



INSTITUTO DE ECONOMÍA
FACULTAD DE CIENCIAS ECONÓMICAS
Y ADMINISTRATIVAS

Tarea N°1

Econometrics Theory - EAE350B-1

Estudiante [Valentina Andrade](#)
Profesor Raimundo Soto
Ayudantes Sofía Valdivia y Gabriel Zelpo

domingo 29, mayo 2022

Resumen

El siguiente reporte tiene por objetivo presentar los análisis realizados en la Tarea N°1 del ramo Teoría Econométrica I dictado por el profesor Raimundo Soto. En la primera parte abordamos propiedades de la estimación de Mínimos Cuadrados Ordinarios, junto con un diálogo con el Teorema Central del Límite. En la parte 2 discutimos de manera más concreta los supuestos y resultados de la teoría asintótica. Se puede acceder al repositorio en GitHub de la tarea [aquí](#)

Índice general

Parte 1	2
Pregunta 1	2
Pregunta 2	2
Pregunta 3	5
Pregunta 4	6
Pregunta 5	8
Pregunta 6	9
Pregunta 7	10
Pregunta 8	12
Pregunta 9	14
Pregunta 10	16
Pregunta 11	17
Pregunta 12	20
Parte 2	21
Referencias	24

Parte 1

Pregunta 1

Genere un vector con 1.000 observaciones de datos aleatorios que provengan de una distribución normal de media 10.000 y desviación estándar de 50, llámelo Y . Luego, genere otra variable, llamada C , tal que $C = \beta \cdot Y$ con $\beta = 0.9$.

Cuadro 1: Proceso generador de datos

Y	C
9989.401	8990.461
9978.267	8980.440
9987.168	8988.451
10038.723	9034.851
10046.105	9041.494
9942.086	8947.877

Pregunta 2

Usando los datos anteriores, se ha realizado una regresión lineal de C en Y .

Cuadro 2: Modelo de regresión lineal

	Modelo 1
(Intercept)	0.000 (0.000)
Y	0.900*** (0.000)
R^2	1.000
Adj. R^2	1.000
Num. obs.	1000
F statistic = 4.044168e+30	

La estimación de $\hat{\beta}$ es 0.9. Numéricamente, este es prácticamente el valor original, pues la distancia entre 0.9 y la estimación es de $-2.8865799 \times 10^{-15}$, es decir, prácticamente cero. En el Cuadro 4 podemos ver el intervalo de confianza al 95%, el cual muestra tener una amplitud muy pequeña pues este es menor a $10 \cdot e^{-10}$

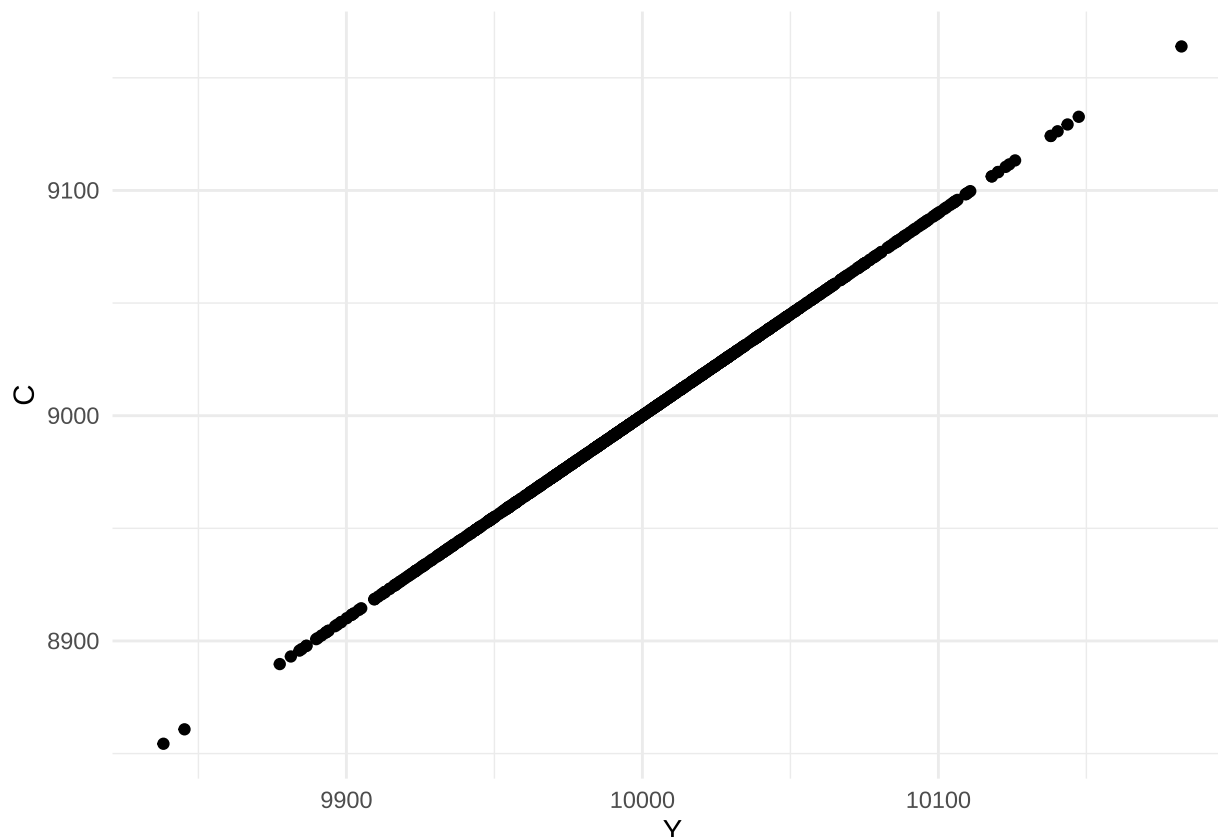
Cuadro 3: Intervalo de confianza

parámetro	2.5 %	97.5 %	amplitud del intervalo
(Intercept)	0.0	0.0	0
Y	0.9	0.9	0

El cuadro 1 nos señala que el R^2 del modelo es 1. Notar que en este caso, al ser un predictor tenemos que $R^2 = \text{cor}(Y, C)^2$. Respecto a la significancia global del modelo, el estadístico F es igual a 1.2577998×10^{28} asociado al valor-p 0 indicando que no existe evidencia estadística de que C no posea efecto sobre Y .

En síntesis, podemos ver cómo se da la relación entre C con Y en la siguiente gráfica. Tal como podemos ver en la figura, la relación es prácticamente perfecta. Esto se debe a que la simulación

corresponde a una relación determinística entre C e Y , donde Y corresponde a una combinación lineal de C . En este modelo hay ausencia de un efecto aleatorio pues pese a que hicimos una muestra random el proceso generador de los datos, la variable dependiente es generada únicamente de la misma variable dependiente.



Pese a que este análisis es muy simple, de aquí en adelante será importante precisar los siguientes puntos.

En *OLS* cuando se estiman los parámetros buscaremos minimizar suma de los residuos al cuadrado $u'u$. En este caso obtuvimos

$$\sum_{i=0}^n \hat{u}_i = 0$$

Cuando ponemos esta suma de residuos en contraste a la suma total de cuadrados SST , podemos saber cuanto de la varianza de la variable dependiente es explicada por el conjunto de predictores, o en otras palabras cuál es la suma de cuadrados explicados SSE . Evidentemente con ello podemos notar que si toda la varianza de Y proviene de C , entonces $SST = SSE$. Como R^2 se define como el ratio entre estas dos, evidentemente este será 1.

Sabemos que el estadístico F es una prueba de significancia conjunta un grupo de variables (q) que pertenecen a k parámetros de un modelo, y que busca probar que esas q de las k variables **no** tienen un efecto parcial sobre y . Este tipo de estadísticos son muy utilizados cuando se requiere testear restricciones de exclusión en los modelos de regresión, elemento que volveremos a evaluar en el punto 11. En términos generales, se buscará que la suma de residuos cuadrados del modelo restringido

(SSR_r) sea mayor que el modelo sin restringir (SSR_{nr}), siendo coherentes con la estimación MCO dado que un mayor número de predictores incorporados en el modelo aumentará la bondad de ajuste ($\nearrow R^2, \searrow SSR$). En consecuencia, se probará si esta diferencia entre los SSR de los modelos es estadísticamente distinto. Formalmente,

$$F_{(n,k,gl)} \equiv \frac{(SSC_r - SSC_{nr})/q}{SSC_{nr}/(n - k - 1)} \quad (1)$$

Donde SSR indica la suma de residuos cuadrados, n el número de observaciones del modelo, k el número de parámetros ($gl = n - k - 1$) y q el número de restricciones. Los subíndices r y nr indican el modelo con restricciones q en parámetros y sin restricciones, respectivamente. El modelo sin restricciones, cuando no se explicita cuál es corresponde a un modelo *nulo* sin predictores.

A su vez, sabemos que si el estadístico F es calculado a partir de la suma de residuos cuadrados, y el coeficiente de determinación R^2 lo podemos calcular como $R^2 = 1 - \frac{SSR}{SST}$, donde SST es la suma total de los residuos, podemos decir que

$$F_{(q,k,gl)} = \frac{(R_{nr}^2 - R_r^2)/q}{1 - R_{nr}^2/(n - k - 1)} \quad (2)$$

Recordemos que dado que el R_r^2 no tiene predictores, este es cero

$$F_{(q,k,gl)} = \frac{(R_{nr}^2/k)}{1 - R_{nr}^2/(n - k - 1)} \quad (3)$$

Es por ello que los resultados obtenidos, por lo general, estarán muy conectados. Como se puede notar su hilo conductor es la minimización de la SSR , por lo que entender qué hay detrás de ellos será la clave para entender qué va pasando en las siguientes preguntas.

Pregunta 3

Generamos un vector aleatorio y lo hemos llamado ϵ_1 tal que este provenga de una distribución normal con media 0 y desviación estándar de 5. Además, se ha generado otra variable llamada $Y_1 = Y + \epsilon_1$ y hemos estimado una regresión lineal de C en Y_1

¿Cuál es el estimador $\hat{\beta}$? ¿Es éste distinto o igual a 0.9? ¿Cuál es el intervalo de confianza del estimador al 95%? ¿Cuál es el R^2 de la regresión? ¿Por qué? ¿Es significativo el modelo de acuerdo al test F ? Explique por qué obtiene esos resultados.

Cuadro 4: Modelo de regresión lineal

	Modelo 2
(Intercept)	99.260*** (28.606)
Y1	0.890*** (0.003)
R ²	0.990
Adj. R ²	0.990
Num. obs.	1000
F statistic = 99710.81, df = 1, gl = 998	

Como podemos ver en el cuadro la estimación de $\hat{\beta}$ es 0.8900697. Respecto a su distancia con el valor que originalmente tomamos para la regresión, si bien no es igual, aun así son muy similares, difieren en 0.0099303.

Respecto al intervalo de confianza, podemos ver la siguiente tabla

Cuadro 5: Intervalo de confianza

parámetro	2.5 %	97.5 %	amplitud_del_intervalo
(Intercept)	43.1251939	155.3939196	112.2687257
Y1	0.8844571	0.8956824	0.0112253

El R^2 es 0.9897996 y no es 1 pues, a diferencia de antes, hemos agregado un componente aleatorio a la variable dependiente, ϵ_1 . Entonces, cuando hacemos la estimación ahora existe una parte de la varianza de C que no es explicada completamente por Y , y por ello, la suma de residuos al cuadrado SSR crece. En la primera regresión SSR comprende es $-1.7149457 \times 10^{-23}$, mientras que en la segunda regresión SSR es $-3.6026737 \times 10^{-14}$.

¿Es significativo el modelo de acuerdo al test F ?

El valor reportado es $F(9.6841697 \times 10^4, 1, 998)$, donde la primera coordenada indica el valor observado de F (F, k, gl). A partir de estos valores se puede obtener su *valor-p* que es $2.2 \cdot e^{-16}$ el que nos indica que se rechaza la hipótesis nula que indica que los predictores en conjunto no tienen un efecto sobre C .

Ya hemos comentado sabemos que existe un diálogo entre estos distintos resultados, por la sencilla razón de que todos provienen de la misma fuente: SSR . En este tercer inciso se nos ha solicitado que el C esté compuesto por Y y un error. Y luego cuando hacemos la estimación, queremos saber cuál es el efecto marginal de Y sobre C . Evidentemente de toda la varianza de C no todo puede ser explicado por Y , pues está compuesta la variable dependiente por un error.

Formalmente $C = Y_1 + \epsilon_1 = 0.9 \cdot Y + \epsilon_1$.

En el fondo, si imaginamos que no podemos saber *cuál es ese error aleatorio* que está en Y , formalmente estamos mostrando cómo se comporta en una regresión la parte sistemática de la no sistemática, y como pese a ello OLS intenta llegar a la mejor estimación lineal insesgada (las estimaciones son muy similares en presencia de un error).

Pregunta 4

Genere la variable $Y_2 = Y_1 + 100$ y estime una regresión lineal de C en Y_2 . ¿Cuál es el estimador $\hat{\beta}$? ¿Cómo cambia el valor de la constante respecto del caso anterior? ¿Cuál es el R^2 de la regresión? ¿Cuál es el intervalo de confianza del estimador al 95%? ¿es significativo el modelo de acuerdo al test F ? Explique por qué obtiene esos resultados.

Cuadro 6: Modelo de regresión lineal

	Modelo 3
(Intercept)	10.253 (28.892)
Y2	0.890*** (0.003)
R^2	0.990
Adj. R^2	0.990
Num. obs.	1000
F statistic = 99710.81, df = 1, gl = 998	

La estimación de β es 0.8900697. El valor de la constante respecto del caso anterior es idéntico al del modelo de regresión de C en Y_1 . En efecto:

```
##          Y1
## 0.8900697
```

```
##          Y2
## 0.8900697
```

El R^2 es 0.9897996, que es igual a la bondad de ajuste del modelo 2. Pasa lo mismo con los intervalos de confianza

Cuadro 7: Intervalo de confianza

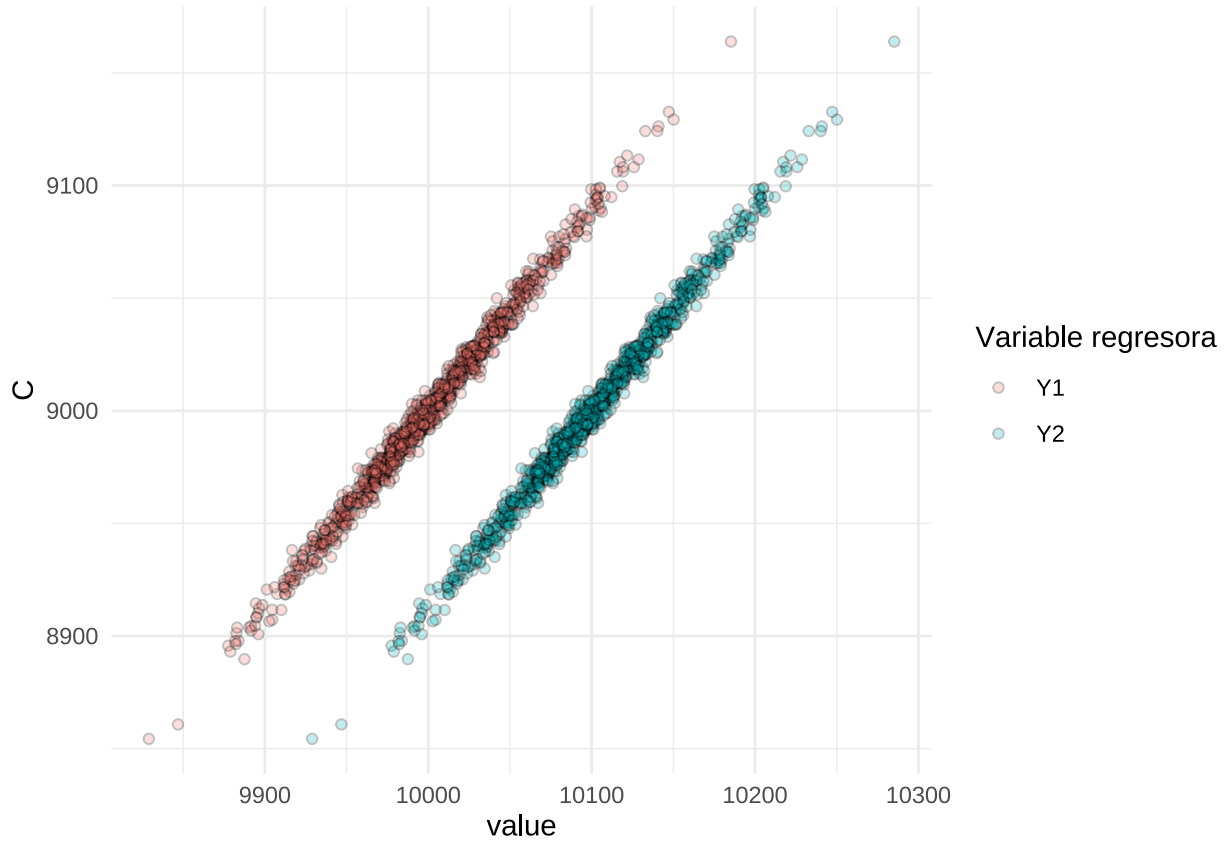
parámetro	2.5 %	97.5 %	amplitud_del_intervalo
(Intercept)	-46.4430373	66.9482048	113.3912421
Y2	0.8844571	0.8956824	0.0112253

¿Es significativo el modelo de acuerdo al test F ?

El valor reportado es $F(9.6841697 \times 10^4, 1, 998)$, donde la primera coordenada indica el valor observado de F (F , k , gl). A partir de estos valores se puede obtener su *valor-p* que es $2.2 \cdot e^{-16}$ el que nos indica que se rechaza la hipótesis nula que indica que los predictores en conjunto no tienen un efecto sobre Y . Estos resultados son los mismos que antes.

Los resultados en términos de ajuste coinciden con la regresión anterior pues lo único que se hizo fue trasladar en 100 puntos la variable dependiente. Esto es similar a cuando pensamos transformaciones

funcionales en los modelos de regresión. Estos no cambian las propiedades de *OLS*, sino que re-escalen las unidades de medida, muchas veces para mejorar interpretaciones. Veamos esto de forma gráfica:



Antes de cerrar la pregunta 3 y 4 es importante señalar que este es una forma de estudiar qué pasa con las propiedades de *OLS* cuando la variable dependiente tiene un error asociado. Consideremos que la primera regresión representa el modelo original. Luego tenemos que Y^* (valor verdadero), entonces $Y = Y^* + e_1 \Rightarrow e_1 = Y - Y^*$. Con ello el modelo original se escribe como

$$C = \beta \cdot Y^* + u$$

Mientras, que el modelo que estimamos se escribe como

$$C = \beta \cdot Y + u + e_o$$

Si tomamos por supuesto el error clásico de variable donde $cov(Y^*, e_1) = 0$, $cov(Y, e_1) = \sigma_e^2$ y $cov(Y, C) = \beta \cdot \sigma_e^2$

Con ello notaremos que $\hat{\beta}$ es sesgado (lo que se llama sesgo de atenuación), y su consistencia depende de la relación que se tiene con e_1 .

$$plim(\hat{\beta}) = plim\left(\frac{cov(C, Y)}{Var(Y)}\right) = \frac{cov(y^*, c) + cov(e_2, c)}{var(y^*) + var(e_1) + 2cov(y^*, e_2)} = \frac{cov(y^*, u + \beta \cdot y^*)}{var(y^*) + var(e_1)} = \beta \cdot \frac{var(y^*)}{var(y^*) + var(e_1)}$$

Donde el último término corresponde al sesgo de atenuación del error en la variable dependiente. Si la varianza del error aleatorio pesa poco, entonces $\text{var}(y^*) > \text{var}(y)$, por lo que la inconsistencia será más pequeña. De hecho, la varianza de e_1 es de cerca de un 4.9937269, mientras que la de Y_1 es 49.4426943

Pregunta 5

Se ha generado un vector aleatorio y lo llamamos ϵ_2 tal que este provenga de una distribución normal con media 0 y desviación estándar de 50. Además, se ha generado otra variable llamada $C1 = C + \epsilon_2$ y se estimó una regresión lineal de C_1 en Y .

Cuadro 8: Modelo de regresión lineal

	Modelo 3
(Intercept)	-385.333886 (312.213403)
Y	0.938496*** (0.031217)
R ²	0.475237
Adj. R ²	0.474711
Num. obs.	1000
F statistic = 793.9376, df = 1, gl = 998	

En este caso podemos decir que en promedio, el efecto parcial que tiene Y sobre C es de 0.9 puntos ($\hat{\beta} = 0.9$), con un 99% de confianza. Este valor es estadísticamente distinto de cero, y numéricamente es 0.9384958, por lo que tiene una distancia muy cercana con 0.9.

De forma idéntica a la pregunta anterior:

Cuadro 9: Intervalo de confianza

parámetro	2.5 %	97.5 %	amplitud_del_intervalo
(Intercept)	-998.003937	227.3361652	1225.3401018
Y	0.877237	0.9997546	0.1225176

Respecto a la bondad de ajuste podemos decir que el R^2 es de 0.4752371, es decir cerca de un 40% de la varianza de Y es explicado por este nuevo C , es decir, una proporción *muchísimo* menor que en los casos anteriores.

El valor reportado es $F(903.8113422, 1, 998)$, donde la primera coordenada indica el valor observado de F (F, k, gl). A partir de estos valores se puede obtener su *valor-p* que es $2.2 \cdot 10^{-16}$ el que nos indica que se rechaza la hipótesis nula que indica que los predictores en conjunto no tienen un efecto sobre Y . Estos resultados son los mismos que antes.

A partir de este ejercicio podemos discutir las implicancias de un error introducido en la variable dependiente. Formalmente, tenemos un $e_2 = C - C^*$, donde el asterisco nos indica el valor verdadero. El modelo original ha sido

$$C* = \beta \cdot Y + u$$

Y ahora estimamos

$$C = \beta \cdot Y + e_2 + u$$

Si el valor esperado de este error aleatorio sigue siendo cero $E(e_2) = 0$ y $cov(e_2, u) = 0$, entonces la estimación seguirá siendo insesgada y consistente. Para demostrarlo, llamemos al error compuesto $v = e_2 + u$. Partamos viendo que pasa con el valor esperado del error

$$E(v) = E(u + e_2) = E(u) + E(e_2) = 0$$

Mientras que para la varianza es algo más complejo.

$$Var(v|Y) = Var(u + e_2|Y) = Var(v|Y) + Var(e_2|Y) + 2Cov(u, e_2|X)$$

Si $cov(u, e_2|X) \neq 0$, entonces $\hat{\beta}$ está sesgado (positivamente si esa covarianza es positiva, y negativamente si esa covarianza es negativa). Notemos que en este caso la estimación da prácticamente igual a la original, por lo que podemos descartar esta fuente de covarianza

Si $cov(u, e_2|X) = 0$ entonces nos queda $Var(v|Y) + Var(e_2|Y)$. Notemos que estas dos varianzas serán las del modelo, y si estas son finitas (o incluso si somos más restrictivos, homocedásticas), entonces $\sigma_u^2 + \sigma_e^2$. Entonces, esta estimación tendrá una mayor varianza, por lo que será *ineficiente*. Es por eso que el error estándar es *muchísimo más grande que en los casos anteriores*. Como ya hemos discutido, la varianza es un componente importante para entender los estadísticos de ajuste. En este caso, existirá una proporción de la varianza de la variable dependiente que la variable independiente no podrá capturar.

Cerremos con ver qué pasa con la consistencia:

$$plim(\hat{\beta}) = plim\left(\frac{cov(C, Y)}{Var(Y)}\right) = \frac{cov(c, y* + e_2)}{var(y)} = \frac{cov(c, y*)}{var(y)} + \frac{cov(c, e_2)}{var(y)} = \beta + \frac{cov(c, e_2)}{var(y)}$$

Entonces si $\frac{cov(c, e_2)}{var(y)} = 0$ es insesgado nuestro coeficiente. Dado los resultados, este es el caso más probable. Pero de lo que no nos salvamos, es de la ineficiencia de la estimación.

Pregunta 6

Generamos un vector con 1.000 observaciones de datos aleatorios que provienen de una distribución normal de media 150 y desviación estándar de 25, y lo hemos llamado Z . Además generamos otra variable llamada C_2 , tal que $C_2 = 0.9 \cdot Y - 0.25 \cdot Z$. Estime la regresión de C_2 en Y y Z . ¿Por qué obtiene esos estimadores? ¿Cuál es el R^2 de la regresión? ¿Cuál es el intervalo de confianza de los estimadores al 95%? ¿es significativo el modelo de acuerdo al test F ? Explique por qué obtiene esos resultados.

Cuadro 10: Modelo de regresión lineal

	Modelo 4
(Intercept)	0.000000 (0.000000)
Y	0.900000*** (0.000000)
Z	-0.250000*** (0.000000)
R ²	1.000000
Adj. R ²	1.000000
Num. obs.	1000
F statistic = 793.9376, df = 1, gl = 998	

Cuadro 11: Intervalo de confianza

parámetro	2.5 %	97.5 %	amplitud_del_intervalo
(Intercept)	-998.003937	227.3361652	1225.3401018
Y	0.877237	0.9997546	0.1225176

Cuadro 12: Estadísticos de ajuste

r.squared	F statistic	p.value
1	3.541437e+29	0

Primero que todo, vemos que la construcción de C_2 en función de Y y Z es determinística pues no existe la presencia de una perturbación. Al igual que en la pregunta N°2, esto genera que el R^2 sea 1, y que globalmente las variables sean estadísticamente significativas (ver test F). Además, los intervalos de confianza poseen amplitud muy pequeña y centrada en el valor real de 0.9 y -0.25. Es decir, en ausencia de incertidumbre respecto a la estimación, pues sabemos a ciencia cierta que C_2 se conforma de una combinación lineal de Y e Z . Entonces, cuando estimamos C_2 y queremos ver cuánto de su varianza es explicada por ambas variables, evidentemente será completamente por ellas.

Pregunta 7

Generamos un $C_3 = C_2 + \epsilon_2$ y estimamos la regresión de C_3 en Y y Z . Discutiremos por qué se obtiene esos estimadores, junto con el ajuste global del modelo.

Cuadro 13: Modelo de regresión lineal

	Modelo 5
(Intercept)	-385.177890 (312.895288)
Y	0.938488*** (0.031246)
Z	-0.250517*** (0.060071)
R ²	0.481969
Adj. R ²	0.480930
Num. obs.	1000
F statistic = 397.9, df = 2, gl = 997	

Como podemos ver, por cada unidad que aumente Y , en promedio C_3 aumenta en un 0.9 puntos, ceteris paribus. Podemos decir con un 95% de confianza que se rechaza la hipótesis nula de que Y no tiene un efecto sobre C_3 . Por otro lado, por cada unidad que aumente Z , en promedio C_3 disminuye en un 0.16 puntos, ceteris paribus. Podemos decir con un 95% de confianza que se rechaza la hipótesis nula de que Z no tiene un efecto sobre C_3 .

Cuadro 14: Intervalo de confianza

parámetro	2.5 %	97.5 %	amplitud_del_intervalo
(Intercept)	-998.003937	227.3361652	1225.3401018
Y	0.877237	0.9997546	0.1225176

r.squared	F statistic	p.value
0.4819692	463.798	0

Nuevamente tenemos una variable dependiente medida con error, por lo que los análisis anteriores también se extrapolan a este caso. Es decir, los estimadores que obtenemos siguen siendo insesgados pero ineficientes. Al igual que antes se muestra que tenemos una bondad de ajuste en términos de R^2 similar al ejercicio 5. Esto se debe a que C esta conformado por una partición de ambas variables más el error, por lo que a no ser que ingresemos una variable que sepa capturar ese error aleatorio (lo cuál no siempre tiene sentido si la investigación no se enfoca a captar ese residuo), entonces solo en ese caso aumentaría la bondad de ajuste. Notemos además, que en presencia de dos variables dependientes, el estadístico de ajuste no “castiga” la estimación multivariada debido a que ambas variables aportan información sobre la variación de la variable dependiente.

Ahora bien, pese a lo indicado sobre el error en variable dependiente y el aumento de la incertidumbre en la estimación, las estimaciones obtenidas están cerca de los valores 0.9 y -0.25 del ejercicio anterior. Matemáticamente, tenemos una perturbación pues partiendo por $C_3 = C_2 + e_2$, y luego reemplazando C_2 en términos de Y y Z tenemos que $C_3 = 0.9 * Y - 0.25 * Z + e_2$.

Pregunta 8

Haremos dos regresiones separadas. La primera de C_3 en Y y la segunda de C_3 en Z .

Cuadro 15: Modelos de regresión lineal simple

	Modelo 6	Modelo 7
(Intercept)	-460.762836 (314.924800)	9008.737231*** (12.626704)
Y	0.942281*** (0.031488)	
Z		-0.303037*** (0.082830)
R ²	0.472933	0.013234
Adj. R ²	0.472404	0.012245
Num. obs.	1000	1000

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Cuadro 16: Intervalo de confianza para Modelo 6

parámetro	2.5 %	97.5 %	amplitud_del_intervalo
(Intercept)	-1078.7535795	157.227908	1235.9814879
Y	0.8804903	1.004072	0.1235816

r.squared	F statistic	p.value
0.4729325	895.4956	0

Cuadro 17: Intervalo de confianza para Modelo 7

parámetro	2.5 %	97.5 %	amplitud_del_intervalo
(Intercept)	8983.959295	9033.5151669	49.5558715
Z	-0.465579	-0.1404957	0.3250833

r.squared	F statistic	p.value
0.0132342	13.38484	0.0002669

En términos descriptivos podemos decir que el primer modelo, C_3 en Y es mejor en términos de ajuste. Si bien los dos modelos son significativos en términos del test F, el segundo modelo está muy cerca de no serlo. Lo anterior se debe principalmente a que en la construcción de C_3 impacta de mayor manera el valor de Y que de Z dado a los coeficientes en su construcción, que son 0.9 y -0.25 respectivamente.

De forma adicional podemos mostrar las correlaciones en donde se observa la alta asociación lineal de C_3 e Y y la baja entre C_3 y Z lo que apoya a los resultados obtenidos.

Cuadro 18: Tabla de correlaciones

	C2	Y	Z
C2	1.0000000	0.9898154	-0.1711092
Y	0.9898154	1.0000000	-0.0291093
Z	-0.1711092	-0.0291093	1.0000000

Ahora bien, de manera sustantiva podemos notar un resultado clave. En términos generales los parámetros estimados para la regresión múltiple y simple son similares pero no exactamente iguales. En la regresión múltiple el efecto parcial de Y sobre C es de 0.938488, mientras que en la simple es de 0.9422812. Si bien son valores similares, no son exactamente iguales. La razón es que entre Y y Z existe una pequeña correlación cercana a 0.02. Entonces con ello podemos mostrar las implicancias del *Teorema de Frish-Waugh-Lovell*.

Tenemos el siguiente modelo real

$$C = \beta_1 Y + \beta_2 Z + u$$

Cuando estimamos la regresión esto se ve matricialmente de la siguiente forma

$$\begin{bmatrix} Y'Y & Y'Z \\ Z'Y & Z'Z \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} Y'C \\ Z'C \end{bmatrix} \quad (4)$$

Modelos encontrar que

$$\hat{\beta}_1 = (Y'Y)^{-1}Y'C - (Y'Y)^{-1}Y'Z\hat{\beta}_2$$

Análogamente

$$\hat{\beta}_2 = (Z'Z)^{-1}Z'C - (Z'Z)^{-1}Z'Y\hat{\beta}_1$$

Entonces, cuando estimamos la regresión por *OLS* en términos multivariados tenemos que los parámetros se calculan de la forma tan de que la regresión de Y sobre C se calcula corrigiendo o descontando por $(Y'Y)^{-1}Y'Z\hat{\beta}_2$. En otras palabras, esto considera por un lado la *relación* que establecen las variables independientes, o como se indica en Greene(2003,p.148) la regresión de las variables Z parcializadas y luego puestas en conjunto en matriz de todas las variables Y . Esto solo sería cero si Y no está linealmente relacionada con Z , lo que sabemos que no es así, pues no son completamente ortogonales. La otra opción es que $\hat{\beta}_2 = 0$, es decir, que no sea estadísticamente significativo, pero como podemos ver tanto en el modelo 5 como 7, la relación entre Z y C es estadísticamente significativa.

Entonces, lo que estamos pasando por *alto* cuando hacemos regresiones lineales separadas es que las variables dependientes comparten varianza, por lo que para poder interpretar su aporte marginal de manera rigurosa, debemos descontarlas de la estimación del parámetro para tener el *efecto neto* de Y , que es lo que frecuentemente llamamos **ceteris paribus**

Pregunta 9

Generamos un vector aleatorio y lo llamamos ϵ_3 tal que este provenga de una distribución normal con media 10 y desviación estándar de 5. Generemos además $W = 0.9 \cdot Y + \epsilon_3$. Compute la correlación entre W e Y .

Además, generamos un vector aleatorio y lo llamamos ϵ_4 tal que este provenga de una distribución normal con media 0 y desviación estándar de 50. Creamos la variable $C_4 = 0.9 \cdot Y - 0.25 \cdot W + \epsilon_4$. Realizamos la regresión de C_4 en Y y W . Explique por qué obtiene esos resultados.

Cuadro 19: Matriz de correlaciones

	W	Y
W	1.0000000	0.9936095
Y	0.9936095	1.0000000

A partir de la matriz de correlaciones vemos que existe una fuerte relación entre Y y W . Ahora generaremos la variable dependiente C_4 a partir de una combinación lineal de Y y W , junto a un término de error. Evidentemente esto nos traerá un problema de multicolinealidad

Cuadro 20: Modelos de regresión lineal múltiple

	Modelo 8
(Intercept)	702.876817* (321.786596)
Y	0.946268*** (0.284532)
W	-0.378003 (0.315928)
R ²	0.265203
Adj. R ²	0.263729
Num. obs.	1000

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

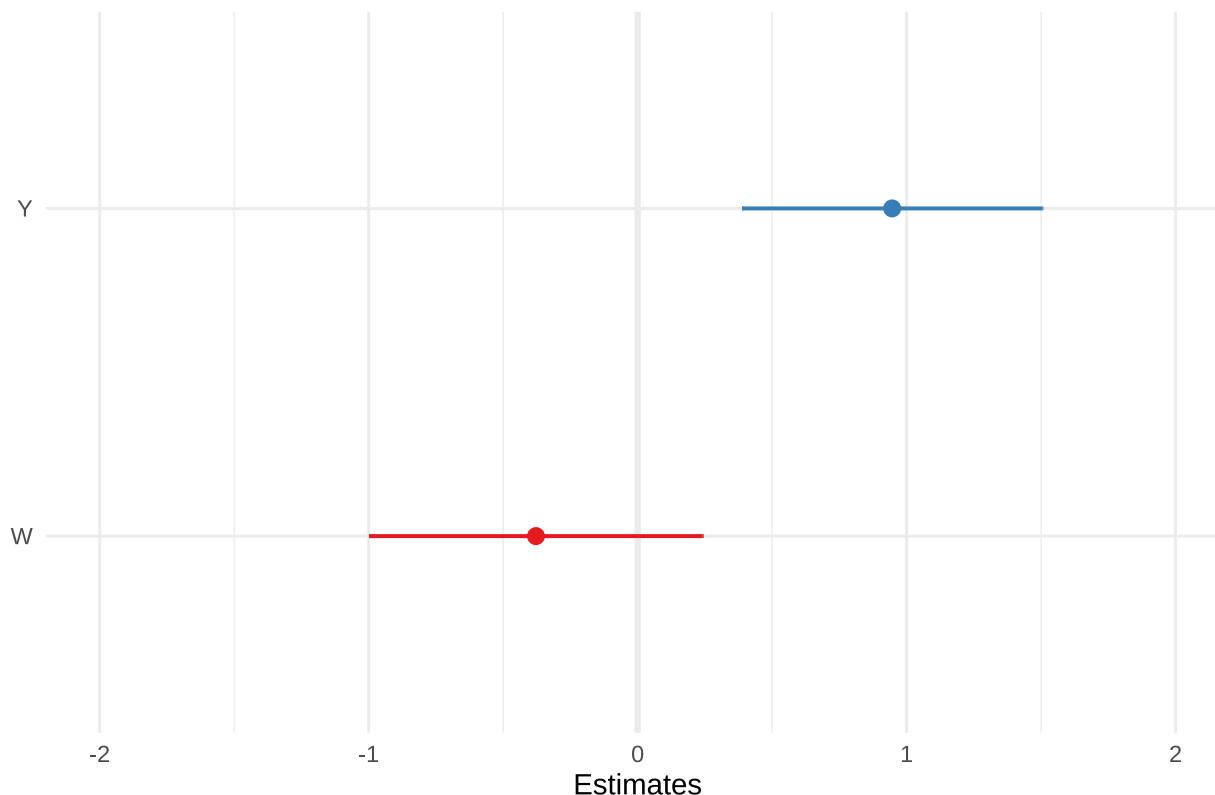
En el cuadro con el modelo 8 podemos notar que en promedio, el efecto parcial que tiene Y sobre C_3 es de 0.9462684 puntos ($\hat{\beta} = 0.9$), con un 99% de confianza. Mientras que no podemos decir lo mismo de W el cual si bien muestra un efecto negativo, este no es estadísticamente significativo. Evidentemente el error estándar ha crecido, lo que se hace evidente en el Cuadro 21 que muestra los intervalos de confianza.

Cuadro 21: Intervalo de confianza para Modelo 6

parámetro	2.5 %	97.5 %	amplitud_del_intervalo
(Intercept)	71.4201018	1334.3335316	1262.913430
Y	0.3879183	1.5046185	1.116700
W	-0.9979626	0.2419563	1.239919

También se expresa en el siguiente *forestplot*, donde ambas variables tienen un gran intervalo de confianza, lo que se debe a la pérdida de precisión en la estimación pues ahora no solo tenemos el error aleatorio del modelo original, sino que un error compuesto que proviene desde W .

C 4



En términos globales vemos que ahora el modelo ha disminuido su capacidad de explicar la varianza de C a un 26.5%. Uno se podría preguntar qué pasa con el test F que sigue siendo significativo. La razón es que si bien F es potente para implementar hipótesis conjuntas, este no distingue efectos parciales. Por lo que tal como dice Wooldridge, también se hace sensible a *multicolinealidad* (Wooldridge, 2017).

r.squared	F statistic	p.value
0.2652034	179.919	0

Recordemos que implica esto. Uno de los requisitos de OLS es tener una matriz de variables aleatorias de rango completo para que pueda ser invertible y definida positiva, pues si no, no sería estimable la regresión. Eso implica que no exista una relación lineal **exacta** entre las variables independientes, lo que se puede probar con si hay o no correlación perfecta (en ese caso, casi lo hay, de no ser por el término de error e_4). Pero Greene profundiza en esto, y más bien indica que la multicolinealidad es un problema y aparece en contextos donde la combinación lineal es una función lineal exacta de más variables. Esto es clave para la parcialización de efectos de las variables aleatorias. Recordemos cómo quedaba demostrado el teorema de *Frish-Waugh*. Este nos dice que descontaremos del coeficiente estimado la relación entre los dos parámetros. Si esta correlación es 1, en el fondo estaremos descontando una parte importante del tamaño de efecto sobre el coeficiente

estimado.

Por último, la multicolinealidad trae un problema no solo de sesgo sino que de ineficiencia. Notemos que si ponemos en una regresión dos veces la misma variable, y sabemos que esa variable tiene un componente de error, en el proceso de estimación, tendremos una mayor suma de residuos, y que como demostramos arriba, si la varianza de la variable con error es muy grande, entonces podríamos enfrentar problemas de consistencia y con ello la inferencia ya no sería válida.

Cuadro 22: Matriz de correlaciones

	C4	W	Y
C4	1.0000000	0.5070028	0.5139536
W	0.5070028	1.0000000	0.9936095
Y	0.5139536	0.9936095	1.0000000

Pregunta 10

Haremos dos regresiones separadas. La primera de C_4 en Y y la segunda de C_4 en W .

Cuadro 23: Modelos de regresión lineal simple

	Modelo 9	Modelo 10
(Intercept)	679.670570* (321.270991)	759.357349* (322.953680)
Y	0.608005*** (0.032123)	
W		0.665964*** (0.035839)
R^2	0.264148	0.257052
Adj. R^2	0.263411	0.256307
Num. obs.	1000	1000

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Cuadro 24: Ajuste global Modelo W

r.squared	F statistic	p.value
0.2570518	345.297	0

Cuadro 25: Ajuste global Modelo W

r.squared	F statistic	p.value
0.2641483	358.2515	0

Como podemos ver ambos modelos poseen coeficientes significativos al 99% de confianza, además de ser globalmente significativos como se ve en el Cuadro 24 y 25.

Ahora bien, para analizar qué está ocurriendo hay dos cosas que tener en cuenta: 1. Recordemos que entre Y y W existía correlación del 0.99 por lo que luego cada uno de forma independiente se correlaciona de forma muy similar con C_4 .

2. Es más, en un principio $C_4 = 0.9Y - 0.25W + e_4$, y reemplazando $W = 0.9Y + e_3$ tenemos que $C_4 = 0.9Y - 0.25(0.9Y + e_3) + e_4 = 0.675Y + e'$, con $e' = e_4 - 0.25e_3$, por lo que finalmente C_4 es una combinación lineal de Y más una perturbación.

Con eso, es evidente que estas regresiones lineales simple están replicando una regresión lineal similar, solo que la de W sobre C tiene un término de error adicional. Pero no es casual que sus coeficientes parciales sean prácticamente los mismos (cerca de 0.6), y que sus coeficientes de determinación también. En ese sentido, la diferencia con el modelo múltiple está que los modelos simples no nos otorgan una visión de como afecta Y a C , o como W afecta a C por separado, pues no estamos controlado por su varianza compartida. En ese sentido, tenemos un *sesgo de variable omitida*, toda vez que supieramos que el modelo real se compone de $Y + Z$. Como indica Greene (2003) ese sesgo de variable omitida será cada vez mayor si es que las variable en el modelo y la que no está están fuertemente relacionadas. Esto se escribe como

$$E(\bar{\beta}_1) = \beta_1 + \delta_2 \cdot \beta_2$$

donde δ_2 es $(Y'Y)^{-1}YW$ y $\bar{\beta}$ es el vector de coeficientes de la regresión que omite la matriz de W , y β_1, β_2 son los vectores verdaderos de coeficientes de regresión que incluyen a Y, Z . Luego el sesgo de la variable omitida esta dado por

$$E(\bar{\beta}_1 - \beta_1|Y) = \delta_2 \cdot \beta_2$$

Ahora bien, no debemos olvidar que este modelo está mal especificado dado que Y y W son una combinación lineal funcional. Entonces, teóricamente si el modelo real solo incluye a Y , entonces no habría sesgo de variable omitida y el modelo estimado simple es el correcto.

Pregunta 11

Hacemos una la regresión de C_3 en Y y Z pero omite se omite la constante. Explique por qué obtiene esos resultados.

Cuadro 26: Modelos de regresión lineal

	Modelo sin intercepto	Modelo con intercepto
Y	0.900040*** (0.000914)	0.938488*** (0.031246)
Z	-0.254800*** (0.059985)	-0.250517*** (0.060071)
(Intercept)		-385.177890 (312.895288)
R ²	0.999971	0.481969
Adj. R ²	0.999971	0.480930
Num. obs.	1000	1000

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

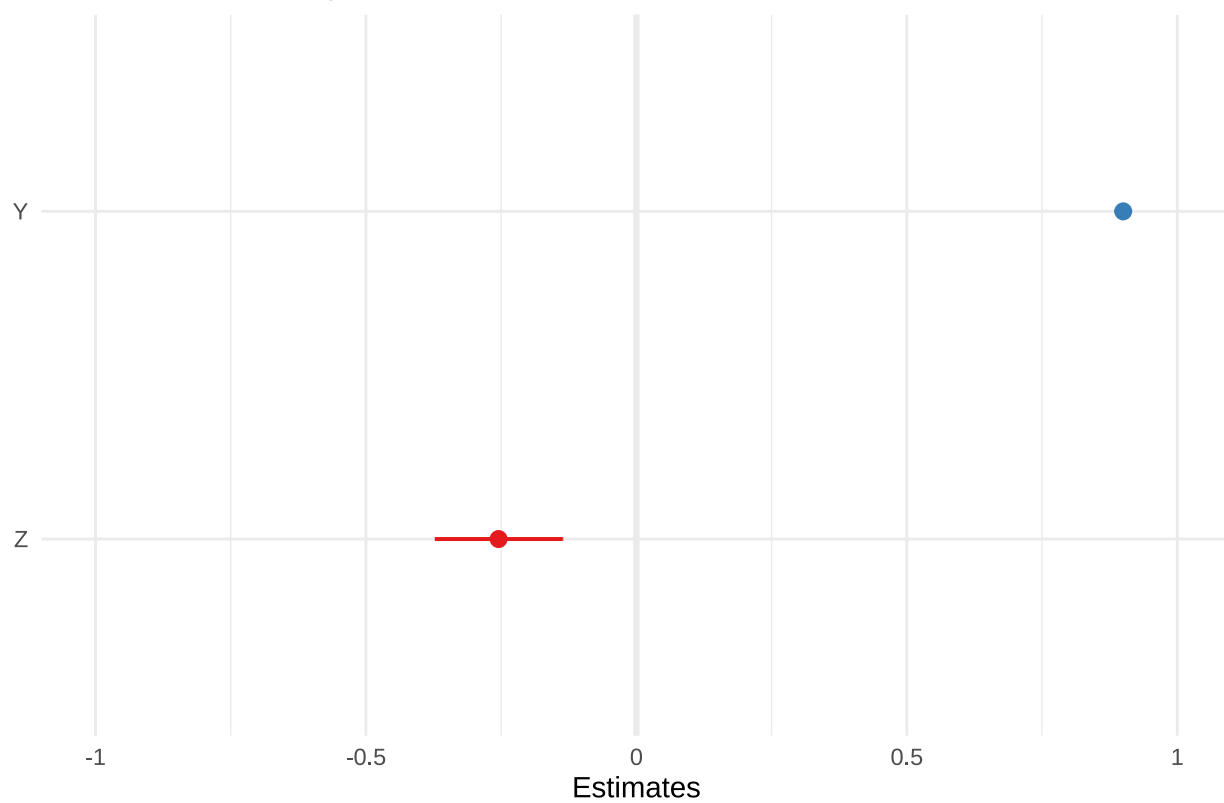
En términos descriptivos, se obtiene que la estimación asociada a Y es más precisa en el modelo sin constante que el modelo anterior con el intercepto. Esto se debe a que en el modelo con intercepto la estimación asociada no era significativa lo que hace que al no considerarla en el ajuste se obtengan mejores resultados en términos de significancia.

Ahora bien, profundicemos en esto. La pendiente en una regresión que contiene el término constante puede ser obtenida transformando los datos en su desviación promedio, y luego, regresando la variable dependiente en esta nueva forma de las variables independientes “desviadas” (*mean deviation form*, como dice Greene). Sea X una constante i en la primera columna de la matriz de variables aleatorias. La solución del intercepto β_0 en este caso es la pendiente de la regresión que contiene el término constante

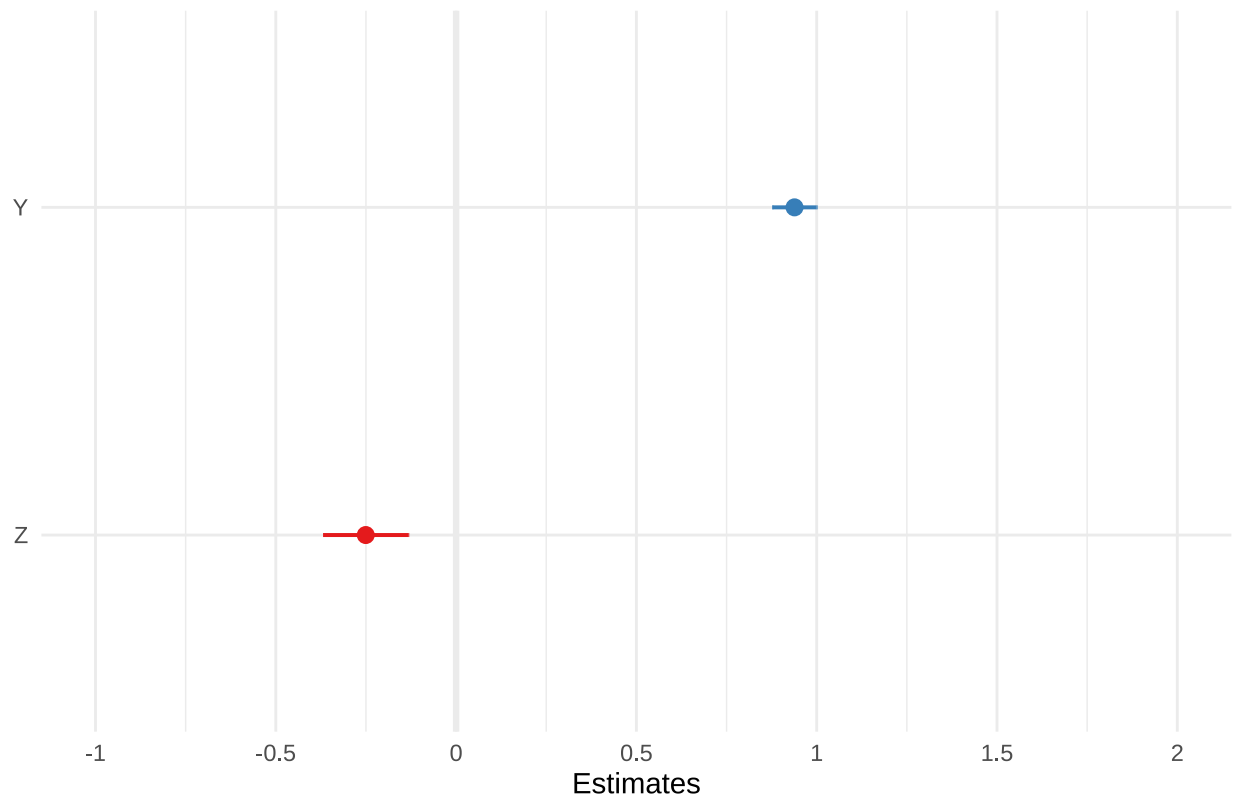
$$(1)X = X - i(i'i)^{-1}i'X(2)X = X - i\bar{X}(3)X = M^0X$$

Así, podemos notar que existirá una versión modificada del *residual marker*, por lo que si omitimos esto en la regresión, básicamente lo que se transforma no solo el $\beta_{1,2}$, sino que también la suma de residuos que se obtienen de *OLS*. Es por ello, que si se omite la pendiente, tendremos un nuevo ajuste del modelo, dado que este si es relevante para la *SSR*. Veamos como se ve gráficamente

Modelo sin intercepto



Modelo con intercepto



Pregunta 12

Se eliminaron las últimas 950 observaciones de su muestra. Haga las siguientes regresiones:

- (a) C en Y ;
- (b) C en Y_1 ;
- (c) C_1 en Y ;
- (d) C_3 en Y y Z ;
- (e) C_4 en Y y W

Explique por qué obtiene esos resultados.

Cuadro 27: Modelos de regresión - Muestra acotada

	(A)	(B)	(C)	(D)	(E)
(Intercept)	0.000000 (0.000000)	88.702792 (107.965672)	-175.016916 (1327.649911)	-292.429998 (1327.283749)	1547.297687 (1368.860046)
Y	0.900000*** (0.000000)		0.917437*** (0.132865)	0.925543*** (0.132829)	
Y1		0.891114*** (0.010805)			
W					0.578863*** (0.152052)
R ²	1.000000	0.992993	0.498324	0.502860	0.231918
Adj. R ²	1.000000	0.992847	0.487873	0.492503	0.215916
Num. obs.	50	50	50	50	50

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

En el cuadro 27 podemos ver un resultado muy interesante respecto a un aspecto que muchas veces confunde en los modelos de regresión. Podemos notar que las estimaciones sobreviven en términos de tamaño efecto y significancia, aún así cuando hemos reducido la muestra. Evidentemente no son *exactamente los mismo*, pues el costo que se tiene al reducir el *tamaño de muestra* es la pérdida de precisión. Así, si uno mira cifras significativas, ya al segundo decimal las estimaciones difieren (al menos desde el modelo (B) en adelante). Probablemente lo más notorio está en el ajuste del modelo, donde en algunos casos la variación es de 2%, en promedio. Empero, la idea que se quiere calcar es que, al menos en distribución, los parámetros de la regresión siguen siendo similares. La importancia de ello está en saber distinguir entonces, la importancia y diferencia entre Teoremas Centrales de Límite (TCL) y Leyes de Grandes Números (LGN). Este último ejercicio tiene que ver con TCL y no LGN.

A partir del Teorema Central del Límite de *Lindeber-Feller* podemos decir que la convergencia en distribución nos permite relacionar estadísticamente los parámetros de la muestra y la población. En particular podemos decir que sea cual sea la distribución de la población, si extraemos una muestra aleatoria de *gran tamaño* y que esa selección es independiente de sacar otra (no hay regla de selección), entonces la distribución del parámetro en la muestra tiende a una normal también. Si esto es así, entonces las estimaciones de los parámetros son estadísticamente equivalentes a las poblacionales. Lo interesante entonces, es que estas últimas regresiones nos dicen que 50 de tamaño de muestra es lo suficientemente grande para seguir obteniendo **en distribución** parámetros similares. Lo que perdemos eso sí, es indudablemente precisión pero al costo de ser más eficientes.

Parte 2

El código que generalos dos loops se pueden encontrar en el archivo *resultados__problem2.R*. Además, sus resultados están en *problem2.pdf*

Respecto al análisis podemos decir que un aspecto importante que discutíamos al finalizar la parte 1 tiene que ver con qué depende la distribución de los parámetros estimados de *OLS*.

La distribución de los parámetros estimados no depende del tamaño de la muestra. En las primeras tres láminas podemos notar que tenemos una distribución normal con tamaño de muestra 50, 100 y 500, y podemos notar cómo no cambia el valor esperado de los coeficientes. Esto nos dice que para tanto para 50 como para 500 de tamaño de muestra podemos hablar de una estimación que converge a un mismo valor estadísticamente. Como comentábamos anteriormente, la particularidad de esta **convergencia en distribución** no nos dice si este valor existe en el límite y que es, sino que nos permite relacionar estadísticamente los parámetros de la muestra y la población, aspecto muy relevante para la inferencia. Entonces, si la selección de la muestra es aleatoria, realmente entre elegir una muestra de 50 o 500, el *trade-off* del investigador será un clivaje entre eficiencia o precisión.

Si bien la convergencia no está dada por este tipo de teoremas, si aportan información muy importante para estimación de *OLS*. La razón es que no siempre tendremos muestras asintóticas, por lo que tendremos que restringir nuestro análisis a lo que si vemos. Los Teoremas Centrales de Límite de todas formas nos aportarán para saber como se comporta **en distribución** los parámetros, sus varianzas y los test asociados en muestra pequeña. Una pregunta importante es ¿es este análisis sólo válido para distribuciones normales? La respuesta es no. Los teoremas de *Lindeberg y Feller* y *Lindeberg y Levy* nos permiten decir que sea cual sea la distribución de la variable aleatoria, si extraemos una muestra lo suficientemente grande, y que esa selección sea independiente, entonces podemos decir que la media de la muestra de nuestros parámetros tiende a una distribución normal, y con un primer momento con media poblacional y un segundo momento de varianza finita. Entonces, la respuesta a si la distribución de los resultados depende de la forma de distribución de la variable aleatoria, la respuesta es que no (podemos notar esto fácilmente comparando los resultados para β_1 y β_2 , por ejemplo, en la página 1).

Respecto al error y su distribución, parece importante destacar un punto indicado en el *TCL de Lindeberg-Feller*. Al final, es difícil asegurar que la varianza sea la misma en cada muestra de variables aleatorias, lo que nos lleva tener distribuciones de error distinta en cada una de ellas. Pero, independiente de eso, tanto la teoría como los resultados de las figuras nos muestran que pese las distintas distribuciones de los errores, mientras sus varianzas sean finitas, entonces se puede demostrar que se cumple el principio de dominancia: esto es, que independiente de que algún valor se “dispare” y se produce una gran dispersión y error, eso no implica que la muestra no vaya a convergen en distribución a una normal. Más adelante cuándo discutamos las implicancias en LGN se entenderá mejor.

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(0, \sigma^2(X'X)^{-1})$$

De hecho, las propiedades de los estimadores en *OLS* en muestras pequeñas son una buena guía para obtener un rango de probables resultados de una muestra de datos. Eso nos dice que finalmente “*los cimientos*” del método estimación no tiene que ver en gran medida al tamaño de muestra. Es así como en un contexto asintótico, donde por ejemplo, las variables son estocásticas y por tanto los errores podrían ir cambiando en cada experimento que se haga, *OLS* no pierde sus propiedades fundamentales. Por ejemplo, las *Grenander* estableció condiciones para incluso en contextos tan “adversos” donde no se sabe como distribuye exactamente el error o si realmente el ponderador de las variables aleatorias converge a algo. Lo que dice estas condiciones es bastante razonable, e incluso posible de inferir de nuestros resultados:

1. La suma de cuadrados continua creciendo en la medida en que crece el tamaño de muestra

(más información se incorpora para estimar). En nuestra figura de la página 35 y 36 se nota claramente. La muestra aumentó y se obtuvo más precisión respecto a la distribución

2. No hay dominancia de los valores extremos o las observaciones individuales se hacen menos importantes. Por ejemplo, en la figura página 33 si bien se tienen un par de casos bajo el 0.9 la estimación de β no cambia.
3. La condición de rango completo se seguirá cumpliendo cuando la muestra aumenta de tamaño, es decir, agregar más muestra no implica que vayamos a perder información pues las observaciones son linealmente independientes (*prácticamente imposible*).

Entonces, así y todo, nuestros coeficiente seguirán siendo sesgados y consistentes. Entonces, ¿cuál es la condición o teoría límite en la que sin ella probablemente no podamos hacer todas las generalizaciones que hacemos con *OLS*? Esta es la **Ley de Grandes Números**, la que nos muestra las condiciones **necesarias y suficientes** para que el experimento no se desvíe del valor esperado o en otras palabras para que en el límite exista nuestro estimador, y que este sea un evento en la distribución que converge al valor verdadero.

Como podemos ver en las primeras 30 páginas, la distribución tiende a una normal con momentos bien definidos en términos de valor esperado, varianza, simetría y curtosis. Es decir, pese a que dos muestras puedan ser de tamaño 500, solo cuando aumentamos las repeticiones del experimento la distribución de los parámetros tiende a una normal.

Podemos también descartar el argumento sobre que este resultado se debe a la función de distribución de probabilidad del error. En línea con Kolmogorov 2 de la Ley Fuerte de Grandes Números podemos decir que a medida que el experimento se repite más veces, en promedio la desviación del estimador β respecto a su valor esperado cada vez será menor. Y esto, con probabilidad 1. Lo importante que nos aporta este teorema es que no solo si las variables no son idénticamente distribuidas (da igual que no sean *Bernoulli*, como proponía el Teorema de Borel), sino que solo basta con que los primeros momentos estén definidos y las variables aleatorias sean independientes.

Notemos que en la página 18-45 tenemos una distribución del error *Poisson* en la 21-52 una *Uniforme* y en la 9-36 una *Normal*, con distintos tamaños de muestra y el resultado es el mismo, lo único en común es que el experimento se ha repetido 500 veces. Con ello podemos notar que tanto β_1 como β_2 no solo convergen en todos al mismo valor, sino que lo más impresionante es que en la distribución se ve que el “valor verdadero” está en la zona más probable de la función de distribución de probabilidad (quizás donde menos se nota es en la uniforme). Así y todo, ambos coeficientes están dentro de los valores reales (una manera simple de verlo es verificar si la línea roja coincide o no con la distribución de $\beta_1, 2$). Para cerrar, no es trivial haber nombrado el par de soluciones para las distintas distribuciones del error, y esto es porque, inclusive si obviamos el término constante, las propiedades se mantienen. Como demostramos en la parte 1, esto solo re-escala la estimación en general, pero MCO sigue siendo MELI.

Referencias

Greene, W. H. (2003). Econometric analysis.

Wooldridge, J. M. (2017). Introductory econometrics: A modern approach