

Econometría 06169
Examen Parcial #2
Cali, lunes 5 de Octubre de 2009

Profesores: Julio César Alonso y Carlos Giovanni González

Estudiante: _____

Código: _____

Instrucciones:

1. Lea cuidadosamente todas las preguntas e instrucciones.
2. Este examen consta de **9** páginas; además, deben tener 2 páginas de fórmulas.
3. El examen consta de 4 preguntas que suman un total de 100 puntos. El valor de cada una de las preguntas esta expresado al lado de cada pregunta.
4. Escriba su respuesta en las hojas suministradas, marque cada una de las hojas con su nombre. **NO** responda en las hojas de preguntas.
5. El examen está diseñado para dos horas, pero ustedes tienen 3 horas para trabajar en él.
6. Recuerde que no se tolerará ningún tipo de deshonestidad académica. En especial usted no puede emplear ningún tipo de ayuda diferente a la que se le entrega con este examen.
7. El uso de calculadoras está prohibido.
8. No se aceptarán reclamos de respuestas escritas en lápiz.
9. Al finalizar su examen entregue sus hojas de respuesta, así como las hojas de preguntas.
10. Asigne su tiempo de forma eficiente!

Buena Suerte!

1 Falso o Verdadero (15 puntos en total, 5 puntos cada subparte)

Diga si las siguientes afirmaciones son falsas o verdaderas y explique brevemente su respuesta. (No se dará ningún crédito a respuestas sin justificación.)

- a) Considere el siguiente modelo: $y_i = \beta_0 + \sum_{k=1}^5 \beta_k D_{k,i} + \varepsilon_i$, donde y_i es el rendimiento de la acción de suramericana para el día i y $D_{k,i}$ toma el valor de uno si i es el k -ésimo día laboral de la semana y cero en caso contrario. Además, suponga que se emplea una muestra para estimar este modelo que está compuesta por los últimos 2000 días hábiles en los que el mercado accionario estuvo abierto. Además *es un hecho*, demostrado por muchos estudios, que “cuando los rendimientos de la acción de Suramericana de los lunes tienden a ser grandes, el rendimiento del viernes es alto. Y cuando el viernes es bajo el lunes también será bajo. Es más existe una correlación entre los rendimientos de esos dos días de 0.9”. Sobre este modelo un investigador afirma que “Dada la evidencia que existe sobre la relación entre los retornos de la acción de Suramericana de los lunes y los viernes, se tendrá un problema de multicolinealidad en el modelo” ¿Es esta afirmación verdadera o falsa?
- b) Considere el caso de una regresión simple: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. Donde $Var(X_i) = i$, ε_i cumple los supuestos del Teorema de Gauss-Markov y $Cov[\varepsilon_i X_i] = Cov[\varepsilon_i \beta_{j,i}] = 0$ ($j = 0, 1$). En este caso podemos afirmar que el estimador MCO para β_0 será insesgado pero no será eficiente.
- c) Suponga el modelo $\ln y_i = \beta_1 + \beta_2 \ln x_i + \varepsilon_i$, en donde ε_i es el término de error y $var(\varepsilon_i) = \sigma^2 D_i$, donde D_i es una variable dummy que toma el valor de uno si el individuo es profesional y cero en caso contrario. Un estudiante afirma: “Se puede transformar el modelo dividiendo por $\sqrt{D_i}$, de modo que los estimadores MCO de los β 's sean insesgados y eficientes.

2 Selección Múltiple (15 puntos en total, 5 puntos cada subparte)

Determine cuál de las siguientes respuestas es la correcta. Escoja la mejor opción y explique en dos o tres líneas su respuesta. (No se dará ningún crédito a respuestas sin justificación.)

- i. Un econométra desea estimar el efecto de la educación y el género sobre el salario de los trabajadores del sector privado en una pequeña República caribeña. Para lo cual estima el siguiente modelo de regresión lineal múltiple $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 D_{1i} + \beta_3 D_{2i} + \varepsilon_i$; donde, Y_i ; X_{1i} ; D_{1i} (1 si es mujer y 0 en caso contrario) y D_{2i} (1 si es hombre y 0 en caso contrario) representan el salario, los años de educación y variables dummy para el género, respectivamente. Por lo tanto, para la estimación de los estimadores MCO se cumple:
- $\hat{\beta} = (X^T X)^{-1} X^T Y$
 - $\det(X^T X) = |X^T X| = 0$
 - $|R| = \prod_{i=1}^{k-1} \lambda_i = 1$
 - Ninguna de las anteriores
- ii. En el siguiente modelo $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, donde la varianza del término de error es $var(\varepsilon_i) = X_{1i} * \sigma^2$; para todo $i=1, \dots, N$. Si estimamos el modelo original, la existencia de heterocedasticidad es un problema porque:

- a. En su forma habitual, MCO supone homocedasticidad y calcula las estimaciones de los parámetros del modelo en base a ese supuesto erróneo.
 - b. En su forma habitual, MCO supone homocedasticidad y calcula las estimaciones de varianza del estimador en base a ese supuesto erróneo.
 - c. Sesga el estimador MCO de los parámetros asociados a pendientes e intercepto.
 - d. Todas las anteriores
- iii. Un econométra desea estimar el siguiente modelo $Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$ con datos de una serie de tiempo ($t=1, \dots, 120$). Dado que $E[\varepsilon_i \varepsilon_j] = \rho \sigma^2$, donde $j = i+1$ y $|\rho| < 1$; existe un problema de autocorrelación y por lo tanto el estimador de la matriz de varianzas y covarianzas es _____. Y además, si se realiza un test de DW para contrastar la existencia de autocorrelación que arroja un valor de $DW = 1,2$. Entonces, la autocorrelación será _____. El apartado se completa con:
- a. Sesgado, autocorrelación negativa
 - b. Sesgado, autocorrelación positiva
 - c. Ineficiente, autocorrelación negativa
 - d. Ninguna de las anteriores

3 (30 puntos)

La Organización Internacional del Trabajo (OIT) desea conocer el efecto de los sindicatos sobre la productividad por trabajador en el sector manufacturero del país Nueva Granada, el cual ha sido uno de los más afectados por la crisis económica. Para esto contrata dos econométristas, quienes se basan en algunas regularidades empíricas, para establecer que la producción por trabajador (y_t) depende del capital por trabajador (k_t), ambas variables medidas en millones de dólares constantes de 2000, y la proporción total de trabajadores que están afiliados a un sindicato (P_t). El modelo que sugiere la teoría emplear es:

$$\ln(y_t) = \beta_0 + \beta_1 \ln(k_t) + \beta_2 P_t + \beta_3 P_t^2 + e_t$$

Responda brevemente cada una de las siguientes preguntas empleando la información que se presenta al final del examen (tabla 1 y tabla 2).

- a) Escriba el modelo estimado por los econométrista, presentado en la tabla 1. **(2 puntos)**
- b) ¿Cree usted que el econométrista cometió algún error o no, al estimar ese modelo? Sea lo más claro posible y argumente su respuesta. **(7 puntos)**
- c) El investigador, estimó dos modelos. con miedo de tener un problema econométrico construye otro modelo y lo estima. Este segundo modelo se reporta en la Tabla 2. ¿Tiene razón en estimar este segundo modelo? (Sea lo más preciso y emplee únicamente la información disponible) **(5 puntos)**
- d) ¿Cómo fue corregido el problema si es que existía uno? ¿Cuál modelo debería emplear para sacar las conclusiones del estudio? **(7 puntos)**
- e) Interprete los coeficientes del modelo que crea más conveniente **(9 puntos)**

4 (40 puntos)

El jefe de estudios económicos de una multinacional productora de papel desea estimar la función de demanda del mercado de papel. Para tal fin, se desea emplear la siguiente función de demanda:

$$Q_i = \beta_1 + \beta_2 P_i + \beta_3 \frac{1}{PI} + \beta_4 Y_i^2 + \varepsilon_i \quad (1)$$

Donde el Q_i corresponde al logaritmo del número de resmas de papel tamaño carta que desea comprar el hogar i al precio P_i (medido en miles de pesos). Además, PI es el precio de la impresora más común en el mercado (este precio es igual para todos los hogares) y Y_i representan el ingreso de cada hogar (medido en pesos). Por otro lado, se cuenta con una muestra de tamaño 15. Finalmente, ε_i es el término de error que cumple las siguientes características: $E[\varepsilon] = 0$, $Var[\varepsilon] = PI^2 \sigma^2 I_{15}$ (I_{15} corresponde a la matriz identidad de tamaño 15).

A partir de la información anterior conteste las siguientes preguntas:

- Si el jefe de estudios económicos emplea el método de mínimos cuadrados ordinarios (MCO), ¿existe algún problema que pudiese no permitir obtener conclusiones “adecuadas” sobre el efecto del precio sobre la demanda de resmas de papel? Sea lo más claro posible. **(10 puntos)**
- El econométrista a cargo del estudio recolecta la información obtener estimadores MELI para el modelo (1). El obtiene las siguientes matrices que corresponden al equivalente de la matriz $X^T X$ y $X^T y$. NOTA: Para la construcción de estas matrices, el modelo se organizó lo más parecido al modelo original:

$$X^T X = \begin{bmatrix} 15 & 0 & 10 \\ 0 & 5 & 0 \\ 10 & 0 & 15 \end{bmatrix} \quad X^T y = \begin{bmatrix} 45 \\ -15 \\ -20 \end{bmatrix}$$

Explique claramente a que corresponde cada uno de los elementos de la matriz $X^T X$. (Por ejemplo, explique a partir de que sumatoria sale el 10 que corresponde al último elemento de la matriz $X^T X$, y así sucesivamente con cada elemento de la matriz) **Expresar su respuesta en términos de las variables en el modelo original y no en términos de variables reparametrizadas** (por ejemplo P_i). **(6 puntos)**

- Encuentre los estimadores MELI para los coeficientes que se puedan recuperar del **modelo (1)**. **MUESTRE** claramente el valor estimado y su correspondiente parámetro poblacional. **(12 Puntos)**
- Interprete el significado de cada uno de los coeficientes estimados. **(6 Puntos)**
- El jefe de estudios económicos de la compañía después de realizar las estimaciones encuentra que:
 - Según el modelo teórico más empleado para la estimación de la demanda, es necesario incluir en el modelo el tamaño de la familia.
 - El ingreso de las personas no fue medido con la precisión requerida. De hecho, se encontró que existe un error de medición que es aleatorio.

De acuerdo a estos hallazgos, que puede esperar el jefe de estudios sobre los resultados obtenidos hasta ahora y en especial sobre las conclusiones que se saquen. Sea lo más claro posible, explique todas sus afirmaciones **(6 Puntos)**

Resultados de EasyReg:

Tabla 1

Dependent variable:			
Y = ln[Y]			
Characteristics:			
Ln[Y]			
First observation = 1(=1951)			
Last observation = 55(=2005)			
Number of usable observations: 55			
Minimum value: 5.5870000E+000			
Maximum value: 1.9781000E+001			
Sample mean: 1.3118156E+001			
X variables:			
X(1) = ln(K)			
X(2) = P			
X(3) = 1			
WARNING: The effective degrees of freedom is only 28.			
Therefore, the estimation results may be unreliable!			
Model:			
$Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + U,$			
where U is the error term, satisfying			
$E[U X(1),X(2),X(3)] = 0.$			
OLS estimation results			
Parameters	Estimate	t-value	H.C. t-value
	(S.E.)	(H.C. S.E.)	
	[p-value]	[H.C. p-value]	
b(1)	33,984131	33.984	41.543
	(0.01000)	(0.01003)	
	[0.00000]	[0.00000]	
b(2)	25,560146	28.339	-2.030
	(0.01962)	(0.02759)	

		[0.00430]	[0.04235]
b(3)	0.2134671	0.983	0.929
		(0.21715)	(0.22985)
		[0.32560]	[0.35304]
Notes:			
1: S.E. = Standard error			
2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.			
3: The two-sided p-values are based on the normal approximation.			
Effective sample size (n):		55	
Variance of the residuals:		0.12151531	
Standard error of the residuals (SER):		0.34859046	
Residual sum of squares (RSS):		3.40242872	
(Also called SSR = Sum of Squared Residuals)			
Total sum of squares (TSS):		788.65192071	
R-square:		0.9957	
Adjusted R-square:		0.8726	
Overall F test: $F(2,53) = 32245,98$			
p-value = 0.00000			
Significance levels:	10%	5%	
Critical values:	2.5	3.34	
Conclusions:	reject	reject	
Test for first-order autocorrelation:			
Durbin-Watson test = 3,0532			
Jarque-Bera/Salmon-Kiefer test = 0.094518			
Null hypothesis: The errors are normally distributed			
Null distribution: Chi-square(2))			

p-value = 0.95384

Significance levels: 10% 5%

Critical values: 4.61 5.99

Conclusions: accept accept

Breusch-Pagan test = 2.200154

Null hypothesis: The errors are homoskedastic

Null distribution: Chi-square(2)

p-value = 0.33285

Significance levels: 10% 5%

Critical values: 4.61 5.99

Conclusions: accept accept

Tabla 2. Resultados de EasyReg

Dependent variable:

$$Y = \ln[Y] + .082733LAG1[\ln[Y]]$$

Characteristics:

$$\ln[Y].082733x LAG1[\ln[Y]]$$

First observation = 1(=1951)

Last observation = 55(=2005)

Number of usable observations: 55

Minimum value: 5.5870000E+000

Maximum value: 1.9781000E+001

Sample mean: 1.3118156E+001

X variables:

$$X(1) = \ln[K] + .082733 LAG1[\ln[K]]$$

$$X(2) = P + .082733LAG1[P]$$

$$X(3) = 1$$

WARNING: The effective degrees of freedom is only 28.

Therefore, the estimation results may be unreliable!

Model:

$$Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + U,$$

where U is the error term, satisfying

$$E[U|X(1),X(2),X(3)] = 0.$$

OLS estimation results

Parameters	Estimate	t-value (S.E.)	H.C. t-value (H.C. S.E.)
		[p-value]	[H.C. p-value]
b(1)	2.2968185	1.502 (0.19765) [0.00000]	1.360 (0.21817) [0.00000]
b(2)	31,91243	28.339 (0.01962) [0.00330]	-2.030 (0.02759) [0.03235]
b(3)	0.3134671	0.983 (0.21715) [0.32560]	0.929 (0.22985) [0.35304]

Notes:

- 1: S.E. = Standard error
- 2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
- 3: The two-sided p-values are based on the normal approximation.

Effective sample size (n):	55
Variance of the residuals:	0.10845059
Standard error of the residuals (SER):	0.32931837
Residual sum of squares (RSS):	3.14506713
(Also called SSR = Sum of Squared Residuals)	
Total sum of squares (TSS):	3.38921543
R-square:	0.0720
Adjusted R-square:	0.0720

Overall F test: $F(2,53) = 145445,98$

p-value = 0.00000

Significance levels: 10% 5%

Critical values: 2.5 3.34

Conclusions: reject reject

Test for first-order autocorrelation:

Durbin-Watson test = 1,896

Jarque-Bera/Salmon-Kiefer test = 0.743108

Null hypothesis: The errors are normally distributed

Null distribution: Chi-square(2)

p-value = 0.68966

Significance levels: 10% 5%

Critical values: 4.61 5.99

Conclusions: accept accept

Breusch-Pagan test = 1,93412

Null hypothesis: The errors are homoskedastic

Null distribution: Chi-square(2)

p-value = 0.92285

Significance levels: 10% 5%

Critical values: 4.61 5.99

Conclusions: accept accept

Prof: Julio César Alonso Carlos Giovanni González

$$X^T X = \begin{bmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} & \cdots & \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i}X_{2i} & \cdots & \sum_{i=1}^n X_{1i}X_{ki} \\ \sum_{i=1}^n X_{2i}^2 & \ddots & \sum_{i=1}^n X_{2i}X_{ki} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^n X_{ki}^2 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i X_{1i} \\ \sum_{i=1}^n y_i X_{2i} \\ \vdots \\ \sum_{i=1}^n y_i X_{ki} \end{bmatrix} \quad y^T y = \sum_{i=1}^n y_i^2$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 = y^T y - n\bar{y}^2 \quad SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad s^2 = \frac{SSE}{n-k} = \frac{y^T y - \hat{\beta}^T X^T y}{n-k}$$

$$Var[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{\beta}^T X^T y - n\bar{y}^2 \quad t = \frac{\hat{\beta}_i - c}{s_{\hat{\beta}_i}}$$

$$F_c = \frac{(c - R\hat{\beta})^T (R(X^T X)^{-1} R^T)^{-1} (c - R\hat{\beta}) / r}{SSE / (n-k)}$$

$$F_c = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)} = \frac{MSR}{MSE}$$

$$F_c = \frac{(SSE_R - SSE_U) / r}{SSE_U / (n-k)} \quad R^2 = \frac{SSR}{SST}$$

$$\hat{\beta}_i \pm t_{\frac{\alpha}{2}, n-k} s_{\hat{\beta}_i} \quad \bar{R}^2 = 1 - (1-R^2) \frac{n-1}{n-k}$$

$$\hat{y}_p = x_p^T \hat{\beta}, \quad x_p^T = (1 \quad x_{1p} \quad x_{2p} \quad \cdots \quad x_{kp})$$

$$\hat{y}_p \pm t_{\frac{\alpha}{2}, n-k} \sqrt{\sigma^2 x_p^T (X^T X)^{-1} x_p}$$

$$\hat{y}_p \pm t_{\frac{\alpha}{2}, n-k} \sqrt{\sigma^2 [1 + x_p^T (X^T X)^{-1} x_p]}$$

$$\hat{\beta}_j^E = \hat{\beta}_j \frac{s_{X_j}}{s_y}, \quad j = 2, 3, \dots, k \quad E_j = \hat{\beta}_j \frac{\bar{X}_j}{\bar{y}}$$

Test de Heteroscedasticidad

Goldfeld y Quand: $F_{GQ} = \frac{SSE_2}{SSE_1} \sim F_{(n-d-2k, n-d-2k)}$

Prof: Julio César Alonso Carlos Giovanni González

Breush-Pagan: $\frac{\hat{\epsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta Z_i + \mu_i, BP = \frac{SSR}{2} \sim \chi_g^2$

White: $\hat{\epsilon}_i^2 = \gamma + \sum_{m=1}^k \sum_{j=1}^k \delta_s X_{mi} X_{ji} + \mu_i, W_a = nR^2 \sim \chi_g^2$

d_l y d_u para el test de DW al nivel de significancia del 5%

N	k-1=1		k-1=2		k-1=3	
	d _l	d _u	d _l	d _u	d _l	d _u
50	1.50	1.59	1.46	1.63	1.42	1.67
60	1.55	1.62	1.51	1.65	1.48	1.69
95	1.64	1.69	1.62	1.71	1.60	1.73
100	1.65	1.69	1.63	1.72	1.61	1.74

Test de Autocorrelación

Durbin-Watson $DW \approx 2(1 - \hat{\rho})$

Ho	Sí	Decisión
$H_0 : \rho = 0$	$d_u < DW < 4 - d_u$	A
No auto +	$0 < DW < d_l$	R
No auto -	$4 - d_l < DW < 4$	R

Área de indecisión $d_l < DW < d_u$ y $4 - d_u < DW < 4 - d_l$

$h = \hat{\rho} \sqrt{\frac{n}{1 - n(\widehat{Var}(\hat{\alpha}))}}$ donde

$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + \alpha y_{t-1} + \epsilon_t$

Cantidades Importantes

$\sqrt{2} = 1.414$
 $\sqrt{10} = 3.162$

$\sqrt{3} = 1.732$
 $\sqrt{13} = 3.606$

$\sqrt{5} = 2.236$

$\sqrt{7} = 2.646$

Econometría 06169
Examen Parcial #2
Respuestas Sugeridas
Cali, lunes 5 de Octubre de 2009

Profesores: Julio César Alonso y Carlos Giovanni González

Estudiante: _____
 Código: _____

Instrucciones:

1. Lea cuidadosamente todas las preguntas e instrucciones.
2. Este examen consta de 5 páginas; además, deben tener 3 páginas de fórmulas.
3. El examen consta de 4 preguntas que suman un total de 100 puntos. El valor de cada una de las preguntas esta expresado al lado de cada pregunta.
4. Escriba su respuesta en las hojas suministradas, marque cada una de las hojas con su nombre. NO responda en las hojas de preguntas.
5. El examen está diseñado para una hora, pero ustedes tienen 2 horas para trabajar en él.
6. Recuerde que no se tolerará ningún tipo de deshonestidad académica. En especial usted no puede emplear ningún tipo de ayuda diferente a la que se le entrega con este examen.
7. Al finalizar su examen entregue sus hojas de respuesta, así como las horas de preguntas.
8. Asigne su tiempo de forma eficiente!

Buena Suerte!

1 Falso o Verdadero (15 puntos en total, 5 puntos cada subparte)

Diga si las siguientes afirmaciones son falsas o verdaderas y explique brevemente su respuesta. (No se dará ningún crédito a respuestas sin justificación.)

- a) Considere el siguiente modelo: $y_i = \beta_0 + \sum_{k=1}^5 \beta_k D_{k,i} + \varepsilon_i$, donde y_i es el rendimiento de la acción de suramericana para el día i y $D_{k,i}$ toma el valor de uno si i es el k -ésimo día laboral de la semana y cero en caso contrario. Además, suponga que se emplea una muestra para estimar este modelo que está compuesta por los últimos 2000 días hábiles en los que el mercado accionario estuvo abierto. Además *es un hecho*, demostrado por muchos estudios, que “cuando los rendimientos de la acción de Suramericana de los lunes tienden a ser grandes, el rendimiento del viernes es alto. Y cuando el viernes es bajo el lunes también será bajo. Es más existe una correlación entre los rendimientos de esos dos días de 0.9”. Sobre este modelo un investigador afirma que “Dada la evidencia que existe sobre la relación entre los retornos de la acción de Suramericana de los lunes y los viernes, se tendrá un problema de multicolinealidad en el modelo” ¿Es esta afirmación verdadera o falsa?

La afirmación es Falsa. Si bien el modelo tendrá un problema de multicolinealidad perfecta,

la razón no es la que da la afirmación. La razón es que $\sum_{k=1}^5 D_{k,i} = 1$.

- b) Considere el caso de una regresión simple: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. Donde $Var(X_i) = i$, ε_i cumple los supuestos del Teorema de Gauss-Markov y $Cov[\varepsilon_i, X_i] = Cov[\varepsilon_i, \beta_{j,i}] = 0$ ($j = 0, 1$). En este caso podemos afirmar que el estimador MCO para β_0 será insesgado pero no será eficiente.

La afirmación es Falsa. Porque los estimadores MCO siguen siendo MELI en presencia de una variable explicatoria estocástica que no está relacionada con el término de error o con los betas.

- c) Suponga el modelo $\ln y_i = \beta_1 + \beta_2 \ln x_i + \varepsilon_i$, en donde ε_i es el término de error y $var(\varepsilon_i) = \sigma^2 D_i$, donde D_i es una variable dummy que toma el valor de uno si el individuo es profesional y cero en caso contrario. Un estudiante afirma: “Se puede transformar el modelo dividiendo por $\sqrt{D_i}$, de modo que los estimadores MCO de los β 's sean insesgados y eficientes.

Falso. En general uno esperaría resolver el problema empleando Mínimos Cuadrados Ponderados (MCP). De tal forma que: $\frac{\ln y_i}{\sqrt{D_i}} = \frac{\beta_1}{\sqrt{D_i}} + \beta_2 \frac{\ln x_i}{\sqrt{D_i}} + \frac{\varepsilon_i}{\sqrt{D_i}}$. Hasta aquí parecería razonable la afirmación, pero noten que se puede tener que D_i sea cero. Y será imposible dividir por cero. Por eso, para este caso esa transformación es imposible.

2 Selección Múltiple (15 puntos en total, 5 puntos cada subparte)

Determine cuál de las siguientes respuestas es la correcta. Escoja la mejor opción y explique en dos o tres líneas su respuesta. (No se dará ningún crédito a respuestas sin justificación.)

- i. Un econométra desea estimar el efecto de la educación y el género sobre el salario de los trabajadores del sector privado en una pequeña República caribeña. Para lo cual estima el siguiente modelo de regresión lineal múltiple $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 D_{1i} + \beta_3 D_{2i} + \varepsilon_i$; donde, Y_i ; X_{1i} ; D_{1i} (1 si es mujer y 0 en caso contrario) y D_{2i} (1 si es hombre y 0 en caso contrario) representan el salario, los años de educación y variables dummy para el género, respectivamente. Por lo tanto, para la estimación de los estimadores MCO se cumple:
- $\beta = (X^T X)^{-1} X^T Y$
 - $\det(X^T X) = |X^T X| = 0$
 - $|R| = \prod_{i=1}^{k-1} \lambda_i = 1$
 - Ninguna de las anteriores

Respuesta: b, existe un problema de multicolinealidad perfecta ($\det(X^T X) = |X^T X| = 0$), derivado de una mala especificación del modelo, el econométra cayó en la trampa de la Dummy al incluir dos variables dummy para el género, que tiene dos características: 1. Hombre, 2 mujer, y por lo tanto siguiendo la regla debió incluir J-1 variables dummy.

- ii. En el siguiente modelo $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$, donde la varianza del término de error es $\text{var}(\varepsilon_i) = X_{1i} * \sigma^2$; para todo $i=1, \dots, N$. Si estimamos el modelo original, la existencia de heterocedasticidad es un problema porque:
- En su forma habitual, MCO supone homocedasticidad y calcula las estimaciones de los parámetros del modelo en base a ese supuesto erróneo.
 - En su forma habitual, MCO supone homocedasticidad y calcula las estimaciones de varianza del estimador en base a ese supuesto erróneo.
 - Sesga el estimador MCO de los parámetros asociados a pendientes e intercepto.
 - Todas las anteriores

Respuestas: b, no podemos suponer homocedasticidad pues en la especificación del modelo conocemos que $\text{var}(\varepsilon_i) = X_{1i} * \sigma^2$; y a priori sabemos que existe un problema de heterocedasticidad. Si estimamos el modelo original suponiendo homocedasticidad este supuesto será erróneo ya que la varianza del estimador estará sesgada.

- iii. Un econométra desea estimar el siguiente modelo $Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$ con datos de una serie de tiempo ($t=1, \dots, 120$). Dado que $E[\varepsilon_i \varepsilon_j] = \rho \sigma^2$, donde $j = i+1$ y $|\rho| < 1$; existe un problema de autocorrelación y por lo tanto el estimador de la matriz de

varianzas y covarianzas es _____. Y además, si se realiza un test de DW para contrastar la existencia de autocorrelación que arroja un valor de DW = 1,2. Entonces, la autocorrelación será _____. El apartado se completa con:

- Sesgado, autocorrelación negativa
- Sesgado, autocorrelación positiva
- Ineficiente, autocorrelación negativa
- Ninguna de las anteriores

Respuesta: d, al existir autocorrelación la matriz de varianzas y covarianzas es sesgada. Pero, el modelo no presenta intercepto por lo que el DW no tienen sentido y no se puede determinar el tipo de autocorrelación.

3 (30 puntos)

La Organización Internacional del Trabajo (OIT) desea conocer el efecto de los sindicatos sobre la productividad por trabajador en el sector manufacturero del país Nueva Granada, el cual ha sido uno de los más afectados por la crisis económica. Para esto contrata dos econométristas, quienes se basan en algunas regularidades empíricas, para establecer que la producción por trabajador (y_t) depende del capital por trabajador (k_t), ambas variables medidas en millones de dólares constantes de 2000, y la proporción total de trabajadores que están afiliados a un sindicato (P_t). El modelo que sugiere la teoría emplear es:

$$\ln(y_t) = \beta_0 + \beta_1 \ln(k_t) + \beta_2 P_t + \beta_3 P_t^2 + e_t$$

Responda brevemente cada una de las siguientes preguntas empleando la información que se presenta al final del examen (tabla 1 y tabla 2).

- a) Escriba el modelo estimado por los econométrista, presentado en la tabla 1. (2 puntos)

El modelo estimado por los econométristas debería ser el siguiente:

$$\ln(y_t) = \beta_0 + \beta_1 \ln(k_t) + \beta_2 P_t + e_t$$

- b) ¿Cree usted que el econométrista cometió algún error o no, al estimar ese modelo? Sea lo más claro posible y argumente su respuesta. (7 puntos)

Noten que en este caso no hay porque esperar multicolinealidad perfecta. en el modelo teórico planteado. Las variables P_t y P_t^2 no tiene una relación lineal. Por eso omitir la variable P_t^2 argumentando que existe multicolinealidad no es una razón válida. (menos 7 puntos si el estudiante argumentó que este era el problema).

El econométrista si incurrió en el error de omitir una variable relevante al momento de estimar el modelo. (5 puntos) De hecho es muy probable que la covarianza entre P_t y P_t^2 sea igual a cero y por tanto no necesariamente existirá un sesgo por omisión de variables. (menos 2 puntos sino se resalta esto último)

Noten que la presencia de autocorrelación en el modelo de la tabla 1 no implica un error del econométrista, es un problema de la muestra. Pero en ningún momento la presencia de un

problema econométrico como la autocorrelación es fruto de un error del econométrico. (menos dos puntos por considerar la autocorrelación como un error)

- c) El investigador, estimó dos modelos. con miedo de tener un problema econométrico construye otro modelo y lo estima. Este segundo modelo se reporta en la Tabla 2. ¿Tiene razón en estimar este segundo modelo? (Sea lo más preciso y emplee únicamente la información disponible) **(5 puntos)**
 Como se observa en los resultados la prueba DW muestra la presencia de autocorrelación negativa para el primer modelo estimado, ya que el Durbin Watson es igual a 3,0532, es decir es mayor a 2. (Se necesitaba hacer la prueba formal 4 puntos). Esto hace razonable estimar el segundo modelo (1 punto).
- d) ¿Cómo fue corregido el problema si es que existía uno? ¿Cuál modelo debería emplear para sacar las conclusiones del estudio? **(7 puntos)**

TABLA CON CORRECCIÓN DE DURBIN DONDE DW=1,896.

El problema fue corregido por el método de Diferencias Generalizadas (explicar brevemente en qué consiste esta corrección). Dado que probablemente no conocían el coeficiente de autocorrelación debieron haberlo estimado primero, obteniendo un resultado de: $\hat{\rho} = 0.082733$ Con este valor se puede transformar el modelo de la siguiente manera:

$$w = \beta_0^* + \beta_1 Z + \beta_2 I + e_t^*$$

Donde $w = \ln(y_t) - \hat{\rho} \ln(y_{t-1})$; $Z = \ln(k_t) - \hat{\rho} \ln(k_{t-1})$; $I = P_t - \hat{\rho} P_{t-1}$; $\beta_0^* = \beta_0(1 - \hat{\rho})$
 (2 puntos)

El modelo corregido está libre de problemas, pues ya no hay autocorrelación, ya que el DW=1,99 es muy cercano a 2 (se necesitaba realizar la prueba formal, menos 1 punto). Así se debe emplear el segundo modelo para sacar conclusiones (4 Puntos).

- e) Interprete los coeficientes del modelo que crea más conveniente **(9 puntos)**

$\hat{\beta}_0 = 0.3134$ No tiene interpretación económica. (3 Puntos)

$\hat{\beta}_1 = 2.29$ Es significativo al 99% de confianza. Representa la elasticidad del producto per cápita respecto al capital per cápita. Por tanto indica que por un aumento de 1% en el capital per cápita, la producción per cápita incrementa en 2.29%. (3 Puntos)

$\hat{\beta}_2 = 31.91$ Es significativo al 99% de confianza. Indica que por un aumento de un punto porcentual en la proporción total de trabajadores sindicalizados, la producción per cápita incrementa en 3191%. (3 Puntos)

4 (40 puntos)

El jefe de estudios económicos de una multinacional productora de papel desea estimar la función de demanda del mercado de papel. Para tal fin, se desea emplear la siguiente función de demanda:

$$Q_i = \beta_1 + \beta_2 P_i + \beta_3 \frac{1}{PI} + \beta_4 Y_i^2 + \varepsilon_i \quad (1)$$

Donde el Q_i corresponde al logaritmo del número de resmas de papel tamaño carta que desea comprar el hogar i al precio P_i (medido en miles de pesos). Además, PI es el precio de la impresora más común en el mercado (este precio es igual para todos los hogares) y Y_i representan el ingreso de cada hogar (medido en pesos). Por otro lado, se cuenta con una muestra de tamaño 15. Finalmente, ε_i es el término de error que cumple las siguientes características: $E[\varepsilon] = 0$, $Var[\varepsilon] = PI^2 \sigma^2 I_{15}$ (I_{15} corresponde a la matriz identidad de tamaño 15).

A partir de la información anterior conteste las siguientes preguntas:

- a) Si el jefe de estudios económicos emplea el método de mínimos cuadrados ordinarios (MCO), ¿existe algún problema que pudiese no permitir obtener conclusiones “adecuadas” sobre el efecto del precio sobre la demanda de resmas de papel? Sea lo más claro posible. **(10 puntos)**

El modelo **¡Error! No se encuentra el origen de la referencia.** no puede ser estimado directamente, pues PI es un valor constante para todos los hogares, en este caso podemos reparametrizar el modelo de la siguiente manera:

$$Q_i = \beta_1 + \beta_3 \frac{1}{PI} + \beta_2 P_i + \beta_4 Y_i^2 + \varepsilon_i$$

$$Q_i = \gamma + \beta_2 P_i + \beta_4 Y_i^2 + \varepsilon_i$$

Donde $\gamma = \beta_1 + \beta_3 \frac{1}{PI}$.

Noten que en este caso todos los supuestos del teorema de Gauss-Markov se cumplen. Es decir:

- > Relación lineal entre la variable dependiente y los regresores.
- > Los regresores deben ser no estocásticos y linealmente independientes entre sí.
- > Los errores deben:
 - Tener media cero
 - Varianza constante
 - Y no estar autocorrelacionados

Esto implica que los estimadores MCO del modelo reparametrizado serán MELI.

Nota: Si no se menciona nada de lo hasta aquí discutido (i.e. no se reparametriza correctamente el modelo y no se reconoce que no hay problemas en el error), se tendrá un error en el siguiente punto. Por tanto no se dará ningún crédito en el punto b. No se penalizará dos veces este error. Así en esta pregunta no se quitará puntos

por esto.) No obstante si se afirma que hay heteroscedasticidad si se descontará 3 punto en esta pregunta y los correspondientes puntos en la parte b.

Por otro lado, dado que la muestra es pequeña (n=15) puede existir un problema al hacer inferencias. Si el error no sigue una distribución normal, entonces no se puede hacer inferencia y por tanto no se podrán sacar conclusiones sobre el efecto del precio sobre la demanda de resmas de papel. (10 puntos). (si no se reconoce la posible no normalidad de los errores, no se dará nada de crédito.)

- b) El econometrista a cargo del estudio recolecta la información obtener estimadores MELI para el modelo (1). El obtiene las siguientes matrices que corresponden al equivalente de la matriz $X^T X$ y $X^T y$. NOTA: Para la construcción de estas matrices, el modelo se organizó lo más parecido al modelo original:

$$X^T X = \begin{bmatrix} 15 & 0 & 10 \\ 0 & 5 & 0 \\ 10 & 0 & 15 \end{bmatrix} \quad X^T y = \begin{bmatrix} 45 \\ -15 \\ -20 \end{bmatrix}$$

Explique claramente a que corresponde cada uno de los elementos de la matriz $X^T X$. (Por ejemplo, explique a partir de que sumatoria sale el 10 que corresponde al último elemento de la matriz $X^T X$, y así sucesivamente con cada elemento de la matriz) **Expresé su respuesta en términos de las variables en el modelo original y no en términos de variables reparametrizadas** (por ejemplo P_i). (6 puntos)

En este caso

$$n = 15, \sum_{i=1}^{15} P_i = \sum_{i=1}^{15} P_i Y_i^2 = 0, \sum_{i=1}^{15} Y_i^2 = 10, \sum_{i=1}^{15} (P_i)^2 = 5, \text{ y } \sum_{i=1}^{15} Y_i^4 = 15$$

1 punto por cada una de estas sumatorias

Nota: Si no se reparametriza correctamente el modelo, se tendrá un error y por tanto no se dará ningún crédito en este punto. No se penalizará dos veces este error). Si este es su caso, no se tendrá en cuenta si tiene bien el tamaño de la muestra.

- c) Encuentre los estimadores MELI para los coeficientes que se puedan recuperar del **modelo ¡Error! No se encuentra el origen de la referencia.. MUESTRE** claramente el valor estimado y su correspondiente parámetro poblacional. (12 Puntos)

En este caso tenemos que:

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 3/25 & 0 & -2/25 \\ 0 & 1/5 & 0 \\ -2/25 & 0 & 3/25 \end{bmatrix} \begin{bmatrix} 45 \\ -15 \\ -20 \end{bmatrix} = \begin{bmatrix} 7 \\ -3 \\ -6 \end{bmatrix}$$

Es decir,

$$\hat{\beta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\beta}_2 \\ \hat{\beta}_4 \end{bmatrix} = \begin{bmatrix} 7 \\ -3 \\ -6 \end{bmatrix}$$

Noten que β_1 y β_2 no se pueden recuperar. Si no muestra a que corresponde cada coeficiente estimado se restarán 4 puntos.

- d) Interprete el significado de cada uno de los coeficientes estimados.(6 Puntos)

$\hat{\gamma} = 7$ No tiene interpretación económica. (2 puntos)

$\hat{\beta}_2 = -3$. Un aumento de mil pesos en el precio de la resma de papel generará una caída en la demanda de papel del 300%. (2 puntos)

$\hat{\beta}_4 = -6$. Un aumento de un peso en el ingreso de las familias, provocará una disminución de 1200Y_i % (2 puntos)

- e) El jefe de estudios económicos de la compañía después de realizar las estimaciones encuentra que:

- Según el modelo teórico más empleado para la estimación de la demanda, es necesario incluir en el modelo el tamaño de la familia.
- El ingreso de las personas no fue medido con la precisión requerida. De hecho, se encontró que existe un error de medición que es aleatorio.

De acuerdo a estos hallazgos, que puede esperar el jefe de estudios sobre los resultados obtenidos hasta ahora y en especial sobre las conclusiones que se saquen. Sea lo más claro posible, explique todas sus afirmaciones (6 Puntos)

1. Se ha omitido una variable del modelo, que muy probablemente está relacionada con el ingreso de la familia, y por tanto se puede obtener estimadores sesgados e inconsistentes (2 puntos)
2. Por otro lado al medir una variable explicatoria con error, se puede obtener estimadores sesgados e inconsistentes (2 puntos)
3. Finalmente, es importante mencionar que en todo caso es imposible hacer inferencia sobre los coeficientes antes de estos hallazgos o después, por tanto no se pueden sacar conclusiones sobre los valores poblacionales. (2 puntos)

**Resultados de EasyReg:
Tabla 1**

Dependent variable:
Y = ln[Y]

Characteristics:
Ln[Y]
First observation = 1(=1951)
Last observation = 55(=2005)
Number of usable observations: 55
Minimum value: 5.587000E+000
Maximum value: 1.9781000E+001
Sample mean: 1.3118156E+001

X variables:
X(1) = ln(K)
X(2) = P
X(3) = 1

WARNING: The effective degrees of freedom is only 28.
Therefore, the estimation results may be unreliable!

Model:
Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + U,
where U is the error term, satisfying
E[U|X(1),X(2),X(3)] = 0.

OLS estimation results

Parameters	Estimate	t-value (S.E.)	H.C. t-value (H.C. S.E.)
		[p-value]	[H.C. p-value]
b(1)	33,984131	33.984 (0.01000)	41.543 (0.01003)
		[0.00000]	[0.00000]
b(2)	25,560146	28.339 (0.01962)	-2.030 (0.02759)

	[0.00430]	[0.04235]
b(3)	0.2134671	0.983
	(0.21715)	(0.22985)
	[0.32560]	[0.35304]

Notes:
1: S.E. = Standard error
2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
3: The two-sided p-values are based on the normal approximation.

Effective sample size (n): 55
Variance of the residuals: 0.12151531
Standard error of the residuals (SER): 0.34859046
Residual sum of squares (RSS): 3.40242872
(Also called SSR = Sum of Squared Residuals)
Total sum of squares (TSS): 788.65192071
R-square: 0.9957
Adjusted R-square: 0.8726

Overall F test: F(2,53) = 32245,98
p-value = 0.00000
Significance levels: 10% 5%
Critical values: 2.5 3.34
Conclusions: reject reject

Test for first-order autocorrelation:
Durbin-Watson test = 3,0532

Jarque-Bera/Salmon-Kiefer test = 0.094518
Null hypothesis: The errors are normally distributed
Null distribution: Chi-square(2)

p-value = 0.95384
 Significance levels: 10% 5%
 Critical values: 4.61 5.99
 Conclusions: accept accept

Breusch-Pagan test = 2.200154
 Null hypothesis: The errors are homoskedastic
 Null distribution: Chi-square(2)
 p-value = 0.33285
 Significance levels: 10% 5%
 Critical values: 4.61 5.99
 Conclusions: accept accept

Tabla 2. Resultados de EasyReg

Dependent variable:
 $Y = \ln[Y] + .082733LAG1[\ln[Y]]$
 Characteristics:
 $\ln[Y].082733x LAG1[\ln[Y]]$
 First observation = 1(=1951)
 Last observation = 55(=2005)
 Number of usable observations: 55
 Minimum value: 5.5870000E+000
 Maximum value: 1.9781000E+001
 Sample mean: 1.3118156E+001

X variables:
 $X(1) = \ln[K] + .082733 LAG1[\ln[K]]$
 $X(2) = P + .082733LAG1[P]$
 $X(3) = 1$
 WARNING: The effective degrees or freedom is only 28.
 Therefore, the estimation results may be unreliable!

Model:
 $Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + U,$

where U is the error term, satisfying
 $E[U|X(1),X(2),X(3)] = 0.$

OLS estimation results

Parameters	Estimate	t-value (S.E.)	H.C. t-value (H.C. S.E.)
		[p-value]	[H.C. p-value]
b(1)	2.2968185	1.502 (0.19765) [0.00000]	1.360 (0.21817) [0.00000]
b(2)	31,91243	28.339 (0.01962) [0.00330]	-2.030 (0.02759) [0.03235]
b(3)	0.3134671	0.983 (0.21715) [0.32560]	0.929 (0.22985) [0.35304]

Notes:

- 1: S.E. = Standard error
- 2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
- 3: The two-sided p-values are based on the normal approximation.

Effective sample size (n): 55
 Variance of the residuals: 0.10845059
 Standard error of the residuals (SER): 0.32931837
 Residual sum of squares (RSS): 3.14506713
 (Also called SSR = Sum of Squared Residuals)
 Total sum of squares (TSS): 3.38921543
 R-square: 0.0720
 Adjusted R-square: 0.0720
 Overall F test: $F(2,53) = 145445,98$

p-value = 0.00000

Significance levels: 10% 5%

Critical values: 2.5 3.34

Conclusions: reject reject

Test for first-order autocorrelation:

Durbin-Watson test = 1,896

Jarque-Bera/Salmon-Kiefer test = 0.743108

Null hypothesis: The errors are normally distributed

Null distribution: Chi-square(2))

p-value = 0.68966

Significance levels: 10% 5%

Critical values: 4.61 5.99

Conclusions: accept accept

Breusch-Pagan test = 1,93412

Null hypothesis: The errors are homoskedastic

Null distribution: Chi-square(2)

p-value = 0.92285

Significance levels: 10% 5%

Critical values: 4.61 5.99

Conclusions: accept accept

Econometría 06169
Examen Parcial #2
Cali, lunes 5 de Octubre de 2009

Profesores: Julio César Alonso y Carlos Giovanni González

Estudiante: _____

Código: _____

Instrucciones:

1. Lea cuidadosamente todas las preguntas e instrucciones.
2. Este examen consta de **9** páginas; además, deben tener 2 páginas de fórmulas.
3. El examen consta de 4 preguntas que suman un total de 100 puntos. El valor de cada una de las preguntas esta expresado al lado de cada pregunta.
4. Escriba su respuesta en las hojas suministradas, marque cada una de las hojas con su nombre. **NO** responda en las hojas de preguntas.
5. El examen está diseñado para dos horas, pero ustedes tienen 3 horas para trabajar en él.
6. Recuerde que no se tolerará ningún tipo de deshonestidad académica. En especial usted no puede emplear ningún tipo de ayuda diferente a la que se le entrega con este examen.
7. El uso de calculadoras está prohibido.
8. No se aceptarán reclamos de respuestas escritas en lápiz.
9. Al finalizar su examen entregue sus hojas de respuesta, así como las hojas de preguntas.
10. Asigne su tiempo de forma eficiente!

Buena Suerte!

1 Falso o Verdadero (15 puntos en total, 5 puntos cada subparte)

Diga si las siguientes afirmaciones son falsas o verdaderas y explique brevemente su respuesta. (No se dará ningún crédito a respuestas sin justificación.)

- a) Considere el siguiente modelo: $y_i = \beta_0 + \sum_{k=1}^4 \beta_k D_{k,i} + \varepsilon_i$, donde y_i es el rendimiento de la acción de suramericana para el día i y $D_{k,i}$ toma el valor de uno si i es el k -ésimo día laboral de la semana y cero en caso contrario. Además, suponga que se emplea una muestra para estimar este modelo que está compuesta por los últimos 2000 días hábiles en los que el mercado accionario estuvo abierto. Además **es un hecho**, demostrado por muchos estudios, que “cuando los rendimientos de la acción de Suramericana de los lunes tienden a ser grandes, el rendimiento del viernes es alto. Y cuando el viernes es bajo el lunes también será bajo. Es más existe una correlación entre los rendimientos de esos dos días de 0.9”. Sobre este modelo un investigador afirma que “Dada la evidencia que existe sobre la relación entre los retornos de la acción de Suramericana de los lunes y los viernes, se tendrá un problema de multicolinealidad en el modelo” ¿Es esta afirmación verdadera o falsa?
- b) Considere el caso de una regresión simple: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. Donde $Var(X_i) = i$, ε_i cumple los supuestos del Teorema de Gauss-Markov y $Cov[\varepsilon_i, X_i] = Cov[\varepsilon_i, \beta_{j,i}] = 0$ ($j = 0, 1$). En este caso podemos afirmar que el estimador MCO para β_0 será sesgado pero consistente.
- c) Suponga el modelo $\ln y_i = \beta_1 + \beta_2 \ln x_i + \varepsilon_i$, en donde ε_i es el término de error y $var(\varepsilon_i) = \sigma^2 D_i$, donde D_i es una variable dummy que toma el valor de uno si el individuo es profesional y cero en caso contrario. Un estudiante afirma: “Se puede transformar el modelo dividiendo por D_i , de modo que los estimadores MCO de los β 's sean insesgados y eficientes.

2 Selección Múltiple (15 puntos en total, 5 puntos cada subparte)

Determine cuál de las siguientes respuestas es la correcta. Escoja la mejor opción y explique en dos o tres líneas su respuesta. (No se dará ningún crédito a respuestas sin justificación.)

- i. Un econométra desea estimar el siguiente modelo $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$; donde, $X_2 = 0.95 X_3$. para descartar un problema econométrico el investigador decide realizar la siguiente prueba $K(X) = \frac{\sqrt{\lambda_{max}}}{\sqrt{\lambda_{min}}}$. El resultados de la prueba seria:
- $k(x) = 1$
 - $k(x) = 0$
 - $k(x) = 25$
 - Ninguna de las anteriores
- ii. En el siguiente modelo $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$; ($t=1, \dots, N$), las perturbaciones presentan heteroscedasticidad si:
- $cov(X_t * \varepsilon_t) = 0$; para todo $t=1, \dots, N$.
 - $var(\varepsilon_t) = t * \sigma^2$; para todo $t=1, \dots, N$.
 - $var(\varepsilon_t) = c * \sigma^2$; para todo $t=1, \dots, N$.

- d. $\varepsilon_t = 10 + \mu_t$; con $var(\varepsilon_t) = 5$, para todo $t=1, \dots, N$.
- iii. Un econométra es contratado para estimar la cantidad demandada de carros de una de las marcas más populares “HIUNDAI” que tienen quince cedes en todo el país, y por lo tanto, el econométra tiene información de corte transversal. La cantidad demandada se estimara en función de ciertas variables relevantes X, pero este incluye algunas variables irrelevantes Z. Y con la información anterior estima el siguiente modelo $\hat{Y}_i = 100 + 0,4X_{1i} + 1,7X_{2i} + 0,2Z_{1i} + 1,2Z_{2i}$, para probar la significancia individual del parámetro que acompaña a Z_{1i} (una de las variables irrelevantes) tenemos un valor P igual a 0,05. ¿Cuál de las siguientes afirmaciones es correcta?:
- El parámetro que acompaña a Z_{1i} es significativo al 95%.
 - El parámetro que acompaña a Z_{1i} no es significativo al 95%.
 - El t calculado no sirve por la presencia de variables irrelevantes.
 - Ninguna de las anteriores

3 (30 puntos)

Usted es contratado por el Ministerio de educación para realizar un estudio sobre la tasa de retorno (rentabilidad privada) de la educación en su país. El Ministro le encarga que su propuesta se base en la teoría del capital humano, específicamente que estime una modificación de la ecuación de salarios propuesta por Mincer (1974). Es decir, se propone el siguiente modelo teórico:

$$\ln(W_i) = \gamma_0 + \gamma_1 educ_i + \gamma_2 exp_i + \gamma_3 gene_i + \gamma_4 gene_i^2 + \mu_i$$

Para realizar dicha estimación el Ministerio le entrega la Encuesta Nacional de Hogares con las siguientes variables: salario mensual en pesos (W_i), los años de educación ($educ_i$), la edad ($edad_i$), la experiencia laboral, medida en años (exp_i) y el género, mujer =1, ($gene_i$).

Además, el jefe del equipo técnico del Ministerio le informa que por estudios previos realizados por el Staff, posiblemente la $var(\mu_i) = X_i\sigma^2$.

- El econométrista estimó el modelo reportado en la tabla 1. Escriba el modelo estimado por el. (2 puntos)
- ¿Cree usted que el econométrista cometió algún error o no, al estimar ese modelo? Sea lo más claro posible y argumente su respuesta. (7 puntos)
- El investigador, con miedo de tener un problema econométrico debido a la información suministrada por el staff del Ministerio construye otro modelo y lo estima. Este segundo modelo se reporta en la Tabla 2. Escriba el segundo modelo estimado por el econométrista. (2 puntos)
- Dada toda la información disponible en este examen, ¿Cuál de los dos modelos debería emplear el investigador? Sea lo más claro posible y en caso de poder hacer pruebas efectúelas. (10 puntos)
- Independientemente de sus respuestas anteriores y de la presencia o no de problemas, interprete los coeficientes del modelo reportado en la Tabla 1. Teniendo en cuenta su significancia (9 puntos).

4 (40 puntos)

La división de mercadeo de una empresa productora de cepillos de dientes desea estimar la elasticidad precio de la demanda de su producto estrella en Colombia. Para tal fin, emplean datos para los últimos 15 meses y el siguiente modelo:

$$Q_t = \beta_1 + \beta_2 \ln(P_t) + \beta_3 \frac{1}{Av_t} + \beta_4 IPI_t^2 + \varepsilon_t \quad (1)$$

Donde el Q_t corresponde al número de paquetes de 100 cepillos de dientes que fueron comprados en Colombia el mes t al precio P_t (medido en miles de pesos). Además, Av_t corresponde al número de avisos publicitarios en la televisión (el número de avisos ha sido igual para todos los meses). Finalmente, IPI_t es el índice de producción industrial en Colombia para el mes t que se emplea como proxy del nivel de actividad económica (el mes uno de la muestra corresponde a la base de este índice, es decir ($IPI_1 = 100$)). Finalmente, ε_t es el término de error que cumple las siguientes características: $E[\varepsilon] = 0$, $Var[\varepsilon] = Av_t^2 \sigma^2 I_{15}$ (I_{15} corresponde a la matriz identidad de tamaño 15).

A partir de la información anterior conteste las siguientes preguntas:

- a) Si el jefe de estudios económicos emplea el método de mínimos cuadrados ordinarios (MCO), ¿existe algún problema que pudiese no permitir obtener conclusiones “adecuadas” sobre la elasticidad precio de la demanda de su producto estrella? Sea lo más claro posible. **(10 puntos)**
- b) El econometrista a cargo del estudio recolecta la información obtener estimadores MELI para el modelo (1). El obtiene las siguientes matrices que corresponden al equivalente de la matriz $X^T X$ y $X^T y$. NOTA: Para la construcción de estas matrices, el modelo se organizó lo más parecido al modelo original:

$$X^T X = \begin{bmatrix} 15 & 0 & 5 \\ 0 & 10 & 0 \\ 5 & 0 & 5 \end{bmatrix} \quad X^T y = \begin{bmatrix} 10 \\ -15 \\ 20 \end{bmatrix}$$

Explique claramente a que corresponde cada uno de los elementos de la matriz $X^T X$. (Por ejemplo, explique a partir de que sumatoria sale el 5 que corresponde al último elemento de la matriz $X^T X$, y así sucesivamente con cada elemento de la matriz) **Expresar su respuesta en términos de las variables en el modelo original y no en términos de variables reparametrizadas** (por ejemplo P_t). **(6 puntos)**

- c) Encuentre los estimadores MELI para los coeficientes que se puedan recuperar del **modelo (1)**. **MUESTRE** claramente el valor estimado y su correspondiente parámetro poblacional. **(12 Puntos)**
- d) Interprete el significado de cada uno de los coeficientes estimados. **(6 Puntos)**
- e) El jefe de estudios económicos de la compañía después de realizar las estimaciones encuentra que:
 - Según el modelo teórico más empleado para la estimación de la demanda, es necesario incluir en el modelo el número de cepillos vendidos en los Estados Unidos.
 - El IPI no fue medido con la precisión requerida. De hecho, se encontró que existe un error que es siempre del 5% en su medición.

De acuerdo a estos hallazgos, que puede esperar el jefe de estudios sobre los resultados obtenidos hasta ahora y en especial sobre las conclusiones que se saquen en torno a la elasticidad deseada. Sea lo más claro posible, explique todas sus afirmaciones **(6 Puntos)**

Resultados de EasyReg:

Tabla 1

Dependent variable:			
Y = ln[W]			
Characteristics:			
Ln[W]			
First observation = 1			
Last observation = 2000			
Number of usable observations: 2000			
Minimum value: 5.5870000E+000			
Maximum value: 1.9781000E+001			
Sample mean: 1.3118156E+001			
X variables:			
X(1) = EDUC			
X(2) = EXP			
X(3) = GENE			
X(4) = 1			
Model:			
Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + b(4)X(4) + U,			
where U is the error term, satisfying			
E[U X(1),X(2),X(3), X(4)] = 0.			
OLS estimation results			
Parameters	Estimate	t-value	H.C. t-value
		(S.E.)	(H.C. S.E.)
		[p-value]	[H.C. p-value]
b(1)	0,1524131	33.984	41.543
		(0.01000)	(0.01003)
		[0.00000]	[0.00000]
b(2)	0,026146	28.339	-2.030
		(0.01962)	(0.02759)
		[0.00430]	[0.04235]

b(3)	0.0024671	0.983	0.929
		(0.21715)	(0.22985)
		[0.032560]	[0.035304]
b(4)	0,1260146	28.339	-2.030
		(0.01962)	(0.02759)
		[0.0230]	[0.02235]
Notes:			
1: S.E. = Standard error			
2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.			
3: The two-sided p-values are based on the normal approximation.			
Effective sample size (n):	2000		
Variance of the residuals:	0.12151531		
Standard error of the residuals (SER):	0.34859046		
Residual sum of squares (RSS):	3.40242872		
(Also called SSR = Sum of Squared Residuals)			
Total sum of squares (TSS):	788.65192071		
R-square:	0.9957		
Adjusted R-square:	0.8726		
Overall F test: $F(2,22) = 32245,98$			
p-value = 0.00000			
Significance levels:	10%	5%	
Critical values:	2.5	3.34	
Conclusions:	reject	reject	
Test for first-order autocorrelation:			
Durbin-Watson test = 1,894			
Jarque-Bera/Salmon-Kiefer test = 0.094518			

Null hypothesis: The errors are normally distributed		
Null distribution: Chi-square(2)		
p-value = 0.95384		
Significance levels:	10%	5%
Critical values:	4.61	5.99
Conclusions:	accept	accept
Breusch-Pagan test = 13,93412		
Null hypothesis: The errors are homoskedastic		
Null distribution: Chi-square(2)		
p-value = 0.003285		
Significance levels:	10%	5%
Critical values:	4.61	5.99
Conclusions:	reject	reject

Tabla 2. Resultados de EasyReg

Dependent variable:
$Y = \ln[Y]/X$
Characteristics:
$\ln[Y]/X$
First observation = 1
Last observation = 2000
Number of usable observations: 2000
Minimum value: 5.5870000E+000
Maximum value: 1.9781000E+001
Sample mean: 1.3118156E+001
X variables:
X(1) = EDUC/X
X(2) = EXP/X
X(3) = GENE/X
X(4) = 1/X
Model:

$$Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + b(4)X(4) + U,$$

where U is the error term, satisfying

$$E[U|X(1),X(2),X(3), X(4)] = 0.$$

OLS estimation results

Parameters	Estimate	t-value (S.E.) [p-value]	H.C. t-value (H.C. S.E.) [H.C. p-value]
b(1)	0.12968185	1.502 (0.19765) [0.00000]	1.360 (0.21817) [0.00000]
b(2)	0,011243	28.339 (0.01962) [0.00330]	-2.030 (0.02759) [0.03235]
b(3)	0.0314671	0.983 (0.21715) [0.32560]	0.929 (0.22985) [0.35304]
b(4)	0.154671	0.983 (0.21715) [0.032560]	0.929 (0.22985) [0.035304]

Notes:

- 1: S.E. = Standard error
- 2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
- 3: The two-sided p-values are based on the normal approximation.

Effective sample size (n):	2000
Variance of the residuals:	0.10845059
Standard error of the residuals (SER):	0.32931837
Residual sum of squares (RSS):	3.14506713
(Also called SSR = Sum of Squared Residuals)	
Total sum of squares (TSS):	3.38921543

R-square:	0.0720	
Adjusted R-square:	0.0720	
Overall F test: $F(2,22) = 23225,98$		
p-value = 0.00000		
Significance levels:	10%	5%
Critical values:	2.5	3.34
Conclusions:	reject	reject
Test for first-order autocorrelation:		
Durbin-Watson test = 1,96		
Jarque-Bera/Salmon-Kiefer test = 0.743108		
Null hypothesis: The errors are normally distributed		
Null distribution: Chi-square(2)		
p-value = 0.68966		
Significance levels:	10%	5%
Critical values:	4.61	5.99
Conclusions:	accept	accept
Breusch-Pagan test = 2,93412		
Null hypothesis: The errors are homoskedastic		
Null distribution: Chi-square(2)		
p-value = 0.28590		
Significance levels:	10%	5%
Critical values:	4.61	5.99
Conclusions:	accept	accept

Prof: Julio César Alonso Carlos Giovanni González

$$X^T X = \begin{bmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} & \cdots & \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i}X_{2i} & \cdots & \sum_{i=1}^n X_{1i}X_{ki} \\ \sum_{i=1}^n X_{2i}^2 & \ddots & \sum_{i=1}^n X_{2i}X_{ki} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^n X_{ki}^2 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i X_{1i} \\ \sum_{i=1}^n y_i X_{2i} \\ \vdots \\ \sum_{i=1}^n y_i X_{ki} \end{bmatrix} \quad y^T y = \sum_{i=1}^n y_i^2$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 = y^T y - n\bar{y}^2 \quad SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad s^2 = \frac{SSE}{n-k} = \frac{y^T y - \hat{\beta}^T X^T y}{n-k}$$

$$Var[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{\beta}^T X^T y - n\bar{y}^2 \quad t = \frac{\hat{\beta}_i - c}{s_{\hat{\beta}_i}}$$

$$F_c = \frac{(c - R\hat{\beta})^T (R(X^T X)^{-1} R^T)^{-1} (c - R\hat{\beta}) / r}{SSE / (n-k)}$$

$$F_c = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)} = \frac{MSR}{MSE}$$

$$F_c = \frac{(SSE_R - SSE_U) / r}{SSE_U / (n-k)} \quad R^2 = \frac{SSR}{SST}$$

$$\hat{\beta}_i \pm t_{\frac{\alpha}{2}, n-k} s_{\hat{\beta}_i} \quad \bar{R}^2 = 1 - (1-R^2) \frac{n-1}{n-k}$$

$$\hat{y}_p = x_p^T \hat{\beta}, \quad x_p^T = (1 \quad x_{1p} \quad x_{2p} \quad \cdots \quad x_{kp})$$

$$\hat{y}_p \pm t_{\frac{\alpha}{2}, n-k} \sqrt{\sigma^2 x_p^T (X^T X)^{-1} x_p}$$

$$\hat{y}_p \pm t_{\frac{\alpha}{2}, n-k} \sqrt{\sigma^2 [1 + x_p^T (X^T X)^{-1} x_p]}$$

$$\hat{\beta}_j^E = \hat{\beta}_j \frac{s_{X_j}}{s_y}, \quad j = 2, 3, \dots, k \quad E_j = \hat{\beta}_j \frac{\bar{X}_j}{\bar{y}}$$

Test de Heteroscedasticidad

Goldfeld y Quand: $F_{GQ} = \frac{SSE_2}{SSE_1} \sim F_{(n-d-2k, n-d-2k)}$

Prof: Julio César Alonso Carlos Giovanni González

Breush-Pagan: $\frac{\hat{\epsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta Z_i + \mu_i, BP = \frac{SSR}{2} \sim \chi_g^2$

White: $\hat{\epsilon}_i^2 = \gamma + \sum_{m=1}^k \sum_{j=1}^k \delta_s X_{mi} X_{ji} + \mu_i, W_a = nR^2 \sim \chi_g^2$

d_l y d_u para el test de DW al nivel de significancia del 5%

N	k-1=1		k-1=2		k-1=3	
	d _l	d _u	d _l	d _u	d _l	d _u
50	1.50	1.59	1.46	1.63	1.42	1.67
60	1.55	1.62	1.51	1.65	1.48	1.69
95	1.64	1.69	1.62	1.71	1.60	1.73
100	1.65	1.69	1.63	1.72	1.61	1.74

Test de Autocorrelación

Durbin-Watson $DW \approx 2(1 - \hat{\rho})$

Ho	Sí	Decisión
$H_0 : \rho = 0$	$d_u < DW < 4 - d_u$	A
No auto +	$0 < DW < d_l$	R
No auto -	$4 - d_l < DW < 4$	R

Área de indecisión $d_l < DW < d_u$ y $4 - d_u < DW < 4 - d_l$

$h = \hat{\rho} \sqrt{\frac{n}{1 - n(\widehat{Var}(\hat{\alpha}))}}$ donde

$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + \alpha y_{t-1} + \epsilon_t$

Cantidades Importantes

$\sqrt{2} = 1.414$
 $\sqrt{10} = 3.162$

$\sqrt{3} = 1.732$
 $\sqrt{13} = 3.606$

$\sqrt{5} = 2.236$

$\sqrt{7} = 2.646$

Econometría 06169
Examen Parcial #2
Respuestas Sugeridas
Cali, lunes 5 de Octubre de 2009

Profesores: Julio César Alonso y Carlos Giovanni González

Estudiante: _____

Código: _____

Instrucciones:

1. Lea cuidadosamente todas las preguntas e instrucciones.
2. Este examen consta de 5 páginas; además, deben tener 3 páginas de fórmulas.
3. El examen consta de 4 preguntas que suman un total de 100 puntos. El valor de cada una de las preguntas esta expresado al lado de cada pregunta.
4. Escriba su respuesta en las hojas suministradas, marque cada una de las hojas con su nombre. NO responda en las hojas de preguntas.
5. El examen está diseñado para una hora, pero ustedes tienen 2 horas para trabajar en él.
6. Recuerde que no se tolerará ningún tipo de deshonestidad académica. En especial usted no puede emplear ningún tipo de ayuda diferente a la que se le entrega con este examen.
7. Al finalizar su examen entregue sus hojas de respuesta, así como las horas de preguntas.
8. Asigne su tiempo de forma eficiente!

Buena Suerte!

1 Falso o Verdadero (15 puntos en total, 5 puntos cada subparte)

Diga si las siguientes afirmaciones son falsas o verdaderas y explique brevemente su respuesta. (No se dará ningún crédito a respuestas sin justificación.)

- a) Considere el siguiente modelo: $y_i = \beta_0 + \sum_{k=1}^4 \beta_k D_{k,i} + \varepsilon_i$, donde y_i es el rendimiento de la acción de suramericana para el día i y $D_{k,i}$ toma el valor de uno si i es el k -ésimo día laboral de la semana y cero en caso contrario. Además, suponga que se emplea una muestra para estimar este modelo que está compuesta por los últimos 2000 días hábiles en los que el mercado accionario estuvo abierto. Además *es un hecho*, demostrado por muchos estudios, que “cuando los rendimientos de la acción de Suramericana de los lunes tienden a ser grandes, el rendimiento del viernes es alto. Y cuando el viernes es bajo el lunes también será bajo. Es más existe una correlación entre los rendimientos de esos dos días de 0.9”. Sobre este modelo un investigador afirma que “Dada la evidencia que existe sobre la relación entre los retornos de la acción de Suramericana de los lunes y los viernes, se tendrá un problema de multicolinealidad en el modelo” ¿Es esta afirmación verdadera o falsa?

Falso, pues recuerden que la multicolinealidad se presenta cuando existe cierto grado de relación lineal entre las variables explicativas. En este modelo las variables explicativas son variables dummy, las cuales no tienen ninguna relación lineal entre ellas.

- b) Considere el caso de una regresión simple: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. Donde $\text{Var}(X_i) = i$, ε_i cumple los supuestos del Teorema de Gauss-Markov y $\text{Cov}[\varepsilon_i, X_i] = \text{Cov}[\varepsilon_i, \beta_{j,i}] = 0$ ($j = 0, 1$). En este caso podemos afirmar que el estimador MCO para β_0 será sesgado pero consistente.

La afirmación es Falsa. Porque los estimadores MCO siguen siendo MELI en presencia de una variable explicatoria estocástica que no está relacionada con el término de error o con los betas.

- c) Suponga el modelo $\ln y_i = \beta_1 + \beta_2 \ln x_i + \varepsilon_i$, en donde ε_i es el término de error y $\text{var}(\varepsilon_i) = \sigma^2 D_i$, donde D_i es una variable dummy que toma el valor de uno si el individuo es profesional y cero en caso contrario. Un estudiante afirma: “Se puede transformar el modelo dividiendo por D_i , de modo que los estimadores MCO de los β 's sean insesgados y eficientes.

Falso. En general uno esperaría resolver el problema empleando Mínimos Cuadrados Ponderados (MCP). De tal forma que: $\frac{\ln y_i}{\sqrt{D_i}} = \frac{\beta_1}{\sqrt{D_i}} + \beta_2 \frac{\ln x_i}{\sqrt{D_i}} + \frac{\varepsilon_i}{\sqrt{D_i}}$. Como D_i es una variable dummy se tiene que $\sqrt{D_i} = D_i$. Hasta aquí parecería razonable la afirmación, pero

noten que se puede tener que D_i sea cero. Y será imposible dividir por cero. Por eso, para este caso esa transformación es imposible.

2 Selección Múltiple (15 puntos en total, 5 puntos cada subparte)

Determine cuál de las siguientes respuestas es la correcta. Escoja la mejor opción y explique en dos o tres líneas su respuesta. (No se dará ningún crédito a respuestas sin justificación.)

- i. Un econométra desea estimar el siguiente modelo $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$; donde, $X_2 = 0.95X_3$. para descartar un problema econométrico el investigador decide realizar la siguiente prueba $K(X) = \frac{\sqrt{\lambda_{max}}}{\sqrt{\lambda_{min}}}$. El resultados de la prueba sería:
- $k(x) = 1$
 - $k(x) = 0$
 - $k(x) = 25$
 - Ninguna de las anteriores

Respuesta: d, ninguna de las anteriores. Existe un problema de multicolinealidad perfecta ya que $X_2 = 0.95X_3$; y por lo tanto, el resultado de la prueba debió ser $k(x) > 30$, confirmando la existencia de multicolinealidad perfecta.

- ii. En el siguiente modelo $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$; ($t=1, \dots, N$), las perturbaciones presentan heteroscedasticidad si:
- $cov(X_t * \varepsilon_t) = 0$; para todo $t=1, \dots, N$.
 - $var(\varepsilon_t) = t * \sigma^2$; para todo $t=1, \dots, N$.
 - $var(\varepsilon_t) = c * \sigma^2$; para todo $t=1, \dots, N$.
 - $\varepsilon_t = 10 + \mu_t$; con $var(\varepsilon_t) = 5$, para todo $t=1, \dots, N$.

Respuesta: b, ya que σ^2 esta multiplicado por t ($t=1, \dots, N$) la varianza del error es distinta para cada observación.

- iii. Un econométra es contratado para estimar la cantidad demandada de carros de una de las marcas más populares "HIUNDAI" que tienen quince cedes en todo el país, y por lo tanto, el econométra tiene información de corte transversal. La cantidad demandada se estimara en función de ciertas variables relevantes X, pero este incluye algunas variables irrelevantes Z. Y con la información anterior estima el siguiente modelo $\hat{Y}_i = 100 + 0,4X_{1i} + 1,7X_{2i} + 0,2Z_{1i} + 1,2Z_{2i}$, para probar la significancia individual del parámetro que acompaña a Z_{1i} (una de las variables irrelevantes) tenemos un valor P igual a 0,05. Cuál de las siguientes afirmaciones es correcta:
- El parámetro que acompaña a Z_{1i} es significativo al 95%.

- El parámetro que acompaña a Z_{1i} no es significativo al 95%.
- El t calculado no sirve por la presencia de variables irrelevantes.
- Ninguna de las anteriores

Respuesta: d, no se puede hacer inferencia sobre los parámetros porque la muestra es de quince individuos. Por lo tanto, debemos probar que los errores se distribuyen normalmente para hacer inferencia.

3 (30 puntos)

Usted es contratado por el Ministerio de educación para realizar un estudio sobre la tasa de retorno (rentabilidad privada) de la educación en su país. El Ministro le encarga que su propuesta se base en la teoría del capital humano, específicamente que estime una modificación de la ecuación de salarios propuesta por Mincer (1974). Es decir, se propone el siguiente modelo teórico:

$$\ln(W_i) = \gamma_0 + \gamma_1 educ_i + \gamma_2 exp_i + \gamma_3 gene_i + \gamma_4 gene_i^2 + \mu_i$$

Para realizar dicha estimación el Ministerio le entrega la Encuesta Nacional de Hogares con las siguientes variables: salario mensual en pesos (W_i), los años de educación ($educ_i$), la edad ($edad_i$), la experiencia laboral, medida en años (exp_i) y el género, mujer =1, ($gene_i$).

Además, el jefe del equipo técnico del Ministerio le informa que por estudios previos realizados por el Staff, **posiblemente la $var(\mu_i) = X_i \sigma^2$** .

- El econométra estimó el modelo reportado en la tabla 1. Escriba el modelo estimado por él. (2 puntos)

La especificación del modelo sería:

$$\ln(W_i) = \gamma_0 + \gamma_1 educ_i + \gamma_2 exp_i + \gamma_3 gene_i + \mu_i$$

- ¿Cree usted que el econométra cometió algún error o no, al estimar ese modelo? Sea lo más claro posible y argumente su respuesta. (7 puntos)

Noten que en general no necesariamente existe un problema de multicolinealidad perfecta al incluir una variable y su respectivo cuadrado. Pero en este caso, al incluirse en el modelo una variable dummy al cuadrado ($gene_i^2$), se tiene que $gene_i^2 = gene_i$ (1 al cuadrado es 1 y 0 al cuadrado es 0) (5 puntos, Ojo si no explican porque existe multicolinealidad perfecta no se darán estos 5 puntos.). Por tanto el modelo tiene multicolinealidad perfecta y por tanto no se ha cometido ningún error al excluir esa variable del modelo. (2 puntos).

- El investigador, con miedo de tener un problema econométrico debido a la información suministrada por el staff del Ministerio construye otro modelo y lo estima. Este segundo modelo se reporta en la Tabla 2. Escriba el segundo modelo estimado por el econométra. (2 puntos)

La especificación del modelo sería:

$$\frac{\ln(W_i)}{X_i} = \gamma_0 \frac{1}{X_i} + \gamma_1 \frac{\text{educ}_i}{X_i} + \gamma_2 \frac{\text{exp}_i}{X_i} + \gamma_4 \frac{\text{gene}_i}{X_i} + \frac{\mu_i}{X_i}$$

- d. Dada toda la información disponible en este examen, ¿Cuál de los dos modelos debería emplear el investigador? Sea lo más claro posible y en caso de poder hacer pruebas efectúelas. (10 puntos)

Pruebas en el primer modelo: presencia de heteroscedasticidad. (3 puntos)

Pruebas en el segundo modelo: no hay presencia de heteroscedasticidad se corrige por el método de mínimos cuadrados ponderados. (3 puntos)

Decisión: Se elige el segundo modelo, no presenta problemas de heteroscedasticidad. (4 puntos).

Noten que si $\text{var}(\mu_i) = X_i \sigma^2$, el econometrista debía dividir por la raíz de X_i y no por X_i . No obstante, el problema fue solucionado, lo cual implica que la varianza del error debía ser: $\text{var}(\mu_i) = X_i^2 \sigma^2$. Seguramene, el econometrista al estimar el modelo de la Tabla 2, se había dado cuenta de este problema antes. (esta última discusión no era necesaria)

- e. Independientemente de sus respuestas anteriores y de la presencia o no de problemas, interprete los coeficientes del modelo reportado en la Tabla 1. Teniendo en cuenta su significancia (9 puntos).

$$E[\ln(W_i)] = \begin{cases} (\gamma_0 + \gamma_3) + \gamma_1 \text{educ}_i + \gamma_2 \text{exp}_i & \text{Si el individuo es mujer} \\ \gamma_0 + \gamma_1 \text{educ}_i + \gamma_2 \text{exp}_i & \text{o.w} \end{cases}$$

$\hat{\gamma}_0 = 0,1260146$. Significativo al 90% de confianza. Representa el logaritmo del salario de los hombres que no depende ni de la educación, ni de la experiencia. (2 puntos)

$\hat{\gamma}_1 = 0,1524131$. Significativo al 99% de confianza. Indica que por un aumento en un año de educación del individuo, el salario aumenta en 15,24131%. (2 puntos)

$\hat{\gamma}_2 = 0,026146$. Significativo al 99% de confianza. Indica que por un aumento en un año de experiencia del individuo, el salario aumenta en 2,6146%. (2 puntos)

$\hat{\gamma}_3 = 0,0024671$. Significativo al 90% de confianza. Representa la diferencia promedio, en el logaritmo del salario, entre los hombres y las mujeres. (3 puntos)

4 (40 puntos)

La división de mercadeo de una empresa productora de cepillos de dientes desea estimar la elasticidad precio de la demanda de su producto estrella en Colombia. Para tal fin, emplean datos para los últimos 15 meses y el siguiente modelo:

$$Q_t = \beta_1 + \beta_2 \ln(P_t) + \beta_3 \frac{1}{\text{Avi}} + \beta_4 \text{IPI}_t^2 + \varepsilon_t \quad (1)$$

Donde el Q_t corresponde al número de paquetes de 100 cepillos de dientes que fueron comprados en Colombia el mes t al precio P_t (medido en miles de pesos). Además, Avi

corresponde al número de avisos publicitarios en la televisión (el número de avisos ha sido igual para todos los meses). Finalmente, IPI_t es el índice de producción industrial en Colombia para el mes t que se emplea como proxy del nivel de actividad económica (el mes uno de la muestra corresponde a la base de este índice, es decir ($\text{IPI}_1 = 100$)). Finalmente, ε_t es el término de error que cumple las siguientes características: $E[\varepsilon] = 0$, $\text{Var}[\varepsilon] = \text{Avi}^2 \sigma^2 I_{15}$ (I_{15} corresponde a la matriz identidad de tamaño 15).

A partir de la información anterior conteste las siguientes preguntas:

- a) Si el jefe de estudios económicos emplea el método de mínimos cuadrados ordinarios (MCO), ¿existe algún problema que pudiese no permitir obtener conclusiones “adecuadas” sobre la elasticidad precio de la demanda de su producto estrella? Sea lo más claro posible. (10 puntos)

El modelo **¡Error! No se encuentra el origen de la referencia.** no puede ser estimado directamente, pues Avi es un valor constante para todos los hogares, en este caso podemos reparametrizar el modelo de la siguiente manera:

$$Q_t = \beta_1 + \beta_3 \frac{1}{\text{Avi}} + \beta_2 \ln(P_t) + \beta_4 \text{IPI}_t^2 + \varepsilon_t$$

$$Q_t = \gamma + \beta_2 \ln(P_t) + \beta_4 \text{IPI}_t^2 + \varepsilon_t$$

Donde $\gamma = \beta_1 + \beta_3 \frac{1}{\text{Avi}}$.

Noten que en este caso todos los supuestos del teorema de Gauss-Markov se cumplen. Es decir:

- Relación lineal entre la variable dependiente y los regresores.
- Los regresores deben ser no estocásticos y linealmente independientes entre sí.
- Los errores deben:
 - Tener media cero
 - Varianza constante
 - Y no estar autocorrelacionados

Esto implica que los estimadores MCO del modelo reparametrizado serán MELI. (4 puntos).

Por otro lado, dado que la muestra es pequeña ($n=15$) puede existir un problema al hacer inferencias. Si el error no sigue una distribución normal, entonces no se puede hacer inferencia y por tanto no se podrán sacar conclusiones sobre el efecto del precio sobre la demanda de resmas de papel. (4 puntos).

- b) El econometrista a cargo del estudio recolecta la información obtener estimadores MELI para el modelo (1). El obtiene las siguientes matrices que corresponden al equivalente de la matriz $X^T X$ y $X^T y$. NOTA: Para la construcción de estas matrices, el modelo se organizó lo más parecido al modelo original:

$$X^T X = \begin{bmatrix} 15 & 0 & 5 \\ 0 & 10 & 0 \\ 5 & 0 & 5 \end{bmatrix} \quad X^T y = \begin{bmatrix} 10 \\ -15 \\ 20 \end{bmatrix}$$

Explique claramente a que corresponde cada uno de los elementos de la matriz $X^T X$. (Por ejemplo, explique a partir de que sumatoria sale el 5 que corresponde al último elemento de la matriz $X^T X$, y así sucesivamente con cada elemento de la matriz) **Expresé su respuesta en términos de las variables en el modelo original y no en términos de variables reparametrizadas** (por ejemplo P_i). (6 puntos)

En este caso

$$n = 15, \quad \sum_{i=1}^{15} \ln(P_i) = \sum_{i=1}^{15} \ln(P_i) IPI_i^2 = 0, \quad \sum_{i=1}^{15} IPI_i^2 = 5, \quad \sum_{i=1}^{15} (\ln(P_i))^2 = 10, \quad y \quad \sum_{i=1}^{15} IPI_i^4 = 5$$

1 punto por cada una de estas sumatorias

- c) Encuentre los estimadores MELI para los coeficientes que se puedan recuperar del **modelo ¡Error! No se encuentra el origen de la referencia. MUESTRE** claramente el valor estimado y su correspondiente parámetro poblacional. (12 Puntos)

En este caso tenemos que:

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 1/10 & 0 & -1/10 \\ 0 & 1/10 & 0 \\ -1/10 & 0 & 3/10 \end{bmatrix} \begin{bmatrix} 10 \\ -15 \\ 20 \end{bmatrix} = \begin{bmatrix} -1 \\ -3/2 \\ 5 \end{bmatrix}$$

Es decir,

$$\hat{\beta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\beta}_2 \\ \hat{\beta}_4 \end{bmatrix} = \begin{bmatrix} -1 \\ -3/2 \\ 5 \end{bmatrix}$$

Noten que β_1 y β_2 no se pueden recuperar. Si no muestra a que corresponde cada coeficiente estimado se restarán 4 puntos

- d) Interprete el significado de cada uno de los coeficientes estimados. (6 Puntos)

$\hat{\gamma} = -1$ No tiene interpretación económica. (2 puntos)

$\hat{\beta}_2 = -3/2$. Un aumento del 1% en el precio provocará una caída en la demanda de 0.015 paquetes de cepillos de 100 unidades. (2 puntos)

$\hat{\beta}_4 = 5$. Un aumento de una unidad en el IPI provocará un aumento de $10IPI_i$ paquetes de cepillos de 100 unidades. (2 puntos)

- e) El jefe de estudios económicos de la compañía después de realizar las estimaciones encuentra que:

- Según el modelo teórico más empleado para la estimación de la demanda, es necesario incluir en el modelo el número de cepillos vendidos en los Estados Unidos.
- El IPI no fue medido con la precisión requerida. De hecho, se encontró que existe un error que es siempre del 5% en su medición.

De acuerdo a estos hallazgos, que puede esperar el jefe de estudios sobre los resultados obtenidos hasta ahora y en especial sobre las conclusiones que se saquen en torno a la elasticidad deseada. Sea lo más claro posible, explique todas sus afirmaciones (6 Puntos)

1. Se ha omitido una variable del modelo, que muy probablemente no está relacionada con las variables en el modelo, y por tanto se puede esperar que no exista un sesgado por omisión de variables (2 puntos)
2. Por otro lado, al medir una variable explicatoria con un error determinístico, no existirá ningún problema en la estimación de los coeficientes. Es decir seguirán siendo MELI. (2 puntos)
3. Finalmente, es importante mencionar que en todo caso es imposible hacer inferencia sobre los coeficientes antes de estos hallazgos o después, por tanto no se pueden sacar conclusiones sobre los valores poblacionales. (2 puntos)

Resultados de EasyReg:

Tabla 1

Dependent variable:			
Y = ln[W]			
Characteristics:			
Ln[W]			
First observation = 1			
Last observation = 2000			
Number of usable observations: 2000			
Minimum value: 5.587000E+000			
Maximum value: 1.9781000E+001			
Sample mean: 1.3118156E+001			
X variables:			
X(1) = EDUC			
X(2) = EXP			
X(3) = GENE			
X(4) = 1			
Model:			
Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + b(4)X(4) + U,			
where U is the error term, satisfying			
E[U X(1),X(2),X(3), X(4)] = 0.			
OLS estimation results			
Parameters	Estimate	t-value (S.E.)	H.C. t-value (H.C. S.E.) [p-value] [H.C. p-value]
b(1)	0,1524131	33.984 (0.01000) [0.00000]	41.543 (0.01003) [0.00000]
b(2)	0,026146	28.339 (0.01962) [0.00430]	-2.030 (0.02759) [0.04235]

b(3)	0.0024671	0.983 (0.21715) [0.032560]	0.929 (0.22985) [0.035304]
b(4)	0,1260146	28.339 (0.01962) [0.0230]	-2.030 (0.02759) [0.02235]
Notes:			
1: S.E. = Standard error			
2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.			
3: The two-sided p-values are based on the normal approximation.			
Effective sample size (n):		2000	
Variance of the residuals:		0.12151531	
Standard error of the residuals (SER):		0.34859046	
Residual sum of squares (RSS):		3.40242872	
(Also called SSR = Sum of Squared Residuals)			
Total sum of squares (TSS):		788.65192071	
R-square:		0.9957	
Adjusted R-square:		0.8726	
Overall F test: F(2,22) = 32245,98			
p-value = 0.00000			
Significance levels:		10% 5%	
Critical values:		2.5 3.34	
Conclusions:		reject reject	
Test for first-order autocorrelation:			
Durbin-Watson test = 1,894			
Jarque-Bera/Salmon-Kiefer test = 0.094518			

Null hypothesis: The errors are normally distributed
 Null distribution: Chi-square(2)
 p-value = 0.95384
 Significance levels: 10% 5%
 Critical values: 4.61 5.99
 Conclusions: accept accept

Breusch-Pagan test = 13,93412
 Null hypothesis: The errors are homoskedastic
 Null distribution: Chi-square(2)
 p-value = 0.003285
 Significance levels: 10% 5%
 Critical values: 4.61 5.99
 Conclusions: reject reject

Tabla 2. Resultados de EasyReg

Dependent variable:
 $Y = \ln[W]/X$
 Characteristics:
 $\ln[W]/X$
 First observation = 1
 Last observation = 2000
 Number of usable observations: 2000
 Minimum value: 5.5870000E+000
 Maximum value: 1.9781000E+001
 Sample mean: 1.3118156E+001

X variables:
 $X(1) = EDUC/X$
 $X(2) = EXP/X$
 $X(3) = GENE/X$
 $X(4) = 1/X$

Model:

$Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + b(4)X(4) + U$,
 where U is the error term, satisfying
 $E[U|X(1), X(2), X(3), X(4)] = 0$.

OLS estimation results

Parameters	Estimate	t-value	H.C. t-value
		(S.E.)	(H.C. S.E.)
		[p-value]	[H.C. p-value]
b(1)	0.12968185	1.502	1.360
		(0.19765)	(0.21817)
		[0.00000]	[0.00000]
b(2)	0,011243	28.339	-2.030
		(0.01962)	(0.02759)
		[0.00330]	[0.03235]
b(3)	0.0314671	0.983	0.929
		(0.21715)	(0.22985)
		[0.32560]	[0.35304]
b(4)	0.154671	0.983	0.929
		(0.21715)	(0.22985)
		[0.032560]	[0.035304]

Notes:
 1: S.E. = Standard error
 2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
 3: The two-sided p-values are based on the normal approximation.

Effective sample size (n): 2000
 Variance of the residuals: 0.10845059
 Standard error of the residuals (SER): 0.32931837
 Residual sum of squares (RSS): 3.14506713
 (Also called SSR = Sum of Squared Residuals)
 Total sum of squares (TSS): 3.38921543

R-square: 0.0720
Adjusted R-square: 0.0720

Overall F test: $F(2,22) = 23225,98$

p-value = 0.00000

Significance levels: 10% 5%

Critical values: 2.5 3.34

Conclusions: reject reject

Test for first-order autocorrelation:

Durbin-Watson test = 1,96

Jarque-Bera/Salmon-Kiefer test = 0.743108

Null hypothesis: The errors are normally distributed

Null distribution: Chi-square(2)

p-value = 0.68966

Significance levels: 10% 5%

Critical values: 4.61 5.99

Conclusions: accept accept

Breusch-Pagan test = 2,93412

Null hypothesis: The errors are homoskedastic

Null distribution: Chi-square(2)

p-value = 0.28590

Significance levels: 10% 5%

Critical values: 4.61 5.99

Conclusions: accept accept