Análisis de Regresión Múltiple: Estimación

Carlos Velasco[1]

[1] Departamento de Economía
Universidad Carlos III de Madrid

Econometría I
Máster en Economía Industrial
Universidad Carlos III de Madrid
Curso 2007/08

Resumen

Limitaciones del Análisis de Regresión Simple

- Difícil extraer conclusiones *ceteris paribus* (justificando el supuesto RLS.3): mejor controlar más factores para hacer una análisis causal.
- Sólo consigue explicar una parte limitada de la variabilidad de y en función de una única x.
- Sólo puede incorporar una determinada relación funcional entre la x y la y (en función de x, $\log x$, etc.).

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

donde *expend* es el gasto por estudiante; *avgscore* es el resultado del test estandarizado en la escuela secundaria y *avginc* es la renta familiar media.

- El objeto de interés para efectos de evaluación de políticas es el efecto cp de *expend* sobre *avgscore*.
- Este modelo incluye *avginc* para controlar su efecto sobre *avgscore*.
- Esto es importante porque *avginc* está correlado positivamente con el gasto por estudiante (determinados por los impuestos locales sobre la renta y la propiedad).
- En RLS, *avginc* se incluiría en el término de error, que estaría correlado con *expend,* y por tanto el estimador de MCO estaría sesgado.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- β_0 es el término constante.
- β_1 mide el efecto sobre y de un cambio en x_1, manteniendo otros factores constantes.
- β_2 mide el efecto sobre y de un cambio en x_2, manteniendo otros factores constantes.

- Es posible que x_1 y x_2 estén relacionados perfectamente:

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

  pero no de forma lineal, donde u contiene otros factores que afectan al consumo aparte de la renta (*inc*).
- Renta es el único factor que afecta al *consumo*, y aparentemente un modelo de RLS es suficiente. Pero renta influye sobre *consumo* mediante dos funciones, $x_1 = inc$ y $x_2 = inc^2$.
- Mecánicamente se estima como un modelo de Regresión Múltiple habitual, pero la interpretación de los parámetros cambia: x_2 no puede estar constante si x_1 cambia.
- En este caso la propensión marginal a consumir es

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

- Para calcular efectos parciales es importante la definición de los regresores.

- Generalización de RLS.3:

$$E\left(u|x_1, x_2\right) = 0.$$

El valor esperado de u debe ser igual para todas las combinaciones de x_1 y x_2 : implica que los otros factores que afectan a y deben estar no relacionados (en media) con x_1 y x_2.

- EJEMPLO ecuación de salarios: los niveles medios habilidad deben ser iguales para todas las combinaciones de educación y experiencia en la población de trabajadores.
- EJEMPLO resultados estudiantes: otros factores que afecten a la nota (características del instituto o del estudiante), NO deben estar, en promedio, relacionados con la financiación ni con el ingreso familiar medio.
- EJEMPLO modelo cuadrático: en este caso

$$E\left(u|inc, inc^2\right) = E\left(u|inc\right) = 0.$$

Motivación para la regresión múltiple

Modelo con k variables independientes

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

- $k + 1$ parámetros.
- β_0 es el término constante.
- β_j mide el efecto sobre y de un cambio en x_j, manteniendo otros factores constantes (parámetros de pendiente).
- u : otros factores que afectan y y no son $x_1, x_2, \ldots, x_k$.

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 ceoten + \beta_3 ceoten^2 + u$$

- β_1 : elasticidad cp del salario respecto a *sales*.
- Si $\beta_3 = 0$: $100\beta_2$ es aproximadamente el porcentaje cp de incremento del salario cuando *ceoten* aumenta en una unidad.
- Si $\beta_3 \neq 0$, la interpretación es más compleja.
- La ecuación es lineal en los parámetros $\beta_0, \beta_1, \ldots, \beta_3$, pero no tiene porqué serlo en las variables: $\log(salary)$, $\log(sales)$, $ceoten^2$.

- Generalización de RLS.3:

$$E\left(u|x_1, x_2, \ldots, x_k\right) = 0.$$

- El valor esperado de u debe ser igual para todas las combinaciones de $x_1, x_2, \ldots, x_k$.
- También implica que se han especificado correctamente todas las relaciones funcionales entre las $x's$ y la y.
- Este supuesto garantiza que MCO está insesgado, y si se omite una variable de $x_1, x_2, \ldots, x_k$ ocasionará un sesgo.

Mecánica e Interpretación de Mínimos Cuadrados Ordinarios

- Ecuación estimada por MCO:

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

- MCO elige los valores de $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ que minimizan la suma de cuadrados de los residuos.
  Dadas n observaciones $\{(x_{i1}, x_{i2}, y_i), i = 1, \ldots, n\}$ se eligen $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ para hacer

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \right)^2$$

  tan pequeño como sea posible.

- x_{ij} : i -th observación de la variable j, $i = 1, \ldots, n$; $j = 1, \ldots, k$.
- x_{i1}, x_{i2} : 1,2, nombre de la variable.

$$x_{i1} = educ_i; \quad x_{i2} = exper_i.$$

- Ecuación estimada por MCO:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k.$$

- MCO elige los valores de $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ que minimizan la suma de cuadrados de los residuos.
  Dadas n observaciones $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i), i = 1, \ldots, n\}$ se eligen $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2 \ldots, \hat{\beta}_k$ para hacer

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik} \right)^2$$

  tan pequeño como sea posible.

Condiciones de primer orden:

$$
\begin{aligned}
\sum_{i=1}^{n}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}-\cdots-\hat{\beta}_k x_{ik}\right) &= 0\\
\sum_{i=1}^{n} x_{i1}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}-\cdots-\hat{\beta}_k x_{ik}\right) &= 0\\
\vdots \quad & \quad \vdots\\
\sum_{i=1}^{n} x_{ik}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}-\cdots-\hat{\beta}_k x_{ik}\right) &= 0.
\end{aligned}
$$

- Las condiciones de primer orden reproducen las condiciones

$$E(u) = 0, \quad E(x_j u) = 0, \;\; j = 1, \ldots, k,$$

que se obtienen a partir del supuesto $E(u|x_1, x_2, \ldots, x_k) = 0$.
- Además deben resolverse de forma única.
- Regresión mínimo cuadrática, función de regresión muestral:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k.$$

- $\hat{\beta}_0$: estimador MCO del término constante.
- $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$: estimadores MCO de las pendientes.

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

- $\hat{\beta}_0$: valor predicho de y cuando $x_1 = 0$ y $x_2 = 0$.
- $\hat{\beta}_1, \hat{\beta}_2$: interpretación ceteris paribus:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2.$$

- Si $\Delta x_2 = 0$ (mantenemos x_2 constante) entonces

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1.$$

- Si $\Delta x_1 = 0$, (mantenemos x_1 constante) entonces

$$\Delta\hat{y} = \hat{\beta}_2 \Delta x_2.$$

- Regresión Múltiple:

$$colGPA = \beta_0 + \beta_1 hsGPA + \beta_2 ACT + u$$

  - *colGPA*: college grade point average.
  - *hsGPA* : high school GPA.
  - *ACT* : achievement test score.
  - $n = 141$, `GPA1`

- Regresión simple:

$$colGPA = \gamma_0 + \gamma_1 ACT + v$$

No permite comparar estudiantes con el mismo *hsGPA*!!

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

- $\hat{\beta}_0$: valor predicho de y cuando $x_1 = 0, \ldots, x_k = 0$.
- $\hat{\beta}_j,\ j = 1, \ldots, k$: interpretación ceteris paribus:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \cdots + \hat{\beta}_k \Delta x_k.$$

- Si $\Delta x_2 = \cdots = \Delta x_k = 0$ (mantenemos $x_2, \ldots, x_k$ constantes) entonces

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1.$$

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

- $n = 526$, `WAGE1`.
- Variable dependiente en logaritmos: el salario aumenta en un $\left(\beta_j \cdot 100\right)$ % si $\Delta x_j = 1$

- El análisis de regresión múltiple proporciona interpretaciones *ceteris paribus* incluso si los datos no se han recogido de forma *ceteris paribus*.
- No se muestrea a personas con igual *hsGPA*, pero posiblemente con diferentes resultados *ACT*.
- Los datos se recogen de forma aleatoria de una población sin restricciones en los valores de *hsGPA* o de *ACT*.
- Regresión múltiple permite realizar en situaciones no experimentales los mismos análisis que hacen otros científicos en situaciones de laboratorio bajo control.

Valor ajustado para la observación i :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}.$$

Residuo para la observación i :

$$\begin{aligned} \hat{u}_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}. \end{aligned}$$

$$\bar{\hat{u}} = 0$$

$$\widehat{Cov}_n\left(u, x_j\right) = 0, \quad j = 1, \ldots, k$$

El punto

$$(\bar{x}_1, \bar{x}_2, , \ldots \bar{x}_k, \bar{y})$$

pertenece la recta de regresión MCO.

En principio no son necesarias las fórmulas para $\hat{\beta}_j$.
Caso $k = 2$:

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

donde $\hat{r}_{i1}$ son los residuos MCO de una regresión simple de x_1 sobre x_2, con la misma muestra.

1. Regresamos la variable x_1 sobre la variable x_2.
   Obtenemos los residuos $\hat{r}_1$ (la parte de x_1 que no puede ser explicada por x_2, incorrelada con x_2) : $\hat{r}_1$ es la parte de x_1 filtrada de x_2.
2. Regresamos y sobre los residuos $\hat{r}_1$, lo que nos dará el efecto parcial de x_1 sobre y, corregido por x_2.
   (Como $\hat{r}_{i1}$ tienen media cero, $\hat{\beta}_1$ es equivalente a una regresión habitual con constante).

Hay dos casos especiales en los que una regresión simple de y sobre x_1 produce el mismo estimador MCO para x_1 que una regresión múltiple sobre x_1 y x_2, es decir

$$\tilde{\beta}_1 = \hat{\beta}_1$$

en

$$\begin{aligned} \tilde{y} &= \tilde{\beta}_0 + \tilde{\beta}_1 x_1 \\ \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2. \end{aligned}$$

1. El efecto parcial de x_2 sobre y es cero en la muestra, es decir $\hat{\beta}_2 = 0$ [iguales condiciones de primer orden].
2. x_1 y x_2 están incorreladas en la muestra [el filtrado no tiene efecto].

Participación en Planes de Pensiones 401(k) (`401K`)

$$prate = \beta_0 + \beta_1 mrate + \beta_2 age + u$$

donde *prate* es la tasa de participación y *mrate* es la tasa *match*, que es la parte del plan que contribuye la empresa (la otra parte la cubre el trabajador). *age* es la edad del plan de pensiones.

¿Qué ocurre si no controlamos por *age*? Dependerá de la correlación entre *mrate* y *age*.

- **Suma de Cuadrados Totales**, SST:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 .$$

- **Suma de Cuadrados Explicada**, SSE:

$$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 .$$

- **Suma de Cuadrados de los Residuos**, SSR:

$$\sum_{i=1}^{n} \hat{u}_i^2 .$$

$$SST = SSE + SSR.$$

- Coeficiente de Determinación or R-cuadrado

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

- Propiedad, recordando que $\bar{\hat{y}} = \bar{y}$,

$$R^2 = \hat{\rho}^2_{y,\hat{y}} = \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{\left(\sum_{i=1}^n (y_i - \bar{y})^2\right)\left(\sum_{i=1}^n (\hat{y}_i - \bar{y})^2\right)}.$$

- R^2 nunca desciende cuando se añade una nueva variable en la regresión: no es un buen criterio para decidir si añadir una nueva variable al modelo.

- Ejemplo: College GPA (`GAP1`)

$$colGPA = \beta_0 + \beta_1 hsGPA + \beta_2 ACT + u.$$

- Ejemplo: Arrestos (`CRIME1`)

$$narr86 = \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 ptime86 + \beta_4 qemp86 + u$$

- A veces una teoría económica o el sentido común imponen $\beta_0 = 0$.
- Se quiere una función de regresión estimada de la forma

$$\tilde{y} = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \cdots + \tilde{\beta}_k x_k,$$

  por lo que si $x_1 = x_2 = \cdots = x_k = 0$ entonces $\tilde{y} = 0$.
- En este caso MCO minimiza

$$\sum_{i=1}^{n} \left(y_i - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_k x_{ik} \right)^2 .$$

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

RLM.1 (*Modelo lineal en parámetros*). En el modelo para la población, la variable dependiente y se relaciona con la variable independiente x y el error u mediante

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

donde β_0,$\beta_1, \ldots, \beta_k$ son parámetros desconocidos y u es una perturbación o error aleatorio.

Esta expresión se la conoce como el modelo verdadero (para la población) y es importante para interpretar los parámetros.

RLM.2 (*Muestreo aleatorio*). Para estimar los parámetros se dispone de una muestra de tamaño n, $\left(x_{i1}, \ldots, x_{ik,} y_i\right)$, $i = 1, \ldots, n$ del modelo poblacional,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i, \quad i = 1, \ldots, n.$$

Esta expresión es importante para deducir las propiedades de los EMCO de β_0,$\beta_1, \ldots, \beta_k$.

RLM.3 (*Media Condicional cero*).

$$E\left(u|x_1, x_2, \ldots, x_k\right) = 0.$$

- Este supuesto puede fallar si se especifica incorrectamente la forma funcional, por ejemplo se omite una variable que debía estar en el modelo o aparece en niveles cuando debía estar en logaritmos.
- También afecta a RLM.3 omitir un factor importante que está correlado con $x_1, x_2, \ldots, x_k$.
- Otros problemas son los errores de medida, y la simultaneidad (y se determina conjuntamente con alguna variable explicativa.).
- Si RLM.3 es cierta, se dice que tenemos **variables explicativas exógenas**, y en caso contrario **endógenas**.

RLM.4 (*No multicolinealidad perfecta*). En la muestra (y por tanto en la población), ninguna de las variables explicativas es constante, y no hay una relación lineal exacta entre las variables explicativas.

- Sólo afecta a las $x's$, por lo que es muy diferente de RLM.3.
- Es más compleja que RLS.4 porque afecta a las relación entre todas las variables explicativas.
- Si una variable independiente es una combinación lineal exacta de otras variables independientes, entonces hay multicolinealidad **perfecta**: no se puede hacer MCO.
- No impide que haya correlación entre las $x's$, sólo requiere que no sea perfecta.

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

- Esperamos que *expend* (gastos en educación) esté relacionado con *avginc* (renta familiar media), ya que los distritos con mayor nivel de renta tienden a gastar más dinero por alumno.
- La idea es precisamente incluir *avginc* precisamente porque estaría correlado con *expend*.
- Sin embargo, es extraño encontrar una muestra donde *expend* esté perfectamente correlado con *avginc*.

- Una variable es un múltiplo de otra: se incluyen dos variables sobre la misma magnitud en distintas unidades.
- Aunque funciones de un mismo regresor aparezcan en la ecuación eso no implica multicolinealidad: x y x^2 están perfectamente relacionadas, pero no perfectamente relacionadas *linealmente*.
- En cambio tendremos problemas en

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2 \log\left(inc^2\right) + u$$

pero no en

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2 [\log(inc)]^2 + u.$$

- También hay problemas en

$$voteA = \beta_0 + \beta_1 expendA + \beta_2 expendB + \beta_3 totexpend + u$$

porque $x_3 = x_1 + x_2$.

- En todos estos casos no se pueden calcular los estimadores MCO.
- Solución: especificar con cuidado el modelo: eliminar la(s) variable(s) redundante(s).
- Otra causa: el tamaño muestral n es demasiado pequeño, $n < k + 1$.

Valores esperados de los estimadores MCO

Teorema 3.1 (Insesgadez de MCO). Bajo los supuestos RLM.1.-5

$$E\left(\hat{\beta}_j\right) = \beta_j, \quad j = 0, \ldots, k,$$

para cualquier valor de los parámetros poblacionales β_j.

- El supuesto clave es RLM.3
- La propiedad de insesgadez no dice nada sobre el resultado de la estimación en un caso particular, si no que es una propiedad del *método* de estimación MCO.

- O **sobre-especificación** del modelo de regresión múltiple:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

con los supuestos RLM.1-4.
- Sin embargo x_3 no tiene efecto sobre y, una vez que se consideran x_1 y x_2, por lo que $\beta_3 = 0$,

$$E(y|x_1, x_2, x_3) = E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

- No se sabe que $\beta_3 = 0$: se incluye x_3 en la ecuación estimada,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3.$$

- <u>¿Cuál es el efecto sobre la insesgadez de $\hat{\beta}_1$ y $\hat{\beta}_2$?</u> NINGUNO. El Teorema 3.1 es válido para cualquier valor de β_j, incluyendo $\beta_j = 0$. Además $\mathrm{E}\left(\hat{\beta}_3\right) = \beta_3 = 0$.
- <u>¿Cuál es el efecto entonces sobre la estimación MCO?</u>

Ahora se omite una variable que sí que pertenece al modelo poblacional: se **excluye una variable relevante** o **infra-estimación del modelo:** caso particular de **error de especificación.**

► Esto conllevará un **sesgo** en los estimadores MCO.

- Modelo verdadero:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

  y se satisfacen RLM.1-4.
- Suponemos que el objeto de interés es β_1, el efecto parcial de x_1 sobre y.
- Para obtener estimadores insesgados deberíamos incluir x_2 en la regresión, pero no lo hacemos,

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 .$$

- Modelo verdadero:

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

- Modelo estimado:

$$wage = \beta_0 + \beta_1 educ + v,$$

donde $v = \beta_2 abil + u$, y el estimador de β_1 en este modelo es $\tilde{\beta}_1$.

- Deducción del valor esperado de $\tilde{\beta}_1$ (regresión simple):

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)\, y_i}{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2}.$$

- Ahora sustituimos

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

y obtenemos

$$\begin{aligned} & \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)\,(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i) \\ = \quad & \beta_1 \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2 + \beta_2 \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)\, x_{i2} + \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)\, u_i. \end{aligned}$$

Valores esperados de los estimadores MCO

Sesgo por omisión de variables: el caso simple (2)

- Por tanto, usando que $E(u_i) = 0$,

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)\, x_{i2}}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2},$$

  por lo que en general $\tilde{\beta}_1$ está sesgado para estimar β_1.
- El cociente que multiplica β_2 es el coeficiente de la pendiente de la regresión de x_2 sobre x_1, usando la misma muestra:

$$\tilde{x}_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1,$$

  y por tanto

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_2 \tilde{\delta}_1$$

  y **el sesgo de variables omitidas** es:

$$\text{Sesgo}\left(\tilde{\beta}_1\right) = E\left(\tilde{\beta}_1\right) - \beta_1 = \beta_2 \tilde{\delta}_1.$$

Hay dos casos en los que $\tilde{\beta}_1$ es **insesgado**:

1. Si $\beta_2 = 0$, por lo que x_2 no aparece en la ecuación.
2. Si $\tilde{\delta}_1 = 0$, es decir la covarianza muestral entre x_1 y x_2 es cero. Ya sabemos que si x_1 y x_2 están incorrelados muestralmente, $\hat{\beta}_1 = \tilde{\beta}_1$, y $\hat{\beta}_1$ es insesgado.
[Sin condicionar en x_{i2} $\tilde{\beta}_1$ es insesgado si $E(x_2|x_1) = E(x_2)$.]

Signo del sesgo:

$$\text{Sesgo}\left(\tilde{\beta}_1\right) = E\left(\tilde{\beta}_1\right) - \beta_1 = \beta_2 \tilde{\delta}_1 .$$

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	sesgo positivo	sesgo negativo
$\beta_2 < 0$	sesgo negativo	sesgo positivo

En general es posible tener idea del signo de β_2, aunque a veces es más difícil conocer el de $\text{Corr}(x_1, x_2)$.

- Sesgo positivo, $E\left(\tilde{\beta}_1\right) > \beta_1$, sesgo hacia arriba, sobre-estima en media.
- Sesgo negativo, $E\left(\tilde{\beta}_1\right) < \beta_1$, sesgo hacia abajo, infra-estima en media.

`WAGE1`, $n = 526$

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 abil + u.$$

Si omitimos *abil* porque no tenemos datos:

- Podemos esperar que $\beta_2 > 0$.
- Y que $\mathrm{Corr}(x_1, x_2) > 0$.
- Por lo que el sesgo será positivo: $\tilde{\beta}_1$ sobreestimará en media el efecto parcial de educación (porque también incluye el efecto parcial de *abil*).

$$avgscore = \beta_0 + \beta_1\ expend + \beta_2 povrate + u.$$

Si no tenemos datos sobre la tasa de pobreza, *povrate*,

- Podemos esperar que $\beta_2 < 0$.
- Y que $\text{Corr}(x_1, x_2) < 0$.
- Por lo que el sesgo será positivo.
- Puede que $\beta_1 = 0$, pero en una regresión simple $\tilde{\beta}_1$ tendrá valores en media positivos y se concluirá que el efecto del gasto es positivo.

- Obtener el signo del sesgo cuando hay múltiples regresores es más complicado.
- Si se omite un sólo regresor, resulta en que todos los estimadores MCO están sesgados.
- Modelo verdadero:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

  y se satisfacen RLM.1-4.
- Se omite x_3, y se ajusta

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

- x_2 y x_3 están incorrelados, pero x_1 está correlado con x_3.
- En general $\tilde{\beta}_1$ y $\tilde{\beta}_2$ estarán sesgados (excepto cuando x_1 y x_2 están incorrelados también).

Valores esperados de los estimadores MCO

Sesgo por omisión de variables: casos más generales (2)

Si x_1 y x_2 están incorrelados, se puede estudiar el sesgo como si x_2 estuviese fuera de la regresión:

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_3 \frac{\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right) x_{i3}}{\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)^2},$$

por lo que en general $\tilde{\beta}_1$ está sesgado para estimar β_1.

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + u.$$

Si omitimos *abil*:

- Los EMCO de β_1 y β_2 cuando se omite *abil* estarán sesgados.
- Asumimos que *exper* y *abil* están incorrelados y además, como aproximación, que *educ* y *exper* también están incorrelados.
- Como $\beta_3 > 0$ y *educ* y *abil* están correlados positivamente, entonces $\tilde{\beta}_1$ tendría un sesgo positivo, al igual que cuando *exper* no está en el modelo.

Varianza de los Estimadores MCO

- Además de saber que la distribución muestral de $\hat{\beta}_j$ está centrada, es importante saber su variabilidad alrededor de β_j.
- Además de RLM.1-4, añadimos un nuevo supuesto para simplificar los cálculos y obtener resultados de optimalidad.

RLM.5 (*Homocedasticidad Condicional*): u tiene varianza, condicional en $x_1, \ldots, x_k$ constante,

$$Var\left(u|x_1, \ldots, x_k\right) = \sigma^2.$$

- Si no se cumple, estamos en presencia de **heteroscedasticidad**.

Varianza de los Estimadores MCO

- Los supuestos RLM.1-5 se denominan los supuestos de **Gauss-Markov** [para datos de sección cruzada].
- RLM.3 y RLM.5 se pueden escribir como, $\mathbf{x} = (x_1, \ldots, x_k)$,

$$\begin{aligned} E(y|\mathbf{x}) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \\ Var(u|\mathbf{x}) &= \sigma^2 \end{aligned}$$

- Es, junto con RLM.3, más débil que el supuesto de independencia.
- σ^2 se le llama la varianza del error o perturbación.

Teorema 3.2 (*Varianzas muestrales de los estimadores MCO de las pendientes*). Bajo supuestos RLM.1-5, condicional en los valores de las variables independientes,

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}, \quad j = 1, \ldots, k$$

donde

$$SST_j = \sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)^2$$

es la variación total de x_j, y R_j^2 es el R-cuadrado de la regresión de x_j sobre todas las demás variables independientes (y el término constante).

- **Varianza del error,** σ^2 : cuanto más *ruido*, mayor $Var\left(\hat{\beta}_j\right)$.
  - Es una propiedad de la población, no depende de n, y es desconocido.
  - La única forma de reducirlo es introducir más variables explicativas, pero eso no es siempre deseable ni posible.
- **Variación total de** x_j, $\mathbf{SST}_j$: cuanto mayor, menor $Var\left(\hat{\beta}_j\right)$.
  - Es difícil incrementarla, a parte de aumentar n.
  - $\mathrm{SST}_j = 0$ no está permitido por RLM.4.

Relación lineal entre las variables independientes

- R_j^2 : cuanto mayor, más alto es $Var\left(\hat{\beta}_j\right)$. No aparece en regresión simple.
  Si R_j^2 está próximo a 1, quiere decir que x_j se puede explicar en gran parte por las otras variables independientes. El caso extremo $R_j^2 = 1$ no está permitido por RLM.4.
  La alta correlación entre regresores se denomina **multicolinealidad.**
- La **Multicolinealidad** aumenta la **varianza** de los EMCO, como lo hace una muestra pequeña, todo lo demás igual.
- Generalmente no hay más remedio que recoger más datos, porque eliminar variables conlleva introducir **sesgos.**

Varianza de los estimadores MCO (5)

Componentes de la Varianza (3)

- Si hay muchos regresores correlados midiendo el mismo efecto (diversas clases de gasto en estudiantes), entonces se estimará el efecto de cada uno con alta varianza inevitablemente.
- Por otro lado, si en

  $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

  x_2 y x_3 están muy correlados, entonces $Var\left(\hat{\beta}_2\right)$ y $Var\left(\hat{\beta}_3\right)$ serán altas.
- Pero la correlación entre x_2 y x_3 no tendrá efecto directo sobre $Var\left(\hat{\beta}_1\right)$:
  - Si x_1 está incorrelado con x_2 y x_3, entonces $Var\left(\hat{\beta}_1\right) = \sigma^2 / SST_1$, igual que en regresión simple.
  - Introducir muchos controles no tiene porqué afectar a la estimación del parámetro de interés.

Varianza de los estimadores MCO (5)

Componentes de la Varianza (3)

- Si hay muchos regresores correlados midiendo el mismo efecto (diversas clases de gasto en estudiantes), entonces se estimará el efecto de cada uno con alta varianza inevitablemente.
- Por otro lado, si en

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

  x_2 y x_3 están muy correlados, entonces $Var\left(\hat{\beta}_2\right)$ y $Var\left(\hat{\beta}_3\right)$ serán altas.
- Pero la correlación entre x_2 y x_3 no tendrá efecto directo sobre $Var\left(\hat{\beta}_1\right)$:
  - Si x_1 está incorrelado con x_2 y x_3, entonces $Var\left(\hat{\beta}_1\right) = \sigma^2 / SST_1$, igual que en regresión simple.
  - Introducir muchos controles no tiene porqué afectar a la estimación del parámetro de interés.

Hay un tradeoff entre **sesgo y varianza** a la hora de decidir si incluir una variable o no en una regresión.

Modelo verdadero, con supuestos Gauss-Markov,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

Consideramos dos estimadores:

- $\hat{\beta}_1$, en regresión múltiple,

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

- $\tilde{\beta}_1$, en regresión simple,

$$\tilde{y} = \beta_0 + \tilde{\beta}_1 x_1.$$

- Si $\beta_2 \neq 0$, $\tilde{\beta}_1$ estará **sesgado**, por lo que desde el punto de vista del sesgo, $\hat{\beta}_1$ será preferido.

Otro criterio más: **varianza**:

$$
\begin{aligned}
Var\left(\hat{\beta}_1\right) &= \frac{\sigma^2}{SST_1\left(1-R_1^2\right)} \\
Var\left(\tilde{\beta}_1\right) &= \frac{\sigma^2}{SST_1}
\end{aligned}
$$

por lo que

$$
Var\left(\tilde{\beta}_1\right) < Var\left(\hat{\beta}_1\right),
$$

a menos que x_1 y x_2 estén incorrelados en la muestra (igual var).

Varianza de los estimadores MCO (8)

Varianzas en modelos mal especificados (3)

- Si x_1 y x_2 NO están incorrelados:

1. Si $\beta_2 \neq 0$, $\tilde{\beta}_1$ es sesgado, $\hat{\beta}_1$ es insesgado y $Var\left(\tilde{\beta}_1\right) < Var\left(\hat{\beta}_1\right)$.
2. Si $\beta_2 = 0$, $\tilde{\beta}_1$ y $\hat{\beta}_1$ son insesgados y $Var\left(\tilde{\beta}_1\right) < Var\left(\hat{\beta}_1\right)$.

En caso (2) preferimos $\tilde{\beta}_1$: introducimos una variable irrelevante.
El caso (1) es más complicado: se suele preferir $\hat{\beta}_1$:

- Se puede calcular el tamaño del sesgo, pero este no se reduce al aumentar n, al contrario que $Var\left(\hat{\beta}_1\right)$.
- Además el cálculo condicional en x_2 no tiene en cuenta que el tamaño del error en el modelo corto será mayor ($\sigma^2 \uparrow$).

Varianza de los estimadores MCO (9)

Estimación de σ^2

- $\sigma^2 = E\left(u^2\right)$, 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, un primer estimador sería $n^{-1}SSR = n^{-1}\sum_{i=1}^{n}\hat{u}_i^2$, pero este no es insesgado.
- Los residuos $\hat{u}_i$ satisfacen $k+1$ restricciones.
- Por tanto los residuos tienen $n-k-1$ **grados de libertad**:

  $n-k-1 =$ núm. de observaciones $-$ núm. de parámetros estimados.
- El estimador insesgado de σ^2 que hace el ajuste por los grados de libertad es

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

Teorema 3.3. (*Estimación insesgada de* σ^2). Bajo las condiciones de Gauss-Markov RLM.1.5,

$$E\left(\hat{\sigma}^2\right) = \sigma^2.$$

- Estimador de σ, o **error estándar de regresión**, o **SER**,

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}.$$

- $\hat{\sigma}$ puede aumentar o disminuir cuando se incluye una variable adicional en la regresión.

- **Desviación estándar** de $\hat{\beta}_j$:

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

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

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

- Sólo es válida en presencia de RLM.5 (hay que cambiar $sd\left(\hat{\beta}_j\right)$ y $se\left(\hat{\beta}_j\right)$).

Eficiencia de MCO: Teorema de Gauss-Markov

- Justifica la utilización de MCO sobre otros métodos de estimación.
- Una justificación de MCO: es **insesgado** bajo RLM.1-4.
- Si limitamos la clase de estimadores competidores, entonces MCO es el mejor.
- Bajo RLM.1-5, el estimador $\hat{\beta}_j$ de β_j será el estimador lineal insesgado óptimo (**ELIO**, BLUE).
- **Estimador:** regla que se aplica a una muestra para producir una estimación.
- **Lineal**:

$$\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$$

  donde w_{ij} puede ser una función de los valores muestrales de todas la variables independientes.
- **Insesgado:** $E\left(\hat{\beta}_j\right) = \beta_j$.
- **Óptimo**: de menor varianza.

Teorema de Gauss-Markov

Teorema 3.4. (*Teorema de Gauss-Markov*). Bajo los supuestos RLM.1-5, $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ son los estimadores lineales insesgados óptimos (**ELIOs**) de $\beta_0, \beta_1, \ldots, \beta_k$, respectivamente.

Si los supuestos habituales se cumplen en un problema dado, no necesitamos considerar ningún estimador insesgado de la forma $\tilde{\beta}_j$, MCO es el mejor.

Si algún supuesto falla, entonces G-M no se cumple:

- Si falla RLM.3, entonces los EMCO dejan de ser insesgados.
- Si falla RLM.5, entonces hay estimadores lineales insesgados con menor varianza que MCO en presencia de heterocedasticidad (MCG).