Examen – Soluciones
Econometría I: Magíster en Economía
Universidad de Santiago de Chile
Semestre 1, 2019

Nombre de alumna/alumno y firma:	
Pregunta 1 (Máximo 15 puntos)	
Pregunta 2 (Máximo 25 puntos)	
Pregunta 3 (Máximo 20 puntos)	

Instrucciones Generales

1. Tiene 90 minutos para responder a esta examen, dividida en 60 puntos.
2. Póngale nombre a todas las hojas.
3. Se permite el uso de calculadoras, siempre y cuando no cuenten con dispositivos de comunicación.
4. Las tablas de probabilidad necesarias están incluidas en la prueba. Si no sale la cantidad exacta de grados de libertad necesaria para un cálculo determinado, utilice la cantidad de grados de libertad más cercana.

Buena suerte!!

1. **Estimación de Parametros en un Modelo Lineal** [17 puntos]. Considere el modelo:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \qquad \text{para } i = 1, 2, \ldots, N,$$

donde se supone que $u_i \sim \mathcal{N}(0, \sigma^2 I_N)$, y donde I_N es una matriz de identidad de $N \times N$. Suponga que $E(x_k u_k) = 0$ para $k = \{1, 2, 3\}$, y que las observaciones i provienen de una muestra independiente e idénticamente distribuida. Plantea tres estimadores para el vector de parametros $\beta = (\beta_1, \beta_2, \beta_3)'$, específicamente el estimador de mínimos cuadrados ordinarios, el estimador de método de momentos, y el estimador de máxima verosimilitud. En esta pregunta es fundamental indicar:

   a) La función o sistema de ecuaciones que el estimador busca maximizar/minimizar/resolver,

   b) cómo el estimador interactua con los supuestos del modelo planteado anteriormente,

   c) y resuelva el estimador en cada caso (respectivamente $\widehat{\beta}_{MCO}$, $\widehat{\beta}_{MM}$ y $\widehat{\beta}_{MV}$)

   Note que no es necesario separar su respuesta en tres secciones (a, b y c). Solamente asegure que su respuesta cubre estos tres puntos.

   Respuesta: El estimador de MCO busca minimizar:

$$\begin{aligned} \widehat{\beta} &= \arg\min_{\beta} \sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} (y_i - x_i'\beta)^2 \\ &= \arg\min_{\beta} u'u = (y - X\beta)'(y - X\beta). \end{aligned}$$

   que es una función de pérdida definida por la suma de los errores al cuadrado. El estimador de MV busca maximizar:

$$\mathcal{L}(\beta, \sigma^2 | y, X) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(y - X\beta)^2}{2\sigma^2}.$$

   que es la función de verosimilitud dado los datos observados. Y el estimador de

MM busca resolver el siguiente sistema de ecuaciones:

$$\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - x_i\hat{\beta}_{MM}) = 0$$

que se basa en el supuesto de $E(xu) = 0$.

El estimador $\widehat{\beta}_{MCO}$ utiliza como componente central el supuesto de media condicionada nula $E(u|X) = 0$, y además requiere el hecho que X es de rango completo. MM requiere los mismos supuestos. El estimador de MV requiere media condicionada nula y que X se de rango completo, pero además requiere la independencia de las observaciones, y la normalidad del componente de error (y por ende y) para poder especificar la función de verosimilitud. **Nota de corrección: Aquí no se debe especificar que necesitamos $V(u|X) = \sigma^2 I$ o normalidad para MCO o MM, ya que estos supuestos solamente so ocupan para inferencia, no para estimación.**

En cada caso, los estimadores son:

$$\begin{aligned}
\frac{\partial u'u}{\partial \beta} &= \frac{\partial}{\partial \beta}(y - X\beta)'(y - X\beta) \\
&= \frac{\partial}{\partial \beta}(y'y - \beta'X'y - y'X\beta + \beta'X'X\beta) \\
&= \frac{\partial}{\partial \beta}(y'y - 2\beta'X'y + \beta'X'X\beta) \\
&= -2X'y + 2X'X\beta, \\
-2X'y + 2X'X\beta &= 0 \\
\Rightarrow X'X\beta = X'y & \\
\hat{\beta}_{MCO} &= (X'X)^{-1}X'y,
\end{aligned}$$

para MCO.

$$\hat{\beta}_{MM} = \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} x_i y_i = (X'X)^{-1}X'y,$$

para método de momentos, y:

$$\ell = -\frac{N}{2}ln2\pi - \frac{N}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}u'u$$

$$\begin{aligned}
\frac{\partial\ell(\beta,\sigma^2)}{\partial\beta} &= 0 \\
\frac{\partial(-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta))}{\partial\beta} &= 0 \\
\Rightarrow \frac{1}{\sigma^2}X'(y-X\beta) &= 0 \\
\beta &= (X'X)^{-1}X'y \qquad (1)
\end{aligned}$$

para MV.

2. **Interpretación y Selección de Modelos** [25 puntos]. Considere el siguiente modelo:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i \qquad \text{para } i = 1, 2, \ldots, N,$$

donde se supone que $u_i \sim \mathcal{N}(0, \sigma^2 I_N)$, y donde I_N es una matriz de identidad de $N \times N$. Suponga que $E(x_k u_k) = 0$ para $k = \{1, 2, 3, 4\}$, y que las observaciones i provienen de una muestra independiente e idénticamente distribuida. Además, $x_{1i} = 1$ para cada i (un vector de $N \times 1$ igual a 1 en cada celda), y x_{2i} es una variable binaria, que toma tanto valores de 0 y 1. Para una aplicación cuando $N = 10,000$, se estima un modelo de mínimos cuadrados ordinarios, reportando las siguientes estimaciones para el vector de parametros β y matriz de varianza-covarianza V:

$$\widehat{\beta} = \begin{pmatrix} 7 \\ 4 \\ 5 \\ 9 \end{pmatrix} \qquad \widehat{V}(\widehat{\beta}) = \begin{pmatrix} 5 & -1 & 2 & -4 \\ -1 & 4 & 1 & 2 \\ 2 & 1 & 2 & 3 \\ -4 & 2 & 3 & 6 \end{pmatrix}$$

Para las respuestas a continuación, los suguientes hechos pueden ser útiles: el percentil 97.5 de la distribución normal estandarizada es 1.96, el percentil 95 de la distribución normal estandarizada es 1.645, el percentil 95 de la distribución $\chi^2(1)$ es 3.84, el percentil 95 de la distribución $\chi^2(2)$ es 5.99, y el percentil 95 de la distribución $\chi^2(3)$ es 7.81. Pista: como la cantidad de observaciones N es grande, los valores críticos del χ^2 reemplazan los valores críticos de la distribución F, y los valores críticos de la normal estandarizada reemplazan los valores críticos de la distribución t.

a) (3 puntos) Interprete las coeficientes $\widehat{\beta}_1$ y $\widehat{\beta}_2$ en este modelo.

Solución: Aquí $\widehat{\beta}_1$ es el término de constante. Su interpretación es que cuando $x_2 i = x_{3i} = x_{4i} = 0$, estimamos que el promedio de y en la población es igual a $\widehat{\beta}_1$, o igual a 7 unidades. Y dado que x_{2i} es una variable binaria, interpretamos a $\widehat{\beta}_2$ como el estimador de la diferencia promedio de y para unidades cuya $x_{2i} = 1$ y aquellas unidades con $x_{2i} = 0$, todo lo demás constante. Es decir:

$$\beta_2 = E[y_i | x_{2i} = 1, x_{3i}, x_{4i}] - E[y_i | x_{2i} = 0, x_{3i}, x_{4i}].$$

En este caso, estimamos que la diferencia es de 4 unidades.

b) (3 puntos) Construya un intervalo de confianza de 95 % para el parametro β_3.

Solución: En este caso, sabemos que

$$z_k = \left(\frac{\hat{\beta}_k - \beta_k}{\sqrt{v_{kk}}} \right) \sim \mathcal{N}(0,1) \text{ para } k = 1, \ldots, K. \tag{2}$$

Entonces, para β_3, tenemos que $\left(\frac{\hat{\beta}_3 - \beta_3}{\sqrt{v_{33}}} \right)$ es aproximadamente normal con promedio 0, y desviación estándar 1. Ahora,

$$
\begin{aligned}
P(-1{,}96 < z_k < 1{,}96) &= 0{,}95 \\
P\left(-1{,}96 < \frac{\hat{\beta}_3 - \beta_3}{\sqrt{v_{33}}} < 1{,}96\right) &= 0{,}95 \\
P(-1{,}96\sqrt{v_{33}} < \hat{\beta}_3 - \beta_3 < 1{,}96\sqrt{v_{33}}) &= 0{,}95 \\
P(-1{,}96\sqrt{v_{33}} < \beta_3 - \hat{\beta}_3 < 1{,}96\sqrt{v_{33}}) &= 0{,}95 \\
P(\hat{\beta}_k - 1{,}96\sqrt{v_{33}} < \beta_3 < \hat{\beta}_3 + 1{,}96\sqrt{v_{33}}) &= 0{,}95 \qquad (3)
\end{aligned}
$$

y con los valores estimados de la pregunta, el intervalo de confianza de 95 % es

$$\hat{\beta}_3 \pm 1{,}96\sqrt{v_{33}} = 5 \pm 1{,}96 \times \sqrt{2} = [2{,}22; 7{,}77].$$

c) (7 puntos) Realice un test de forma matricial de $H_0 : \beta_3 + \beta_4 = 10$.

Solución: Escribimos la hipótesis nula de $\beta_3 + \beta_4 = 10$ como:

$$H_0 : H\beta = \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \beta_3 + \beta_4 = 10$$

Y utilizamos

$$\hat{\theta}_{MCO} = H\hat{\beta}_{MCO} = \hat{\beta}_3 + \hat{\beta}_4$$

y $\hat{V}(\hat{\theta}_{MCO}|X) = H\hat{V}(\hat{\beta}_{MCO}|X)H' =$

$$\begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{v}_{11} & \hat{v}_{12} & \hat{v}_{13} & \hat{v}_{14} \\ \hat{v}_{21} & \hat{v}_{22} & \hat{v}_{23} & \hat{v}_{24} \\ \hat{v}_{31} & \hat{v}_{32} & \hat{v}_{33} & \hat{v}_{34} \\ \hat{v}_{41} & \hat{v}_{42} & \hat{v}_{43} & \hat{v}_{44} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \hat{v}_{33} + 2\hat{v}_{34} + \hat{v}_{44}$$

Para hacer el test de $H_0 : \theta = \beta_3 + \beta_4 = 10$ contra la alternativa $H_1 : \theta =$

$\beta_3 + \beta_4 \neq 10$, construimos la estadística de prueba:

$$\begin{aligned} v &= \frac{1}{p}(\hat{\theta}_{MCO} - \theta^0)'[\hat{V}(\hat{\theta}_{MCO}|X)]^{-1}(\hat{\theta}_{MCO} - \theta^0) \\ &= \frac{(\hat{\beta}_3 + \hat{\beta}_4 - 10)^2}{\hat{v}_{33} + 2\hat{v}_{34} + \hat{v}_{44}} \\ &\sim \chi^2(1) \;\; \text{bajo } H_0 : \theta = \beta_3 + \beta_4 = 10. \end{aligned}$$

Y rechazamos la nula a un nivel de significancia de 5 % si la probabilidad de observar el valor de v se ubica más arriba del percentil 95 % de la distribución $\chi^2(1)$ (notamos que en la pregunta se indica que se puede utilizar la χ^2 como la aproximación asintfotica a la distribución F. Calculamos el valor observado de la estadística de prueba como:

$$v = \frac{(5 + 9 - 10)^2}{2 + 2 \times 3 + 6} = 1{,}143.$$

Como $1{,}143 < 3{,}84$, no rechazamos la nula que $\beta_3 + \beta_4 = 10$.

d) (4 puntos) Suponga que además sabe que se estimó una regresión de la siguiente forma, donde $\widehat{u}^2$ son los residuos al cuadrado:

$$\begin{aligned} \widehat{u}^2 &= 98{,}538 + 0{,}772 x_{2i} + 1{,}204 x_{3i} + 1{,}421 x_{4i} \\ & \quad\; (1{,}983) \;\; (1{,}391) \qquad (2{,}810) \qquad (1{,}390), \quad R^2 = 0{,}000155, \end{aligned}$$

en base a esta información, realice un test de Breusch-Pagan, e interprete sus resultados.

Solución: Aquí la clave es reconocer que podemos realizar el test conjunto: $H_0 : \gamma_2 = \gamma_3 = \gamma_4$ (donde γ es el coeficiente sobre cada regresor en la proyección de los residuos al cuadrado sobre los regresores) mediante un test F del modelo utilizando el R^2. Esto es con la estadística de prueba:

$$v = \left(\frac{N - K}{K - 1}\right)\left(\frac{R^2}{1 - R^2}\right),$$

y bajo la nula, v se distribuye aproximadamente $\chi^2(K - 1) = \chi^2(3)$. Por la tanto, calculamos el valor de la estadística de prueba como:

$$v = \frac{9996}{3} \times \frac{0{,}000155}{1 - 0{,}000155} = 0{,}52.$$

Como $0{,}52 < 7{,}81$, no rechazamos la nula $H_0 : \gamma_2 = \gamma_3 = \gamma_4$, y concluimos que no hay evidencia para sugerir heteroscedasticidad utilizando el test de Breusch-Pagan.

e) (5 puntos) Ahora, considere el caso en que $E(x_{4i}u_i) \neq 0$, pero existe otra variable z_{1i} que cumple con $E(z_{1i}u_i) = 0$ y $E(z_{1i}x_{4i}) \neq 0$. Plantee un estimador de variables instrumentales y de mínimos cuadrados en dos etapas que le permita estimar el vector de parametros β de forma consistente. ¿Cómo se comparan estos estimadores?

Solución: El estimador de variables instrumentales ($\widehat{\beta}_{IV}$) es:

$$\widehat{\beta}_{IV} = (Z'X)^{-1}(Z'y)$$

donde X es una matriz que consiste de las variables x_1, x_2, x_3 y x_4, y Z es una matriz que consiste de las variables x_1, x_2, x_3 y z_1. El estimador de MC2E consiste en utilizar $\widehat{x}_{4i}$ en vez de la variable endógena x_{4i}, donde $\widehat{x}_{4i}$ viene de una proyección de x_{4i} sobre z_{1i} (y cada otra variable exógena. Especificamente, $\widehat{\beta}_{MC2E}$ es:

$$\begin{aligned}
\widehat{\beta}_{MC2E} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y \\
&= [(Z(Z'Z)^{-1}Z'X)'(Z(Z'Z)^{-1}Z'X)]^{-1}(Z(Z'Z)^{-1}Z'X)'y \\
&= [(X'Z(Z'Z)^{-1}Z')(Z(Z'Z)^{-1}Z'X)]^{-1}(X'Z(Z'Z)^{-1}Z')y \qquad (4) \\
&= [X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y \\
&= [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y.
\end{aligned}$$

Nota de corrección: No es necesario plantear todo esto. Basta indicar la primera línea. Justamente en el caso de que la cantidad de variables endógenas es igual a la cantidad de instrumentos, el estimador $\widehat{\beta}_{IV}$ es igual al estimador $\widehat{\beta}_{MC2E}$. Como aquí tenemos que la cantidad de ambas es igual a 1, tenemos:

$$\begin{aligned}
\widehat{\beta}_{MC2E} &= [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y \\
&= (Z'X)^{-1}(Z'Z)(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'y \\
&= (Z'X)^{-1}(Z'Z)(Z'Z)^{-1}Z'y \\
&= (Z'X)^{-1}Z'y \\
&= \widehat{\beta}_{IV}.
\end{aligned}$$

f) (3 puntos) Suponga que también exista otra variable $z_{2i} = 2z_{1i}+1$, que por definición cumple con $E(z_{2i}u_i) = 0$ y $E(z_{2i}x_{4i}) \neq 0$. ¿Hay alguna manera de utilizar esta variable adicional para ganar poder y/o realizar contrastes que no se podrían realizar en la parte e)? ¿Por qué, o por qué no?
Solución: No. Tipicamente al tener más variables instrumentales, tenemos una situación de sobre-identificación, y ganamos varias cosas, incluyendo la posibilidad de hacer contrastes de sobre-identificación. Pero aquí es importante notar que z_{2i} es una función lineal exacta de z_{1i} y un término constante, y por ende NO agrega más información independiente por sobre z_{1i}.

3. **Otras Detalles Teóricas** [18 puntos]. Para cada pregunta a continuación responda apoyando su respuesta con las fórmulas y/o derivaciones necesarias.

a) (6 puntos) Considere el estimador de mínimos cuadardos ordinarios $\widehat{\beta}_{MCO} = (X'X)^{-1}X'y$. Bajo los supuestos clásicos con errores normales, cuál es la expectativa y la varianza de este estimador?
La expectativa de $\widehat{\beta}$ es:

$$\begin{aligned}
E(\widehat{\beta}_{MCO}|X) &= E(Ay|X) \\
&= AE(y|X) \\
&= AX\beta \\
&= (X'X)^{-1}X'X\beta \\
&= \beta
\end{aligned}$$

donde utilizamos el supuesto que $E(y|X) = X\beta$ (o que $E(u|X) = 0$). Y la varianza de $\widehat{\beta}$ es:

$$\begin{aligned} V(\widehat{\beta}|X) &= AV(y|X)A' = A(\sigma^2 I)A' = \sigma^2 AA' \\ &= \sigma^2 (X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

donde utilizamos el supuesto que $V(y|X) = \sigma^2 I$, y las propiedades de la varianza.

b) (6 puntos) Cuando existe error de medición aditiva en la variable independiente, ¿por qué se dice que existe un sesgo de atenuación en el estimador MCO (bajo los supuestos clásicos de errores en variables)?

Decimos que hay un sesgo de atenuación, porque bajo los supuestos clásicos, el valor estimado para $\widehat{\beta}$ *siempre* estará sesgado hacía cero, es decir siempre será más pequeño en magnitud, que el parametro verdadero. Esto se ve con los dos siguientes resultados:

$$\begin{aligned} &\text{plim}\widehat{\beta} = \tfrac{\beta}{1+(\sigma_e^2/\sigma_{x^*}^2)} < \beta \quad \text{para } \beta > 0 \text{ y } \sigma_e^2 > 0 \\ &\text{plim}\widehat{\beta} = \tfrac{\beta}{1+(\sigma_e^2/\sigma_{x^*}^2)} > \beta \quad \text{para } \beta < 0 \text{ y } \sigma_e^2 > 0 \end{aligned}$$

c) (6 puntos) ¿Cómo se compara la varianza asintótica de $\widehat{\beta}_{MCO}$ robusto a heteroscedasticidad con la varianza asintótica de $\widehat{\beta}_{MCO}$ clusterizada? ¿Cuál proporcione inferencia más robusto a factores idiosincráticos que impacta la varianza/correlación entre observaciones?

El estimador de la varianza asintótica robusto a heteroscedasticidad para $\widehat{\beta}$ es:

$$\widehat{avar}(\widehat{\beta}) = (X'X)^{-1}\left(\sum_{i=1}^{2} \hat{u}_i^2 x_i x_i'\right)(X'X)^{-1}$$

y el estimador de la varianza asintótica clusterizada es:

$$\widehat{avar}(\widehat{\beta}) = (X'X)^{-1}\left(\sum_{c=1}^{C} X_c'\widehat{u}_c\widehat{u}_c' X_c\right)(X'X)^{-1}.$$

La segunda varianza es más robusta a mis-specificaciones del modelo, ya que permite que la varianza no es común entre individuos (al igual a la varianza robusto a heteroscedasticidad), pero también permite una correlación no cero del término u para observaciones i dentro del mismo cluster c, algo que la varianza robusto a heteroscedasticidad no permite.