j|^2 \leq \|\mathbf{x}\|^2$ y luego $\mathbb{E}(|x_j|^2) \leq \mathbb{E}(\|\mathbf{x}\|^2)$ que es finita por hipótesis. También $\mathbb{E}\left(|y|^2\right)$ es finita por hipótesis, y queda probado.

Además, $\mathbb{E}(\boldsymbol{x}\boldsymbol{x}')$ existe por hipótesis.

Deuda

Resta ver que $\mathbb{E}\left(u^2\right) < \infty$,

$$
\begin{aligned}
\mathbb{E}\left(u^2\right) &= \mathbb{E}\left(\left(y - \boldsymbol{x}'\boldsymbol{\beta}_{OL}\right)^2\right) \\
&= \mathbb{E}\left(y^2\right) - 2\mathbb{E}\left(y\boldsymbol{x}'\right)\boldsymbol{\beta}_{OL} + \boldsymbol{\beta}_{OL}'\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\boldsymbol{\beta}_{OL} \\
&= \mathbb{E}\left(y^2\right) - 2\mathbb{E}\left(y\boldsymbol{x}'\right)\left(\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1}\mathbb{E}(\boldsymbol{x}y) \\
&+ \mathbb{E}(\boldsymbol{x}y)'\left(\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1}\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\left(\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1}\mathbb{E}(\boldsymbol{x}y) \\
&= \mathbb{E}\left(y^2\right) - \mathbb{E}\left(y\boldsymbol{x}'\right)\left(\mathbb{E}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1}\mathbb{E}(\boldsymbol{x}y) \\
&\leq \mathbb{E}\left(y^2\right) \quad \text{(¿por qué?)} \\
&< \infty
\end{aligned}
$$