ECONOMETRÍA: MODELO DE REGRESIÓN NO LINEAL- ARCH-GARCH.

San Antonio de Los Altos 30 de julio de 2007

Autor: Andrés E Reyes Polanco[1].

Esta monografía está organizada de la siguiente forma: la primera parte se introduce la importancia del tema, la segunda parte se define lo que es un modelo no lineal, tanto los modelos de serie de tiempo como los modelos causales uniecuacionales de regresión no lineal; la III se refiere a conocimientos previos de los métodos de estimación no lineales tales como la expansión de Taylor, mínimos cuadrados no lineales y máxima verosimilitud indicando, las propiedades de cada uno de los estimadores obtenidos por cada método, el problema de parametrización y la transformación de Box-Cox. En el siguiente punto, el IV se plantea los contrastes: con restricciones lineales y no lineales: Wald, máxima verosimilitud, y él de los multiplicadores de Lagrange . El punto V se plantean varios test de no linealidad en los modelos causales. La VI parte trata del modelo ARCH y GARCH los métodos de estimación y contraste de hipótesis, los modelos relacionados con ellos, tales como: TARCH, INGARCH, se estudia una prueba basada el exponente de Liapunov para detectar no linealidad, y el test BDS. Al final hay un anexo del uso del SPSS para resolver algunos problemas de estimación no lineal planteados en esta monografía.

Palabras claves: regresión no lineal, expansión de Taylor, mínimos cuadrados no lineales, máxima verosimilitud, contrastes restringidos, contraste de no linealidad, serie de tiempo no lineal: ARCH, GARCH-TARCH- INGARCH. Exponente de Liapunov BDS.

INTRODUCCIÓN.

Como señala Lachtermacher et al [2] en la practica los modelos del tipo lineal en serie de tiempo tales como ARIMA(p,d,q) o los modelos causales de regresión lineal, no siempre resultan los más adecuados para analizar y predecir acuradamente un proceso real. Por tal motivo se han propuestos modelos no lineales con la consecuencia de desarrollar métodos de estimación apropiados para estos casos así como los test que permitan validar los resultados. La justificación de estos desarrollos obedece a varias razones: una de ella es que hay un gran número de fenómenos que por su naturaleza son altamente volátiles en el tiempo, tanto en el campo económico financiero como en otros campos como la epidemiología, las telecomunicaciones, el mercado energético, otra razón es que las relaciones entre un fenómeno y los otros que se toman como explicativos se ha observado que estas relaciones son más que proporcionales, es decir, no lineales.

[1] Profesor Asociado UCV.
[2] Lachtermacher et al ( ) Backpropagation in Time Series Forecasting-Journal of Forescasting-Vol 14 pag 338-393

Las relaciones entre los diferentes actores de la economía, donde cada uno afecta el comportamiento del otro desdibujan la posibilidad del empleo de modelos uniecuacionales y multiecuacionales lineales para explicar y predecir la evolución de las variables microeconómicas y macroeconómicas en el tiempo, de ahí el interés cada vez más marcado en los modelos de regresión y de serie de tiempo no lineal que están incluido en los nuevos textos de econometría como Greene; W (1999), Gugarati; D (2006), Pindych; R y Rubinfeld; D (2001). En esta monografía nos avocaremos al estudio de este tipo de modelos no lineales tanto en regresión como en las series de tiempo, concretamente los dos más familiares: el modelo ARCH y el GARCH y algunas de sus variantes.

II.-MODELOS NO LINEALES.

Un modelo no lineal clásico en econometría es el modelo generalizado del consumo en función de la renta dado por:

$C_t = \alpha + \beta P_t^{\gamma} + \varepsilon_t$

Donde C_t es el consumo en el período t, P_t es el producto interno bruto en ese mismo período, ε_t es la perturbación aleatoria en el período t que se asume con distribución normal no necesariamente homoscedástica. Los parámetros son: α, β y γ, como puede observarse este último parámetro es un exponente.

Otro modelo no lineal es la generalización de la función de producción de Cobb-Douglas propuesto por Zellner et al (1970):

$\ln Y_t + \theta P Y_t = \ln \alpha + \beta(1-\delta) C_t + \beta\delta \ln T_t + \varepsilon_t$.

Donde Y_t, C_t, T_t son respectivamente, la producción, el factor capital y el factor trabajo en el período t; α, β, δ son los parámetros a estimar y finalmente ε_t es la perturbación aleatoria en el período t que se asume con distribución normal $N(0, \sigma^2)$.

En serie de tiempo tenemos como un ejemplo de modelo no lineal sin término de promedio móvil él propuesto por Byers el al (1995):

$$Y_t = \mu + \sum_{j=1}^{p} \varphi_j Y_{t-j} + \sum_{j=1}^{p}\sum_{k=1}^{q} \beta_{jk} Y_{t-j} \alpha_{t-k} + \alpha_t \quad (1)$$

Este modelo se conoce como bilineal sin término de promedio móvil en donde $\{Y_t\}$ y $\{\alpha_t\}$ son sucesiones de variables aleatorias, μ, φ_j y β_{jk} son parámetros que deben ser estimados.

Este modelo tiene una parte lineal autoregresiva de orden p representada por $\mu + \sum_{j=1}^{p} \varphi_j Y_{t-j} + \alpha_t$ y una parte no lineal $\sum_{j=1}^{p}\sum_{k=1}^{q} \beta_{jk} Y_{t-j} \alpha_{t-k}$, si $\beta_{jk} = 0 \quad \forall (j,k)$ entonces estaríamos en presencia de un $AR(p)$. Si la serie de tiempo es hetoroscedástica condicional, es decir con varianza condicional de las perturbaciones diferentes en el tiempo, entonces estos autores mencionan que tal modelo es inapropiado su aplicación directa. Kuldeep Kumar (1986) propuso un modelo más simplificado que el anterior dado por:

$$Y_t = \sum_{j=1}^{p}\sum_{k=1}^{q}\beta_{jk}Y_{t-j}\alpha_{t-k} + \alpha_t \quad (2)$$

Si $\beta_{jk} = 0 \ \ \forall j > k$ el modelo se llama superdiagonal, si $\beta_{jk} = 0 \ \ \forall j < k$ el modelo se llama subdiagonal y si $\beta_{jk} = 0 \quad \forall j \neq k$ el modelo se llama diagonal. Un caso particular del modelo dado en (2) es:

$$Y_{tt} = \beta Y_{t-j}\alpha_{t-k} + \alpha_t$$

Empleando este modelo, el método de Monte Carlo y utilizando el momento de tercer orden logra distinguir entre un modelo bilineal y un *AR*, *MA* o *ARMA* y de forma similar si se trata de un modelo bilineal diagonal o no.
En estos tipos de modelos como los descritos surgen tres problemas fundamentales a la hora de conocer si la serie proviene de un proceso no lineal.

1. Poseer alguna medida, cuya valoración nos de una condición necesaria de la presencia de no linealidad tanto en modelos causales como los de serie de tiempo.
2. Estimar los parámetros asociados al modelo no lineal especificado.
3. Tener una medida de bondad de ajuste del modelo a los datos empíricos.

Para el primer caso hay varias propuestas que dan condiciones necesarias del comportamiento no lineal de series temporales, entre ellos están: el estadístico de Brock, Dechert y Scheinkman BDS, el exponente de Lyapunov, multiplicadores de Lagrange (LM) y finalmente el RBF. El problema que se presenta es la cantidad de datos requeridos para aplicar alguna de estas técnicas. Un ejemplo de aplicación de un caso real del estadístico BDS está en el artículo ya citado de Byers et al (1995), un desarrollo del empleo del exponente de Lyapunov se encuentra en Cline; D.B.H (2006), el empleo de los multiplicadores de Lagrange se encuentra por ejemplo en Medeiros; M.C et al (2003) y finalmente el uso de test construidos con redes neuronales (RBF) Blake; A.P et al (2003). Los test DBS y exponente de Lyapunov se puede aplicar a la serie de tiempo observada sin hacer referencia a un modelo particular, por tanto puede ser una buena estrategia determinar primero si la serie en cuestión responde a un proceso no lineal, bien sea estocástico o deterministico. En el caso de modelos causales están los trabajos de MacKinnon et al (1983) y más recientemente los de Schimek, M (2000) y Samorov, A et al (2006).
El segundo problema de estimación, hay en forma general los siguientes métodos: linealización del modelo empleando la expansión de Taylor, mínimos cuadrados no lineales, máximo verosimilitud o el método generalizado de los momentos. Estos métodos son aplicados por igual a modelos de regresión no lineal como los de serie de tiempo. Un método es preferido a otro tomando en cuenta las propiedades asintóticas de los estimadores obtenidos y las facilidades computacionales que presentan.
Finalmente, entre las medidas de bondad de ajusten están el coeficiente de determinación que en el caso de regresión no lineal hay que interpretarlo de una forma diferente que cuando se emplea en el modelo lineal. Adicionalmente los diferentes test que permiten validar el modelo propuestos. Cuando hay restricciones en los parámetros, bien sean estas lineales o no, están entre otras, las pruebas o contrastes de Wald, la razón de verosimilitud y los multiplicadores de Lagrange.

III.-MÉTODOS DE ESTIMACIÓN NO LINEAL.

En este punto, veremos tres métodos de estimación aplicados a la regresión no lineal y posteriormente a las series temporales: el método de expansión de Taylor, mínimos cuadrados no lineales y finalmente el método de máxima verosimilitud. En la bibliografía especializada de econometría se encuentran además de estos métodos: el método de los mínimos cuadrados no lineales en dos etapas, empleo de variables instrumentales no lineales y redes neuronales, este último aplicado a las series de tiempo.

Ejemplo 1:
Consideremos dos pares de modelos de regresión no lineales:

a) $Y = \beta_0 \prod_{i=1}^{p} X_i^{\beta_i} e^{\varepsilon}$ b) $Y = \beta_0 \prod_{i=1}^{p} X_i^{\beta_i} + \varepsilon$

a1) $Y = e^{\beta_0 + \sum_{i=1}^{p} \beta_i X_i + \varepsilon}$ b1) $Y = e^{\beta_0 + \sum_{i=1}^{p} \beta_i X_i} + \varepsilon$

En donde Y es una variable aleatoria observable endógena, X_i son p variables exógenas, $\beta_0, \beta_1, ...\beta_p$ son los parámetros a estimar ε es una variable aleatoria, generalmente se asume que tiene una distribución normal.
Los modelos de regresión no lineales: a) y a1) son linealizable mediante una transformación logaritmica directa, esto es son intrínsecamente lineales, en efecto al tomar logaritmo neperiano en ambos modelos se obtiene:

a) $\ln Y = \ln \beta_0 + \sum_{i=1}^{p} \beta_i \ln X_i + \varepsilon$ a1) $\ln Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \varepsilon$

Al aplicar los mínimos cuadrados las ecuaciones normales obtenidas son lineales en los parámetros.
No así los modelos b) y b1) pero se pueden linealizar mediante el empleo de la expansión de Taylor.[3] Si en estos modelos se aplica los mínimos cuadrados ordinarios, las ecuaciones normales son no lineales en los parámetros

---

[3] Sea $f: R^n \to R$ una función doblemente diferenciable en un punto $\overset{*}{X}$, esto es : existen el vector gradiente $\nabla f(\overset{*}{X})$ y la matriz Hessiana $H(\overset{*}{X})$ y una función $\alpha: R^n \to R$, entonces $f(.)$ puede expresarse como:

$$f(X) = f(\overset{*}{X}) + \nabla f(\overset{*}{X})^T (X - \overset{*}{X}) + \frac{1}{2}(X - \overset{*}{X})^T H(\overset{*}{X})(X - \overset{*}{X}) + \left\| X - \overset{*}{X} \right\|^2 \alpha(\overset{*}{X}; X - \overset{*}{X})$$

$$\lim_{X \to \overset{*}{X}} \alpha(\overset{*}{X}; X - \overset{*}{X}) \to 0 \quad (\nabla f(.)^T = (\partial f / \partial x_1, \partial f / x_2 .. \partial f / x_n); \; H(.) = \{\partial^2 f / \partial x_i \partial x_j\})$$

IIIa.- Expansión de Taylor:

Para resolver estos problemas, en donde no se puede linealizar directamente la ecuación de regresión mediante una transformación, podemos emplear la expansión de Taylor dada como:

$$Y_i \cong f(X_i;\overset{\circ}{\beta}) + \varepsilon_i = f(X_i;\overset{\circ}{\beta}) + \nabla f(X_i;\overset{\circ}{\beta})^T(\beta - \overset{\circ}{\beta}) + \varepsilon_i$$

Donde $X_i = (x_{i0}, x_{i1}, x_{i2}, x_{i3}...x_{ip})$ es el vector asociado a la observación i-ésima y $\beta = (\beta_0, \beta_1, \beta_2...\beta_p)$ es el vector de parámetros, perteneciente al espacio paramétrico $p+1$ dimensional B, $f(X_i;\beta)$ es una función continua doblemente diferenciable en B, entonces:

$$\nabla f(X_i;\beta)^T = (\partial f(x_{i0};\beta)/\partial\beta_0, \partial f(x_{i1};\beta)/\partial\beta_1 . \partial f(x_{i2};\beta)/\partial\beta_2 .... \partial f(x_{ip};\beta)/\partial\beta_p)\ ;\ i = 1,2....n$$

Luego podemos escribir $Y_i = f(X_i;\beta) + \varepsilon_i$, empleando la expansión de Taylor como:

$$Y_i \cong f(X_i;\overset{\circ}{\beta}) + \sum_{j=1}^{p} \partial f(X_i;\overset{\circ}{\beta})/\partial\beta_j (\beta_j - \overset{\circ}{\beta}_j) + \varepsilon_i \quad (3)$$

Donde $\overset{\circ}{\beta}$ es un valor inicial del vector de parámetros, si despejamos y agrupamos convenientemente obtenemos lo siguiente:

$$Y_i - f(X_i;\overset{\circ}{\beta}) + \sum_{j=1}^{p} \overset{\circ}{\beta}_j \partial f(X_i;\overset{\circ}{\beta})/\partial\beta_j \cong \sum_{j=1}^{p} \beta_j \partial f(X_i;\overset{\circ}{\beta})/\partial\beta_j + \varepsilon_i \quad (4)$$

Las derivadas evaluadas en $\overset{\circ}{\beta}$ para cada observación la denotamos como Z°_{ij}, esto es:

$$\partial f(X_i;\overset{\circ}{\beta})/\partial\beta_j = Z^{\circ}_{ij};\ j = 0,1,2...p;\ i = 1,2...n$$

Y $\partial f(X_i;\beta)/\partial\beta_j = Z_{ij};\ j = 0,1,2...p;\ i = 1,2...n$

El miembro derecho de la ecuación (4) lo denotamos por Y_i^0:

$$Y_i - f(X_i;\overset{\circ}{\beta}) + \sum_{j=1}^{p} \overset{\circ}{\beta}_j \partial f(X_i;\overset{\circ}{\beta})/\partial\beta_j \cong Y_i - f(X_i;\overset{\circ}{\beta}) + Z_i^{\circ} = Y_i^0$$

El miembro izquierdo lo expresamos como: $\sum_{j=1}^{p} \beta_j \partial f(X_i;\overset{\circ}{\beta})/\partial\beta_j + \varepsilon_i = \sum_{j=1}^{p} \beta_j Z_{ij} + \varepsilon$

Ahora, podemos escribir la ecuación (3) como:

$$Y_i^0 = \sum_{j=1}^{p} \beta_j Z_{ij} + \varepsilon_i$$

En notación matricial es:

$$Y = Z\beta + \varepsilon$$

A partir de aquí podemos aplicar cualquier criterio norma L_p, por ejemplo los mínimos cuadrados ordinarios:

$$\underset{\beta=\overset{*}{\beta}}{Min} \left\| Y - Z\beta \right\|_{p=2} \quad (5)$$

Entonces, obtenemos como estimador mínimo cuadrático dado el valor inicial $\overset{\circ}{\beta}$, lo que sigue:

$\overset{*}{\beta} = (Z^T Z)^{-1} Z^T Y$ (6)

La pregunta obligada es ¿cuál valor inicial? La repuesta está en la nota[4]

Si se asume que ε es normal $N(0,\sigma^2 I)$ y si existe la matriz inversa $(Z^T Z)^{-1}$ y el estimador de σ^2 está dado por:

$\sigma^{*2} = \varepsilon^{*T}.\varepsilon^{*}/(n-p')$ donde $\varepsilon^{*} = Y - Z\overset{*}{\beta}$ $p' = p+1$, entonces el estimador converge asintóticamente[5] a la distribución normal $\overset{*}{\beta} \overset{d}{\approx} N(\beta, \sigma^{*2}(Z^T Z)^{-1})$

Ejemplo 2:

Consideremos el modelo: $Y_i = f(X_i;\beta) + \varepsilon_i; Y_i = \alpha e^{\beta X_i} + \varepsilon_i; i = 1,2..n$ y consideremos unos valores iniciales de los parámetros: $\overset{\circ}{\alpha}$ y $\overset{\circ}{\beta}$ y evaluamos la función y sus derivadas en estos valores:

$f(X_i;\overset{\circ}{\alpha},\overset{\circ}{\beta}) = \overset{\circ}{\alpha} e^{\overset{\circ}{\beta} x}$ $\quad \partial f(X_i,\overset{\circ}{\alpha},\overset{\circ}{\beta})/\partial\alpha = e^{\overset{\circ}{\beta} X_i}$ $\quad \partial f(X_i,\overset{\circ}{\alpha},\overset{\circ}{\beta})/\partial\beta = \overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i}$

Evaluamos ahora los componentes de la suma: $\sum_{j=1}^{p} \overset{\circ}{\beta}_j \partial f(X_i;\overset{\circ}{\beta})/\partial\beta_j$ dados como:

$\overset{\circ}{\alpha}\partial f(X_i,\overset{\circ}{\alpha},\overset{\circ}{\beta})/\partial\alpha = \overset{\circ}{\alpha} e^{\overset{\circ}{\beta} X_i}$ $\quad \overset{\circ}{\beta}\partial f(X_i,\overset{\circ}{\alpha},\overset{\circ}{\beta})/\partial\beta = \overset{\circ}{\beta}\overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i}$

Con estos resultados podemos obtener:

$Y_i - f(X_i;\overset{\circ}{\beta}) + \sum_{j=1}^{p} \overset{\circ}{\beta}_j \partial f(X_i;\overset{\circ}{\beta})/\partial\beta_j \cong Y_i - f(X_i;\overset{\circ}{\beta}) + Z_i^{\circ} = Y_i^0$

Que en nuestro ejemplo es:

$Y_i - \overset{\circ}{\alpha} e^{\overset{\circ}{\beta} x} + \overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i} + \overset{\circ}{\beta}\overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i} = Y^0$

---

[4] El primer valor inicial nos dará una primera estimación de los parámetros al resolver el problema (6), este valor lo empleamos como nuevo valor inicial y resolvemos nuevamente el problema (6) y así sucesivamente hasta que se estabilicen los estimadores. Esto se puede expresar como sigue: Sea $\beta_j^{*(K)}$ el estimador logrado en la iteración (K) y $\beta_j^{*(K+1)}$ el obtenido en el paso (K+1), el proceso termina si para un $\delta \geq 0$, suficientemente pequeño:

$\left|\left(\beta_j^{*(K+1)} - \beta_j^{*(K)}\right)/\beta_j^{*(K)}\right| < \delta$ para $j = 1,2...p$.

[5] Consideremos una sucesión de variables aleatorias $\{X_n\}$, con la misma función de distribución $F_{X_n}(x;\theta_1,\theta_2...\theta_k)$ para toda variable aleatoria de la sucesión y, una variable aleatoria X con distribución $F_X(x;\theta_1,\theta_2...\theta_k)$, si la sucesión $\{X_n\}$ converge en probabilidad a X entonces, la sucesión de funciones de distribuciones : $F_{X_n}(x;\theta_1,\theta_2...\theta_k)$ converge a la función de distribución $F_X(x;\theta_1,\theta_2...\theta_k)$, esta última se llama distribución límite y se escribe:

$\lim_{n\to\infty} F_{Xn}(x;\theta_1,\theta_2..\theta_k) = F(x,\theta_1,\theta_2..\theta_k)$

Ahora considerando $\alpha \partial f(X_i, \overset{\circ}{\alpha}, \overset{\circ}{\beta})/\partial\alpha = \alpha e^{\overset{\circ}{\beta} X_i}$ y $\beta \partial f(X_i, \overset{\circ}{\alpha}, \overset{\circ}{\beta})/\partial\beta = \beta \overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i}$ entonces obtenemos:

$$\sum_{j=1}^{p} \beta_j \partial f(X_i; \overset{\circ}{\beta})/\partial\beta_j = \alpha e^{\overset{\circ}{\beta} X_i} + \beta \overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i}$$

Al considerar la ecuación:

$$Y_i - f(X_i; \overset{\circ}{\beta}) + \sum_{j=1}^{p} \overset{\circ}{\beta}_j \partial f(X_i; \overset{\circ}{\beta})/\partial\beta_j \cong \sum_{j=1}^{p} \beta_j \partial f(X_i; \overset{\circ}{\beta})/\partial\beta_j$$

Finalmente obtenemos:

$$Y_i - \overset{\circ}{\alpha} e^{\overset{\circ}{\beta} x} + \overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i} + \overset{\circ}{\beta} \overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i} = \alpha e^{\overset{\circ}{\beta} X_i} + \beta \overset{\circ}{\alpha} X_i e^{\overset{\circ}{\beta} X_i}$$

La ecuación linealizada es:

$$Y_i^0 \cong \alpha Z_{i1}^{\circ} + \beta Z_{i2}^{\circ} + \varepsilon .$$

Partiendo de esta última ecuación aplicamos los mínimos cuadrados ordinarios para obtener los estimadores de los parámetros: α, β .

IIIb.- Mínimos cuadrados no lineales MCNL:

Consideremos el modelo general $Y = f(X;\beta) + \varepsilon$ donde $f(X;\beta)$ es no lineal en los parámetros y ε es un vector de variables aleatorias independientes e idénticamente distribuidas. El problema es encontrar los estimadores de los parámetros β tal que minimicen a:

$$\|Y - f(X;\beta)\|_{p=2} = \sum_{i=1}^{n} (Y_i - f(X_i;\beta))^2 = S(\beta)$$

Tomando derivadas parciales respecto a los parámetros obtenemos:

$$\partial S(\beta)/\partial\beta = -2\sum_{i=1}^{n} (Y_i - f(X_i;\beta)) \partial f(X_i;\beta)/\partial\beta \quad (5)$$

O escrito de otra forma:

$$\partial S(\beta)/\partial\beta_j = -2\sum_{i=1}^{n} (Y_i - f(X_i;\beta)) \partial f(X_i;\beta)/\partial\beta_j \quad j = 1,2,3...p$$

Las derivadas quedarán como funciones no lineales en los parámetros, por lo que habrá que emplear los algoritmos de optimización no lineal para resolver las ecuaciones normales resultantes.

Igualando la ecuación dada en 5) al vector nulo, obtenemos las siguientes ecuaciones normales:

$$\sum_{i=1}^{n} (Y_i - f(X_i;\beta^*)) \partial f(X_i;\beta^*)/\partial\beta = 0_{p+1}$$

$$\sum_{i=1}^{n} Y_i \partial f(X_i;\beta^*)/\partial\beta = \sum_{i=1}^{n} f(X_i;\beta^*)) \partial f(X_i;\beta^*)/\partial\beta$$

Que en notación matricial es:

$$\left[\partial f(X,\beta^*)/\partial\beta\right]^T Y = \left[\partial f(X,\beta^*)/\partial\beta\right]^T f(X;\beta^*) \quad (6)$$

Ahora, hacemos $\left[\partial f(X,\beta^*)/\partial\beta\right]^T = X^T; f(X;\beta^*) = X\beta^*$. Entonces la ecuación (6) se puede escribir en notación matricial como:

$$X^T Y = X^T X \beta^*$$

El problema fundamental es que tanto $\left[\partial f(X,\beta^*)/\partial\beta\right]^T$; como $f(X;\beta^*)$ no son generalmente lineales en β^* y por tanto la solución de las ecuaciones normales rara vez

se puede obtener de forma directa, entonces se requiere de algoritmos[6] eficientes en termino de convergencia entre los que se encuentran lo que genéricamente se llaman métodos de direcciones factibles, el método de Newton-Raphson y otros[7]

Ejemplo 3:

$Y_i = \alpha e^{\beta X_i} + \varepsilon_i; i = 1,2..n \qquad \sum_{i=1}^{n} (Y_i - \alpha e^{\beta X_i})^2; j = 1,2$

$\partial S(\alpha,\beta)/\partial\alpha = -2\sum_{i=1}^{n}(Y_i - \alpha e^{\beta X_i})e^{\beta X_i} \quad \partial S(\alpha,\beta)/\partial\beta = -2\sum_{i=1}^{n}(Y_i - \alpha e^{\beta X_i})X_i e^{\beta X_i}$

$-2\sum_{i=1}^{n}(Y_i - \overset{*}{\alpha} e^{\overset{*}{\beta} X_i})e^{\overset{*}{\beta} X_i} = 0$

$-2\sum_{i=1}^{n}(Y_i - \overset{*}{\alpha} e^{\overset{*}{\beta} X_i})X_i e^{\overset{*}{\beta} X_i} = 0$

Puede observarse con este ejemplo que las ecuaciones son no lineales y no tienen solución inmediata.

Un procedimiento para resolver las ecuaciones normales anteriores , si se cuenta con un valor inicial $\beta^0 = \beta_t^*$, es el siguiente:

$$\beta_{t+1}^* = \left[\sum_{i=1}^{n} X_i^0 X_i^{0T}\right]^{-1}\left[\sum_{i=1}^{n} X_i^0 \left(Y_i^0 - f_i^0 + X_i^{0T}\beta_t^*\right)\right]$$

$$\beta_{t+1}^* = \beta_t^* + \left[\sum_{i=1}^{n} X_i^0 X_i^{0T}\right]^{-1}\left[\sum_{i=1}^{n} X_i^0 \left(Y_i^0 - f_i^0\right)\right]$$

$$\beta_{t+1}^* = \beta_t^* + (X^{0T}X^0)^{-1}X^{0T}e^0$$

El procedimiento termina cuando $X^{0T}e^0$ está próximo a cero. Green, W (1999)

---

[6] Un algoritmo es un procedimiento que consiste en generar de forma iterativa un conjunto de valores de acuerdo a ciertas reglas establecidas, mediante el cual se obtiene una solución que puede ser única o no.

[7] Ver Bazaraa y Shetty (1979) Nonlinear Programming, John Wiley and Sons-pág 361-434. Avriel (2003) Nonlinear Programming, Dover. Pág. 216 y siguientes. Uno de los algoritmos más empleados de Newton-Raphson que consiste en lo siguiente: Supongamos que deseamos obtener el mínimo de la función dada por $f(x)$ la cual es continua y doblemente diferenciable y supongamos que poseemos un valor inicial X_n^* , entonces:

$$f(X) \cong f(\overset{*}{X}_n) + \nabla f(\overset{*}{X}_n)^T(X - \overset{*}{X}_n) + \frac{1}{2}(X - \overset{*}{X}_n)^T H(\overset{*}{X}_n)(X - \overset{*}{X}_n) = G(X)$$

Usando la primera condición de existencia de un punto extremo :

$\partial G(X)/\partial X = \nabla f(\overset{*}{X}_n)^T + H(\overset{*}{X}_n)(X - \overset{*}{X}_n) = 0_n$. Entonces, definimos un nuevo valor:

$X_{n+1} = \overset{*}{X}_n - H(\overset{*}{X}_n)^{-1}\nabla f(\overset{*}{X})$ Este algoritmo es útil cuando la función es no cuadrática, de serla se obtendrá la solución en la primera iteración. El algoritmo termina cuando se considere que la solución se ha estabilizado, tal como lo indicamos en la nota 4.

IIIb.1.-Propiedades de los estimadote MCNL.

1.-El estimador MCNL: β^* en general es una función no lineal del vector de observaciones Y por tanto, y lo será también del vector de perturbaciones ε.
2.-Como consecuencia de lo anterior, no siempre el estimador MCNL es insesgado.
3.-$\left[\partial f(X,\beta^*)/\partial\beta\right]^T \varepsilon^* = 0$, donde $\varepsilon^* = Y - f(X;\beta^*)$.
4.-Como el problema de optimización es no lineal, no se puede garantizar que la solución obtenida como estimador, sea única.
5.-Para estudiar la distribución asintótica de β^*, consideramos el concepto de convergencia en probabilidad[8]. Como indicamos anteriormente: $\left[\partial f(X,\beta^*)/\partial\beta\right] = X$,

esto es: $X = \begin{bmatrix} \partial f(X_1,\beta^*)/\partial\beta_0 & \partial f(X_1,\beta^*)/\partial\beta_1 & . & . & \partial f(X_1,\beta^*)/\partial\beta_p \\ \partial f(X_2,\beta^*)/\partial\beta_0 & \partial f(X_2,\beta^*)/\partial\beta_1 & . & . & \partial f(X_2,\beta^*)/\partial\beta_p \\ . & . & . & . & . \\ . & . & . & . & . \\ \partial f(X_n,\beta)/\partial\beta_0 & \partial f(X_n,\beta^*)/\partial\beta_1 & . & . & \partial f(X_n,\beta^*)/\partial\beta_p \end{bmatrix}$

Entonces si:

$$P\lim_{n\to\infty} X^T X = P\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\left(\partial f(X_i;\overset{*}{\beta})/\partial\beta\right)^T\left(\partial f(X_i;\overset{*}{\beta})/\partial\beta\right)\right] = Q_0$$

Donde Q_0 es una matriz positiva definida, esto es: $Q_0 \perp X^T Q_0 X > 0$
Y además, se verifica:

$$P\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i\,\varepsilon_i = 0 \quad \text{y} \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\,\varepsilon_i \overset{d}{\to} N(0_n,\sigma^2 Q_0^{-1})$$

Entonces:

$$\overset{*}{\beta} \overset{d}{\to} N(\beta,\frac{\sigma^2}{n}Q_0^{-1})$$

Ejemplo 4.
Consideremos el siguiente modelo: $Y_t = \alpha e^{\beta X_t} + \varepsilon_t$ y supongamos que se ha obtenido el ajuste por mínimos cuadrados no lineales: $\overset{*}{Y}_t = \overset{*}{\alpha}\, e^{\overset{*}{\beta} X_t}$; para obtener el estimador de la matriz de varianza covarianza de los estimadores procedemos de la siguiente forma: calculamos las correspondientes derivadas parciales: $\partial\overset{*}{Y}_t/\partial\overset{*}{\alpha} = e^{\overset{*}{\beta} X_t}$; $\partial\overset{*}{Y}_t/\partial\overset{*}{\beta} = \overset{*}{\alpha}\, e^{\overset{*}{\beta} X_t}$ y con ellas definimos el vector gradiente y su transpuesta:

$$\nabla f(\overset{*}{\alpha},\overset{*}{\beta}) = \begin{bmatrix} e^{\overset{*}{\beta} X_t} \\ \overset{*}{\alpha}\, e^{\overset{*}{\beta} X_t} \end{bmatrix}; \quad \nabla f(\overset{*}{\alpha},\overset{*}{\beta})^T = \begin{bmatrix} e^{\overset{*}{\beta} X_t} & \overset{*}{\alpha}\, e^{\overset{*}{\beta} X_t} \end{bmatrix}$$

[8] Una sucesión de variables aleatorias ς_n converge en probabilidad a una constante c si para cualquier $\varepsilon > 0$ se verifica que: $\lim_{n\to\infty} P\left(|\varsigma_n - c| < \varepsilon\right) = 1$. Esto quiere decir, que a partir de $n > N$ el suceso $|\varepsilon - c| < \varepsilon$, es un suceso seguro. Esto se escribe como: $P\lim_{n\to\infty}\varsigma_n = c$.

Con ellos obtenemos la matriz Q_0 .El estimador de la matriz de varianza covarianza es:

$$\overset{*}{\sigma}_{\varepsilon}^{2}\left[\sum_{t=1}^{n}\nabla f(\overset{*}{\alpha},\overset{*}{\beta})^{T}\nabla f(\overset{*}{\alpha},\overset{*}{\beta})\right]^{-1}=\overset{*}{\sigma}_{\varepsilon}^{2}\begin{bmatrix}\sum_{t=1}^{n}e^{2\overset{*}{\beta}X_t} & \overset{*}{\alpha}\sum_{t=1}^{n}X_t e^{2\overset{*}{\beta}X_t}\\ \overset{*}{\alpha}\sum_{t=1}^{n}X_t e^{2\overset{*}{\beta}X_t} & \overset{*}{\alpha}^{2}\sum_{t=1}^{n}X_t^{2}e^{2\overset{*}{\beta}X_t}\end{bmatrix}^{-1}$$

Donde: $\sigma_t^{*2}=\sum_{t=1}^{n}\varepsilon_t^{*2}/n$ y $\varepsilon_t^{*}=Y_t-\overset{*}{Y}_t$

III.c.- Máxima verosimilitud.

Consideremos que se tiene una muestra de n variables aleatorias independientes e idénticamente distribuidas: $X_1, X_2,...X_n$ proveniente de una población con función de distribución: $F_X(x;\theta_1,\theta_2..\theta_k)$ Entonces, la función de densidad conjunta se expresa como:

$$f_{X1,X2,..Xn}(x_1,x_2,...x_n;\theta_1,...\theta_k)=\prod_{i=1}^{n}f_{Xi}(x_i;\theta_1,\theta_2..\theta_k)$$

Tomando logaritmo neperiano, obtenemos la función log-verosimilitud:

$$Lnf_{X1,X2,..Xn}(x_1,x_2,...x_n;\theta_1,...\theta_k)=\sum_{i=1}^{n}Lnf_{Xi}(x_i;\theta_1,\theta_2..\theta_k)$$

Considerando la condición necesaria para la existencia de un punto extremo, obtenemos los estimadores:

$$\partial Lnf_{X1,X2,..Xn}(x_1,x_2,...x_n;\theta^*{}_1,...\theta^*{}_k)/\partial\theta_j=\sum_{i=1}^{n}\partial Lnf_{Xi}(x_i;\theta^*{}_1,\theta^*{}_2..\theta^*{}_k)/\partial\theta_j=0;\ j=1,2..k$$

Encontremos el estimador máximo verosímil de vector de parámetro β y σ^2 del modelo: $Y_i=f(X_i;\beta)+\varepsilon$ asumiendo que el vector de perturbaciones aleatorias se distribuye como $\varepsilon\overset{D}{\cong}N(0,\sigma^2 I_{n:n})$. Entonces, Y_i tiene una distribución $N(f(X_i;\beta);\sigma^2)$ Consideremos bajo el supuesto de normalidad la función de densidad marginal de cada Y_i dada por:

$$g_{Yi}(Y_i;\sigma_\varepsilon,\beta)=\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left\{-\frac{1}{2\sigma^2{}_\varepsilon}\left(Y_i-f(X_i;\beta)\right)^2\right\}$$

Luego la función de verosimilitud es:

$$\prod_{i=1}^{n}g_{Yi}(Y_i;\sigma_\varepsilon;\beta)=\left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^n\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(Y_i-f(X_i;\beta)\right)^2\right\}$$

La función de log-verosimilitud es:

$$Ln\prod_{i=1}^{n}g_{Yi}(Y_i;\sigma_\varepsilon;\beta)=Ln\left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^n+\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(Y_i-f(X_i;\beta)\right)^2\right\}$$

Consideramos ahora las condiciones necesarias para la existencia de un punto extremo, entonces obtendremos los estimadores de β y σ^2:

$$\partial Ln\prod_{i=1}^{n}g_{Yi}(Y_i;\sigma_\varepsilon;\beta^*)/\partial\beta=\left\{-\frac{1}{2\sigma_\varepsilon^2}\partial\sum_{i=1}^{n}\left(Y_i-f(X_i;\beta^*\right)^2/\partial\beta\right\}=0_{p+1}\ (1)$$

$$\partial Ln\prod_{i=1}^{n}g_{Yi}(x_i;\sigma_\varepsilon;\beta)/\partial\sigma^2=\partial[Ln\left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^n+\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(Y_i-f(X_i;\beta)\right)^2\right\}]/\partial\sigma^2=0\ \ (2)$$

El primer conjunto de ecuaciones está formado por $p+1$ ecuaciones generalmente no lineales, por tanto se empleará las mismas técnicas que se emplean para los MCNL, de la ecuación (2) se obtiene el estimador de σ^2 una vez obtenido el del vector β y este es:

$$\overset{*}{\sigma}^2 = \sum_{i=1}^{n}\left(Y_i - f(X_i, \overset{*}{\beta}\right)^2 / n$$

Ejemplo 5.

Veremos como ejemplo la estimación de los parámetros de modelo logit. La variable aleatoria discreta Y_i tiene una ley de probabilidad de Bernoulli con probabilidad: P_i y Q_1 tal que $P_i = P(Y_i = 1) = \exp X_i\beta/(1+\exp X_i\beta)$ y $Q_i = P(Y_i = 0) = 1/(1+\exp X_i\beta)$, $i = 1,2...n$ donde se tiene p regresores que explican el comportamiento de Y_i, por tanto: $X_i = (1, X_{i1}, X_{i2}, X_{i3},..X_{ip})$ y $\beta = (\beta_0, \beta_1, \beta_2,...\beta_p)^T$ se desea estimar tanto P_i como β . Se toma una muestra de tamaño n y definimos la función de verosimilitud como:

$$L(Y; P_i) = \prod_{i=1}^{n} P_i^{Y_i}(1-P_i)^{1-Y_i}$$

Tomando logaritmo neperiano obtenemos:

$$LnL = \sum_{i=1}^{n} Y_i LnP_i + \sum_{i=1}^{n}(1-Y_i)Ln(1-P_i)$$

$LnL = \sum_{i=1}^{n} Y_i LnP_i - \sum_{i=1}^{n} Y_i Ln(1-P_i) + \sum_{i=1}^{n} Ln(1-P_i)$ Agrupando convenientemente tenemos:

$$LnL = \sum_{i=1}^{n} Y_i Ln\left[P_i/(1-P_i)\right] + \sum_{i=1}^{n} Ln(1-P_i)\,(1)$$

Considerando que: $Q_i = 1-P_i = 1/(1+\exp X_i\beta)$ y $Ln\left[P_i/(1-P_i)\right] = X_i\beta$ sustituyendo en (1) obtenemos:

$$LnL = \sum_{i=1}^{n} Y_i X_i\beta - \sum_{i=1}^{n} Ln(1+\exp X_i\beta)$$

Para obtener el estimador de β usamos la condición necesaria para tener un punto extremo:

$$\partial LnL/\partial\overset{*}{\beta} = \partial(\sum_{i=1}^{n} Y_i X_i\beta)/\partial\overset{*}{\beta} - \partial(\sum_{i=1}^{n} Ln(1+\exp X_i\beta))/\partial\overset{*}{\beta} = 0_{p+1}$$

$$\partial LnL/\partial\overset{*}{\beta}_j = \sum_{i=1}^{n} Y_i X_{ij} - \sum_{i=1}^{n} X_{ij}\exp X_i\overset{*}{\beta}/(1+\exp X_i\overset{*}{\beta}) = 0;\ j = 0,1,2...p$$

Este problema hay que resolverlo mediante el método de Newton, una vez obtenido el resultado se estima:

$\overset{*}{P}_i = \exp X_i\overset{*}{\beta}/(1+\exp X_i\overset{*}{\beta}); i = 1,2...n$.

IIIc1.-Propiedades.

Las propiedades generales de los estimadores de máxima verosimilitud son:

1.- Dado una sucesión de estimadores de máxima verosimilitud $\{\beta^*_{MV}\}$ y sea β el parámetro a estimar de una población con función de distribución: $G_X(X;\beta)$ entonces:

$P\lim_{n\to\infty}\{\beta^*_{MV}\} = \beta$. Esto quiere decir, que el estimador MV es consistente.

2.-Invarianza. El estimador de máxima verosimilitud es invariante respecto a una transformación, esto es: dado el parámetro β y una transformación $\varphi(\beta)$, si β^*_{MV} es el estimador de β, entonces $\varphi(\beta^*_{Mv})$ lo es de $\varphi(\beta)$.

3.- La distribución de $\{\beta^*_{MV}\}$ tiende asintóticamente a una distribución normal: $N(\beta;[I(\beta)]^{-1})$ donde $I(\beta) = -E(\partial^2 \ln L/\partial\beta\partial\beta^T) = E[(\partial \ln L/\partial\beta)(\partial \ln L/\beta)^T]$[9]. Se puede demostrar que la matriz de varianza covarianza del estimador de máxima verosimilitud es la dada.[10]

4.-El estimador de máxima verosimilitud es un estimador eficiente y alcanza la cota de Cramer Rao[11] bajo ciertos supuestos (ver nota 10): $VAR(\overset{*}{\beta}) \geq [I(\beta)]^{-1}$ si $E(\overset{*}{\beta}) = \beta$ y en general $VAR(\overset{*}{\beta}) \geq \left(\partial E(\beta)/\partial \overset{*}{\beta}\right)[I(\beta)]^{-1}$

Ejemplo 6.
Consideremos el mismo modelo del ejemplo 4: $Y_t = \alpha e^{\beta X_t} + \varepsilon_t$ y supongamos que el ajuste se obtuvo mediante el método de máxima verosimilitud: $\overset{*}{Y}_t = \overset{*}{\alpha}\, e^{\overset{*}{\beta} X_t}$. Para obtener la estimación de la matriz de varianza covarianza debemos partir de la función de verosimilitud siguiente:

$$L = \left[\frac{1}{2\pi\sigma_\varepsilon^2}\right]^{n/2} \exp\left\{\left(-\frac{1}{2\sigma_\varepsilon^2}\right)\sum_{i=1}^{n}\left(Y_i - \alpha e^{\beta X_i}\right)^2\right\}$$

La función log-verosimilitud es:

$$LnL = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma_\varepsilon^2 + \left\{\left(-\frac{1}{2\sigma_\varepsilon^2}\right)\sum_{i=1}^{n}\left(Y_i - \alpha e^{\beta X_i}\right)^2\right\}$$

Obtengamos las segundas derivadas evaluadas en: $\left(\overset{*}{\alpha}, \overset{*}{\beta}, \sigma_\varepsilon^{*2}\right)$

$$\partial^2 LnL/\partial\alpha^2 = \frac{1}{\sigma_\varepsilon^{*2}}\sum_{i=1}^{n} e^{2\overset{*}{\beta} X_i} \;;\; \partial^2 LnL/\partial\beta^2 = \frac{1}{\sigma_\varepsilon^{*2}}\overset{*}{\alpha}^2\sum_{i=1}^{n} X_i^2 e^{2\overset{*}{\beta} X_i} \;;\; \partial^2 LnL/\partial\alpha\partial\beta = \frac{1}{\sigma_\varepsilon^{*2}}\overset{*}{\alpha}\sum_{i=1}^{n} X_i\, e^{2\overset{*}{\beta} X_i}$$

---

[9] Esto es: $VAR(\partial \ln L/\partial\beta) = I(\beta)^{-1}$ Un estimador de $[I(\beta)]^{-1}$ es $[I^*(\beta^*_{MV})]^{-1} = [\partial^2 \ln L(\beta^*_{MV})/\partial\beta^*_{MV}\partial\beta^{*T}_{MV}]$ (ver Greene. W, Theil .H).

[10] Demostración que: $VAR(\partial \ln L/\partial\beta) = I(\beta)$. Partimos de los siguientes supuestos generales: 1.-El rango de las variables aleatoria Y_i, no depende del vector de parámetro β. 2.-La función de densidad $g_Y(.;\beta)$ posee derivadas de al menos de tercer orden con respecto a β y están acotadas. Entonces se verifica que $\int_S (\partial LnL^*/\partial\beta) L^* dx = 0_{p+1}$. Diferenciando esta ecuación obtenemos: $\int_S \{(\partial^2 LnL^*/\partial\beta\partial\beta)L^* + (\partial LnL^*/\partial\beta)(\partial LnL^*/\partial\beta)^T L^*\} dx = 0_{p+1}$. De aquí se obtiene que: $VAR(\partial LnL^*/\partial\beta) = -\int_S [\partial^2 LnL^*/\partial\beta\partial\beta] L^* dx = -E[\partial^2 LnL^*/\partial\beta\partial\beta]$.(Dhrymes, P; Theil., H; Greene, W)

[11] Ver Theil (1971) pag:385-387

$\partial^2 LnL/\partial\sigma_\varepsilon^{*2}\partial\alpha = 0$; $\partial^2 LnL/\partial\sigma_\varepsilon^{*2}\partial\beta = 0$; $\partial^2 LnL/\partial(\sigma_\varepsilon^{*2})^2 = \frac{n}{2\sigma_\varepsilon^{*4}}$ Luego la estimación de la matriz de varianza covarianza de los estimadores de máxima verosimilitud es:

$$\left[I(\overset{*}{\alpha},\overset{*}{\beta},\sigma_\varepsilon^{*2})\right]^{-1} = \sigma_\varepsilon^{*2}\begin{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} e^{2\overset{*}{\beta}X_i} & \overset{*}{\alpha}\sum_{i=1}^{n} X_i e^{2\overset{*}{\beta}X_i} \\ \overset{*}{\alpha}\sum_{i=1}^{n} X_i e^{2\overset{*}{\beta}X_i} & \overset{*}{\alpha}^2\sum X_i^2 e^{2\overset{*}{\beta}X_i} \end{bmatrix}^{-1} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \frac{\sigma_\varepsilon^{*2}}{n} \end{bmatrix}$$

III.d Parametrización,

Ahora veremos la aplicación del método de máxima verosimilitud aplicado al problema de la parametrización de la variable dependiente en un modelo no lineal. Esta supone un cambio de variable[12]. Veamos la función de producción propuesto por Zellner et al (1970) y visto en la página 2:

$\ln Y_t + \theta Y_t = \ln\alpha + \beta(1-\delta)C_t + \beta\delta\ln T_t + \varepsilon_t$ Esta función puede escribirse como la suma: $g(Y_i;\theta) = f(X_i;\beta) + \varepsilon_i$

Donde: $LnY_i + \theta Y_i = g(Y_i;\theta)$ y $f(X_i;\alpha,\beta,\delta) = Ln\alpha + \beta(1-\delta)LnC_i + \beta\delta LnT_i$

La función de densidad de Y_i asumiendo que las perturbaciones aleatorias ε_i se distribuyen según una normal es:

$$f_Y(y_i, g(.), f(.)) = \left|\partial g(Y_i,\theta)/\partial Y_i\right|(2\pi\sigma^2)^{-1/2}\exp -\frac{1}{2\sigma^2}[(g(Y_i,\theta) - f(X_i,\alpha,\beta,\delta)]^2$$

La función de verosimilitud es:

$$L = \prod_{i=1}^{n} f_Y(y_i, g(.), f(.)) = \prod_{i=1}^{n}\{\left|\partial g(Y_i,\theta)/\partial Y_i\right|(2\pi\sigma^2)^{-1/2}\exp -\frac{1}{2\sigma^2}\left[g(Y_i,\theta) - f(X_i,\alpha,\beta,\delta)\right]^2\}$$

La función log-verosimilitud es:

$$LnL = \sum_{i=1}^{n}\ln f_Y(y_i, g(.), f(.)) = \sum_{i=1}^{n}\ln\left|\partial g(Y_i,\theta)/\partial Y_i\right| + \ln(2\pi\sigma^2)^{-n/2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[g(Y_i,\theta) - f(X_i,\alpha,\beta,\delta)\right]^2$$

---

[12] El cambio de variable consiste en el siguiente problema: Supongamos que tenemos dos variables aleatorias: (X_1, X_2) con función de densidad: $f_{X1,X2}(x_1,x_2;\Theta)$, consideremos una transformación biunívoca y continua dada por: $Z = h_1(X_1,X_2)$ y $W = h_2(X_1,X_2)$ y supongamos que las funciones h_1 y h_2 son continuas y diferenciables en $S \subseteq R^2$ tal que existe el jacobiano: $J = \begin{vmatrix} \partial Z/\partial X_1 & \partial Z/\partial X_2 \\ \partial W/\partial X_1 & \partial W/\partial X_2 \end{vmatrix}$, supongamos además, que existe la transformación inversa dada por: $X_1 = g_1(x_1,z)$ y $X_2 = g_2(x_2,w)$, entonces:

$$P(a \le X_1 \le b; c \le X_2 \le d) = \int_a^b\int_c^d f_{X1,X2}(x_1,x_2;\Theta)dx_1dx_2 =$$

$$\iint_S f\left[g_1(x_1,z)g_2(x_2,w)\right]\left|J\right|dzdw$$

Aplicando la condición necesaria para la existencia de un punto extremo obtenemos:

$$\partial LnL/\partial \overset{*}{\alpha} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left[g(Y_i,\theta) - f(X_i,\overset{*}{\alpha},\beta,\delta)\right]\left(1/\overset{*}{\alpha}\right) = 0$$

$$\partial LnL/\partial \overset{*}{\beta} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left[g(Y_i,\theta) - f(X_i,\overset{*}{\alpha},\overset{*}{\beta},\delta)\right]\left[(1-\overset{*}{\delta})LnC_i + \overset{*}{\delta}\,LnT_i\right] = 0$$

$$\partial LnL/\partial \overset{*}{\delta} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\left[g(Y_i,\theta) - f(X_i,\overset{*}{\alpha},\overset{*}{\beta},\overset{*}{\delta})\right]\left[-\overset{*}{\beta}\,LnC_i + \overset{*}{\beta}\,LnT_i\right] = 0$$

Ahora para obtener el estimador de θ partimos de:

$$\partial LnL/\partial\theta == \sum_{i=1}^{n}\partial \ln\left|\partial g(Y_i,\theta)/\partial Y_I\right|/\partial\theta + -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[g(Y_i,\theta) - f(X_i,\alpha,\beta,\delta)\right]\partial g(Y_i;\theta)/\partial\theta$$

Donde: $\left|\partial g(Y_i;\theta)/\partial Y_i\right| = (1/Y_i + \theta)$, $\partial \ln\left|\partial g(Y_i;\theta)/\partial Y_i\right|\partial\theta = 1/(1/Y_i + \theta)$ y $\partial g(Y_i;\theta)/\partial\theta = 1$.

Luego:

$$\partial LnL/\partial \overset{*}{\theta} == \sum_{i=1}^{n}(Y_i/(1+Y_i\,\overset{*}{\theta})) + -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[g(Y_i,\overset{*}{\theta}) - f(X_i,\alpha,\beta,\delta)\right] = 0$$

Finalmente obtenemos el estimador de σ^2:

$$\partial LnL/\partial \overset{*}{\sigma^2} = -n/2\overset{*}{\sigma^2} + -\frac{1}{2\overset{*}{\sigma^4}}\sum_{i=1}^{n}\left[g(Y_i,\theta) - f(X_i,\alpha,\beta,\delta)\right]^2 = 0$$

Nuevamente estamos en presencia de un sistema de ecuaciones no lineales lo que obliga emplear algunos de los algoritmos propuestos para este tipo de problemas.

Ejemplo 7 con SPSS.

Consideremos que el índice de precio al mayorista *IPM* se puede explicar por la tasa de cambio *TCN* , la masa monetaria *M*2*M* y el tiempo *t* , el modelo es:

$$IPM = \beta_0 M2M^{\beta_1} TCN^{\beta_2} e^{\beta_3 TIEMPO + \varepsilon}$$

Se cuenta con una base de datos mensual desde Enero de 1980 a Diciembre del 2002. Para resolver el problema se empleó el software SPSS . El paquete exige que se ingrese un valor inicial para cada parámetro (Ver apéndice SPSS). Estos valores se van cambiando buscando mejorar el modelo, tomando como criterio que el cuadrado medio de la regresión mejore y que se estabilicen las estimaciones de los parámetros. Para aclarar esto veamos el siguiente cuadro:

RESULTADOS DE CINCO PRUEBAS
CAMBIANDO LOS VALORES INICIALES

Beta0	0,5	0,5	0,01	0,5	0,5
Beta1	0,4	0,4	0,4	0,4	0,1
Beta2	0,5	0,5	0,5	0,5	0,2
Beta3	1,0	0,5	0,5	0,05	0,1
SCR/gl	-289E+38	120988,606	120812	550540,449	550540,49
SCE/gl	4,253E+36	6327,17	6329	10,23	10,233
R^2	-	-	-	0,998	0,998

El cuadro muestra los valores iniciales dados a los parámetros en cinco pruebas, se podrá notar que las dos últimas son más convincentes. En general, puede ocurrir que la solución óptima no sea única o que se obtenga un óptimo local. A continuación presentamos dos resultados seleccionados. El primero, los valores iniciales de los parámetros fueron beta0=0,05; beta1=0,2; beta2=0,1; beta3=0,1. Para el segundo estos valores fueron: beta0=0,05; beta1=0,01; beta2=0,01; beta3=0,01

Estimaciones de los parámetros

Parámetro	Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior
beta0	,003	,001	,002	,005
bata1	,383	,016	,351	,414
beta2	,657	,012	,633	,681
beta3	,001	,000	,000	,002

Estimaciones de los parámetros

Parámetro	Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior
beta0	,003	,001	,002	,005
bata1	,383	,016	,351	,414
beta2	,657	,012	,633	,681
beta3	,001	,000	,000	,002

Se puede observar que las estimaciones obtenidas son idénticas. El modelo ajustado es:

$$IPM^{*} = 0{,}03M2M^{0{,}383}TCN^{0{,}657}e^{0{,}001TIEMPO}$$

Además de lo indicado para seleccionar la estimación del modelo, debe tenerse presente los supuestos teóricos del área donde se está aplicando, puesto que puede ocurrir que las estimaciones por más que luzcan satisfactorias, debe verificarse si no contradice algún supuesto importante, propio del fenómeno que se quiere modelar.

Correlaciones de las estimaciones de los parámetros

	beta0	bata1	beta2	beta3
beta0	1,000	-,983	-,265	,829
bata1	-,983	1,000	,123	-,790
beta2	-,265	,123	1,000	-,652
beta3	,829	-,790	-,652	1,000

ANOVA[a]

Origen	Suma de cuadrados	gl	Medias cuadráticas
Regresión	2202161,8	4	550540,449
Residual	2783,301	272	10,233
Total sin corrección	2204945,1	276	
Total corregido	1494274,2	275	

Variable dependiente: IPMG

a. R cuadrado = 1 - (Suma de cuadrados residual) / (Suma corregida de cuadrados) = ,998.

En los modelos no lineales el coeficiente de determinación no tiene la misma connotación que en los lineales, sin embargo se puede tomar para seleccionar el mejor ajuste después de elegir los valores iniciales. De acuerdo al valor obtenido de $R^2 = 0{,}998$ podemos pensar que el modelo se adecua bien a los datos. La suma de cuadrados del residual lo emplearemos para hacer contraste de hipótesis.

IIIe.-Transformación de Box-Cox.

La transformación de Box Cox consiste en lo siguiente. Asumamos el modelo dado por:

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j f(X_j) + \varepsilon$$

Cada regresor $f(X_j)$ se modifica por; $Z_j = \left(X_j^{\lambda_0} - 1\right)/\lambda_0$, donde $\lambda \in [-1,1]$ si

$\lambda = -1; f(X_j) = \left(X_j^{-1} - 1\right)/(-1) = -\frac{1}{X_j} + 1$, esto nos conduce a un modelo de regresión lineal clásico, donde hay que tomar como regresores los inversos.

Si

$\lambda = 1; f(X_j) = \left(X_j^{1} - 1\right)/1$

Si

$\lambda = 0; \lim_{\lambda \to 0} f(X_j) = \lim_{\lambda \to 0}\left(X_j^{\lambda} - 1\right)/\lambda = 0/0$. Esta indeterminación se resuelve mediante la regla de L´ Hospital, obteniendo lo siguiente:

$\lim_{\lambda \to 0} df(X_j)/d\lambda = \lim_{\lambda \to 0} d\left(X_j^{\lambda} - 1\right)/d\lambda/(d\lambda/d\lambda) = \lim_{\lambda \to 0} 1.X_j^{\lambda} \log X_j = LogX_j$

En los tres casos podemos escribir el modelo como:

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j Z_j + \varepsilon$$

Si se asume que la variable a explicar tiene la forma $g(Y) = (Y^{\lambda} - 1)/\lambda$ y procedemos de la misma forma obtenemos, para el caso de $\lambda = 0$ el modelo log-lineal dado por:

$$LogY = \beta_0 + \sum_{j=1}^{p} LogX_j + \varepsilon$$

En la practica se puede variar el valor del parámetro λ y mediante el empleo de los mínimos cuadrados ordinarios encontrar el modelo que mejor se ajusta a los datos, sin embargo como anota Novales (1992) el proceder de esta forma afecta la estimación de la matriz de varianza covarianza. Un tratamiento más amplio de esta transformación se encuentra en Greene (1999).

IV.-Contrastes Restringidos.

Tanto para modelos lineales como no lineales se proponen hipótesis restrictiva con respecto a los parámetros. Puede darse uno de estos casos: a) el modelo es lineal y el conjunto de restricciones también es lineal, este un caso frecuente en diseño de experimento b) el modelo es lineal y el conjunto de restricciones al menos una restricción es no lineal c) el modelo es no lineal y el conjunto de restricciones es lineal, un ejemplo el modelo de producción de Cobb-Douglas donde se establece una o dos restricciones lineales sobre los parámetros d) el modelo y las restricciones son no lineales.

Como ejemplo del primer caso tenemos el siguiente modelo de consumo en función de la renta con dos rezagos:

$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \varepsilon_t$

Donde Y_t es el consumo en el período t, X_t y X_{t-1} corresponden a las rentas del período t y $t-1$ respectivamente, ε_t es la perturbación aleatoria que asumimos normal y $\beta_0, \beta_1, \beta_2$ son los parámetros. Como puede observarse el modelo es lineal. Las restricciones pueden ser: $\beta_1 \geq \beta_2$ y $\beta_1 + \beta_2 = 1$, la primera restricción podemos escribirla como: $\beta_1 - \beta_2 - h = 0$

Esto último permite escribir las restricciones en forma matricial como:

$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ h \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Entonces, tenemos un modelo lineal con restricciones lineales.

Del segundo caso tenemos el modelo que es el mismo que el anterior pero con restricciones no lineales:

$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \varepsilon_t$

El conjunto de restricciones ahora es:

$\beta_1^2 + \beta_2^2 = 1$ y $\beta_1 . \beta_2 = 0$

En el caso el modelo es no lineal y el conjunto de restricciones lineales tenemos el ejemplo de la función de consumo dada por:

$C_t = \beta_0 + \beta_1 R_t^{\lambda} + \beta_2 C_{t-1} + \varepsilon_t$

Donde C_t es el consumo del período t, C_{t-1} es el consumo del período anterior y R_t es la renta. Las restricciones son:

$\beta_1 / (1 - \beta_2) = 1$

$\lambda = 1$

Por último tenemos como ejemplo de un modelo y restricciones no lineales el siguiente:

$C_t = \beta_0 + \beta_1 R_t^{\lambda} + \beta_2 C_{t-1} + \varepsilon_t$

$\beta_1^2 + \beta_2^2 = 1$y $\beta_1 . \beta_2 = 0$

Entonces, en forma general el problema es: dado el modelo $Y = G(X;\beta) + \varepsilon$ realizar el contraste de hipótesis: $H_0 : R\beta = r$ donde R es una matriz en el caso de ser las restricciones lineales y r el vector de valores o $H_0 : R(\beta) = r$ donde $R(\beta)$ es un conjunto de ecuaciones no lineales. En este último caso no se verifica necesariamente que: $ER(\overset{*}{\beta}) = R(E(\overset{*}{\beta}))$ pero si se demuestra la consistencia, esto es: $P\lim_{n\to\infty} R(\overset{*}{\beta}) = R(P\lim_{n\to\infty} \overset{*}{\beta})$.

Para realizar el contraste: $H_0 : R\beta = r$ ó $H_0 : R(\beta) = r$ hay varios test: el uso del estadístico F cuando se tiene una muestra suficientemente grande, en teoría para muestras infinitas, el contraste de Wald (W) el cuál solo requiere obtener el estimador del vector de parámetro sino considerar la restricción formulada como hipótesis, el método de la razón máxima verosimilitud (MV)que implica el cálculo del estimador tanto del modelo no restringido, como del modelo con las restricciones formuladas en la hipótesis y finalmente, el contraste del multiplicador de Lagrange o prueba de C.R Rao (ML).

Podemos empezar con el caso más sencillo: dado el modelo lineal $Y = X\beta + \varepsilon$ y la hipótesis $H_0 : R\beta = r$. El estimador mínimo cuadrático sin considerar las restricciones es: $\beta^* = (X^T X)^{-1} X^T Y$. Si se trata de una sola restricción, esto es: R es un vector fila: v, y asumiendo que las perturbaciones aleatorias son normales: $\varepsilon_t \overset{d}{\approx} N(0, \sigma^2)$ entonces para medir la discrepancia entre lo observado y la hipótesis nula, podemos emplear la prueba t: que bajo la hipótesis nula es:

$$t = (v\overset{*}{\beta} - v\beta) / \sqrt{(v\ (\sigma_\varepsilon^{*2}(X^T X)^{-1})v^T} = (v\overset{*}{\beta} - r) / \sqrt{(v\ (\sigma_\varepsilon^{*2}(X^T X)^{-1})v^T}$$

En el caso de que R es una matriz de l filas, esto es: l restricciones lineales entonces podemos emplear la prueba F para tamaño de muestras suficientemente grandes.

$$F = (n-k)\left(R\overset{*}{\beta} - r\right)^T \left(R(X^T X)^{-1} R^T\right)^{-1} \left(R - r\right) / l\varepsilon^{*T}\varepsilon$$

Bajo la hipótesis nula el estadístico se distribuye como una $F_{l,;n-k}$. En cualquiera de los dos casos, bajo la hipótesis nula, la esperanza de la discrepancia d debe ser igual al vector nulo, esto es en forma general:

$$E(d) = E(R\overset{*}{\beta} - R\beta) = 0_l$$

Si se desea obtener el estimador del modelo restringido, entonces el problema es:
$\underset{\beta}{Min} \|Y - X\beta\|_{p=2}$
Sujeto a:
$R\beta - r = 0_l$
Empleando los multiplicadores de Lagrange[13] obtenemos:

[13] Es una técnica que permite resolver problemas de programación no lineal restringido, esto es, dada una función: $F(x_1, x_2, ...x_n)$ la cual, por ejemplo, quiere minimizar y, un conjunto de restricciones: $G_i(x_1, x_2, ...x_n) - b_i = 0_n$. Para $i = 1,2...m$ definimos entonces una nueva función llamada lagrangiana dada por:

$$L(x_1, x_2, ...x_n; \lambda_1, \lambda_2, ...\lambda_m) = F(x_1, x_2, ...x_n) + \sum_{i=1}^{m} \lambda_i (G_i(x_1, x_2, ...x_n) - b_i)$$

Si las restricciones fuesen del tipo $G_i(x_1, x_2, ...x_n) - b_i \geq 0$, se restaría a cada restricción una nueva variable no negativas llamadas variables de holguras h_i obteniendo lo siguiente:

$$L(x_1, x_2, ...x_n; \lambda_1, \lambda_2, ...\lambda_m) = F(x_1, x_2, ...x_n) + \sum_{i=1}^{m} \lambda_i (G_i(x_1, x_2, ...x_n) - b_i - h_i)$$

$$\underset{\beta,\lambda}{Min}$$

$$L(\beta,\lambda) = (Y - X\beta)^T (Y - X\beta) + 2\lambda(R\beta - r)$$

Las condiciones necesarias para la existencia de un punto extremo son:

$$\partial L(\overset{*}{\beta},\lambda)/\partial\beta = 2X^T(Y - X\overset{*}{\beta}_{ML}) + 2\lambda R^T = 0_{p+1}$$

$$\partial L(\overset{*}{\beta},\overset{*}{\lambda})/\partial\lambda = 2(R\overset{*}{\beta}_{ML} - r) = 0_l$$

De estas ecuaciones normales obtenemos:

$$\begin{bmatrix} X^T X & R^T \\ R & 0 \end{bmatrix}\begin{bmatrix} \overset{*}{\beta}_{ml} \\ \overset{*}{\lambda} \end{bmatrix} = \begin{bmatrix} X^T Y \\ r \end{bmatrix}$$

Simplificando la notación:

$B\overset{*}{\gamma} = b$ de donde: $\overset{*}{\gamma} = B^{-1}b$. La suma de los cuadrados de los errores para un modelo sin restricciones lo expresamos como:

$$SCE = \left\| Y - X\overset{*}{\beta} \right\|_{p=2}$$

Y la suma de cuadrados de los errores del modelo restringido lo a anotamos como:

$$SCE^* = \left\| Y - X\overset{*}{\beta}_{ML} \right\|_{p=2}$$

La variabilidad de Y recogida por el modelo restringido está limitada por el conjunto de restricciones, por tanto se verifica que:

$SCE^* \geq SCE$

En efecto podemos descomponer el vector de los residuos del modelo restringido como sigue:

$$e_R^* = Y - X\overset{*}{\beta}_{ML} = Y - X\overset{*}{\beta} + X(\overset{*}{\beta} - \overset{*}{\beta}_{ML}) = \overset{*}{e} + X(\overset{*}{\beta} - \overset{*}{\beta}_{ML})$$

$$e_R^{*T} e_R^* \geq \overset{*}{e}^T \overset{*}{e}$$

La igualdad se cumple si $\overset{*}{\beta}_{ML} = \overset{*}{\beta}$.

Cuando el modelo es no lineal y las restricciones son lineales el procedimiento puede partir de la linealización del modelo y proceder de forma similar como se ha indicado para hacer el contraste, el problema se presenta en la situación de no linealidad de las restricciones.

IVa.-Prueba F asintótica.

Tomando en cuenta la relación $SCE^* \geq SCE$, se pude construir el contraste F tomando en cuenta que el tamaño de la muestra es suficientemente grande como para que se justifique la aplicación de este test desde el punto de vista asintótico, teóricamente esto significa que se hace tan grande como se quiera:

Esta nueva función es la que emplearemos para encontrar la solución óptima. Este método tiene validez cuando el número de restricciones m es menor al número de variables n.

$$\partial L(x_1, x_2, ...x_n; \lambda_1, \lambda_2, ...\lambda_m)/\partial x_j = 0;\ j = 1,2..n; \forall x_j = x_j^*$$

$$\partial L(x_1, x_2, ...x_n; \lambda_1, \lambda_2, ...\lambda_m)/\partial \lambda_j = 0;\ j = 1,2..m; \forall \lambda_j = \lambda_i^*$$

$F = (n-k)(SCE^* - SCE)/lSCE$

Como se recordará p es el número de regresores, n es el número de observaciones y l el número de restricciones. Bajo la hipótesis nula $H_0 : R(\beta) = r$, el estadístico F converge asintóticamente a una distribución $F_{l;n-p}$. La hipótesis nula se rechaza si: $P(F_{l,n-p} \geq F^*) \leq \alpha$, siendo α el nivel de significación.

IVb.-Prueba de Wald.

Esta prueba se usa indistintamente en el caso del modelo lineal o no lineal, la prueba se construye tomando en cuenta la discrepancia: $R(\overset{*}{\beta}) - r$ y su varianza, esto es:

$$W = (R(\overset{*}{\beta}) - r)^T \left[Var(R(\overset{*}{\beta}) - r) \right]^{-1} (R(\overset{*}{\beta}) - r)$$

En el caso de un modelo lineal con restricciones lineales, esto es: $Y = X\beta + \varepsilon$, $H_0 : R\beta = r$, tenemos que: $\overset{*}{\beta} = (X^T X)^{-1} X^T Y$, y asumiendo que ε tiene una distribución $N(0, \sigma^2 I)$ la estimación de la matriz de varianza covariaza es $\overset{*}{Var}(\overset{*}{\beta}) = \sigma_\varepsilon^{*2} (X^T X)^{-1}$ luego $\overset{*}{Var}(R\overset{*}{\beta} - r) = R\ \overset{*}{Var}(\overset{*}{\beta}) R^T = \sigma_\varepsilon^{*2} R(X^T X)^{-1} R^T$ por tanto el contraste es:

$$W = \sigma_\varepsilon^{*2} (R\overset{*}{\beta} - r)^T (R(X^T X)^{-1} R^T)^{-1} (R\overset{*}{\beta} - r)$$

Este estadístico se distribuye como una χ_l^2 con tantos grados de libertad como restricciones estén presentes en: $H_0 : R\beta = r$.

Cuando el modelo es lineal: $Y = X\beta + \varepsilon$ y las restricciones no lineales: $H_0 : R(\beta) = r$, una vez obtenido el estimador: $\overset{*}{\beta} = (X^T X)^{-1} X^T Y$ debemos obtener: $Var(R(\overset{*}{\beta}) - r)$, en este caso empleamos en primer lugar la expansión de Taylor como una aproximación de $R(\overset{*}{\beta})$ dada por:

$$R(\overset{*}{\beta}) = R(\beta) + \left[\partial R(\beta^*) / \partial \beta \right] (\overset{*}{\beta} - \beta)$$

Donde $\left[\partial R(\beta) / \partial \beta \right]$ es una matriz de dimensión: lxk, la varianza $Var(R(\overset{*}{\beta}) - r)$ es ahora:

$$Var(R(\overset{*}{\beta})) = \left[\partial R(\beta^*) / \partial \beta \right] Var(\overset{*}{\beta} - \beta) \left[\partial R(\beta^*) / \partial \beta \right]^T$$

La estimación de la matriz de varianza covarianza es:

$$\overset{*}{Var}(R(\overset{*}{\beta})) = \sigma_\varepsilon^{*2} \left[\partial R(\beta^*) / \partial \beta \right] (X^T X)^{-1} \left[\partial R(\beta^*) / \partial \beta \right]^T$$

Ahora suponemos que el modelo es no lineal $Y = G(X;\beta) + \varepsilon$ y las restricciones lineales: $H_0 : R\beta = r$. El estimador $\overset{*}{\beta}$ se obtiene empleando el método de máxima verosimilitud, asumiendo que ε tiene una distribución $N(0, \sigma^2 I)$. Por tanto:

$$Var(R\overset{*}{\beta} - r) = R Var(\overset{*}{\beta}) R^T = R\left[I(\beta) \right]^{-1} R^T$$

Cuya estimación es:

$$\overset{*}{Var}(R\overset{*}{\beta} - r) = R\overset{*}{Var}(\overset{*}{\beta}) R^T = R\left[I(\overset{*}{\beta}) \right]^{-1} R^T$$

Finalmente si el modelo y el conjunto de restricciones es no lineal: $Y = G(X;\beta) + \varepsilon$ **,** $H_0 : R(\beta) = r$ entonces, si se emplea el estimador máximo verosímil obtenemos:

$$Var(R(\overset{*}{\beta})) = \left[\partial R(\beta)/\partial\beta\right]\left[I(\beta)\right]^{-1}\left[\partial R(\beta)/\partial\beta\right]^T$$

De forma general el estadístico tiene la forma en el caso de restricción lineal:

$$W = \left[R\overset{*}{\beta} - r\right]\left[Var(R\overset{*}{\beta})\right]^{-1}\left[R\overset{*}{\beta} - r\right]$$

Y en el caso de restricción no lineal:

$$W = \left[\partial R(\overset{*}{\beta})/\partial\beta\right]\left[Var(R(\overset{*}{\beta}))\right]^{-1}\left[\partial R(\overset{*}{\beta})/\beta\right]$$

W sigue, bajo la hipótesis nula una distribución χ_l^2. La hipótesis nula se rechaza si: $P(\chi_l^2 \geq W^* | H_0) \leq \alpha$

Ejemplo 8. $Y_t = \alpha e^{\beta X_t} + \varepsilon_t$ donde ε_t es una variable aleatoria $N(0,\sigma^2)$ y supongamos como hipótesis $H_0 = \alpha^2 + \beta^2 = 1$, luego $R(\beta) = \alpha^2 + \beta^2$ y $r = 1$Asumamos que los estimadores son de máxima verosimilitud, Como se vio en el ejemplo 6:

$$\left[I(\overset{*}{\alpha},\overset{*}{\beta},\sigma_\varepsilon^{*2})\right]^{-1} = \sigma_\varepsilon^{*2}\begin{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} e^{2\overset{*}{\beta}X_i} & \overset{*}{\alpha}\sum_{i=1}^{n} X_i e^{2\overset{*}{\beta}X_i} \\ \overset{*}{\alpha}\sum_{i=1}^{n} X_i e^{2\overset{*}{\beta}X_i} & \overset{*}{\alpha}^2 \sum X_i^2 e^{2\overset{*}{\beta}X_i} \end{bmatrix}^{-1} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \frac{\sigma_\varepsilon^{*2}}{n} \end{bmatrix}$$

$$\left[I(\overset{*}{\alpha};\overset{*}{\beta}\right]^{-1} = \overset{*}{\sigma}_\varepsilon^{\,2}\begin{bmatrix} \sum_{t=1}^{n} e^{2\overset{*}{\beta}X_t} & \overset{*}{\alpha}\sum_{t=1}^{n} X_t e^{2\overset{*}{\beta}X_t} \\ \overset{*}{\alpha}\sum_{t=1}^{n} X_t e^{2\overset{*}{\beta}X_t} & \overset{*}{\alpha}^2 \sum_{t=1}^{n} X_t^2 e^{2\overset{*}{\beta}X_t} \end{bmatrix}^{-1}$$

Donde**:** $\partial R(\overset{*}{\alpha})/\partial\alpha = 2\overset{*}{\alpha}$ $\partial R(\overset{*}{\beta})/\partial\beta = 2\overset{*}{\beta}$ por tanto:

$$W = \overset{*}{\sigma}_\varepsilon^{\,2}\left[2\overset{*}{\alpha}, 2\overset{*}{\beta}\right]\begin{bmatrix} \sum_{t=1}^{n} e^{2\overset{*}{\beta}X_t} & \overset{*}{\alpha}\sum_{t=1}^{n} X_t e^{2\overset{*}{\beta}X_t} \\ \overset{*}{\alpha}\sum_{t=1}^{n} X_t e^{2\overset{*}{\beta}X_t} & \overset{*}{\alpha}^2 \sum_{t=1}^{n} X_t^2 e^{2\overset{*}{\beta}X_t} \end{bmatrix}^{-1}\left[2\overset{*}{\alpha}, 2\overset{*}{\beta}\right]^T$$

Dependiendo del valor de W, rechazamos o no la hipótesis.

IVc.-Contraste de los multiplicadores de Lagrange (ML).

Como sabemos siempre se cumple que $SCE^* \geq SCE$, por tanto el contraste partirá de esta relación. El método no cambia sustancialmente si el modelo es o no lineal, el cambio se nota cuando: $H_0 : R\beta = r$ o $H_0 : R(\beta) = r$**,** esto es: si las restricciones son

lineales o no. Entonces, consideremos el modelo $Y = G(X;\beta) + \varepsilon$ bajo la restricción: $H_0 : R\beta = r$

El problema es:

$$\underset{\beta,\lambda}{Max}\, LnL(\beta;\lambda)_R = \underset{\beta,\lambda}{Max}\left[LnL(\beta;\lambda) + \lambda^T (R\beta - r)\right]$$

Las condiciones necesarias para la existencia de un punto extremo son:

$$\partial LnL(\overset{*}{\beta}_R;\lambda)_R / \partial \beta = \left[\partial LnL(\overset{*}{\beta};\lambda)/\partial\beta + R^T\lambda\right] = 0_p$$

$$\partial LnL(\overset{*}{\beta}_R;\overset{*}{\lambda})_R / \partial \lambda = \left[R\beta - r\right] = 0_l$$

Si la hipótesis nula es cierta entonces $H_0 : R\beta = r$, entonces: $R^T\lambda$ debe ser muy próximo al vector nulo, por tanto:

$$\partial LnL(\overset{*}{\beta}_R;\lambda)_R / \partial \beta \cong \left[\partial LnL(\overset{*}{\beta};\lambda)/\partial\beta\right] \cong 0_p$$

Es decir las dos funciones de máxima verosimilitud, la del modelo restringido y la del modelo sin restricción deben estar próximas. El estadístico tiene la forma

$$ML = n(\partial LnL(\overset{*}{\beta}_R;\lambda)_R / \partial\beta)^T I(\overset{*}{\beta}_R)^{-1} (\partial LnL(\overset{*}{\beta}_R;\lambda)_R / \partial\beta) / e_R^{*T} e_R^* \quad (1)$$

En el caso del contraste para la restricción no lineal: $H_0 : R(\beta) = r$ el problema es:

$$\underset{\beta,\lambda}{Max}\, LnL(\beta;\lambda)_R = \underset{\beta,\lambda}{Max}\left[LnL(\beta;\lambda) + \lambda^T (R(\beta) - r)\right]$$

$$\partial LnL(\overset{*}{\beta}_R;\lambda)_R / \partial \beta = \left[\partial LnL(\overset{*}{\beta};\lambda)/\partial\beta + \lambda^T \partial R(\overset{*}{\beta})/\partial\beta\right] = 0_p$$

$$\partial LnL(\overset{*}{\beta}_R;\overset{*}{\lambda})_R / \partial \lambda = \left[R(\overset{*}{\beta}) - r\right] = 0_l$$

Si $H_0 : R(\beta) = r$ es cierta entonces la discrepancia $R(\overset{*}{\beta}) - r$ debe estar muy próxima al vector nulo y por tanto $\lambda^T \partial R(\beta)/\partial\beta$ debe ser nulo luego, el estadístico nuevamente es igual que el dado en (1). Este estadístico tiende asintóticamente a $\chi^2_{lgl}.\chi^2_l$. La hipótesis nula se rechaza si: $P(\chi^2_l \geq ML^* | H_0) \leq \alpha$

Ejemplo 9. Consideremos el modelo lineal $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ con las restricciones $H_0 : \beta_1^2 + \beta_2 = 1$. (El desarrollo se deja al lector)

IVd.-Contraste de la razón de verosimilitud[14].

Considerando la nota explicativa 14 dada al pie de página sabemos que en forma general $\lambda_n = \sup_{\theta \in \Omega_0} L(x_1, x_2 ... x_n; \theta) / \sup_{\theta \in \Omega} L(x_1, x_2 ... x_n; \theta \leq 1$ Luego para cualquier tipo de modelo o de restricciones (lineales o no) se verifica que el logaritmo de la función de verosimilitud asociada al modelo restringido es mayor que el logaritmo de la función de verosimilitud asociada al modelo libre de restricciones. Esto es:

$LnL(\overset{*}{\beta}_R) \geq LnL(\overset{*}{\beta})$

Donde:

$$LnL(\beta) = Jacoviano - \frac{n}{2} Ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \varepsilon^T \varepsilon$$

El estimador Máximo verosímil $\sigma^{*2} = e^{*T} e / n$

$$LnL(\overset{*}{\beta}) = Jacoviano - \frac{n}{2} Ln(2\pi) - \frac{n}{2} \ln \sigma^{*2} - \frac{n}{2}$$

$$LnL(\overset{*}{\beta}_R) = Jacoviano - \frac{n}{2} Ln(2\pi) - \frac{n}{2} \ln \sigma_R^{*2} - \frac{n}{2}$$

El estadístico de la razón de verosimilitud es:

$$RMV = LnL(\overset{*}{\beta}_R) - LnL(\overset{*}{\beta}) = \frac{n}{2} (Ln\sigma_R^{*2} - Ln\sigma^{*2})$$

Bajo la Hipótesis nula $2RMV$ se distribuye asintóticamente como una χ_l^2. La hipótesis nula se rechaza si: $P(\chi_l^2 \geq RMV^* | H_0) \leq \alpha$

Ejemplo 10.

Retomemos el ejemplo 7 y consideremos ahora el modelo del IPM dado por:

$$IPM = \beta_0 M2M^{\beta_1} TCN^{\beta_2} e^{\beta_3 TIEMPO + \varepsilon}$$

Y consideremos el siguiente conjunto de restricciones como H_0: $\beta_1 + \beta_2 = 1$, $\beta_1 \geq 0$ y $\beta_2 \geq 0$. Ahora los resultados usando SPSS son:

Estimaciones de los parámetros

Parámetro	Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior
beta0	,176	3,9E+013	-7,674E+013	8E+013
beta1	,402	1,9E+009	-3704114643	4E+009
beta2	,598	1,5E+014	-2,985E+014	3E+014
beta3	-4,029	29217483	-57521159,6	6E+007

[14] En general, el contraste de razón de verosimilitud consiste en lo siguiente. Consideremos la hipótesis nula $H_0 : \theta \in \Omega_0$ y la hipótesis alterna $H_1 : \theta \in \Omega - \Omega_0$ y sea la función de verosimilitud $L(x_1, x_2 ... x_n; \theta)$ asociada a una muestra aleatoria $X_1, X_2 ... X_n$ proveniente de una población $F_X(x, \Omega_0)$. Se define la razón de verosimilitud a: $\lambda_n = \sup_{\theta \in \Omega_0} L(x_1, x_2 ... x_n; \theta) / \sup_{\theta \in \Omega} L(x_1, x_2 ... x_n; \theta)$. El test consiste en elegir una región crítica $W \subseteq R^n$, tal que para un valor $\partial \in [0,1]$, $\lambda_n \leq \partial$ para todo $X \in S$ y $\lambda_n \geq \partial$ para todo $X \notin S$. $\partial = \sup \{X \in S | \theta \in \Omega_0\} \leq \alpha$. Se rechaza la hipótesis nula si $\lambda_n \leq \partial$ 0 equivalentemente si $X \in S$.

Correlaciones de las estimaciones de los parámetros

	beta0	beta1	beta2	beta3
beta0	1,000	,998	-1,000	-,997
beta1	,998	1,000	-,998	-1,000
beta2	-1,000	-,998	1,000	,997
beta3	-,997	-1,000	,997	1,000

ANOVA[a]

Origen	Suma de cuadrados	gl	Medias cuadráticas
Regresión	,501	3	,167
Residual	2204944,6	273	8076,720
Total sin corrección	2204945,1	276	
Total corregido	1494274,2	275	

Variable dependiente: IPMG

a. R cuadrado = 1 - (Suma de cuadrados residual) / (Suma corregida de cuadrados) = ..

Contratemos las hipótesis empleando el estadístico $2RMV$ y tomando los datos de la matriz de ANOVA del ejemplo 7 referente al residual y de la misma forma, el residual de la matriz ANOVA de este ejemplo, obtenemos:

$$2RMV = 2(LnL(\overset{*}{\beta}_R) - LnL(\overset{*}{\beta})) = n(Ln\sigma_R^{*2} - Ln\sigma^{*2}) = nLn(\frac{\sigma_R^{*2}}{\sigma^{*2}})$$

Esto es equivalente a $2RMV \cong nLn(SCE_R^* / SCE) = 275Ln(2.204.944,6/2783,301) = 1846,263$ $P(\chi^2_{3gl} \geq 1846,263) = 0$. Esto implica rechazar la hipótesis nula: que los estimadores obtenidos en el ejemplo 7 cumplen con todas las restricciones impuestas en dicha hipótesis .

Es importante tomar en cuenta que el SPSS permite guardar las derivadas de los parámetros evaluadas en cada observación. Con los cálculos necesarios se puede entonces emplear los contrastes que requieren de la matriz de información.

Ejemplo 11

Ahora redefinamos el modelo IPM como:

$IPM = \beta_0 M2M^{\beta_1} TCN^{\beta_2}$

Entonces la nueva solución es:

Estimaciones de los parámetros

Parámetro	Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior
beta0	,006	,001	,005	,008
beta1	,327	,011	,304	,349
beta2	,739	,011	,717	,760

Correlaciones de las estimaciones de los parámetros

	beta0	beta1	beta2
beta0	1,000	-,956	,655
beta1	-,956	1,000	-,847
beta2	,655	-,847	1,000

ANOVA[a]

Origen	Suma de cuadrados	gl	Medias cuadráticas
Regresión	2201115,4	3	733705,146
Residual	3829,658	273	14,028
Total sin corrección	2204945,1	276	
Total corregido	1494274,2	275	

Variable dependiente: IPMG

a. R cuadrado = 1 - (Suma de cuadrados residual) / (Suma corregida de cuadrados) = ,997.

Ahora consideremos las restricciones: $H_0\ \beta_1+\beta_2=1, \beta_1\geq 0$ y $\beta_2\geq 0$ los resultados son:

Estimaciones de los parámetros

Parámetro	Estimación	Error típico	Intervalo de confianza al 95%	
			Límite inferior	Límite superior
beta0	,007	,001	,005	,010
beta1	,356	,014	,328	,385
beta2	,644	,014	,616	,671

Correlaciones de las estimaciones de los parámetros

	beta0	beta1	beta2
beta0	1,000	-,956	,674
beta1	-,956	1,000	-,861
beta2	,674	-,861	1,000

ANOVA[a]

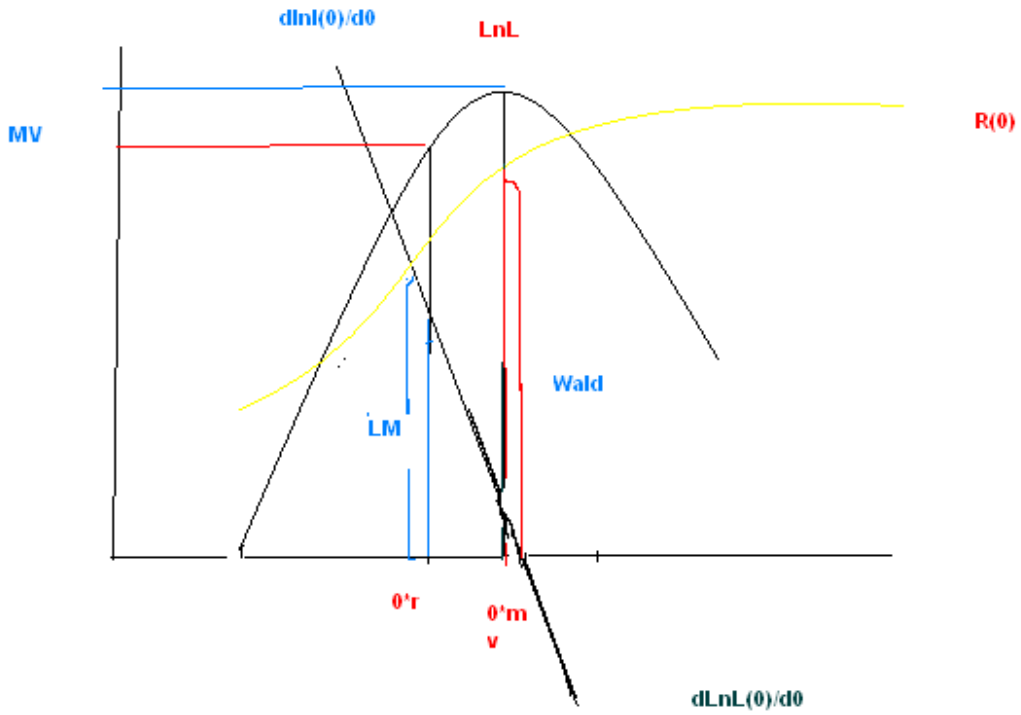
Origen	Suma de cuadrados	gl	Medias cuadráticas
Regresión	2198643,6	2	1099321,810
Residual	6301,476	274	22,998
Total sin corrección	2204945,1	276	
Total corregido	1494274,2	275	

Variable dependiente: IPMG

a. R cuadrado = 1 - (Suma de cuadrados residual) / (Suma corregida de cuadrados) = ,996.

La solución se deja al lector.

El siguiente gráfico muestra los tres últimos contrastes:



Se puede observar en el gráfico que la relación entre los tres contrastes es: el estadístico de Wald siempre dará un valor mayor a los otros dos contrastes auque los tres tienen las mismas propiedades asíntoticas, es decir: son equivalentes asintóticamente. El contraste de Wald solo requiere estimar mediante el método de máxima verosimilitud el parámetro del modelo sin restricciones, el contraste de los multiplicadores de Lagrange necesita solamente la estimación del parámetro del modelo reducido, mientras que el contraste de máxima verosimilitud parte de la estimación del parámetro del modelo sin restricciones y con restricciones requiriendo por tanto de dos estimadores.

V.-Contrastes de Hipótesis de linealidad:

Hay varios test que permiten decidir en términos estadístico si un fenómeno responde a un comportamiento no lineal, bien sea en el caso de un modelo de regresión o de serie de tiempo, algunos de estos test no están disponibles en las versiones de los software

más usados en estadística tales como el SAS o SPSS por ser de desarrollo muy recientes. En este punto veremos algunos de estos test, el primero está basado en la descomposición de la variación total, conduciendo a una prueba F el segundo es el test P_d, y finalmente el test propuesto por Samorov, A.; Spokoiny, V. ;Vial, C (2005).

Va.-Prueba F[15]

Partimos de la descomposición de la variación total en tres componentes, bajo la hipótesis nula que el modelo que mejor se adecua a los datos es un modelo de regresión lineal $H_0 : Y = X\beta + \varepsilon$. Además, asumimos que existen $E(Y/X)$, $E(Y)$, y el ajuste de Y dado como $\overset{*}{Y}$. Entonces la variación total de Y dada por: $\|Y - E(Y)\|_{p=2}$ la descomponemos como:

$$\|Y - E(Y)\|_{p=2} = \left\| (\overset{*}{Y} - E(Y)) + ((E(Y/X) - \overset{*}{Y}) + (Y - E(Y/X)) \right\|_{p=2}$$

$$\|Y - E(Y)\|_{p=2} = \left\| \overset{*}{Y} - E(Y) \right\|_{p=2} + \left\| E(Y/X) - \overset{*}{Y} \right\|_{p=2} + \|Y - E(Y/X)\|_{p=2}$$

De donde:

$$\phi^{*2} = R^{*2} + \|Y - E(Y/X)\|_{p=2} / \|Y - E(Y)\|_{p=2}$$

Si $\phi^{*2} - R^{*2} \geq 0$ indica el grado de error de especificación al asumir que el modelo es lineal.

La prueba F es:

$$F = (n-p)(\phi^{*2} - R^{*2}) / (p-2)(1 - R^{*2})$$

Bajo la hipótesis nula F se distribuye como $F_{p-2;n-p,g.l}$. La hipótesis nula se rechaza si $P(F_{p-2;n-p,g.l} > F) \leq \alpha$.

Vb.-Contraste de P_E

Consideremos dos funciones no lineales: $G_0(X;\beta)$ y $G_1(X;\gamma)$ entonces las hipótesis son:

$H_0 : y = G_0(X;\beta) + \varepsilon_0$

$H_1 : f(y) = G_1(Z;\gamma) + \varepsilon_1$

Apliquemos la idea del contraste J [16], y consideremos que Y se puede expresar como una combinación lineal convexa de $G_0(X;\beta)$ y $G_1(Z;\gamma)$, esto es:

$$Y = (1-\alpha)G_0(X;\beta) + \alpha G_1(Z;\gamma) + \varepsilon \qquad 0 \leq \alpha \leq 1 \quad (1)$$

[15] Este test está en Bolch, B; Huang, C (1974)
Multivariate Statistical Methods for Business and Economics.

[16] El contraste J fue desarrollado por Davidson y MacKinnon en el artículo " Several Tests for Model Specification in the Presence of Alternative Hipotheses" Econometrica Vol 49 pp 241-262-1981

El conjunto de hipótesis anterior se modifica por la hipótesis: $H_0 : \alpha = 0$, entonces hay que estimar el conjunto de parámetros: α, β, γ . Una forma de proceder es estimar uno de los vectores de parámetros por mínimos cuadrados no lineales y luego utilizarlo en la ecuación (1) para estimar por mínimos cuadrados no lineales tanto el otro vector de parámetro y α con la restricción: $0 \le \alpha \le 1$ el siguiente paso hacer el contraste de la hipótesis nula: $H_0 : \alpha = 0$, empleando el estadístico: $\alpha^* / \sqrt{Var(\alpha^*)}$ que sigue una distribución normal estándar. Ahora, haremos una modificación introduciendo en la hipótesis alterna una función no lineal, monótona, continua y diferenciable $h(y)$ de donde se obtiene lo siguiente:

$$(1-\alpha)\left[y - G_0(X;\beta)\right] + \alpha\left[h(y) - G_1(Z;\gamma)\right] = \varepsilon$$

Para facilitar la aplicación práctica se ha propuesto reescribir la última expresión como:

$$y - G_0(X;\beta) = \alpha\left[G_1(Z;\gamma) - h(y)\right] + \alpha\left[y - G_0(X;\beta)\right] + \varepsilon \quad (2)$$

Ahora podemos emplear la expansión de Taylor de $G_0(X;\beta)$, partiendo de un valor inicial β° :

$$G_0(X;\beta) = G_0(X;\beta^\circ) + \nabla G_0(X;\beta^\circ)(\beta - \beta^\circ)$$

Esta expresión se sustituye en (2). Esta modificación se llama contraste P_E .

Vc.-Otro Test de linealidad.

En este punto comentaremos brevemente el trabajo de Samorov el al (2005) que parten del modelo:

$$y = f(X) + \varepsilon. \quad f(X) = \theta^T X_1 + G(X_2) + \varepsilon$$

$$X^T = (X_1^T, X_2^T)$$

La dimensión de cada grupo de variables es: $\dim(X_2) = M$ y $\dim(X_2) = d - M$ donde $M << d$, $G(X_2)$ es una función no lineal desconocida cuya dimensión es mucho menor que de la parte lineal, la distribución de ε es también desconocida. En este artículo se propone la solución de varios problemas asociado al modelo propuesto que se denomina modelo parcialmente lineal. El primer problema es determinar el grado de no linealiad, esto es, el valor de M que puede tomar los valores: 0,1,2...identificando por tanto cuantas variables conforman a $G(X_2)$, el siguiente problema es estimar el vector de parámetro θ y finalmente hay que estimar la función $G(X_2)$. El test que se desarrolla es: $H_0 : y = \theta^T X_1$ versus $H_1 : y = f(X) + \varepsilon$ o equivalentemente: $H_0 : M = 0$ y $H_1 : M > 0$. La idea parte de que si el modelo es lineal, entonces, $\nabla f(X) = \partial f(X) / \partial X$ debe ser constante para todo X por tanto la varianza de $\nabla f(X)$ es una buena medida del grado de no linealidad de las variables X . El contraste lo generalizan como: $H_0 : M \le M_0$ versus $H_0 : M > M_0$, siendo M_0 un valor mínimo prefijado para el cual la hipótesis nula no se rechaza, si $M_0 = 1$, se está indicando que el componente no lineal es univariante. El procedimiento supone la estimación de la varianza, denotémosla por V_m^*, rechazamos la hipótesis $H_0 : M \le M_0$, si $V_{(M_0+1)}^*$ es significativamente diferente de cero. Se asume en todo caso que la dimensión del componente no lineal es relativamente pequeño. Los autores de este trabajo proponen un algoritmo construidos bajo varios supuestos tanto para estimar el vector de parámetro θ como el estimador de la varianza: V_m^*.

VI.-Modelos ARCH y GARCH.

Consideremos el proceso de parámetro discreto $\{Y_t; t \in E\}$ y consideremos además un conjunto de información Ω_t tal que $\Omega_t \subseteq \Omega_{t+1} \subseteq \Omega_{t+2} \subseteq ... \subseteq \Omega$, $\Omega = \{Y_t; t < \infty\}$, entonces, podemos considerar a la esperanza condicional: $E(Y_t / \Omega_t)$ y la varianza condicional: $VAR(Y_t / \Omega_t)$ esta última puede ser no homogénea, es decir diferente para cada periodo. Este caso puede darse tanto en modelos no lineales como lineales. La presencia de varianza condicional no homogenea en los modelos lineales se ha asociado a una mala especificación del mismo y se ha tratado de resolver añadiendo nuevas variables exógenas, pero la práctica ha demostrado que este no es el mejor camino. Tanto la media condicional como la varianza condicional se asocia al corto plazo, mientras que la media no condicional $E(Y_t)$ y la varianza no condicional $VAR(Y_t)$ se asocian al largo plazo.
Consideremos el modelo lineal autoregresivo de primer orden $AR(1)$: $Y_t = \varphi_1 Y_{t-1} + a_t$ donde se asume que $a_t \overset{D}{\approx} N(0, \sigma^2)$.

La media no condicional de este proceso, o sea su media a largo plazo es:
$E(Y_t) = \varphi_1 E(Y_{t-1}) + E(a_t)$ Asumiendo que el proceso es estacionario entonces $\mu(1-\varphi_1) = 0$ mientras que la media a corto plazo o condicional es: $E(Y_t / \Omega_{t-1}) = \varphi_1 E(Y_{t-1} / \Omega_{t-1}) + E(a_t) = \varphi_i Y_{t-1}$ [17]

La varianza no condicional dado que el proceso es estacionario, es constante, en efecto:
$VAR(Y_t) = \varphi_1^2 VAR(Y_{t-1}) + VAR(a_t) = VAR(Y_t) = \varphi_1^2 VAR(Y_t) + \sigma^2$
Luego:

$$VAR(Y_t) = \sigma^2 (1-\varphi_1^2)^{-1}$$

La varianza condicional[18] es:

$$VAR(Y_t / \Omega_t) = VAR(a_t) = \sigma^2$$

Entonces estamos en la situación donde tanto la varianza no condicional como la condicional son iguales para cualquier período, es decir el proceso es homoscedástico.

---

[17] Consideremos en forma general dos vectores aleatorios X_1 y X_2, la esperanza de X_1 condicionada a X_2 es: $E(X_1 / X_2) = \int_S x_1 f_{X_1/X_2}(x_1 / x_2;)dx_1$, por otra parte

$\mu_1 = E(X_1) = \int_S x_1 f_{X_1}(x_1; \Theta)dx_1 =$

$\int_S \int_S x_1 f_{X,X_2}(x_1, x_2; \Theta)dx_1 dx_2 = \int_S \int_S x_1 f_{X/X_2}(x_1 / x_2; \Theta) f_{X_2}(x_2; \Theta)dx_1 dx_2 = \int_S f_{X_2}(x_2; \Theta)(\int x_1 f_{X/X_2}(x_1 / x_2; \Theta)dx_1)dx_2$
$= E(E(X_1 / X_2))$

[18] De la misma forma, consideramos dos vectores aleatorios X_1 y X_2, la varianza de X_1 condicionada a X_2 se obtiene a partir de: $X_1 - \mu_1 = X_1 - E(X_1 / X_2) + E(X_1 / X_2) - \mu_1$. por tanto la varianza de X_1 es:

$VAR(X_1) = VAR(X_1 - E(X_1 / X_2) + E(X_1 / X_2) - \mu) = E(X_1 - E(X_1 / X_2)^2 + E(E(X_1 / X_2) - \mu)^2$
considerando la nota anterior y que el producto cruzado es nulo, concluimos que:
$VAR(X_1) = E(VAR(X_1 / X_2) + VAR(E(X_1 / X_2)$

Pero se puede preguntar que ocurre cuando el proceso autoregresivo tiene una varianza heteroscedástica. De ahí surge la primera repuesta que da R. F. Engle (1982) con el modelo ARCH y la generalización de Bollerslev (1986) el GARCH. De estos trabajos pioneros han surgido nuevas consideraciones expresadas en los aportes de Andrew A. Weiss (1986), Ruey S Tsay (1987), James Rochon (1992) y mucho más, recientes están los de Medeiros M el al (2003), Audrini F (2005) Chandra el al (2006) Cline(2006), Ferlan R (2006) . Lo interesante de estos últimos trabajos es que giran en torno a nuevas metodologías que incluyen métodos no paramétricos y redes neurales para ampliar las posibilidades de los modelos propuestos por ambos pioneros.

VIa.-ARCH

Veremos en primer lugar la propuesta planteada por R. F. Engle (1982), parte de un modelo cuasilineal dado por:

$Y_t = \alpha_t h_t^{1/2}$ y $h_t = \beta_0 + \beta_1 Y_{t-1}^2$ (1)

Donde $Var(\alpha_t) = 1$, y β_0, β_1 parámetros desconocidos. Este es el modelo ARCH más sencillo. Asumiendo normalidad dado el conjunto de información disponible Y_{t-1} la distribución condicional de Y_t / Ω_{t-1} tiene una distribución normal $N(0, h_t)$. La varianza condicional h_t puede expresarse más generalmente como:

$h_t = h(Y_{t-1}, Y_{t-2}, ..., Y_{t-p}; \beta_0, \beta_1, \beta_2 ... \beta_p)$ (2)

Donde de la misma forma que en (1) $\beta_0, \beta_1, \beta_2 ... \beta_p$ son parámetros desconocidos, el valor de p dará el orden del modelo ARCH. Algunas formas de h_t que pueden ser útiles en ciertas aplicaciones son:

$h_t = \exp(\beta_0 + \beta_1 Y_{t-1}^2)$

$h_t = \beta_0 + \beta_1 |Y_{t-1}|$

Si se asume que la esperanza de Y_t es la combinación de variables endógenas y exógenas representadas por el vector de variables Z_t y el vector de parámetros γ; $\gamma^T = (\gamma_0, \gamma_1, ... \gamma_p)$ asociado al vector de variables Z_t, entonces: $E(Y_t / \Omega_{t-1}) = Z_t \gamma$, luego: $Y_t / \Omega_{t-1} \overset{d}{\approx} N(Z_t \gamma_t, g_t)$ con:

$g_t = g(\alpha_{t-1}, \alpha_{t-2}, ..., \alpha_{t-p}; \varphi_0, \varphi_1, \varphi_2 ... \varphi_p)$ (3)

$\alpha_t = Y_t - Z_t \gamma$

Un caso particular es $g_t = \varphi_0 + \varphi_1 \alpha_{t-1}^2$ como varianza condicional $Var(\alpha_t / \alpha_{t-1})$, la varianza no condicional es $Var(\alpha_t) = E(Var(\alpha_t / \alpha_{t-1})) = \alpha_0 + \alpha_1 E(\alpha_{t-1}^2) = \alpha_0 + \alpha_1 Var(\alpha_{t-1})$ si $\{\alpha_t\}$ es un proceso estacionario en varianza entonces $Var(\alpha_t) = \alpha_0 / (1 - \alpha_1)$

Este caso es el modelo de regresión ARCH.

Considerando (1) y (2) se puede generalizar escribiendo:

$h^\circ{}_t = h^\circ(\alpha_{t-1}, \alpha_{t-2}, ..., \alpha_{t-p};, z_{t-1}, z_{t-2} ... z_{t-p}; \varphi_0, \varphi_1, ... \varphi_p; \gamma_0, \gamma_1 ... \gamma_p,) = h_t(z; \gamma) g_t(\alpha; \varphi)$ (4)

O empleando el concepto de conjunto de información disponible: Ω_{t-1} y considerando los vectores de parámetros β y φ entonces:

$h^\circ{}_t = h(\Omega_{t-1}; \alpha; \varphi)$

Para el caso que solo se emplean solamente variables endógenas Ruey S. Tsay (1987) propone lo siguiente:

$$Y_t / \Omega_{t-1} \overset{d}{\approx} D(r_t; g_t)$$

$$r_t = r(Y_{t-1}, Y_{t-2}, ..., Y_{t-p}; \beta_0, \beta_1, \beta_2 ... \beta_p) = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} .... \beta_p Y_{t-p}$$

$$g_t = g(\alpha_{t-1}, \alpha_{t-2}, ..., \alpha_{t-p}; \varphi_0, \varphi_1, \varphi_2 ... \varphi_p) = \varphi_0 + \varphi_1 \alpha_{t-1}^2 + \varphi_2 \alpha_{t-2}^2 + ... + \varphi_p \alpha_{t-p}$$

Donde como siempre:

$$\alpha_t = Y_t - \sum_{i=1}^{p} \varphi_t Y_{t-i}$$

Como sugieren el autor al plantear este modelo: $Y_t / \Omega_{t-1} \overset{d}{\approx} D(r_t; g_t)$ donde $\Omega_{t-1} : \{Y_{t-1}, Y_{t-2} ... Y_{t-p}\}$ indica que el proceso sigue una distribución con media r_t y varianza g_t. Este modelo se expresa como un ARCH(p,q). El problema está en la determinación de los valores de p y q que pueden resultar relativamente grande y que entonces el modelo no cumpliría con el principio de parsimonia. Por tal motivo, este autor propone otro modelo partiendo de un modelo ARMA(p,q) y el empleo de coeficientes aleatorios para el proceso $\{\alpha_t\}$. Esto es:

$$\beta(B)(Y_t - \mu) = \varphi(B)\alpha_t$$

$$\delta_i(B)\alpha_t = w_{0t}(Y_{t-1}^{*}(1) - \mu) + w_t^{*}(B)(Y_t - \mu) + \varepsilon_t$$

Donde: $\beta(B) = 1 - \beta_1 B - ... \beta_p B^p$; $\varphi(B) = 1 - \varphi_1 B - ... \varphi_q B^q$ son polinomios de orden p y q en B respectivamente, con coeficientes constantes tal como ocurre en el clásico modelo ARMA y $\delta_i(B) = 1 - \delta_{1,t} B - ... \delta_{r,t} B^r$; $w^{*}(B) = w_{1,t} B + ... + w_{s,t} B^s$ son polinomios en B de orden r y s con coeficientes aleatorios, $Y_{t-1}^{*}(1)$ representa el pronóstico mínimo cuadrático de Y_t dado el conjunto de información Ω_{t-1}, finalmente ε_t es la perturbación aleatoria o ruido blanco en el período t. Este modelo lo llama Ruey S. Tsay en su artículo: CHARMA(p,q,r,s), aparte de las propiedades y métodos de estimación señala dos aplicaciones importantes aparte de su uso en series heteroscedásticas permite manipular series de tiempo con presencia de valores atípicos y la otra aplicación importante es que permite refinar los coeficientes constantes, tal como llama el autor los parámetros, del modelo ARMA.

VIa1.-Estimación.

Las técnicas de estimación de los parámetros son: el de máxima verosimilitud paramétrico y no paramétrico de Audrino F (2005), mínima α divergencia de Ajay Chandra et al (2006), mínimos cuadrados generalizados factibles Green (1999). El estimador de máxima verosimilitud siguiendo a Engle (1982) se emplea para estimar los parámetros de $h_t = h(Y_{t-1}, Y_{t-2}, ..., Y_{t-p}; \beta_0, \beta_1, \beta_2 ... \beta_p)$ o de la regresión ARCH dado como $E(Y_t / \Omega_{t-1}) = Z_t \gamma$ y $g_t = g(\alpha_{t-1}, \alpha_{t-2}, ..., \alpha_{t-p}; \varphi_0, \varphi_1, \varphi_2 ... \varphi_p)$, en este último consideramos el caso particular $g_t = \varphi_0 + \varphi_1 \alpha_{t-1}^2$.

En el primer caso, se parte de $Y_t / \Omega_{t-1} \overset{d}{\approx} N(0, h_t)$ por tanto:

$$f_{Y_t}(y_t; 0, h_t) = \frac{1}{(2\pi h_t)^{1/2}} \exp\left\{ -\frac{Y_t^2}{2h_t} \right\}$$

La función de log-verosimilitud es:

$$\sum_{t=1}^{n}\ln f_{Y_t}(y_t;0,h_t)=\sum_{t=1}^{n}\ln\frac{1}{(2\pi h)^{1/2}}+\sum_{t=1}^{n}\left\{-\frac{Y_t^2}{2h_t}\right\}$$

$$\sum_{t=1}^{n}\ln f_{Y_t}(y_t;0,h_t)=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{n}\ln h_t+\sum_{t=1}^{n}\left\{-\frac{Y_t^2}{2h_t}\right\}$$

Aplicando la condición necesaria para la existencia de un óptimo:

$$\sum_{t=1}^{n}\partial\ln f_{Y_t}(y_t;0,h_t)/\partial\beta=-\frac{1}{2}\sum_{t=1}^{n}\partial\ln h_t/\partial\beta+\sum_{t=1}^{n}\partial\left\{-\frac{Y_t^2}{2h_t}\right\}/\partial\beta=0$$

$$\sum_{t=1}^{n}\partial\ln f_{Y_t}(y_t;0,h_t)/\partial\beta=-\frac{1}{2}\sum_{t=1}^{n}\frac{1}{h_t}\partial lh_t/\partial\beta+\sum_{t=1}^{n}\partial\left\{-\frac{Y_t^2}{2h_t}\right\}/\partial\beta=0$$

Consideremos el caso particular:

$h_t=\beta_0+\beta_1Y_{t-1}^2+\beta_2Y_{t-2}^2....\beta_pY_{t-p}^2$, entonces, ahora definimos $z_T=(1,Y_{t-1}^2,Y_{t-2}^2,...Y_{t-p}^2)$ y $\beta=(\beta_0,\beta_1,...\beta_p)$ La matriz de información está dada como:

$$I(\beta)=\frac{1}{2n}\sum_{t=1}^{n}(z_t^Tz_t/h_t)$$

El segundo caso es:

$E(Y_t/\Omega_{t-1})=Z_t\gamma$

$g_t=\varphi_0+\varphi_1\alpha_{t-1}^2+\varphi_2\alpha_{t-2}^2+...+\varphi_p\alpha_{t-p}^2$

$\alpha_t=Y_t-Z_t\gamma$

Asumiendo normalidad, esto es: $Y_t/\Omega_{t-1}\overset{d}{\approx}N(Z_t\gamma_t,g_t)$ empleamos nuevamente el estimador de máximo verosimilitud.

$$f_{Y_t}(y_t;Z_t\gamma,g_t)=\frac{1}{(2\pi g_t)^{1/2}}\exp\left\{-\frac{(Y_t-Z_t\gamma)^2}{2g_t}\right\}=\frac{1}{(2\pi g_t)^{1/2}}\exp\left\{-\frac{\alpha_t^2}{2g_t}\right\}$$

$$\prod_{t=1}^{n}f_{Y_t}(y_t;Z_t\gamma,g_t)=\prod_{t=1}^{n}\frac{1}{(2\pi g_t)^{1/2}}\exp\left\{-\frac{\alpha_t^2}{2g_t}\right\}$$

La función log-verosimilitud es:

$$Ln\prod_{t=1}^{n}f_{Y_t}(y_t;Z_t\gamma,g_t)=\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{n}\ln g_t+\sum_{t=1}^{n}\left\{-\frac{\alpha_t^2}{2g_t}\right\}$$

La condición necesaria de existencia de un punto extremo es:

$$\partial Ln\prod_{t=1}^{n}f_{Y_t}(y_t;Z_t\gamma,g_t)/\partial\gamma=-\frac{1}{2}\sum_{t=1}^{n}\partial\ln g_t/\partial\gamma+\sum_{t=1}^{n}\partial\left\{-\frac{\alpha_t^2}{2g_t}\right\}/\partial\gamma=0$$

La matriz de información del estimador de γ está dada por:

$$I^*(\gamma)=\frac{1}{n}\sum_{t=1}^{n}Z_t^TZ_t(g_t^{-1}+2\alpha_t^2\sum_{j=1}^{p}\varphi_j^2g_{t-j}^{-2})$$

La matriz de información conjunta es:

$$I(\varphi;\gamma)=\frac{1}{n}\sum_{t=1}^{n}E\left(\frac{1}{2g_t^{\,2}}\frac{\partial g_t}{\partial\varphi}\frac{\partial g_t}{\partial\gamma}\right)$$ [19]

En vez de emplear los estimadores máximos verosímiles se puede emplear el siguiente procedimiento para el caso $g_t=\varphi_0+\varphi_1\alpha_{t-1}^2$:
1.-Se hace la regresión de Y_t sobre Z_t para obtener por mínimos cuadrados γ^* y el error e_t^*.

2.-Defina la regresión: $e_t^{*2}=\varphi_0+\varphi_1 e_t^{*2}+\xi_t$, y obtenga los estimadores mínimos cuadráticos: φ_0^* y φ_1^* como valores iniciales, que lo denotamos como el vector: $\varphi^*=\left[\varphi_0^*,\varphi_1^*\right]$.

3.-Calcule el ajuste $f_t=\varphi_0^*+\varphi_1^* e_{t-1}^{*2}$ y defina la regresión de $\left[(e_t^{*2}/f_t)-1\right]$ en $(1/f_t)$ y (e_{t-1}^{*2}/f_t) y obtenemos los estimadores mínimos cuadráticos, y con ellos definimos el vector: θ_α^* el primer estimador del vector $\left[\varphi_0,\varphi_1\right]$ es:
$\varphi^{\circ\circ}=\varphi^*+\theta_\alpha^*$
Puede observarse que f_t es un estimador de $g_t=\varphi_0+\varphi_1\alpha_{t-1}^2$; e_{t-1}^{*2} estima a α_{t-1}^2
La matriz de varianza covarianza asintótica de $\varphi^{\circ\circ}$ es $2(R^TR)^{-1}$ donde R es la matriz de los regresores, bajo el supuesto de normalidad.

4.-Recalcule f_t usando $\varphi^{\circ\circ}$ para t=1,2…n-1 y con ello defina:
$$r_t=\left[\frac{1}{f_t}+2\left(\frac{\varphi_1^{\circ\circ}}{f_{t+1}}\right)\right]^{1/2}\qquad s_t=\frac{1}{f_t}-\frac{\varphi_1^{\circ\circ}}{f_{t-1}}\left[\frac{e_{t+1}^{*2}}{f_{t+1}}-1\right]$$
Calcule la estimación de la regresión de $e_t s_t/r_t$ en $Z_t r_t$ los estimadores obtenidos lo denotamos por el vector: θ_γ^*. El primer estimador de γ es:

$\gamma^{\circ\circ}=\gamma^*+\theta_\gamma^*$
Este procedimiento se itera para refinar los resultados, lo estimadores obtenidos son asintóticamente equivalentes a los de máxima verosimilitud a pesar que no necesariamente cinciden. Este procedimiento se puede generalizar para:
$g_t=\varphi_0+\varphi_1\alpha_{t-1}^2+\varphi_2\alpha_{t-2}^2+\ldots+\varphi_p\alpha_{t-p}^2$

Teorema (R. F. Engle (1982))

El proceso ARCH de orden p, con parámetros no negativos $\beta_0\geq 0,\beta_1\geq 0,\ldots\beta_p\geq 0$, es estacionario en covarianza si, y solamente si, la ecuación característica tiene todas sus raíces fuera de un circulo de radio unitario. La varianza estacionaria está dada por:
$E(Y_t^2)=\beta_0/(1-\sum_{j=1}^{p}\beta_j)$

[19] Bajo el supuesto que el modelo de regresión ARCH sea simétrico y regular entonces $I(\varphi;\gamma)=0$, ver R. F. Engle (1982).

VIa2.-Otros métodos de estimación.

Entre los otros métodos de estimación están: Audrino, F (2005) en donde no se asume el supuesto de normalidad de las perturbaciones aleatorias también conocidazas como innovaciones, aunque la propuesta se refiere a un ARCH(1), indica el autor que es posible extenderlo a modelos con $p>1$. El método de estimación es no paramétrico y se basa en una transformación logarítmica y emplear luego la estimación de verosimilitud local. Entonces consideremos el modelo:

$Y_t = \alpha_t h^{1/2}$ y $h_t = \beta_0 + \beta_1 Y_{t-1}^2$

$Y_t^2 = \alpha_t^2 h; \quad Log(Y_t^2) = Log\alpha_t^2 + Logh = \beta + Log\alpha_t^2 + Logh_t - \beta$

Donde $\beta = E\left[\log\alpha_t^2\right]$, haciendo las transformaciones:

$V_t = Log(Y_t^2)$ y $U_t = Log(\alpha_t^2) - \beta$

$G(Y_{t-1}) = \beta + Logh_t$

Entonces:

$V_t = G(Y_{t-1}) + U_t$

Tomando en cuenta que: $\beta = E\left[\log\alpha_t^2\right]$ entonces: $E(U_t) = 0$, la sucesión de variables aleatorias: $\{U_t\}$, son independientes e idénticamente distribuidas e independientes de la sucesión de variables aleatorias: $\{V_t; s<t\}$. El problema es estimar $G(Y_{t-1})$, para ello se emplea el local constante log-verosimilitud, dado una muestra de tamaño n definimos:

$$LogL(y, G_y) = \sum_{t=2}^{n} W_t(Y))\rho(V_t, G)$$

Donde: $W_t(Y) = W((y - Y_{t-1})/h_n)$ [20] cumple con la función de ponderar las observaciones, dándole mayor peso a las cercanas a y $\rho(V_t, G_y) = -\log((f_{U_t} - G_y))$, f_{Ut} es la función de densidad, h_n es el ancho de banda global, la secuencia $\{h_n\}$ satisface con:

$\lim_{n\to\infty} h_n \to 0$, $n\lim_{n\to\infty} h_n \to \infty$, $h_n > 0$, $\forall n \geq 2$

El estimador local constante log-verosimilitud es:

$$G^*_y = \min_{Gy} \sum_{t=2}^{n} w_t(Y)(-\log(f_{Ut} - G_y))$$

El autor demuestra la unicidad del estimador asumiendo normalidad de U_t, la consistencia y las condiciones de normalidad asintótica. Los resultados los lleva a cabo mediante la técnica de simulación.

Bajo el supuesto de normalidad y asumiendo que la densidad estacionaria de Y_t es continua y positiva, acotada uniformemente y otros supuestos demuestra que la varianza asintótica es:

$$\sigma^2 = \frac{\int_{-\infty}^{\infty} W^2(u)du}{dy}$$

[20] $W(.)$ es el núcleo (kernel) que en general cumple con: $W(.) \geq 0$, $\int_{-\infty}^{\infty} W(z)dz = 1$ es una función simétrica no negativa y acotada

El otro método es él de mínima α divergencia desarrollado por Ajay Chandra, J et al (2006), el problema es determinar la forma de la distribución de probabilidad de α_t en el modelo:

$Y_t = \alpha_t h^{1/2}$, $h_t = \beta_0 + \beta_1 Y_t^2 + \beta_2 Y_t^2 .... \beta_p Y_{t-p}^2$;

Donde $\{\alpha_t\}$es una sucesión de variables aleatorias independientes e igualmente distribuidas con función de densidad $g(\alpha;\Omega)$. Hagamos el cambio:

$Y_t^2 = \alpha_t^2 h_t$ (4)

Denotemos por $Z_t = Y_t^2$; $W_{t-1} = (1, Y_{t-1}^2, Y_{t-2}^2, ... Y_{t-p}^2)$ y, $\beta^T = (\beta_0, \beta_1, \beta_2 ... \beta_p)$, entonces: $h_t = W_{t-1}\beta$ luego podemos escribir (4) como:

$Z_t = W_{t-1}\beta + W_{t-1}\beta(\alpha_t^2 - 1)$

Si hacemos: $\varsigma_t = W_{t-1}\beta(\alpha_t^2 - 1)$ el modelo queda finalmente como:

$Z_t = W_{t-1}\beta + \varsigma_t$

Se cumple que $E(\varsigma_t / \Omega_{t-1}) = 0$. El vector de estimadores β^* se puede obtener partiendo de :

$$Min \sum_{t=p+1}^{n} \left(Z_t - W_{t-1}\beta\right)^2$$

Los residuos están dados por: $\alpha_t^* = Y_t / h_t^*$, donde: $h_t^* = \beta_0^* + \beta_1^* Y_t^2 + \beta_2^* Y_t^2 .... \beta_p^* Y_{t-p}^2$

Encontremos ahora una función f_θ como ajuste de la verdadera función de densidad $g(\alpha;\Omega)$ utilizando la mínima α divergencia, en general se parte de la definición dada por:

$$D_\alpha(f_\theta, g(\alpha;\Omega)) = \int_{-\infty}^{\infty} K_\kappa\left\{f_\theta(\alpha)/g(\alpha;\Omega)\right\} g(\alpha;\Omega) d\beta$$

Donde $K_\kappa(x) = \left\{4/(1-\kappa^2)\right\}\left\{1 - x^{(1+\kappa)/2}\right\}$; $-1 < \kappa < 1$. El estimador mínimo α divergente es:

$$\overset{*}{\theta}{}_n^{\kappa} = \min_{\theta\in\Theta} \overset{*}{D}_\kappa(f_\theta, \overset{*}{g}_n(\alpha;\Omega)),\ \overset{*}{D}_\kappa(f_\theta, \overset{*}{g}_n(\alpha;\Omega)) = \int_{-\infty}^{\infty} K_\kappa\left\{f_\theta(\alpha)/\overset{*}{g}_n(\overset{*}{\alpha};\Omega)\right\} \overset{*}{g}_n(\alpha;\Omega) d\alpha$$

$$g_n^*(\overset{*}{\alpha};\Omega) = \frac{1}{nc_n}\sum_{t=2}^{n} W\left(\frac{\left(x - \overset{*}{\alpha}_t\right)}{c_n}\right); (4)$$

$c_n = n^{-\lambda}$; $1/4 < \lambda < 1/3$

Para que (4) sea más fácil de computar , los autores consideran como ejemplo el núcleo de Epanechbokov: $W(x) = 0{,}75(1 - x^2); |x| \le 1$.

VI3a.-Test de hipótesis.

Consideremos la varianza condicional $g_t = \varphi_0 + \varphi_1 \alpha_{t-1}^2 + \varphi_2 \alpha_{t-2}^2 + ... + \varphi_p \alpha_{t-p}^2$, o de forma más general: $Var(\alpha_t / \alpha_{t-1}, \alpha_{t-2}, ... \alpha_{t-p}) = g_t(\alpha_{t-1}, ... \alpha_{t-p}; \varphi_0 ... \varphi_{t-p})$ lineal o no, por ejemplo puede ser exponencial. La hipótesis nula es: $H_0 : \varphi_1 = \varphi_2 = ... \varphi_p = 0$. Estoa hipótesis mantiene que la varianza condicional es constante. Engle (1982) propone el siguiente procedimiento, consideremos el vector $e_t = (1, \alpha_{t-1}^*, \alpha_{t-2}^*, ... \alpha_{t-p}^*)$, donde α_j^* son los

residuos, asumimos que existe $\partial g_t / \partial \varphi$ y la esperanza : $E(e^t e)$[21], por tanto la matriz de información es:

$$I(\alpha^*) = \frac{1}{2}\left(\frac{\partial g^0 / \partial \alpha}{g^0}\right) E(e^t e)\left(\frac{\partial g^0 / \partial \alpha}{g^0}\right)^t$$

El estadístico para realizar el contraste es:

$$ML = \frac{1}{2} f^0 e (e^t e)^{-1} e^t f^0$$

Donde g^0 es el vector columna: $\left\{e_t^2 / g^0 - 1\right\}$, este estadístico, bajo la hipótesis nula, tiende asintóticamente a una χ^2 con p grados de libertad. La hipótesis nula se rechaza si la probabilidad de que el valor del estadístico exceda a una χ_p^2 sea menor que el nivel de significación. Una forma expedita consiste en considerar los residuos y plantear el modelo $\alpha_t^{*2} = \phi_0 + \phi_1 \alpha_{t-1}^{*2} + \phi_2 \alpha_{t-2}^{*2} + ... + \phi_p \alpha_{t-p}^{*2} + \varepsilon_t$ y emplear la prueba F para contrastar la hipótesis $H_0 : \varphi_1 = \varphi_2 = ... \varphi_p = 0$ y la pruba t para cada uno de los parámetros, asumiendo como hipótesis nula $H_0 : \varphi_j = 0, \forall j = 1,2...p$. Una forma de aproximarse al problema es, partiendo de un AR(p), estudiar los residuos y ver si hay un patrón de agrupamiento de los mismos.

Ejemplo 12

En este ejemplo emplea el índice de precio del consumidor en Venezuela desde Enero de 1980 a Diciembre de 2002, en otros países se considera que este tipo de índice presenta hetoroscedasticidad condicional. Lo primero que se realizó fue considerar un ARIMA(1,1,1)[22] como modelo tentativo, obteniendo los siguientes resultados.

Descripción del modelo

	Tipo de modelo
ID del modelo IPMG Model_1	ARIMA(1,1,1)

[21] Obsérvese que para cada valor de t se obtiene un vector columna, por tanto e es una matriz.

El programa es SPSS es:

[22]
```
 GET DATA /TYPE=ODBC /CONNECT=
  'DSN=Excel Files;DBQ=D:\ARTÌCULOS\DATOS
ARCH.xls;DriverId=790;MaxBufferSiz'+
  'e=2048;PageTimeout=5;'
  /SQL = 'SELECT  RE1C,  `RE-1C` AS RE1C1,  `RE-2C` AS RE2C,  `RE-3C`
AS RE3'+
  'C,  `RE-4C` AS RE4C FROM  `Hoja2$`'
  /ASSUMEDSTRWIDTH=255
  .
CACHE.
DATASET NAME DataSet2 WINDOW=FRONT.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT RE1C
  /METHOD=ENTER RE1C1 RE2C RE3C RE4C
  /SCATTERPLOT=(*SRESID ,RE1C ) (*ZRESID ,RE1C )
  /RESIDUALS DURBIN HIST(ZRESID) NORM(ZRESID)
  /SAVE COOK RESID .
```

De acuerdo a los siguientes resultados, el modelo no se adecua satisfactoriamente, si se observa el coeficiente de determinación $R^2 = 1{,}0$ (este valor suele ser sospechoso, rara vez un modelo explican la variabilidad de los datos perfectamente); el valor del índice de información Bayesiano: $BIC = 1{,}020$. El estadístico observado de Ljung-Box conduce a rechazar la hipótesis nula que las correlaciones de los residuos es nula, es decir: los residuos no provienen de un ruido blanco.

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo							Ljung-Box Q(18)			Número de valores atípicos
		R-cuadrado estacionaria	R-cuadrado	RMSE	MAPE	MaxAPE	MaxAE	BIC normalizado	Estadísticos	GL	Sig.	
IPMG-Model_1	0	,585	1,000	1,615	8,141	151,892	10,857	1,020	46,664	16	,000	0

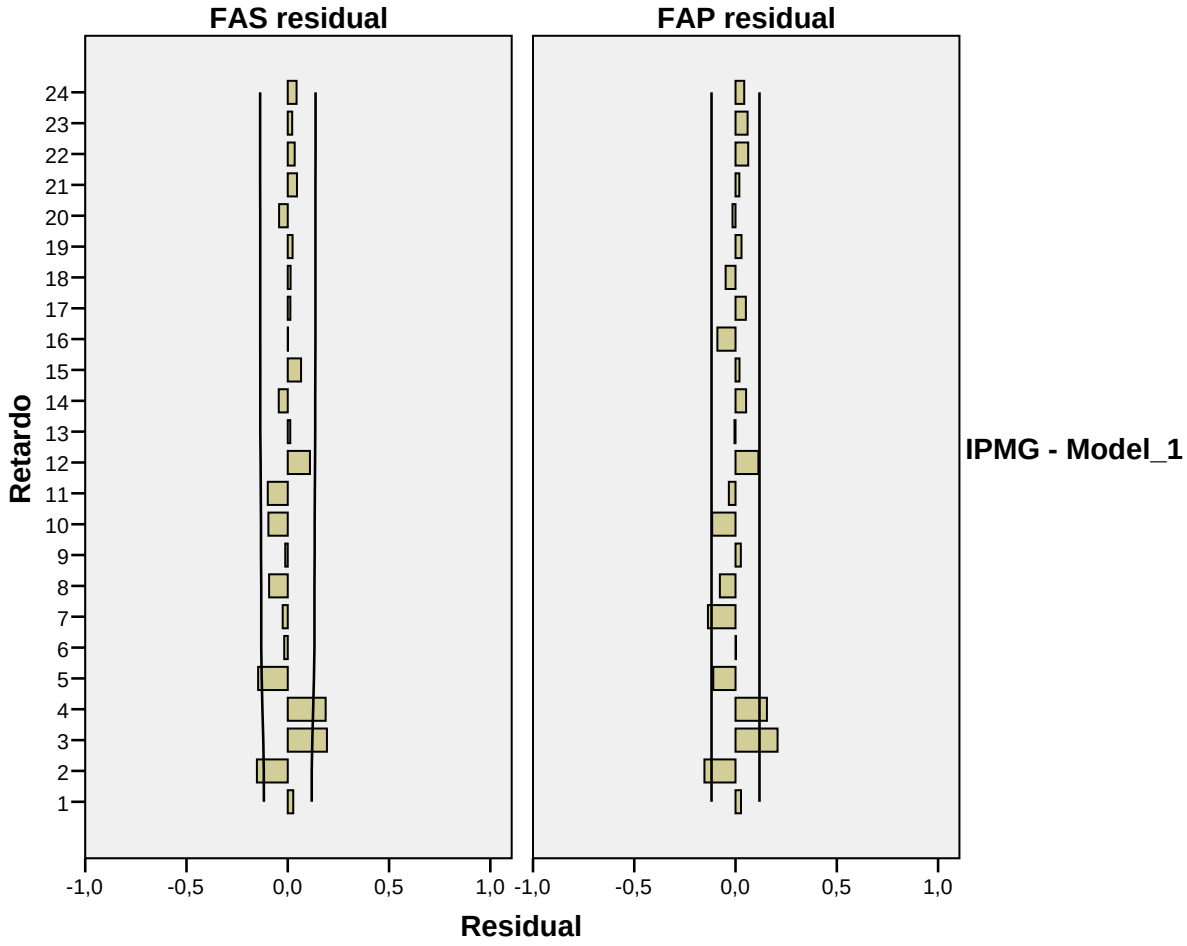
Parámetros del modelo ARIMA

					Estimación	ET	t	Sig.
IPMG-Model_1	IPMG	Sin transformación	Constante		1,067	,516	2,069	,040
			AR	Retardo 1	,868	,039	22,289	,000
			Diferencia		1			
			MA	Retardo 1	,286	,078	3,682	,000

Si observamos la tabla anterior, el modelo ajustado es:

$(1-0{,}868B)(1-B)Y_t^* = 1{,}067 + 0{,}286\alpha_{t-1}$

La prueba t nos indica que todos los parámetros son significativamente diferentes de cero, sin embargo esto no valida el modelo. Si se observa las funciones de autocorrelacio y autocorrelación parcial de los residuos algunas salen de los límites, lo que coincide con el resultado del test de Ljung-Box.



Para ilustrar la construcción de un ARCH proseguiremos con el ejemplo. Ahora, construimos el modelo de regresión considerando los residuos, la variable dependiente es α_t^{*2} y los regresores: $\alpha_{t-1}^{*2}, \alpha_{t-2}^{*2}, \alpha_{t-3}^{*2}, \alpha_{t-2}^{*2}$; esto es:

$$\alpha_t^{*2} = \beta_0 + \beta_1 \alpha_{t-1}^{*2} + \beta_2 \alpha_{t-2}^{*2} + \beta_3 \alpha_{t-3}^{*2} + \beta_4 \alpha_{t-2}^{*2} + \varepsilon$$

Los resultados se presentan a continuación:

Variables introducidas/eliminadas(b)

Modelo	Variables introducidas	Variables eliminadas	Método
1	RE4C, RE2C, RE3C, RE1C1(a)	.	Introducir

a Todas las variables solicitadas introducidas
b Variable dependiente: RE1C

Resumen del modelo[b]

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Estadísticos de cambio					Durbin-Watson
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Sig. del cambio en F	
1	,543[a]	,294	,284	9,48563	,294	27,757	4	266	,000	2,107

a. Variables predictoras: (Constante), RE4C, RE2C, RE3C, RE1C1

b. Variable dependiente: RE1C

ANOVA[b]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	9990,156	4	2497,539	27,757	,000[a]
	Residual	23933,939	266	89,977		
	Total	33924,095	270			

a. Variables predictoras: (Constante), RE4C, RE2C, RE3C, RE1C1

b. Variable dependiente: RE1C

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Estadísticos de colinealidad	
		B	Error típ.	Beta			Tolerancia	FIV
1	(Constante)	,486	,603		,806	,421		
	RE1C1	,143	,059	,144	2,422	,016	,750	1,332
	RE2C	,035	,057	,036	,618	,537	,802	1,247
	RE3C	,277	,055	,287	4,987	,000	,801	1,248
	RE4C	,249	,057	,258	4,343	,000	,749	1,335

a. Variable dependiente: RE1C

De acuerdo a los resultados de la prueba F que se muestra en la tabla ANOVA se rechaza la hipótesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. De acuerdo a la prueba t solo la constante y el parámetro asociado a α_{t-2}^{*2} no son significativamente diferentes de cero. Se puede aumentar el número de regresores para aumentar el valor de R^2 y ver si mejora el modelo. En todo caso, es fácil verificar que si realizamos la operación $nR^2 = 271x0{,}294 = 76$ y calcular $P(\chi_4^2 > 76) \cong 0$ rechazamos igualmente la hipótesis nula anterior. De acuerdo el valor del estadístico Durbin Watson la autocorrelación de los residuos es aproximadamente 0,0535 que es un valor no significativo. Observando el índice de condición se puede concluir que no hay presencia de colinealidad.

Diagnósticos de colinealidad[a]

Modelo	Dimensión	Autovalor	Indice de condición	Proporciones de la varianza (Constante)	RE1C1	RE2C	RE3C	RE4C
1	1	2,256	1,000	,04	,07	,07	,07	,08
	2	,858	1,621	,95	,04	,02	,01	,03
	3	,725	1,765	,01	,23	,30	,28	,21
	4	,660	1,849	,00	,13	,43	,44	,14
	5	,501	2,121	,00	,52	,19	,20	,54

a. Variable dependiente: RE1C

Estadísticos sobre los residuos[a]

	Mínimo	Máximo	Media	Desviación típ.	N
Valor pronosticado	,4893	44,1138	2,2681	6,08281	271
Valor pronosticado tip.	-,292	6,879	,000	1,000	271
Error típico del valor pronosticado	,589	6,632	,857	,964	271
Valor pronosticado corregido	,4911	62,1935	2,3835	7,04578	271
Residuo bruto	-35,55740	100,70521	,00000	9,41511	271
Residuo tip.	-3,749	10,617	,000	,993	271
Residuo estud.	-4,940	11,017	-,005	1,077	271
Residuo eliminado	-61,75787	108,44568	-,11531	11,30054	271
Residuo eliminado estud.	-5,174	14,913	,012	1,268	271
Dist. de Mahalanobis	,046	130,986	3,985	17,065	271
Distancia de Cook	,000	3,864	,052	,356	271
Valor de influencia centrado	,000	,485	,015	,063	271

a. Variable dependiente: RE1C

VI4a.-Modelos relacionados: ARCH-M, TARCH, AR-ARCH.

En este punto veremos brevemente algunos modelos relacionados con el ARCH, tales como ARCH-M , TARCH y finalmente el AR-ARCH. El modelo ARCH-M consiste en introducir en el modelo la varianza condicional como regresor. Un caso de tal modelo es:

$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 g_t + \varepsilon_t$,

$g_t = \varphi_0 + \varphi_1 \alpha_{t-1}^2 + \varphi_2 \alpha_{t-2}^2 + ... + \varphi_p \alpha_{t-p}^2$

Otro modelo que se emplea en situación de asimetría es el ARCH que dependen de un umbral, estos modelos se llaman TARCH. Su forma es por ejemplo:

$Y_t = \alpha_t g_t^{1/2}$

$g_t = \varphi_0 + (\varphi_1 + \delta d_{t-1})\alpha_{t-1}^2$

$\varphi_0 > 0; \varphi_1 > 0, \ \varphi_1 + \delta < 1$

$d_{t-1} = 1; si\, \alpha_{t-1} < 0$, si $d_{t-1} = 0; si\, \alpha_{t-1} > 0$

La varianza condicional puede tener la forma:

$g_t = \varphi_0 + (\varphi_1 + \delta d_{t-1})\alpha_{t-1}^2 + \gamma g_{t-1}$

Entonces:

$\varphi_0 > 0; \varphi_1 > 0, \gamma > 0$ y $\varphi_1 + \gamma + \delta/2 < 1$

La variable indicadora d_{t-1} indica si el umbral está presente o no, toma el valor uno cuando lo está, y esto ocurre cuando la innovación es negativa, por tanto el efecto de la varianza condicional es mayor, un ejemplo de esto se da en el mercado de capitales, que una mala noticia sobre el desenvolvimiento del mercado conduce a una mayor volatilidad de los títulos valores.

Finalmente tenemos el modelo AR-ARCH, el modelo AR-ARMA lo desarrolla por primera vez Emanuel Parzen[23], este autor clasifica las series de tiempo en tres categorías, series con tendencias, estacionalidad y ciclos que denomina de memoria larga, series que cumplen con la estacionalidad en media y varianza que llama series de memoria corta y finalmente serie formadas exclusivamente con ruido blanco. Una de las características de las series de tiempo con memoria larga es que sus autocorrelaciones muestrales decrecen muy lentamente. Para lograr que una serie de memoria larga se convierta en memoria corta en vez de diferenciar la serie como propone la metodología de Box-Jenkins, es aumenta el número de parámetros AR sin importar el principio de parcimonia, y la parte MA la deja solo en casos especiales. Para seleccionar el orden del AR utiliza el criterio de información de Akaike. Otra propuesta para el estudio de series de tiempo de memoria larga es emplear diferenciación no entera, esto es: ∇^d donde d toma cualquier valor real, generalmente entre -1 y 1, este modelo se llama AFIRMA.

El modelo AR-ARCH consiste entonces, en un modelo con una parte autoregresiva y otra autoregresiva con heteroscedásticidad condicional.. Un caso particular de AR-ARCH lo presenta Cline, D.B.H (2006), en el cual considera el umbral. El modelo AR-ARCH lo define como: dada la sucesión de errores aleatorios $\{\alpha_t^*\}$ independientes e idénticamente distribuidas, con función de densidad simétrica a rededor de cero y positiva en R, también se asume que $E(|\alpha_t^*|^r) < \infty$. Consideremos ahora la sucesión de variables aleatorias: $\{\alpha_t\}$, entonces, el modelo AR-ARCH es:

$$\alpha_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i \alpha_{t-i} + (\phi_0 + \sum_{i=1}^{p} \phi_i^2 \alpha_{t-i}^2)^{1/2} \alpha_t^*$$

La parte autoregresiva es: $\varphi_0 + \sum_{i=1}^{p} \varphi_i \alpha_{t-i}$ y la parte ARCH es $(\phi_0 + \sum_{i=1}^{p} \phi_i^2 \alpha_{t-i}^2)^{1/2} \alpha_t^*$, donde la varianza condicional o volatilidad es $(\phi_0 + \sum_{i=1}^{p} \phi_i^2 \alpha_{t-i}^2) = h_t$. El modelo es ergódico y tiene una distribución estacionario si:

$$\left(\sum_{i=1}^{p} |\varphi_i|\right)^2 + \left(\sum_{i=1}^{p} \phi_i\right) E(\alpha_t^*) < 1$$

El AR-ARCH bajo la condición de umbral (TAR-ARCH) de orden p con rezago $k \le p$ es:

$$\alpha_t = \varphi_{10} + \sum_{i=1}^{p} \varphi_{1i} \alpha_{t-i} + (\phi_{10} + \sum_{i=1}^{p} \phi_{1i}^2 \alpha_{t-i}^2)^{1/2} \alpha_t^* \text{ si } \alpha_{t-k} \le 0$$

$$\alpha_t = \varphi_{20} + \sum_{i=1}^{p} \varphi_{2i} \alpha_{t-i} + (\phi_{20} + \sum_{i=1}^{p} \phi_{2i}^2 \alpha_{t-i}^2)^{1/2} \alpha_t^* \text{ si } \alpha_{t-k} \ge 0$$

[23] Parzen, E (1982)
ARARMA Models for Time Series Analysis and Forecasting.
Journal of Forecasting. Vol 1pp 67-82

VIb.-GARCH.

El modelo GARCH es una generalización del ARCH, presentado por Bollerslev (1986), el modelo es: dado el proceso estocástico: $\{\alpha_t; t \in E\}$ entonces la distribución dado el conjunto de información Ω_{t-1} es:

$\alpha_t / \Omega_{t-1} \overset{d}{\approx} N(0, h_t)$

La varianza condicional es:

$h_t = \varphi_0 + \varphi_1 \alpha_{t-1}^2 + \varphi_2 \alpha_{t-2}^2 + ... + \varphi_q \alpha_{t-q}^2 + \gamma_1 h_{t-1} + \gamma_2 h_{t-2} + ... + \gamma_p h_{t-p}$

Donde:

$p \geq 0, q > 0$, $\varphi_0 > 0$, $\varphi_i \geq 0$ para $i = 1,2...q$; $\gamma_i \geq 0$ para $i = 1,2...p$. Este es GARCH(p,d), cuando $p = 1$ y $q = 1$ la varianza condicional es:

$h_t = \varphi_0 + \varphi_1 \alpha_{t-1}^2 + \gamma_1 h_{t-1}$

Si $p = 0$, y $q \neq 0$ entonces el modelo es un ARCH(q)[24], si consideramos ahora, como en el caso del ARCH que:

$Y_t = Z_t \theta + \alpha_t$

Z_t es el vector de variables endógenas y θ es el vector de parámetros. Consideremos los polinomios $\Phi(B) = (\varphi_1 B + \varphi_2 B^2 + ... + \varphi_q B^q)$ y $\Gamma(B) = (\gamma_1 B + \gamma_2 B^2 + ... \gamma_p B^p)$ donde B es el operador de retardo: $B^x w_t = w_{t-x}$. Ahora escribimos la varianza condicional como:

$h_t = \varphi_0 + \Phi(B)\alpha_t^2 + \Gamma(B) h_t$

Si se verifica que todas las raices de $1 - \Gamma(z) = 0$ caen fuera de un círculo de radio unitario, la varianza condicional resulta un ARCH (∞), esto es:

$h_t = \varphi_0 (1 - \Gamma(1))^{-1} + \Phi(B)(1 - \Gamma(B))^{-1} \alpha_t$

$h_t = \varphi_0 (1 - \sum_{i=1}^{p} \gamma_i)^{-1} + \sum_{i=1}^{\infty} \delta_i \alpha_{t-i}^2$

Teorema. Bollerslev (1986):

El GARCH(p,q) definido anteriormente es estacionario en sentido amplio con $E(\alpha_t) = 0$, $Var(\alpha_t) = \varphi_0 (1 - \Phi(1) + \Gamma(1))^{-1}$, $Cov(\alpha_t, \alpha_s) = 0$ $s \neq t$ si y solamente si $\Phi(1) + \Gamma(1) < 1$.

Bollerslev, menciona una forma alternativa de expresar el modelo GARCH:

$$\alpha_t = \varphi_0 + \sum_{i=1}^{q} \varphi_i \alpha_{t-i}^2 + \sum_{j=1}^{p} \beta_j \alpha_{t-j} + \sum_{j=1}^{p} \beta_j \nu_{t-j} + \nu_t$$

Donde:

$\nu_t = \alpha_t^2 - h_t = (\eta_t^2 - 1) h_t$; $\eta_t \overset{i.i.d}{\approx} N(0,1)$. De acuerdo a esto el GARCH(p,q) debe interpretarse como un ARMA en α_t^2 de orden $m = \max(p, q)$ y p respectivamente.

La función de autocorrelación y autocorrelación parcial es como sigue:

La covarianza es:

$\gamma_n = Cov(\alpha_t^2; \alpha_{t-n}^2) = \sum_{i=1}^{q} \varphi_i \gamma_{n-i} + \sum_{i=1}^{p} \beta_j \gamma_{n-i} = \sum_{i=1}^{m} \kappa_i \gamma_{n-i}$ para $n \geq p+1$; $\kappa_i = \varphi_i + \beta_i$ $i = 1,2...q$

Las ecuaciones de Yule-Walker están dada por:

[24] En el artículo de Engle utiliza p en vez de q. La estructura del ARCH es similar a un MA (q) definido sobre los cuadrados de las innovaciones, mientras que el GARCH se asocia al ARMA.

$$\rho_n = \gamma_n / \gamma_0 = \sum_{i=1}^{m} \kappa_i \rho_{n-i} \quad n \geq p+1$$

Denotemos por ϕ_{kk} la k-ésima correlación parcial de α_t^2, para determinar cada una definimos el siguiente sistema de ecuaciones con k incognitas y k ecuaciones:

$$\rho_n = \sum_{i=1}^{k} \phi_{ki} \rho_{n-i} \quad n = 1,2..k$$

Si el proceso es un ARCH(q), entonces:

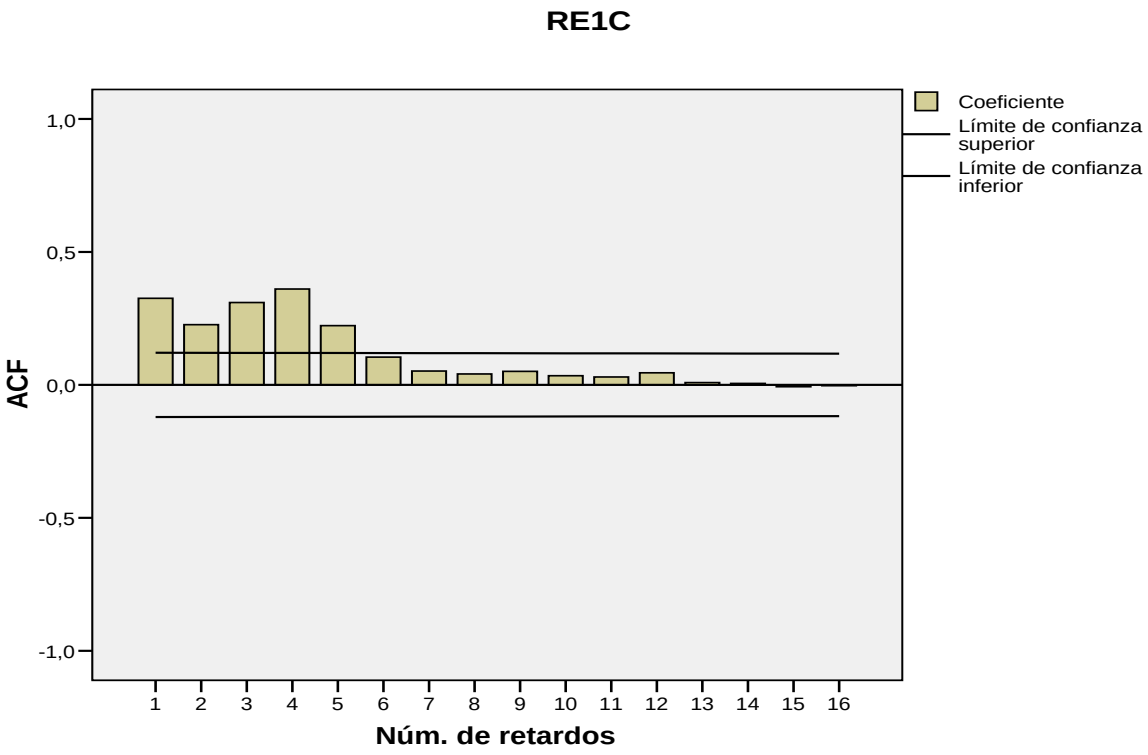
$\phi_{kk} \neq 0$ si $k < q$

$\phi_{kk} = 0$ si $k > q$

Tanto ρ_n como ϕ_{kk} deben estimarse con los datos observados. Se espera que el comportamiento de un modelo GARCH tenga similitud con el ARCH. En la práctica el primer modelo requerirá un número de término menor al ARCH.

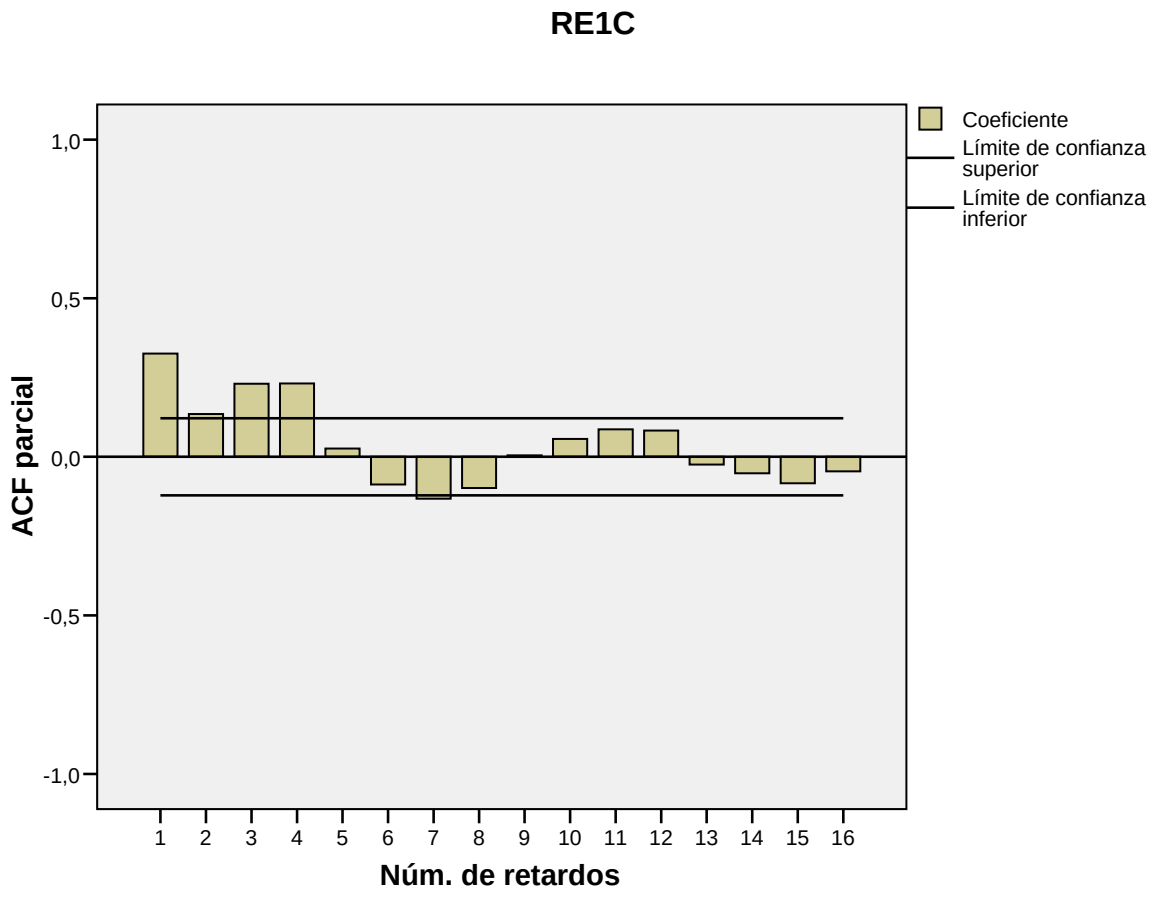
Ejemplo 13.

Tomando la información del ejemplo anterior, representamos las funciones de autocorrelación y autocorrelación parcial de α_t^{*2}



Se puede observar que tanto la función de autocorrelación como la función de autocorrelación parcial hay varios valores que salen de los límites de confianza, por

tanto son significativamente diferente de cero. Un modelo tentativo para la varianza condicional es un ARCH(4). Si se emplea un GARCH debe esperarse un número menor de parámetros a estimar.



VI1b.-Estimación de los parámetros de GARCH-generalización.

La estimación de los parámetros parte del supuesto de que las innovaciones tienen una distribución normal, la exposición se hará considerando los comentarios que hace Green(1999) al artículo original de Bollerslev (1986). Asumiendo que α_t sigue una distribución normal $N(0,h_t)$ Partimos entonces, que la función de máxima verosimilitud es:

$$Ln\prod_{t=1}^{n} f_{Y_t}(y_t;Z_t\gamma,g_t) = \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\ln h^2{}_t + \sum_{t=1}^{n}\left\{-\frac{\alpha_t^2}{h_t^2}\right\}$$

Donde:

$$h_t = \varphi_0 + \varphi_1\alpha_{t-1}^2 + \varphi_2\alpha_{t-2}^2 + ... + \varphi_{t-q}\alpha_{t-q}^2 + \gamma_1 h_{t-1} + \gamma_2 h_{t-2} + ... + \gamma_p h_{t-p}$$

$$\alpha_t = Y_t - Z_t\theta$$

Entonces, el problema es estimar tanto el vector de parámetros: $(\varphi_0,\varphi_1,...\varphi_q,\gamma_1,\gamma_2...\gamma_p)=(\varphi_0,\varphi,\gamma)$ y el vector de parámetros asociado a las variables endógenas θ. Para simplificar la notación haremos:

$$Ln\prod_{t=1}^{n} f_{Y_t}(y_t;Z_t\gamma,g_t)=\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{n}\ln h^2{}_t+\sum_{t=1}^{n}\left\{-\frac{\alpha_t^2}{h_t^2}\right\}=\sum_{t=1}^{n}l_t(\omega)$$

Donde: $\omega=(\theta,\varphi_0,\varphi,\gamma)=(\theta;\lambda)$

$$\partial l_t(\omega)/\partial\lambda=-\frac{1}{2}\left[\frac{1}{h_t^2}-\frac{\alpha_t^2}{\left(h_t^2\right)^2}\right]\partial h_t^2/\partial\lambda=\frac{1}{2}\left(\frac{1}{h_t^2}\right)\partial h_t^2/\partial\gamma\left(\frac{\alpha_t^2}{h_t^2}-1\right)=\frac{1}{2}\left(\frac{1}{h_t^2}\right)f_t v_t=B$$

Para encontrar el estimador, aplicamos la condición de existencia de un punto extremo:

$$\partial l_t(\omega^*)/\partial\lambda=-\frac{1}{2}\left[\frac{1}{h_t^2}-\frac{\alpha_t^2}{\left(h_t^2\right)^2}\right]\partial h_t^2/\partial\lambda=\frac{1}{2}\left(\frac{1}{h_t^2}\right)\partial h_t^2/\partial\gamma\left(\frac{\alpha_t^2}{h_t^2}-1\right)=\frac{1}{2}\left(\frac{1}{h_t^2}\right)f_t v_t=0$$

La resolución de la ecuación anterior exige el empleo de algoritmos de optimización no lineal.

La varianza asintótica de los estimadores se obtiene de: $\left(B^T B\right)^{-1}$, que es equivalente de partir del cálculo de: $\partial^2 l\left(\omega^*\right)/\partial\lambda\partial\lambda$ el cual requiere de la aplicación de algún algoritmo, Bollerslev (1986) propone el desarrollo del algoritmo BHHH. (Berndt,Hall,Hall and Hausman). Green (1999) propone el método de Newton. Aquí estamos asumiendo que no hay parámetro θ de la regresión.

Para estimar el vector de parámetros θ de la regresión y la matriz de varianza consideramos que α_t sigue una distribución normal $N(0,h_t)$ $E(\partial^2 l_t(\omega)/\partial\lambda\partial\theta)=0$

$$\partial l_t(\omega)/\partial\theta=\frac{\alpha_t Z_t}{h_t^2}+\frac{1}{2}\left(\frac{1}{h_t^2}\right)v_t\left(\partial h_t/\partial\theta\right)$$

$$\partial l_t(\omega^*)/\partial\theta=\frac{\alpha_t Z_t}{h_t^2}+\frac{1}{2}\left(\frac{1}{h_t^2}\right)v_t\left(\partial h_t/\partial\theta\right)=0$$

La matriz de varianza covarianza se obtiene de $\partial^2 l_t(\omega)/\partial\theta\partial\theta)$ y su obtención requiere de la aplicación de técnicas de optimización no lineal.

Cuando el supuesto de normalidad de α_t no es apropiado, se ha propuesto como técnica de estimación la del pseudos-máxima verosimilitud donde se ajusta la matriz de varianza covarianza de los estimadores. (Green 1999)

VI2b.-Contraste de hipótesis de los parámetros de GARCH.

Para efectuar el contraste de hipótesis Bollerslev (1986) propone el uso del contraste de multiplicadores de Lagrange sin descartar el posible uso del contraste de Wald. Para ello descompone la varianza condicional como sigue:

$h_t=\omega x=\omega_1 x_1+\omega_2 x_2$

La hipótesis nula es: $H_0:\omega_2=0$, esto quiere decir que los parámetros asociados a $h_{t-1};i=1,2...p$ son nulos y si la hipótesis no se rechaza entonces el modelo es un ARCH.

Recordemos que el contraste de multiplicadores de Lagrange para el caso del modelo de regresión restringido parte de

$$\underset{\beta,\lambda}{Max}\, LnL(\beta;\lambda)_R = \underset{\beta,\lambda}{Max}\left[LnL(\beta;\lambda) + \lambda^T(R\beta - r)\right]$$

De donde es test es:

$$ML = n(\partial LnL(\overset{*}{\beta}_R;\lambda)_R/\partial\beta)^T I(\overset{*}{\beta}_R)^{-1}(\partial LnL(\overset{*}{\beta}_R;\lambda)_R/\partial\beta)/e_R^{*T}e_R^*$$

En este caso el estadístico tiene la forma:

$$ML = \frac{1}{2}g_0^T X_0(X_0^T X_0)^{-1}X_0^T g_0$$

Donde:

$$g_0 = \left(\alpha_1^2 h_1^{-1} - 1, \alpha_2^2 h_2^{-1} - 1, ..\alpha_n^2 h_n^{-1} - 1\right)^T$$

$$X_0 = (h_1\partial h_1/\partial\omega, h_2\partial h_2/\partial\omega, ...h_n\partial h_n/\partial\omega)^T$$

Están evaluados bajo la hipótesis nula.

Este estadístico se distribuye con una χ^2 con tantos grados de libertad como elementos tiene el vector ω_2, un test equivalente es:

$$ML = nR^2$$

Donde R^2 es el coeficiente de correlación múltiple entre g_0 y X_0 que de la misma forma se distribuye con una χ^2 con tantos grados de libertad como elementos tiene el vector ω_2.

Es importante el comentario que hace el autor, para la hipótesis nula de que el modelo es un ARCH(q), el contraste de los multiplicadores de Lagrange, usando como hipótesis alternativas GARCH(r,q) o ARCH(q+r) el resultado se confunde. Siguiendo a Green, el contraste para ARCH(q) frente a GARCH(p,q) es exactamente el mismo que él de ARCH(q) frente a ARCH(p+q).

VI3b Modelos relacionados. INGARCH

En los últimos años hay un creciente interés en aquellos procesos tanto de conjunto de índice como espacio discreto, en especial los procesos de Poisson entre cuyas aplicaciones están: cierto tipo de transacciones que se realiza en el mercado y problemas relacionados con epidemiología, de ahí surge el estudio del modelo INGARCH. Los autores Ferland, R et al (2006) presentan un estudio bastante completo en su artículo donde estudian las propiedades generales de este modelo y luego lo particularizan a un modelo INGARCH(1,1) y finalmente dan una aplicación empleando los datos epidemiológico de la infección de campylobacterosis. Ahora cambiamos el proceso $\alpha_t/\Omega_{t-1} \overset{d}{\approx} N(0,h_t)$ por $\alpha_t/\mathrm{P}_{t-1} \overset{d}{\approx} P(h_t)$ lo que quiere decir que la distribución condicional del proceso dado el conjunto de información "poissoniano", es una distribución de Poisson.

El modelo INGARCH (p,q) se define como:

$$\alpha_t/\mathrm{P}_{t-1} \overset{d}{\approx} P(h_t)\ t \in Z$$

$$h_t = \varphi_0 + \varphi_1\alpha_{t-1}^2 + \varphi_2\alpha_{t-2}^2 + ... + \varphi_{t-q}\alpha_{t-q}^2 + \gamma_1 h_{t-1} + \gamma_2 h_{t-2} + ... + \gamma_p h_{t-p}$$

$p \geq 0, q > 0$, $\varphi_0 > 0$, $\varphi_i \geq 0$ para $i = 1,2...q$; $\gamma_i \geq 0$ para $i = 1,2...p$. Empleando los polinomios en función del operador de retardo: $\Phi(B) = (1 + \varphi_1 B + \varphi_2 B^2 + ... + \varphi_q B^q)$ y

$\Gamma(B) = (1 + \gamma_1 B + \gamma_2 B^2 + ... \gamma_p B^p)$ y asumiendo que las raíces de ambos polinomios caen fuera de un círculo unitario y además $\Gamma(1) = \sum_{i=1}^{p} \gamma_i < 1$, entonces la varianza condicional puede escribirse como sigue:

$h_t = \Gamma^{-1}(B)(\varphi_0 + \Phi(B)\alpha_t) = \varphi_0 \Gamma^{-1}(1) + H(B)\alpha_t$

$H(B) = \Phi(B)\Gamma^{-1}(B) = \sum_{i=1}^{\infty} \psi_i B^i$

Los coeficientes ψ_i son dado por el desarrollo de $\Phi(B)\Gamma^{-1}(B)$. Para que un proceso estacionario de segundo orden satisfaga la definición de INGARCH (p,q) es necesario que $\Gamma(1) - \Phi(1) > 0$ o equivalentemente que $0 \le \sum_{i=1}^{q} \varphi_i + \sum_{i=1}^{p} \gamma_i < 1$.

Para construir INGARCH los autores proceden de la siguiente forma: consideran dos sucesiones de variables aleatorias independientes y distribuidas como una Poisson $\{U_t : t \in Z\}$ y $\{Z_{t,i,j}; t \in Z \wedge (i,j) \in N\}$, la primera sucesión tiene media común: $\psi_0 = \varphi_0 / \Gamma(1)$. Las variables U y Z son variables independientes. Ahora se define la sucesión $\{\alpha_t^{(n)}; t \in Z\}$:

1) $\alpha_t^{(n)} = 0$ para $n < 0$

2) $\alpha_t^{(n)} = U_t$ para $n = 0$

3) $\alpha_t^{(n)} = U_t + \sum_{i=1}^{n} \sum_{j=1}^{X_{t-1}^{(n-i)}} Z_{t-i,i,j}$ para $n > 0$.

Algunas propiedades de esta sucesión de variables aleatorias son:

1) Si $\Gamma(1) - \Phi(1) > 0$, entonces tiene al menos un límite seguro.
2) La sucesión es un proceso estrictamente estacionario para cada n.
3) Si $\Gamma(1) - \Phi(1) > 0$, entonces los dos primeros momentos son finitos.
4) INGARCH(p,d) está relacionado con el ARMA $(\max\{p,q\}, p)$

El modelo INGARCH(1,1) tiene la forma:

$\alpha_t / \mathrm{P}_{t-1} \overset{d}{\approx} P(h_t)\ t \in Z$

$h_t = \varphi_0 + \varphi_1 \alpha_{t-1}^2 + \gamma_1 h_{t-1}$

Las propiedades más importantes son en este caso:

1) la media es: $\mu = \varphi_0 / (1 - \varphi_1 - \gamma_1)$
2) $Var(\alpha_t) = \mu(1 - (\varphi_1 + \gamma_1)^2 + \varphi_1^2) / (1 - (\varphi_1 + \gamma_1)^2)$
3) La función de autocorrelación : $\gamma(r) = \varphi_1(1 - \gamma_1(\varphi_1 + \gamma_1)(\varphi_1 + \gamma_1)^{r-1} \mu / (1 - (\varphi_1 + \gamma_1)^2)$

La estimación de los parámetros de modelo INGARCH es similar al método empleado en GARCH, solo debemos cambiar la función de verosimilitud que en el segundo asume normalidad de α_t. En el caso INGARCH es por tanto:

$L(\Theta) = \prod_{t=1}^{n} \frac{e^{-\lambda_t} \lambda_t^{\alpha_t}}{\alpha_t!}$

VII.-Constraste de linealidad: DBS.

En los últimos años han surgido un conjunto de test para contrastar la hipótesis nula de linealidad de la serie de tiempo versus la alternativa de no linealidad. Uno de lo más reciente es el que presenta Escanio. J.C (2006). En su artículo hace un recuento de los diferentes trabajos que tratan el tema. Después de hacer una breve crítica de los test basado en las correlaciones de los residuos para detectar la buena especifición de un modelo con datos temporales, y de otra naturaleza como los basados en núcleos, presenta su test que parte del concepto de transformada de Fourier explicando y demostrando posteriormente las propiedades del mismo y finalmente presenta un ejemplo vía simulación. Se menciona este trabajo, al inicio de esta parte de la monografía dada su importancia, sin embargo escapa del nivel de la misma. Por tanto presentaremos uno de los test más usados como lo es el test BDS.

BDS.

$$\gamma(2)=\{(Y_0,Y_1),(Y_1,Y_2),(Y_2,Y_3)....(Y_{t-2},Y_{t-1},)\}$$

$$\gamma(3)=\{(Y_0,Y_1,Y_2),(Y_1,Y_2,Y_3),(Y_2,Y_3,Y_4)....(Y_{t-2},Y_{t-1},Y_{t-2})\}$$

$$C_n(\varepsilon)=\sum_{\substack{j,i=1\\ j\neq i}}^{T(n)}\theta\{\varepsilon-\|y_i(n)-y_j(n)\|\}/T(n)(T(n)-1)$$

$\|y_i(n)-y_j(n)\|$ es la distancia euclidea entre $y_i(n)$ y $y_j(n)$.

$$\theta\{\varepsilon-\|y_i(n)-y_j(n)\|\}=1 \text{ si } \|y_i(n)-y_j(n)\|<\varepsilon$$

$$\theta\{\varepsilon-\|y_i(n)-y_j(n)\|\}=0 \text{ si } \|y_i(n)-y_j(n)\|>\varepsilon$$

$$C_n(\varepsilon)=\alpha\varepsilon^n$$

$$W_n(\varepsilon,N)=\sqrt{N}(C_n(\varepsilon)-C_1(\varepsilon)^n)/\sigma_n^*(\varepsilon)$$

El primer test (BDS) debido a los autores Brock, Dechert y Scheinkman asumen como hipótesis nula que la serie responde a una sucesión de variables aleatorias independientes e idénticamente distribuidas en donde para una dimensión n, $C_n(\varepsilon)=C_1(\varepsilon)^n$. La hipótesis alterna es que la serie responde a un sistema determinístico o estocástico no lineal. Para contrastar la hipótesis nula proponen el estadístico:

$$W_n(\varepsilon,N)=\sqrt{N}\,(C_n(\varepsilon)-C_1(\varepsilon)^n)/\sigma_n^*(\varepsilon)$$

Dado: $C_n(\varepsilon)$, $C_1(\varepsilon)^n$, $\sigma_n^*(\varepsilon)$ es la desviación estándar de $(C_n(\varepsilon)-C_1(\varepsilon)^n)$. Esta viene dada como:

$$\sigma_n^*(\varepsilon)=4\{h^n+2\,4\sum_{j=1}^{n-1}h^{m-j}c^{2j}+(n-1)^2c^{2n}-n^2hc^{2n-2}\}$$

Para obtener el valor de h simplificaremos la notación de la función $\theta\{\varepsilon-\|x_i(n)-x_j(n)\|\}$ como $H(i,j)$.

$$h=2\sum_{t=1}^{n}\sum_{s=t+1=s+1}^{n}\sum^{n}(H(t,s)H(s,r)+H(t,r)H(r,s)+H(s,t)H(t,r))/(n(n-1)(n-2))$$

Bajo la hipótesis nula este estadístico se distribuye como una normal N(0,1). Si el valor del estadístico es tal que $P\left(W_n(\varepsilon,N)\geq\left|W_n^*(\varepsilon,N)\right|\right)<\alpha$ entonces rechazamos la hipótesis nula. El otro test conocido como test residual de Brock, consiste en utilizar un AR(p) para ajustar la serie, si el coeficiente de correlación espacial de la serie de los residuos y de la serie x son significativamente diferentes entonces sospechamos la existencia de una estructura caótica.

El cociente entre $C_n(\varepsilon)$ y ε cuando ε tiende a cero da una nueva dimensión conocida como dimensión de correlación:

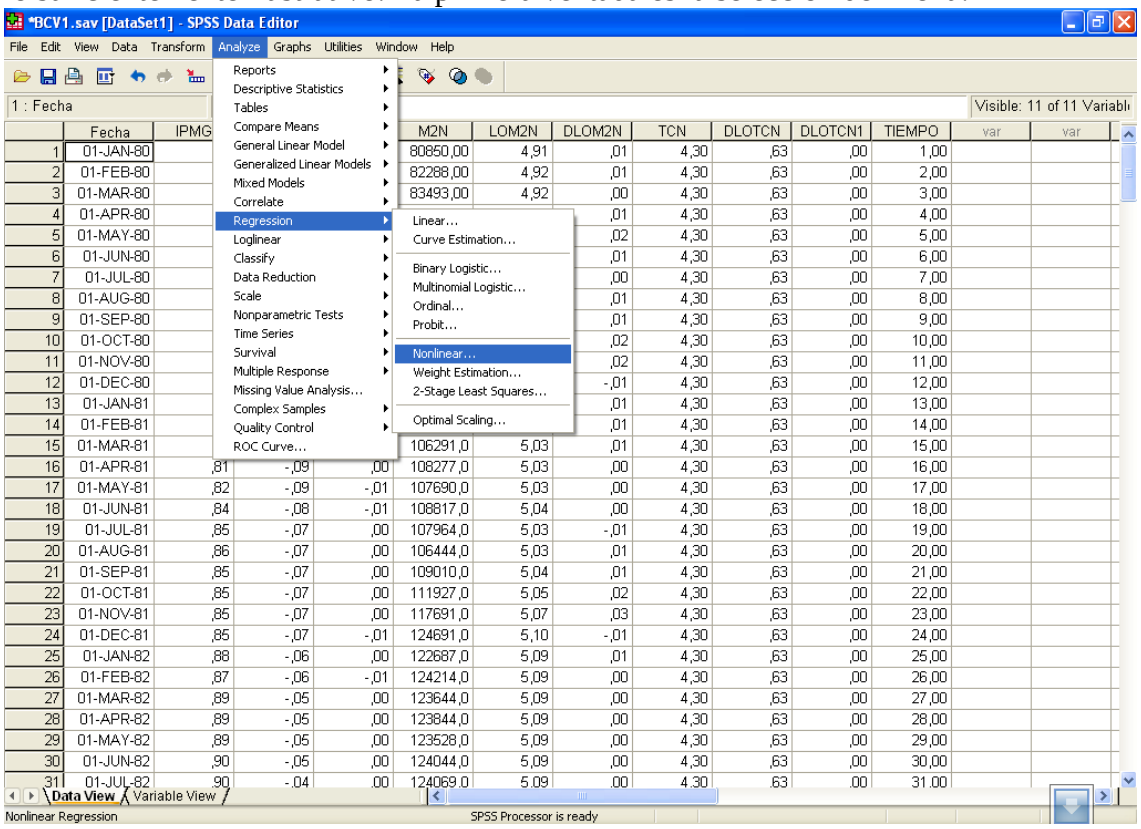
$$D(n)=\lim_{\varepsilon\to 0} C_n(\varepsilon)/\varepsilon$$

Si el proceso es aleatorio D(n) tiende a infinito.

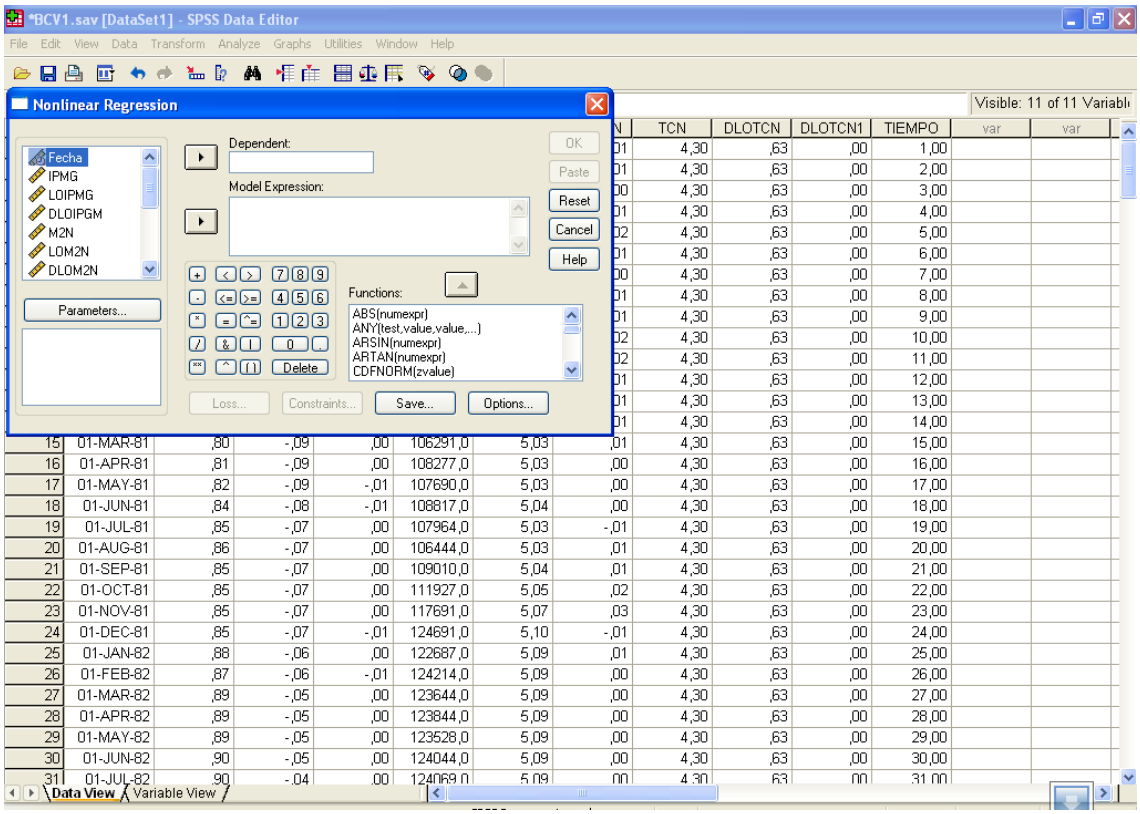
(Internet**:**Gimeno R. Matilla Garcia,M. Rodríguez Ruiz, J; Sanz Carnero B. Perez Espartero, Ana).

ANEXO 1 SPSS: REGRESIÓN NO LINEAL LIBRE Y RESTRINGIDA.

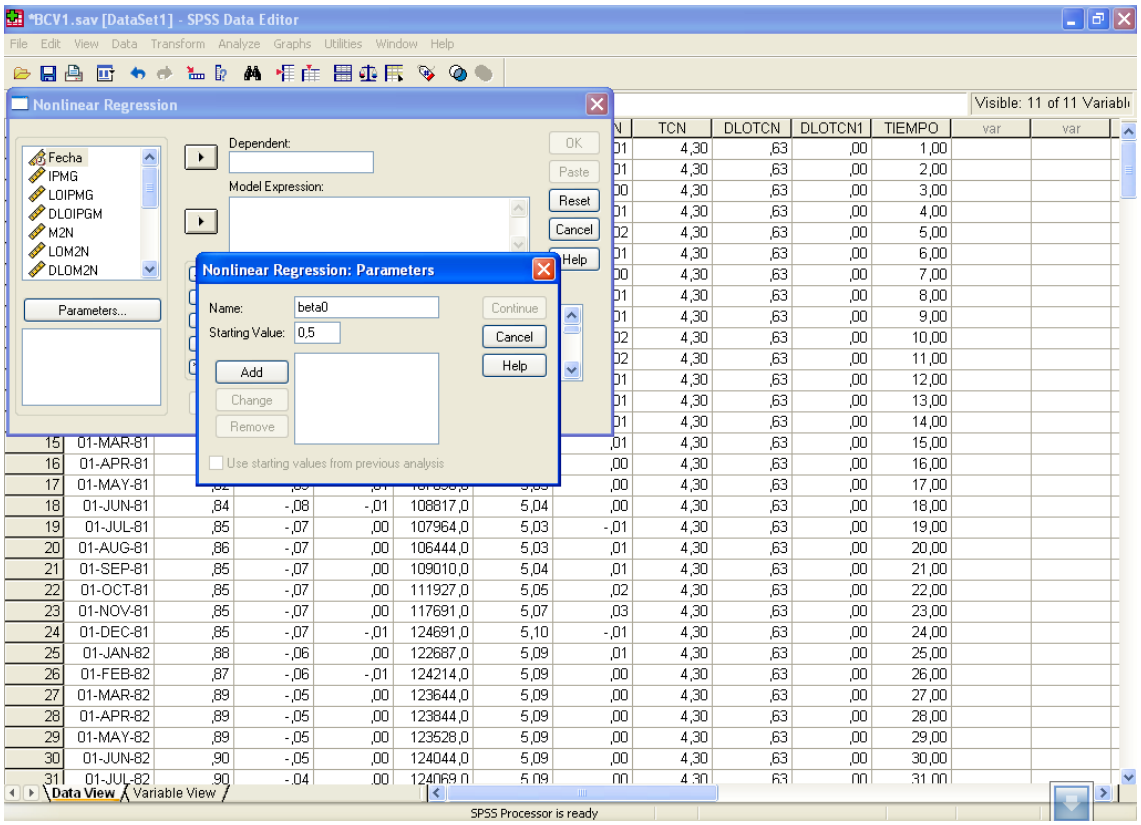
El siguiente anexo presenta las ventanas del SPSS para el problema de regresión no lineal, la visualización de la mismas debe ser suficiente para un lector avisado, sin embargo nos tomaremos la libertad de explicar la misma, sin tener la pretensión de dictar un curso de SPSS que los hay mejores, en mucho. Suponemos que el usuario conoce el manejo de la base de datos de este paquete tan popular entre alumnos y especialista, por tanto, aunque parezca trivial la explicación que sigue lo haremos para aquellos que no están tan familiarizado, aunque insistimos, ver la secuencia debería ser lo suficientemente ilustrativo. La primera ventada es la selección del menú.



Al seleccionar ***Nonlinear***, obtendrá una nueva ventana donde aparece la lista de variables.



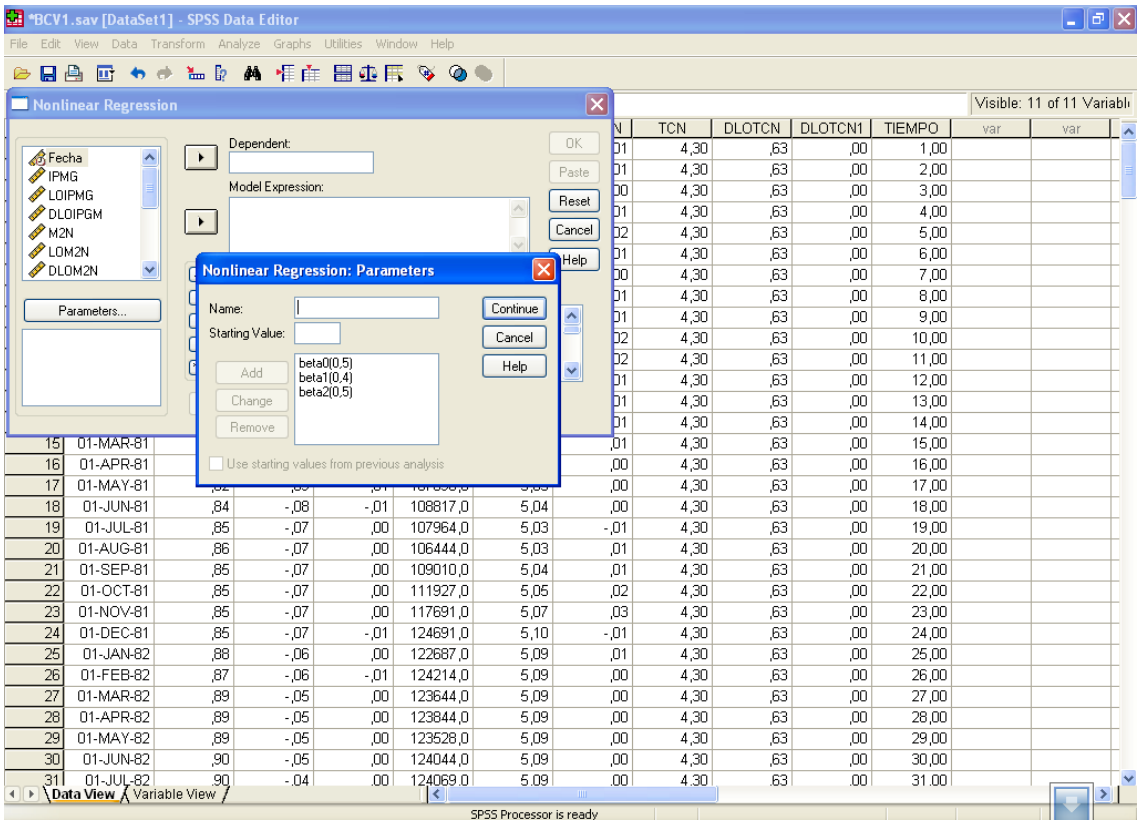
Además una casilla donde debe especificar los parámetros, al activar el botón ***parameters***, aparecerá una nueva ventana donde debe colocar el nombre del primer parámetro y el valor inicial, al terminar debe activar el botón ***Add,*** luego escribir el nombre del segundo parámetro y su valor adicional y, activar nuevamente el botón ***Add.*** Este procedimiento continua tanta veces como parámetro tiene el modelo.

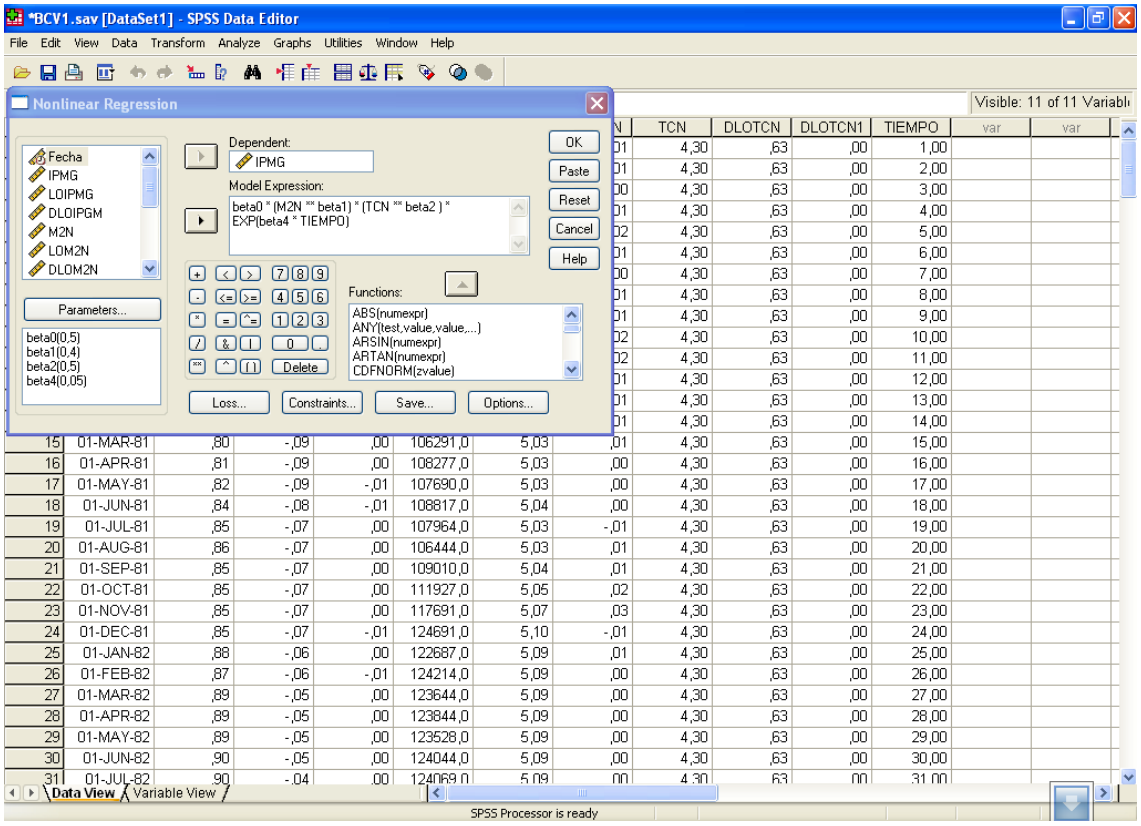


Al finalizar la escritura de todos los parámetros active continuar: ***Continue***. Con este último paso se presentara en la ventana ***parameers*** la lista de los nombres de los parámetros con los valores iniciales seleccionados entre paréntesis. Si desea cambiar o remover algún parámetro, debe activar nuevamente la ventana ***parameters***, al hacerlo aparecerá la ventana con los parámetros que utilizó para signarle nombres y valor iniciales y podrá usar los botones ***remove*** y ***change*** Esta misma ventana permite cambiar o remover un parámetro o su valor inicial.

Una vez realizado este procedimiento, escriba el nombre de la variable dependiente en la casilla correspondiente.

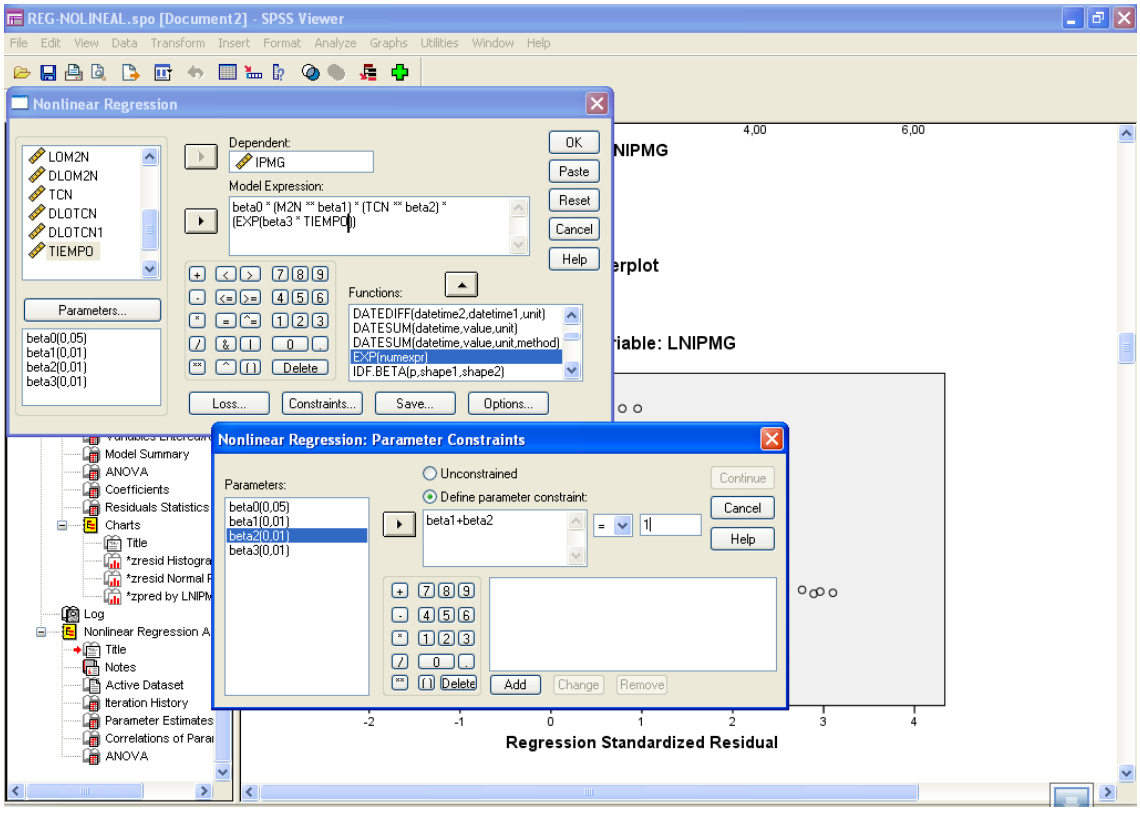
En la casilla **Model expression** introduzca los parámetros indicando si están multiplicado las variables o si son exponentes de las mismas, para ello emplee el recuadro de números y símbolos o la listas de funciones. En la listas de símbolos cuenta con paréntesis, el símbolo de multiplicar con un asterisco etc. En el caso de necesitar alguna función preestablecida en la lista de las mismas debe recordar que debe indicarle el argumento, este aparecerá con un símbolo de interrogación hasta tanto usted no lo defina.

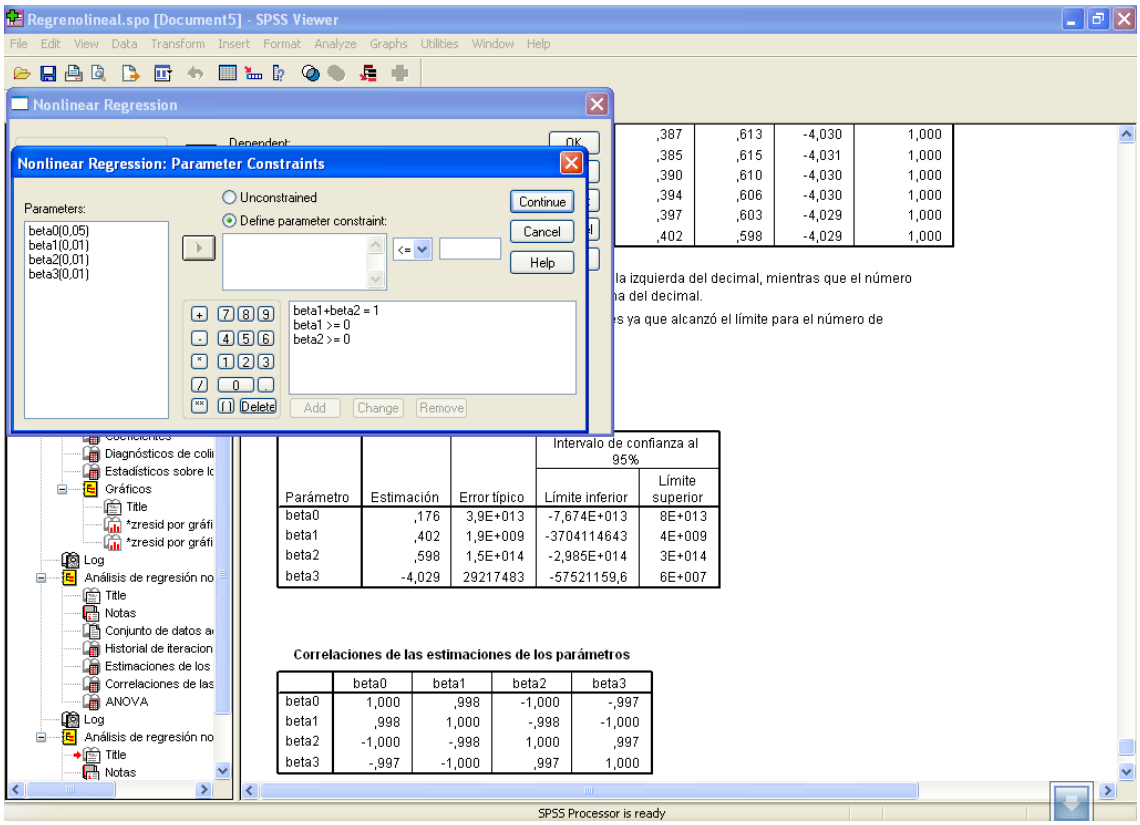

*BCV1.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Nonlinear Regression
Dependent:
Model Expression:
Fecha
IPMG
LOIPMG
DLOIPGM
M2N
LOM2N
DLOM2N
Parameters...
OK
Paste
Reset
Cancel
Help
Nonlinear Regression: Parameters
Name:
Starting Value:
Add
Change
Remove
beta0(0,5)
beta1(0,4)
beta2(0,5)
Continue
Cancel
Help
Use starting values from previous analysis
Visible: 11 of 11 Variabl
TCN DLOTCN DLOTCN1 TIEMPO var var
Data View Variable View
SPSS Processor is ready


*BCV1.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Nonlinear Regression
Dependent:
IPMG
Model Expression:
beta0 * (M2N ** beta1) * (TCN ** beta2) *
EXP(beta4 * TIEMPO)
Fecha
IPMG
LOIPMG
DLOIPGM
M2N
LOM2N
DLOM2N
Parameters...
beta0(0,5)
beta1(0,4)
beta2(0,5)
beta4(0,05)
Functions:
ABS(numexpr)
ANY(test,value,value,...)
ARSIN(numexpr)
ARTAN(numexpr)
CDFNORM(zvalue)
Delete
Loss...
Constraints...
Save...
Options...
OK
Paste
Reset
Cancel
Help
Visible: 11 of 11 Variabl
TCN DLOTCN DLOTCN1 TIEMPO var var
Data View Variable View
SPSS Processor is ready

Concluido lo señalado active el botón ***OK***.
Si el modelo está especificado con una o varias restricciones sobre los parámetros active el botón ***constraints*** y marque el redondillo: ***Define parameters constraints***, entonces podrá escribir cada restricción con la ayuda del cuadro de números y símbolos. Al lado derecho de la casilla donde escribe la forma de la restricción, tiene para seleccionar el tipo de disigualda y el valor númérico con la se compara la restricción.
Repita este procedimiento tanta veces como restricciones tiene el modelo. Al finalizar le dará al botón continuar. Hay que tener cuidado en el momento de especificar las restricciones pues el paquete tiene ciertas limitaciones si en la ventana se seleccionó como ***Options*** **: Estimation Method Levenberg-Marquardt,** que es un método de estimación que está entre el método de la expansión de Taylor y el método de la pendiente descendiente.

Los otros botones que están en la primera ventana son: 1) **Reset,** que borra todo lo escrito si se quiere modificar sustancialmente el modelo. 2) **Save,** que permite guardar en la base de datos los valores predichos por el modelo, los residuos y las derivadas de los parámetros evaluadas en cada punto.3)**Loss,** permite definir una función de pérdida o penalización. 4) **Paste,** que permite obtener el programa del procedimiento empleado.




Regrenolineal.spo [Document5] - SPSS Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help
Nonlinear Regression
Dependent:
OK
Nonlinear Regression: Parameter Constraints
Parameters:
beta0(0,05)
beta1(0,01)
beta2(0,01)
beta3(0,01)
Unconstrained
Define parameter constraint:
Continue
Cancel
Help
beta1+beta2 = 1
beta1 >= 0
beta2 >= 0
Add
Change
Remove
Delete
Diagnósticos de coli
Estadísticos sobre lc
Gráficos
Title
*zresid por gráfi
*zresid por gráfi
Log
Análisis de regresión no
Title
Notas
Conjunto de datos a
Historial de iteracion
Estimaciones de los
Correlaciones de las
ANOVA
Log
Análisis de regresión no
Title
Notas
,387 ,613 -4,030 1,000
,385 ,615 -4,031 1,000
,390 ,610 -4,030 1,000
,394 ,606 -4,030 1,000
,397 ,603 -4,029 1,000
,402 ,598 -4,029 1,000
la izquierda del decimal, mientras que el número
na del decimal.
es ya que alcanzó el límite para el número de
Intervalo de confianza al 95%
Parámetro Estimación Error típico Límite inferior Límite superior
beta0 ,176 3,9E+013 -7,674E+013 8E+013
beta1 ,402 1,9E+009 -3704114643 4E+009
beta2 ,598 1,5E+014 -2,985E+014 3E+014
beta3 -4,029 29217483 -57521159,6 6E+007
Correlaciones de las estimaciones de los parámetros
beta0 beta1 beta2 beta3
beta0 1,000 ,998 -1,000 -,997
beta1 ,998 1,000 -,998 -1,000
beta2 -1,000 -,998 1,000 ,997
beta3 -,997 -1,000 ,997 1,000
SPSS Processor is ready

REFERENCIA BIBLIOGRÁFICA.

Audrino, F. (2005)
Local Likelihood for non parametric ARCH(1) Models.
Journal of Time Series Analysis Vol 26, Nº 2 pag. 252-278.

Avriel, M (2003)
Nonlinear Programming-Analysis and Methods
Dover Publications. INC. –New York.

Bazaraa, M.S.; Shetty, C.M (1979)
NonLinear Programming. Theory and Algorithms.
Jonh –Wiley and Sons.-New York.-USA.

Blake, A.P and Kapetanios. G (2003)
Pure significance test of unit root hypothesis against nonlinear alternatives.
Journal of Time Series Analysis Vol 24, Nº3 pag. 254-267.

Bollerslev, T. (1986)
Generalized Autoregressive Conditional Heteroskedasticity.
Journal of Econometrics 31; pag. 307-327.

Byers, J.D and Peel, D.A (1995)
Foecasting Industrial Production Using Non-Linear Methods.
Journal of Forecasting, Vol 14, pag. 325-336.

Chandr,. A and Taniguchi Masanobu (2006)
Minimum α-divergence Estimation for ARCH Models.
Journal of Time Series Analysis Vol 27, Nº1 pag. 19-39.

Cline, D.B.H (2006)
Evaluating the Lyapunov Exponent and Existence of Moments for Threshold AR-ARCH Models.
Journal of Time Series Analysis Vol 28, Nº2 pag. 242-260.

Dhrymes, P.J (1970)
Econometrics-Statistical foundations and Applications.
Harper International Edition.-New York.

Draper, N.R and Smith.H (1981)
Applied Regression Analysis. 2ª Edition.
John Wiley and Sons.Inc. New York.

Engle, R.F (1982)
Autoregresive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.
Econometrica, Vol 50, Pag 987-1007.

Escanciano, J.C (2006)
Goodness of Fit Test for Linear and Nonlinear Time Series Models.
Journal of American Statistical Association. Vol 101, Nº 474 pag. 531-541.

Ferland, R. ; Latour, A. ;Oraichi, D(2006)
Integer-Valued GARCH Process.
Journal of Time Series Analysis Vol 27, Nº2 pag. 923-942.

Green, W.H (1999)
Análisis Econométrico. 3ª Edición.
Prentice Hall-Madrid-España.

Gujarati, D.N (2004)
Econometría 4ª Edición.
McGraw Hill-México.

Koopman, J.S. ; Ooms, M. ; Carnero,M.A. (2007)
Periodic Seasonal Reg-AFIRMA-GARCH Models for Daily Electricity Spot Prices.
Journal of American Statistical Association. Vol 102, Nº 477 pag. 16-27.

Kumar, K (1986)
On The Identification of Some Bilinear Time Series Models.
Journal of Time Series Analysis Vol 7, Nº2 pag. 117-122.

Lai, D.;Chen, G (2003)
Distribution of the Estimated Lyapunov Exponents from Noisy Time Series.
Journal of Time Series Analysis Vol 24, Nº 6 pag. 706-720.

Maddala,G,S (1986)
Introduccion a la Econometría. 2ª Edición.
Prentice Hall-Madrid-España.

Marcelo, M.C. ; Veiga, A (2003)
Diagnostic Checking in a Flexible Nonlinear Time Series Model.
Journal of Time Series Analysis Vol 24, Nº 4 pag. 461-481.

Novales, A (1993)
Econometría. 2ª Edición.
McGraw Hill.-México.

Pindyck, R.S. ; Rubinfeld, D.L (2000)
Econometría:-Modelos y Pronósticos. 4ª Edición.
McGraw Hill.-México.

Samoro, A.; Spokoiny, V. ;Vial, C (2005)
Component Identification and Estimation in Nonlinear High_Dimension Regression Models by Structural Adaptation.
Journal of American Statistical Association. Vol 100, Nº 470 pag. 429-445.

Rochon, J.(1992)
ARMA Covariance Structure with Time Heteroscedasticity for Repeated Measures Experiments.
Journal of American Statistical Association. Vol 87, Nº 419 pag. 777-784.

Theil, H (1971)
Principles of Econometrics.
John Wiley and Sons.Inc. New York.

Tsay, R.S (1987)
Conditional Heteroscedasticic Time Series Models.
Journal of American Statistical Association. Vol 82, Nº 398 pag. 590-604.

Zellner, A.,and M. Geisel(1970)
Analysis of Distributed Lag Models wiht Application to the Consumption Funtion
Econometrica, 38, pag 865-888.