## Problem set 4

Duarte Baracat, Agustín; Oviedo, Nicolás; Sánchez, Guido; Sosa, Juan Bautista

Economía Aplicada 2020 - UdeSA

## **Ejercicio 1**

Generamos variables de una manera muy similar a lo visto en clase. Luego, regresamos la variable *wage* en *intelligence, education*, *a* y *b*. Los resultados de la regresión se presentan en la primera columna de la Tabla 1. Al haber agregado *education*, hemos agregado una variable irrelevante para el modelo. A continuación, mostramos cómo aumentar el tamaño de la muestra, pasando de 100 observaciones a 1000, disminuye los errores estándar para todos los regresores, como se ve en la segunda columna de dicha tabla. La cuarta columna de la Tabla 1 contiene los verdaderos valores de los parámetros.

#### Tabla 1 Regresiones con Multicolinealidad y Verdaderos Valores

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| VARIABLES | wage | wage | wage | *true values* |
|  |  |  |  |  |
| intelligence | 2.0039\*\*\* | 1.9967\*\*\* | 2.1048\*\*\* | *2* |
|  | (0.0060) | (0.0023) | (0.1152) |  |
| education | -0.0225 | 0.0145 | -0.4896 | *-* |
|  | (0.0288) | (0.0108) | (0.5522) |  |
| a | 0.9832\*\*\* | 0.9934\*\*\* | 0.5052 | *1* |
|  | (0.0313) | (0.0109) | (0.6000) |  |
| b | 1.9963\*\*\* | 1.9925\*\*\* | 1.8964\*\*\* | *2* |
|  | (0.0206) | (0.0069) | (0.3936) |  |
| Constant | 19.9864\*\*\* | 19.8237\*\*\* | 20.3174 | *20* |
|  | (1.0831) | (0.3595) | (20.7385) |  |
|  |  |  |  |  |
| Observations | 100 | 1,000 | 100 | *∞* |
| R-squared | 0.9999 | 0.9999 | 0.9657 | *1* |

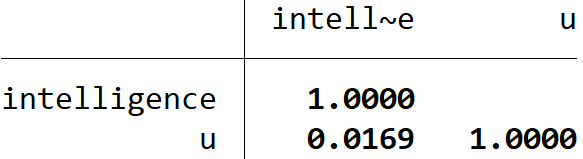
Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

En segundo lugar, generamos exactamente las mismas variables que antes, con la excepción de que ahora la variable u la construimos con una varianza de 20, mientras que antes era 1. Con 100 observaciones nuevamente, los resultados de la regresión se exponen en la columna 3. Como puede apreciarse, los errores estándar de todos los regresores aumentan significativamente.

Finalmente, mostramos que los errores del modelo poblacional son ortogonales a nuestra variable de interés, regresando los primeros sobre la segunda. Esto lo podemos hacer gracias a que conocemos el verdadero proceso generador de datos. La Tabla 2 muestra la correlación entre las variables mientras que la Tabla 3 despliega los resultados de dicha regresión.

#### Tabla 2 Variable Ortogonal al Error

**

Como se puede observar en la tabla 3, al correr la regresión de *intelligence* sobre u, el coeficiente de este último no es significativamente distinto de cero.

#### Tabla 3

|  |  |
| --- | --- |
|  | *(1)* |
| *VARIABLES* | *intelligence* |
|  |  |
| *u* | *0.0435* |
|  | *(0.2596)* |
| *Constant* | *504.4182\*\*\** |
|  | *(6.5360)* |
|  |  |
| *Observations* | *100* |
| *R-squared* | *0.0003* |

*Standard errors in parentheses*

*\*\*\* p<0.01, \*\* p<0.05, \* p<0.1*

## **Ejercicio 2**

1. Se espera que los coeficientes sean distintos, dado que ** está sesgado. La razón del sesgo es que esa especificación omite variables relevantes altamente correlacionadas con la variable de interés.
2. En este caso, **va a seguir estando sesgado pero el sesgo va a ser pequeño. En la segunda regresión, como la correlación entre X1 y las dos variables muy correlacionadas es casi nula, la varianza del estimador  no se va a ver muy afectada, por lo que esperamos que los regresores sean similares.
3. Suponiendo que la variable X4 es una variable irrelevante a la hora de explicar el score, esperamos que  también sea insesgado, entonces serán muy similares.
4. Esperaríamos que el error estándar de  sea mayor al de *.* Por un lado, como X2 y X3 tienen un pequeño efecto sobre Y, incluirlas disminuiría la varianza del error. Sin embargo, al tener un alto grado de colinealidad con X1, esto aumentaría la varianza del estimador  significativamente.
5. Es probable que el error estándar de  sea menor al de *.* Podríamos pensar que la varianza del error en el modelo con solo X1 es alta, dado que estamos dejando variables muy relevantes en el error, por lo cual la varianza de ** sería alta. Por otro lado, la varianza del error en el segundo modelo con X1, X2 y X3 debería ser más chica dado que agregamos más variables relevantes para explicar Y. El alto grado de colinealidad entre X2 y X3 puede aumentar la varianza de , pero el efecto debería ser pequeño al haber poca correlación entre estas variables y X1.
6. El error estándar de  será mayor que el de . Suponiendo que X4 no está correlacionada con Y, agregarla no disminuiría la varianza del error. Pero, como posiblemente tenga cierta correlación con X1, elevaría el grado de colinealidad de X1 con los regresores, por lo cual aumentaría la varianza del estimador .
7. Estimamos un modelo de score en base a *assistance* y otras dos variables, a y b. Al ingresar un error no sistemático de medición en *assistance* con media cero, como podemos observar en la Tabla 4, el coeficiente de *assistance\_error* será sesgado hacia el cero.

#### Tabla 4

|  |  |  |
| --- | --- | --- |
|  | *(1)* | *(2)* |
| *VARIABLES* | *score* | *score* |
|  |  |  |
| *assistance* | *1.9996\*\*\** |  |
|  | *(0.0007)* |  |
| *a* | *0.9932\*\*\** | *1.5025\*\*\** |
|  | *(0.0109)* | *(0.5346)* |
| *b* | *1.9928\*\*\** | *1.4784\*\*\** |
|  | *(0.0069)* | *(0.3372)* |
| *assistance\_error* |  | *1.4390\*\*\** |
|  |  | *(0.0286)* |
| *Constant* | *19.8194\*\*\** | *297.2857\*\*\** |
|  | *(0.3597)* | *(15.2527)* |
|  |  |  |
| *Observations* | *1,000* | *1,000* |
| *R-squared* | *0.9999* | *0.7192* |

Por último, estimamos un modelo de *score* en base a *assitance, a* y *b*, y luego la variable dependiente score\_error, con error de medición, contra los mismos regresores. Observamos que el valor de los coeficientes no cambia mucho (dado que no afecta la insesgadez), pero aumenta significativamente los errores estándar.

#### Tabla 5

|  |  |  |
| --- | --- | --- |
|  | *(1)* | *(2)* |
| *VARIABLES* | *score* | *score\_error* |
|  |  |  |
| *assistance* | *1.9996\*\*\** | *1.9899\*\*\** |
|  | *(0.0007)* | *(0.0534)* |
| *a* | *0.9932\*\*\** | *0.5533* |
|  | *(0.0109)* | *(0.8473)* |
| *b* | *1.9928\*\*\** | *2.4982\*\*\** |
|  | *(0.0069)* | *(0.5347)* |
| *Constant* | *19.8194\*\*\** | *25.2363* |
|  | *(0.3597)* | *(27.8956)* |
|  |  |  |
| *Observations* | *1,000* | *1,000* |
| *R-squared* | *0.9999* | *0.5846* |

**Código:**

*clear all*

*cd "C:\Users\juanb\OneDrive\Documentos\Juan\UdeSA\Economía Aplicada\Tutoriales\Tutorial 4"*

*\*EJERCICIO 1*

*\*a)*

*\*Primero un modelo con 100 observaciones*

*set obs 100*

*set seed 150*

*gen intelligence=int(invnormal(uniform())\*50+500)*

*\*sum intelligence*

*\*kdensity intelligence, norm*

*gen education=int(intelligence/5+invnormal(uniform())\*3)*

*corr education intelligence*

*gen a=int(invnormal(uniform())\*3+10)*

*gen b=int(invnormal(uniform())\*5+5)*

*gen u=int(invnormal(uniform())\*1+20)*

*\*Verdadero modelo*

*gen wage=2\*intelligence+a+2\*b+u*

*\*Corremos el modelo con la variable education*

*reg wage intelligence education a b*

*est store mco1*

*outreg2 using prueba1, word dec(4) label*

*clear*

*\*Ahora con 1000 observaciones*

*set obs 1000*

*set seed 150*

*gen intelligence=int(invnormal(uniform())\*50+500)*

*\*sum intelligence*

*\*kdensity intelligence, norm*

*gen education=int(intelligence/5+invnormal(uniform())\*3)*

*corr education intelligence*

*gen a=int(invnormal(uniform())\*3+10)*

*gen b=int(invnormal(uniform())\*5+5)*

*gen u=int(invnormal(uniform())\*1+20)*

*\*Verdadero modelo*

*gen wage=2\*intelligence+a+2\*b+u*

*reg wage intelligence education a b*

*est store mco2*

*outreg2 using prueba1, word dec(4) label*

*esttab mco1 mco2, se*

*\*b)*

*clear*

*set obs 100*

*set seed 150*

*gen intelligence=int(invnormal(uniform())\*50+500)*

*\*sum intelligence*

*\*kdensity intelligence, norm*

*gen education=int(intelligence/5+invnormal(uniform())\*3)*

*corr education intelligence*

*gen a=int(invnormal(uniform())\*3+10)*

*gen b=int(invnormal(uniform())\*5+5)*

*gen u=int(invnormal(uniform())\*20+20)*

*\*Verdadero modelo*

*gen wage=2\*intelligence+a+2\*b+u*

*reg wage intelligence education a b*

*est store mco3*

*outreg2 using prueba1, word dec(4) label*

*esttab mco1 mco2 mco3, se*

*\*3)*

*corr intelligence u*

*reg intelligence u*

*outreg2 using prueba2, word dec(4) label*

*\*EJERCICIO 2*

*\*g)i)*

*clear*

*set obs 1000*

*set seed 150*

*gen assistance=int(invnormal(uniform())\*50+500)*

*gen error\_medición=int(invnormal(uniform())\*80)*

*gen assistance\_error = assistance + error\_medición*

*gen a=int(invnormal(uniform())\*3+10)*

*gen b=int(invnormal(uniform())\*5+5)*

*gen u=int(invnormal(uniform())\*1+20)*

*gen score=2\*assistance+a+2\*b+u*

*\*Regresión sin error de medición*

*reg score assistance a b*

*est store proof1*

*outreg2 using prueba3, word dec(4) label*

*\*Regresión con error de medición*

*reg score assistance\_error a b*

*est store proof2*

*outreg2 using prueba3, word dec(4) label*

*esttab proof1 proof2, se*

*\*g)ii)*

*clear*

*set obs 1000*

*set seed 150*

*gen assistance=int(invnormal(uniform())\*50+500)*

*gen error\_medición=int(invnormal(uniform())\*80)*

*gen a=int(invnormal(uniform())\*3+10)*

*gen b=int(invnormal(uniform())\*5+5)*

*gen u=int(invnormal(uniform())\*1+20)*

*gen score\_error=2\*assistance+a+2\*b+u+error\_medición*

*gen score=2\*assistance+a+2\*b+u*

*\*Regresión sin error de medición*

*reg score assistance a b*

*est store proof3*

*outreg2 using prueba4, word dec(4) label*

*\*Regresión con error de medición*

*reg score\_error assistance a b*

*est store proof4*

*outreg2 using prueba4, word dec(4) label*

*esttab proof3 proof4, se*