

Econometría 06216
Examen Final
Respuestas Sugeridas
Cali, Martes 18 de mayo de 2010

Profesores: Julio César Alonso C. --- Carlos Giovanni González E. --- Ana Isabel Gallego L.

Estudiante: _____
Código: _____

Instrucciones:

1. Lea cuidadosamente todas las preguntas e instrucciones.
2. Este examen consta de **11** páginas; además, deben tener 2 páginas de fórmulas.
3. El examen consta de 3 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 4 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. ¡Asgne su tiempo de forma eficiente!

¡Suerte!

I. Selección Múltiple (50 puntos en total, 1 punto por cada subparte)

Seleccione la opción **más indicada** en la hoja de respuestas que encontrará al final de este examen. Sólo se considerarán respuestas que sean consignadas en la hoja de respuestas. (No es necesario justificar su respuesta)

1. What is the number of degrees of freedom for a simple linear regression with 20 observations?
 - a. 21
 - b. 20
 - c. 19
 - d. 18
 - e. 17
2. Which of the following statements is TRUE?
 - a. the Durbin-Watson statistic can be used if the assumption of normality is valid.
 - b. if the residuals in a regression analysis of time-ordered data are positively correlated, the value of the Durbin-Watson statistic should be near 4.
 - c. if the Durbin-Watson statistic takes a value near 0, then the errors are unlikely to be autocorrelated.
 - d. if the Durbin-Watson statistic takes a value near 0 then the regression model probably violates the regression assumption of independence of errors.
 - e. a and d.
3. Suppose the following model $y_i = \beta_2 D_i x_{it} + \beta_3 W_i x_{it} + \varepsilon_i$ where D_i and W_i are dummy variables, and $D_i = \begin{cases} 1 & \text{female} \\ 0 & \text{o.w} \end{cases}$ and $W_i = \begin{cases} 1 & \text{male} \\ 0 & \text{o.w} \end{cases}$, therefore the β 's estimates will most likely be:
 - a. Inconsistent.
 - b. Perfectly multicollinear.
 - c. Unbiased.
 - d. Biased.
 - e. Efficient.

4. The use of the following dummies variables $D1_t = \begin{cases} 1 & \text{Quarter I} \\ 0 & \text{o.w} \end{cases}$ and $D2_t = \begin{cases} 1 & \text{Quarter II} \\ 0 & \text{o.w} \end{cases}$ and $D3_t = \begin{cases} 1 & \text{Quarter III} \\ 0 & \text{o.w} \end{cases}$ and $D4_t = \begin{cases} 1 & \text{Quarter IV} \\ 0 & \text{o.w} \end{cases}$ related to the unemployment index time series in the same model specification that originally included an intercept and two explanatory variables will cause:
 - a. Inconsistent estimates.
 - b. Biased and efficient estimates.
 - c. A much better approximation.
 - d. None of the above.
5. The Chow test can be used when we want to impose equal values (for the before and after samples):
 - a. On all the parameters of the model.
 - b. Only on the slopes.
 - c. Only on the intercepts.
 - d. All of the above.
6. Assume you want to predict the sales of a product (y) using the relationship between sales and amount of advertising expenditure (x). Consider a simple regression model $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, where ε_i follows the usual assumptions. If we want to predict the sales of a new product with a given amount x_p of advertising expenditure, then the following holds:
 - a. The closer x_p is to \bar{x} the less accurate the predictions will be.
 - b. The length of the estimated prediction interval could be decreased by considering a value of x_p closer to \bar{x} .
 - c. The length of the estimated prediction interval is independent of x_p .
 - d. The length of the estimated prediction interval could be decreased by using a smaller number of data points.
 - e. None of the above.

7. If in a simple linear regression model it is assumed that the intercept parameter is equal to 0, then:
- The regression line will pass through the origin.
 - The regression line will pass through the point (1,1).
 - The regression line will pass through the point (0,-1).
 - The slope of the line will also be equal to 0.
 - a and b are correct.
8. In a multiple regression model with the usual assumptions, if we eliminate a predictor variable:
- The explained variability will increase.
 - The unexplained variability will increase or stay the same.
 - The total variability will increase.
 - The coefficient of determination (R^2) increases.
 - a and d are correct.
9. A professor wants to test if the number of hours of study and the number of exercises completed during the course by a student has an effect on the grade of the student in the course. He considers 43 students and fits a multiple regression model with two predictor variables (numbers of hours of study and exercises completed) and a constant term. The following is obtained:

Source	Sum Squares	SS
Regression		70
Residual Errors		
Total		80

- The coefficient of determination (R^2) is 0.95.
 - After conducting the global F-test we conclude that the model is statistically useful.
 - None of the two variables have an effect on the grade of the course.
 - Both variables are significant.
 - None of the above.
10. Suppose that a sample of 15 data follows the multiple regression model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ with the usual assumptions. Individual tests of significance have been done for both variables as well as the global regression test. The results are:

$$H_0 : \beta_1 = 0, t = 0.5$$

$$H_0 : \beta_2 = 0, t = 5.6$$

$$H_0 : \beta_1 = \beta_2 = 0, F = 225,$$

therefore:

- The results indicate that the regression model is useful for predicting, but x_{1i} is individually not significant.
 - The results indicate that both variables x_{1i} and x_{2i} are not significant.
 - Both the individual tests and the global test indicate that the model is not useful for predicting.
 - The results indicate that we must apply a transformation of the response variable.
 - The model has a relatively small coefficient of determination (R^2)
11. A microeconomist wants to determine how corporate sales are influenced by capital and wage spending by large companies. A statistical analyst discovers that capital spending has a significant inverse relationship with wage spending. What should this microeconomist be particularly concerned about in developing a multiple regression model?
- Autocorrelation.
 - Collinearity.
 - Normality of error terms.
 - Heteroscedasticity.
 - None of the above
12. If the Durbin-Watson statistic has a value close to 4, which assumption is violated?
- Normality
 - independence of the sample
 - homoscedasticity
 - all of the above
 - none of the above
13. The reason why estimators have a sampling distribution is that
- economics is not a precise science.
 - individuals respond differently to incentives.
 - in real life you typically get to sample many times.
 - the values of the explanatory variable and the error term differ across samples.
 - all of the above

14. To decide whether $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ or $\ln(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ fits the data better, you cannot consult the regression R^2 because:
- $\ln(y_i)$ may be negative for $0 < Y < 1$.
 - the SSTs are not measured in the same units between the two models.
 - the slope no longer indicates the effect of a unit change of X on Y in the log-linear model.
 - the regression R^2 can be greater than one in the second model.
 - None of the above
15. The slope estimator $\hat{\beta}_1$ has a smaller standard error, other things equal, if
- there is more variation in the explanatory variable X.
 - there is a large variance of the error term
 - the sample size is smaller
 - the intercept, β_0 , is small.
 - the intercept, β_0 , is zero.
16. The general answer to the question of choosing the scale of the variables is
- dependent on you impulse.
 - to make the regression results easy to read and to interpret.
 - to ensure that the regression coefficients always lie between -1 and 1.
 - irrelevant because regardless of the scale of the variable, the regression coefficient's value is unaffected.
 - to ensure that the regression coefficients are as closed as possible to an elasticity.
17. Si $X_{1,i} = 0.9X_{2,i} + \varepsilon_i$, el modelo $y_i = \alpha + \beta X_{1,i} + \gamma X_{2,i} + \mu_i$ tiene problemas de:
- Autocorrelación
 - Multicolinealidad perfecta
 - Multicolinealidad no perfecta
 - Heteroscedasticidad
 - No es posible saber sin hacer pruebas

18. Si el p-valor de la prueba de Durbin-Watson para probar $H_0 : \rho = 0$ es 0.17 y el p-valor del DW para probar $H_0 : \rho \geq 0$ es 0.08, con un 10% de significancia, y según para lo que la prueba ha sido diseñada, podemos concluir de ésta que:
- Hay autocorrelación positiva de primer grado.
 - Hay autocorrelación negativa de primer grado.
 - No hay autocorrelación de primer grado.
 - Nos encontramos con un proceso autoregresivo de un proceso superior a 0.
 - Nos encontramos con un proceso autoregresivo de un proceso inferior a 0.
19. Una matriz $X^T X$ de 5x5, con $X_{1,1} = 130$, está asociada a una regresión con:
- Cinco pendientes
 - Cinco parámetros
 - 130 observaciones
 - A y B
 - A y C
20. Suponga que está usando EasyReg y estima un modelo con Heteroscedasticidad. Y decide emplear la matriz consistente de Heteroscedasticidad de White. Entonces, el estadístico calculado en la prueba de significancia conjunta sigue una distribución:
- z
 - F
 - χ^2
 - t
 - Ninguno, no se pueden hacer pruebas de significancia conjunta.
21. El modelo $y_i = \alpha + \beta_1 X_{1,i} + \beta_2 y_{i-1} + \beta_3 y_{i-2} + \beta_4 y_{i-3} + \mu_i$ con seguridad, tiene problemas de autocorrelación de:
- Primer orden
 - Segundo orden

- c. Tercer orden
 - d. Primer, segundo y tercer orden
 - e. No es posible saber
22. Un econometrista está convencido de la necesidad de uno de los siguientes supuestos para que el estimador MCO de β sea insesgado:
- a. Que la varianza del término de error sea constante a lo largo de la muestra
 - b. Que los parámetros del modelo sean constantes a lo largo de la muestra
 - c. Que la distribución de los errores sea normal
 - d. Que la covarianza entre cualquier par de errores sea igual a cero
 - e. Ninguno de los anteriores.
23. Una deficiencia importante, entre otras, de los Modelos de Probabilidad Lineal, cuando se le compara con los Modelos Logit y Probit es:
- a. El permitir valores pronosticados, por fuera del Intervalo [0, 1].
 - b. La Función de Distribución de probabilidad que utilizan.
 - c. La Probabilidad varía linealmente, en la misma proporción constante, ante un cambio unitario de la variable independiente.
 - d. Los coeficientes estimados no tienen una interpretación intuitiva.
 - e. a. y c. son ciertas
24. Si una dummy (D_i que toma valores 1 y 0) incluida en el modelo causa Heteroscedasticidad del tipo $\sigma_i^2 = \sigma^2 D_i^2$
- a. Se puede detectar la heteroscedasticidad con la prueba de Golfeld y Quandt, y se puede corregir por MCP.
 - b. Se puede detectar la heteroscedasticidad con la prueba de Golfeld y Quandt, pero no se puede corregir por MCP.
 - c. No se puede detectar la heteroscedasticidad con la prueba de Golfeld y Quandt, pero se puede corregir por MCP.
 - d. No se puede detectar la heteroscedasticidad con la prueba de Golfeld y Quandt, y no se puede corregir por MCP.
 - e. Ninguna de las anteriores.
25. En un sistema de ecuaciones simultáneas, si no existe un problema de simultaneidad, entonces:
- a. Los MCO están sesgados mientras que los EMC2E serán consistentes, pero no eficientes.
 - b. Los MCO son consistentes pero no eficientes, mientras que los EMC2E son eficientes y consistentes.
 - c. Los MCO son eficientes y consistentes mientras que los EMC2E serán inconsistentes, pero eficientes.
 - d. Los MCO son eficientes y consistentes mientras que los EMC2E serán consistentes, pero no eficientes.
 - e. Ninguna de las anteriores
26. El objetivo en la estimación de un modelo de variable dependiente cualitativa por el método de Máxima Verosimilitud será
- a. Encontrar un estimador para β que maximice el valor promedio de que y ocurra dadas las variables explicativas en X
 - b. Encontrar un estimador para β que minimice la probabilidad de que y ocurra dadas las variables explicativas en X
 - c. Encontrar un estimador para β que minimice la suma de los residuos al cuadrado dadas las variables explicativas en X
 - d. Encontrar un estimador para β que maximice la probabilidad de que y ocurra dadas las variables explicativas en X
 - e. Ninguna de las anteriores
27. The sample average of the OLS residuals for a regression model with k parameters and k-1 explanatory variables is

- a. some positive number since OLS uses squares.
 - b. zero
 - c. unobservable since the population regression function is unknown.
 - d. dependent on whether the explanatory variable is mostly positive or negative
 - e. is also MELI.
28. Autocorrelation of the OLS regression errors
- a. results in incorrect OLS standard errors.
 - b. makes the OLS estimator (for slopes and constant term) inconsistent, but not unbiased.
 - c. is not a problem in cross-sections since the data can always be "reordered."
 - d. all of the above.
 - e. a) and c) are correct.
29. The distinction between endogenous and exogenous variables is
- a. that exogenous variables are determined inside the model and endogenous variables are determined outside the model.
 - b. dependent on the sample size: for $n > 100$, endogenous variables become exogenous.
 - c. depends on the distribution of the variables: when they are normally distributed, they are exogenous, otherwise they are endogenous.
 - d. whether or not the variables are correlated with the error term.
 - e. Exogenous variables produce the so called "simultaneity problem".
30. In the linear probability model, the interpretation of the slope coefficient is
- a. the change in the probability associated with a unit change in X, holding other regressors constant.
 - b. not all that meaningful since the dependent variable is either 0 or 1.
 - c. the change in probability that $Y=1$ associated with a one percentage change in X, holding others regressors constant.
 - d. the response in the dependent variable to a percentage change in the regressor.
 - e. 100 times % the change in the probability that $Y=1$ associated with a unit change in X, holding others regressors constant.
31. Si no rechazamos la nula del test de Hausman para la forma estructural de un sistema conformado por la función inversa de demanda de carros y la oferta inversa de carros, lo mejor que podemos hacer es:
- a. Estimar las ecuaciones de la forma estructural por MC2E.
 - b. Estimar las ecuaciones de la forma estructural por MCO.
 - c. Estimar las ecuaciones de la forma estructural por MCP.
 - d. Estimar las ecuaciones reducidas por MC2E.
 - e. Nada, porque hay problemas de simultaneidad.
32. The OLS residuals ($\hat{\epsilon}_i$)
- a. can be calculated using the errors from the regression model.
 - b. are unknown since we do not know the population regression function.
 - c. should not be used in practice since they indicate that your regression does not run through all your observations.
 - d. are MELI thanks to the Gauss-Markov theorem.
 - e. None of the above
33. In the presence of heteroskedasticity, and assuming that the usual least squares assumptions hold, the OLS estimator is
- a. Efficient
 - b. BLUE
 - c. unbiased and consistent
 - d. unbiased but not consistent
34. The overall regression F-statistic tests the null hypothesis that
- a. all slope coefficients are zero.
 - b. all coefficients are zero.
 - c. the intercept in the regression and at least one, but not all, of the slope coefficients is zero.
 - d. the slope coefficient of the variable of interest is zero, but that the other slope coefficients are not.
 - e. a and b are correct.

35. When there are omitted variables in the regression, which are determinants of the dependent variable, then
- you cannot measure the effect of the omitted variable, but the estimator of your included variables are unaffected.
 - this has no effect on the estimator of your included variables because the other variable is not included.
 - this will always bias the OLS estimator of the included variables.
 - the OLS estimator is biased if the omitted variable is correlated with at least one of the included variables.
 - We will have the famous "simultaneity bias".
36. All of the following are true, with the exception of one condition:
- a high R^2 or \bar{R}^2 does not mean that the regressors are a true cause of the dependent variable.
 - a high R^2 or \bar{R}^2 does not mean that there is no omitted variable bias.
 - a high R^2 or \bar{R}^2 always means that an added variable is statistically significant.
 - a high R^2 or \bar{R}^2 does not necessarily mean that you have the most appropriate set of regressors.
 - none of the above
37. La autocorrelación en los datos es un problema porque:
- El método de MCO asume que los datos no son correlacionados y calcula los estimadores puntuales de los parámetros de la regresión de manera acorde.
 - Sesga los estimadores puntuales del error estándar del parámetro.
 - Los errores están autocorrelacionados y no son linealmente independientes.
 - (a) y (b) son correctas.
 - Ninguna de las anteriores.
38. El test de White **NO** está diseñado para detectar:
- Heteroscedasticidad, si ésta es causada por variables rezagadas que ya están incluidas en el modelo.
 - Heteroscedasticidad, si ésta es causada por variables no incluidas en el modelo.
 - Heteroscedasticidad, si ésta es causada por dummies incluidas en el modelo.
 - Heteroscedasticidad, si ésta es causada por la primera variable independiente del modelo.
 - B y C son correctas.
39. Si en un modelo de regresión lineal múltiple no se cumple el supuesto de que las X 's son no estocásticas, entonces tenemos un problema de:
- Multicolinealidad
 - Heteroscedasticidad
 - Autocorrelación
 - Error de medición
 - Ninguno de los anteriores
40. En un sistema de ecuaciones cuando la ecuación uno presenta $k=2$, $g=4$ y la ecuación dos presenta $k=3$, $g=6$. Podemos afirmar entonces que:
- Ecuación uno y dos subidentificadas
 - Ecuación uno y dos sobreidentificadas
 - Ecuación uno y dos perfectamente identificadas
 - Ecuación uno subidentificada y ecuación dos sobreidentificada
 - No se puede afirmar nada sobre el problema de identificación
41. Si el "*p-value*", asociado a la prueba de Normalidad de Jarque-Bera ($\alpha = 0.05$), toma el valor de 0.5, podemos concluir a partir de dicho estadístico que:
- Se acepta la Hipótesis alterna de Normalidad
 - Se rechaza la hipótesis nula de No Normalidad
 - Existe evidencia para rechazar la hipótesis nula de Normalidad
 - No existe evidencia para rechazar la hipótesis nula de Normalidad
 - Ninguna de las anteriores.
42. El P-valor asociado con una prueba de hipótesis es:
- La probabilidad de que la hipótesis nula sea falsa.
 - La probabilidad de que la hipótesis nula sea verdadera.

- El menor nivel de significancia al que la hipótesis nula será rechazada.
 - El mayor nivel de significancia asociado a que la hipótesis nula sea rechazada.
 - Ninguna de las anteriores
43. Es un supuesto del test de DW, para probar correlación serial:
- El modelo de regresión incluye una constante.
 - La autocorrelación sigue un proceso auto-regresivo de orden uno.
 - La ecuación no incluye una variable dependiente rezagada como regresor.
 - Todas las anteriores.
 - a y c son correctas.
44. La diferencia entre un coeficiente estandarizado y un coeficiente sin estandarizar es:
- Que el primero siempre es positivo.
 - Que el primero siempre es más grande
 - Que el primero no tiene interpretación
 - Todas las anteriores
 - Ninguna las anteriores
45. Si una variable empleada en un modelo de regresión presenta un error de medición, entonces:
- Los estimadores MCO de los coeficientes son siempre insesgados.
 - Los estimadores MCO de los coeficientes no siempre serán insesgados.
 - Los estimadores MCO de los coeficientes siempre serán consistentes.
 - Los estimadores MCO de los coeficientes siempre serán inconsistentes.
 - Ninguna las anteriores
46. El test de Goldfeld-Quandt es:
- Una prueba de heteroscedasticidad que se basa en la idea de que las varianzas de los residuos son iguales entre todas las observaciones, entonces la varianza de una sección de la muestra debe ser igual a la varianza de otra sección de la muestra.
 - Una prueba de heteroscedasticidad que separa la muestra en dos sub-muestras y calcula un estadístico F a partir de las dos diferentes regresiones de las sub-muestras.
 - Una prueba que concluye que la sub-muestra con el SSE más alto de los dos es la que presenta mayor varianza.
 - Todas las anteriores.
 - a y b son correctas.
47. El denominado problema de "simultaneidad" se presenta cuando:
- Existe un co-movimiento entre la variable dependiente y alguna de las independientes.
 - Existe correlación entre el término de perturbación estocástico y una de las variables independientes.
 - En un modelo de Regresión lineal múltiple, una variable independiente afecta, de manera simultánea, a la variable dependiente y a otra independiente.
 - Todas las anteriores
 - Ninguna de las anteriores
48. Bajo el cumplimiento estricto de los supuestos del teorema de Gauss-Markov, **NO** se puede decir de la matriz $(X^T X)^{-1}$ que:
- Las entradas de la diagonal toman valores estrictamente no negativos.
 - Las entradas de la diagonal son iguales.
 - Los elementos fuera de la diagonal pueden ser negativos.
 - Es una matriz de rango completo.
 - Ninguna de las anteriores.
49. Si se rechaza la nula del test de Jarque-Bera en un modelo estimado por MCO con una muestra grande, ¿cuál de las siguientes **NO** se debe hacer?:
- Pruebas de significancia individual
 - Pruebas de significancia conjunta
 - Prueba de Goldfeld y Quandt
 - Todas se pueden hacer
 - Ninguna se puede hacer

50. Si encontramos que una ecuación está sub-identificada en un sistema de ecuaciones simultáneas, entonces:

- a. Podemos estimar los parámetros de forma reducida
- b. Podemos estimar los coeficientes de esa ecuación de tal forma que eliminemos el problema de simultaneidad

- c. No existe forma de encontrar los parámetros estructurales de esa ecuación
- d. Podemos estimar los parámetros de esa ecuación por el método de MC2E
- e. Todas las anteriores

II. (20 puntos)

En un mes habrá elecciones presidenciales en un país del este de Beirut y las encuestas continuamente muestran resultados contradictorios. Actualmente parece que las elecciones se definirán en segunda vuelta entre los dos candidatos más populares, el candidato del partido de la A y el candidato del partido B. Usted es contratado por una prestigiosa firma de consultoría para realizar un estudio sobre la votación en segunda vuelta de las elecciones presidenciales de este año. La variable dependiente toma los valores de 1 si en segunda vuelta el individuo vota por el candidato del partido de la A y 0 en caso contrario. Cada una de las variables independientes es una representación a escala (Escala de 1 a 10) de la propia opinión del individuo (i) respecto a lo que sentía sobre una cuestión particular. Las variables independientes son: X₂ seguridad, X₃ derechos y legalidad, X₄ empleo, y X₅ toma el valor de 1 si es hombre. La información proviene de una encuesta realizada a nivel nacional a 1500 personas mayores de edad. Las siguientes ecuaciones representan la elección del individuo:

$$prob(Y_i = 0) = [1 - F(\beta'X_i)] \tag{1}$$

$$prob(Y_i = 1) = F(\beta'X_i) \tag{2}$$

Donde F(.) es la función de distribución logística.

- a. Explique qué representan las dos ecuaciones anteriores. Escriba de una forma alternativa (a las ecuaciones 1 y 2) el modelo estimado y presentado en la tabla 1. **(5 puntos, 3 por la explicación y 2 por escribir correctamente el modelo).**
- b. Encuentre los valores que fueron reemplazados por XXX en la Tabla 1. No es necesario efectuar el cálculo, pero sí mostrar con qué cantidades se puede encontrar dicho número **(6 puntos, 2 por cada XXX)**
- c. Un compañero de universidad le manifiesta que él también está realizando un trabajo similar y que la estimación debería realizarse por el método MCO. Escriba cinco argumentos para contradecir a su compañero y ratificar la estimación que usted hizo. **(5 puntos)**
- d. Teniendo en cuenta la significancia, interprete los coeficientes estimados **(4 puntos)**

III. (30 puntos)

La división de investigaciones económicas del Departamento Nacional de Planeación (DNP) de una pequeña República Andina ha determinado que la mejor forma de determinar la posibilidad de no cumplimiento con el servicio de la deuda pública de la república en el periodo t está dado por el siguiente sistema de ecuaciones:

$$pd_t = \alpha_1 + \alpha_2 \ln(stock_t) + \alpha_3 TES_t + \alpha_4 i_t + \mu_t \tag{1}$$

$$TES_t = \beta_1 + \beta_2 Deficit_t + \beta_3 i_t + \varepsilon_t \tag{2}$$

$$REC_t = \gamma_1 + \gamma_3 pd_t + \gamma_4 i_t + \zeta_t \tag{3}$$

En equilibrio tendremos que

$$REC_t = TES_t = Q_t \tag{4}$$

donde pd_t es una variable dummy que toma el valor de uno si la república entra en default (no puede cumplir con el servicio de la deuda) en el periodo t y cero en caso contrario. $stock_t$ y $Deficit_t$ representan el saldo de la deuda y el déficit fiscal (medido como porcentaje del PIB) del gobierno central de la república en el periodo t. Además, i_t representa la tasa de interés de los bonos del tesoro que financian el gobierno central para el periodo t (medida en puntos porcentuales). REC_t y TES_t representan las cantidades demandas de bonos de tesorería del gobierno de la república (medido en billones de dólares) y las cantidades ofrecidas de bonos para financiar el déficit del gobierno en el periodo t (medido en billones de dólares), respectivamente. Finalmente, ε_t , μ_t y ζ_t representan términos de error con media cero, independientemente distribuidos y con varianza constante. Es decir, su respectiva matriz de varianzas y covarianzas es $\sigma^2 I$.

- a. Interprete los siguientes coeficientes: α_2 , γ_4 y β_2 . **(6 puntos, 2 puntos cada uno)**
- b. Explique claramente con qué métodos de puede estimar cada una de las expresiones (1) a (4). **(4 puntos, 1 punto cada uno)**
- c. El director de la investigación está convencido que puede encontrar estimadores MELI para los coeficientes de la ecuación (2) empleando MCO. Explique claramente si el investigador puede tener la razón. **(5 puntos)**

Independientemente de su respuesta en la pregunta anterior, si lo cree necesario, suponga que no hay ningún problema econométrico asociado a estimar (2) empleando MCO. La división de investigaciones económicas encontró las siguientes matrices (conservando el orden de las variables en (2)) para el modelo (2).

$$X^T X = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 30 & -5 \\ 0 & -5 & 1 \end{bmatrix} \quad X^T y = \begin{bmatrix} 400 \\ 5 \\ 1/2 \end{bmatrix}$$

- d. El economista en jefe quiere conocer las siguientes cantidades. Estime esas cantidades si es posible encontrarlas a partir de la información disponible. **(5 puntos)**
 - $E[Deficit_t]$
 - $Var[i_t]$
 - $E[\ln(stock_t)]$
- e. Encuentre los estimadores MCO para (2). **MUESTRE** claramente sus cálculos y el valor estimado para cada parámetro poblacional. **(5 Puntos)**
- f. **Interprete los coeficientes estimados (5 Puntos)**

Tabla 1. Resultados de EasyReg.

```

Logit Model:
Dependent variable:
Y = Y
A Probit or Logit model is suitable.
X(1) = X2
X(2) = X3
X(3) = X4
X(4) = X5
X(5) = 1
Frequency of y = 1:56.84%
Frequency of y = 0:43.16%
Model: P(Y=1|x) = F(b(1)x(1)+..+b(5)x(5))
Chosen option: (Logit model)
Newton iteration successfully completed after 5 iterations
Last absolute parameter change = 0.0000
Last percentage change of the likelihood = 0.0000
Maximum likelihood estimation results
Variable          ML estimate of b(.) (t-value)
x(1)=X2           b(1)= 0.352135 (2.19)
.....           (S.E.= XXX)
                  [p-value = 0.00000]
x(2)=X3           b(2)= 0.0063 (7.08)
                  [p-value = 0.00000]
x(3)=X4           b(3)= -0.032 (-7.08)
                  [p-value = 0.00000]
x(4)=X5           b(4)= -0.257 (-7.08)
                  [p-value = 0.00000]
x(5)=1            b(5)= 0.375 (7.08)
                  [p-value = 0.00000]
[The two-sided p-values are based on the normal approximation]
XXX: -4.02924561069E+002
Sample size (n): XXX
If the model is correctly specified then the maximum likelihood
parameter estimators b(1),...b(8), minus their true values, times the
square root of the sample size n, are (asymptotically) jointly normally
distributed with zero mean vector and variance matrix:
3.07967990E+01 1.63386352E+00 1.06820808E+00 -1.01057199E+00 1.04188402E-01 -2.05734023E-01
2.06271994E-03 -4.18669664E+01
1.63386352E+00 4.15511299E+00 2.74846838E-01 2.00021109E-01 1.72016431E-02 1.69643787E-01 -
9.66182876E-04 -2.19981364E+01
1.06820808E+00 2.74846838E-01 1.57445332E-01 1.50137000E-02 -1.64734159E-02 -2.51871553E-02 -
7.13180078E-04 -7.04283234E+00
-1.01057199E+00 2.00021109E-01 1.50137000E-02 1.36449462E+00 -2.07359582E-01 -1.33412298E-02
7.83170778E-04 -1.57680627E+01
1.04188402E-01 1.72016431E-02 -1.64734159E-02 -2.07359582E-01 3.42025860E-01 -7.38244671E-03
9.96897177E-04 5.04897801E-01
-2.05734023E-01 1.69643787E-01 -2.51871553E-02 -1.33412298E-02 -7.38244671E-03 7.69303200E-01 -
2.22935699E-02 -3.08276475E+00
2.06271994E-03 -9.66182876E-04 -7.13180078E-04 7.83170778E-04 9.96897177E-04 -2.22935699E-02
7.71997869E-04 1.15550636E-01
-4.18669664E+01 -2.19981364E+01 -7.04283234E+00 -1.57680627E+01 5.04897801E-01 -
3.08276475E+00
1.15550636E-01 5.47947784E+02
    
```

Econometría 06216
Examen Final
Respuestas Sugeridas
Cali, Martes 18 de mayo de 2010

Profesores: Julio César Alonso --- Carlos Giovanni González --- Ana Isabel Gallego

Estudiante: _____
Código: _____

Instrucciones:

1. Lea cuidadosamente todas las preguntas e instrucciones.
2. Este examen consta de **11** páginas; además, deben tener 2 páginas de fórmulas.
3. El examen consta de 3 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 4 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!

¡Suerte!

I. Selección Múltiple (50 puntos en total, 1 punto por cada subparte)

*Seleccione la opción **más indicada** en la hoja de respuestas que encontrará al final de este examen. Sólo se considerarán respuestas que sean consignadas en la hoja de respuestas. (No es necesario justificar su respuesta)*

1. What is the number of degrees of freedom for a simple linear regression with 20 observations?
 - a. 21
 - b. 20
 - c. 19
 - d. **18**
 - e. 17
2. Which of the following statements is TRUE?
 - a. the Durbin-Watson statistic can be used if the assumption of normality is valid.
 - b. if the residuals in a regression analysis of time-ordered data are positively correlated, the value of the Durbin-Watson statistic should be near 4.
 - c. if the Durbin-Watson statistic takes a value near 0, then the errors are unlikely to be autocorrelated.
 - d. if the Durbin-Watson statistic takes a value near 0 then the regression model probably violates the regression assumption of independence of errors.
 - e. **a and d.**
3. Suppose the following model $y_i = \beta_2 D_i x_{i1} + \beta_3 W_i x_{i1} + \varepsilon_i$ where D_i and W_i are dummy variables, and $D_i = \begin{cases} 1 & \text{female} \\ 0 & \text{o.w} \end{cases}$ and $W_i = \begin{cases} 1 & \text{male} \\ 0 & \text{o.w} \end{cases}$, therefore the β 's estimates will most likely be:
 - a. Inconsistent.
 - b. Perfectly multicollinear.
 - c. Unbiased.
 - d. **Biased.**
 - e. Efficient.

4. The use of the following dummies variables

$$D1_i = \begin{cases} 1 & \text{Quarter I} \\ 0 & \text{o.w} \end{cases} \quad \text{and}$$

$$D2_i = \begin{cases} 1 & \text{Quarter II} \\ 0 & \text{o.w} \end{cases} \quad \text{and}$$

$$D3_i = \begin{cases} 1 & \text{Quarter III} \\ 0 & \text{o.w} \end{cases} \quad \text{and}$$

$$D4_i = \begin{cases} 1 & \text{Quarter IV} \\ 0 & \text{o.w} \end{cases} \quad \text{related to the}$$

unemployment index time series in the same model specification that originally included an intercept and two explanatory variables will cause:

- a. Inconsistent estimates.
 - b. Biased and efficient estimates.
 - c. A much better approximation.
 - d. **None of the above.**
5. The Chow test can be used when we want to impose equal values (for the before and after samples):
- a. **On all the parameters of the model.**
 - b. Only on the slopes.
 - c. Only on the intercepts.
 - d. All of the above.
6. Assume you want to predict the sales of a product (y) using the relationship between sales and amount of advertising expenditure (x). Consider a simple regression model $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, where ε_i follows the usual assumptions. If we want to predict the sales of a new product with a given amount x_p of advertising expenditure, then the following holds:
- a. The closer x_p is to \bar{x} the less accurate the predictions will be.
 - b. **The length of the estimated prediction interval could be decreased by considering a value of x_p closer to \bar{x} .**
 - c. The length of the estimated prediction interval is independent of x_p .
 - d. The length of the estimated prediction interval could be decreased by using a smaller number of data points.
 - e. None of the above.

7. If in a simple linear regression model it is assumed that the intercept parameter is equal to 0, then:
- The regression line will pass through the origin.
 - The regression line will pass through the point (1,1).
 - The regression line will pass through the point (0,-1).
 - The slope of the line will also be equal to 0.
 - a and b are correct.
8. In a multiple regression model with the usual assumptions, if we eliminate a predictor variable:
- The explained variability will increase.
 - The unexplained variability will increase or stay the same.
 - The total variability will increase.
 - The coefficient of determination (R^2) increases.
 - a and d are correct.
9. A professor wants to test if the number of hours of study and the number of exercises completed during the course by a student has an effect on the grade of the student in the course. He considers 43 students and fits a multiple regression model with two predictor variables (numbers of hours of study and exercises completed) and a constant term. The following is obtained:
- | Source | Sum Squares | SS |
|-----------------|-------------|----|
| Regression | | 70 |
| Residual Errors | | |
| Total | | 80 |
- The coefficient of determination (R^2) is 0.95.
 - After conducting the global F-test we conclude that the model is statistically useful.
 - None of the two variables have an effect on the grade of the course.
 - Both variables are significant.
 - None of the above.
10. Suppose that a sample of 15 data follows the multiple regression model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ with the usual assumptions. Individual tests of significance have been done for both variables as well as the global regression test. The results are:

$$H_0 : \beta_1 = 0, t = 0.5$$

$$H_0 : \beta_2 = 0, t = 5.6$$

$$H_0 : \beta_1 = \beta_2 = 0, F = 225,$$

therefore:

- The results indicate that the regression model is useful for predicting, but x_{1i} is individually not significant.
 - The results indicate that both variables x_{1i} and x_{2i} are not significant.
 - Both the individual tests and the global test indicate that the model is not useful for predicting.
 - The results indicate that we must apply a transformation of the response variable.
 - The model has a relatively small coefficient of determination (R^2)
11. A microeconomist wants to determine how corporate sales are influenced by capital and wage spending by large companies. A statistical analyst discovers that capital spending has a significant inverse relationship with wage spending. What should this microeconomist be particularly concerned about in developing a multiple regression model?
- Autocorrelation.
 - Collinearity.
 - Normality of error terms.
 - Heteroscedasticity.
 - None of the above
12. If the Durbin-Watson statistic has a value close to 4, which assumption is violated?
- Normality
 - independence of the sample
 - homoscedasticity
 - all of the above
 - none of the above
13. The reason why estimators have a sampling distribution is that
- economics is not a precise science.
 - individuals respond differently to incentives.
 - in real life you typically get to sample many times.
 - the values of the explanatory variable and the error term differ across samples.
 - all of the above

14. To decide whether $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ or $\ln(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$ fits the data better, you cannot consult the regression R^2 because:
- $\ln(y_i)$ may be negative for $0 < Y < 1$.
 - the SSTs are not measured in the same units between the two models.
 - the slope no longer indicates the effect of a unit change of X on Y in the log-linear model.
 - the regression R^2 can be greater than one in the second model.
 - None of the above
15. The slope estimator β_1 has a smaller standard error, other things equal, if
- there is more variation in the explanatory variable X .
 - there is a large variance of the error term
 - the sample size is smaller
 - the intercept, β_0 , is small.
 - the intercept, β_0 , is zero.
16. The general answer to the question of choosing the scale of the variables is
- dependent on you impulse.
 - to make the regression results easy to read and to interpret.
 - to ensure that the regression coefficients always lie between -1 and 1.
 - irrelevant because regardless of the scale of the variable, the regression coefficient's value is unaffected.
 - to ensure that the regression coefficients are as closed as possible to an elasticity.
17. Si $X_{1,i} = 0.9X_{2,i} + \varepsilon_i$, el modelo $y_i = \alpha + \beta X_{1,i} + \gamma X_{2,i} + \mu_i$ tiene problemas de:
- Autocorrelación
 - Multicolinealidad perfecta
 - Multicolinealidad no perfecta
 - Heteroscedasticidad
 - No es posible saber sin hacer pruebas

18. Si el p-valor de la prueba de Durbin-Watson para probar $H_0 : \rho = 0$ es 0.17 y el p-valor del DW para probar $H_0 : \rho \geq 0$ es 0.08, con un 10% de significancia, y según para lo que la prueba ha sido diseñada, podemos concluir de ésta que:
- Hay autocorrelación positiva de primer grado.
 - Hay autocorrelación negativa de primer grado.
 - No hay autocorrelación de primer grado.
 - Nos encontramos con un proceso autoregresivo de un proceso superior a 0.
 - Nos encontramos con un proceso autoregresivo de un proceso inferior a 0.
19. Una matriz $X^T X$ de 5×5 , con $X_{11} = 130$, está asociada a una regresión con:
- Cinco pendientes
 - Cinco parámetros
 - 130 observaciones
 - A y B
 - A y C
20. Suponga que está usando EasyReg y estima un modelo con Heteroscedasticidad. Y decide emplear la matriz consistente de Heteroscedasticidad de White. Entonces, el estadístico calculado en la prueba de significancia conjunta sigue una distribución:
- z
 - F
 - χ^2
 - t
 - Ninguno, no se pueden hacer pruebas de significancia conjunta.
21. El modelo $y_i = \alpha + \beta_1 X_{1,i} + \beta_2 y_{i-1} + \beta_3 y_{i-2} + \beta_4 y_{i-3} + \mu_i$, con seguridad, tiene problemas de autocorrelación de:
- Primer orden
 - Segundo orden

- c. Tercer orden
 - d. Primer, segundo y tercer orden
 - e. No es posible saber
22. Un econometrista está convencido de la necesidad de uno de los siguientes supuestos para que el estimador MCO de β sea insesgado:
- a. Que la varianza del término de error sea constante a lo largo de la muestra
 - b. Que los parámetros del modelo sean constantes a lo largo de la muestra
 - c. Que la distribución de los errores sea normal
 - d. Que la covarianza entre cualquier par de errores sea igual a cero
 - e. Ninguno de los anteriores.
23. Una deficiencia importante, entre otras, de los Modelos de Probabilidad Lineal, cuando se le compara con los Modelos Logit y Probit es:
- a. El permitir valores pronosticados, por fuera del Intervalo $[0, 1]$.
 - b. La Función de Distribución de probabilidad que utilizan.
 - c. La Probabilidad varía linealmente, en la misma proporción constante, ante un cambio unitario de la variable independiente.
 - d. Los coeficientes estimados no tienen una interpretación intuitiva.
 - e. a. y c. son ciertas
24. Si una dummy (D_i que toma valores 1 y 0) incluida en el modelo causa Heteroscedasticidad del tipo $\sigma_i^2 = \sigma^2 D_i^2$.
- a. Se puede detectar la heteroscedasticidad con la prueba de Golfeld y Quandt, y se puede corregir por MCP.
 - b. Se puede detectar la heteroscedasticidad con la prueba de Golfeld y Quandt, pero no se puede corregir por MCP.
 - c. No se puede detectar la heteroscedasticidad con la prueba de Golfeld y Quandt, pero se puede corregir por MCP.
 - d. No se puede detectar la heteroscedasticidad con la prueba de Golfeld y Quandt, y no se puede corregir por MCP.
 - e. Ninguna de las anteriores.
25. En un sistema de ecuaciones simultáneas, si no existe un problema de simultaneidad, entonces:
- a. Los MCO están sesgados mientras que los EMC2E serán consistentes, pero no eficientes.
 - b. Los MCO son consistentes pero no eficientes, mientras que los EMC2E son eficientes y consistentes.
 - c. Los MCO son eficientes y consistentes mientras que los EMC2E serán inconsistentes, pero eficientes.
 - d. Los MCO son eficientes y consistentes mientras que los EMC2E serán inconsistentes, pero no eficientes.
 - e. Ninguna de las anteriores.
26. El objetivo en la estimación de un modelo de variable dependiente cualitativa por el método de Máxima Verosimilitud será
- a. Encontrar un estimador para β que maximice el valor promedio de que y ocurra dadas las variables explicativas en X
 - b. Encontrar un estimador para β que minimice la probabilidad de que y ocurra dadas las variables explicativas en X
 - c. Encontrar un estimador para β que minimice la suma de los residuos al cuadrado dadas las variables explicativas en X
 - d. Encontrar un estimador para β que maximice la probabilidad de que y ocurra dadas las variables explicativas en X
 - e. Ninguna de las anteriores
27. The sample average of the OLS residuals for a regression model with k parameters and k-1 explanatory variables is

- a. some positive number since OLS uses squares.
 - b. zero
 - c. unobservable since the population regression function is unknown.
 - d. dependent on whether the explanatory variable is mostly positive or negative
 - e. is also MELI.
28. Autocorrelation of the OLS regression errors
- a. results in incorrect OLS standard errors.
 - b. makes the OLS estimator (for slopes and constant term) inconsistent, but not unbiased.
 - c. is not a problem in cross-sections since the data can always be "reordered."
 - d. all of the above.
 - e. a) and c) are correct.
29. The distinction between endogenous and exogenous variables is
- a. that exogenous variables are determined inside the model and endogenous variables are determined outside the model.
 - b. dependent on the sample size: for $n > 100$, endogenous variables become exogenous.
 - c. depends on the distribution of the variables: when they are normally distributed, they are exogenous, otherwise they are endogenous.
 - d. whether or not the variables are correlated with the error term.
 - e. Exogenous variables produce the so called "simultaneity problem".
30. In the linear probability model, the interpretation of the slope coefficient is
- a. the change in the probability associated with a unit change in X, holding other regressors constant.
 - b. not all that meaningful since the dependent variable is either 0 or 1.
 - c. the change in probability that $Y=1$ associated with a one percentage change in X, holding others regressors constant.
 - d. the response in the dependent variable to a percentage change in the regressor.
 - e. 100 times % the change in the probability that $Y=1$ associated with a unit change in X, holding others regressors constant.
31. Si no rechazamos la nula del test de Hausman para la forma estructural de un sistema conformado por la función inversa de demanda de carros y la oferta inversa de carros, lo mejor que podemos hacer es:
- a. Estimar las ecuaciones de la forma estructural por MC2E.
 - b. Estimar las ecuaciones de la forma estructural por MCO.
 - c. Estimar las ecuaciones de la forma estructural por MCP.
 - d. Estimar las ecuaciones reducidas por MC2E.
 - e. Nada, porque hay problemas de simultaneidad.
32. The OLS residuals ($\hat{\epsilon}_i$)
- a. can be calculated using the errors from the regression model.
 - b. are unknown since we do not know the population regression function.
 - c. should not be used in practice since they indicate that your regression does not run through all your observations.
 - d. are MELI thanks to the Gauss-Markov theorem.
 - e. None of the above
33. In the presence of heteroskedasticity, and assuming that the usual least squares assumptions hold, the OLS estimator is
- a. Efficient
 - b. BLUE
 - c. unbiased and consistent
 - d. unbiased but not consistent
34. The overall regression F-statistic tests the null hypothesis that
- a. all slope coefficients are zero.
 - b. all coefficients are zero.
 - c. the intercept in the regression and at least one, but not all, of the slope coefficients is zero.
 - d. the slope coefficient of the variable of interest is zero, but that the other slope coefficients are not.
 - e. a and b are correct.

35. When there are omitted variables in the regression, which are determinants of the dependent variable, then
- you cannot measure the effect of the omitted variable, but the estimator of your included variables are unaffected.
 - this has no effect on the estimator of your included variables because the other variable is not included.
 - this will always bias the OLS estimator of the included variables.
 - the OLS estimator is biased if the omitted variable is correlated with at least one of the included variables.
 - We will have the famous "simultaneity bias".
36. All of the following are true, with the exception of one condition:
- a high R^2 or \bar{R}^2 does not mean that the regressors are a true cause of the dependent variable.
 - a high R^2 or \bar{R}^2 does not mean that there is no omitted variable bias.
 - a high R^2 or \bar{R}^2 always means that an added variable is statistically significant.
 - a high R^2 or \bar{R}^2 does not necessarily mean that you have the most appropriate set of regressors.
 - none of the above
37. La autocorrelación en los datos es un problema porque:
- El método de MCO asume que los datos no son correlacionados y calcula los estimadores puntuales de los parámetros de la regresión de manera acorde.
 - Sesga los estimadores puntuales del error estándar del parámetro.
 - Los errores están autocorrelacionados y no son linealmente independientes.
 - (a) y (b) son correctas.
 - Ninguna de las anteriores.
38. El test de White **NO** está diseñado para detectar:
- Heteroscedasticidad, si ésta es causada por variables rezagadas que ya están incluidas en el modelo.
 - Heteroscedasticidad, si ésta es causada por variables no incluidas en el modelo.
 - Heteroscedasticidad, si ésta es causada por dummies incluidas en el modelo.
 - Heteroscedasticidad, si ésta es causada por la primera variable independiente del modelo.
 - B y C son correctas.
39. Si en un modelo de regresión lineal múltiple no se cumple el supuesto de que las X 's son no estocásticas, entonces tenemos un problema de:
- Multicolinealidad
 - Heteroscedasticidad
 - Autocorrelación
 - Error de medición
 - Ninguno de los anteriores
40. En un sistema de ecuaciones cuando la ecuación uno presenta $k=2$, $g=4$ y la ecuación dos presenta $k=3$, $g=6$. Podemos afirmar entonces que:
- Ecuación uno y dos subidentificadas
 - Ecuación uno y dos sobreidentificadas
 - Ecuación uno y dos perfectamente identificadas
 - Ecuación uno subidentificada y ecuación dos sobreidentificada
 - No se puede afirmar nada sobre el problema de identificación
41. Si el "*p-value*", asociado a la prueba de Normalidad de Jarque-Bera ($\alpha = 0.05$), toma el valor de 0.5, podemos concluir a partir de dicho estadístico que:
- Se acepta la Hipótesis alterna de Normalidad
 - Se rechaza la hipótesis nula de No Normalidad
 - Existe evidencia para rechazar la hipótesis nula de Normalidad
 - No existe evidencia para rechazar la hipótesis nula de Normalidad
 - Ninguna de las anteriores.
42. El P-valor asociado con una prueba de hipótesis es:
- La probabilidad de que la hipótesis nula sea falsa.
 - La probabilidad de que la hipótesis nula sea verdadera.

- El menor nivel de significancia al que la hipótesis nula será rechazada.
 - El mayor nivel de significancia asociado a que la hipótesis nula sea rechazada.
 - Ninguna de las anteriores
43. Es un supuesto del test de DW, para probar correlación serial:
- El modelo de regresión incluye una constante.
 - La autocorrelación sigue un proceso auto-regresivo de orden uno.
 - La ecuación no incluye una variable dependiente rezagada como regresor.
 - Todas las anteriores.
 - a y c son correctas.
44. La diferencia entre un coeficiente estandarizado y un coeficiente sin estandarizar es:
- Que el primero siempre es positivo.
 - Que el primero siempre es más grande
 - Que el primero no tiene interpretación
 - Todas las anteriores
 - Ninguna las anteriores
45. Si una variable empleada en un modelo de regresión presenta un error de medición, entonces:
- Los estimadores MCO de los coeficientes son siempre insesgados.
 - Los estimadores MCO de los coeficientes no siempre serán insesgados.
 - Los estimadores MCO de los coeficientes siempre serán consistentes.
 - Los estimadores MCO de los coeficientes siempre serán inconsistentes.
 - Ninguna las anteriores
46. El test de Goldfeld-Quandt es:
- Una prueba de heteroscedasticidad que se basa en la idea de que las varianzas de los residuos son iguales entre todas las observaciones, entonces la varianza de una sección de la muestra debe ser igual a la varianza de otra sección de la muestra.
 - Una prueba de heteroscedasticidad que separa la muestra en dos sub-muestras y calcula un estadístico F a partir de las dos diferentes regresiones de las sub-muestras.
 - Una prueba que concluye que la sub-muestra con el SSE más alto de los dos es la que presenta mayor varianza.
 - Todas las anteriores.
 - a y b son correctas.
47. El denominado problema de "simultaneidad" se presenta cuando:
- Existe un co-movimiento en la variable dependiente y alguna de las independientes.
 - Existe correlación entre el término de perturbación estocástico y una de las variables independientes.
 - En un modelo de Regresión lineal múltiple, una variable independiente afecta, de manera simultánea, a la variable dependiente y a otra independiente.
 - Todas las anteriores
 - Ninguna de las anteriores
48. Bajo el cumplimiento estricto de los supuestos del teorema de Gauss-Markov, **NO** se puede decir de la matriz $(X^T X)^{-1}$ que:
- Las entradas de la diagonal toman valores estrictamente no negativos.
 - Las entradas de la diagonal son iguales.
 - Los elementos fuera de la diagonal pueden ser negativos.
 - Es una matriz de rango completo.
 - Ninguna de las anteriores.
49. Si se rechaza la nula del test de Jarque-Bera en un modelo estimado por MCO con una muestra grande, ¿cuál de las siguientes **NO** se debe hacer?:
- Pruebas de significancia individual
 - Pruebas de significancia conjunta
 - Prueba de Goldfeld y Quandt
 - Todas se pueden hacer
 - Ninguna se puede hacer

50. Si encontramos que una ecuación está sub-identificada en un sistema de ecuaciones simultáneas, entonces:

- a. Podemos estimar los parámetros de forma reducida
- b. Podemos estimar los coeficientes de esa ecuación de tal forma que eliminemos el problema de simultaneidad

c. No existe forma de encontrar los parámetros estructurales de esa ecuación

- d. Podemos estimar los parámetros de esa ecuación por el método de MC2E
- e. Todas las anteriores

II. (20 puntos)

En un mes habrá elecciones presidenciales en un país del este de Beirut y las encuestas continuamente muestran resultados contradictorios. Actualmente parece que las elecciones se definirán en segunda vuelta entre los dos candidatos más populares, el candidato del partido de la A y el candidato del partido B. Usted es contratado por una prestigiosa firma de consultoría para realizar un estudio sobre la votación en segunda vuelta de las elecciones presidenciales de este año. La variable dependiente toma los valores de 1 si en segunda vuelta el individuo vota por el candidato del partido de la A y 0 en caso contrario. Cada una de las variables independientes es una representación a escala (Escala de 1 a 10) de la propia opinión del individuo (i) respecto a lo que sentía sobre una cuestión particular. Las variables independientes son: X₂ seguridad, X₃ derechos y legalidad, X₄ empleo, y X₅ toma el valor de 1 si es hombre. La información proviene de una encuesta realizada a nivel nacional a 1500 personas mayores de edad. Las siguientes ecuaciones representan la elección del individuo:

$$prob(Y_i = 0) = [1 - F(\beta'X_i)] \tag{1}$$

$$prob(Y_i = 1) = F(\beta'X_i) \tag{2}$$

Donde F(.) es la función de distribución logística.

- a. Explique qué representan las dos ecuaciones anteriores. Escriba de una forma alternativa (a las ecuaciones 1 y 2) el modelo estimado y presentado en la tabla 1. **(5 puntos, 3 por la explicación y 2 por escribir correctamente el modelo).**

Noten que las ecuaciones están escritas como probabilidades condicionales. Por lo tanto, la ecuación 1 representa la probabilidad condicional de que un individuo (i) vote en segunda vuelta por el candidato del partido B. Mientras, la ecuación 2 representa lo contrario, la probabilidad condicional de que un individuo (i) vote en segunda vuelta por el candidato del partido de la A.

Noten que el modelo estimado es un logit, que se puede escribir de la siguiente forma:

$$Y_i = \frac{1e^{\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i}}}{1 + e^{\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i}}} + \mu_i = \Lambda(\beta'X_i) + \mu_i$$

- b. Encuentre los valores que fueron reemplazados por XXX en la tabla 1. No es necesario efectuar el cálculo, pero sí mostrar con qué cantidades se puede encontrar dicho número **(6 puntos, 2 por cada XXX)**

XXX= 2.19/0.352135=0.16079224. Error estándar.
 XXX= Log likelihood
 XXX= 1500. Tamaño de la muestra.

- c. Un compañero de universidad le manifiesta que él también está realizando un trabajo similar y que la estimación debería realizarse por el método MCO. Escriba cinco argumentos para contradecir a su compañero y ratificar la estimación que usted hizo. **(5 puntos)**

El método de estimación que propone el compañero sería por MCO al estimar un MLP. Las desventajas de estimar un modelo con variable dependiente discreta por el Método de MCO, son:

- 1- Predicciones no acotadas (0-1)
- 2- Heteroscedasticidad
- 3- No normalidad de las perturbaciones
- 4- R2 está subestimado y no tiene sentido en muchos casos
- 5- la varianza de las perturbaciones puede ser negativa

Por lo tanto, al ser un modelo de variable dependiente cualitativa, la estimación debe realizarse por el método de máxima verosimilitud, usando un probit o un logit, como en este caso.

- d. Teniendo en cuenta la significancia, interprete los coeficientes estimados (4 puntos)

Los parámetros no se pueden interpretar de forma convencional como lo hacemos con el Modelo de Regresión Lineal Múltiple. Aquí los parámetros no tienen una interpretación intuitiva y por eso sólo podemos mencionar si el parámetro es significativo y el signo (o el efecto que tiene la variable explicativa sobre la variable dependiente) (0.8 cada uno)

$\hat{\beta}_1$ = significativo al 99%. A medida que aumenta la percepción de seguridad, la probabilidad de que un individuo vote en segunda vuelta por el candidato del partido de la A aumenta.

$\hat{\beta}_2$ = significativo al 99%. A medida que aumenta la percepción de derechos y legalidad, la probabilidad de que un individuo vote en segunda vuelta por el candidato del partido de la A aumenta.

$\hat{\beta}_3$ = significativo al 99%. A medida que aumenta la percepción de empleo, la probabilidad de que un individuo vote en segunda vuelta por el candidato del partido de la A disminuye.

$\hat{\beta}_4$ = significativo al 99%. Los hombres presentan una probabilidad mas baja frente a las mujeres de votar en segunda vuelta por el candidato del partido A.

$\hat{\beta}_5$ = significativo al 99%. No tiene interpretación.

. Cada una de las variables independientes es una representación a escala (Escala de 1 a 10) de la propia opinión del individuo (i) respecto a lo que sentía sobre una cuestión particular. Las variables independientes son: X₂ seguridad, X₃ derechos y legalidad, X₄ empleo, y X₅ toma el valor de 1 si es hombre.

III. (30 puntos)

La división de investigaciones económicas del Departamento Nacional de Planeación (DNP) de una pequeña República Andina ha determinado que la mejor forma de determinar la posibilidad de no cumplimiento con el servicio de la deuda pública de la república en el periodo t está dado por el siguiente sistema de ecuaciones:

$$pd_t = \alpha_1 + \alpha_2 \ln(stock_t) + \alpha_3 TES_t + \alpha_4 i_t + \mu_t \tag{1}$$

$$TES_t = \beta_1 + \beta_2 Deficit_t + \beta_3 i_t + \varepsilon_t \tag{2}$$

$$REC_t = \gamma_1 + \gamma_3 pd_t + \gamma_4 i_t + \zeta_t \quad (3)$$

En equilibrio tendremos que

$$REC_t = TES_t = Q_t \quad (4)$$

donde pd_t es una variable dummy que toma el valor de uno si la república entra en default (no puede cumplir con el servicio de la deuda) en el periodo t y cero en caso contrario. $stock_t$ y $Deficit_t$ representan el saldo de la deuda y el déficit fiscal (medido como porcentaje del PIB) del gobierno central de la república en el periodo t. Además, i_t representa la tasa de interés de los bonos del tesoro que financian el gobierno central para el periodo t (medida en puntos porcentuales). REC_t y TES_t representan las cantidades demandas de bonos de tesorería del gobierno de la república (medido en billones de dólares) y las cantidades ofrecidas de bonos para financiar el déficit del gobierno en el periodo t (medido en billones de dólares), respectivamente. Finalmente, ε_t , μ_t y ζ_t representan términos de error con media cero, independientemente distribuidos y con varianzas constante. Es decir, su respectiva matriz de varianzas y covarianzas es $\sigma^2 I$.

a. Interprete los siguientes coeficientes: α_2 , γ_4 y β_2 . (6 puntos, 2 puntos cada uno)

En este caso es importante reconocer que la expresión (1) representa la relación que existe entre la probabilidad de entrar en mora (default) y en logaritmo del stock de capital y la oferta de TES. Por otro lado, (2) representa la oferta de TES y (3) representa la demanda de TES. Así, la expresión (1) es la función de default, (2) es la oferta de TES y (3) es la demanda de TES. Por tanto,

- α_2 se puede interpretar de la siguiente manera. Ante un aumento del uno por ciento en el stock de deuda como porcentaje del PIB para el periodo t, la función de default cambiará en $(\alpha_2/100)*100$ puntos porcentuales. Es decir, esa función se desplazará α_2 puntos porcentuales.
- β_2 se puede interpretar de la siguiente manera. Ante un aumento de un punto porcentual en el déficit como porcentaje del PIB, la oferta de TES aumentará en β_2 billones de dólares. Es decir, la oferta se desplazará en β_2 billones de dólares, cuando el déficit aumenta en un punto porcentual.
- γ_4 se puede interpretar de la siguiente manera. Ante un aumento de un punto porcentual en la tasa de interés, las cantidades demandas de TES cambiarán en γ_4 billones de dólares. En otras palabras, tendremos un desplazamiento sobre la curva de demanda de TES de γ_4 billones de dólares cuando la tasa de interés aumenta en un punto porcentual.

Nota: No se dará crédito si al interpretar no se distingue entre las funciones y los correspondientes desplazamientos sobre la curva o de la curva. Ya sea de manera implícita o explícita. Por ejemplo, de manera implícita, se puede mostrar que la curva se desplaza cuando se menciona que la demanda cambia y que hay un movimiento sobre la curva cuando se dice que las cantidades demandadas cambiaron.

b. Explique claramente con qué métodos de puede estimar cada una de las expresiones (1) a (4). (4 puntos, 1 punto cada uno)

En este caso las variables endógenas son: pd_t , i_t y $REC_t = TES_t = Q_t$. (3 variables endógenas)

Las variables exógenas son: $\ln(stock_t)$, $Deficit_t$

En ese orden de ideas tenemos que:

Ecuación	k_t	g_t	$k_t ? g_t - 1$	¿Sub, perfectamente o sobre identificada?	Método de estimación
(1)	1	3	1<2	Subidentificada	no se puede estimar
(2)	1	2	1=1	Perfectamente identificada (probablemente)	MCI o MC2E
(3)	2	3	2=2	Perfectamente identificada (probablemente)	MCI o MC2E
(4)			Es una identidad		No se requiere estimar parámetros

c. El director de la investigación está convencido que puede encontrar estimadores MELI para los coeficientes de la ecuación (2) empleando MCO. Explique claramente si el investigador puede tener la razón. (5 puntos)

Es importante tener en cuenta que (2) es una ecuación de la forma estructural, por eso se esperaría un sesgo de simultaneidad que harían los MCO sesgados. La única manera que podemos tener que los estimadores MCO fuesen insesgados y por tanto MELI es que i_t no esté relacionada con ε_t . En otras palabras, sólo si $Cov(i_t, \varepsilon_t)=0$.

Por tanto el investigador puede tener razón si se cumple que $Cov(i_t, \varepsilon_t)=0$. En caso que esto no se cumpla, el investigador no tendrá la razón.

Independientemente de su respuesta en la pregunta anterior, si lo cree necesario, suponga que no hay ningún problema econométrico asociado a estimar (2) empleando MCO. La división de investigaciones económicas encontró las siguientes matrices (conservando el orden de las variables en (2)) para el modelo (2).

$$X^T X = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 30 & -5 \\ 0 & -5 & 1 \end{bmatrix} \quad X^T y = \begin{bmatrix} 400 \\ 5 \\ 1/2 \end{bmatrix}$$

d. El economista en jefe quiere conocer las siguientes cantidades. Estime esas cantidades si es posible encontrarlas a partir de la información disponible. (5 puntos)

- $E[Deficit_t]$
- $Var[i_t]$

- $E[\ln(stock_t)]$

En este caso tenemos que:

- $E[Deficit_t]$ puede ser estimado con $\overline{Deficit} = \frac{\sum Deficit_t}{n} = \frac{0}{100} = 0$ (2 puntos)
- $Var[i_t]$ puede ser estimado con $s_i^2 = \frac{\sum (i_t - \bar{i})^2}{n-1} = \frac{\sum (i_t)^2}{100-1} = \frac{1}{100-1} = \frac{1}{99}$ (2 puntos)
- $E[\ln(stock_t)]$ no se puede estimar, pues no tenemos información de la variable $\ln(stock_t)$ de la matriz $X^T X$ reportada. (1 punto)

e. Encuentre los estimadores MCO para (2). **MUESTRE** claramente sus cálculos y el valor estimado para cada parámetro poblacional. (5 Puntos)

En este caso tenemos que:

$$\hat{\beta}_{\text{hat}} = (X^T X)^{-1} \cdot X^T y$$

$$\hat{\beta} = \begin{bmatrix} 1/100 & 0 & 0 \\ 0 & 1/5 & 1 \\ 0 & 1 & 6 \end{bmatrix} \begin{bmatrix} 400 \\ 5 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 4 \\ 3/2 \\ 8 \end{bmatrix}$$

f. Interprete los coeficientes estimados (5 Puntos)

$\hat{\beta}_1 = 4$. La parte de la oferta de bonos del tesoro que no depende de las demás variables es de 4 billones de dólares. (1 punto)

$\hat{\beta}_2 = 3/2$. Ante un aumento de un punto porcentual del déficit como porcentaje del PIB, la oferta de TES aumentará (se desplazará) en 1.5 billones de dólares. (nota: no se dará crédito si no es claro que se trata de un desplazamiento de la oferta). (2 puntos)

$\hat{\beta}_3 = 8$. Un aumento de un punto porcentual en la tasa de interés aumentará las cantidades ofrecidas de TES en 8 billones de dólares. (nota: no se dará crédito si no es claro que se trata de un desplazamiento sobre la curva). (2 puntos)

Tabla 1. Resultados de EasyReg.

```

Logit Model:
Dependent variable:
Y = Y
A Probit or Logit model is suitable.
X(1) = X2
X(2) = X3
X(3) = X4
X(4) = X5
X(5) = 1
Frequency of y = 1:56.84%
Frequency of y = 0:43.16%
Model: P(Y=1|x) = F(b(1)x(1)+..+b(5)x(5))
Chosen option: (Logit model)
Newton iteration successfully completed after 5 iterations
Last absolute parameter change = 0.0000
Last percentage change of the likelihood = 0.0000
Maximum likelihood estimation results
Variable          ML estimate of b(.) (t-value)
x(1)=X2           b(1)= 0.352135 (2.19)
                  (S.E. = XXX)
                  [p-value = 0.00000]
x(2)=X3           b(2)= 0.0063 (7.08)
                  [p-value = 0.00000]
x(3)=X4           b(3)= -0.032 (-7.08)
                  [p-value = 0.00000]
x(4)=X5           b(4)= -0.257 (-7.08)
                  [p-value = 0.00000]
x(5)=1            b(5)= 0.375 (7.08)
                  [p-value = 0.00000]
[The two-sided p-values are based on the normal approximation]
XXX: -4.02924561069E+002
Sample size (n): XXX
If the model is correctly specified then the maximum likelihood
parameter estimators b(1),...b(8), minus their true values, times the
square root of the sample size n, are (asymptotically) jointly normally
distributed with zero mean vector and variance matrix:
3.07967990E+01 1.63386352E+00 1.06820808E+00 -1.01057199E+00 1.04188402E-01 -
2.05734023E-01
2.06271994E-03 -4.18669664E+01
1.63386352E+00 4.15511299E+00 2.74846838E-01 2.00021109E-01 1.72016431E-02
1.69643787E-01 -
9.66182876E-04 -2.19981364E+01
1.06820808E+00 2.74846838E-01 1.57445332E-01 1.50137000E-02 -1.64734159E-02 -
2.51871553E-02 -
7.13180078E-04 -7.04283234E+00
-1.01057199E+00 2.00021109E-01 1.50137000E-02 1.36449462E+00 -2.07359582E-01 -
1.33412298E-02
7.83170778E-04 -1.57680627E+01
1.04188402E-01 1.72016431E-02 -1.64734159E-02 -2.07359582E-01 3.42025860E-01 -
    
```

7.38244671E-03
9.96897177E-04 5.04897801E-01
-2.05734023E-01 1.69643787E-01 -2.51871553E-02 -1.33412298E-02 -7.38244671E-03
7.69303200E-01 -
2.22935699E-02 -3.08276475E+00
2.06271994E-03 -9.66182876E-04 -7.13180078E-04 7.83170778E-04 9.96897177E-04 -
2.22935699E-02
7.71997869E-04 1.15550636E-01
-4.18669664E+01 -2.19981364E+01 -7.04283234E+00 -1.57680627E+01 5.04897801E-01 -
3.08276475E+00
1.15550636E-01 5.47947784E+02