Examen II

Econometría Básica 2017-II
Paul Rodríguez Lesmes

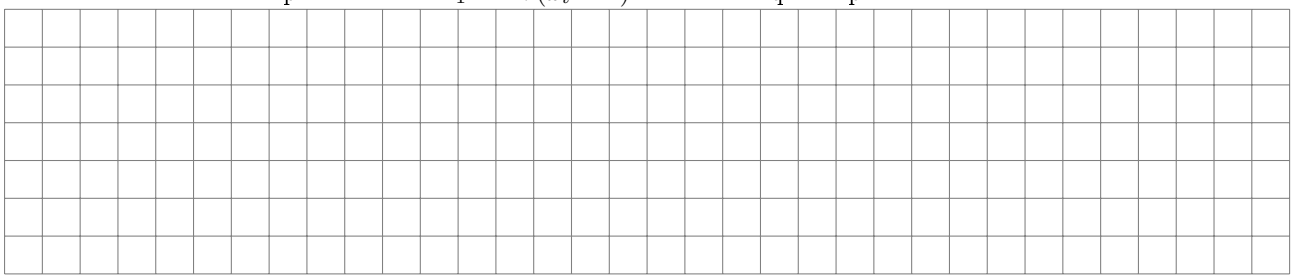
18/10/2017

Indicaciones

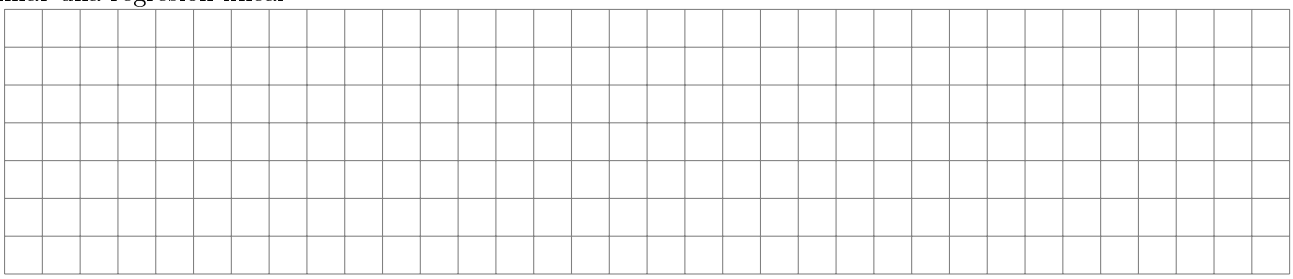
- Para poder hacer reclamos, responda este examen con tinta permanente
- Siga exactamente la notación que se encuentra en su enunciado. De lo contrario se penalizará con el 50 % de la nota de cada punto en la que se presente
- Debe responder en los espacios indicados
- El examen tiene dos partes, la primera es sobre conceptos, y una segunda donde hay tres opciones. **Sólo debe responder una de ellas**, si responde más de una, sólo se calificará la primera que haya sido respondida. Se sugiere que lea muy bien las 3 posibilidades para determinar cuál responder.
- El examen consta de 55 puntos, distribuidos en las preguntas según se expresa en cada una de ellas. La nota del examen se calcula a partir de la siguiente función: mín(5, $\frac{puntos}{10}$)

1. Parte I, única: Conceptos (32 puntos)

Pregunta 1 (2pt). La función de densidad de un proceso discreto $x_t \in \{0, 1\}$, que se repite T veces, es de la forma $P_1^N \cdot (1 - P_1)^{T-N_1}$, donde $N_1 = \sum_{t=1}^{T} x_t$ (el número de veces que se observa $x_t = 1$). Escriba la función de log-verosimilitud adecuada para estimar $P_1 = Pr(x_t = 1)$ asumiendo que el proceso es i.i.d.



Pregunta 2 (2pt). Responda [1pt] y justifique [1pt] si el siguiente enunciado es verdadero, falso, o incierto: el método de máxima verosimilitud tiene una fuerte desventaja frente a mínimos cuadrados ordinarios al momento de estimar una regresión lineal



Pregunta 3 (2pt). ¿Qué es el modelo de probabilidad lineal?

Pregunta 4 (2pt). ¿Qué es un efecto marginal en la media?, ¿por qué no se pueden interpretar directamente los coeficientes $\hat{\beta}$ de un modelo probit o logit?

Pregunta 5 (2pt). ¿Qué es sensibilidad y efectividad en el contexto de la predicción de variables binarias?

Pregunta 6 (2pt). ¿Según lo visto en clase, cuál es la diferencia fundamental entre una base de datos estructurada como una serie de tiempo y una de corte transversal? Mencione al menos un concepto que se basa en la estructura de esta clase de datos

Pregunta 7 (2pt). ¿Qué es un modelo de rezagos distribuidos? Presente la ecuación de uno de estos modelos.

Pregunta 8 (2pt). ¿Qué son la estacionalidad y la tendencia de una serie de tiempo?, ¿cómo se puede obtener el componente ***irregular*** de la serie?

Pregunta 9 (2pt). ¿Qué es un modelo AR(1)? Escriba su ecuación y mencione qué condición debe cumplirse para que sea estacionario.

Pregunta 10 (2pt). ¿Qué es un modelo MA(1)? Escriba su ecuación.

Pregunta 11 (2pt). En la literatura de evaluación de impacto, ¿qué es un contrafactual?, ¿para qué sirve?

Pregunta 12 (2pt). Intuitivamente, explique por qué el modelo de diferencia en diferencias requiere observar un grupo tratamiento y control en dos momentos del tiempo para estimar un efecto causal.

Pregunta 13 (2pt). En el modelo de diferencia en diferencias, ¿cuál es el supuesto de tendencias paralelas?

Pregunta 14 (2pt). Intuitivamente, explique cómo funciona el modelo de regresión discontinua para estimar un efecto causal.

Pregunta 15 (4pt). ¿Cuál era el objetivo de su proyecto final?

2. Parte II, opción 1: Modelos de elección discreta (23 puntos)

La base de datos college2 contiene 1341 estudiantes graduados del High School and Beyond Survey, quienes reportaron asistir a la universidad en octubre de 1980 después de graduarse en el mes de junio. Las variables de la base se describen así: COLLEGE: 1 si el estudiante fue a la universidad, 0 en caso contrario. FEMALE: 1 si el individuo es mujer, 0 si es hombre TEST: Puntaje del test. Para estimar el modelo, se propone el siguiente modelo logit:

$$\Pr(COLLEGE_i = 1|X) = \Lambda(\beta_0 + \beta_1 FEMALE + \beta_2 TEST + \beta_3 INCOME)$$

(4 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(margins); library(car);
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# Leer la base de datos
college2 = read_csv("college2.txt")
attach(college2)

#(i) Estimar un modelo logit
logit  =glm(COLLEGE ~ FEMALE + TEST + INCOME  , family = binomial(link="logit"))

#(ii) Esta línea calcula [2pt] ____________________________________
logit_meff  =summary(margins(logit  , vcov = hccm ))

#(iii) Esta línea obtiene los efectos marginales para un ingreso de 25k y de 75k

logit_meff  =summary(margins(logit  , at = [2pt] ____________________________
                      , vcov = hccm))
```

Los resultados del ejercicio anterior se presentan a continuación con algunas líneas omitidas. Teniéndole como referencia, responda las preguntas que se presentan después de los resultados

```
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
#(i) logit  =glm(COLLEGE ~ FEMALE + TEST  , family = binomial(link="logit"), data=college2)

Call:  glm(formula = COLLEGE ~ FEMALE + TEST, family = binomial(link = "logit"),
    data = college2)

Coefficients:
(Intercept)       FEMALE         TEST
    -5.0068       0.5527       0.1036

Degrees of Freedom: 1340 Total (i.e. Null);  1338 Residual
Null Deviance:      1786
Residual Deviance: 1570         AIC: 1576

# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
#(ii) logit_meff
 factor     AME     SE       z       p    lower   upper
 FEMALE  0.1116 0.0583  1.9151  0.0555 -0.0026  0.2258
 INCOME  0.0023 0.0013  1.7723  0.0764 -0.0002  0.0048
   TEST  0.0187 0.0031  5.9724  0.0000  0.0125  0.0248
```

```
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
#
#(iii) logit_meff2
   factor INCOME     AME      SE       z       p    lower   upper
 FEMALE      25  0.1196  0.0622  1.9216  0.0547 -0.0024  0.2416
 FEMALE      75  0.1066  0.0561  1.9009  0.0573 -0.0033  0.2166
 INCOME      25  0.0024  0.0015  1.6540  0.0981 -0.0005  0.0053
 INCOME      75  0.0022  0.0012  1.8887  0.0589 -0.0001  0.0044
   TEST      25  0.0200  0.0031  6.4139  0.0000  0.0139  0.0261
   TEST      75  0.0178  0.0035  5.1505  0.0000  0.0110  0.0246
```

1. [4pt] Cuando se tiene una variable dependiente binaria, ¿cuál es la ventaja de utilizar un logit en vez de un modelo de probabilidad lineal?
2. [4 pt] Escriba el problema de maxima verosimilitud que debe solucionar el computador para estimar el vector β del modelo logit propuesto
3. [4 pt] Interprete los efectos marginales en la media presentados en el ejercicio
4. [4 pt] Interprete los efectos marginales para un ingreso de 25k y de 75k
5. [3 pt] Sugiera cómo mejorar el modelo empírico propuesto

3. Parte II, opción 2: Modelos de series de tiempo (23 puntos)

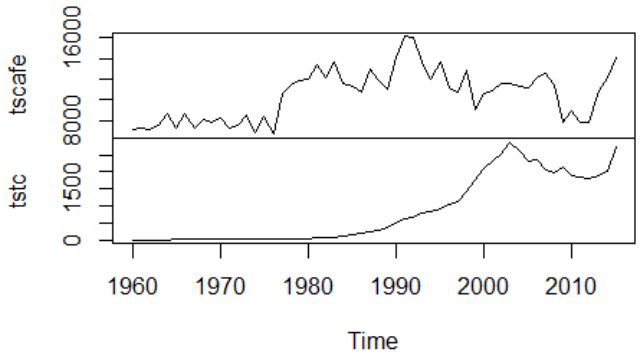
La base de datos mits.csv muestra la producción anual de café en Colombia ($Coffe_t$,por miles de sacos de café verde de 60Kg) , y el promedio anual de tasa de cambio COP/USD ($EXrate_t$) desde 1950 hasta 2016. Utilizando estos datos, se onsidera el siguiente modelo.

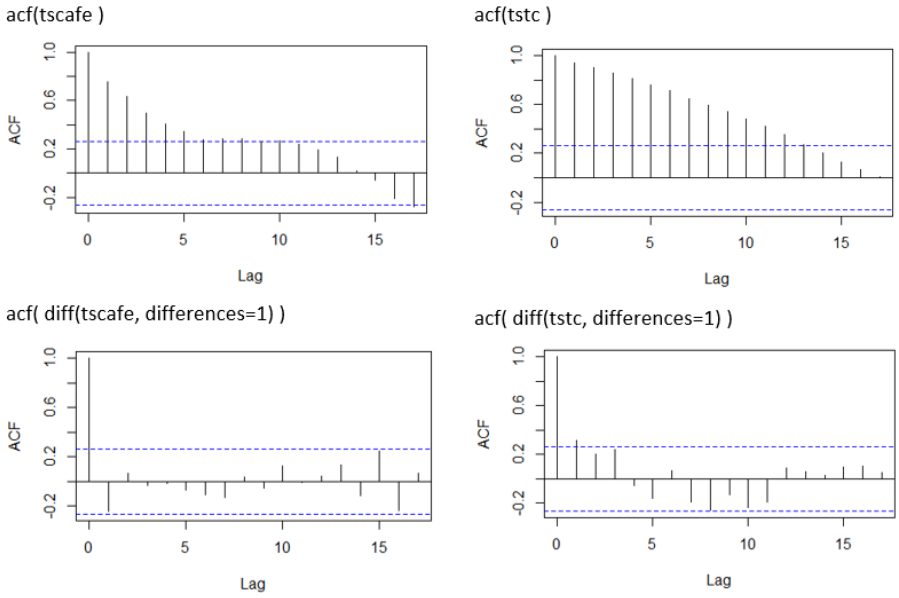
$$\Delta Coffe_t = \beta_0 + \beta_1 \Delta EXrate_t + \beta_2 \Delta Exrate_{t-1} + \delta t + u_t$$

(4 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(stargazer); library(dynlm); library(car); library(lmtest)
# ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# (i) Graficar ambas series de tiempo
plot(mits)

# (ii) Se caculan las funciones de [2pt] _____________________________________
#
# ___________________________________________________________________________

acf(tscafe )
acf( diff(tscafe, differences=1) )

acf(tstc )
acf( diff(tstc, differences=1) )

# (iii) Correr dos versiones del modelo: la primera línea (r1) incluye
#       una versión son el rezago de la diferencia en el tiempo de
#       EXRate, la segunda (r2) una versión con dicha variable

r1=dynlm(diff(tscafe,1) ~ diff(tstc,1)                    + trend(mits) , data=mits)

r2=dynlm(diff(tscafe,1) ~ [2pt]_____________________________________________, data=mits)

stargazer(r1,r2,type="text")
```

Los resultados del ejercicio anterior se presentan a continuación con algunas líneas omitidas. Teniéndole como referencia, responda las preguntas que se presentan después de los resultados





	Dependent variable: diff(tscafe, 1)	
	(1)	(2)
diff(tstc, 1)	0.604	−1.285
	(1.400)	(1.828)
lag(diff(tstc, 1), 1)		1.663***
		(0.505)
trend(mits)	−0.548	−5.764
	(14.177)	(14.481)
Constant	113.108	219.341
	(458.270)	(461.224)
Observations	55	54
R2	0.004	0.027
Residual Std. Error	1,630.646 (df = 52)	1,621.552 (df = 50)
F Statistic	0.095 (df = 2; 52)	0.455 (df = 3; 50)
Note:		*p<0.1; **p<0.05; ***p<0.01

1. [4pt] ¿Qué nos indica las funciones de autocorrelación parcial graficadas en la parte (ii) del código?
2. [4pt] Basado en los resultados de este ejercicio, ¿por qué se utiliza la primera diferencia de ambas variables?
3. [4 pt] ¿Qué es un modelo ARIMA(p,d,q)?, ¿cuál sería el orden del modelo que se fue estimado para este ejercicio?
4. [4 pt] Interprete los resultados de los dos modelos estimados. ¿Cuál modelo es preferido?
5. [3 pt] Sugiera cómo mejorar el modelo empírico propuesto

4. Parte II, opción 3: Evaluación de impacto (23 puntos)

4.1. [11.5pt] Diferencia en Diferencias

Para determinar el rol de los salarios mínimos sobre el empleo, cierto estudio considera el número de empleos en las cadenas de restaurantes en los diferentes estados de EEUU. En particular, a principios de 1988 se introduce el salario mínimo para el sector en cierto estado (estado «tratamiento»), mientras que esta clase de políticas sólo serían consideradas años después en otros estados de dicho país, en particular, de un estado vecino geográficamente (estado «control»). Gracias a un intenso trabajo de recolección de datos, se cuenta con información del número de empleos en el restaurante i en el año t, (variableE_{it}), para un grupo de establecimientos a ambos lados de la frontera entre los dos estados. Si un restaurante está del lado del estado «tratado», se define la variable binaria $Treat_i = 1$, mientras que si está en el lado del estado control se tiene $Treat_i = 0$. Inicialmente contamos con dos años de información; si la observación es para 1988 tenemos $After_t = 1$, mientras que si es para 1987 tendríamos $After_t = 0$. Con base a estos datos, se construye la siguiente tabla:

Cuadro 1: Número de empleos promedio por restaurante

	$t = 1987$	$t = 1988$
Estado tratado	5	5.5
Estado control	4.3	4.0

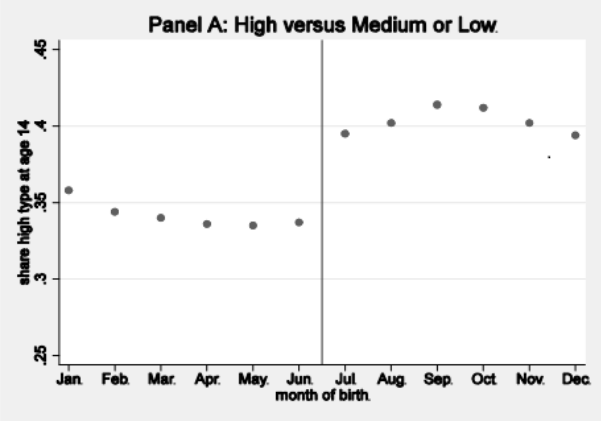
1. [4pt] Con base en la tabla, ¿cuál es el efecto causal de la introducción de los salarios mínimos?
2. [2pt] ¿Qué ecuación se utilizaría para estimar el efecto causal si tuviésemos más estados control, e información en más de dos periodos de tiempo?
3. [2pt] ¿Qué supuestos son necesarios para verificar la validez del ejercicio anterior? Explique a qué se refieren en este contexto.
4. [3.5pt] Interprete los resultados del ejercicio. ¿Considera que este ejercicio es el adecuado para estimar el impacto de esta política?

4.2. [11.5pt] Regresión discontinua

Al final de la escuela primara en Alemania, los niños pueden entrar a dos tipos diferentes de «trayectorias» educativas, la alta (*Gymnasium*) y el resto (*Realschule* y *Hauptschule*). La enseñanza en la trayectoria alta es más intensa y más académica que la impartida en los otros tipos de colegios. La trayectoria alta es tradicionalmente reservada para los más brillantes, y hay una fuerte competencia para ingresar. Dustmann, Puhani y Schoenberg (2014) buscan evaluar los efectos de largo plazo de estudiar en uno de los colegios de élite en este país.

En Alemania, los niños usualmente empiezan a estudiar cuando tienen 7 años, y el calendario oficial empieza en el segundo semestre del año (como los colegios calendario B). Por tal motivo, los nacidos en Julio normalmente empiezan a estudiar un año después que los que nacieron en Junio. Para los autores, los niños que entran un poco mayores a la primaria tienen una ventaja sobre los demás, lo que se refleja en una mayor probabilidad de entrar a los colegios de trayectoria alta (ver la gráfica a la izquierda de la siguiente figura). Esta misma regla se puede usar para ver si debido a la fecha de nacimiento hay un salto en los salarios de los hombres (30 años o más) sin ajustar por experiencia (gráfica derecha-arriba) y ajustando por la misma (gráfica derecha-abajo). Basados en esta observaciones, los autores estiman el efecto que tiene estudiar en la trayectoria alta frente a las otras opciones sobre dichos salarios (tabla presentada en la figura).

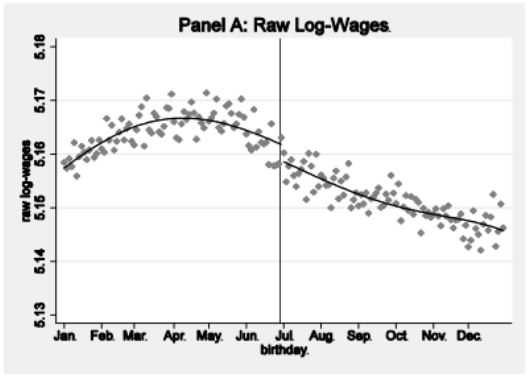
Track Attendance at Age 14 and Birth Month

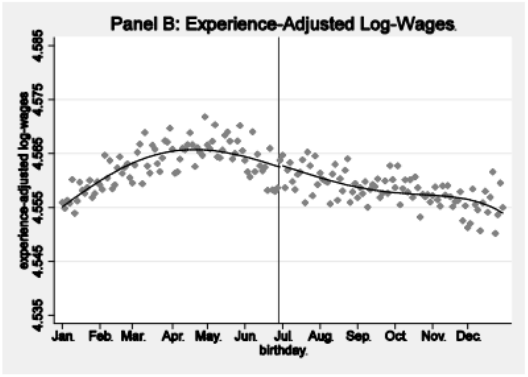


Panel B: Two-Sample Two-Stage Least Squares Estimates

	(1) Jul-Jun, none
(i) Wages net of experience (age 30 and higher)	
Coeff.	-0.024
(s.e.)	(0.025)

Date of Birth and Wages (Men only)





1. [4pt] ¿Corresponde este ejercicio a un *sharp RDD* o a un *fuzzy RDD*?, ¿cuál es la diferencia entre ambas técnicas?
2. [2pt] ¿Qué ecuación utilizaría para estimar el efecto del salto en la probabilidad de asistir a un colegio de trayectoria alta debido a la fecha de nacimiento? (Primera etapa)
3. [2pt] ¿Qué ecuación utilizaría para estimar el efecto de asistir a un colegio de trayectoria alta en los salarios? (Segunda etapa, correspondiente a la tabla presentada en la figura)
4. [3.5pt] Interprete los resultados del ejercicio. ¿Qué explicaciones pueden explicar estos resultados?