Tarea 5 - Estadística Aplicada II
Equipo 3

Roberto Tello Ayala - 154706 Jesús Antonio Piñera - 157909
Manuel Quintero - 159889 José Pliego - 157103

Álvaro Rangel Ochoa - 156063

24 de noviembre de 2019

Ajuste un modelo de regresión lineal simple a los datos del siguiente cuadro, en el cual se presentan las variables Y = Indice de la Bolsa y X = Producto Nacional Bruto de un cierto país (PNB).
Realice el **análisis de residuos** y, en particular, **verifique** el supuesto de no-correlación de los errores. De existir autocorrelación, **corríjala** y compare los errores estándar de los coeficientes para los modelos SIN y CON corrección por autocorrelación.

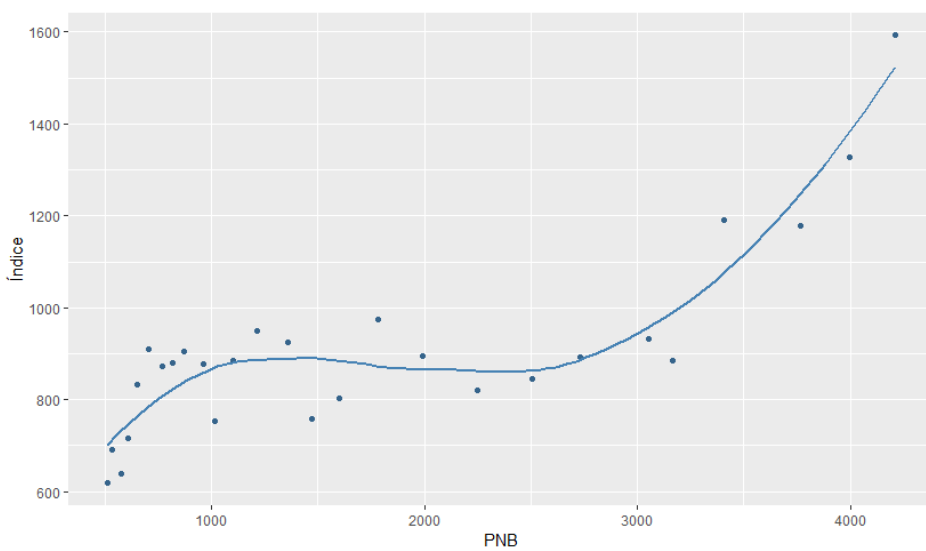
Tabla 1. Datos

Año	PNB	Indice
1990	515	618.04
1991	534	691.55
1992	575	639.76
1993	607	714.81
1994	650	834.05
1995	705	910.88
1996	772	873.60
1997	816	879.12
1998	873	906.00
1999	964	876.72
2000	1,016	735.19
2001	1,103	884.76
2002	1,213	950.71
2003	1,359	923.88
2004	1,473	759.37—

2005	1,598	802.49
2006	1,783	974.92
2007	1,991	894.63
2008	2,250	820.23
2009	2,508	844.40
2010	2,732	891.41
2011	3,053	932.92
2012	3,166	884.36
2013	3,406	1,190.34
2014	3,765	1,178.48
2015	3,998	1,328.23
2016	4,209	1,592.76

Consideraremos el siguiente modelo de regresión lineal simple:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{1}$$



Gráfica 1. PNB contra Índice

Utilizando el método de MCO se obtienen los siguientes resultados: el intercepto tiene un valor de $b_0 = 664.8781$ lo cual nos indica que para un PNB de 0 el índice de la bolsa tiene un valor de 664.8781. Similarmente, la pendiente nos muestra $b_1 = 0.1385$ cuánto incrementa el índice ante un aumento unitario del PNB. El modelo obtenido es

$$\hat{\text{Indice}}_t = 664.8781 + 0.1385\text{PNB}_t$$

Asimismo,se obtienen p-values de 1.24e-13 y 1.21e-06 para los estimadores de β_0 y β_1 respectivamente. Así, se puede concluir que los 2 son estadísticamente significativos (para niveles de significancia muy bajos). Asimismo, para el estadístico F se tiene un p-value: 1.207e-06, indicando que el modelo posee (aparentemente) capacidad explicativa. Es decir, se rechaza la hipótesis nula $H_0 : Y = \beta_0 + \epsilon$. Finalmente, se obtiene $R^2 = 0.6172$ y $\bar{R}^2 = 0.6018$.

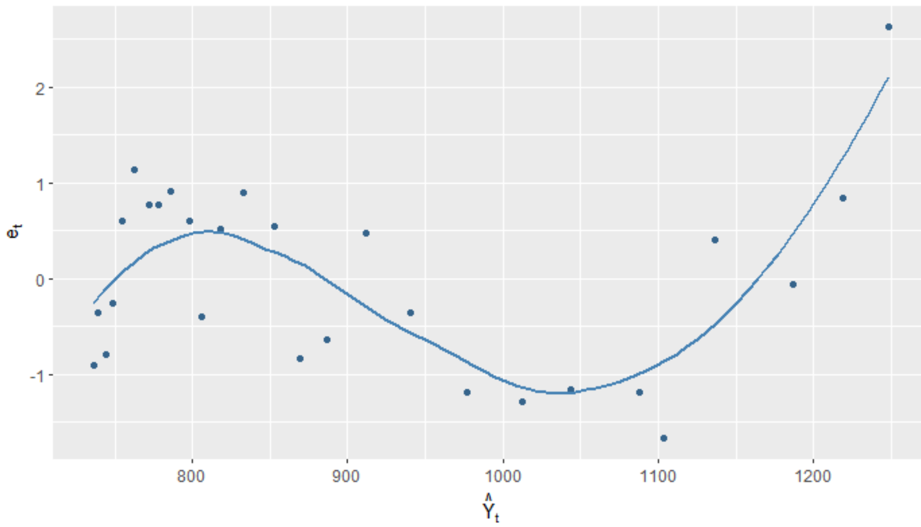
Tabla 2. Resultados obtenidos de EViews con el modelo (1)

Dependent Variable: INDICE
Method: Least Squares
Date: 04/30/19 Time: 14:55
Sample: 1990 2016
Included observations: 27

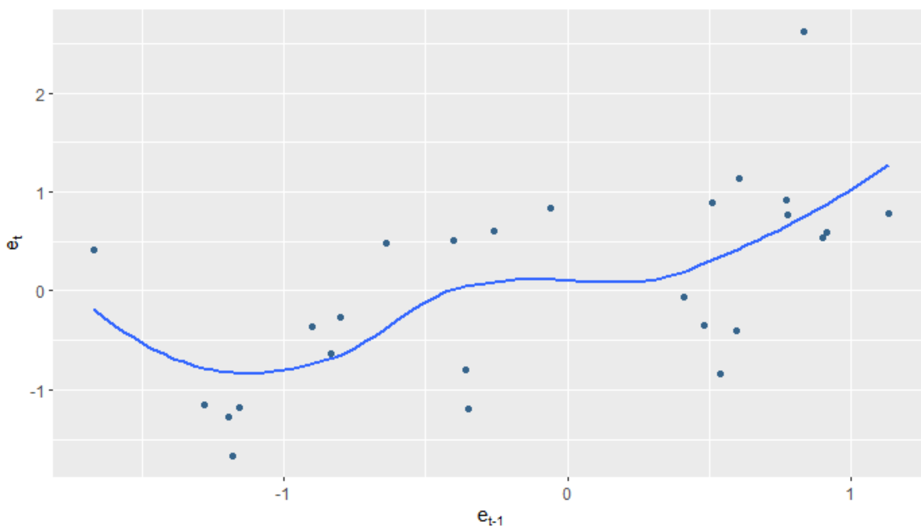
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	664.8781	46.04952	14.43833	0.0000
PNB	0.138554	0.021826	6.348217	0.0000

R-squared	0.617151	Mean dependent var	909.3189
Adjusted R-squared	0.601837	S.D. dependent var	207.9815
S.E. of regression	131.2367	Akaike info criterion	12.66307
Sum squared resid	430576.6	Schwarz criterion	12.75906
Log likelihood	-168.9514	Hannan-Quinn criter.	12.69161
F-statistic	40.29986	Durbin-Watson stat	0.701461
Prob(F-statistic)	0.000001		

Por otro lado, realizando un análisis de residuos, se obtuvieron los siguientes resultados. En la gráfica 2 se observa $\hat{Y}$ contra los residuos. Es claro que los residuos muestran cierta tendencia, similar a aquella que se observa en la gráfica 1. Asimismo, al graficar e_{t-1} contra e_t se observa una clara correlación positiva entre los residuos (gráfica 3). Esto indica que existe autocorrelación entre los errores, lo cual podría invalidar los resultados obtenidos anteriormente.



Gráfica 2. $\hat{Y}_t$ contra e_t del modelo (1)



Gráfica 3. e_t contra e_{t-1} estandarizados del modelo (1)

Para analizar de manera numérica la violación al supuesto de no correlación de los errores, se utiliza el estadístico de Durbin y Watson. Se obtiene el valor de $DW = 0.7014$. Al nivel de significancia del 1 %, con $k = 1$ el número de regresores y $n = 27$ el tamaño de la muestra, obtenemos los límites $d_L = 1.088$ y $d_U = 1.232$. Dado que el estadístico de Durbin-Watson es $0.7014 < d_L = 1.088$, entonces se rechaza la hipótesis nula $H_0 : \rho = 0$ y optamos por $H_A : \rho > 0$. Es decir, los errores están correlacionados positivamente. Por lo tanto, los resultados obtenidos anteriormente se invalidan.

Dado el problema encontrado, se asume un modelo autoregresivo de primer orden. Es decir, dado el modelo inicial (1), los errores cumplen que

$$\epsilon_t = \rho\epsilon_{t-1} + \delta_t$$

donde δ_t satisface los supuestos clásicos de MCO. Se utiliza, pues, el siguiente modelo transformado

$$Y_t - \rho Y_{t-1} = (1-\rho)\beta_0 + \beta_1(X_t - \rho X_{t-1}) + \delta_t$$

Primero, se utilizaron los métodos de Hildreth-Lu y Cochrane-Orcutt para solucionar el problema de autocorrelación, los cuales arrojaron resultados prácticamente idénticos. Se presentan los resultados del procedimiento de Hildreth-Lu utilizando el modelo transformado

$$Y_t^{'} = \beta_0^{'} + \beta_1^{'} X_t^{'} + \delta_t \tag{2}$$

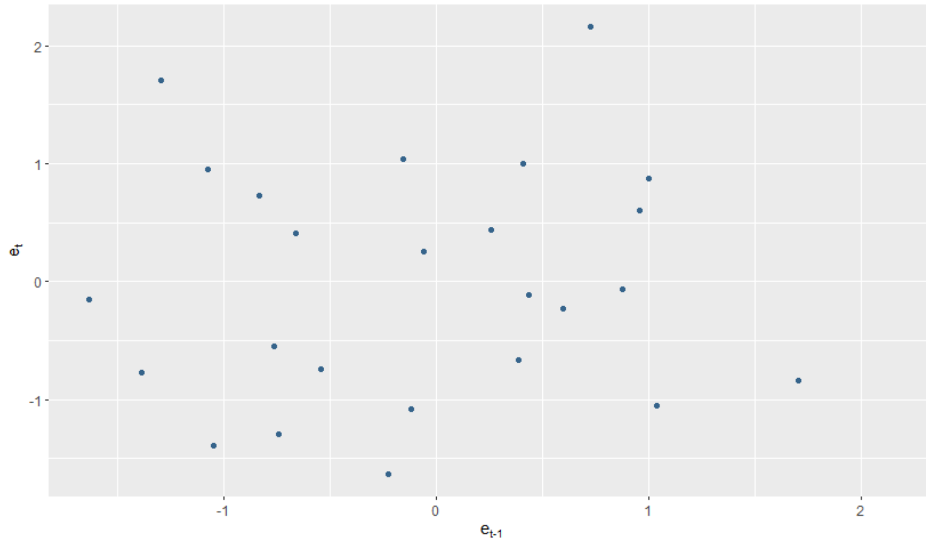
donde δ_t es idéntica a aquella en (2), $Y_t^{'} = Y_t - \rho Y_{t-1}$, $X_t^{'} = (X_t - \rho X_{t-1})$, y $\beta_0^{'} = (1-\rho)\beta_0$.

Con este modelo se obtiene lo siguiente: $b_0^{'} = 177.2869$, con un p-value de 0.0001; $b_1^{'} = 0.1760$ con un p-value de 0.0031; un estadístico $F = 10.82$ con un p-value de 0.0031; y finalmente $R^2 = 0.3107$ y $\bar{R}^2 = 0.2820$ [1]. La fórmula está dada por

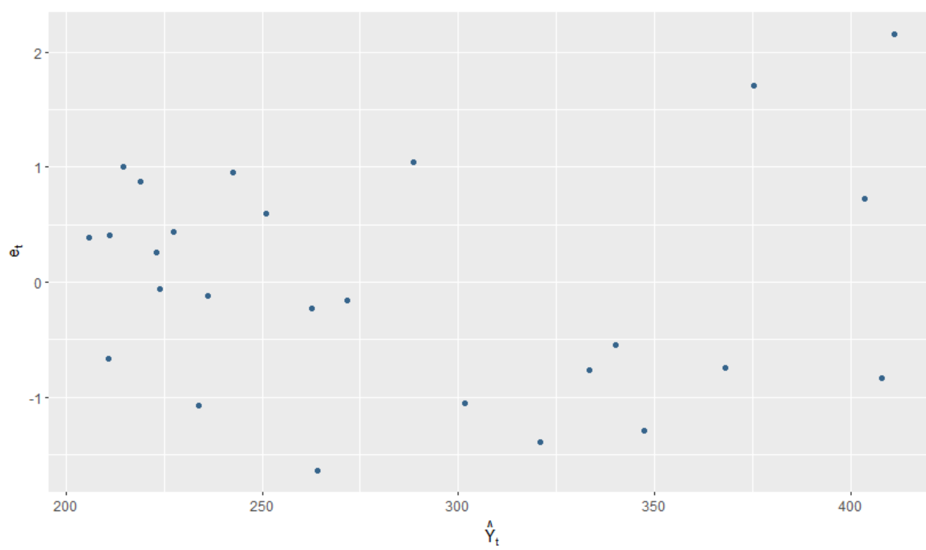
$$\hat{\text{Indice}}_t^{'} = 177.2869 + 0.1760\,\text{PNB}_t$$

Además, al calcular el estadístico de Durbin y Watson se obtiene el valor de $DW = 1.6923$, el cual se encuentra por encima de $d_U = 1.232$ y por debajo de $4 - d_U = 2.768$. Es decir, no se rechaza la hipótesis nula $H_0 : \rho^{'} = 0$. De hecho, si se observa la gráfica 4, no parece haber una relación aparente entre los residuos y ellos mismos un periodo anterior. Asimismo, tampoco parece haber relación aparente entre las estimaciones y los residuos (gráfica 5). Esto nos indica que, estadísticamente hablando, no hay autocorrelación aparente. Así, nuestros resultados son válidos.

[1] Se obtuvo también, a través del método iterativo de Cochrane-Orcutt, el valor apróximado de ρ dado por $\hat{\rho} = 0.7203$



Gráfica 4. $e_t^{'}$ contra $e_{t-1}^{'}$ estandarizados del modelo (2) utilizando Hildreth-Lu



Gráfica 5. $e_t^{'}$ estandarizados contra $\hat{Y}_t^{'}$ del modelo (2) utilizando Hildreth-Lu

Ahora bien, para mejorar el análisis, se utiliza un método autorregresivo de orden 1 con máxima verosimilitud [2], AR(1), que hace uso de la ecuación:

$$Y_t = \beta_0(1-\rho) + \beta_1 X_t - \beta_1 \rho X_{t-1} + \rho Y_{t-1} + u_t \tag{3}$$

[2] La paquetería de EViews utilizó este método de forma predeterminada al solicitar un modelo AR(1), así que se decidió utilizar dicho método en lugar de MCO.

Tabla 3. Resultados obtenidos de EViews con un modelo autoregresivo de primer orden

Dependent Variable: INDICE
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 04/30/19 Time: 15:15
Sample: 1990 2016
Included observations: 27
Convergence achieved after 9 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	589.5055	157.0658	3.753238	0.0010
PNB	0.188558	0.059890	3.148394	0.0045
AR(1)	0.710900	0.170395	4.172061	0.0004
SIGMASQ	9794.692	4114.880	2.380310	0.0260

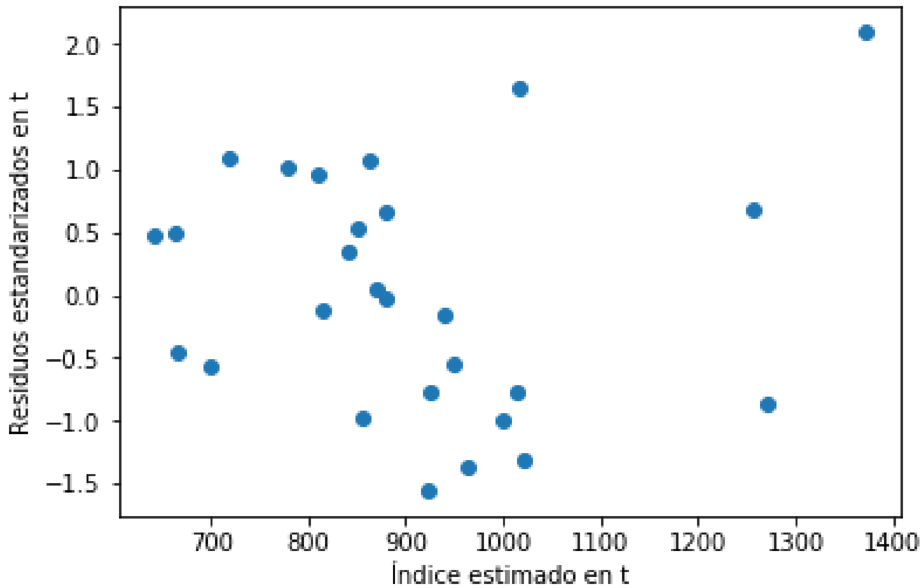
R-squared	0.764857	Mean dependent var	909.3189
Adjusted R-squared	0.734186	S.D. dependent var	207.9815
S.E. of regression	107.2293	Akaike info criterion	12.34984
Sum squared resid	264456.7	Schwarz criterion	12.54182
Log likelihood	-162.7229	Hannan-Quinn criter.	12.40693
F-statistic	24.93762	Durbin-Watson stat	1.691324
Prob(F-statistic)	0.000000		

Inverted AR Roots	.71

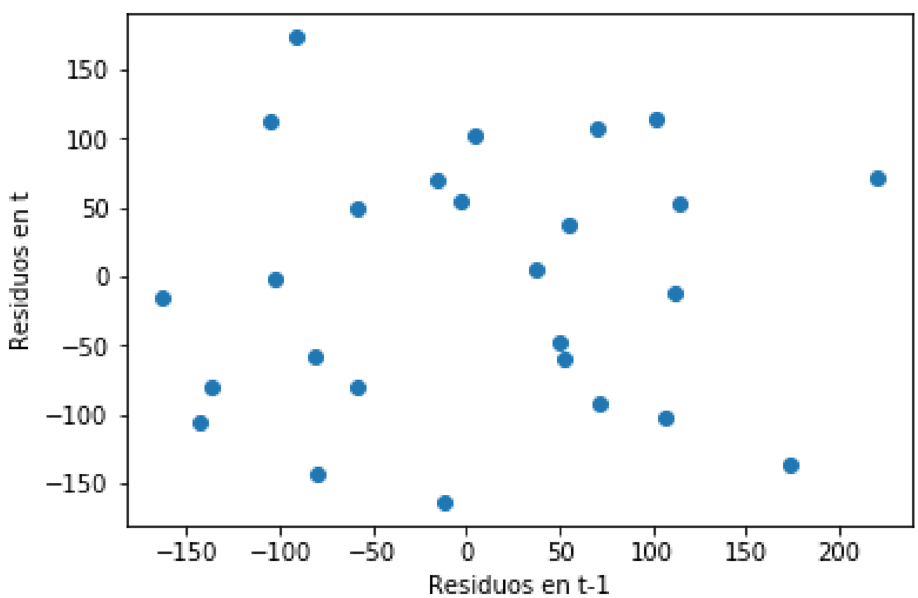
Al estimar los valores obtenemos:
$\hat{\rho} = 0.7109$
$b_0 = 589.5055$
$b_1 = 0.1886$
Por lo que el modelo quedaría como sigue:

$$\text{Índice}_t = 170.4260 + 0.1886(\text{PNB}_t - 0.7109\,\text{PNB}_{t-1}) + 0.7109\,\text{Índice}_{t-1}$$

Este modelo tiene una $R^2 = 0.7649$ y una $\bar{R}^2 = 0.7342$. El estadístico de Durbin y Watson tiene un valor de 1.6013 el cual se encuentra entre $d_U = 1.232$ y $4 - d_U = 2.768$, por lo que no se rechaza $H_0 : \rho = 0$. Además, en la Gráfica 6 podemos observar que los residuos no parecen presentar ningún comportamiento sistemático y en la Gráfica 7 no se aprecia algún tipo de correlación entre los residuos. Por lo tanto, los resultados obtenidos son válidos. Si se compara con el modelo obtenido con los métodos de Hildreth y Lu y Cochrane y Orcutt, se obtiene una R^2 y $\bar{R}^2$ mucho mayores, arreglando de la misma forma el problema de autocorrelación existente en el modelo inicial (1). Así, nuestro modelo es apropiado para explicar el índice utilizando un modelo AR(1) y la variable explicativa PNB.



Gráfica 6. e_t contra $\hat{Y}_t$ con el Método AR(1)



Gráfica 7. e_t contra e_{t-1} con el Método AR(1)