

UNIVERSIDAD AUTONOMA DE NUEVO LEON
FACULTAD DE AGRONOMIA



AUTOCORRELACION ENTRE LAS PERTURBACIONES
 U_t PARA $l = f$ (PIB)

OPCION V (CASO TEORICO-PRACTICO)
QUE PARA OBTENER EL TITULO DE
INGENIERO AGRONOMO EN DESARROLLO RURAL

PRESENTA:

SAUL GUZMAN CORONADO

JULIO DE 1987.

040.519
FA1
1987
c.5

T
QA278
.2
G8
C.1

A

040.519
FA1
1987
C.5

T
QA278
.2
G8
C.1



1080063945

UNIVERSIDAD AUTONOMA DE NUEVO LEON

FACULTAD DE AGRONOMIA



AUTOCORRELACION ENTRE LAS PERTURBACIONES

U_t PARA $I = f$ (PIB)

OPCION V (CASO TEORICO-PRACTICO)

QUE PARA OBTENER EL TITULO DE:

INGENIERO AGRONOMO EN DESARROLLO RURAL

P R E S E N T A:

SAUL GUZMAN CORONADO



11007368 *[Signature]*

MARIN, N.L.

JULIO DE 1987

[Handwritten signature]

T
QA278
-2
G8

040.519
FA1
197
C.5



Biblioteca Central
Magna Solidaridad

F. tesis



DEDICATORIAS

A MIS PADRES.

Dos grandes en verdad; pioneros de mi educación
y forjadores de mi anhelo de progreso.

A MIS HERMANOS.

Cuyo ejemplo, tezhón y voluntad trazaron la senda
a seguir, logrando en mi el término de mi formación -
profesional.

A MIS AMIGOS.

Que de una u otra manera contribuyeron alentando
mi espíritu de lucha.

AGRADECIMIENTOS

AL LIC. EFRAIN RIVERA CARRILLO, y al

ING. M.C. JOSE I. MENDEZ PEREZ.

Por su noble y mesurada participación en el presente trabajo.

A TODOS MIS MAESTROS.

Que en el largo recorrido de mi camino escolar, fueron ejemplo de bonanza.

A todos aquellos que hacen posible, con su - esfuerzo y sudor, la educación de los pobres.

INDICE GENERAL

	Página
I.- INTRODUCCION	1
II.- LA AUTOCORRELACION	2
III.- CONSECUENCIAS DE LA AUTOCORRELACION	4
IV.- ALGUNOS METODOS PARA DETECTAR LA AUTOCORRELA- CION	5
IV. 1.- El Método Gráfico	5
IV. 2.- El Método Analítico: Estadístico Dur- bin-Watson	5
IV. 2.1 Proceso de la Prueba Durbin-Watson	8
IV. 2.2 Forma de Remediar la Autocorrela- ción	9
V.- DESARROLLO DEL PROBLEMA	10
V. 1.- Eliminación del Problema	13
V. 1.1 Primera Transformación de los Datos. .	13
V. 1.2 Segunda Transformación de los Datos. .	16
VI.- CONCLUSIONES	19
VII.- BIBLIOGRAFIA	20

I.- INTRODUCCION

El propósito central del presente trabajo consiste en mostrar si existe autocorrelación (correlación serial) entre la inversión (I) y el Producto Interno Bruto (PIB); y en segundo término, si existe, eliminarla usando el método analítico del estadístico Durbin-Watson (d)

Cuando la correlación serial es eliminada, los datos del problema pueden ser usados para hacer predicciones, utilizando el método de los Mínimos Cuadrados Ordinarios (CMO). Sin embargo, en este caso, no se realizan por dos razones principalmente: a).- Porque el número de observaciones incluidas es muy reducido; y b).- Porque se trabaja únicamente con una variable explicatoria, lo que no permite comparar o relacionar la influencia del PIB respecto a otras sobre la inversión.

El trabajo está dividido en dos partes: la primera, del punto II al IV, contiene las bases teóricas, someramente tratadas para entender y lograr la realización del problema; la segunda contiene en sí el desarrollo del problema que nos ocupa.

II.- LA AUTOCORRELACION

Gujarati (1978) define la autocorrelación como la "Correlación existente entre los miembros de una serie de observaciones ordenadas en el tiempo (como las cifras de series de tiempo) o en el espacio (como las cifras de corte transversal)" (1).

El análisis de regresión lineal clásico supone la no existencia de autocorrelación en las perturbaciones U_i , lo cual se expresa simbólicamente de la siguiente manera:

$$E (U_i U_j) = 0 \quad \forall \quad i \neq j$$

Es decir, "el término de perturbación perteneciente a una observación no está influenciado por el término de perturbación perteneciente a otra" (1). Por el contrario, en el caso de existir autocorrelación entre las perturbaciones U_i , simbólicamente se expresa así:

$$E (U_i U_j) \neq 0 \quad \forall \quad i \neq j$$

O sea, la perturbación de una observación está influenciado por la perturbación de otra, fenómeno que se presenta por alguno de los factores siguientes: inercia, sesgo de especificación (caso de variables excluidas y mal especificación de la función), el fenómeno de la telaraña y manipulación de datos.

Cuando Maddala (1985), analiza los factores causales de la autocorrelación, menciona cuatro razones por las que se pue

de presentar la correlación serial, haciendo énfasis principalmente sobre el caso de las variables excluidas (3).

La correlación serial se presenta más frecuentemente sobre las series de tiempo con respecto a las de corte transversal (1).

III.- CONSECUENCIAS DE LA AUTOCORRELACION.

El análisis de regresión lineal clásico asume la no existencia de autocorrelación entre las perturbaciones U_i , siendo los Mínimos Cuadrados Ordinarios (CMO) los mejores estimadores; más, cuando el problema de la autocorrelación serial se presenta, los estimadores CMO ya no son los mejores estimadores (ya no son eficientes), aunque se mantengan todos los demás supuestos invariables (que son: insesgados y consistentes). Los autores antes citados nos advierten sobre algunas consecuencias si se persiste en aplicar dichos estimadores: primero, los intervalos de confianza serán más anchos de lo necesario y la prueba de significancia menos fuerte; segundo, se tiende a subestimar la verdadera varianza residual (σ^2) y los errores estándar; y tercero, los estimadores tienden a dar una visión distorsionada de los valores poblacionales, es decir, que los estimadores CMO se vuelven sensibles a las fluctuaciones (1) (3).

IV.- ALGUNOS METODOS PARA DETECTAR LA AUTOCORRELACION

Hasta la actualidad los métodos más comunes para diagnosticar la existencia o no de autocorrelación en un modelo de regresión son: el gráfico y el analítico.

IV. 1.- El Método Gráfico.- Consiste en relacionar los errores e_t (\hat{U}_i) con el tiempo; los e_t se obtienen mediante el método de CMO. Una ventaja de este método es su simplicidad para dibujarlo sin que el número de variables explicatorias influya; de esta manera se compara el resultado gráfico con los patrones de autocorrelación, proporcionando una prueba visual en caso de existir.

IV. 2.- El Método Analítico (Estadístico d).- Maddala señala como método mas usado la prueba Durbin-Watson (3). Gujarati afirma lo mismo al mencionar que es el mas apropiado para tales casos, y lo define así:

$$d = \frac{\sum_{t=2}^{t=n} (e_t - e_{t-1})^2}{\sum_{t=1}^{t=n} e_t^2}$$

Se puede observar que en el numerador del estadístico Durbin-Watson (d), el número de observaciones es n-1 (t=2) por haberse perdido una de ellas al tomar las diferencias consecutivas. Una gran ventaja del estadístico d consiste en estar basado en los residuos estimados que se calculan automáticamente en el análisis de regresión. Por eso hoy es común incluir

en los informes de análisis de regresión al d al lado de R^2 y \bar{R}^2 , razones t, etc. (1).

Los supuestos del estadístico d son los siguientes (1):

- a).- En el modelo de regresión, su intercepto no debe pasar por el origen.
- b).- Las variables explicatorias X's no son estocásticas, o son fijas en muestras repetidas.
- c).- Las perturbaciones U_t se generan mediante un esquema autorregresivo de primer orden:

$$U_t = \rho U_{t-1} + \varepsilon_t$$

- d).- En el modelo de regresión no deben incluirse las variables dependientes como variables explicatorias. Cuando este supuesto se viole, Durbin ha desarrollado el estadístico h para verificar la autocorrelación en dichos modelos.

La distribución de probabilidad exacta del estadístico d es difícil de encontrar, pues el cálculo del mismo depende tanto de los ε_t como de los X's dados, en comparación con las pruebas t, F o X^2 , cuyo cálculo es más sencillo porque se tiene un valor crítico único que lleva a rechazar o aceptar la hipótesis nula. Para resolver este problema, Durbin y Watson establecieron un límite inferior (d_L) y un límite superior (d_U) de valores de tal manera que:

- a).- Si el d calculado cae por fuera de estos valores críticos puede tenerse una decisión sobre la posible presencia de correlación serial positiva o negativa

b).- Esos límites dependen únicamente del número de observaciones (N), del número de variables explicatorias, y del valor que tomen dichas variables explicatorias (1).

Los límites se obtienen desarrollando el estadístico d a partir de:

$$d = \frac{\sum e_t^2 + \sum e_{t-1}^2 - \sum e_t e_{t-1}}{\sum e_t^2}$$

pero como $\sum e_t^2$ y $\sum e_{t-1}^2$ difieren entre sí tan sólo por una observación, se consideran aproximadamente iguales, de tal manera que si $\sum e_t^2 = \sum e_{t-1}^2$, entonces:

$$d = 2 \left(1 - \frac{\sum e_t e_{t-1}}{\sum e_t^2} \right); \text{ ahora bien, como}$$

$$\hat{\epsilon} = \frac{\sum e_t e_{t-1}}{\sum e_t^2}, \text{ se puede substituir en la anterior}$$

ecuación del estadístico d y tenemos:

$$d = 2 (1 - \hat{\epsilon}) \text{ de donde se concluye:}$$

- 1.- Si $\hat{\epsilon} = 0$; $d = 2$. Es decir, si no hay correlación serial de primer orden positiva o negativa, se espera que $d = 2$.
- 2.- Si $\hat{\epsilon} = 1$; $d = 0$. O sea, existe correlación serial positiva, y mientras más cerca esté d de 0, mayor será la evidencia de la existencia de correlación serial positiva.
- 3.- Si $\hat{\epsilon} = -1$; $d = 4$. Quiere decir que hay correlación serial negativa entre los valores residuales. Entre más cerca este d de 4, mayor será la evidencia de autocorrelación negativa.

IV. 2.1.- Proceso de la Prueba Durbin-Watson.

- 1.- Correr la regresión de CMO y obtener los residuos e_t .
- 2.- Calcular el d .
- 3.- Encontrar los valores críticos d_L y d_U para el tamaño de la muestra y el número de variables explicatorias dadas.
- 4.- Si la hipótesis nula (H_0) es la de que no hay correlación serial positiva, entonces si:

$$\begin{array}{lll} d < d_L & : & \text{rechace } H_0 \\ d > d_U & : & \text{no rechace } H_0 \\ d \leq d_L & \text{ y } & d \leq d_U & : & \text{la prueba no es concluyente.} \end{array}$$

- 5.- Si H_0 es la de que no hay correlación serial negativa, entonces si:

$$\begin{array}{lll} d > 4 - d_L & : & \text{rechace } H_0 \\ d < 4 - d_U & : & \text{no rechace } H_0 \\ 4 - d_U \leq d \leq 4 - d_L & : & \text{la prueba no es concluyente.} \end{array}$$

- 6.- Si H_0 es de dos colas, es decir, que no hay correlación serial positiva o negativa, entonces si:

$$\begin{array}{lll} d < d_L & : & \text{rechace } H_0 \\ d > 4 - d_L & : & \text{rechace } H_0 \\ d_U < d < 4 - d_U & : & \text{no rechace } H_0 \\ d_L \leq d \leq d_U & : & \text{la prueba no es concluyente} \\ 4 - d_U \leq d \leq 4 - d_L & : & \text{la prueba no es concluyente} \end{array}$$

Cuando el estadístico d cae en la zona de indecisión no se puede concluir. Entonces, en estos casos, se recurre a

otras pruebas, a la adición de cifras, o bien a cambiar la serie (1).

IV. 2.2.- Forma de Remediar la Autocorrelación (2) (4).

La forma más simple de remediar la autocorrelación consiste en transformar los valores originales de las variables Y_t y X_t observados, usando el siguiente procedimiento:

$$Y'_t = Y_t - \hat{e} Y_{t-1}$$

$$X'_t = X_t - \hat{e} X_{t-1} \quad , \quad \text{donde:}$$

$$Y'_t \text{ y } X'_t = \text{valores transformados}$$

$$Y_{t-1} \text{ y } X_{t-1} = \text{valor observado menos una observación}$$

$$\hat{e} = \text{Coeficiente de autocorrelación muestral de primer orden.}$$

Como se puede notar, se pierde la primera observación, por lo que, si se desea recuperar, se emplea la siguiente fórmula:

$$Y'_1 = Y_1 \sqrt{1-e^2}$$

$$X'_1 = X_1 \sqrt{1-e^2}$$

Este proceso se repite hasta obtener los resultados deseados, pero comúnmente con la primera o segunda transformación de los datos originales se obtienen resultados satisfactorios. Para nuestro caso, hubo necesidad de recurrir a una segunda transformación de los datos.

V.- DESARROLLO DEL PROBLEMA

Los datos utilizados son los siguientes:

Inversión (I) y Producto Interno Bruto (PIB), a precios constantes de 1970.

Período	Y_t	X_t
1970	100,956	444,271
1971	96,041	462,804
1972	106,148	520,860
1973	122,327	544,307
1974	143,619	577,568
1975	150,851	609,976
1976	147,397	635,831
1977	146,938	657,722
1978	164,472	711,983
1979	193,418	777,163
1980	235,974	841,855
1981	272,782	908,765
1982	194,485	903,839
1983	146,050	856,174
1984	154,901	885,928
Σ	2,375,359	10,339,046
Medias	158,357.26	689,269.73

FUENTE: El Mercado de valores, NAFINSA, Año XLVI,
No. 12, Marzo 24 de 1986.

Se procede a realizar la regresión donde:

Años	Y	X	Y ²	X ²	XY
1970	-57401.26	-244998.73	3,294,904,649.58	60,024,297,701.61	14,063,235,800.39
1971	-62316.26	-226465.73	3,883,316,260.38	51,286,726,864.43	14,112,497,311.76
.
.
.
1984	-4456.26	196658.27	19,858,253.18	38,674,575,159.29	-876,360,382.27
Σ	0	0	33,586,859,906.82	375,458,484,898.79	83,872,511,760.94

$$\hat{\beta}_1 = \frac{\Sigma x y}{\Sigma x^2} = \frac{83,872,511,760.94}{375,458,484,898.79} = 0.2233$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 158,357.26 - (0.2233)(689,269.73) = \underline{4443.33}$$

Procedemos al cálculo de las perturbaciones, con la línea de regresión obtenida por mínimos cuadrados:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$= 4443.33 + (0.2233) x$$

Y $e_t = y - \hat{Y}$, Los residuos de la regresión.

Años	Y	e_t	e_t^2	e_t^{-1}	$(e_t^{-1})^2$	$e_t e^{-1}$
1970	103,649.04	-2693.04	7,252,464.44	-9053.42	81,964,413.69	31,633,686.63
1971	107,787.46	-11746.46	137,979,322.53	-2856.91	8,161,934.74	171,537,901.57
.
.
.
1984	202,271.05	-48370.05	2,339,661,737.0	1206.93	1,456,680.02	2,398,041,001.44
Σ	0	11,950,663,928.37	9,419,793,273.72	6,072,459,690.68	9,611,002,191.37	

$$\Sigma e_t^2 - 1 = 11,950,663,928.37 - 2,339,661,737.0 = \underline{9,611,002,191.37}$$

Una vez obtenidos los resultados de la regresión se procede a diagnosticar la existencia o no de autocorrelación, para lo cual se calcula el estadístico d

$$d = \frac{\frac{\Sigma_{t=1}^n (e_t - e_{t-1})^2}{\Sigma_{t=1}^n e_t^2}}{\frac{\Sigma_{t=2}^n e_t e_{t-1}}{\Sigma_{t=2}^n e_t^2}} = \frac{9,419,793,273.72}{11,950,663,928.37} = \underline{0.78}$$

$$\hat{\rho} = \frac{\frac{\Sigma_{t=2}^n e_t e_{t-1}}{\Sigma_{t=2}^n e_t^2}}{\frac{\Sigma_{t=2}^n e_{t-1}^2}{\Sigma_{t=2}^n e_t^2}} = \frac{6,072,459,690.68}{9,611,002,191.37} = \underline{0.63}$$

$$R^2 = 1 - \frac{\sum e_t^2}{\sum y_t^2} = 1 - \frac{11,950,663,928.37}{33,586,859,906.82} = \underline{0.35}$$

$$\bar{R}^2 = 1 - (1-R^2) \frac{n-1}{n-k} = 1 - (1-0.35) \frac{15-1}{15-2} = 1-0.69 = \underline{0.31}$$

Donde: el R^2 es el coeficiente de determinación múltiple, que nos explica la proporción de la variable Y_t explicada por las variables explicatorias conjuntamente (en este caso por una variable explicatoria), y \bar{R}^2 es el R^2 ajustado (1).

Comparando el valor encontrado de d con los límites establecidos para el caso de una variable explicatoria y $n=15$, se puede apreciar claramente que $d < d_L$ ($d_L = 1.08$), por lo que, según el estadístico, existe correlación serial positiva.

V.1.- Eliminación del Problema.

V.1.1.- Primera transformación de los datos.

En virtud de haber encontrado autocorrelación en los \hat{U}_t (e_t), es necesario hacer la siguiente transformación de los datos originales para eliminar dicha autocorrelación

Años	Y_t	X_t	Y'_t	X'_t
1970	100,956	444,271		
1971	96,041	462,804	32,438.72	182,913.27
.
.
.
1984	153,901	885,928	61,889.50	346,538.37
		Σ	953,226.31	4,284,064.96
		MEDIAS	63,548.42	285,604.33

Los valores de y'_1 y x'_1 , los recuperamos de la siguiente manera:

$$y'_1 = y_t \sqrt{1 - \hat{\rho}^2} = 100,956 (0.776) = \underline{78,341.85}$$

$$x'_1 = x_t \sqrt{1 - \hat{\rho}^2} = 444,271 (0.776) = \underline{344,754.30}$$

El proceso se repite, obteniendo los siguientes resultados:

Años	y	x	y ²	x ²	xy
1970	14,793.43	59,149.97	218,845,571.16	3,498,718,951.0	875,030,940.69
1971	-31,109.70	-102,691.06	967,813,434.09	10,545,453,803.92	3,194,688,069.28
.
.
.
1984	1,658.92	60,934.05	2,752,015.60	3,712,958,449.40	-101,084,714.22
Σ	0	0	12,070,804,732.0	49,011,452,448.48	14,465,391,972.72

$$\hat{\beta}_1 = \underline{0.295} \quad \hat{\beta}_0 = - \underline{20,704.85}$$

Nuestra ecuación de regresión será:

$$\hat{Y} = - 20,704.85 + 0.295 X'_t$$

Años	\hat{Y}	e_t	e_t^2	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	$e_t e_{t-1}$
1970	80,997.66	- 2655.81	7,053,326.75	- -	- -	- -
1971	33,254.56	- 815.84	665,594.90	1,839.97	3,385,489.60	2,166,716.03
.
.
.
1984	81,523.97	-19,634.47	385,512,412.18	20,729.08	429,694,757.70	792,516,911.60
Σ	0	7,800,896,429.85	9,123,935,793.78	3,042,646,607.15		

$$\sum_{t=2}^n e_{t-1}^2 = 7,415,384,017.67$$

$$d = \underline{L. 16} \quad R^2 = \underline{0.64}$$

$$\hat{e} = 0.41 \quad \bar{R}^2 = \underline{0.62}$$

El valor del estadístico d se encuentra en la zona de indecisión donde no podemos concluir, por lo que es necesario nuevamente recurrir a una segunda transformación de los datos originales.

V.1.2.- Segunda transformación de los datos

Años	Y'_t	X'_t	Y''_t	X''_t
1970	78,341.85	344,754.30		
1971	32,438.72	182,913.27	318.57	41,564.00
.
.
.
1984	61,889.50	346,538.38	52,244.47	228,968.65

Σ 580,727.41 2,638,651.09
 MEDIAS 38,715.16 175,910.07

$Y''_t = 71,291.08$; $X''_t = 313,726.41$

Años	Y	X	X^2	Y^2	XY
1970	32,575.92	137,816.34	18,993,343,570.99	1,061,190,563.84	4,489,494,066.53
1971	-38,396.54	-134,346.07	18,048,866,524.44	1,474,298,123.62	5,158,430,967.90
.
.
.
1984	13,529.31	53,058.58	2,815,212,911.61	183,042,229.07	717,845,976.97
Σ	0	0	52,834,927,142.15	11,633,455,768.47	16,301,451,991.86

$\hat{\beta}_1 = 0.308$; $\hat{\beta}_0 = -15,465.14$

La ecuación de regresión será entonces:

$$\hat{Y} = - 15,465.14 + 0.308 X^*_t$$

Años	Y	e_t	e_t^2	e_{t-1}	$(e_t - e_{t-1})^2$	$e_t \cdot e_{t-1}$
1970	81,162.59	-9871.51	97,446,709.68			
1971	2,663.42	2981.99	8,892,264.36	12,853.50	165,212,462.30	- 29,436,744.11
.
.
.
1984	55,057.20	-2812.73	7,911,450.05	13,958.61	194,842,793.10	47,173,251.16
Σ		0	6,603,893,988.66		11,273,752,752.29	914,338,533.0

$$\hat{\Sigma}_{t=2} e_{t-1}^2 = 6,595,982,538.61$$

$$d = \underline{1.70}$$

$$\hat{c} = \underline{0.13}$$

$$R^2 = \underline{0.56}$$

$$\bar{R}^2 = \underline{0.53}$$

El d calculado ahora es mayor que d_u ($d_u = 1.36$) de la tabla, por lo que se afirma que no existe correlación serial. Eliminada ésta los datos pueden ser usados para predicciones, cálculo que en el presente trabajo no se realiza por los aspectos antes mencionados.

Respecto al valor de R^2 podemos decir que, el PIB, a pesar de no ser una variable altamente significativa para explicar el comportamiento de la inversión, sí tiene influencia sobre ella.

VI.- CONCLUSIONES

Siguiendo el modelo del estadístico Durbin-Watson, encontramos que en el problema existe autocorrelación positiva, dado que el d calculado (0.78) es menor que el d_L (1.08); por lo tanto, rechazamos H_0 .

Una vez detectada la presencia de autocorrelación se procedió, enseguida, a realizar la transformación correspondiente con la finalidad de eliminarla, lo que se consigue con la segunda transformación de los datos.

VII. BIBLIOGRAFIA

- 1.- Gujarati, D. (1978), *Econometría Básica*. Edit. Mc Graw-Hill Latinoamericana, S.A., Colombia. pp. 33 - 37, 100 - 105, 215 - 238.
- 2.- Johnston, J., (1970). *Métodos de Econometría*, Vices-Vives, España. pp. 197 - 199.
- 3.- Maddala, G.S., (1985). *Econometría*, Edit. Mc Graw - Hill, México. pp. 76 - 92, 112 - 114, 124 - 126, 287-305.
- 4.- Merrill, W., Fox, K., (1972), *Introducción a la Estadística Económica*, Amorrortu Editores, Argentina. pp. - 432 - 439.

