Notas de Clase: Econometría

Clase: Regresión Múltiple
Fecha: 24 Abril de 2023
Profesor Ricardo Pasquini. FCE Universidad Austral

Regresión Múltiple

Motivación e interpretación

- Nos interesa agregar al modelo una cantidad arbitraria de variables explicativas. Por eso vamos a estudiar los modelos del tipo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon \tag{1}$$

  - Donde $x_1, x_2, \ldots, x_k$ representan k variables explicativas.
- Para la interpretación tomemos un ejemplo. Anteriormente estudiamos el modelo del ingreso (ing) de las personas en base a la los años de educación ($educ$). Ahora incorporamos adicionalmente la experiencia (exp) :

$$ing = \beta_0 + \beta_1 educ + \beta_2 exp + \epsilon \tag{2}$$

  - Interpretamos β_1 : este coeficiente mide el efecto de un año de educación, una vez que *ya tenemos en cuenta* el efecto de los años de experiencia. También podemos interpretar como "manteniendo la experiencia constante". En la jerga también se dice "controlando por experiencia".
  - Comparemos contra el modelo sin experiencia ($ing = \beta_0 + \beta_1 educ + \epsilon$) : En ese caso β_1 capturaba el efecto de un año extra de educación . Pero qué pasaría si, por ejemplo, la gente que se educa más siempre tiene más experiencia? Ya que un año extra de educación también implica mayor experiencia, entonces lo que capturaría β_1 entonces es el efecto del año extra de educación pero conteniendo el efecto de la experiencia.
  - Por lo tanto, la diferencia con el modelo extendido es que nos permite aislar el efecto propio de la educación cuando la experiencia está constante. En palabras de Wooldridge:
  - The power of multiple regression analysis is that it allows us to do in nonexperimental environments what natural scientists are able to do in a controlled laboratory setting: keep other factors fixed.

Estimador y Propiedades Estadísticas

- Las estimaciones del modelo de los parámetros (i.e., $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$) se obtienen a partir de buscar los valores que minimizan la suma de los errores estimados al cuadrado. Los estimadores que surgen de este método se conocen como estimadores de Mínimos Cuadrados Ordinarios o Ordinary Least Squares (OLS).
- ¿Qué sabemos sobre las propiedades estadísticas de estos estimadores?

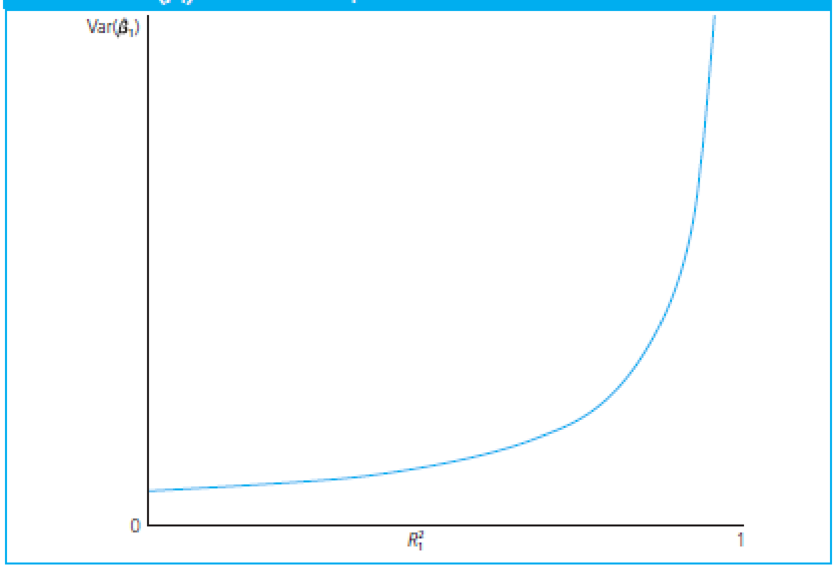
1. Sabemos que, al igual que en el caso de Regresion Simple, estos estimadores son *insesgados* (i.e., podemos confiar que su valor esperado coincide con el valor poblacional).
2. Sabemos que la varianza toma la siguiente forma (no lo demostramos):

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)} \tag{3}$$

   donde:

   - σ^2 es la varianza del error (n.b., que es desconocida pero estimable como ya vimos anteriormente)
   - SST_j es la suma de los cuadrados totales de la j-esima variable explicativa ($SST_j = \sum_{i=1}^{n}(x_{i,j} - \bar{x}_j)^2$). La suma de los cuadrados totales es una media de varianza, en este caso de la variable j-esima.
   - R_j^2 [Atención que esto es algo nuevo] es una medida del tipo R^2 de bondad de ajuste que ya vimos. Pero en este caso, el subindice j se usa para indicar que esta medida es el R^2 de otra regresión (no la que estamos analizando). R_j^2 se refiere a la bondad de ajuste del modelo del modelo que busca explicar x_j en función del resto de las variables explicativas del modelo. Es decir, es una medida de cuánto de la variabilidad de la j-esima variable explicativa puede explicarse por el resto de las $(k-1)$ variables que incorporamos al modelo.
   - Notar que la definición de varianza implica: i) que la varianza del estimador será mayor si el modelo en general tiene más error (más varianza en el error), ii) que tener más variabilidad en la variable explicativa jesima es bueno, ya que reduce la varianza del estimador del coeficiente j-esimo. Esto se puede intuir como que mayor variabilidad en la variable explicativa permite (todo lo demás constante) captar más variabilidad en la variable a explicar, y iii) que mientras más se pueda explicar la variable jesima con las demas explicativas, peor es para la variabilidad del coeficiente. Esto se puede interpretar como que la variable jesima sería más redundante, y a razón de esto el modelo no puede asignarle el efecto adecuado.
   - En relación al último efecto (el de la correlación entre las variables explicativas), Wooldridge agrega una gráfico ilustrativo para ver el efecto de este efecto sobre la varianza (véase a continuación). Nótese que cuando la relación es muy alta, el modelo no puede identificar el efecto de x_j ya que la varianza se dispara:

FIGURE 3.1 $Var(\hat{\beta}_1)$ as a function of R_1^2.



Inclusión de variables irrelevantes

Incluir una variable irrelevante no implicará a un sesgo en la estimación.

- Supongamos que el modelo poblacional (el generador de los datos) es:

$$y = \beta_0 + \beta_1 x_1 + \epsilon \tag{4}$$

- Pero a la hora de estimar el modelo, agrego una variable x_2 como explicativa y lo estimo:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \epsilon \tag{5}$$

- Sin embargo, notemos que, en realidad, lo que se hizo no implica que se asumió un modelo incorrecto, ya que el modelo poblacional es equivalente al modelo estimado pero cuando el coeficiente de efecto de x_2 es 0:

$$y = \beta_0 + \beta_1 x_1 + 0 x_2 + \epsilon \tag{6}$$

- De esta manera, no hay inconsistencia entre el modelo poblacional y el estimado, y al estimar el modelo extendido, deberíamos esperar que el valor de $\hat{\beta}_2$ sea cercano a 0.
- Aplican los resultados estadísticos que ya vimos, que asumen que el modelo estimado es el correcto, entre ellos que los estimadores son insesgados.

La inclusión de una variable irrelevante implica una pérdida de eficiencia en la estimación

- Si agregamos variables de más a un modelo, la varianza de los estimadores aumentará.
  - Este efecto se da porque, si aumentamos una variable explicativa, por definición aumenta la capacidad que esa variable explique a las otras variables que ya están en el modelo (i.e., aumenta el valor de R_j^2)
    - Recuerden que $Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}$
  - Si la varianza aumenta, aumenta el error estándar.
  - Si aumenta el error estándar es menos probable que rechacemos una hipótesis (el valor del estadístico T del test de hipótesis se reduce cuando el error estándar aumenta)

Omisión de variables relevantes

La omisión de una variable relevante implica un sesgo en los coeficientes estimados

- Supongamos que el modelo poblacional (el generador de los datos) es:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \tag{7}$$

- Pero a la hora de estimar el modelo no cuento con x_2 como explicativa o no la tengo en cuenta:

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \epsilon \tag{8}$$

donde estamos usando la notación con el tilde, sólo para señalar que estamos estimando un modelo distinto del poblacional.

- Se puede demostrar que la estimación estará sesgada. Es relativamente fácil demostrar que:

$$E[\tilde{\beta_1}] = \beta_1 + \beta_2 \delta_1 \tag{9}$$

  donde δ_1 es el coeficiente de la pendiente de una regresión de x_2 sobre x_1. (i.e., una regresión del tipo $x_2 = \delta_0 + \delta_1 x_1 + \epsilon$)

  - Es decir, el coeficiente para β_1 que se espera de la estimación que omite x_2 no le pega al valor de β_1 en términos esperados, sino que está sesgado por la magnitud de $\beta_2 \delta_1$.
  - Nótese que la magnitud del sesgo se anularía si:
    - i) x_2 no tiene efecto en y (i.e., en la población $\beta_2 = 0$), o bien,
    - ii) x_2 no tiene relación con x_1. (i.e., $\delta_1 = 0$.)