Examen Final
Econometría I: Magíster en Economía
Universidad de Santiago de Chile
Semestre 1, 2017

Nombre de alumna/alumno y firma:	
Pregunta 1 (Máximo 30 puntos)	
Pregunta 2 (Máximo 30 puntos)	

Instrucciones Generales

1. Tiene 120 minutos para responder a este examen, dividida en 60 puntos.
2. Póngale nombre a todas las hojas.
3. **No se puede realizar preguntas acerca de la interpretación del enunciado del examen.** Si hay alguna duda después de haber leído el anunciado, en su respuesta explicite cualquier supuesto adicional que cree que es necesario tomar para responder a la pregunta.
4. Se permite el uso de calculadoras, siempre y cuando no cuenten con dispositivos de comunicación.
5. En caso de copia, se sancionará de acuerdo a lo estipulado en el programa del curso.

Buena suerte!!

1. **El Modelo Lineal Clásico** [30 puntos] Consideremos el modelo clásico de regresión lineal con errores normalmente distribuidas:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i. \tag{1}$$

Aquí $y \sim \mathcal{N}(X\beta, \sigma^2 I)$, y se cuenta con una muestra de N observaciones independientes $i = 1, \ldots, N$. Por lo tanto, cada variable $(y_i, x_{1i}, x_{2i}, x_{3i})$ es un vector de $N \times 1$. La variable x_{1i} es un vector con cada elemento igual a 1, de modo que β_1 es un término de intercepto.

   (a) [5 puntos] Proponga un estimador insesgado $\widehat{\beta}$ para el vector de parametros β, y demuestre que este estimador es insesgado.
   Solución: La opción más fácil es utilizar MCO: $\hat{\beta} = (X'X)^{-1}X'y$. Este estimador es insesgado dado que la expectativa condicional de $\hat{\beta}_{MCO}$ es:

$$E(\hat{\beta}_{MCO}|X) = AE(y|X) = AX\beta = (X'X)^{-1}X'X\beta = \beta.$$

   Se puede demostrar que también $E(\hat{\beta}_{MCO}) = \beta$ sin condicionar con la ley de expectativas iteradas (no es necesario para recibir todos los puntos).

   (b) [5 puntos] Considere la hipótesis nula $\beta_2 + \beta_3 = 2$, y suponga que se ha estimado ecuación 1 por MCO, resultando en un vector de parametros $\widehat{\beta}_{MCO}$, y matriz de varianza covarianza asociado con los parametros $\widehat{V}(\widehat{\beta}_{MCO})$. Describa cómo realizar el test múltiple de forma matricial utilizado una matriz H que permite aislar los componentes relevantes del vector de parametros y la matriz de varianza-covarianza del modelo. *Pista*: Se sugiere escribir la estadística de prueba para realizar el test de hipótesis, y después explicar cómo se puede formar esta estadística utilizando la matriz H propuesta.
   Solución: Escribimos la hipótesis nula de $\beta_2 + \beta_3 = 2$ como:

$$H_0 : H\beta = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \beta_2 + \beta_3 = 2$$

   Y utilizamos

$$\hat{\theta}_{MCO} = H\hat{\beta}_{MCO} = \hat{\beta}_2 + \hat{\beta}_3$$

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

$$\begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{v}_{11} & \hat{v}_{12} & \hat{v}_{13} \\ \hat{v}_{12} & \hat{v}_{22} & \hat{v}_{23} \\ \hat{v}_{13} & \hat{v}_{23} & \hat{v}_{33} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \hat{v}_{22} + 2\hat{v}_{23} + \hat{v}_{33}$$

Para hacer el test de $H_0 : \theta = \beta_2 + \beta_3 = 2$ contra la alternativa $H_1 : \theta = \beta_2 + \beta_3 \neq 2$, construimos la estadística de prueba:

$$\begin{aligned} v &= \frac{1}{p}(\hat{\theta}_{MCO} - \theta^0)'[\hat{V}(\hat{\theta}_{MCO}|X)]^{-1}(\hat{\theta}_{MCO} - \theta^0) \\ &= \frac{(\hat{\beta}_2 + \hat{\beta}_3 - 2)^2}{\hat{v}_{22} + 2\hat{v}_{23} + \hat{v}_{33}} \\ &\sim F(1, N-K) \text{ bajo } H_0 : \theta = \beta_2 + \beta_3 = 2. \end{aligned}$$

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 $F(1, N-K)$.

(c) [5 puntos] Demuestre que se puede re-parametrizar el modelo para comprobar la nula descrita en la parte (b) considerando un solo parametro en una regresión.

Solución: [Ésta no es la única manera de resolver la pregunta] Queremos realizar el test: $H_0 : \beta_2 + \beta_3 = 2 \leftrightarrow \beta_2 + \beta_3 - 2 = 0$. Entonces, para poder escribir un modelo que nos permite comprobar la nula directamente, primera restamos $2x_{3i}$ de ambos lados:

$$y_i - 2x_{3i} = \beta_1 + \beta_2 x_{2i} + (\beta_3 - 2)x_{3i} + u_i.$$

Ahora, restamos $\beta_2 x_{3i}$ del segundo término y lo sumamos al tercer término para tener:

$$y_i - 2x_{3i} = \beta_1 + \beta_2(x_{2i} - x_{3i}) + (\beta_2 + \beta_3 - 2)x_{3i} + u_i.$$

Esto sugiere que para comprobar la prueba directamente en una regresión, tenemos que estimar la regresión de $(y_i - 2x_{3i})$ sobre un constante, $(x_{2i} - x_{3i})$, y x_{3i}, y consideramos si el término sobre x_{3i} podría ser igual a cero.

(d) [5 puntos] Ahora, suponga que los errores no son identícamente distribuidas con varianza σ^2 para cada i, sino distribuida según $\sigma^2 x_{3i}$. Describa cómo se podría realizar un test de Mínimos Cuadrados Generalizados Factibles para tomar en cuenta esta heteroscedasticidad. Solución:

- Formamos $y_i^* = \frac{y_i}{\sqrt{x_{3i}}}$ y $x_i^* = \frac{x_i}{\sqrt{x_{3i}}}$ $\forall k = 1, \ldots, K$
- Ahora, con el modelo transformado $y_i^* = x_i^{*\prime}\beta + u_i^*$, tenemos $V(y_i^*|X) = \frac{\sigma_i^2}{x_{3i}} = \frac{\sigma^2 x_{3i}}{x_{3i}} = \sigma^2$ para cada $i = 1, \ldots, N$
- Y ahora, el modelo transformado cumple con homoscedasticidad, y lo estimamos utilitando MCO
- Los parametros estimados serán consistentes y asintóticamente eficientes
- Notemos que la transformación da más peso a observaciones cuya varianza es menor

(e) [6 puntos] Considerando el modelo lineal de regresión, explique brevemente la diferencia entre los errores estándares robustos a heteroscedasticidad, robustos a clusterización, y los errores estándares clásicos. Apoye su respuesta con fórmulas relevantes, y expliqué en qué circunstancias los errores estándares estimados se difieren entre sí.

Solución: La diferencia entre los tres tipos de errores estándar es el supuesto que cada tipo implica entre la relación entre no-observables en el modelo. Errores estándares tradicionales suponen que cada observación es independente y el término u es idénticamente distribuida. Errores estándar robustos a heteroscedasticidad mantienen el supuesto de independencia, pero permiten que la varianza de cada no observables es distinto. Por último, errores estándares clusterizadas permiten dependencia adentro (pero no entre) clusters. Errores estándares clusterizadas son idénticas a errores estándares robustos a heteroscedasticidad solo si cada observación es su propio cluster, y los dos serán idénticas a errores estándares normales si cada observación tiene la misma varianza. Los errores estándares tradicionales se estiman a partir de la siguiente matriz de varianza covarianza: $\widehat{V}(\hat{\beta}_{MCO}|X) = \widehat{\sigma}^2(X'X)^{-1}$. Los de heteroscedasticidad son: $\widehat{avar}(\hat{\beta}_{MCO}) = (X'X)^{-1}\left(\sum_{i=1}^N \hat{u}_i^2 x_i x_i'\right)(X'X)^{-1}$, y los clusterizados son: $\widehat{V}/C = (X'X)^{-1}\left(\sum_{c=1}^C X_c'\widehat{u}_c\widehat{u}_c' X_c\right)(X'X)^{-1}$

(f) [4 puntos] Describe como se puede estimar el coeficiente $\widehat{\beta}_2$ mediante una "regresión de residuos", donde primero se quitan el impacto de x_{1i} y x_{3i} del modelo de interés. *Nota*: No es necesario escribir la fórmula de $\widehat{\beta}_2$, sino describir el proceso a seguir en este caso.

Solución: Primero, hacemos una regresión de y_i sobre x_{1i} y x_{3i}, y formamos los residuos de MCO $\widehat{r}_y$. Después hacemos una regresión de x_{2i} sobre x_{1i} y

x_{3i} y formamos los residuos $\widehat{r}_x$. Por último, estimamos la regresión:

$$\widehat{r}_y = b\widehat{r}_x + \eta_i$$

por MCO (donde esta regresión no tiene coeficiente). Aquí el parametro estimado $\widehat{b}$ es idéntica al parametro $\widehat{\beta}_2$ en el modelo original.

2. **Aplicación Empírica** [30 puntos] Considere la siguiente ecuación, donde se quiere estimar la relación causal entre más años de escolaridad en la población adulta con los níveles de salud de cada persona:

$$salud_i = \alpha + \beta educacion_i + u_i. \tag{2}$$

Suponga que se cuenta con una muestra representativa e independiente de $N = 1.000$ observaciones de la población adulta. Para cada observación i, se cuenta con la cantidad total de educación final de la persona en años ($educacion_i$), y una medida de salud que es un puntaje continuo asignado en base a un control de salud ($salud_i$). Suponga que ambas variables están medidas sin error, y para simplicidad suponga que la relación correcta entre salud y educación es lineal.

   (a) [6 puntos] ¿Cuál(es) problema(s) podría(n) existir al tratar de inferir causalidad en la ecuación 2 si se estima con MCO? Explique el/los problema/s en términos de las variables en la ecuación 2.
   Solución: Bajo lo estipulado, el próblema más grave que queda es endogeneidad por variables omitidas (o simultaneadad de la relación observada). Probablemente habrán muchas variables que están omitidas que están relacionados tanto con educación como salud (por ejemplo acceso a información,...), y esto va a provocar sesgos (e inconsistencia) en los parametros estimados.

   (b) [5 puntos] ¿Qué tipo de sesgo podría provocar el/los problema/s descrito/s en la parte (a)? ¿Se podrá inferir algo de la dirección del sesgo?
   Solución: Si podemos saber la dirección de correlación entre las variables omitidas y salud y educación por lo menos vamos a poder inferir la dirección del sesgo (pero no el tamaño). Aunque es más díficil dado que hay más de una variable omitida, podemos afirmar la dirección del sesgo si estamos dispuestos a considerar todas las variables omitidas como una grande variable omitida. Por ejemplo, imaginemos que pensamos que en términos globales el impacto promedio de los omitidos sobre educación es positiva (eg acceso a información, ingreso, capacidad de procesar información, etc) y también está positivamente relacionado con salud. Entonces, definiendo como W un indice de todos los factores omitidos relevantes tenemos que:

$$p \lim_{N\to\infty} \widehat{\beta} = \beta + (p \lim_{N\to\infty} \widehat{\delta})\beta_W$$

   donde δ es la relación entre el no observable W y educación, y β_W es su relación con salud. Entonces, si los dos son positivos, habrá un sesgo positivo en el estimador (es decir, sobre-estimamos β).

(c) [6 puntos] En vez de estimar la ecuación 2 con MCO, se sugiere estimar con mínimos cuadrados en dos etapas (MC2E), utilizando como instrumentos para la educación de cada individuo dos variables: $educM_i$ y $educP_i$: la educación de la madre y el padre (o los dos cuidadores principales) de cada individuo. Escriba el nuevo modelo de MC2E (ambas ecuaciones) que se necesitará para estimar el parametro β, y la fórmula del estimador de MC2E que resulta de estas ecuaciones.

Solución: El sistema de ecuaciones es:

$$salud_i = \alpha + \beta educacion_i + u_i \quad (3)$$

$$educacion_i = \pi_1 + \pi_2 educM_i + \pi_3 educP_i + \nu_i \quad (4)$$

donde la primera línea es la segunda etapa, y la segunda linea es la primera etapa. El estimador de mínimos cuadrados en dos etapas es:

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

(d) [4 puntos] ¿Cuáles son los supuestos requeridos para la consistencia del estimador de MC2E? Explique los supuestos en términos de las variables utilizadas en el modelo.

Solución: Estimación mediante variables instrumentales requiere los supuestos de relevancia y validez. El supuesto de relevancia dice que los instrumentos deben estar correlacionados con la variable endógena y el supesuto de validez dice que los instrumentos no deberían estar correlacionados con el término omitido u. En este caso, el supuesto de relevancia implica que la educación de los p/madres de una persona debe estar correlacionado con su propia educación, y el supuesto de relevancia dice que la educación de los p/madres no debe estar correlacionado con no observables importantes en la segunda etapa entre salud y educación.

(e) [5 puntos] ¿Estos supuestos son razonables en este contexto? ¿Por qué o por qué no?

Solución: El supuesto de relevancia es muy razonable ya que (en promedio) hay un alto componente inter-generacional de educación. Y de hecho, se puede comprobar este supesuto en los datos estimando la primera etapa, donde probablemente pensaríamos que $\pi_1 > 0$ y $\pi_2 > 0$. El segundo supuesto parece mucho menos razonable. Si estamos preocupados que hay no observables importantes correlacionados con la educación de una persona ¿por qué no deben estar correlacionados con la educación de sus p/madres? Como un ejemplo muy simple, si las madres con más educación tienen más información acerca de distintos tipos de tratamientos de salud, y además más información acerca de opciones educacionales, el parametro β seguirá siendo inconsistente incluso cuando estimamos con MC2E.

(f) [4 puntos] ¿El estimador resultante de la estimación de MC2E ($\widehat{\beta}_{MC2E}$) es un estimador insesgado de β? ¿Por qué o por qué no?

Solución: Un estimador de variables instrumentales *nunca* es insesgado. Solo puede ser consistente. Si nos fijamos en la segunda etapa del estimador de MC2E, utilizamos el valor predicho para la variable endógena, y se estima la siguiente ecuación:

$$y = \widehat{X}\beta + (u + (X - \widehat{X})\beta).$$

El insesgadez (muestras pequeñas) viene del segundo término del término de error: $(X - \widehat{X})\beta$. Este término solo desaparece asintóticamente.