Examen
Econometría I: Magíster en Economía
Universidad de Santiago de Chile

1. **Teoría** [20 puntos] Para esta pregunta consideramos un modelo (poblacional) de regresión lineal de la siguiente forma:

$$y_i = \beta_1 + \beta_2 x_{2i} + u_i, \tag{1}$$

donde $i = 1, \dots, N$.

   (a) (5 puntos) ¿Cómo se puede formar un estimador de Mínimos Cuadrados Ordinarios para los parametros β_1 y β_2? Se explicito en términos de la función que se resuelve al estimar con MCO.
   Respuesta: Se minimiza la suma de los errores al cuadrado:

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

   El vector de parametros $\hat{\beta}$ ($\hat{\beta}_1$ y $\hat{\beta}_2$) proveniente de esta ecuación es el estimador MCO.

   (b) (6 puntos) Si el modelo poblacional (1) es un modelo correcto para el proceso de interés ¿por qué se eligiría estimar con MCO en vez de un estimador lineal alternativa? Se sugiere apoyar su respuesta con una fórmula o derivación relevante.
   Respuesta: El estimador es el mejor estimador lineal insesgado, y por ende es el estimador más eficiente en la clase de estimadores lineales (además de ser un estimador insesgado). Se puede demostrar formalmente que el estimador es el mejor estimador lineal insesgado utilizando una derivación como la derivación en las láminas 25 y 26 de clase VII: A.

   (c) (6 puntos) Ahora, imagine que la variable independiente x_{2i} está medido con error, y en vez de observar x_{2i} se observe $k_i = x_{2i} + \nu_i$. El error de medición introducida por ν_i es homoscedásitic, no está correlacionada con x_{2i}, y tampoco con el error del modelo u_i. Derive el sesgo en el parametro estimado $\hat{\beta}_2$ si se estima utilizando la variable observada k_i en vez de la variable de interés x_{2i}. **Respuesta: Sabemos que el estimador MCO si estimamos con k_i es:**

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'y = \frac{\sum_{i=1}^{N} k_i y_i}{\sum_{i=1}^{N} k_i^2} = \frac{\frac{1}{N}\sum_{i=1}^{N} k_i y_i}{\frac{1}{N}\sum_{i=1}^{N} k_i^2}.$$

Ahora, como revisado en la clase VIII B (endogeneidad), introducimos $k = x_{2i} + \nu_i$:

$$p \lim_{N\to\infty} \hat{\beta}_{MCO} = \frac{p\lim \frac{1}{N}\sum_{i=1}^{N}(x_{2i}+\nu_i)(k_i\beta + u_i)}{p\lim \frac{1}{N}\sum_{i=1}^{N}(x_{2i}+\nu_i)^2}$$

$$= \frac{\left(p\lim \frac{1}{N}\sum_{i=1}^{N} x_{2i}^2\right)\beta + p\lim \frac{1}{N}\sum_{i=1}^{N} x_{2i}u_i + \left(p\lim \frac{1}{N}\sum_{i=1}^{N} x_{2i}\nu_i\right)\beta + p\lim \frac{1}{N}\sum_{i=1}^{N} u_i\nu_i}{p\lim \frac{1}{N}\sum_{i=1}^{N} x_{2i} + 2p\lim \frac{1}{N}\sum_{i=1}^{N} x_{2i}\nu_i + p\lim \frac{1}{N}\sum_{i=1}^{N} \nu_i^2}$$

$$= \frac{E(x_{2i}^2)\beta + E(x_{2i}u_i) + E(x_{2i}e_i)\beta + E(u_i\nu_i)}{E(x_{2i}) + 2E(x_{2i}\nu_i) + E(\nu_i^2)} = \frac{E(x_{2i}^2)\beta + +0 + 0 + 0}{E(x_{2i}^2) + 0 + E(\nu_i^2)}$$

$$= \left(\frac{\sigma_{x_2}^2}{\sigma_{x_2}^2 + \sigma_\nu^2}\right)\beta = \frac{\beta}{1 + (\sigma_\nu^2/\sigma_{x_2}^2)} \neq \beta \qquad \textbf{si } \sigma_\nu^2 > 0$$

Entonces, el sesgo será: $\beta - \frac{\beta}{1+(\sigma_\nu^2/\sigma_{x_2}^2)}$

(d) (3 puntos) Imagine, además que se sabe que hay poca variación de la variable x_{1i}, y se sospecha que el error de medición es grande. ¿A qué conclusión se puede llegar acerca del sesgo que derivaste en la parte 1(d)? ¿Por qué?

Respuesta: Mientrás más grande el error de medición, y más

pequeño la variación en la variable x_{2i}, más grande será el sesgo. Especfícamente, es un sesgo de atenuación, y por lo tanto el valor estimado sería cada vez más cerca a cero. Esto es porque en vez de estar estimando con señal que viene de la variable verdadera x_{2i}, se utiliza más del ruído que viene de ν_i

2. **Aplicación Empírica** [25 puntos] Para un análisis del efecto de insumos sobre resultados educativos, contamos con una base de datos representativa de los puntajes en una evaluación de 584 alumno/as de la educación básica, y varias variables de interés. Se estima una serie de regresiones de los puntajes que cada estudiante saca en la evaluación (que varia entre 0 y 100) sobre las siguientes variables:

   - Profesor(a) tiene post-titulo = 1 si la profesor(a) lo tiene, y cero si no
   - Profesor(a) tiene magíster = 1 si la profesor(a) lo tiene, y cero si no
   - Tamaño de sala de clases = número de estudiantes en la sala
   - Inversión por estudiante = candidad total gastado por el colegio en cada alumno/a (medido en unidades de \$10.000)

   La tabla 1 presenta los resultados de cuatro regresiones distintas utilizando estas variables, y la variable dependiente "puntaje".

   (a) (6 puntos) En la columna (1), se estima la siguiente regresión:

   $$puntaje_i = \beta_1 + \beta_2 ProfPostTitulo_i + \beta_3 ProfMagister_i + \beta_4 TamSala_i + u_i$$

   Utilizando los valores de la tabla 1, interprete el parametro $\hat{\beta}_3$ y el parametro $\hat{\beta}_4$. Explica brevemente cómo se interpretan los parametros en términos econométrics, y cuál es su implicancia en el mundo real.
   Respuesta: La coeficiente estimada $\hat{\beta}$ da el efecto de la variable independiente sobre la variable dependiente. Es el cambio en unidades de la variable dependiente cuando se aumenta un una unidad la variable independiente manteniendo fijas todas las

Tab. 1: Disponibilidad de Recursos y Rendimiento Escolar

	(1) Puntaje	(2) Puntaje	(3) Puntaje	(4) Puntaje
Constante	66.301***	66.593***	39.380***	39.479***
	[1.479]	[1.461]	[9.688]	[9.703]
Profesor(a) tiene post-titulo	0.605			0.709
	[1.269]			[1.262]
Profesor(a) tiene magíster	1.911			1.705
	[1.536]			[1.528]
Tamaño de Sala de Clases	-1.857***	-1.861***	-0.772**	-0.784**
	[0.070]	[0.069]	[0.390]	[0.390]
Inversión por estudiante			0.541***	0.534***
			[0.191]	[0.191]
Observaciones	584	584	584	584
R-Cuadrado	0.90865	0.90838	0.90964	0.90986

Nota: cada columna representa una regresión diferente del puntaje en la evaluación sobre las variables incluidas en el modelo. En cada columna, se presenta la estimación de cada coeficiente de regresión, y abajo en parentesis su error estándar.

otras variables en el modelo. β_3 sugiere que un(a) profesor(a) que tiene magíster entrega en promedio 1,9 puntos más que un(a) profesora sin magíster y postítulo, pero no es un efecto significativo. β_4 sugiere que por cada estudiante adicional en la sala de clases, el rendimiento promedio cae (de forma estadísticamente significativa) en 1,85 puntos (ambos valores de la columna 1).

(b) (8 puntos) Imagine que en el modelo escrito arriba y estimado en la columna (1) que se quiere testear la hipótesis de que **la suma** de los dos parametros que refieren al nivel de educación del/la profesor(a) son iguales a cero. Realiza este test utilizando un test F, y la notación matricial visto en clases. Asegure de escribir la nula, calcular la estadística de prueba, y concluir si se debe rechazar o no recahzar la nula. [*pista:* Recuerde que la estadística de prueba $v = \left(\frac{1}{p}\right)(\hat{\theta}_{MCO} - \theta^0)'[\hat{V}(\hat{\theta}_{MCO}|X)]^{-1}(\hat{\theta}_{MCO} - \theta^0)$ está distribuída según un F con p y $N-K$ grados de libertad, donde p son la cantidad de restricciones en el modelo.] La matriz de varianza covarianza $\hat{V}(\hat{\beta}_{MCO})$ de este modelo es:

$$\begin{pmatrix} & \beta_2 & \beta_3 & \beta_4 & \beta_1 \\ \beta_2 & 1.610 & & & \\ \beta_3 & .1968 & 2.358 & & \\ \beta_4 & -.0060 & .0067 & .0048 & \\ \beta_1 & -.0834 & -.3420 & -.0982 & 2.1882 \end{pmatrix}$$

Respuesta: Escribimos la hipótesis nula de $\beta_2 + \beta_3 = 0$ como:

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

Y utilizamos

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

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

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

Para hacer el test de $H_0 : \theta = \beta_2 + \beta_3 = 0$ contra la alternativa $H_1 : \theta = \beta_2 + \beta_3 \neq 0$, construimos la estadística de prueba:

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

dado que $p = 1$. Ahora, el valor de v es: $\frac{(0.605+1.911)^2}{1.610+2.358+2\times0.1968} = 1.451$, y el valor crítico para un F con 1 y 580 grados de libertad es (según la tabla) aproximadamente 3.84. Por lo tanto, no se rechaza la nula a un nivel de significancia de 5%.

(c) (7 puntos) Ahora, consideremos los modelos de las columnas (2) y (3). En ambos modelos se incluye la variable “tamaño de sala de clase” pero las coeficientes son diferentes en cada columna. Cómo se puede explicar que el valor estimado en la columna 2 es difierente al valor de columna 3? A partir de los cambios de coeficientes y la fórmula del sesgo de variables omitidas, infiere la relación que existe entre las variables “tamaño de sala de clase” y “inversión por estudiante”.

Respuesta: El hecho de que las estimaciones son diferentes implica que (a) inversión es relevante en el modelo, y (b) inversión está correlacionada con el tamaño de la sala de clases. De la fórmula de variables omitidas tenemos que:

$$p \lim_{N\to\infty} \widehat{\beta}_4 = \beta_4 + (p \lim_{N\to\infty} \widehat{\delta})\beta_5$$

donde δ viene de la regresión: $inversion_i = tamSala_i\delta + e_i$, y $\hat{\beta}_4$ es el parametro estimado en columna dos. Comparando entre las dos columnas, vemos que cuando se estima sin inversión, el parametro es sesgado hacía abajo (más negativa). Como parece razonable (de columna 3) pensar que $\beta_5 > 0$, de esto podemos inferir que $\delta < 0$ explicando el sesgo negativo estimado en columna 2.

(d) (3 puntos) ¿Cómo se puede testear si los errores estándares de los modelos estimados aquí son heteroscedásticos? Nombre y explique un test formal para comprobar esto.

Respuesta: Se debe nombrar y explicar el test de White o el test de Breusch Pagan.

(e) (1 punto) Si se encuentra que los errores no son homoscedásticos, ¿cuál es una técnica que se puede aplica para hacer inferencia válida en la presencia de heteroscedasticidad?

Respuesta: Se acepta errores estándares robustos a heteroscedasticidad, o MCGF.

3. **Varias Preguntas de Respuesta Corta (6 líneas o menos)** [15 puntos]

(a) (3 puntos) En la ecuación:

$$salario_i = \beta_0 + \beta_1 educacion_i + \beta_2 experiencia_i + \beta_3 experiencia_i^2 + u_i$$

la variable $educacion_i$ probablemente está relacionado con muchas variables omitidas en el termino de error u_i. ¿Le parece razonable utilizar la educación de la madre de cada persona como variable instrumental para su educación?

Respuesta: Posiblemente... Para ser un buen instrumento neceistamos que la variable instrumental (educación de madre) está correlacionado con la variable endógena, pero no correlacionado con u_i. Parece bastante razonable suponer que la educación de la madre va a estar correlacionado con la educación del niño/a (es

decir el instrumento es relevante). Sin embargo, posiblemente la educación de la madre también va a estar correlacionada con u_i. La validez del instrumento depende de esta correlación.

(b) (3 puntos) ¿Cómo funciona el test de razón de verosimilitudes?

Respuesta: Se estima dos modelos de maxima verosimilitud: uno no restringido y otro restringido donde se restringe a los parametros a ser iguales a la cantidad estipulada en la hipótesis nula. La verosimilitud no restringido sería mayor a igual a la verosimilitud restringido, siendo igual solo en el caso de que la nula es cierto. El test entonces consiste en comprobar formalmente si la verosimilitud no restringida es estadísticamente mayor que la verosimilitud restringida al nivel deseado.

(c) (3 puntos) En el siguiente modelo, PIB refiere al Proucto Interno Bruto de Chile, $cobrePeso$ refiere al precio de cobre en el mercado internacional medido en pesos, y $cobreDolares$ refiere al precio de cobre en el mercado internacional medido en dolares.

$$PIB_t = \beta_1 + \beta_2 cobrePeso_t + \beta_3 cobreDolares_t + \cdots + u_i$$

¿Qué problema habrá al estimar este modelo? ¿Cómo se podría resolver este problema?

Respuesta: Multicolinealidad exacta. Como el precio de cobre en dolares en el mercado internacional es simplemente igual al precio en pesos multiplicado por un constante (tipo de cambio), no se puede estimar el modelo con las dos variables. Para solucionar el problema, tenemos que sacar una de las dos variables.

(d) (3 puntos) ¿Bajo qué condición se puede interpretar una coeficiente de regresión como un estimador causal?

Respuesta: Exogeneidad estricta. Si todas las variables incluidas en el modelo no están relacionado con el término de error, las coeficientes son estimadores causales.

(e) (3 puntos) ¿Cómo se difieren los supuestos del modelo clásico de regresión lineal asintótica y la versión de muestra finita? ¿Cuáles son más exigentes? [NOTA: No es necesario escribir los supuestos de cada modelo.]

Respuesta: La diferencia central depende del comportamiento de la parte no observada del modelo clásico: u**. En la versión de muestra finita se supone que** $u \sim \mathcal{N}(0, \sigma^2)$ **en todas las muestras, incluyendo muestras chicas. En la versión asintótica del modelo, sólo se supone que** $X'u/\sqrt{N}$ **es normal en su distribución asintótica, que quiere decir en muestras muy grandes. Esto es un supuesto mucho más débil, ya que viene gratis de los teoremas de límite central.**