

TRABAJO DE ECONOMETRIA (Multicolinealidad)

10.2 Considérese el conjunto de datos hipotéticos de la tabla 10.10. Supóngase que desea ajustar el modelo

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

TABLA 10.10

Y	X_2	X_3
-10	1	1
-8	2	3
-6	3	5
-4	4	7
-2	5	9
0	6	11
2	7	13
4	8	15
6	9	17
8	10	19
10	11	21

A la información.

a).- ¿Se puede estimar las tres incógnitas? ¿Porqué si o por que no?

No se puede estimar las tres incógnitas debido a que existe multicolinealidad perfecta, puesto que el coeficiente de correlación r_{23} es la unidad es decir la correlación entre X_2 y X_3 es igual a 1.

b).- Si no se puede hacer; ¿Qué funciones lineales de estos parámetros, las funciones estimables, se puede estimar? Muéstrense los cálculos necesarios.

Para poder estimar estos parámetros partimos de que:

$$Y = \alpha_1 + \alpha_2 X_{2i} + u_i$$

Donde:

$$\alpha_1 = \beta_1 - \beta_3$$

$$\alpha_2 = \beta_2 + 2\beta_3$$

$$X_{3i} = 2X_{2i} - 1$$

Por lo tanto reemplazando $X_{3i} = 2X_{2i} - 1$ en $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ nos quedaría:

$$Y = \beta_1 + \beta_2 X_{2i} + \beta_3 (2X_{2i} - 1) + u_i$$

10.3 Refiérase al ejemplo de la mortalidad infantil analizado en el capítulo 8. Dicho ejemplo implicó hacer la regresión de la tasa de mortalidad infantil (MI) sobre el PIB per cápita (PIBPC) y la tasa de alfabetización para las mujeres (TAM).

Ahora, supóngase que se añade la variable tasa de fertilidad total (TFT). Lo anterior da los siguientes resultados de la regresión:

Dependent variable: CM				
Variable	Coefficient	Std. Error	t-Statistic	Prob
C	168,3078	32,89165	5,117003	0,0000
PGNP	-0,005511	0,001878	-2,934275	0,0047
FLR	-1,768029	0,248017	-7,128663	0,0000
TFR	12,86864	4,190533	3,070883	0,0032
R-squared	0.747372	Mean dependent var		141.5000

p value = (0.0000) (0.0065) (0.0000)

$R^2 = 0.7077$ $\bar{R}^2 = 0.6981$

Los coeficientes de β_1 y β_3 cambian significativamente debido a la introducción de una nueva variable sin embargo el coeficiente de β_2 permanece siendo casi el mismo, de igual manera ocurre con las varianzas β_1 y β_3 es decir las varianzas de estos se incrementan y por lo tanto la precisión es menor.

b).- ¿Vale la pena añadir la variable TFT al modelo? ¿Por qué?

Si porque debido a la introducción de esta nueva variable obtenemos un $R^2=0.7473$ mayor al $R^2=0.7077$ del modelo sin incluir la nueva variable, esto nos diría que el segundo modelo parece ajustarse mejor, en resumen diríamos que la Mortalidad Infantil esta explicada por estas tres variables en su conjunto.

c).- Puesto que todos los coeficientes t individuales son estadísticamente significativos, ¿se puede decir que no existe un problema de colinealidad en el presente caso?

No existe un problema de colinealidad, para afirmar esto nos basamos en la primera regla empírica de la detección de colinealidad que nos señala que existe colinealidad cuando hay un R^2 alto y varias razones “t” poco significativas.

10.5 Considérese el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + \beta_5 X_{t-3} + \beta_6 X_{t-4} + u_t$$

Donde Y= consumo, X= ingreso y t= tiempo. El modelo anterior postula que el gasto de consumo en el tiempo t es función no solamente del ingreso en el tiempo t, sino también del ingreso en períodos anteriores. Por tanto, el gasto de consumo en el primer trimestre del 2000 es función del ingreso en ese trimestre y en los cuatro trimestres de 1999. Tales modelos se denominan modelos de rezago distribuido.

a).- ¿Se esperaría la presencia de multicolinealidad en tales modelos y por qué?

Sí, se esperaría la presencia de multicolinealidad porque las variables ingreso y tiempo tienen una relación directa debido a que, a medida que transcurre el tiempo se espera que el ingreso se incremente.

b).- Si se espera la presencia de colinealidad, ¿cómo se resolvería el problema?

Tomando en cuenta que existe dos soluciones para resolver este problema que son:

1.-No hacer nada

2.- Eliminar la variable colineada

Pero en este caso no existe la posibilidad de eliminar la variable colineada ya que esta es relevante en el modelo y esto ocasionaría un sesgo de especificación, por lo tanto optamos por la posibilidad de no hacer nada.

10.6 Considérese en ejemplo ilustrativo de la sección 10.6. ¿Cómo se interpretaría la diferencia en la propensión marginal a consumir obtenida de 10.6.1 y 10.6.4?

10.6.1

$$\hat{Y}_t = 24.7747 + 0.9415 X_{2i} - 0.0424 X_{3i}$$

ee= (6.7525) (0.8229) (0.0807)

t= (3.6690) (1.1442) (-0.5261)

$R^2 = 0.7077$ $\bar{R}^2 = 0.6981$ g de 1=2

10.6.4

$$\hat{Y}_t = 24.4545 + 0.5091 X_{2i}$$

ee= (6.4138) (0.0357)

t = (3.8128) (14.2432) $R^2=0.9621$

La diferencia que se puede observar es que en el primer modelo donde el consumo está en función del ingreso y la riqueza tenemos una propensión marginal al consumo alta de 0.9415 debido a que existe multicolinealidad entre las variables explicativas ingreso y riqueza, mientras que en el segundo modelo donde el consumo está explicado únicamente por el ingreso tiene una propensión marginal al consumo de 0.5091, que es menor al primer modelo

b).- ¿Será $\hat{\beta}_1$ igual a $\hat{\alpha}_1$ o \hat{Y}_1 o a alguna combinación de estos?

No es igual $\hat{\beta}_1$ a $\hat{\alpha}_1$ o \hat{Y}_1 porque $\hat{\beta}_1$ es igual a $\beta_1 + \beta_3 X_{3i}$

c).- ¿Será $\text{var}(\hat{\beta}_2) = \text{var}(\hat{\alpha}_2)$ y $\text{var}(\hat{\beta}_3) = \text{var}(\hat{Y}_3)$?

Si porque no existe la presencia de multicolinealidad debido a que existe homocedasticidad es decir, la varianza permanece constante.

10.12 Establézcase si las siguientes afirmaciones son ciertas, falsas o inciertas y explique la respuesta.

a).- A pesar de la presencia de multicolinealidad perfecta, los estimadores MCO son MELI. (F)

Es falso porque cuando existe multicolinealidad perfecta sus varianzas son infinitas y por lo tanto ya no cumple con la condición de que son estimadores MELI es decir con varianza mínima.

b).- En los casos de alta multicolinealidad, no es posible evaluar la significancia individual de una o más coeficientes de regresión parcial. (V)

Si es verdadero porque las pruebas “t” de uno de uno o más coeficientes son poco significativas debido a que la varianza tiende a inflarse y por lo tanto no hay precisión de los estimadores.

c).- Si una regresión auxiliar muestra que una R_i^2 particular es alta, entonces hay evidencia clara de colinealidad. (V)

Es verdadero porque nos muestra que existe una relación fuerte entre las variables explicativas del modelo.

d).- Las correlaciones altas entre parejas de regresoras no sugieren la existencia de alta multicolinealidad. (F)

Es Falso porque un coeficiente de correlación alto nos indica que sí existe multicolinealidad, es decir existe una relación alta pero no determinística entre todas o algunas variables del modelo.

e).- La multicolinealidad es inofensiva si el objetivo del análisis es solamente la predicción. (V)

Es verdadero porque para predecir se desea conocer el comportamiento de la variable en el futuro y no se necesitan datos realmente comprobados.

f).- Entre mayor sea el FIV, ceteris paribus, más grandes son las varianzas de los estimadores MCO. (V)

Es verdadero porque entre mayor sea el Factor Inflador de Varianza mayor será la varianza de los estimadores de Mínimos Cuadrados Ordinarios (MCO) sin embargo sus estimadores siguen siendo MELI.

g).- La tolerancia (TOL) es una medida de multicolinealidad mejor que el FIV.

(F)

Esto es falso porque la tolerancia (TOL), el FIV y otros métodos podrían utilizarse únicamente como “expedición de pesca”, ya que no puede decirse cuáles de ellos funcionan en una aplicación particular.

h).- No podrá obtenerse un valor R^2 elevado en una regresión múltiple si todos los coeficientes parciales de pendiente no son estadísticamente significativos, a nivel individual, con base en la prueba t usual. (F)

Es falso porque aún cuando las pruebas “t” individuales sean poco significativas se puede obtener un R^2 alto, es decir individualmente las variables independientes no explican a la variable dependiente pero si explican de mejor manera en su conjunto.

i).- En la regresión de Y sobre X_2 y X_3 , supóngase que hay poca variabilidad en los valores de X_3 . Esto aumentaría la var $\hat{\beta}_3$. En el extremo si todas las X_3 fueran idénticas, var $\hat{\beta}_3$ fueran infinitas. (F)

Es falso porque si existe poca variabilidad en los valores de X_3 nos indica que existe una mayor precisión por lo tanto las varianzas serán mínimas.

$$R = \begin{bmatrix} \mathbf{X}_2 & \mathbf{1} & \mathbf{r}_{23} & \dots & \mathbf{r}_{2k} \\ \mathbf{X}_3 & \mathbf{r}_{32} & \mathbf{1} & \dots & \mathbf{r}_{3k} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{X}_k & \mathbf{r}_{k2} & \mathbf{r}_{k3} & \dots & \mathbf{1} \end{bmatrix}$$

¿Cómo se averigua, a partir de la matriz de correlación, si

a).- hay colinealidad perfecta?

Se detecta que hay colinealidad perfecta cuando al menos una de las correlaciones entre pares de variables es igual a 1.

b).- hay colinealidad menos que perfecta?

Se detecta que hay colinealidad menos que perfecta cuando por lo menos una o alguna correlación entre pares de variables explicativa esta entre 0.8 y 0.95.

c).- las X no están correlacionadas?

Pista: puede utilizarse $|R|$ para responder estas preguntas, donde $|R|$ denota el determinante de R.

Nos damos cuenta que las X no están correlacionadas cuando el coeficiente de correlación global (R) tiende a acercarse a cero.

10.24 Con base en la información anual para el sector manufacturero de Estados Unidos durante el período 1899-1922, Dougherty obtuvo los siguientes resultados de regresión:

$$\log \hat{Y} = 2.81 - 0.53 \log K + 0.91 \log L + 0.047t$$

$$ee = (1.38) \quad (0.34) \quad (0.14) \quad (0.021)$$

$$R^2 = 0.97 \quad F = 189.8$$

Donde Y= índice de producción real, K=índice de insumo de capital real, L= índice de insumo trabajo real, t= tiempo o tendencia.

Utilizando la misma información, obtuvo también la siguiente regresión:

$$\log \hat{Y} = -0.11 + 0.11 \log(K / L) + 0.006t$$

$$ee = (0.03) \quad (0.15) \quad (0.006)$$

$$R^2 = 0.65 \quad F = 19.5$$

a) ¿Existe multicolinealidad en la regresión (1)? ¿Cómo se sabe?

Si existe multicolinealidad debido a que las pruebas de significancia individual en su mayoría son poco significativas, y como tenemos un R^2 alto nos indica que individualmente las variables independientes no explican el índice de producción real pero que en forma global las variables explican al modelo en su conjunto.

b) En la regresión (1), ¿cuál es el signo a priori de log K? ¿Concuerdan los resultados con esta expectativa? ¿Por qué sí o por qué no?

El signo a priori del índice de insumo de capital real debe ser positivo ya que la producción y el capital tienen una relación directa, es decir si se incrementa el capital se espera que la producción también se incrementa y viceversa.

c) ¿Cómo justificaría usted la forma funcional de la regresión (1)? (Pista: función de producción de Cobb-Douglas).

Esta forma funcional de la producción a través del tiempo se justifica de mejor manera con un modelo logarítmico que con un modelo lineal, debido a que la recta de regresión se ajusta mejor a los datos.

d) Interpretese la regresión (1). ¿Cuál es el papel de la variable de tendencia en esta regresión?

Es importante introducir en esta regresión la variable tendencia o de tiempo ya que tanto el capital como el trabajo son variables en el largo plazo.

e) ¿Cuál es la lógica detrás de la estimación de la regresión (2)?

$$H_a : (\beta_2 + \beta_3) \neq 1$$

$$n = 24$$

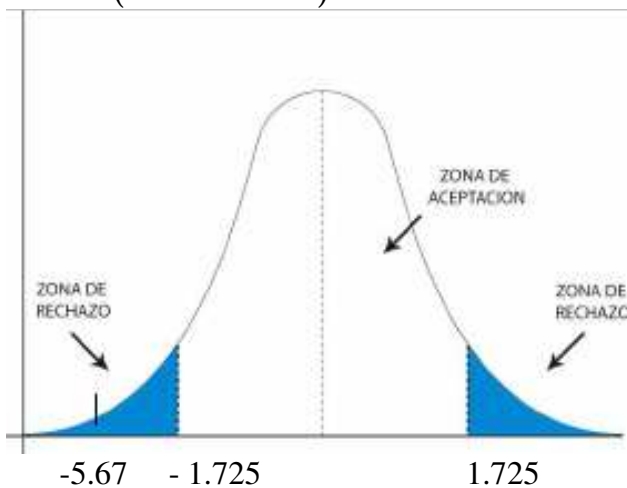
$$g \text{ de } l = 20$$

$$\alpha = 0.05$$

$$t_d = 1.725$$

$$t_c = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{ee(\hat{\beta}_2 + \hat{\beta}_3)}$$

$$t_c = \frac{(0.11 + 0.006) - 1}{(0.15 + 0.006)} = -5.67$$



No existe suficiente evidencia estadística para aceptar la $H_0 : (\beta_2 + \beta_3) = 1$ con un 95% de confianza debido a que el t calculado está fuera de la zona de aceptación por lo tanto se acepta la $H_a : (\beta_2 + \beta_3) \neq 1$, y económicamente se diría que no existen rendimientos constantes a escala.

h).- ¿Son comparables los valores R^2 de las dos regresiones? ¿Por qué si o porque no? ¿Cómo puede hacerse comparables, si no lo son en la forma actual?

No, no son comparables los R^2 y debido que para que sean comparables las variables de los dos modelos deben ser iguales, y en este caso no lo son.

Para tener la misma variable dependiente se realiza lo siguiente:

$$\log Y - \log L = -0.11 + 0.11 \log K - 0.11 \log L + 0.006t$$

$$\log Y = -0.11 + 0.11 \log K - 0.11 \log L + 0.006t + \log L$$

$$\log Y = -0.11 + 0.11 \log K - \log L(0.11 - 1) + 0.006t$$

$$\log Y = -0.11 + 0.11 \log K + 0.81 \log L + 0.006t$$

A partir de este modelo se puede comparar los R^2 ya que las variables dependientes ahora sí son las mismas.

10.26 Klein y Goldberger intentaron ajustar el siguiente modelo de regresión a la economía de Estados Unidos:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

donde Y = consumo, X_2 = ingreso salarial, X_3 = ingreso no salarial, no procedente del campo y X_4 = ingreso procedente del campo. Pero, puesto que se espera que X_2, X_3 y X_4 sean altamente colineales, ellos obtuvieron las siguientes estimaciones de β_3 y β_4 del análisis de corte transversal: $\beta_3 = 0.75 \beta_2$ y $\beta_4 = 0.625 \beta_2$.

Utilizando estas estimaciones, ellos reformularon su función de consumo de la siguiente forma:

$$Y_i = \beta_1 + \beta_2 (X_{2i} + 0.75 X_{3i} + 0.625 X_{4i}) + u_i = \beta_1 + \beta_2 Z_i + u_i$$

$$\text{donde } Z_i = X_{2i} + 0.75 X_{3i} + 0.625 X_{4i}$$

Tabla 10.11

Año	Y	X2	X3	X4
-----	---	----	----	----

a) **Ajústese el modelo modificado a los datos de la tabla 10.11 y obténgase estimaciones de β_1 a β_4 .**

Con los datos de la tabla anterior procedemos a correr el modelo en el EViews, y obtenemos los siguientes resultados:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Z	0.503074	0.052527	9.577428	0.0000
C	20.72698	7.153282	2.897549	0.0134
R-squared	0.884312	Mean dependent var	87.12143	
Adjusted R-squared	0.874671	S.D. dependent var	18.64313	
S.E. of regression	6.600008	Akaike info criterion	6.743582	
Sum squared resid	522.7212	Schwarz criterion	6.834876	
Log likelihood	-45.20508	F-statistic	91.72712	
Durbin-Watson stat	0.796342	Prob(F-statistic)	0.000001	

$Y_i = \beta_1 + \beta_2 Z_i + u_i$, en donde:

$Z = (X_{2i} + 0.75X_{3i} + 0.625X_{4i})$

$$Y_i = 20,726 + 0,5030 Z_i$$

Como β_3 y β_4 dependen de β_2 , entonces solamente al obtener el valor de β_2 podemos saber el valor de los otros estimadores, es así entonces como se obtienen los datos siguientes:

$$B_2 = 0,5030 * X_{2i} = \mathbf{0,5030 X_{2i}}$$

$$B_3 = 0,5030 * 0,75 * X_{3i} = \mathbf{0,3773 X_{3i}}$$

$$B_4 = 0,5030 * 0,625 * X_{4i} = \mathbf{0,3144 X_{4i}}$$

Restaurando el modelo con los datos anteriores obtenemos la siguiente regresión.

$$Y_i = \mathbf{20,726 + 0,5030X_{2i} + 0,3773X_{3i} + 0,3144X_{4i}}$$

b) **¿Cómo se interpretaría la variable Z?**

Dado un incremento unitario en Z se estima que el consumo también se incrementara en 1.1947 dólares porque $Z = 0,5030 X_{2i} + 0,3773 X_{3i} + 0,3144 X_{4i}$ además se puede ver que existe una relación directa entre las variables.

10.27 La tabla 10.12 proporciona cifras sobre importaciones, PNB e índice de precios al consumidor (IPC) para Estados Unidos, durante el período 1970-1998. Se pide considerar el siguiente modelo:

$$\ln \text{Importaciones}_t = \beta_1 + \beta_2 \ln \text{PNB}_t + \beta_3 \ln \text{IPC}_t + u_t$$

Tabla 10.12

observación	IPC	PNB	Importaciones
1970	38,8	1039,7	39866
1971	40,5	1128,6	45579
1972	41,8	1240,4	55797
1973	44,4	1385,5	70499
1974	49,3	1501	103811
1975	53,8	1635,2	98185
1976	56,9	1823,9	124228
1977	60,6	2031,4	151907
1978	65,2	2295,9	176002
1979	72,6	2566,4	212007
1980	82,4	2795	249750
1981	90,9	3131,2	235007
1982	96,5	3259,2	247642
1983	99,6	3534,9	268901
1984	103,9	3932,7	332418
1985	107,6	4213	338088
1986	109,6	4452,9	368425
1987	113,6	4742,5	409765

Dependent Variable: IMPOR

Method: Least Squares

Date: 04/25/07 Time: 18:16

Sample: 1970 1998

Included observations: 29

IMPOR=C(1)+C(2)*PNB+C(3)*LNIPC

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.737832	0.777294	2.235746	0.0342
C(2)	1.170334	0.402363	2.908650	0.0073
C(3)	0.267969	0.564399	0.474785	0.6389
R-squared	0.982883	Mean dependent var	12.48633	
Adjusted R-squared	0.981566	S.D. dependent var	0.905137	
S.E. of regression	0.122891	Akaike info criterion	-1.257343	
Sum squared resid	0.392656	Schwarz criterion	-1.115899	
Log likelihood	21.23148	Durbin-Watson stat	0.523606	

$$\ln \hat{\text{Im portaciones}}_t = 1.738 + 1.170 \ln \text{PNB}_t + 0.268 \ln \text{IPC}_t$$

$$ee = (0.78) \quad (0.40) \quad (0.56)$$

$$t = (2.24) \quad (2.91) \quad (0.47)$$

b) ¿Se sospecha que hay multicolinealidad en los datos?

Si existe multicolinealidad en los datos porque al observar los resultados nos damos cuenta que los valores “t” son poco significativos a pesar de que tenemos un R^2 alto, es decir nos indica que individualmente las variables independientes no explican las variaciones en la importaciones pero si lo hacen de manera global.

c) Efectúense las regresiones:

$$1) \ln \text{Im portaciones}_t = A_1 + A_2 \ln \text{PNB}_t$$

Dependent Variable: IMPOR

Method: Least Squares

Date: 04/25/07 Time: 18:18

Sample: 1970 1998

Included observations: 29

IMPOR=C(1)+C(2)*PNB

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.395046	0.283816	4.915316	0.0000
C(2)	1.360637	0.034708	39.20240	0.0000
R-squared	0.982735	Mean dependent var	12.48633	
Adjusted R-squared	0.982095	S.D. dependent var	0.905137	
S.E. of regression	0.121115	Akaike info criterion	-1.317676	
Sum squared resid	0.396060	Schwarz criterion	-1.223380	
Log likelihood	21.10630	Durbin-Watson stat	0.514920	

A partir del modelo corrido construimos la regresión que es la siguiente.

$$\ln \hat{\text{Im portaciones}}_t = 1.395 + 1.361 \ln \text{PNB}_t$$

$$ee = (0.284) \quad (0.035)$$

$$t = (4.915) \quad (39.20)$$

$$2) \ln \text{Im portaciones}_t = \beta_1 + \beta_2 \ln \text{IPC}_t$$

Dependent Variable: IMPOR

Method: Least Squares

Date: 04/25/07 Time: 18:18

$$\ln \hat{I}mportaciones_t = 3.903 + 1.903 \ln \hat{IPC}_t$$

$$ee = (0.253) (0.056)$$

$$t = (15.43) (34.11)$$

$$3) \ln PNB_t = C_1 + C_2 \ln IPC_t$$

Dependent Variable: PNB
 Method: Least Squares
 Date: 04/25/07 Time: 18:19
 Sample: 1970 1998
 Included observations: 29
 PNB=C(1)+C(2)*LNIPC

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.849908	0.107112	17.27073	0.0000
C(2)	1.397327	0.023628	59.13983	0.0000
R-squared	0.992339	Mean dependent var	8.151538	
Adjusted R-squared	0.992056	S.D. dependent var	0.659462	
S.E. of regression	0.058779	Akaike info criterion	-2.763606	
Sum squared resid	0.093283	Schwarz criterion	-2.669310	
Log likelihood	42.07229	Durbin-Watson stat	0.365873	

A partir del modelo corrido construimos la regresión que es la siguiente.

$$\ln PNB_t = 1.85 + 1.40 \ln \hat{IPC}_t$$

$$ee = (0.11) (0.02)$$

$$t = (17.27) (59.14)$$

Con base en estas regresiones, ¿qué se puede decir sobre la naturaleza de la multicolinealidad en la información?

En base a estas regresiones se puede decir que el producto nacional bruto (PNB) y el índice de precios al consumidor (IPC) para Estados Unidos están altamente correlacionados, es por esta razón que existe multicolinealidad, es decir en conjunto estas dos variables explican las variaciones en las importaciones,

d) Supóngase que existe multicolinealidad en los datos, pero que $\hat{\beta}_2$ y $\hat{\beta}_3$ son significativos individualmente al nivel del 5% y que la prueba global F es también significativa. En este caso, ¿debe preocupar el problema de colinealidad?

A pesar de que $\hat{\beta}_2$ y $\hat{\beta}_3$ son significativos individualmente al nivel del 5% y que la prueba global F existe multicolinealidad pero esto no se considera un problema grave ya que se encontraría en un rango muy bajo de 0 a 0.5 es decir existe una multicolinealidad baja.

EJERCICIO

Datos

PIB	CAPITAL	TRABAJO	TIEMPO
9,7937	20,787	11,113	1
10,2285	21,257	11,262	2
11,0429	21,727	11,424	3
11,3861	22,2294	11,605	4
12,0335	22,7317	11,81	5
11,7738	23,203	11,679	6
11,8811	23,6742	11,504	7
13,2206	21,2198	11,668	8
14,4959	24,7653	11,891	9
15,3725	25,5707	11,829	10
16,9171	26,1899	11,891	11

Dependet Variable: LNPNB
 Method: Least Squares
 Date: 04/26/07 Time: 09:31
 Sample: 2001 2021
 Included observations: 21

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.674319	2.575361	-0.261835	0.7966
LNTRABAJO	0.403773	1.137211	0.355056	0.7269
LNCAPITALL	0.628648	0.155539	4.041736	0.0008
TIEMPO	0.039676	0.008026	4.943527	0.0001
R-squared	0.991804	Mean dependent var	2.888762	
Adjusted R-squared	0.990358	S.D. dependent var	0.443670	
S.E. of regression	0.043566	Akaike info criterion	-3.259419	
Sum squared resid	0.032267	Schwarz criterion	-3.060462	
Log likelihood	38.22389	F-statistic	685.7265	
Durbin-Watson stat	0.479947	Prob(F-statistic)	0.000000	

Pruebas para la detección de la Multicolinealidad

1.- Un R^2 alto y varias razones “t” poco significativas

En esta regresión tenemos un $R^2=0.99$ y el valor “t” de la prueba de significancia individual del trabajo es poco significativo por lo que determinamos que si existe multicolinealidad ya que individualmente el trabajo no explica las variaciones en el PNB pero conjuntamente las variables independientes si explican las variaciones en el PNB.

2.- Altas correlaciones simples entre pares de variables

Correlaciones

	LNPIB	LNTRABAJO	LNCAPITAL	TIEMPO
LNPIB	1.000000	0.977958	0.986022	0.990562
LNTRABAJO	0.977958	1.000000	0.970110	0.975764
LNCAPITALL	0.986022	0.970110	1.000000	0.970396
TIEMPO	0.990562	0.975764	0.970396	1.000000

Analizando las correlaciones entre pares de variables explicativas trabajo-capital, trabajo-tiempo, capital-tiempo se puede concluir diciendo que, de acuerdo a esta prueba si existe multicolinealidad ya que estas son correlaciones altas.

3.- Examen de correlaciones parciales

$$R_{global}^2=0.992$$

Debido a que el R_{global}^2 es mayor a los valores de las correlaciones parciales queda comprobado con esta prueba una vez más que si existe multicolinealidad entre las variables explicativas del modelo.

Matriz de correlaciones parciales

C1 C2 C3 C4

Dependent Variable: LNTRABAJO

Method: Least Squares

Date: 04/26/07 Time: 10:23

Sample: 2001 2021

Included observations: 21

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNCAPITALL	0.060124	0.028956	2.076419	0.0525
TIEMPO	0.004139	0.001347	3.072043	0.0066
C	2.236741	0.083510	26.78403	0.0000
R-squared	0.961368	Mean dependent var	2.485329	
Adjusted R-squared	0.957076	S.D. dependent var	0.043584	
S.E. of regression	0.009030	Akaike info criterion	-6.445028	
Sum squared resid	0.001468	Schwarz criterion	-6.295811	
Log likelihood	70.67280	F-statistic	223.9691	
Durbin-Watson stat	1.219962	Prob(F-statistic)	0.000000	

Regresión 2

Dependent Variable: LNCAPITALL

Method: Least Squares

Date: 04/26/07 Time: 10:09

Sample: 2001 2021

Included observations: 21

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNTRABAJO	3.214045	1.547879	2.076419	0.0525
TIEMPO	0.023125	0.010873	2.126903	0.0475
C	-4.865026	3.730408	-1.304154	0.2086
R-squared	0.952940	Mean dependent var	3.377305	
Adjusted R-squared	0.947712	S.D. dependent var	0.288717	
S.E. of regression	0.066020	Akaike info criterion	-2.466154	
Sum squared resid	0.078456	Schwarz criterion	-2.316937	
Log likelihood	28.89462	F-statistic	182.2470	
Durbin-Watson stat	0.328898	Prob(F-statistic)	0.000000	

Regresión 3

Dependent Variable: TIEMPO

Method: Least Squares

Date: 04/26/07 Time: 10:11

Sample: 2001 2021

Included observations: 21

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNTRABAJO	83.10117	27.05078	3.072043	0.0066
LNCAPITALL	8.685160	4.083477	2.126903	0.0475
C	-224.8661	53.95576	-4.167602	0.0006
R-squared	0.961732	Mean dependent var	11.00000	
Adjusted R-squared	0.957480	S.D. dependent var	6.204837	
S.E. of regression	1.279457	Akaike info criterion	3.462312	
Sum squared resid	29.46618	Schwarz criterion	3.611530	
Log likelihood	-33.35428	F-statistic	226.1848	
Durbin-Watson stat	0.467259	Prob(F-statistic)	0.000000	

Corridos las regresiones con los datos proporcionados obtuvimos los siguientes valores:

$$R_{global}^2 = 0.992$$

el R^2 global es menor al R^2 auxiliar, es decir que existe mayor relación entre la variable dependiente con las explicativas que entre las explicativas mismas.

$$R_{global}^2 = 0.992$$

$$R_j^2 = 0.962 \quad F_3 = 226.18$$

Relacionando las variables explicativas tiempo en función del trabajo y el capital podemos darnos cuenta que el R^2 de la regresión auxiliar es igual a 0.962 y por lo tanto es menor que el R^2 global de 0.992 lo cual podríamos decir que existe un problema de multicolinealidad pero este no es grave ya que el problema es preocupante cuando el R^2 global es menor al R^2 auxiliar, es decir que existe mayor relación entre la variable dependiente con las explicativas que entre las explicativas mismas.

5.- Análisis del Factor Inflador de la Varianza

Para el cálculo del Factor Inflador de la Varianza se procede de la siguiente manera:

$$FIV = \frac{1}{(1 - R_j^2)}$$

Regresión 1

$$FIV = \frac{1}{(1 - 0.961368)} = 25.885$$

Regla práctica

En general si el FIV es mayor o igual a 10 ($\Rightarrow R_j^2 \geq 0.9$) hay evidencia de alta colinealidad

Tomando en cuenta la regla práctica podemos decir que sí existe un alto grado de colinealidad entre las variables explicativas del modelo debido a que el Factor Inflador de la Varianza en este caso es mayor que 10.

Regresión 2

$$FIV = \frac{1}{(1 - 0.952940)} = 21.249$$

Tomando en cuenta la regla práctica podemos decir que sí existe un alto grado de colinealidad entre las variables explicativas del modelo debido a que el Factor Inflador de la Varianza en este caso es mucho mayor que 10.

Regresión 3

$$FIV = \frac{1}{(1 - 0.961732)} = 26.131$$

Tomando en cuenta la regla práctica podemos decir que sí existe un alto grado de colinealidad entre las variables explicativas del modelo debido a que el Factor Inflador de la Varianza en este caso es mucho mayor que 10.

6.- Observar el número de condición y el índice de condición

Diagnóstico de colinealidad

Modelo	Dimensión	Autovalor	Índice de condición	Proporciones de la varianza			
				(Constante)	LNTRABAJ	LNCAPITA	TIEMPO
1	1	3.835	1.000	.00	.00	.00	.00
	2	.164	4.832	.00	.00	.00	.04
	3	.000	118.864	.01	.00	.86	.55

A partir del Índice de condición de $\hat{\beta}_2$ se obtiene el valor de K que es de 23.35

Como este valor esta entre 10 y 30 se puede decir que existe multicolinealidad entre moderada y fuerte.

A partir del Índice de condición de $\hat{\beta}_3$ se obtiene el valor de K que es de 14128.65

Como este valor excede a 1000 se dice entonces que existe una multicolinealidad severa.

A partir del Índice de condición de $\hat{\beta}_4$ se obtiene el valor de K que es de 618607.42

Como este valor excede a 1000 se dice entonces que existe una multicolinealidad severa.

7.- Prueba de Farrar-Glauber

$$H_0 : |R_x| = 1$$

$$H_a : |R_x| = 0$$

con $K' \frac{K' - 1}{2}$ grados de libertad

$$G_d = 9.34840 \text{ con } \alpha / 2$$

$$G_d = 0.215795 \text{ con } 1 - \alpha / 2$$

$$gl = 3$$

$$G_c = - \left[n - 1 - \frac{1}{6} (2K' + 5) \right] \ln |R_x|$$

$$G_c = - \left[21 - 1 - \frac{1}{6} (2 * 3 + 5) \right] \ln |0.0000184693301439|$$

$$G_c = - \left[21 - 1 - \frac{1}{6} (2 * 3 + 5) \right] (-10.89939903)$$

$$G_c = 198.01$$

Interpretación

Debido a que el G calculado es mayor que el G dado, no existe suficiente evidencia estadística para aceptar la hipótesis nula de que $H_0 : |R_x| = 1$ con un 95% de confianza, es decir se acepta la $H_a : |R_x| = 0$ por lo tanto si existe multicolinealidad.

Corrección de la Multicolinealidad

Como medida correctiva para la multicolinealidad nosotras hemos tomado la decisión de “no hacer nada” ya que las variables explicativas tienen una relevancia teórica importante en la especificación del modelo, es decir si eliminaríamos cualquier variable provocaríamos un sesgo de especificación.

