ECONOMETRÍA I

Tema 3: El Modelo de Regresión Lineal Múltiple: estimación

Patricia Moreno
Juan Manuel Rodriguez Poo
Alexandra Soberon
Departamento de Economía

Limitaciones del Análisis de Regresión Simple

Existen diversos problemas que no pueden ser resueltos con el modelo de regresión simple:

- Es complicado extraer conclusiones *ceteris paribus*. A la hora de realizar un análisis causal es mejor controlar más factores que recurrir al supuesto de media condicional cero, $E(u|x) = 0$.
- Si usamos un único regresor, solamente somos capaces de explicar una parte limitada de la variabilidad de y en función de la información proporcionada por esa x.
- Sólo se puede incorporar una relación funcional concreta entre x e y (en función de x, $logx$, etc.)

- **Objetivo**: determinar el impacto de la educación (educ) sobre los salarios (wage) cuando también se conoce la experiencia laboral (exper).

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u.$$

- Este modelo incluye *exper* para controlar su efecto sobre *wage*.
- Incorporar *exper* explícitamente es importante porque la educación y la experiencia laboral están positivamente correlacionados.
- En el modelo de regresión lineal la experiencia está incluida en el término de error por lo que el estimador MCO estaría sesgado.

- El Modelo de Regresión Lineal Múltiple nos permite explicar relaciones económicas en las que intervienen más de dos variables.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u,$$

  donde
  - β_0: término de intercepto.
  - β_j $(j = 1, \cdots, k)$: parámetro de la pendiente. Se interpreta como el efecto parcial sobre y de un cambio en x_j, *ceteris paribus*.
  - u: término de error.
- **Supuesto clave**: $E(u|x_1, \cdots, x_k) = 0$.

Modelo de regresión con 2 variables

Objetivo: explicar el efecto del gasto por estudiante (*expend*) sobre la calificación media en un examen estandarizado (*avgscore*) controlando también por el ingreso familiar medio (*avginc*).

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u,$$

donde

- β_0: intercepto.
- β_1: mide cómo cambia *avgscore* ante un cambio en *expend*, manteniendo todos los demás factores constantes.
- β_2: mide cómo cambia *avgscore* ante un cambio en *avginc*, manteniendo todos los demás factores constantes.

Modelo de regresión múltiple

- **Objetivo**: estimar un modelo en el cual x_1 y x_2 están correlacionados perfectamente pero no de forma lineal.

$$cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u,$$

  donde *cons* es el consumo del hogar e *inc* es el ingreso del hogar.
- Como en este caso estimamos un modelo de regresión múltiple pero la renta es el único factor que afecta al consumo, la interpretación de los parámetros cambia dado que x_2 no puede ser cosntante si x_1 cambia.
- β_1 y β_2 no tienen interpretación por separado.
- Propensión marginal a consumir:

$$\frac{\Delta cons}{\Delta inc} = \beta_1 + 2\beta_2 inc.$$

- Los estimadores MCO resuelven el siguiente problema

$$min_{\beta_0,\beta_1,\beta_2} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\widehat{\beta}_2 \right)^2 .$$

de modo que los estimadores MCO minimizan el promedio de la diferencia al cuadrado entre los valores actuales y_i y los valores predichos (línea estimada).

- **Condiciones de primer orden**:

$$\begin{aligned}
\sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\widehat{\beta}_2 \right) &= 0, \\
\sum_{i=1}^{n} x_{1i} \left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\widehat{\beta}_2 \right) &= 0, \\
\sum_{i=1}^{n} x_{2i} \left(y_i - \widehat{\beta}_0 - x_{1i}\widehat{\beta}_1 - x_{2i}\widehat{\beta}_2 \right) &= 0.
\end{aligned}$$

- Sea la ecuación estimada

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2$$

- Los estimadores de las pendientes se interpretan como efectos parciales o *ceteris paribus*, i.e.,

$$\Delta\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \Delta x_1 + \widehat{\beta}_2 \delta x_2,$$

donde
  - $\widehat{\beta}_1 = \frac{\Delta E[y|x_1,x_2]}{\Delta x_1}$: Si x_1 varía 1 unidad, y varía en promedio β_1 unidades de y cuando x_2 se mantiene constante, i.e., cuando $\Delta x_2 = 0$, $\Delta\widehat{y} = \widehat{\beta}_1 \Delta x_1$.
  - $\widehat{\beta}_2 = \frac{\Delta E[y|x_1,x_2]}{\Delta x_2}$: si x_2 varía 1 unidad, y varía en promedio β_2 unidades de y *ceteris paribus*, i.e., cuando $\Delta x_1 = 0$, $\Delta\widehat{y} = \widehat{\beta}_2 \Delta x_2$.
  - β_0: valor predicho del promedio de y cuando $x_1 = x_2 = 0$.

Interpretación del modelo con términos cuadráticos

- **Modelo**: Cuando queremos modelizar la relación entre x e y considerando la existencia de efectos marginales crecientes o decrecientes el modelo a considerar es:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u,$$

  donde $E(u|x) = 0 \Longrightarrow \quad E(y|x) = \beta_0 + \beta_1 x + \beta_2 x^2.$

- **Interpretación**: cuando x varía 1 unidad, y varía en media en $(\beta_1 + 2\beta_2 x)$ unidades de y, i.e.

$$\frac{\Delta E(y|x)}{\Delta x} = \beta_1 + 2\beta_2 x.$$

- β_1 y β_2 no tienen interpetación por separado.
- Existe un valor crítico ($x^* = -\beta_1/2\beta_2$) en el que el efecto de x sobre $E(y|x)$ cambia de signo.

Modelo de regresión lineal múltiple: forma matricial

- Sea el modelo de regresión con k variables explicativas,

$$Y = X\beta + u.$$

- Y es un vector $n \times 1$ de observaciones de Y.
- X es una matriz $n \times k$ de observaciones de las variables explicativas.
- u es un vector $n \times 1$ de perturbaciones no observables.

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{k1} \\ 1 & X_{12} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \cdots & X_{kn} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

MCO en notación matricial

- **Modelo de regresión ajustado**: $Y = X\widehat{\beta} + \widehat{u}$,
- **Residuos al cuadrado**:

$$\begin{aligned} \widehat{u}'\widehat{u} &= X\widehat{\beta}(Y - X\widehat{\beta})'(Y - X\widehat{\beta}) = (Y' - \widehat{\beta}'X')(Y - X\widehat{\beta}) \\ &= Y'Y - \widehat{\beta}'X'Y - Y'X\widehat{\beta} + \widehat{\beta}'X'X\widehat{\beta} \\ &= Y'Y - 2\widehat{\beta}'X'Y + \widehat{\beta}'X'X\widehat{\beta}, \end{aligned}$$

- **Condición de primer orden**:

$$\frac{\partial(\widehat{u}'\widehat{u})}{\partial\widehat{\beta}} = -2X'(Y - X\widehat{\beta}) = 0_k,$$

Esta c.p.o. es la misma que $X'\widehat{u} = 0$. Reordenando $X'X\widehat{\beta} = X'Y$.

$$\widehat{\beta} = (X'X)^{-1} X'Y.$$

Propiedades algebráicas MCO

- **Valores ajustados**: $\widehat{Y} = X\widehat{\beta}$.
- **Residuos**: $\widehat{u} = Y - \widehat{Y} = Y - X\widehat{\beta}$.

- **Propiedades**:
  - La suma de los residuos MCO es cero: $\sum_{i=1}^{n} \widehat{u}_i = 0$.
  - La media muestral de los residuos es cero. Así $\overline{Y} = \overline{\widehat{Y}}$.
  - La covarianza muestral entre cada una de las variables independientes y los residuos MCO es cero.
  - La línea de regresión MCO siempre va a través de la media de la muestra ($\overline{Y} = X\widehat{\beta}$). Así,

$$X'[Y - X\widehat{\beta}] = X'\widehat{u} = 0_k.$$

Bondad de ajuste

- Dado que la variación total en y_i es igual a la suma de la variación total en $\widehat{y}_i$ y en $\widehat{u}_i$ ($y_i = \widehat{y}_i + \widehat{u}_i$) podemos definir:
  - **Suma Total de los Cuadrados (STC)**:

$$STC = \sum_{i=1}^{n} (y_i - \overline{y})^2 \rightarrow STC = Y'Y - n\overline{Y}^2 = Y'M_iY.$$

  - **Suma de Cuadrados Explicada (SCE)**:

$$SCE = \sum_{i=1}^{n} (\widehat{y}_i - \overline{y}) \rightarrow SCE = \widehat{Y}'\widehat{Y} - n\overline{Y}^2 = \widehat{Y}'M_i\widehat{Y}.$$

  - **Suma de Cuadrados de los Residuos (SCR)**:

$$SCR = \sum_{i=1}^{n} \widehat{u}_i^2 \rightarrow SCR = \widehat{u}'\widehat{u} = Y'MY.$$

donde $M_i = I_n - \imath(\imath'\imath)^{-1}\imath'$ y $M = I - X(X'X)^{-1}X'$.

Bondad de ajuste

- $SST = SSE + SSR$.
- Suponiendo que la variación total en Y no sea cero,

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = 1 - \frac{Y'MY}{Y'M_iY},$$

- R^2 representa la proporción de la variación muestral en Y que es explicada por la línea de regresión MCO.

Bondad de ajuste

- R^2 también puede ser entendido como el coeficiente de correlación al cuadrado entre el valor actual y_i y el valor predicho $\widehat{y}_i$:

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_i - \overline{y})(\widehat{Y}_i - \overline{\widehat{y}})\right)^2}{\left(\sum_{i=1}^{n}(y_i - \overline{y})^2\right)\left(\sum_{i=1}^{n}(\widehat{y}_i - \overline{\widehat{y}})^2\right)}.$$

- **Propiedades** R^2:
  - $0 \leq R^2 \leq 1$.
  - R^2 nunca disminuye cuando otra variable independiente es añadida a la regresión. Normalmente aumenta.
  - R^2 no es un instrumento muy confiable para decidir si incluir o no un nuevo regresor en la regresión dado que normalmente siempre aumenta a medida que se incorporar más variables explicativas.

- En ocasiones la teoría económica impone la restricción de que cuando $X_k = 0$ el valor esperado de Y es cero, es decir, $\beta_0 = 0$. Por ejemplo, si los ingresos son iguales a cero ($X = 0$) los impuestos asociados al trabajo también son iguales a cero ($Y = 0$).
- **Modelo**: $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{k_i}$, $i = 1, \cdots, n$.
- El estimador MCO minimiza el problema

$$\sum_{i=1}^{n} (y_i - \widetilde{\beta}_j x_{ji})^2, \quad j = 1, \cdots, k.$$

- **Condiciones de primer orden**:

$$\sum_{i=1} x_{ji}(y_i - \widetilde{\beta}_j x_{ji}) = 0$$

- **Estimador MCO**:

$$\widetilde{\beta}_j = \frac{\sum_{i=1}^{n} x_{ji} y_i}{\sum_{i=1}^{n} x_{ji}^2}$$

- El ajuste MCO ya no satisface las mismas condiciones que en el caso general:
  - $\widehat{u}$ ya no tienen media muestral cero.
  - Si definimos $R^2 = 1 - SSR/SST$ entonces R^2 puede ser negativa.
  - Si $\beta_0 \neq 0$, los estimadores MCO pueden estar sesgados.

- **Regresión simple**: $\widetilde{y}_i = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_{1i}$.
- **Regresión múltiple**: $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{1i} + \widehat{\beta}_2 x_{2i}$.
- Parcializando obtenemos la siguiente relación

$$\widetilde{\beta}_1 \quad = \quad \widehat{\beta}_1 \frac{\sum_{i=1}^{n} \widehat{r}_{1i} x_{1i}}{\sum_{i=1}^{n} \widehat{r}_{1i}^2} + \widehat{\beta}_k \frac{\sum_{i=1}^{n} \widehat{r}_{1i} x_{ki}}{\sum_{i=1}^{n} \widehat{r}_{1i}^2} + \widehat{\beta}_k \frac{\sum_{i=1}^{n} \widehat{r}_{1i} \widehat{r}_{2i}}{\sum_{i=1}^{n} \widehat{r}_{1i}^2}$$

$$\widetilde{\beta}_1 = \widehat{\beta}_1 + \widehat{\beta}_k \widetilde{\delta}_1.$$

- Normalmente, $\widetilde{\beta}_1 \neq \widehat{\beta}_1$ aunque existen dos casos en que $\widetilde{\beta}_1 = \widehat{\beta}_1$.
  - $\widehat{\beta}_2 = 0$: el efecto parcial de X_2 sobre $\widehat{Y}$ es cero.
  - $\widetilde{\delta}_1$: X_1 y X_2 no están correlacionadas en la muestra.

- **Objetivo**: comparación entre estimación de regresión simple y de regresión múltiple.
- El modelo lineal general $Y = X\widehat{\beta} + \widehat{u}$ se puede escribir como

$$Y = X_r\widehat{\beta}_r + X_s\widehat{\beta}_s + \widehat{u}$$

donde X_r y X_s son submatrices de dimensión $n \times r$ y $n \times s$, respectivamente. $\widehat{\beta}_r$ es un subvector $r \times 1$ que contiene los r primeros parámetros de $\widehat{\beta}$ y $\widehat{\beta}_s$ es un subvector $s \times 1$ que contiene los $s = k - r$ restantes de $\widehat{\beta}$.

Parcialización

- Sean H y M matrices de proyección (simétrica e idempotente)

$$H = X(X'X)^{-1}X' \qquad \text{y} \qquad M = I - X(X'X)^{-1}X'.$$

- Sea la matriz

$$M_s = I_n - X_s(X_s'X_s)^{-1}X_s',$$

multiplicamos el modelo por M_s

$$M_sY = M_sX_r\widehat{\beta}_r + M_sX_s\widehat{\beta}_s + M_s\widehat{u}$$

y dado que $M_sX_s = 0$ y $M\widehat{u} = MMY = MY = \widehat{u}$,

$$M_sY = M_sX_r\widehat{\beta}_r + M_s\widehat{u}.$$

- **Estimador MCO**:

$$\widehat{\beta}_r = (X_r'M_s'M_sX_r)^{-1}X_r'M_s'M_sY = (X_r'M_sX_r)^{-1}X_r'M_sY.$$

- Considerando el caso en que $k = 2$, $\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2$,

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \widehat{r}_{i1} y_i}{\sum_{i=1}^{n} \widehat{r}_{i1}^2},$$

  donde $\widehat{r}_{i1}$ son los residuos de la regresión estimada $\widehat{X}_1 = \widehat{\gamma}_0 + \widehat{\gamma}_2 X_2$.

- Interpretación: sólo la parte de x_{1i} que está incorrelacionada con x_{2i} está relacionada con y_i. Por lo tanto, estamos estimando el efecto de x_1 sobre y después de que x_2 haya sido parcializado.

- **Supuestos**:
  - **RLM.1**:*Linealidad en parámetros*:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u.$$

  - **RLM.2**:*Muestreo aleatorio*. Usamos un muestreo aleatorio con n observaciones $\{(x_{1i}, x_{2i}, \cdots, x_{ki}) : i = 1, 2, \cdots, n\}$ de acuerdo con el modelo poblacional propuesto. Así,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i.$$

  - **RLM.3**:*No hay multicolinealidad perfecta*. Ninguno de los regresores es constentes y tampoco existen relaciones lineales exactas entre las variables independientes.
  - **RLM.4**:*Media condicional cero*: $E(u|x_1, x_2, \cdots, x_k) = 0$. Así, todas las variables explicativas son exógenas.

Valores esperados del estimador MCO

Teorema Bajo los supuestos RML.1-RML.4, el estimador MCO $\widehat{\beta}_j$ es insesgado para β_j

$$E(\widehat{\beta}_j) = \beta_j, \qquad j = 1, \cdots, k,$$

para cualquier valor de los parámetros poblacionales β_j.

- El supuesto clave es RML.3.
- La propiedad de insesgadez no dice nada sobre el resultado de la estimación. Sólo es una propiedad del método de estimación MCO.

Valores esperados del estimador MCO: matricial

- **Supuestos**:
  - **(A.1)** El modelo puede ser escrito como $Y = X\beta + u$, donde Y es un vector $n \times 1$ observado, X es una matriz observada $n \times k$ y u es un vector $n \times 1$ de términos de error no observados.
  - **(A.2)** La matriz X tiene rango k.
  - **(A.3)** Cada error v_i tiene media cero, condicional a toda la matrix X, $E(u_i|X) = 0$.

- **Teorema**: Bajo los supuestos **(A.1)-(A.3)**, el estimador MCO $\widehat{\beta}$ es insesgado para β.

$$E(\widehat{\beta}) = \beta$$

Proof:

$$E(\widehat{\beta}) = \beta + E_X\{(X'X)^{-1}X'E(u|X)\}.$$

- **Sobre-especificación** del modelo de regresión múltiple: cuando una o más regresores están incluidos en el modelo a pesar de que en la población no tienen efecto parcial en y.
- Sea

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

donde X_3 no tiene efecto sobre Y una vez controladas por X_1 y X_2 por lo que $\beta_3 = 0$,

$$E(Y|X_1, X_2, X_3) = E(Y|X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$$

- Efectos de la inclusión de variables irrelevantes: Ninguno. $E(\widehat{\beta}_3) = \beta_3 = 0$ y el estimador MCO sigue siendo insesgado.

- Cuando omitimos una variable relevante (subespecificamos el modelo) incurrimos en un caso particular de **error de especificación**.
- **Efecto de la subespecificación**: los estimadores MCO normalmente serán sesgados. ¿Por qué?
- **Modelo verdadero**:

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 abil_i + u_i.$$

- **Modelo a estimar**:

$$wage_i = \beta_0 + \beta_1 educ_i + v_i, \qquad v_i = \beta_2 abil_i + u_i.$$

- **Estimador MCO**: siendo $x_1 = educ$,

$$\widetilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{1i} - \overline{x}_1)y_i}{\sum_{i=1}^{n}(x_{1i} - \overline{x}_1)^2}.$$

- Reemplazando $\widetilde{\beta}_1$ en el verdadero modelo y siendo $x_2 = abil$,

$$\sum_{i=1}^{n}(x_{1i} - \overline{x}_1)(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i)$$

y usando $E(u_i) = 0$,

$$E(\widetilde{\beta}_1) = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{1i} - \overline{x}_1)x_{2i}}{\sum_{i=1}^{n}(x_{1i} - \overline{x}_1)^2}$$

ó

$$E(\widetilde{\beta}_1) = \beta_1 + \beta_2 \widetilde{\delta}_1,$$

donde $\widetilde{\delta}_1$ es el coeficiente de la pendiente de regresión ($\widetilde{x}_2 = \widetilde{\delta}_0 + \widetilde{\delta}_1 x_1$) usando la misma muestra.

- Así, $\widetilde{\beta}_1$ está sesgado para estimar β_1:

$$Sesgo(\widetilde{\beta}_1) = E(\widetilde{\beta}_1) - \beta_1 = \beta_2 \widetilde{\delta}_1.$$

- Existen dos situaciones en las cuales $\widetilde{\beta}_1$ es insesgado:
  - Si $\beta_2 = 0$, por lo que x_2 no aparece en el modelo poblacional.
  - Si $\widetilde{\delta}_1 = 0$, es decir, la covarianza muestral entre x_1 y x_2 es cero.
- **Signo del sesgo**: $Sesgo(\widetilde{\beta}_1) = E(\widetilde{\beta}_1) - \beta_1 = \beta_2 \widetilde{\delta}_2$.

	$Corr(x_1, x_2) > 0$	$Corr(x_1, x_2) > 0$
$\beta_2 > 0$	Sesgo positivo	Sesgo negativo
$\beta_2 < 0$	Sesgo negativo	Sesgo positivo

Estimadores MCO: Omisión de variables relevantes

- **Objetivo**: determinar el impacto de la omisión de múltiples regresores en el modelo estimado.
- **Modelo poblacional**:

$$Y = X_r\beta_r + X_s\beta_s + u.$$

- **Modelo a estimar**:

$$Y = X_r\beta_r + \epsilon, \qquad \epsilon = X_s\beta_s + u.$$

$$\begin{aligned}\widehat{\beta}_r = (X_r'X_r)^{-1}X_r'Y \Rightarrow \widehat{\beta}_r &= (X_r'X_r)^{-1}X_r'[X_r\beta_r + X_s\beta_s + u] \\ &= \beta_r + (X_r'X_r)^{-1}X_r'X_s\beta_s + (X_r'X_r)^{-1}X_r'u.\end{aligned}$$

- **Sesgo por error de especificación**:

$$E(\widehat{\beta}_r) = \beta_r + \underbrace{(X_r'X_r)^{-1}X_r'X_s\beta_s}_{Sesgo}.$$

- Además de saber que la distribución muestral de $\widehat{\beta}_j$ está centrada, es importante conocer su variabilidad alrededor de β_j.
- Para hallar la varianza debemos añadir el supuesto de homocedasticidad.
- **RLM.5**: *Homocedasticidad condicional*. Dado cualquier valor x para $(x_1, x_2, \cdots, x_k)$, el término de error u tiene la misma varianza, es decir,

$$Var(u|x_1, x_2, \cdots, x_k) = \sigma^2.$$

  Si esto no se cumple, estaríamos en presencia de heterocedasticidad.
- Los cuatro supuestos sobre la insesgadez (**RLM.1-RLM.4**) junto con este supuesto de homocedasticidad son conocidos como los **supuestos de Gauss-Markov**.

Bajo los supuestos **RLM.1-5**, condicional en los valores de las variables independientes,

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \quad j = 1, \cdots, k,$$

donde

- R_j^2 es el R^2 de la regresión de x_j sobre todos los demás regresores (y el término constante).
- SST_j es la variación total de x_j, es decir,

$$SST_j = \sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2.$$

Varianza MCO: notación matricial

$$
\begin{aligned}
E[uu'|X] &= E\left[\left.\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}\begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix}\right| X\right] \\
&= \begin{pmatrix} E(u_1^2|X) & E(u_1u_2|X) & \cdots & E(u_1u_n|X) \\ E(u_1u_2|X) & E(u_2^2|X) & \cdots & E(u_2u_n|X) \\ \vdots & \vdots & \cdots & \vdots \\ E(u_nu_1|X) & E(u_nu_2|X) & \cdots & E(u_n^2|X) \end{pmatrix}.
\end{aligned}
$$

Varianza MCO: notación matricial

Si asumimos que la varianza del término de error es homocedástica,

$$E(uu'|X) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

De este modo,

$$E[uu'|X] = \sigma^2 I_n,$$

donde I_n es una matriz diagonal $n \times n$.

Varianza MCO: notación matricial

$$
\begin{aligned}
Var(\widehat{\beta}|X) &= E\left[(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'|X\right] \\
&= E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}|X\right] \\
&= \left[(X'X)^{-1}X'\underbrace{E(uu'|X)}_{\sigma^2}X(X'X)^{-1}\right] \\
Var(\widehat{\beta}|X) &= \sigma^2(X'X)^{-1}.
\end{aligned}
$$

Varianza de los estimadores MCO

- **Varianza del error,** σ^2: cuando mayor sea σ^2 mayor será $Var(\widehat{\beta}_j)$.
  - Es una propiedad de la población, no depende de n y es desconocido.
  - Para reducirla es necesario introducir más variables explicativas. Problema: en ocasiones no siempre es deseable ni posible.
- **Varianza total de** x_j**,** SST_j: cuanto mayor SST_j, menor será $Var(\widehat{\beta}_j)$.
- **Relación lineal entre los regresores**: un mayor R_j^2 está asociado a una mayor $Var(\widehat{\beta}_j)$.
  - $R_j^2 \simeq 1$: X_j se puede explicar en gran parte por los otros regresores.
  - A una correlación fuerte (pero no perfecta) entre 2 o más regresores se le conoce como **multicolinealidad**. Esta situación aumenta $Var(\widehat{\beta}_j)$, como lo hace una muestra pequeña, todo lo demás igual.

Varianza en los modelos mal especificados

- A la hora de decidir si incluir una variable o no en la regresión existe un trade-off entre el sesgo y la varianza.
- **Modelo verdadero**:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- **Estimador modelo regresión múltiple**:

$$\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2.$$

- **Estimador modelo regresión simple**:

$$\widetilde{Y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 X_1.$$

- Si $\beta_2 \neq 0$, $\widetilde{\beta}_1$ estará sesgado. Desde el punto de vista del sesgo $\widehat{\beta}_1$ será preferido.

Varianza en los modelos mal especificados

- Desde el punto de vista de la varianza:

$$
\begin{aligned}
Var(\widehat{\beta}_1) &= \frac{\sigma^2}{SST_1(1-R_1^2)} \\
Var(\widetilde{\beta}_1) &= \frac{\sigma^2}{SST_1}.
\end{aligned}
$$

  Por lo tanto,

$$Var(\widetilde{\beta}_1) < Var(\widehat{\beta}_1).$$

- Si x_1 y x_2 están correlacionados:

$$Var(\widetilde{\beta}_1) = Var(\widehat{\beta}_1).$$

- Así, si x_1 y x_2 no están incorrelacionados:
  - Si $\beta_2 \neq 0$, $\widetilde{\beta}_1$ es sesgado, $\widehat{\beta}_1$ es insesgado y $Var(\widetilde{\beta}_1) < Var(\widehat{\beta}_1)$.
  - Si $\beta_2 = 0$, $\widetilde{\beta}_1$ y $\widehat{\beta}_1$ son insesgados y $Var(\widetilde{\beta}_1) < Var(\widehat{\beta}_1)$.

- **Modelo verdadero**:

$$Y = X_r\beta_r + X_s\beta_s + u$$

- **Modelo a estimar**:

$$Y = X_r\beta_r + \epsilon, \qquad \epsilon = X_s\beta_s + u.$$

- **Estimador MCO**:

$$\widetilde{\beta}_r = (X_r'X_r)^{-1}X_r'Y$$

- **Varianza del estimador mal especificado**:

$$V(\widetilde{\beta}_r) \;=\; \sigma^2(X_r'M_sX_r)^{-1}.$$

- **Varianza del estimador bien especificado**:

$$\begin{aligned}
V(\widehat{\beta}_r) &= E\left[(\widehat{\beta}_r - E(\widehat{\beta}_r))(\widehat{\beta}_r - E(\widehat{\beta}_r))'\right] \\
&= E\left[(X_r'X_r)^{-1}X_r'uu'X_r(X_r'X_r)^{-1}\right] \\
&= (X_r'X_r)^{-1}X_r'\underbrace{E(uu'|X_r)}_{\sigma^2 I_n}X_r(X_r'X_r)^{-1} = \sigma^2(X_r'X_r)^{-1}
\end{aligned}$$

- **Prueba de la varianza del modelo mal especificado**:

$$\begin{aligned}
V(\widetilde{\beta}_r) &= E\left[(\widetilde{\beta}_r - E(\widetilde{\beta}_r))(\widetilde{\beta}_r - E(\widetilde{\beta}_r))'\right] \\
&= E\left[(X_r'M_sX_r)^{-1}X_r'M_suu'M_sX_r(X_r'M_sX_r)^{-1}\right] \\
&= (X_r'M_sX_r)^{-1}X_r'M_s\underbrace{E(uu'|X_r,X_s)}_{\sigma^2 I_n}M_sX_r(X_r'M_sX_r)^{-1} \\
&= \sigma^2(X_r'M_sX_r)^{-1}X_r'\underbrace{M_sM_s}_{M_s}X_r(X_r'M_sX_r)^{-1} = \sigma^2(X_r'M_sX_r)^{-1}
\end{aligned}$$

- **Inversa de las dos varianzas**:

$$
\begin{aligned}
V(\widehat{\beta}_r)^{-1} - V(\widetilde{\beta}_r)^{-1} &= \sigma^{-2} X_r' X_r - \sigma^{-2} X_r' M_s X_r \\
&= \sigma^{-2} X_r' [I_n - M_s] X_r \\
&= \sigma^{-2} X_r' [X_s (X_s' X_s) X_s'] X_r .
\end{aligned}
$$

Esta matriz de diferencias será nula si $X_s' X_r = 0$.

Estimación del error de varianza

- Hemos visto que $\sigma^2 = E(u_i^2)$, por lo que un estimador insesgado de σ^2 es $n^{-1}\sum_{i=1}^{n} u_i^2$.
- Como los errores no son observables, no podemos estimar σ^2. Sin embargo, si podemos utilizar los residuos $\widehat{u}_i$ para obtener un estimador de σ^2

$$\widetilde{\sigma}^2 = \frac{SSR}{n} = \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i^2.$$

- $\widetilde{\sigma}^2$ es un estimador sesgado dado que $\widehat{u}_i$ satisfacen $k+1$ restricciones.
- El estimador insesgado de σ^2 ajustado por los grados de libertad $(n-k-1)$ es

$$\widehat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{SSR}{n-k-1},$$

$$\widehat{\sigma}^2 = \frac{\widehat{u}'\widehat{u}}{n-k} = \frac{\sum_{i=1}^{n} \widehat{u}_i^2}{n-k},$$

dado que si $M = (X'X)^{-1}X'$,

$$\widehat{u} = MY = M[X\beta + u] = Mu$$

y

$$\widehat{u}'\widehat{u} = (Mu)'Mu = u'M'Mu = u'Mu.$$

Valor esperado de la varianza estimada

Bajo los supuestos **RLM.1-5**,

$$E(\widehat{\sigma}^2) = E\left(\frac{\widehat{u}'\widehat{u}}{n-k}\right) = \sigma^2$$

dado que

$$\begin{aligned} E(\widehat{u}'\widehat{u}) &= E(u'Mu) = E(tr(u'Mu)) = E(tr(Muu')) \\ &= trE(Muu') = tr[ME(uu')] = tr\left[M(\sigma^2 I_n)\right] \\ &= tr\left[\sigma^2 M\right] = \sigma^2 trM = \sigma^2(n-k). \end{aligned}$$

Estimación del error de varianza

- Bajo **RLM.1-5**, $E(\widehat{\sigma}^2) = \sigma^2$.
- **Error estándar de la regresión**: $\widehat{\sigma} = \sqrt{\widehat{\sigma}^2}$.
- **Desviación estándar de** $\widehat{\beta}_j$:

$$sd(\widehat{\beta}_j) = \frac{\sigma}{[SST_j(1 - R_j^2)]^{1/2}}.$$

- **Error estándar de** $\widehat{\beta}_j$:

$$se(\widehat{\beta}_j) = \frac{\widehat{\sigma}}{[SST_j(1 - R_j^2)]^{1/2}}.$$

Teorema de Gauss-Markov

Bajo los supuestos **RLM.1-5**, $\widehat{\beta}_0, \widehat{\beta}_1, \cdots, \widehat{\beta}_k$ son los estimadores lineales insesgados óptimos de $\beta_0, \beta_1, \cdots, \beta_k$, respectivamente, son (**ELIOs**, también conocido como **BLUEs**), es decir, el

- **Estimador**
- **Lineal**
- **Insesgado**
- **Óptimo**.

Entonces, si los supuestos **RLM.1-5** se cumplen el estimador MCO es el mejor.