Más Allá del Modelo de Regresión Lineal

Dante A. Urbina

CONTENIDOS

1. Modelos de Regresión No Lineales
2. Modelos de Respuesta Cualitativa
3. Datos de Panel
4. Modelos Autorregresivos y de Rezagos
5. Modelos de Ecuaciones Simultáneas
6. Series de Tiempo
7. Modelos AR, MA, ARMA y ARIMA

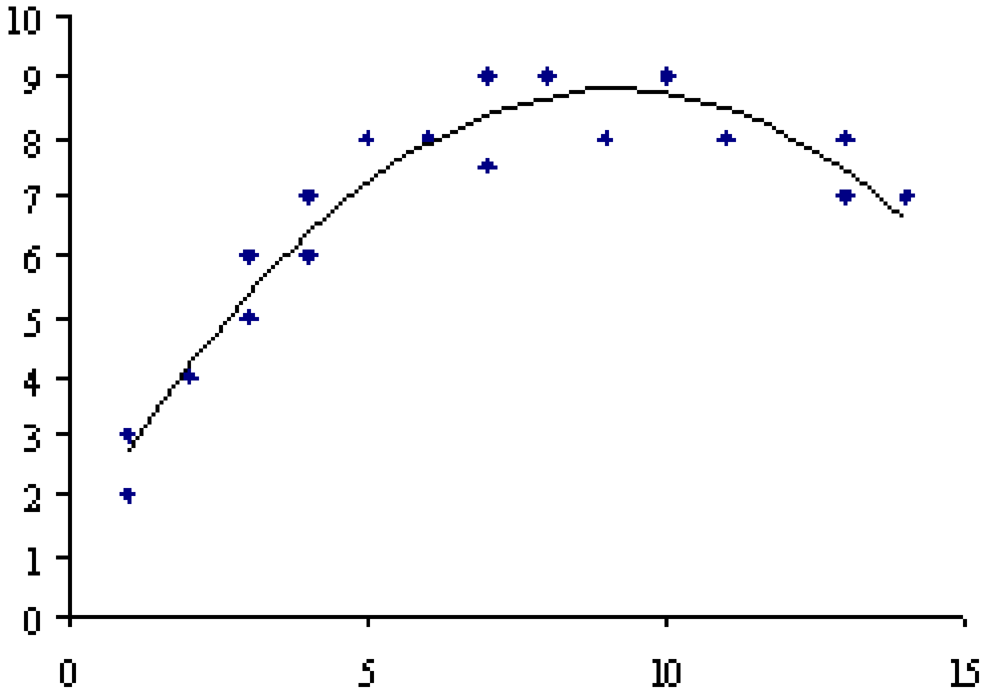
MODELOS DE REGRESIÓN NO LINEALES

Definición

Son todos aquellos modelos en que no se cumple el supuestos de linealidad de parámetros. Por ejemplo:

. $Y_i = \alpha + \beta e^{X_i} + \varepsilon_i$ es un modelo lineal.

. $Y_i = \alpha e^{\beta X_i} + \varepsilon_i$ es un modelo no lineal.

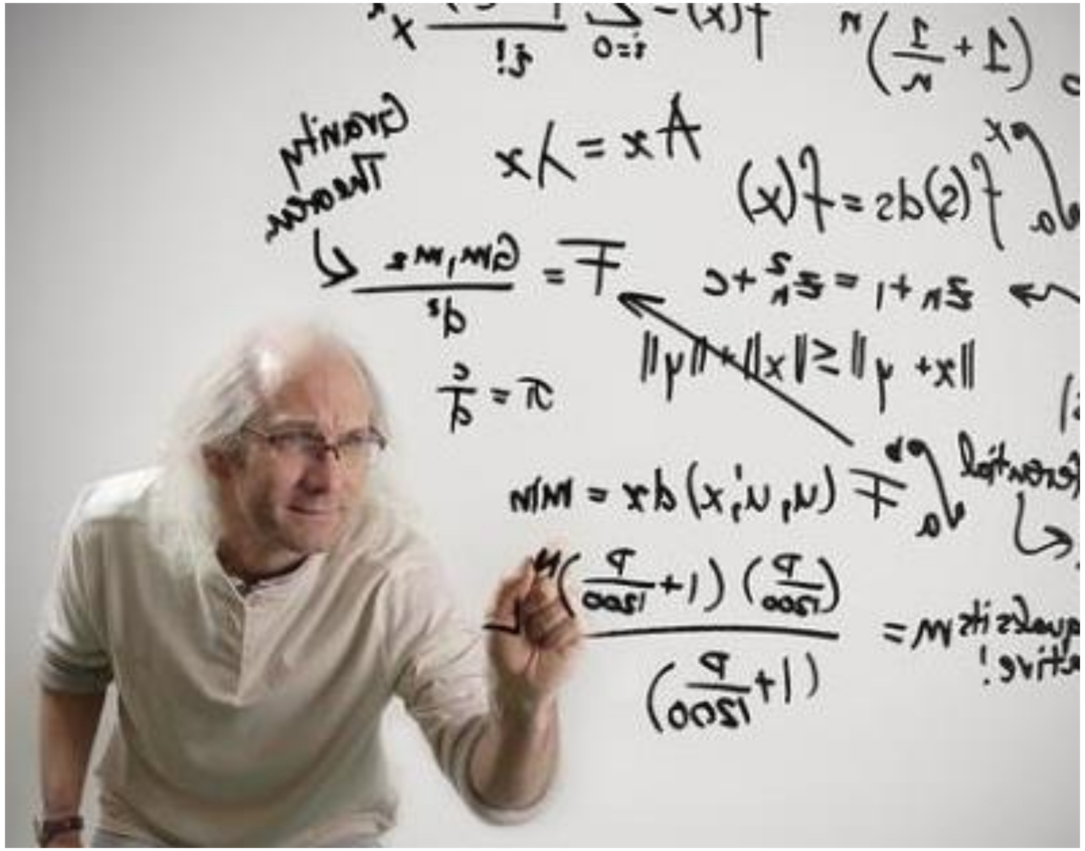


Métodos de estimación

1. Método de prueba y error: Se prueban distintos valores para los parámetros buscando una cada vez menor suma de cuadrados de los errores.

2. Optimización directa: Se aplica un método de iteración más sistemático que el anterior.

3. Linealización iterativa: Se linealiza la ecuación en torno a ciertos valores iniciales de los parámetros y se estima por MCO, para luego hacer lo mismo hasta que no haya diferencia significativa entre los parámetros estimados respecto de los últimos para la iteración.



MODELOS DE RESPUESTA CUALITATIVA

Definición

Son aquellos modelos en los que la variable dependiente es de tipo cualitativo tomando valores numéricos que indican cierta característica. Cuando esta variable dependiente solo toma dos valores (por lo general 0 y 1) se le llama binaria o dicotómica.

En este tipo de modelos lo que se busca es hallar la probabilidad de que cierta característica aparezca, por lo que también se les conoce como modelos de probabilidad.

Modelo Logit

Presenta una función de distribución logística acumulativa para el modelo de regresión planteado:

$$P_i = \frac{1}{1 - e^{-(\alpha+\beta X_i)}} = \frac{e^{(\alpha+\beta X_i)}}{1 + e^{(\alpha+\beta X_i)}}$$

Donde P_i es la probabilidad de que la característica aparezca o el evento ocurra.
Haciendo transformaciones se llega a la expresión:

$$L_i = Ln\frac{P_i}{1 - P_i} = \alpha + \beta X_i$$

Dependent Variable: EPIDURAL
Method: ML - Binary Logit (Quadratic hill climbing)
Sample: 1 130
Included observations: 130
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-7.832855	4.655343	-1.682552	0.0925
AGE	-0.092133	0.042852	-2.150040	0.0316
BABYWGHT	-0.049414	0.190905	-0.258840	0.7958
MW	0.017350	0.010915	1.589540	0.1119
TERM	0.198494	0.126788	1.565557	0.1175

McFadden R-squared	0.062391	Mean dependent var	0.507692
S.D. dependent var	0.501875	S.E. of regression	0.488927
Akaike info criterion	1.376504	Sum squared resid	29.88115
Schwarz criterion	1.486793	Log likelihood	-84.47274
Hannan-Quinn criter.	1.421318	Deviance	168.9455
Restr. deviance	180.1875	Restr. log likelihood	-90.09375
LR statistic	11.24202	Avg. log likelihood	-0.649790
Prob(LR statistic)	0.023975		
Obs with Dep=0	64	Total obs	130
Obs with Dep=1	66		

Modelo Probit

Presenta una función de distribución normal acumulativa para el modelo de regresión planteado:

$$F(P_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+\beta X_i} e^{-z^2/2} dz$$

Donde P_i es la probabilidad de que la característica aparezca o el evento ocurra.

Haciendo transformaciones se llega a la expresión:

$$F^{-1}(P_i) = \alpha + \beta X_i$$

Dependent Variable: EPIDURAL
Method: ML - Binary Probit (Quadratic hill climbing)
Sample: 1 130
Included observations: 130
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-4.882704	2.829543	-1.725616	0.0844
AGE	-0.057577	0.026366	-2.183741	0.0290
BABYWGHT	-0.032537	0.118232	-0.275199	0.7832
MW	0.010644	0.006588	1.615683	0.1062
TERM	0.124892	0.076874	1.624632	0.1042

McFadden R-squared	0.062851	Mean dependent var	0.507692
S.D. dependent var	0.501875	S.E. of regression	0.488847
Akaike info criterion	1.375865	Sum squared resid	29.87148
Schwarz criterion	1.486155	Log likelihood	-84.43123
Hannan-Quinn criter.	1.420680	Deviance	168.8625
Restr. deviance	180.1875	Restr. log likelihood	-90.09375
LR statistic	11.32503	Avg. log likelihood	-0.649471
Prob(LR statistic)	0.023144		

DATOS DE PANEL

Definición

En los modelos de regresión con datos de panel se estudia la misma unidad de corte transversal a lo largo del tiempo, es decir, se toman en cuenta el tiempo y el espacio.

La principal ventaja de utilizar este tipo de modelos es que se toma explícitamente en cuenta la heterogeneidad de las unidades de análisis.

Country	Year	Ease_N	BusFree_N	PropRight_N
1	2005	,70	,70	,30
1	2006	,70	,63	,30
1	2007	,70	,64	,30
1	2008	,71	,64	,30
1	2009	,69	,62	,20
1	2010	,69	,62	,20
1	2011	,69	,62	,20
2	2005	,98	,85	,90
2	2006	,98	,91	,90
2	2007	,98	,89	,90
2	2008	,98	,90	,90
2	2009	,98	,91	,90
2	2010	,98	,90	,90
2	2011	,98	,90	,90

Tipos

1. Panel corto: El número de sujetos de corte transversal es mayor que el número de períodos.

2. Panel largo: El número de sujetos de corte transversal es menor que el número de períodos.

3. Panel balanceado: Cada sujeto de corte transversal tiene el mismo número de observaciones (períodos).

Técnicas de estimación

1. Modelo de MCO agrupados: Se agrupan todas las observaciones de datos de panel y se estima una sola regresión sin discriminar el contexto de los datos. La principal desventaja de esto es que se introducen distorsiones por soslayar la heterogeneidad.

2. Modelo de mínimos cuadrados con variable dicótoma de efectos fijos: Se agrupa el total de observaciones pero se permite que cada unidad de corte transversal tenga su propia variable dicótoma (intercepto) fija.

3. Modelo de efectos fijos dentro del grupo: Se expresa cada variable como una desviación de su valor medio y luego se estima una regresión por MCO esos valores.

4. Modelo de efectos aleatorios: Se supone que los valores del intercepto son una extracción aleatoria de una población mucho mayor de variables explicativas.

MODELOS AUTORREGRESIVOS Y DE REZAGOS

Tipos (1)

1. Modelo de rezagos distribuidos de Koyck: Parte del modelo de rezagos distribuidos infinito:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \varepsilon_i$$

Luego, sea λ la tasa de decaimiento del parámetro distribuido, tal que: $\beta_k = \beta_0 . \lambda^k$, obtenemos el modelo transformado de Koyck:

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + e_i$$

2. Modelo de expectativas adaptativas: Se parte del modelo:

$$Y_t = \beta_0 + \beta_1 X_t^* + \varepsilon_t$$

Donde X_t^* es una variable económica esperada a largo plazo. Luego, sea la hipótesis de expectativas adaptativas: $X_t^* - X_{t-1}^* = \gamma(X_t - X_{t-1}^*)$, se llega a la ecuación:

$$Y_t = \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + e_t$$

Tipos (2)

3. Modelo de ajuste parcial: Se parte del modelo:

$$Y_t^* = \beta_0 + \beta_1 X_t + \varepsilon_t$$

Donde Y_t^* es una variable económica esperada a largo plazo. Luego, sea la hipótesis de ajuste parcial: $Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1})$, se llega a la ecuación:

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + e_t$$

4. Modelo de rezagos distribuidos de Almon: Parte del modelo de rezagos distribuidos finito:

$$Y_t = \alpha + \Sigma_{i=1}^{k} \beta_i X_{t-i} + \varepsilon_t$$

Donde $\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \cdots + \alpha_m i^m$. Y luego se obtiene el modelo transformado de Almon:

$$Y_t = \alpha + \alpha_0 Z_{0t} + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \cdots + \alpha_m Z_{mt} + \varepsilon_t$$

MODELOS DE ECUACIONES SIMULTÁNEAS

Forma estructural y reducida

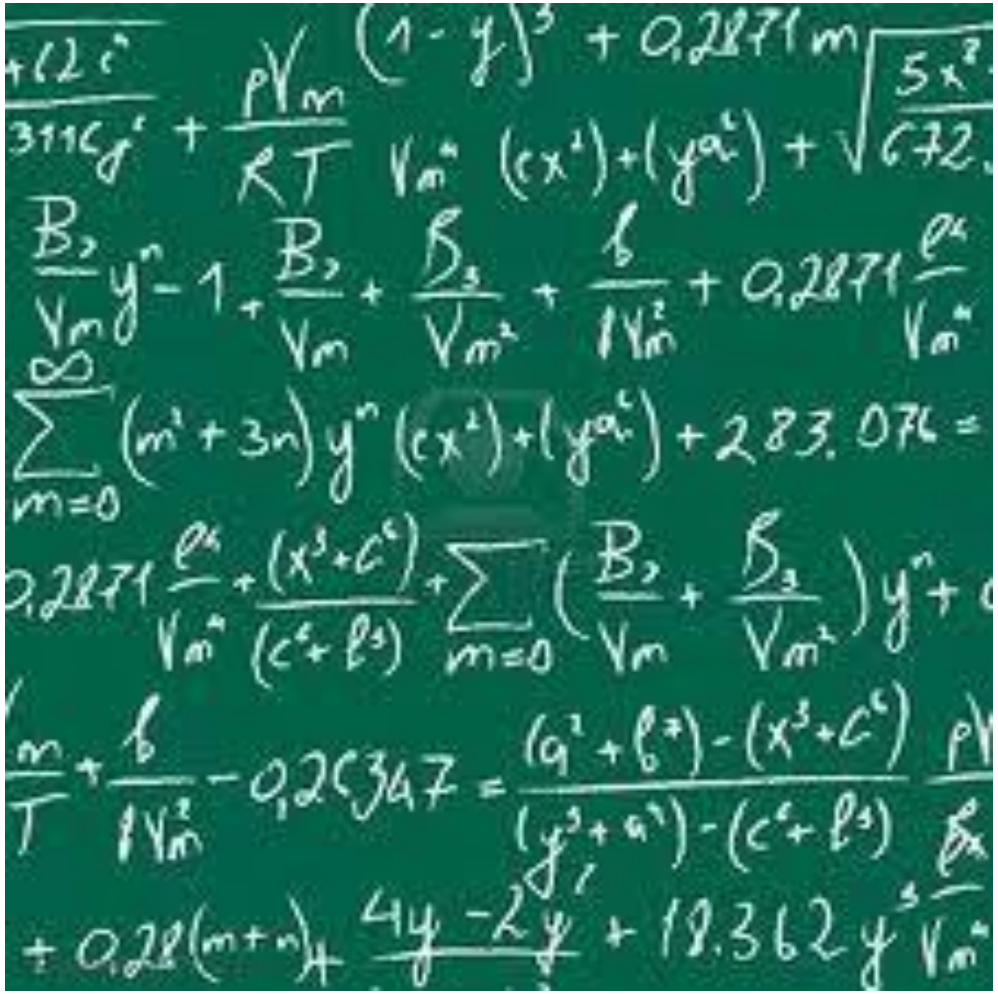
Considerando N variables endógenas y k predeterminadas, la forma estructural correspondiente a un modelo de ecuaciones simultaneas puede escribirse como:

$$y_t^T \Gamma + x_t^T B + \varepsilon_t^T = 0$$

La forma reducida consiste en poner las variables endógenas en función de las predeterminadas llegándose a una expresión de la forma:

$$y_t^T \Gamma = x_t^T \Pi + v_t^T$$

Si se cumplen ciertas restricciones, a partir de la estimación de los parámetros de la forma reducida será posible estimar los parámetros de la forma estructural.



Enfoques para la estimación

1. Enfoque directo: Cada ecuación del modelo se estima como si estuviera aislada sin considerar el resto de ecuaciones del modelo y sin distinguir entre variables endógenas y predeterminadas. Por tanto, el método de estimación idóneo es el de Mínimos Cuadrados Ordinarios (MCO).

2. Enfoque con información incompleta: Se distingue entre variables endógenas y predeterminadas pero, al igual que en el método anterior, las ecuaciones se estiman de manera individual. Bajo este enfoque se aplican los métodos de Mínimos Cuadrados Indirecto (MCI) y Mínimos Cuadrados en Dos Etapas (MC2E).

3. Enfoque con información completa: se estiman en su conjunto y de manera simultanea todas las ecuaciones del modelo. Un método a usar para la estimación en este caso es el de Mínimos Cuadrados en Tres Etapas (MC3E).

SERIES DE TIEMPO

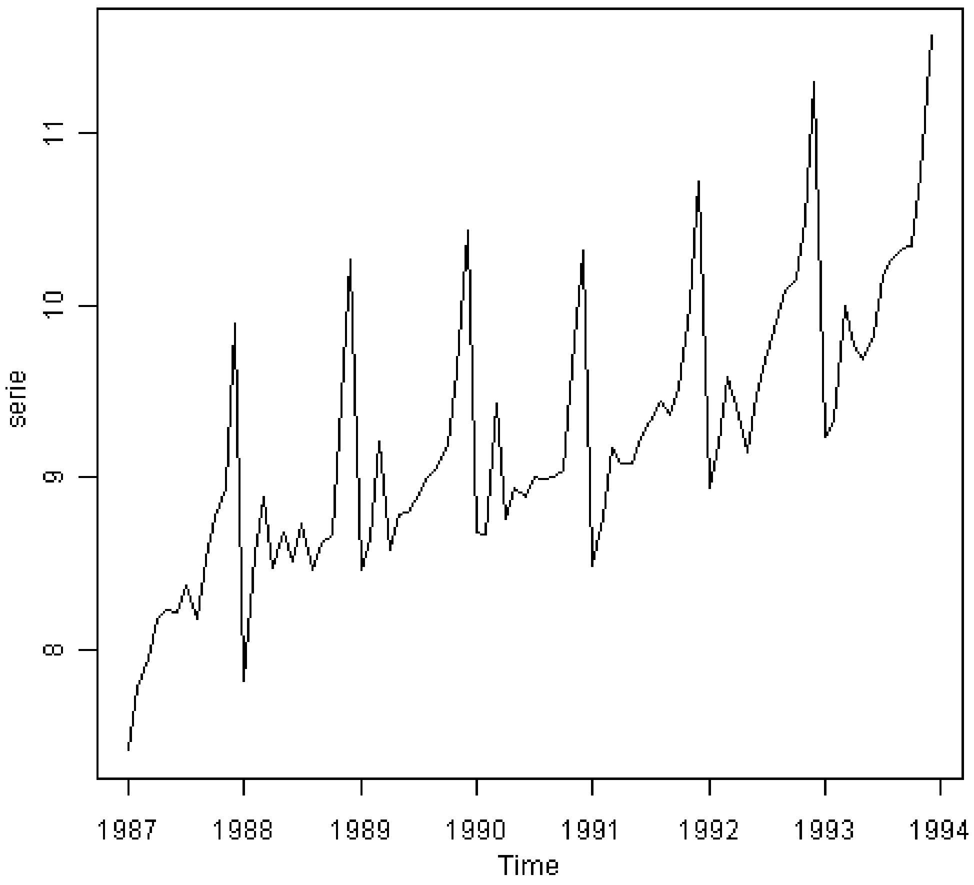
Definición

Una serie de tiempo es aquel grupo de observaciones que contienen los valores de una variable determinada en diferentes momentos del tiempo.

El método clásico identifica componentes para las series de tiempo:
- Tendencia (T).
- Fluctuaciones cíclicas (C).
- Variaciones estacionales (E).
- Variaciones irregulares (I).

De este modo, para cualquier período, el valor de la variable está determinado por:

$$Y = T \text{ x } C \text{ x } E \text{ x } I$$



Estacionariedad

Se refiere a la invariabilidad de los momentos estadísticos de una serie de tiempo. Tiene dos tipos:

1. Estacionariedad débil:

$$E(Y_t) = \mu$$

$$Var(Y_t) = \sigma^2$$

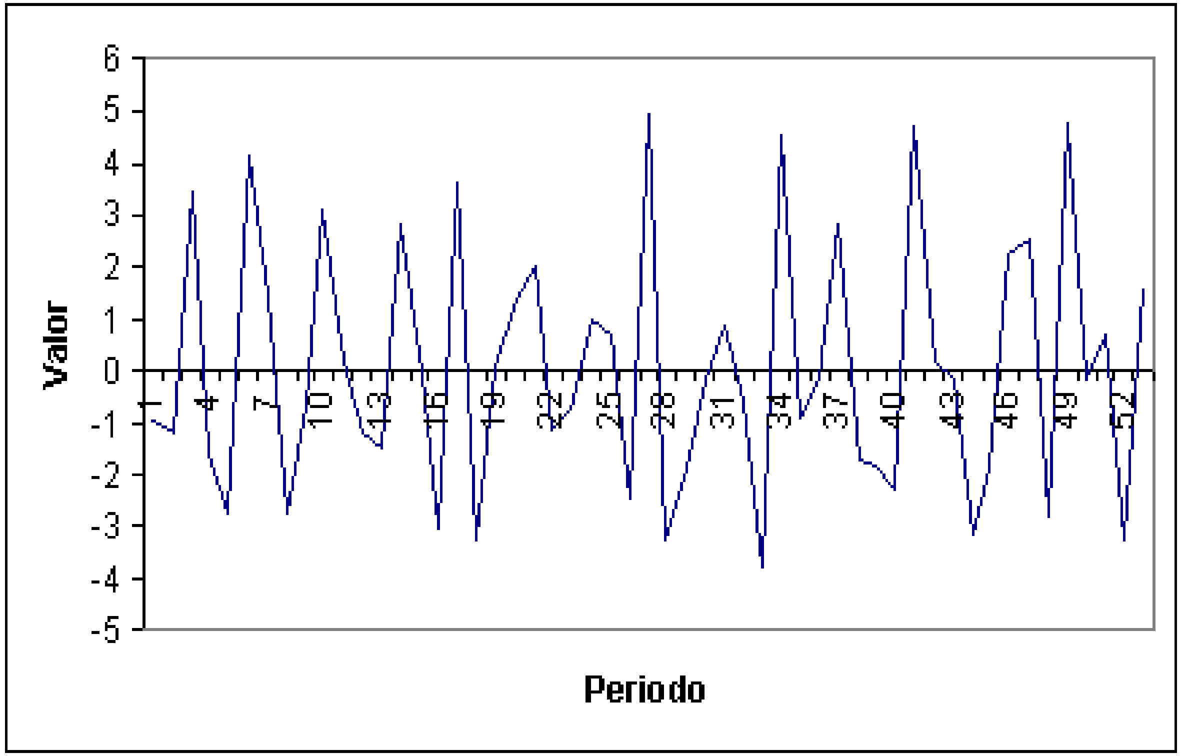
$$Cov(Y_t, Y_{t+k}) = \gamma_k$$

2. Estacionariedad fuerte:

$$E(\varepsilon_t) = 0$$

$$Var(\varepsilon_t) = \sigma^2$$

$$Cov(\varepsilon_t, \varepsilon_{t+k}) = 0$$



Modelo de caminata aleatoria

1. Caminata aleatoria sin deriva o desvío: Es de la forma:

$$Y_t = Y_{t-1} + \varepsilon_t$$

Donde ε_t es "ruido blanco". La solución general de la serie por iteraciones es:

$$Y_t = Y_0 + \Sigma_1^t \varepsilon_t$$

Que no es estacionaria pues:

$$E(Y_t) = Y_0 \qquad Var(Y_t) = t\sigma^2$$

2. Caminata aleatoria con deriva o desvío: Es de la forma:

$$Y_t = \delta + Y_{t-1} + \varepsilon_t$$

Donde ε_t es "ruido blanco". La solución general de la serie por iteraciones es:

$$Y_t = t\delta + Y_0 + \Sigma_1^t \varepsilon_t$$

Que no es estacionaria pues:

$$E(Y_t) = t\delta + Y_0 \qquad Var(Y_t) = t\sigma^2$$

Prueba de raíz unitaria

Se suele realizar por medio del test de Dickey-Fuller Aumentado que contrasta la estructura de hipótesis siguiente:

$$H_0 = La\ serie\ presenta\ raíz\ unitaria$$

$$H_1 = La\ serie\ no\ presenta\ raíz\ unitaria$$

Si el p-valor asociado al estadístico de MacKinnon es mayor al nivel de significación elegido, no se rechaza la hipótesis nula, de modo que la serie sí presentaría raíz unitaria, es decir, no sería estacionaria.

Augmented Dickey-Fuller Unit Root Test on PBI

Null Hypothesis: PBI has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 4 (Automatic - based on SIC, maxlag=4)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.239580	0.8989
Test critical values:	1% level	-4.004836	
	5% level	-3.432566	
	10% level	-3.140059	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PBI)
Method: Least Squares
Date: 09/15/10 Time: 19:07
Sample (adjusted): 1993M06 2009M12
Included observations: 199 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PBI(-1)	-0.057181	0.046129	-1.239580	0.2166
D(PBI(-1))	-0.291186	0.073049	-3.986193	0.0001
D(PBI(-2))	-0.206957	0.066492	-3.112490	0.0021
D(PBI(-3))	-0.420970	0.064758	-6.500637	0.0000
D(PBI(-4))	-0.446095	0.066597	-6.698459	0.0000
C	5.478286	3.982643	1.375540	0.1706
@TREND(1993M01)	0.035226	0.023686	1.487224	0.1386

MODELOS AR, MA, ARMA Y ARIMA

Estructura de los modelos

1. Modelo autorregresivo: Un modelo $AR(p)$ se escribe como:

$$Y_t = \gamma_1 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t$$

2. Modelo de media móvil: Un modelo $MA(q)$ se escribe como:

$$Y_t = \delta + \alpha_1 \varepsilon_t + \alpha_2 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q}$$

3. Modelo ARMA: Un modelo $ARMA(p,q)$ es una combinación de los anteriores y se escribe como:

$$Y_t = \varphi + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \alpha_1 \varepsilon_t + \alpha_2 \varepsilon_{t-1} + \cdots + \alpha_q \varepsilon_{t-q}$$

4. Modelo ARIMA: Es un proceso ARMA que requiere ser diferenciado d veces para ser estacionario, representándose esto como $ARIMA(p,d,q)$.

Metodología Box-Jenkins

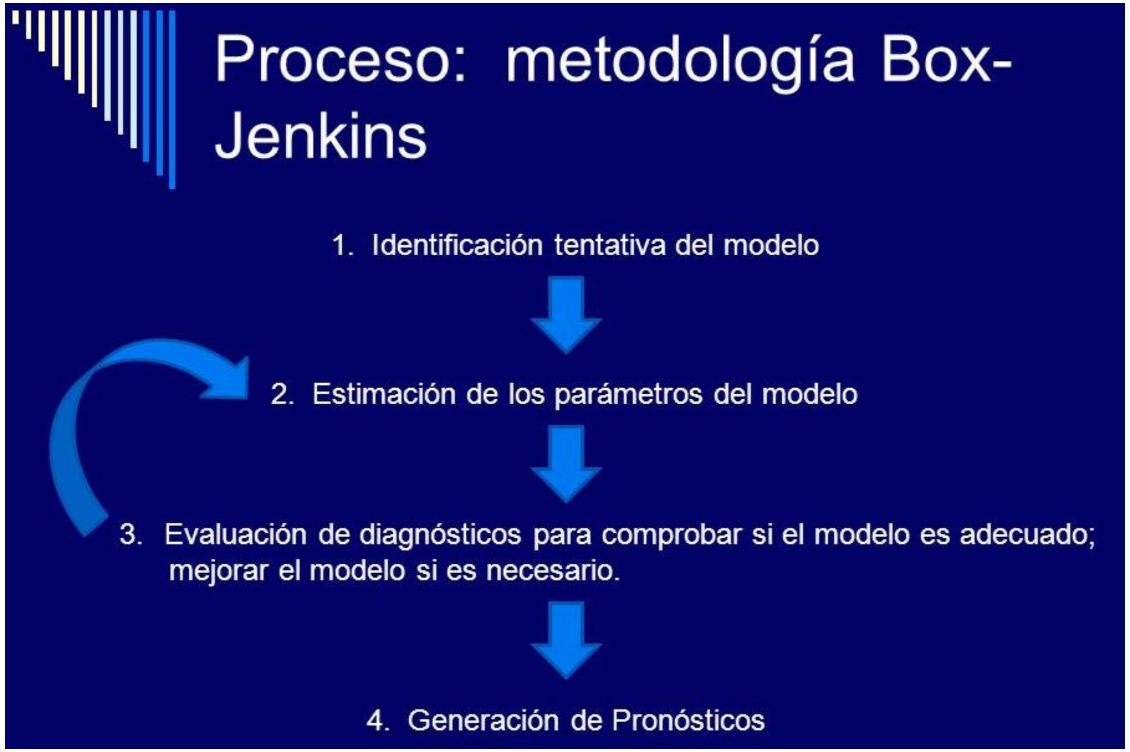
Sirve para determinar qué tipo de proceso *(AR, MA, ARMA o ARIMA)* sigue una determinada serie de tiempo y se desarrolla en términos de cuatro pasos:

1. Identificación.

2. Estimación.

3. Diagnóstico.

4. Pronóstico.



CONCLUSIONES

- Existen en econometría varias herramientas bastante útiles que van más allá del modelo de regresión lineal clásico.

- Determinadas relaciones económicas se ajustan más propiamente a modelos de tipo no lineal o de respuesta cualitativa. Asimismo, en caso se cuente con observaciones temporales para distintas unidades o sujetos de corte transversal, convendrá aplicar datos de panel. Si lo que se tiene más bien es un sistema de ecuaciones habrá que evaluar distintos métodos de estimación.

- Las series de tiempo nos permiten estudiar la trayectoria de determinadas variables económicas a lo largo del tiempo e incluso hacer pronósticos si verificamos estacionariedad. A su vez, por medio de la metodología de Box-Jenkins podemos determinar si un proceso sigue una estructura AR, MA, ARMA o ARIMA.

REFERENCIAS

. Gujarati, D. y Porter, D. (2011). *Econometría*. México: McGraw-Hill.

. Larios, J., Álvarez, V. y Quineche, R. (2014). *Fundamentos de Econometría*. Lima: Universidad San Ignacio de Loyola.

. Novales, A. (1993). *Econometría*. Madrid: McGraw-Hill.

. Sosa, W. (2015). *El Lado Oscuro de la Econometría*. Buenos Aires: Temas.

Profesor Dante A. Urbina:

. Página Web: http://www.danteaurbina.com
. Facebook: http://www.facebook.com/danteaurbina.oficial
. Canal YouTube: http://www.youtube.com/channel/UCCwVIDA-8wV4D_GpYNVecrg