

Examen II

Econometría Básica 2017-II

Paul Rodríguez Lesmes

18/10/2017

Indicaciones

- Tiene 1 hora y 50 minutos para responder todo el examen
- Para poder hacer reclamos, responda este examen con tinta permanente
- Debe responder en los espacios indicados, no se aceptan hojas suplementarias
- El examen consta de 50 puntos, distribuidos en las preguntas según se expresa en cada pregunta.

1. Parte I: Conceptos (12 puntos)

Pregunta 1 (4pt). Cierta investigador estimó el retorno a la educación en EEUU utilizando datos de hombres en 1976. Utilizó una *dummy* que señala si una persona vive cerca o no de una universidad (**nearc4**) como una variable instrumental para los años de educación, **educ**. En una ecuación donde la dependiente es el logaritmo de los salarios, $\log(wage)$, además de la educación se consideraron controles como la experiencia, ingreso de la familia, raza, lugar de residencia y de origen (no se incluye ninguna medida de la habilidad). En la tabla a continuación se presenta en la columna (1) la primera etapa de la estimación, donde la variable dependiente es el número de años de educación del individuo, y las independientes son la dummy de cercanía a la universidad junto a los otros controles ya descritos. En las columnas (2) y (3) se presentan dos regresiones donde están los resultados de estimar el retorno a la educación mediante mínimos cuadrados ordinarios (Col 2), y mediante variables instrumentales (Col 3).

	Dependent variable:		
	educ OLS (1)	log(wage) OLS (2)	instrumental variable (3)
nearc4	0.320*** (0.088)		
educ		0.075*** (0.003)	0.132** (0.055)
other controls	Yes	Yes	Yes

Note: *p < 0.1; **p < 0.05; ***p < 0.01

1. [1pt] ¿Por qué se debe instrumentar la variable **educ** en la regresión $\log(wage) = \delta_0 + \delta_1 educ + controls$, a pesar de estar controlando por otras variables?
2. [2pt] ¿Qué condiciones se requieren para que **nearc4** se pueda utilizar como instrumento de **educ** en la regresión por variables instrumentales de la columna 3?, ¿se cumplen dichas condiciones? Argumente
3. [1pt] Interprete los resultados de la tabla



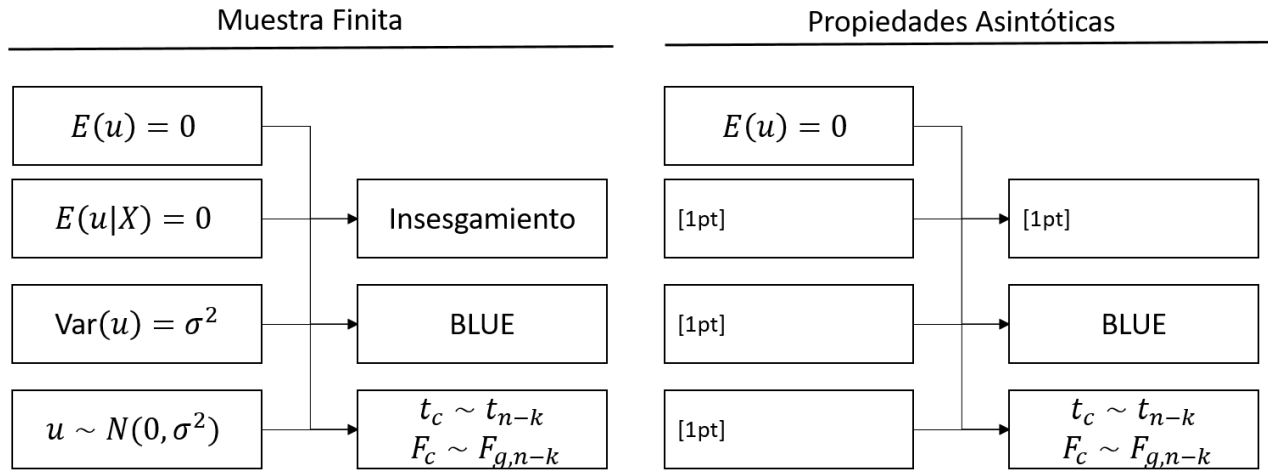
Pregunta 2 (2pt). Si suponemos que los resultados de la siguiente ecuación son causales, donde **marrmale** es 1 para los hombres casados y 0 para los hombres solteros y las mujeres; **marrfem** es 1 para las mujeres casadas y 0 para las solteras y los hombres; y **singfem** es 1 para las mujeres solteras y 0 para las casadas y los hombres ¿es correcta la siguiente interpretación?

Los hombres casados se ganan apx. un 21 % más que las mujeres en general

$$\begin{aligned} \log(\hat{wage}) = & 0,321 & +0,213 \text{ marrmale} & -0,191 \text{ marrfem} \\ & (0,100) & (0,055) & (0,058) \\ & & -0,110 \text{ singfem} & +0,079 \text{ educ} \\ & & (0,026) & (0,997) \end{aligned} \tag{1}$$

- a. Sí
- b. No

Pregunta 3 (6pt). El siguiente cuadro resume el rol de los principales supuestos del modelo de regresión lineal en las propiedades del estimador de mínimos cuadrados ordinarios, bajo el caso de muestra finita (muestras con pocas observaciones) y en el caso de tener una muestra donde aplican las propiedades asintóticas. (1) Complete el cuadro en los espacios requeridos, y (2) realice una breve explicación sobre las ventajas de estos nuevos conceptos frente



2. Parte II: Implementación, teoría, e interpretación (38 puntos)

A continuación se presenta el código de una rutina escrita en el lenguaje R que tiene como objetivo estimar un modelo de regresión lineal y contrastar hipótesis relevantes a la generación eléctrica. Para 145 termoeléctricas ubicadas en 13 ciudades, el archivo *nerlovEX.txt* tiene datos anuales sobre: identificador de ciudad (ciudad), costos totales (TC) en millones de dólares, producción (Q) en billones de kilovatios hora, salarios promedio de los operarios (PL), precio de una tonelada de carbón (PF), y precio del capital (PK) en miles de dólares. Basándose en una función de producción Cobb-Douglas, se estiman los parámetros de dicha función a través de la función de costos óptima:

$$\log TC_i = \beta_0 + \beta_1 \log Q_i + \beta_2 \log PL_i + \beta_3 \log PF_i + \beta_4 \log PK_i + u_i$$

(8 pt) Complete en los espacios indicados, ó si lo prefiere, especifique en otro lenguaje de programación la forma en la que llevaría a cabo el mismo ejercicio partiendo de la misma base de datos.

```
library(foreign); library(sandwich); library(car); library(lfe);
library(multiwayvcov); library(lmtest); library(stargazer)

misDatos=read.table("nerlovEX.txt",header=T,dec=".")
attach(misDatos)
# ~~~~~
# (i) Estimar los coeficientes del modelo
reg=lm(log(TC)~log(Q)+log(PL)+log(PF)+log(PK), data=misDatos)

# ~~~~~
# (ii) Calculo de las pruebas de [2 pt] _____

bptest(reg)
bptest(reg, ~ fitted(reg) + I(fitted(reg)^2) )

# ~~~~~
# (iii) Calculo de los errores estandar (SE) bajo varios supuestos

# Calculo de los SE bajo homoscedasticidad
clas=coeftest(reg)

# Calculo de los SE robustos a la heteroscedasticidad de Huber-White

rob =coeftest(reg, vcov = [2 pt] _____ )

# Calculo de los SE [2 pt] _____
vcov_clust = cluster.vcov(reg, misDatos$ciudad, df_correction = TRUE)
clust=coeftest(reg, vcov_clust)

# Compare them .....
stargazer(reg, reg, reg, type = "text",
          se=list(clas["Std. Error"],
                 rob["Std. Error"],
                 clust["Std. Error"]) )
# ~~~~~
# (iv) Contrastar H0: beta_2+beta_3+beta_4=1

linearHypothesis(reg, c("[2 pt] _____"),
                 test="F", vcov= vcov_clust )
```

Los resultados de la rutina anterior se presentan a continuación con algunas líneas omitidas. Según estos, responda las preguntas que se presentan a continuación.

```
# ~~~~~
# (ii)

> bptest(reg)
      studentized Breusch-Pagan test
data:  reg
BP = 35.552, df = 4, p-value = 3.578e-07

> bptest(reg, ~ fitted(reg) + I(fitted(reg)^2) )
      studentized Breusch-Pagan test
data:  reg
BP = 53.156, df = 2, p-value = 2.867e-12

# ~~~~~
# (iii)
```

	Dependent variable:		
	(1)	log(TC) (2)	(3)
log(Q)	0.720*** (0.017)	0.720*** (0.033)	0.720*** (0.064)
log(PL)	0.436 (0.291)	0.436* (0.246)	0.436 (0.269)
log(PF)	0.427*** (0.100)	0.427*** (0.075)	0.427*** (0.047)
log(PK)	-0.220 (0.339)	-0.220 (0.324)	-0.220 (0.239)
Constant	-3.527** (1.774)	-3.527** (1.719)	-3.527** (1.616)
Observations	145	145	145
R2	0.926	0.926	0.926

Note: *p<0.1; **p<0.05; ***p<0.01

```
# ~~~~~
# (iv)
Linear hypothesis test
Hypothesis:
log(PL) + log(PF) + log(PK) = 1
(... resultados omitidos ...)
  Res.Df Df      F Pr(>F)
1      141
2      140  1 0.6455 0.4231
```