
Economía Aplicada: Problem Set N°3

Milton Bronstein

Felipe García Vassallo

Santiago López

Franco Riottini

1. Ejercicio 1

En este ejercicio realizaremos distintas simulaciones para observar cambios sobre resultados de distintas regresiones

1.1. Aumento de la muestra

| | (1) | (2) |
|--------------|-----------------------|-----------------------|
| | wage | wage |
| intelligence | 2.990*** (0.00510) | 2.999*** (0.00176) |
| a | 1.063*** (0.0536) | 1.020*** (0.0171) |
| b | 2.023*** (0.104) | 1.979*** (0.0327) |
| _cons | 6.799*** (0.883) | 6.500*** (0.282) |
| Observations | 100 | 1000 |
| R-squared | 1.000 | 1.000 |

Standard errors in parentheses

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

En la Tabla 1, podemos ver que cuando aumentamos el tamaño de la muestra los errores estándar disminuyen porque en el denominador de la varianza del coeficiente de la variable explicativa se encuentra el tamaño de la muestra: n .

1.2. Aumento de la varianza del término de error

| | (1) | (2) |
|--------------|-----------------------|----------------------|
| | wage | wage |
| intelligence | 2.990*** (0.00510) | 2.991*** (0.0229) |
| a | 1.063*** (0.0536) | 1.106*** (0.240) |
| b | 2.023*** (0.104) | 0.961* (0.465) |
| _cons | 6.799*** (0.883) | 11.70** (3.959) |
| Observations | 100 | 100 |
| R-squared | 1.000 | 0.994 |

Standard errors in parentheses
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

En este caso, aumentar la varianza del término de error aumenta los errores estándar de los regresores, ya que en el numerador de la varianza del coeficiente de los regresores está la varianza del término de error.

1.3. Aumento de la varianza de una variable explicativa

| | (1) | (2) |
|--------------|-----------------------|-----------------------|
| | wage | wage |
| intelligence | 2.990*** (0.00510) | 1.201*** (0.00942) |
| a | 1.063*** (0.0536) | 1.063*** (0.246) |
| b | 2.023*** (0.104) | 1.035* (0.477) |
| _cons | 6.799*** (0.883) | 190.0*** (3.316) |
| Observations | 100 | 100 |
| R-squared | 1.000 | 0.994 |

Standard errors in parentheses
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Para este ejercicio, aumentamos la varianza de la variable explicativa inteligencia y observamos que al hacerlo el error estándar naturalmente aumenta. También observamos que aumentan los errores estándar de los otros dos regresores.

1.4. El valor de los residuos

Por otro lado, partiendo del mismo ejercicio del inciso anterior, el valor de los residuales aumenta porque el modelo es más impreciso. Pasan de un valor de 100,95 a 2029,84.

Cuadro 1: Valor residuos sin cambios en la varianza

| Source | SS | df | MS | Number of obs | = | 100 |
|----------|-------------------|-----------|-------------------|---------------|---|-----------------|
| | | | | F(3, 96) | > | 99999.00 |
| Model | 363047.562 | 3 | 121015.854 | Prob >F | = | 0.0000 |
| Residual | 100.947757 | 96 | 1.05153914 | R-squared | = | 0.0000 |
| | | | | Adj R-squared | = | 0.9997 |
| Total | 363148.51 | 99 | 3668.16677 | Root MSE | = | 1.0254 |

Cuadro 2: Valor residuos con cambios en la varianza

| Source | SS | df | MS | Number of obs | 100 | 100 |
|----------|-------------------|-----------|-------------------|---------------|-----|----------------|
| | | | | F(3, 96) | > | 5743.26 |
| Model | 364309.55 | 3 | 121436.517 | Prob >F | = | 0.0000 |
| Residual | 2029.83987 | 96 | 21.1441653 | R-squared | = | 0.9945 |
| | | | | Adj R-squared | = | 0.9943 |
| Total | 366339.39 | 99 | 3700.39788 | Root MSE | = | 4.5983 |

1.5. Ortogonalidad de los residuos

Por construcción, el método de mínimos cuadrados ordinarios genera residuos que son ortogonales a los regresores.

1.6. Prediciendo Y con multicolinealidad

La alta multicolinealidad de ciertas variables no afecta la predicción de Y (en Stata realizamos un pequeño ejercicio para mostrar esto) no afecta la predicción de la variable explicada. Solamente pierde potencia el estimador de los coeficientes multicolineales.

1.7. Error no aleatorio en X

Naturalmente, un error de medición sistemático sesga el estimador en una dirección específica, en este caso hacia abajo, y aunque sigue siendo significativo, aumenta el error estándar. En el caso de los otros regresores, también están sesgados, pero se vuelven estadísticamente no significativos.

| | (1) | (2) |
|--------------|-----------------------|----------------------|
| | wage | wage |
| intelligence | 2.990*** (0.00510) | 1.149*** (0.0253) |
| a | 1.063*** (0.0536) | 1.227 (0.678) |
| b | 2.023*** (0.104) | -0.00465 (1.311) |
| _cons | 6.799*** (0.883) | 196.9*** (9.102) |
| Observations | 100 | 100 |
| R-squared | 1.000 | 0.956 |

Standard errors in parentheses
* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

1.8. Error aleatorio en X

La presencia de un error aleatorio en los regresores sesga los regresores apenas, en general hacia abajo excepto en el caso de a , y continúan siendo todos estadísticamente significativos.

| | (1) | (2) |
|--------------|-----------------------|----------------------|
| | wage | wage |
| intelligence | 2.990*** (0.00510) | 2.978*** (0.0246) |
| a | 1.063*** (0.0536) | 1.096*** (0.259) |
| b | 2.023*** (0.104) | 1.236* (0.503) |
| _cons | 6.799*** (0.883) | -7.411 (4.380) |
| Observations | 100 | 100 |
| R-squared | 1.000 | 0.994 |

Standard errors in parentheses
* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

1.9. Error aleatorio y no aleatorio en Y

En la columna 1 de la Tabla 8 se observa la regresión de la columna 1 de la Tabla 1 nuevamente con errores aleatorios en la variable explicada: *wage*. En la columna 2 de la Tabla 8, en cambio, observamos la misma regresión con errores no aleatorios en *wage*. Vemos que el coeficiente estimado de nuestro regresor *intelligence* es similar y se mantiene significativo al 1%. Su error estándar aumenta. En cambio, las estimaciones para a y b dejan de ser significativas. Sus errores estándar también aumentan, al igual que los de la estimación realizada para *intelligence*.

| | (1) | (2) |
|--------------|-----------------------|----------------------|
| | wage | wage |
| intelligence | 2.990*** (0.00510) | 2.992*** (0.0535) |
| a | 1.063*** (0.0536) | 0.907 (0.563) |
| b | 2.023*** (0.104) | 1.929 (1.092) |
| _cons | 6.799*** (0.883) | -9.085 (9.514) |
| Observations | 100 | 100 |
| R-squared | 1.000 | 0.970 |

Standard errors in parentheses
* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

2. Ejercicio 2

A lo largo de este ejercicio tendremos siempre presente la siguiente fórmula, en forma reducida, para el sesgo:

$$E[\hat{\beta}_1] = \beta_1 + \delta_{21}\beta_2$$

donde δ_{21} es el coeficiente de la regresión de X_2 en X_1 . Esta fórmula nos dice que aparecerá un sesgo si y solo si X_1 y X_2 están correlacionadas y además el coeficiente β_2 es significativo estadísticamente.

2.1.

Como decíamos, el sesgo de un coeficiente surge cuando las variables omitidas tienen un efecto sobre la variable explicada y, además, están correlacionadas con los otros regresores. En nuestro caso, como las variables de interés están altamente correlacionadas, sabemos que $\hat{\beta}_1$ y $\tilde{\beta}_1$ serán diferentes.

2.2.

Si X_2 y X_3 no tienen ninguna correlación con X_1 , $\tilde{\beta}_1$ y $\hat{\beta}_1$ no serán distintos.

2.3.

Si asumimos que el modelo presentado al inicio del ejercicio es el verdadero modelo del Data Generating Process, añadir un variable irrelevante no es fuente de sesgo (por lo tanto $\tilde{\beta}_1$ y $\hat{\beta}_1$ serán iguales) pero si generará estimadores más ineficientes.

Asumiendo que no conocemos el DGP, si corriésemos una regresión con un cuarto regresor que es el consumo de chocolate del alumno, no esperaríamos que $\tilde{\beta}_1$ y $\hat{\beta}_1$ sean distintos, ya que resultaría extraño que el consumo de chocolate esté correlacionado con la asistencia a clase. De todas formas, generaría estimadores más ineficientes, por lo que no debería ser incluida.

2.4.

La magnitud del sesgo sería menor que en el punto 1.1, pero igual veríamos resultados distintos entre $\tilde{\beta}_1$ y $\hat{\beta}_1$.

2.5.

Dada

$$V(\hat{\beta}_j) = \frac{\sigma^2}{n(1 - R_j^2)V(X_j)}$$

Los errores estándar de $\hat{\beta}_1$ serán menores que los de $\tilde{\beta}_1$, debido a que estamos en una situación en donde X_2 y X_3 son relevantes para explicar a Y . En este caso, la varianza del $\hat{\beta}_1$ va a disminuir al incluir estas dos variables explicativas porque depuramos el término de error y le quitamos varianza (disminuye σ^2 en la formula de arriba). . De todas formas, también aumenta el R_1^2 al incluir variables porque en la práctica la ortogonalidad perfecta no existe. Igualmente, basándonos en la consigna, prevalece el efecto de menor varianza en el error.

2.6.

Asumimos que el consumo de chocolate no es para nada relevante para explicar la nota en un examen de matemática y que tampoco tiene correlación con la variable de interés (pero que en la realidad no puede ser cero, siempre algún valor \neq a 0 va a tomar). Por lo tanto, el error estándar de $\hat{\beta}_1$ será mayor que el de $\tilde{\beta}_1$, dado que el término R_j^2 , la parte explicada de X_1 por las otras variables explicativas, va a aumentar.