



NOTA TÉCNICA
N.º 037 | 2003

Principales indicadores para el diagnóstico del análisis de regresión lineal

Ana Cecilia Kikut V.
Bernal Laverde M.
Jorge León M.
Evelyn Muñoz S.
Juan Carlos Quirós S.
Carlos Torres G.

Fotografía de portada: "Presentes", conjunto escultórico en bronce, año 1983, del artista costarricense Fernando Calvo Sánchez. Colección del Banco Central de Costa Rica.

Principales indicadores para el diagnóstico del análisis de regresión lineal

Ana Cecilia Kikut V^{*}, Bernal Laverde M[†], Jorge León M[‡], Evelyn Muñoz S[§], Juan Carlos Quirós S^{**}, Carlos Torres G^{††}

Las ideas expresadas en este documento son de los autores y no necesariamente representan las del Banco Central de Costa Rica.

Resumen

En este documento se recopilan los principales indicadores econométricos que deben tomarse en consideración al efectuar un diagnóstico del análisis de regresión lineal por mínimos cuadrados ordinarios. El objetivo de esta recopilación es proveer a los investigadores de un material de apoyo conciso y completo, que puedan tener a mano para efectuar y evaluar este procedimiento en una forma rigurosa y más ágil.

Palabras clave: Análisis de regresión lineal, Modelación económica.

Clasificación JEL: C10, C12, C52.

* Departamento de Investigación Económica. División Económica, BCCR. kikutva@bccr.fi.cr

† Departamento de Investigación Económica. División Económica, BCCR. laverdema@bccr.fi.cr

‡ Departamento de Investigación Económica. División Económica, BCCR. leonmj@bccr.fi.cr

§ Departamento de Investigación Económica. División Económica, BCCR. munozse@bccr.fi.cr

** Departamento de Investigación Económica. División Económica, BCCR. quirossi@bccr.fi.cr

†† Departamento de Investigación Económica. División Económica, BCCR. torresgc@bccr.fi.cr

Main Indicators for Linear Regression Analysis

Ana Cecilia Kikut V^{‡‡}, Bernal Laverde M^{§§}, Jorge León M^{***}, Evelyn Muñoz S^{†††},
Juan Carlos Quirós S^{‡‡‡}, Carlos Torres G^{§§§}

The ideas expressed in this paper are those of the authors and not necessarily represent the view of the Central Bank of Costa Rica.

Key words: Linear regression analysis, Economic modeling.

JEL codes: C10, C12, C52.

^{‡‡} Department of Economic Research. Email address kikutva@bccr.fi.cr

^{§§} Department of Economic Research. Email address laverdema@bccr.fi.cr

^{***} Department of Economic Research. Email address leonmj@bccr.fi.cr

^{†††} Department of Economic Research. Email address munozse@bccr.fi.cr

^{‡‡‡} Department of Economic Research. Email address quirossj@bccr.fi.cr

^{§§§} Department of Economic Research. Email address torresgc@bccr.fi.cr

Presentación

En este documento se recopilan los principales indicadores econométricos que deben tomarse en consideración al efectuar un diagnóstico del análisis de regresión lineal por mínimos cuadrados ordinarios. El objetivo de esta recopilación es proveer a los investigadores de un material de apoyo conciso y completo, que puedan tener a mano para efectuar y evaluar este procedimiento en una forma rigurosa y más ágil.

El formato empleado facilita su consulta y permitirá ir adicionando nuevos temas, conforme sea necesario. Asimismo, queda abierta la posibilidad de agregar indicadores adicionales en cada uno de los apartados, lo que posibilita cubrir una gama mayor de herramientas

Se parte del Análisis de regresión estándar, detallando los supuestos que hay detrás, y los errores que pueden ocurrir si no se cumple cada uno de ellos. De igual forma se señala cómo detectar estos problemas y posibles formas de corregirlos. Se indica además cómo emplear en cada caso el paquete econométrico EViews en las modalidades de ventanas y línea de comandos. Los aspectos que se analizan se clasificaron en dos partes: la primera contempla aquellos supuestos relacionados con la parte sistemática y la segunda los relacionados con la parte aleatoria del análisis de regresión, como lo establece la siguiente especificación general:

$$Y_t = \underbrace{\alpha + \beta X_t}_{\text{Parte sistemática}} + \underbrace{\mu_t}_{\text{Parte aleatoria}}$$

Este material será empleado en la primera parte del taller de EViews que está diseñando el equipo de trabajo de Desarrollo y Análisis Metodológico. Se considera que un complemento de este informe técnico consiste en la documentación de los programas econométricos que efectúan estos procedimientos, los cuales se presentarán en otro informe.

Tabla de contenido

ANÁLISIS DE REGRESIÓN ESTÁNDAR	1
A. ANÁLISIS DE LA PARTE SISTEMÁTICA DE UN MODELO	4
I. MULTICOLINEALIDAD	4
II. ESPECIFICACIÓN	7
III. ESTABILIDAD ESTRUCTURAL	8
B. ANÁLISIS DE LA PARTE ALEATORIA DE UN MODELO	10
I. AUTOCORRELACIÓN	10
II. HETEROCEDASTICIDAD	12

ANÁLISIS DE REGRESIÓN ESTÁNDAR

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	COMANDOS (E-Views)
$E(X_i U_j) = 0$ $\forall i,$ X ' S independientes del comportamiento anatorio X causa Y	Puede darse una violación de los supuestos, y por ende deben ser corregidos. Sin embargo, dependiendo del objetivo de la estimación, algunas violaciones de los supuestos podrían no requerir corrección.	<ul style="list-style-type: none"> • Signos de los Coeficientes 	Deben ser los esperados antes de realizar el cálculo	En la línea de comandos: LS Y C X₁ X₂ Estimación del modelo por ventanas: Se seleccionan las series que intervienen como un objeto ecuación
		<ul style="list-style-type: none"> • Errores Estándar (Std. Error) • Error Estándar de la regresión (S.E of regression) • Suma de los Errores al Cuadrado (Sum squared resid) 	Sirven para analizar la capacidad explicativa del modelo (minimizar)	
		<ul style="list-style-type: none"> • T-statistic: Prueba si la variable es significativa. • Probabilidad: indica la probabilidad de cometer el error de rechazar la hipótesis nula siendo cierta (error de tipo I). 	$H_0 : \beta_i = 0$ $H_1 : \beta_i \neq 0$ Si $t_c > t_i \Rightarrow SRH_0$ $P - valor < 0.05$	Procs / Make Equation O bien: Quick / Estimate Equation

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	COMANDOS (E-Views)
		<ul style="list-style-type: none"> • R^2 es un indicador de la bondad del ajuste del modelo. • R^2 ajustado: Se obtiene a partir del R^2, y se pondera por los grados de libertad. • Durbin-Watson Stat: Indicador de Autocorrelación Serial de Primer orden en los residuos $d = 2(1 - \rho)$ 	<p>R^2 elevado, nos explica la variabilidad de la variable endógena.</p> <p>R^2 Ajustado elevado. Permite comparar la capacidad explicativa de modelos para una misma variable dependiente con diferente número de variables explicativas.</p> <p>Toma valores entre 0 y 4, <u>alrededor de 2</u> No hay correlación serial <u>Cercano a 0</u> Autocorrelación positiva <u>Cercano a 4</u> Autocorrelación negativa</p>	

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	COMANDOS (E-Views)
		<ul style="list-style-type: none"> • Akaike info Criterion • Schwarz criterion Permiten analizar capacidad predictiva y realizar la comparación entre modelos anidados.	A menor valor el modelo es mejor.	
		<ul style="list-style-type: none"> • F-statistic permite contrastar la capacidad explicativa conjunta de las variables introducidas en el modelo. • Prob (F-statistic): Probabilidad de cometer el error de tipo I. 	Elevado $P - valor < 0.05$	

A. ANÁLISIS DE LA PARTE SISTEMÁTICA DE UN MODELO

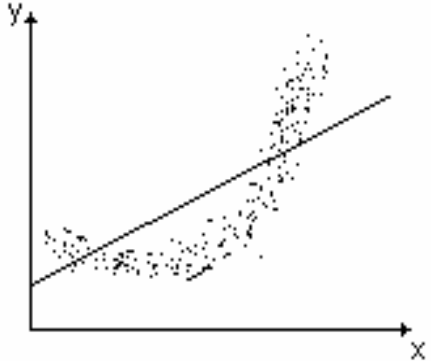
I. MULTICOLINEALIDAD: Cuando los regresores incluidos en un modelo econométrico se encuentran interrelacionados.

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	CORRECCIÓN	COMANDOS (E-Views)
<p>Las variables explicativas deben ser linealmente independientes.</p> <p>$H_0 : E(x_i y_j) = 0, \forall i, j$</p> <p>La multicolinealidad perfecta se da cuando existe una relación exacta entre varios de los regresores del modelo. En este caso de matriz de regresores es singular (no tiene inversa) y no pueden determinarse los parámetros del modelo.</p>	<ul style="list-style-type: none"> • Existe dificultad para conocer el aporte a la explicación de la variable dependiente de cada una de las variables explicativas del modelo • La varianza de los estimadores se encuentra aumentada, lo que implica el rechazo de la significancia individual de los regresores que sí contribuyen a la explicación del modelo. • Los límites de confianza son más amplios • Es un problema de tipo muestral tal vez no tan malo si el 	<p>Modelo globalmente bien estimado (R^2 alto y F significativa) pero todos o algunos regresores individuales no significativos (t - student bajos) en el modelo.</p>	Se rechaza H_0	<ul style="list-style-type: none"> • Suprimir variables cuando sean redundantes y su efecto sea capturado dentro de otra(s) variables(s) del modelo. (Se corre el riesgo de introducir sesgo de especificación del modelo) • Uso de información adicional (ampliar la muestra) si es posible • Usar primeras diferencias de las series (podría causar autocorrelación en los residuos) • Usar razones de las variables, dividiendo todas 	<p>Para identificar si una serie es generadora de multicolinealidad se analiza su capacidad predictiva individual en un modelo de regresión simple en que se use cada regresor como única variable explicativa. Si pasa a ser significativa, es indicio de que genera multicolinealidad.</p>
		<p>Pequeños cambios en los datos pueden producir grandes variaciones en los estimadores de los parámetros.</p>			
		<p>Determinantes de la matriz de correlaciones entre regresores cercano a 0. Si los regresores fueran ortogonales, el</p>	Se rechaza H_0		

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	CORRECCIÓN	COMANDOS (E-Views)
	objetivo es la predicción puesto que no conlleva al incumplimiento de ninguna de las hipótesis en las que se basa el modelo lineal clásico.	determinante tomaría el valor de 1.		entre un factor de escala común.	<u>Comandos</u> sym mcorrel=@cor(grupo de var) Scalar detmcorrel=@det(correl)
		Presencia de signos de coeficientes estimados contrarios a los esperados o de una magnitud poco creíble.	Se rechaza H_0		
		Contraste de multicolinealidad de Farrar-Glauber $G \rightarrow \chi_n^2$ donde $n=k*(k-1)/2$	$H_0 = R_{xx} = 1$ Si $G_{calc} > G_{tab}$ se rechaza H_0		
		Factor de Inflación de la Variancia (FIV) y/o Indices de Tolerancia. $FIV = \frac{1}{(1 - R_{x_i}^2)}$	Cuanto mayor sea FIV mayor es el grado de multicolinealidad de la variable en		Se calculan a partir de regresiones en las que cada variable explicativa es función del resto

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	CORRECCIÓN	COMANDOS (E-Views)
		$IT = (1 - R_{X_i}^2)$	<p>cuestión con alguna(s) de las otras variables en el modelo. En la práctica un FIV mayor a 10 se considera problemático, aunque un FIV alto puede encontrarse por una variancia pequeña y/o un $\sum_{i=1}^n X_i^2$ alto.</p>		de los regresores.

II. **ESPECIFICACIÓN:** Se refiere a la forma en que está formulado el modelo.

SUPUESTO	PROBLEMA (SI NO SE CUMPLE)	INDICADORES	DECISIÓN	SOLUCIÓN AL PROBLEMA	COMANDOS EIEWS
<p>La Especificación del Modelo:</p> <p>El supuesto es que se conoce la especificación correcta del modelo de regresión.</p> $y = X\beta + \varepsilon$ <p>Tipos de errores que pueden cometerse en la especificación de la ecuación estimada:</p> <ol style="list-style-type: none"> 1. Omisión de variables relevantes. 2. Inclusión de Variables superfluas. 3. Mala Especificación. 	<p>Mala especificación:</p> <p>H_0: mala especificación</p> <p>Bajo esta mala especificación los mínimos cuadrados ordinarios, serán inconsistentes, por lo que las inferencias no serán validas.</p>	<p>Ramsey RESET (Regression Specification Error Test)</p> $y = X\beta + \varepsilon$ $H_0 : \varepsilon \rightarrow N(0, \sigma^2 I)$ $H_1 : \varepsilon \rightarrow N(\mu, \sigma^2 I)$ $\mu \neq 0$ <p>El test se basa en una regresión aumentada:</p> $y = X\beta + Z\gamma + \varepsilon$ <p>Donde se prueba que $\delta = 0$, y Z es una matriz de variables no incluidas o las variables X elevadas a algún exponente.</p>	<p>La salida de Eviews muestra:</p> <p>i) La probabilidad de la F estadística.</p> <p>ii) La probabilidad de log de máxima verosimilitud.</p>	<p>Si se comprueba que existe el problema. Se pueden probar otras especificaciones para el modelo (analizando los datos o revisando la teoría).</p>	<p><u>Windows:</u></p> <p>View/Stability Test/Ramsey RESET</p> <p><u>Comando:</u></p> <p>reset(n, options)</p> <p><u>Ecuación:</u></p> <p>eq_name.reset(opt)</p> <p>NOTA: este test solo sirve para MCO.</p>

III. **ESTABILIDAD ESTRUCTURAL:** Analiza la presencia de cambios en la relación que vincula las variables del modelo a lo largo del periodo muestral.

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES		DECISIÓN	CORRECCIÓN	COMANDOS (E-Views)
Los parámetros del modelo son estables durante todo el período muestral.	Disminuye la bondad del ajuste y el poder de pronóstico.	<i>Punto de Quiebre de Chow:</i> Divide la muestra en grupos y estima el modelo para cada uno, comparando las ecuaciones de cada submuestra. Es necesario indicar los puntos dónde se divide la muestra.		H ₀ = El modelo es estable. Se rechaza H ₀ si la probabilidad de los estadísticos F y Log likelihood es menor que 0.05	Incorporar el cambio estructural en el modelo mediante variables dummy (D). Puede ser una dummy para todo el modelo: $y = \alpha + \beta_1 x + \beta_2 D$ O sólo para la variable que causa el cambio estructural: $y = \alpha + \beta_1 x + \beta_2 Dx$	<i>Eq-name.chow</i> <i>obs1 obs2 ...</i> Obs1 obs2 = puntos de quiebre definidos.
		<i>Pronóstico de Chow:</i> Compara los errores del modelo completo con los del modelo con el primer subgrupo (más grande). Se usa cuando el segundo subgrupo es muy pequeño para correr una regresión.				<i>Eq-name.chow(f)</i> <i>obs1</i> Obs1 = único punto de quiebre. (f) = opción para prueba de pronóstico
		<i>Estimación Recursiva:</i> Esta técnica es adecuada para series de tiempo y cuando no	<i>Coefficientes Recursivos:</i> grafica la evolución de cada coeficiente al ir agregando observaciones	Hay estabilidad si los coeficientes no muestran grandes cambios al ir variando la muestra	Si el número de datos lo permite estimar un modelo para cada submuestra.	<i>Eq-name.rls(c)</i> <i>c(1)</i> <i>c(2) ...</i> (c) = opción para coeficientes recursivos <i>c(1) c(2)</i> = coeficientes a graficar

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	CORRECCIÓN	COMANDOS (E-Views)
		se conoce el punto de quiebre. Sólo sirve para MCO.			
		<i>Residuos Recursivos:</i> son los errores de predicción un período hacia delante calculados en cada etapa de la estimación recursiva. Son útiles cuando el modelo no contiene variables dummy.	Hay estabilidad si los residuos se mantienen dentro de las bandas de confianza (de 2 desviaciones estándar).		<i>Eq-name.rls(r)</i> (r) = opción para residuos recursivos
		Cusum: Se construye a partir de la suma acumulada de los residuos recursivos. Son útiles cuando el modelo no contiene variables dummy.	Hay estabilidad (al 95% de significancia) si el estadístico se mantiene dentro de las bandas de confianza.		<i>Eq-name.rls(q)</i> (q) = opción para CUSUM
		Cusum Q: utiliza la suma acumulada de los residuos al cuadrado			<i>Eq-name.rls(v)</i> (r) = opción para CUSUM Q

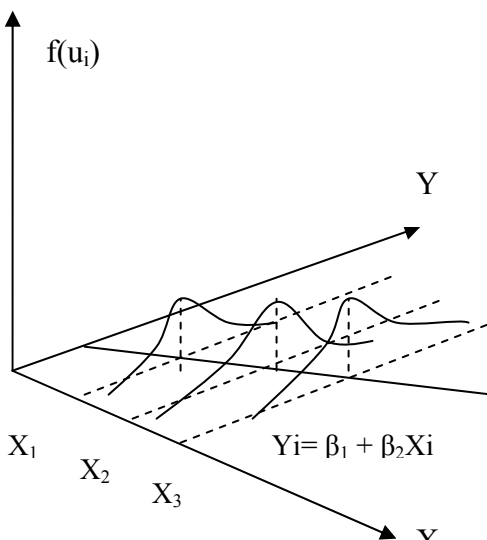
ANÁLISIS DE LA PARTE ALEATORIA DE UN MODELO

I. AUTOCORRELACIÓN: Se presenta cuando los errores del modelo se encuentran correlacionados.

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	CORRECCIÓN	COMANDOS (E-Views)
$E(\mu_i \mu_j) = 0$ no hay autocorrelación en los residuos. AR $u_i = \rho \mu_{i-1} + e_i$ MA $u_i = \phi e_{i-1} + e_i$	H ₀ : existe autocorrelación <u>Consecuencias:</u> los estimadores β son lineales, insesgados y consistentes pero no eficientes (variancia mínima), dejan de ser MELI. Al existir la posibilidad de que la variancia estimada subestime la verdadera variancia, existe la posibilidad de que se sobreestime el R ² y que las pruebas t y F dejan de ser válidas, si se aplican es probable que conduzcan a conclusiones erróneas sobre la significancia estadística de los estimadores.	<ul style="list-style-type: none"> DW (Durbin Watson): esta prueba permite detectar autocorrelación de primer orden cuando la variable dependiente rezagada no se encuentra dentro de los regresores del modelo, en cuyo caso debe recurrirse al Durbin-H, el cual debe programarse en Eviews. <div style="text-align: center;"> </div>	DW < Linf: NSR H ₀ DW > Lsup: SR H ₀ Linf < DW < Lsup: no hay conclusión.	Incluir la variable dependiente rezagada un periodo entre las variables explicativas o agregar un término autorregresivo o de medias móviles.	El cálculo del estadístico DW se encuentra en la salida de la estimación de MCO. LS Y C X1 X2
		<ul style="list-style-type: none"> Correlogramas simple y parcial 	Si al menos una barra (RRR) se sale de los límites (+) → NSRH ₀		Con la barra de herramientas, una vez 'abierta' la serie resid: VIEW/CORRELATION (1)

SUPUESTO	PROBLEMA (si no se cumple)	INDICADORES	DECISIÓN	CORRECCIÓN	COMANDOS (E-Views)
		<ul style="list-style-type: none"> • Q (Ljung-Box-Pierce) Se distribuye como una Chi-cuadrado. 	P-value > 5% SRH ₀ P-value < 5% NSRH ₀ ($\chi^2_c > \chi^2_t$ NSRH ₀)		Eviews lo calcula junto con los correlogramas simple y parcial.
		<ul style="list-style-type: none"> • Runs test (Prueba de las corridas) ~z 	$ z_c > z_t $ NSRH ₀ \uparrow 1.96		
		<ul style="list-style-type: none"> • Gráfico de los residuos 	Se espera que no muestren un comportamiento sistemático, aunque no son un instrumento definitivo para detectar autocorrelación.		En la barra de herramientas una vez 'abierta' la serie resid: VIEW/ACTUAL, FITTED, RESIDUAL/ACTUAL FITTED,RESIDUAL, GRAPH

II. HETEROCEDASTICIDAD: los componentes del vector de errores no tienen igual variancia.

SUPUESTO	PROBLEMA (SI NO SE CUMPLE)	INDICADORES	DECISIÓN	SOLUCIÓN AL PROBLEMA	COMANDOS EIEWS
<p>Las varianzas de los errores de estimación (u_i), condicionales a los valores de las variables explicativas (X_i), son idénticas (homocedásticas):</p> $\text{var}(u_i X_i) = E(u_i^2 X_i) = \sigma^2, \forall i = 1, 2, \dots, n$ 	<p>Mala estimación de la matriz varianzas-covarianzas de los errores mínimo cuadráticos.</p> <p>Los coeficientes de regresión estimados siguen siendo lineales e insesgados.</p> <p>Disminución de la eficiencia del estimador mínimo cuadrático. Éste deja de ser el de mínima varianza entre todos los estimadores lineales e insesgados.</p> <p>No necesariamente es obligatorio corregir por heterocedasticidad,</p>	<p>Prueba de White</p> <p>Ho: $\sigma_i^2 = \sigma^2$ para todo i H₁: no se verifica Ho</p> <p>El estadístico para realizar la prueba (ya sea con términos cruzados o sin ellos) es $\lambda = NR^2$, donde R^2 es el coeficiente de determinación de la regresión auxiliar (con o sin términos cruzados). Y N es el número de datos.</p> <p>Bajo Ho, dicho estadístico se distribuye asintóticamente</p>	<p>Si el valor de la probabilidad asociado al estadístico reportado en la prueba ($\lambda = NR^2$) es menor al 5%, rechazamos Ho (homocedasticidad) y concluimos que el modelo tiene problemas de heterocedasticidad.</p> <p>En caso contrario (si la probabilidad es superior a ese %), no rechazamos Ho y concluimos que no hay heterocedasticidad.</p>	<p>Una solución empírica simple, basada en la transformación Box-Cox, es reestimar el modelo original en logaritmos, para suavizar la dispersión de los valores originales.</p> <p>Otra solución es aplicar Mínimos Cuadrados Generalizados (MCG o ponderados), transformando el modelo original al dividir todas las observaciones de las variables por la desviación típica de los errores.</p>	<p>Para efectuar la prueba de White sobre la regresión auxiliar, con productos cruzados 2 a 2, pulsamos</p> <p>View/Residual TEST/White Heteroskedasticity (cross terms).</p> <p>Para correr la prueba sin términos cruzados pulsamos</p> <p>View/Residual TEST/White Heteroskedasticity (no cross terms).</p> <p>Para solucionar por MCG, conociendo la matriz omega, pulsamos secuencialmente Procs/Make</p>

<i>SUPUESTO</i>	<i>PROBLEMA (SI NO SE CUMPLE)</i>	<i>INDICADORES</i>	<i>DECISIÓN</i>	<i>SOLUCIÓN AL PROBLEMA</i>	<i>COMANDOS EViews</i>
	pero si queremos hacer inferencia estadística sí debemos corregirla	como $\chi^2(p)$, donde p es el número de variables incluidas en la regresión auxiliar, exceptuando el término independiente.		Un tratamiento más avanzado de la heterocedasticidad es el uso de modelos ARCH y GARCH.	Equation, Options, Weighted LS/TLS y en la casilla Weight especificamos la variable de ponderación y pulsamos OK

Referencias

- Carrascal, Ursicino, González Yolanda y Rodríguez Beatriz. (2001). "Análisis Econométrico con EVIEWS". Alfaomega Grupo Editor S.A. México D.F.
- Fernández Viviana. (2000). "Material de repaso:teoría Econométrica I (EAE-350B)" manuscrito, Instituto de Economía, Pontificia Universidad Católica de Chile.
- Green, William (1998). "Análisis Econométrico". Prentice Hall, Tercera Edición.
- Gujarati, Damodar. (1997) "Econometría". Mc Grae Hill, Tercera Edición.
- Kikut Croceri, Otto. (1997). "Análisis de regresión múltiple utilizando EViews 2.0". Consejo Monetario Centroamericano.
- Pena, Bernardo; Estavillo, Julio; Galindo, María Ester; Receta, María José; Zamora, María del Mar. "Cien ejercicios de econometría"