

GUIA PARA LA ESTIMACION DE
MODELOS ARCH

Alfonso Novales
Documento 90-11

Julio 1990

El autor agradece los comentarios recibidos de
Juan Ayuso y M^a Luisa de la Torre.

CONTENIDO

1. Introducción
2. Justificación del modelo de
 heterocedasticidad autoregresiva
3. El modelo univariante ARCH
 - 3.1 Propiedades del modelo AR(1) con
 estructura ARCH
 - 3.2 Momentos de orden superior
 - 3.3 Función de verosimilitud del modelo
 ARCH(p) univariante
4. El modelo ARCH de regresión
 - 4.1 Estimación de máxima verosimilitud del
 modelo ARCH de regresión
 - 4.2 Procedimiento de estimación sugerido
 - 4.3 Contrastes de estructura ARCH
5. El modelo ARCH en media (ARCH-M)
 - 5.1 Estimación del modelo ARCH-M
 - 5.2 Casos particulares
6. El modelo ARCH generalizado (GARCH)
 - 6.1 El modelo GARCH(1,1)
 - 6.2 El modelo GARCH(p,q)
 - 6.3 Identificación de modelos GARCH
 - 6.4 Contrastes de estructura GARCH
 - 6.5 Estimación de máxima verosimilitud del
 modelo GARCH de regresión

1. INTRODUCCION

La creciente disponibilidad de datos procedentes de los mercados financieros hace que un número creciente de investigadores dediquen parte de sus esfuerzos al análisis empírico de las múltiples cuestiones teóricas de interés relativas a tales mercados. Dentro de este área de trabajo, la modelización tipo ARCH ha sido utilizada en múltiples ocasiones, sin duda debido a los aceptables resultados empíricos que ha venido proporcionando.

A pesar de la existencia de software destinado, más o menos específicamente, a la estimación de este tipo de modelos, puede ser preferible, en ocasiones, llevar a cabo la programación de la estimación máximo-verosímil de la estructura ARCH por parte del investigador. Ello se debe a que, en algunos casos, las rutinas de evaluación numérica de las derivadas de primer y segundo orden incorporadas en los paquetes estadísticos sugeridos no son suficientemente precisas como para garantizar la convergencia del proceso de estimación. Por todas estas razones, el investigador puede preferir tener total control del código de programación que utiliza en la estimación del modelo.

En nuestra propia experiencia, hemos comprobado que, en ocasiones, un proceso de estimación de máxima verosimilitud que resultaba no convergente utilizando expresiones analíticas sólo para las primeras derivadas de la función de verosimilitud, ha resultado rápidamente convergente al utilizar dichas expresiones para las primeras y segundas derivadas. También hemos comprobado que los problemas de convergencia del

algoritmo de estimación cuando no se utilizan expresiones analíticas para las derivadas parecen más frecuentes en modelos con pocas variables explicativas. Por el contrario, modelos de regresión con estructura ARCH y con un cierto número de variables explicativas relevantes han alcanzado generalmente la convergencia en la estimación de máxima verosimilitud sin necesidad de recurrir a segundas derivadas analíticas.

Una segunda motivación para este artículo consiste en la utilidad de recoger en un mismo trabajo las diversas especificaciones de modelos con heterocedasticidad condicional.

El propósito de este trabajo es, por tanto, proporcionar una introducción a los modelos ARCH, desde el punto de vista de la estimación de dichos modelos, prestando especial atención a las expresiones analíticas del vector gradiente y la matriz hessiana precisas para utilizar el algoritmo iterativo numérico necesario para la estimación de máxima verosimilitud. Es en este último aspecto donde reside su contribución original, por no aparecer de forma suficientemente explícita dichas derivadas en los artículos habitualmente referidos en esta literatura. La formulación se hace de modo suficientemente genérico como para que cualquier potencial usuario pueda trasladar directamente las expresiones que aquí se proporcionan a su propia aplicación, de modo que pueda programar el algoritmo de estimación de máxima verosimilitud sin mucho esfuerzo.

2. JUSTIFICACION DE LA MODELIZACION DE HETEROCEDASTICIDAD AUTOREGRESIVA

Cuando se considera el modelo univariante:
 $y_t = \rho y_{t-1} + \epsilon_t$ con $|\rho| < 1$ y ϵ_t ruido blanco $\sim N(0, \sigma_\epsilon^2)$
 se tiene que la esperanza matemática del proceso
 estocástico y_t es 0 y que su varianza

es $\frac{\sigma_\epsilon^2}{1-\rho^2}$, donde σ_ϵ^2 es la varianza de ϵ_t .

De modo más general, si el modelo fuese:

$$y_t = \alpha + \rho y_{t-1} + \epsilon_t, \quad |\rho| < 1$$

se tendría que: $E(y_t) = \frac{\alpha}{1-\rho}$ y $Var(y_t) = \frac{\sigma_\epsilon^2}{1-\rho^2}$,

por lo que la inclusión de un término constante
 afecta a la esperanza, pero no a la varianza del
 proceso.

Los términos "esperanza" y "varianza" que
 acabamos de usar son incondicionales, o dicho de
 otro modo, corresponden a la distribución marginal
 de y_t . Alternativamente, podemos preguntarnos
 acerca de la distribución condicional de y_t dado
 el conjunto de información Ω_{t-1} de que se dispone
 en el instante $t-1$ que suponemos que incluye todo
 el pasado de la variable y_t : $\Omega_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$.

En tal caso:

$$E(y_t / y_{t-1}) = \alpha + \rho y_{t-1}$$

por lo que: $y_t - E(y_t / y_{t-1}) = \epsilon_t$

mientras que:

$$\text{Var}(y_t/y_{t-1}) = E_{t-1}(y_t - E(y_t/y_{t-1}))^2 = E_{t-1}\epsilon_t^2 = \sigma_\epsilon^2 < \frac{\sigma_\epsilon^2}{1-\rho^2} = \text{Var}(y_t)$$

donde estamos suponiendo que el modelo univariante está correctamente especificado y, en particular, que y_{t-1} realmente resume toda la información contenida en Ω_{t-1} acerca de y_t . Es importante observar que, aun siendo distintas, las varianzas condicional e incondicional son ambas constantes en el tiempo.

Esta distinción puede generalizarse a todo modelo de regresión: $y_t = x_t'\beta + u_t$, con variables x_t deterministas y u_t ruido blanco, con distribución Normal. La distribución de y_t condicional en Ω_{t-1} es, en este caso, también Normal, con esperanza y varianza: $Ey_t = x_t'\beta$; $\text{Var}(y_t) = \sigma_u^2$. Sin embargo, si u_t tiene autocorrelación de orden 1, por ejemplo: $u_t = \rho u_{t-1} + \epsilon_t$, donde ϵ_t es ruido blanco, entonces la distribución marginal de y_t es:

$$N(x_t'\beta, \sigma_u^2), \text{ con } \sigma_u^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}, \text{ mientras que la}$$

distribución condicional es:

$$y_t/\Omega_{t-1} \sim N(\rho y_{t-1} + (x_t' - \rho x_{t-1}')\beta; \sigma_\epsilon^2)$$

con menor varianza que la distribución marginal (recordemos que $\sigma_\epsilon^2 = \sigma_u^2 \cdot (1-\rho^2)$).

Utilizar como fórmula de previsión la esperanza condicional $E(y_t/\Omega_{t-1})$ resulta óptimo si se quiere minimizar el error cuadrático medio (ECM) de la previsión resultante. Este tipo de

mecanismo de previsión (que puede obtenerse, bajo determinados supuestos, a partir de un modelo univariante Box-Jenkins) genera buenos resultados en comparación con procedimientos más complejos, como la utilización de modelos econométricos de gran tamaño. Sin embargo, no se utiliza la varianza condicional de y_t para tratar de mejorar estas previsiones y ésta puede ser una información relevante.

¿En qué sentido puede utilizarse la varianza de la distribución condicional para mejorar la previsión?. Si dicha varianza es constante y el objetivo es obtener la previsión con menor ECM, esta información no es relevante; sin embargo, si la varianza no es constante, ignorar tal hecho conduciría a estimadores ineficientes y por ello, con intervalos de confianza para la previsión más amplios, así como con mayor variabilidad de la propia previsión puntual, como reflejaría la mayor amplitud del rango de sus valores posibles, para un nivel de confianza dado.

Los textos habituales de Econometría describen los procedimientos para mejorar la eficiencia de las estimaciones en situaciones de heterocedasticidad, especialmente cuando ésta se hace depender explícitamente de variables de dentro o fuera del modelo. Se describen asimismo los contrastes de la hipótesis nula de ausencia de heterocedasticidad. No es tan habitual encontrar referencias a una situación especial de heterocedasticidad que puede ocurrir en datos de series temporales y cuya detección y tratamiento ha sido el objeto de diversos trabajos de investigación en los últimos años.

Supongamos, por ejemplo, que los errores de estimación (residuos) grandes y pequeños de un modelo de regresión aparecen agrupados a intervalos; ello sugiere que la varianza del término de error cambia suavemente en el tiempo, y que podría predecirse a partir de los residuos previos. De este modo, se especificaría una dependencia del tipo: $\sigma_t^2 = f(\epsilon_{t-1}^2, \epsilon_{t-2}^2, \dots)$ donde σ_t^2 denota la varianza de ϵ_t , cambiante en el tiempo, y f es una función continua del vector de residuos pasados, elevados al cuadrado¹.

Existen, además, teorías que requieren la modelización y estimación numérica de variables que representan la volatilidad de y_t (generalmente de su componente no predecible), que pueden asociarse al concepto estadístico de varianza condicional. Así, en Finanzas, las carteras de activos de inversión se escogen a partir de la media y varianza esperadas de las tasas de rendimiento. Cambios de cartera deben explicarse por estos factores; si se especifica un modelo econométrico tradicional para el rendimiento medio esperado, su varianza queda restringida a una constante y no es, por tanto, consistente con el supuesto teórico de partida.

Necesitamos, por tanto, desarrollar modelos econométricos en que se permita que la varianza del proceso de perturbación cambie en el tiempo. Una posible alternativa son los modelos ARCH, que describimos en las secciones siguientes.

¹ Los residuos se incluyen como argumentos de f elevados al cuadrado, para hacerlos comparables con la varianza σ_t^2 .

3. EL MODELO UNIVARIANTE ARCH: (Autoregressive Conditional Heteroskedasticity)

Consideremos la siguiente generalización de un proceso de ruido blanco:

$$y_t = \epsilon_t h_t,$$

donde ϵ_t es ruido blanco, con $\sigma_{\epsilon}^2 = 1$

y h_t es una función contenida en el conjunto de información disponible en el instante $t-1$, Ω_{t-1} .

En consecuencia, se tiene que $E y_t = 0$, y también:

$$\begin{aligned} \text{Var } y_t &= \sigma_t^2 = E y_t^2 = E[E_{t-1}(\epsilon_t^2 h_t^2)] = \\ &= E[h_t^2 \cdot E_{t-1} \epsilon_t^2] = E(h_t^2 \sigma_{\epsilon}^2) = E h_t^2 \end{aligned}$$

$$\text{Var}_{t-1} y_t = E_{t-1}(\epsilon_t^2 h_t^2) = h_t^2 E_{t-1} \epsilon_t^2 = h_t^2$$

Si suponemos, por ejemplo, que:

$$h_t^2 = \delta_0 + \delta_1 y_{t-1}^2, \quad \text{con } \delta_0 > 0, \delta_1 \geq 0,$$

se tiene el modelo ARCH de orden 1, puesto que la varianza condicional depende de un retardo de y_t . En términos de la distribución condicionada de y_t , suponiendo además Normalidad, se tiene el modelo:

$$(1) \quad \left| \begin{array}{l} y_t / \Omega_{t-1} \text{ se distribuye } N(0, h_t^2) \\ h_t^2 = \delta_0 + \delta_1 y_{t-1}^2, \delta_0, \delta_1 > 0 \end{array} \right.$$

o más generalmente: $h_t^2 = h(y_{t-1}, \dots, y_{t-p}, \delta)$.
[Modelo ARCH de orden p].

3.1 Propiedades del modelo AR(1) con estructura ARCH.

Examinemos las propiedades de un proceso univariante AR(1) con estructura ARCH(1). En particular, prestaremos especial interés a las semejanzas y diferencias que presenta con respecto a un proceso AR(1) habitual. Sea el proceso:

$$y_t = \rho y_{t-1} + \epsilon_t, \quad |\rho| < 1, \quad \text{donde} \quad E_{t-1} \epsilon_t = 0,$$

$$E_{t-1} \epsilon_t^2 = h_t^2, \quad \text{y} \quad h_t^2 = \delta_0 + \delta_1 \epsilon_{t-1}^2$$

Se tiene:

$$E y_t = 0 \quad \text{y} \quad \text{Var } y_t = \frac{1}{1 - \rho^2} \sigma_\epsilon^2$$

ahora bien, $\text{Var } \epsilon_t = E \epsilon_t^2 = E E_{t-1} \epsilon_t^2 = E h_t^2$, pero:

$E h_t^2 = \delta_0 + \delta_1 E \epsilon_{t-1}^2 = \delta_0 + \delta_1 E h_{t-1}^2$ y, bajo el supuesto de estacionariedad, se tiene:

$E h_t^2 = E h_{t-1}^2$, por lo que si $|\delta_1| < 1$, entonces:

$$\text{Var } \epsilon_t = E h_t^2 = \frac{\delta_0}{1 - \delta_1}, \quad \text{por lo que:} \quad \text{Var } y_t = \frac{1}{1 - \rho^2} \cdot \frac{\delta_0}{1 - \delta_1}$$

Además: $E \epsilon_t \epsilon_{t-1} = E[E_{t-1} (\epsilon_t \epsilon_{t-1})] = (E \epsilon_{t-1})(E_{t-1} \epsilon_t) = 0$, y lo mismo ocurre para desfases mayores, por lo que ϵ_t no presenta autocorrelación. Sin embargo, las perturbaciones ϵ_t no son independientes, pues sus momentos de segundo orden están correlacionados.

El modelo ARCH(1) puede escribirse también:
 $h_t^2 - \sigma_\epsilon^2 = \delta_0 \cdot (\epsilon_{t-1}^2 - \sigma_\epsilon^2)$, por lo que la varianza condicional del término de error, h_t^2 , excede a su varianza incondicional, σ_ϵ^2 , en aquellos períodos en que el cuadrado del residuo previo excede a su varianza. Es decir, sorpresas altas (positivas o negativas) hacen que la varianza condicional de ϵ_t el período siguiente sea asimismo elevada.

En cualquier caso, es importante observar que tanto la esperanza como la varianza condicionales de y_t dependen de la información disponible en cada instante y no son, por tanto, constantes.

Nótese que puesto que h_t^2 es la varianza condicional de ϵ_t , ha de ser positiva, por lo que los parámetros δ_0 y δ_1 deben ser positivos. Para los resultados anteriores es preciso, además, que $|\delta_1| < 1$. Una estimación de $\delta_0 < 0$, aun pudiendo ser compatible con $h_t \geq 0$ a lo largo de todo el intervalo muestral utilizado, debe generar sospechas de mala especificación de la estructura ARCH.

3.2 Momentos de orden superior

Si suponemos que ϵ_t sigue una distribución Normal, sus momentos de orden impar serán nulos. En cuanto a los momentos de orden par, ya hemos

visto que $E\epsilon_t^2 = \frac{\delta_0}{1-\delta_1}$. Para obtener los de orden

superior, recordemos que el momento de orden 4, respecto al origen, de una variable Normal es igual a 3 veces el cuadrado de su varianza, es decir:

$$\begin{aligned}
E\epsilon_t^4 &= 3[E\epsilon_t^2]^2 = 3[E(E_{t-1}\epsilon_t^2)]^2 = E[3(\delta_0 + \delta_1\epsilon_{t-1}^2)^2] = \\
&= 3\delta_0^2 + 6\delta_0\delta_1 \frac{\delta_0}{1-\delta_1} + 3\delta_1^2 E\epsilon_{t-1}^4
\end{aligned}$$

Bajo el supuesto de estacionariedad, es decir, si $|\delta_1| < 1$, se tiene:

$$E\epsilon_t^4 = \frac{3\delta_0^2(1 + \frac{2\delta_1}{1-\delta_1})}{1-3\delta_1^2} = \frac{3\delta_0^2 \cdot \frac{1+\delta_1}{1-\delta_1}}{1-3\delta_1^2} = 3(\text{Var } \epsilon_t)^2 \cdot \frac{1-\delta_1^2}{1-3\delta_1^2}$$

por lo que el momento de orden 4 de ϵ_t satisface una relación similar al de una variable Normal independiente e idénticamente distribuida, es decir, sin heteroscedasticidad, solo que con un factor adicional que excede a la unidad. Ello quiere decir que las colas de su distribución tienen más densidad que las de la distribución Normal. En particular, los momentos de cuarto orden podrían no existir, lo que ocurrirá si $\delta_1 > \sqrt{1/3}$. De este modo, al aumentar δ_1 , se van teniendo resultados de inexistencia de momentos de orden par.

Por otra parte, la cota $|\delta_1| < 1$ debe satisfacerse para obtener la estacionariedad del modelo, condición que, en general, se expresa:

Teorema .- Un proceso ARCH lineal de orden p con

$$h_t^2 = \delta_0 + \delta_1\epsilon_{t-1}^2 + \dots + \delta_p\epsilon_{t-p}^2, \quad \delta_0 > 0, \quad \delta_1 > 0, \quad \dots, \quad \delta_p > 0,$$

es estacionario en covarianza si y sólo si las raíces de la ecuación característica de h_t^2 caen fuera del círculo unidad.

Dicha estacionariedad se toma como una condición necesaria para aceptar como especificación correcta el modelo ARCH que se ha estimado.

3.3 Función de verosimilitud del modelo ARCH (p) univariante

Cada observación tiene distribución condicionada Normal, aunque la distribución conjunta del vector y de dimensión (Tx1) no es normal. El logaritmo de la función de verosimilitud condicionada puede escribirse:

$$L = -\frac{1}{T} \ln \text{lik} \sim -\frac{1}{T} \cdot \sum_t l_t = -\frac{1}{T} \cdot \sum_t \left[\frac{1}{2} \ln (h_t^2) + \frac{\epsilon_t^2}{2h_t^2} \right]$$

salvo constantes, donde l_t denota el logaritmo neperiano de la contribución del período t a la función de verosimilitud.

$$\frac{\partial l_t}{\partial \delta} = \frac{1}{2h_t^2} \cdot \frac{px1}{\partial h_t^2} \cdot \left[\frac{\epsilon_t^2}{h_t^2} - 1 \right] \quad \text{donde } \delta \text{ es un}$$

vector $px1$, y el resto son escalares, y:

$$\frac{\partial^2 l_t}{\partial \delta \partial \delta'} = -\frac{1}{2h_t^4} \cdot \left[\frac{\partial h_t^2}{\partial \delta} \right] \underbrace{\left[\frac{\partial h_t^2}{\partial \delta'} \right]}_{E(. / \Omega_{t-1})=1} \frac{\epsilon_t^2}{h_t^2} + \underbrace{\left[\frac{\epsilon_t^2}{h_t^2} - 1 \right]}_{E(. / \Omega_{t-1})=0} \frac{\partial}{\partial \delta'} \left[\frac{1}{2h_t^2} \cdot \frac{\partial h_t^2}{\partial \delta} \right]$$

=>por lo que la matriz de información resulta:

$$I_{\delta\delta} = \sum_{t=1}^T \frac{1}{2h_t^4} E \left[\frac{1}{h_t^4} \cdot \begin{bmatrix} \frac{\partial h_t^2}{\partial \delta} \\ \text{px1} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial h_t^2}{\partial \delta} \\ \text{1xp} \end{bmatrix}' \right]$$

que puede estimarse por:

$$\hat{I}_{\delta\delta} = \frac{1}{T} \sum_{t=1}^T \frac{1}{2h_t^4} \cdot \begin{bmatrix} \frac{\partial h_t^2}{\partial \delta} \\ \text{px1} \end{bmatrix} \begin{bmatrix} \frac{\partial h_t^2}{\partial \delta} \\ \text{1xp} \end{bmatrix}'$$

En el caso habitual en que h se supone una función lineal:

$h_t^2 = \delta_0 + \delta_1 \epsilon_{t-1}^2 + \dots + \delta_p \epsilon_{t-p}^2 \Rightarrow h_t^2 = z_t' \delta$ donde
 $z_t' = (1, \epsilon_{t-1}^2, \dots, \epsilon_{t-p}^2)$ y
 $\delta' = (\delta_0, \delta_1, \dots, \delta_p)$, entonces las expresiones

anteriores pueden escribirse:

$$(2) \quad \frac{\partial h_t}{\partial \delta} = \frac{1}{2h_t^2} \cdot z_t \left[\frac{\epsilon_t^2}{h_t^2} - 1 \right] \quad \text{e} \quad \hat{I}_{\delta\delta} = \frac{1}{2T} \sum_{t=1}^T \frac{z_t z_t'}{h_t^4}$$

que resultan útiles al programar la estimación numérica del vector gradiente y la matriz hessiana. Estos, a su vez, se utilizan en un algoritmo iterativo del tipo de "scoring" para obtener el estimador MV de los parámetros del modelo.

De este modo, el algoritmo numérico de estimación quedaría:

$$\hat{\delta}_{i+1} = \hat{\delta}_i - [I_{\delta\delta}]^{-1} \frac{\partial L}{\partial \delta} = \hat{\delta}_i - \left[\sum_1^T \frac{z_t z_t'}{h_t^4} \right]^{-1} \left[\sum_1^T \frac{z_t}{h_t^2} \left[\frac{\epsilon_t^2}{h_t^2} - 1 \right] \right]$$

por lo que la variación a introducir en la estimación del vector δ en cada etapa del proceso iterativo, viene dada por los coeficientes MCO en

una regresión de la variable $\left[\frac{\epsilon_t^2}{h_t^2} - 1 \right]$ sobre el vector $(p+1)$ -dimensional z_t/h_t^2 .

La consideración de una estructura ARCH(P) supone estimar $p+1$ parámetros de dicha estructura, junto con los restantes parámetros del modelo. Para limitar la dimensionalidad del espacio paramétrico, aunque manteniendo p retardos en la estructura ARCH, en ocasiones se utiliza la parametrización de h_t^2 siguiente:

$$h_t^2 = \delta_0 + \delta_1 \sum_1^p w_i \epsilon_{t-i}^2, \text{ en cuyo caso se tiene:}$$

$$\frac{\partial h_t^2}{\partial \delta} = (1, S_t), \text{ donde } S_t = \sum_1^p w_i \epsilon_{t-i}^2, \text{ y el algoritmo}$$

funciona como el de un ARCH(1), con S_t jugando el

papel de ϵ_{t-1}^2 . Sin embargo, hay que notar que la condición de estacionariedad debe comprobarse a partir de las raíces del polinomio:

$$1 - \delta_1 w_1 B - \dots - \delta_1 w_p B^p, \text{ no en la raíz de: } 1 - \delta_1 B.$$

4. EL MODELO ARCH-DE REGRESION

El modelo ARCH de regresión es la extensión natural del modelo ARCH univariante, en el sentido de que es ahora el término de error de un modelo de regresión quien acepta una estructura ARCH. Por otra parte, se contempla la posibilidad de que las variables explicativas del modelo de regresión entren también como variables explicativas del modelo de heteroscedasticidad.

Así, el modelo es:

$$\begin{aligned} y_t / \Omega_{t-1} &\sim N(x_t' \beta, h_t^2) \\ h_t^2 &= h(\epsilon_{t-1}, \dots, \epsilon_{t-p}, \delta) \\ \epsilon_t &= y_t - x_t' \beta \end{aligned}$$

o, más generalmente:

$$h_t^2 = h(\epsilon_{t-1}, \dots, \epsilon_{t-p}, x_t, x_{t-1}, \dots, x_{t-p}, \delta)$$

$$l_t = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln h_t^2 - \frac{1}{2} \frac{\epsilon_t^2}{h_t^2} = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln h_t^2 - \frac{1}{2} \frac{(y_t - x_t' \beta)^2}{h_t^2}$$

Si denotamos por θ el vector (β, σ^2) , se tiene:

$$\frac{\partial L}{\partial \theta} = \frac{1}{2} \sum [(\epsilon_t^2 - h_t^2) h_t^{-4}] \frac{\partial h_t^2}{\partial \theta} - \sum_t \frac{\epsilon_t}{h_t^2} \frac{\partial \epsilon_t}{\partial \theta}$$

donde: $\partial \epsilon_t / \partial \theta = (-x_t, 0)$.

Si no aparecen retardos de la variable endógena como variables explicativas del modelo de regresión, entonces el estimador MCO aplicado a este modelo, satisface el teorema de Gauss-Markov. Sin embargo el estimador MV es más eficiente.

4.1 Estimación de máxima verosimilitud del modelo ARCH de regresión

Para poder utilizar un algoritmo del tipo Newton-Raphson en el cálculo del estimador MV precisamos de la expresión analítica del vector gradiente y la matriz hessiana. Obtendremos por separado las derivadas con respecto a los subvectores β y δ .

Las derivadas de cada término de la función de verosimilitud con respecto al vector β de coeficientes de las variables explicativas vienen dadas por:

$$\frac{\partial l_t}{\partial \beta} = -\frac{1}{2} \frac{1}{h_t^2} \cdot \frac{\partial h_t^2}{\partial \beta} + \frac{1}{2} \frac{\epsilon_t^2}{h_t^4} \cdot \frac{\partial h_t^2}{\partial \beta} + \frac{\epsilon_t x_t}{h_t^2} =$$

$$\frac{\epsilon_t x_t}{h_t^2} + \frac{1}{2h_t^2} \cdot \frac{\partial h_t^2}{\partial \beta} \left(\frac{\epsilon_t^2}{h_t^2} - 1 \right)$$

y, teniendo en cuenta que:

$$\frac{\partial h_t^2}{\partial \beta} = \delta_1 \frac{\partial \epsilon_{t-1}^2}{\partial \beta} + \dots + \delta_p \frac{\partial \epsilon_{t-p}^2}{\partial \beta} = -2 \sum_{i=1}^p \delta_i x_{t-i} \epsilon_{t-i}$$

$$\text{implica: } \frac{\partial l_t}{\partial \beta} = \frac{\epsilon_t x_t}{h_t^2} - \frac{1}{h_t^2} \left(\frac{\epsilon_t^2}{h_t^2} - 1 \right) \sum_{i=1}^p \delta_i x_{t-i} \epsilon_{t-i},$$

$$\text{por lo que } \frac{\partial L}{\partial \beta} = \frac{1}{T} \sum_{t=1}^T \frac{\partial l_t}{\partial \beta} \text{ puede escribirse:}$$

$$\frac{\partial L}{\partial \beta} = \frac{1}{T} \sum_{t=1}^T x_t \epsilon_t \left[(h_t^2)^{-1} - \sum_{j=1}^p \delta_j (h_{t+j}^2)^{-2} (\epsilon_{t+j}^2 - h_{t+j}^2) \right] =$$

$$= \frac{1}{T} \sum_{t=1}^T x_t \epsilon_t S_t$$

donde S_t denota el corchete en la expresión anterior.

Derivando de nuevo, se tiene:

$$\frac{\partial^2 L}{\partial \beta \partial \beta'} = \sum_{t=1}^T \left[- \frac{x_t x_t'}{h_t^2} - \frac{2 \epsilon_t x_t'}{h_t^4} \cdot \frac{\partial h_t^2}{\partial \beta} - \frac{1}{2 h_t^4} \cdot \frac{\epsilon_t^2}{h_t^2} \cdot \left(\frac{\partial h_t^2}{\partial \beta} \right) \left(\frac{\partial h_t^2}{\partial \beta} \right)' + \left(\frac{\epsilon_t^2}{h_t^2} - 1 \right) \frac{\partial}{\partial \beta} \left(\frac{1}{2 h_t^2} \cdot \frac{\partial h_t^2}{\partial \beta} \right) \right]$$

y, tomando esperanzas condicionales en Ω_{t-1} y utilizando las propiedades: $E_{t-1} \epsilon_t = 0$; $E_{t-1} \epsilon_t^2 = h_t^2$, es decir, $E_{t-1}(\epsilon_t^2/h_t^2) = 1$, se tiene la matriz de información:

$$I_{\beta\beta} = \frac{1}{T} E \sum_{t=1}^T \frac{\partial^2 L}{\partial \beta \partial \beta'} =$$

$$= \frac{1}{T} E \sum_{t=1}^T \left[\frac{x_t x_t'}{h_t^2} + \frac{1}{2 h_t^4} \cdot \left[\frac{\partial h_t^2}{\partial \beta} \right] \left[\frac{\partial h_t^2}{\partial \beta} \right]' \right] =$$

$$\begin{aligned}
&= \frac{1}{T} \sum_1^T \left[\frac{x_t x_t'}{h_t^2} + \frac{1}{2h_t^4} \left[2 \sum_1^p \delta_i x_{t-i} \epsilon_{t-i} \right]^2 \right] = \\
&= \frac{1}{T} \sum_1^T \left[\frac{x_t x_t'}{h_t^2} + 2 \sum_1^p \delta_i^2 \frac{\epsilon_{t-i}^2}{h_t^4} \cdot x_{t-i} x_{t-i}' \right]
\end{aligned}$$

donde se han ignorado los términos de productos cruzados $\epsilon_{t-i} \cdot \epsilon_{t-j}$ cuando $i \neq j$. Esta expresión también puede escribirse:

$$\begin{aligned}
I_{\beta\beta} &= \frac{1}{T} \cdot \sum_i^T x_t x_t' \left[(h_t^2)^{-1} + 2 \epsilon_t^2 \sum_{j=1}^p \delta_j^2 (h_{t+j}^2)^{-2} \right] = \\
&= \frac{1}{2} \sum_1^T x_t x_t' r_t^2
\end{aligned}$$

donde r_t^2 denota la expresión dentro del corchete, mientras que las derivadas con respecto al subvector δ :

$$\frac{\partial l_t}{\partial \delta} \text{ y } \frac{\partial^2 l_t}{\partial \delta \partial \delta'} \text{ se obtienen del mismo modo que en el}$$

modelo ARCH simple, es decir:

$$\frac{\partial L}{\partial \delta} = \frac{1}{T} \sum_1^T \frac{1}{2h_t^2} \left[\frac{\epsilon_t^2}{h_t^2} - 1 \right] \frac{\partial h_t^2}{\partial \delta},$$

con $\frac{\partial h_t^2}{\partial \delta} = (1, \epsilon_{t-1}^2, \dots, \epsilon_{t-p}^2) = e_t$, por lo que:

$$\frac{\partial L}{\partial \delta} = \sum_{t=1}^T \frac{1}{2h_t^2} \begin{bmatrix} \epsilon_t^2 \\ \frac{\epsilon_t^2}{h_t^2} - 1 \end{bmatrix} e_t,$$

y

$$-E \begin{bmatrix} \frac{\partial^2 L}{\partial \delta \partial \delta'} \end{bmatrix} = E \frac{1}{2h_t^4} \begin{bmatrix} \frac{\partial h_t^2}{\partial \delta} \\ \frac{\partial h_t^2}{\partial \delta} \end{bmatrix} \begin{bmatrix} \frac{\partial h_t^2}{\partial \delta} \end{bmatrix}',$$

que puede estimarse por:

$$\hat{I}_{\delta\delta} = \frac{1}{2T} \sum_{t=1}^T \frac{e_t e_t'}{h_t^4}$$

donde e_t denota el vector de dimensión $p+1$ antes definido.

Un proceso ARCH se dice simétrico si la función h_t^2 y su vector gradiente (respecto al vector δ) son funciones reales pares en todas sus componentes $\epsilon_{t-1}^2, \dots, \epsilon_{t-p}^2$, mientras que las derivadas parciales de h_t^2 respecto de $\epsilon_{t-p}^2, \dots, \epsilon_{t-p}^2$ son funciones impares en $\delta_1, \dots, \delta_p$.

Lema.- Si un modelo ARCH de regresión con $\delta_0 > 0$, $\delta_1, \dots, \delta_p \geq 0$ es simétrico, se tiene:

$$I_{\delta\beta} = \frac{1}{T} \cdot \sum_{t=1}^T E \left[\frac{1}{2h_t^4} \cdot \begin{bmatrix} \frac{\partial h_t^2}{\partial \delta} \\ \frac{\partial h_t^2}{\partial \delta} \end{bmatrix} \begin{bmatrix} \frac{\partial h_t^2}{\partial \beta} \end{bmatrix}' \right] = 0$$

lo que implica que la estimación de δ y β por separado es asintóticamente eficiente.

Demostración:

$$\frac{\partial^2 l_t}{\partial \delta \partial \beta} = - \frac{1}{2h_t^4} \left[\frac{\epsilon_t^2}{h_t^2} - 1 \right] \frac{\partial h_t^2}{\partial \beta} \cdot \frac{\partial h_t^2}{\partial \delta} - \frac{\epsilon_t x_t}{h_t^4} \cdot \frac{\partial h_t^2}{\partial \delta}$$

$$- \frac{1}{2h_t^2} \cdot \frac{\epsilon_t^2}{h_t^4} \cdot \left[\frac{\partial h_t^2}{\partial \beta} \right] \cdot \left[\frac{\partial h_t^2}{\partial \delta} \right] + \frac{1}{2h_t^2} \cdot \left[\frac{\epsilon_t^2}{h_t^2} - 1 \right] \frac{\partial^2 h_t^2}{\partial \delta \partial \beta}$$

Como:

$$\frac{\partial h_t^2}{\partial \delta} = e_t = (1, \epsilon_{t-1}^2, \dots, \epsilon_{t-p}^2),$$

y:

$$\frac{\partial h_t^2}{\partial \beta} = -2 \sum_{i=1}^p \delta_i x_{t-i} \epsilon_{t-i},$$

ambas derivadas forman parte de Ω_{t-1} . lo mismo que ocurre con h_{t-1}^2 . De este modo, tomando la esperanza matemática del primer término y utilizando la descomposición: $E = E E_{t-1}$, se tiene que dicha esperanza es igual a:

$$-E \left[\frac{1}{2h_t^4} \cdot \frac{\partial h_t^2}{\partial \beta} \frac{\partial h_t^2}{\partial \delta} \cdot E_{t-1} \left[\frac{\epsilon_t^2}{h_t^2} - 1 \right] \right]$$

que es igual a cero, ya que $E_{t-1} \epsilon_t^2 = h_t^2$. Un razonamiento análogo puede hacerse para el segundo término. En cuanto al último término, basta

observar que $\frac{\partial^2 h_t}{\partial \delta \partial \beta}$ también pertenece a Ω_{t-1} , por lo que de nuevo, puede realizarse un argumento como los anteriores.

4.2 Procedimiento de estimación sugerido:

En base a los desarrollos anteriores, puede proponerse la siguiente estrategia de estimación:

- 1) Estimar β por MCO y obtener residuos $\hat{\epsilon}_t$.
- 2) Utilizar $\hat{\epsilon}_t$ para estimar δ por un algoritmo iterativo.
- 3) Utilizar $\hat{\delta}$ para estimar β e iterar.

En cada etapa del proceso se sugiere utilizar el algoritmo de "Scoring":

$$\theta_{i+1} = \theta_i + \left[\begin{array}{c|c} -1 & \\ \hline \hat{I}_{\theta\theta} & \\ \hline & \theta_i \end{array} \right] \cdot \left[\begin{array}{ccc|c} 1 & T & \frac{\partial l_t}{\partial \theta} & \\ \hline - & \Sigma & - & \\ T & 1 & \frac{\partial \theta}{\partial \theta} & \\ \hline & & & \theta_i \end{array} \right]$$

que sólo requiere derivadas primeras de la función de verosimilitud. La descomposición de este algoritmo en las etapas 2 y 3 anteriores, que utilizan el algoritmo para los subvectores β y δ por separado es eficiente gracias al lema anterior.

Puede verse fácilmente que el algoritmo iterativo del vector δ puede escribirse:

$$\hat{\delta}_{i+1} = \delta_i + (Z' Z)^{-1} Z' Y$$

donde:

$$Z_t = (1, \epsilon_{t-1}^2, \dots, \epsilon_{t-p}^2) / h_t^2$$

$$Z' = (Z_1, Z_2, \dots, Z_T) \text{ es } p \times T$$

$$y_t = (\epsilon_t^2 / h_t^2) - 1$$

$$Y = (y_1, y_2, \dots, y_T) \text{ es } T \times 1$$

donde: ϵ_t , h_t^2 y $\hat{\delta}_i$ provienen de la estimación previa. De este modo, cada iteración puede obtenerse mediante una regresión de y_t sobre el vector Z_t . La matriz de información estimada, dividida por el tamaño muestral, se toma como matriz de covarianzas del estimador MV del vector δ .

Dicha matriz de covarianzas no es sino: $2(Z'Z)^{-1}$, donde Z es la matriz de observaciones de las variables explicativas. En muestras finitas, esta matriz no es la proporcionada por la rutina de estimación, pero asintóticamente es válida, ya que en dicha regresión se tiene:

$$\sigma^2 = \text{Var} \left[\begin{array}{c} 2 \\ \epsilon_t^2 \\ h_t^2 \end{array} - 1 \right] \text{ y como } \frac{\epsilon_t}{h_t} \sim N(0,1),$$

$$\text{entonces: } \frac{\epsilon_t^2}{h_t^2} \sim X_1^2 \text{ y } \text{Var} \frac{\epsilon_t^2}{h_t^2} = 2$$

Las iteraciones sobre el vector β pueden hacerse:

$$\hat{\beta}_{i+1} = \hat{\beta}_i + (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\tilde{e}$$

donde se ha efectuado la transformación de

variables: $\tilde{x}_t = x_t r_t$ y $\tilde{e}_t = \epsilon_t \hat{S}_t / r_t$, por lo que el término de la derecha puede de nuevo interpretarse como el vector de coeficientes estimado en una regresión de $\tilde{e}_t$ sobre el vector $\tilde{x}_t$, y la matriz de información estimada para el subvector $\hat{\beta}$ dividida por el tamaño muestral, que no es sino $2(\tilde{X}'\tilde{X})^{-1}$, se toma como matriz de covarianzas del estimador MV del vector β .

Se utiliza en cada uno de los dos casos como criterio de convergencia el estadístico G:

$$G = \frac{\partial L}{\partial \theta} \left[\frac{\partial^2 L}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial L}{\partial \theta} \quad \text{que se distribuye } \chi_p^2, \text{ y}$$

puede probarse que coincide con el R^2 de la última regresión efectuada. Si dicho estadístico toma un valor superior al de las tablas, se interpreta que el cambio en la estimación es grande, y se continúa iterando. En caso contrario, se detiene el proceso de estimación, conservando el último valor obtenido.

Con la parametrización: $h_t^2 = \delta_0 + \delta_1 S_t$ donde

$$S_t = \sum_1^p w_i \epsilon_{t-i}^2, \quad \text{se tiene:}$$

$$\frac{\partial h_t^2}{\partial \delta} = (1, S_t) \quad \text{y} \quad \frac{\partial h_t^2}{\partial \beta} = -2\delta_1 \sum_1^p w_i x_{t-i} \epsilon_{t-i}$$

por lo que se tiene:

$$\frac{\partial L}{\partial \beta} = \frac{1}{T} \sum_1^T \left[\begin{array}{ccc} \frac{\epsilon_t x_t}{h_t^2} & - \frac{\delta_1}{h_t^2} & \left(\frac{\epsilon_t^2}{h_t^2} - 1 \right) \cdot \sum_1^p w_i x_{t-i} \epsilon_{t-i} \end{array} \right]$$

que puede también expresarse:

$$\frac{\partial L}{\partial \beta} = \frac{1}{T} \cdot \sum_1^T \left[\begin{array}{cc} \frac{1}{h_t^2} & - \delta_1 \sum_1^p \frac{w_i}{h_{t+i}^2} \left[\frac{\epsilon_{t+i}^2}{h_{t+i}^2} - 1 \right] \end{array} \right] x_t \epsilon_t$$

que es claramente un vector $k \times 1$, y:

$$I_{\beta\beta} = \frac{1}{T} \sum_1^T \left[\begin{array}{cc} \frac{1}{h_t^2} & + 2 \delta_1^2 \epsilon_t^2 \left[\sum_1^p \frac{w_i}{h_{t+i}^4} \right] \end{array} \right] x_t x_t',$$

mientras que:

$$\partial l_t / \partial \delta = -\frac{1}{2h_t^2} \cdot \left[\begin{array}{c} \frac{\epsilon_t^2}{h_t^2} - 1 \end{array} \right] \left[\begin{array}{c} 1 \\ S_t \end{array} \right]$$

$$I_{\delta\delta} = -\frac{1}{T} \Sigma_1^T \frac{1}{2h_t^4} \left[\begin{array}{cc} 1 & S_t \\ S_t & S_t^2 \end{array} \right]$$

con independencia de la composición del vector de variables explicativas x_t en el modelo de regresión.

El algoritmo de estimación puede, por tanto, escribirse:

$$\left[\begin{array}{c} \hat{\beta} \\ \hat{\delta} \end{array} \right]_{i+1} = \left[\begin{array}{c} \hat{\beta} \\ \hat{\delta} \end{array} \right]_i + T \Sigma_1^T \left[\begin{array}{c} \frac{1}{h_t^2} + \frac{2}{h_t^4} \frac{\epsilon_t^2}{\delta_1^2} \Sigma_1^P \frac{W_i^2}{h_{t+i}^4} x_t x_t' - 1 \\ 0 \\ 0 \\ \frac{1}{2h_t^4} \left[\begin{array}{cc} 1 & S_t \\ S_t & S_t^2 \end{array} \right] \end{array} \right]$$

$$\cdot \left[\begin{array}{c} \partial L / \partial \beta \\ \partial L / \partial \delta \end{array} \right]$$

donde $\frac{\partial L}{\partial \beta} = -\frac{1}{T} \sum_1^T \frac{\partial l_t}{\partial \beta}$, por lo que, en particular:

$$\hat{\delta}_{i+1} = \hat{\delta}_i + \frac{1}{T} \sum_1^T \left[\begin{array}{cc} 1 & S_t \\ S_t & S_t^2 \end{array} \right]^{-1} \left[\begin{array}{c} 1 \\ 2h_t^2 \end{array} \right] \cdot \sum_1^T \frac{1}{2h_t^2} \left[\begin{array}{cc} \frac{\epsilon_t^2}{2} & -1 \\ h_t & \end{array} \right] \left[\begin{array}{c} 1 \\ S_t \end{array} \right]$$

que puede verse como el resultado de una regresión de:

$$\left[\begin{array}{c} 2 \\ \frac{\epsilon_t^2}{2} - 1 \\ h_t \end{array} \right] \text{ sobre el par de variables: } \frac{1}{2} \cdot \left[\begin{array}{c} 1 \\ S_t \end{array} \right]$$

Cuando, como hemos supuesto al comienzo de esta sección, se utiliza la parametrización más amplia:

$$h_t^2 = \delta_0 + \delta_1 \epsilon_{t-1}^2 + \dots + \delta_p \epsilon_{t-p}^2,$$

entonces la matriz que aparece arriba no es

$$\left[\begin{array}{cc} 1 & S_t \\ S_t & S_t^2 \end{array} \right], \quad \text{sino } \frac{1}{T} \sum_1^T \left[\begin{array}{cccc} 1 & \epsilon_{t-1} & \dots & \epsilon_{t-p} \\ 2 & \epsilon_{t-1}^2 & \dots & \epsilon_{t-1} \cdot \epsilon_{t-p} \\ \dots & \dots & \dots & \dots \\ 2 & \epsilon_{t-p}^2 & \dots & \dots \end{array} \right]$$

mientras que en vez del vector $\begin{bmatrix} 1 \\ S_t \end{bmatrix}$ se tiene el vector:

$$\begin{bmatrix} 1 \\ \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-p} \end{bmatrix}$$

de modo que la parametrización reducida puede verse como una reducción de una estructura ARCH(p) a una estructura ARCH(1), donde S_t juega el papel de ϵ_{t-1} .

La idea del procedimiento de estimación consiste en comenzar del vector $\hat{\beta}$ estimado por MCO, generar los residuos y utilizarlos del modo que acabamos de ver para iterar sobre el estimador de δ . La matriz de covarianzas de los coeficientes estimados se obtiene del mismo modo que en la especificación general del modelo que antes vimos.

4.3 Contrastes de estructura ARCH

Con carácter previo a la estimación, deben llevarse a cabo los contrastes de estructura de heteroscedasticidad del tipo ARCH, donde la hipótesis nula es:

$$H_0: h_t^2 = h_0 \text{ constante,}$$

es decir, ausencia de heterocedasticidad, frente

a:

$$H_1: h_t^2 = h(z_t' \delta), \text{ donde } z_t = (1, \epsilon_{t-1}^2, \dots, \epsilon_{t-p}^2)$$

El estadístico utilizado para llevar a cabo el

contraste es: $\frac{1}{2} y' Z (Z' Z)^{-1} Z' y \sim \chi_p^2$

donde $y = [(\frac{e_t^2}{h_0^2} - 1)]$ vector $T \times 1$, aunque

omiteremos las p primeras observaciones; y $Z' = (z_{p+1}', \dots, z_T')$ por lo que Z es de dimensión $(T-p) \times (p+1)$.

Como puede verse, el contraste es condicional en un orden p del modelo ARCH especificado en la hipótesis alternativa H_1 , por lo que debe realizarse para varios valores de dicho parámetro.

El contraste es asintóticamente equivalente a utilizar el producto $T.R^2$ donde R^2 procede de la regresión de e_t^2 sobre una constante y p retardos suyos, todos ellos obtenidos bajo la hipótesis nula; dicho producto se distribuye de acuerdo con una variable χ_p^2 (el número de grados de libertad no incluye la constante). Por esto es que el correlograma de los residuos al cuadrado puede utilizarse en la identificación del orden p del posible modelo ARCH.

Este contraste es válido con independencia de la forma funcional de h_t , siempre que h_t^2 dependa de δ tan sólo a través del producto $z_t' \delta$.

Para contrastar un modelo ARCH de un determinado orden frente a una estructura ARCH de orden inferior, por ejemplo, para contrastar la hipótesis nula:

$$H_0 : \delta_2 = 0$$

en el modelo: $h_t^2 = \delta_0 + \delta_1 \epsilon_{t-1}^2 + \delta_2 \epsilon_{t-2}^2$, se efectúa una iteración a partir de la estimación MV del modelo de orden inferior, permitiendo esta vez que δ_2 sea distinto de cero, y se utiliza el estadístico R^2 de la regresión correspondiente a tal iteración. Equivalentemente, se compara con una χ_1^2 el producto TR^2 , donde R^2 proviene de la regresión de un vector de unos sobre el vector gradiente de la función de verosimilitud.

El interés de este segundo contraste estriba en que el test de los multiplicadores de Lagrange es más sencillo que el test de Wald, si el modelo ya ha sido estimado bajo la hipótesis nula $H_0 : \phi \notin \Phi_0$. En efecto, si denotamos por S^0 la matriz de observaciones del vector de las derivadas parciales calculada para el modelo más general, pero evaluada bajo la hipótesis nula, dicho vector coincide con los multiplicadores de Lagrange, y sus varianzas vienen dadas por la matriz de información, de modo que el test de los multiplicadores de Lagrange utiliza el estadístico:

$$\Phi_{LM} = 1' S^0 (S^0' S^0)^{-1} S^0' 1 = TR_0^2$$

donde R_0^2 es el coeficiente de determinación no centrado de una regresión del vector de unos sobre el vector de derivadas parciales evaluado bajo la hipótesis nula y 1 denota un vector de unos, de dimensión igual al número de parámetros que se pretenden estimar.

Los errores standard que se obtienen de esta regresión son asintóticamente válidos, al igual que en los casos anteriores. En efecto, la matriz de covarianzas estimada no es sino $2(Z'Z)^{-1}$,

donde Z es la matriz de observaciones de las variables explicativas. Aunque en muestras finitas esta matriz no coincide con la proporcionada por la rutina de estimación, es asintóticamente válida, exactamente por las mismas razones antes vistas.

5. EL MODELO ARCH-M

El modelo ARCH-M (Engle, Lilien y Robins (1987)), incorpora el supuesto adicional de que la varianza del término de error influye en cada período sobre el nivel de la variable que se pretende explicar, por lo que representa un nivel superior de generalidad con respecto a los modelos que hasta ahora hemos considerado. Tal supuesto es muy apropiado en modelos financieros que traten de caracterizar la magnitud de las primas de riesgo y_t , definidas como diferencia entre el tipo forward F_t la expectativa actual del tipo futuro, ${}_t r_{t+1}$. La idea es que siendo h_t una medida de la incertidumbre existente en el período t acerca de y_t , un mayor valor de h_t supone, en el caso de los mercados financieros, un mayor riesgo, que podría aparecer reflejado en una mayor prima al riesgo. Ello generaría, una mayor diferencia entre ${}_t r_{t+1}$ y el tipo forward F_t que aparece a partir de la condición de ausencia de arbitraje,² por lo que la magnitud de la diferencia $y_t = F_t - {}_t r_{t+1}$ dependerá, en general, de h_t .

Consideramos, por tanto, el modelo:

$$y_t = \alpha + \beta \cdot g(h_t^2) + u_t, \quad u_t \sim N(0, h_t^2)$$

$$h_t^2 = \delta_0 + \delta_1 \sum_{i=1}^p w_i u_{t-i}^2$$

donde mantenemos la parametrización reducida de h_t^2 antes considerada.

² Según la cual F_t , en términos continuos, es la diferencia entre los tipos a largo y corto plazo hoy vigentes en el mercado. Véase por ejemplo el artículo: "¿Incorporan los tipos del interbancario una evaluación del riesgo?. Documento de Trabajo FEDEA 90-08 J. Ayuso, A. Novales y M.L. de la Torre.

5.1 Estimación del modelo ARCH-M

El logaritmo de la función de verosimilitud es:

$$\begin{aligned}
 L = \ln \text{lik} &= -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_1^T \ln h_t^2 - \frac{1}{2} \sum_1^T \frac{u_t^2}{h_t^2} = \\
 &= -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_1^T \ln h_t^2 - \frac{1}{2} \sum_1^T \frac{(y_t - \alpha - \beta g(h_t^2))^2}{h_t^2}
 \end{aligned}$$

donde:

$$h_t^2 = \delta_0 + \delta_1 \sum_1^p w_i (y_{t-i} - \alpha - \beta g(h_{t-i}^2))^2$$

Es fácil ver que:

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2} \sum_1^T \frac{1}{h_t^2} \cdot \frac{\partial h_t^2}{\partial \theta} + \frac{1}{2} \sum_1^T \frac{[y_t - \alpha - \beta g(h_t^2)]^2}{(h_t^2)^2} \cdot \frac{\partial h_t^2}{\partial \theta}$$

$$\begin{aligned}
& - \Sigma_1^T \frac{u_t}{h_t^2} \begin{bmatrix} -1 - \beta \frac{\partial g(h_t^2)}{\partial \alpha} \\ -g(h_t^2) - \beta \frac{\partial g(h_t^2)}{\partial \beta} \\ -\beta \frac{\partial g(h_t^2)}{\partial \delta_0} \\ -\beta \frac{\partial g(h_t^2)}{\partial \delta_1} \end{bmatrix} \\
& = - \frac{1}{2} \Sigma_1^T \frac{1}{h_t^2} \cdot \begin{bmatrix} 1 - \frac{u_t^2}{h_t^2} \end{bmatrix} \cdot \frac{\partial h_t^2}{\partial \theta} + \beta \Sigma_1^T \frac{u_t}{h_t^2} \frac{\partial g(h_t^2)}{\partial \theta} + \Sigma_1^T \frac{u_t}{h_t^2} \cdot \begin{bmatrix} 1 \\ g(h_t^2) \\ 0 \\ 0 \end{bmatrix}
\end{aligned}$$

donde θ denota el vector : $\theta = (\alpha, \beta, \delta_0, \delta_1)$

Teniendo en cuenta que $\frac{\partial g(h_t^2)}{\partial \theta} = \frac{\partial g(h_t^2)}{\partial h_t^2} \cdot \frac{\partial h_t^2}{\partial \theta}$,

se tiene:

$$\frac{\partial L}{\partial \theta} = \Sigma_1^T \frac{1}{h_t^2} \left[\begin{bmatrix} 1 - \frac{u_t^2}{h_t^2} \end{bmatrix} + \beta u_t \frac{\partial g(h_t^2)}{\partial h_t^2} \right] \frac{\partial h_t^2}{\partial \theta} + \begin{bmatrix} 1 \\ g(h_t^2) \\ 0 \\ 0 \end{bmatrix} \cdot u_t$$

donde hay que incorporar la derivada:

$$\frac{\partial h_t^2}{\partial \theta} = \begin{bmatrix} -2\delta_1 \sum_1^P w_i u_{t-i} \\ -2\delta_1 \sum_1^P w_i g(h_{t-i}^2) u_{t-i} \\ 1 \\ S_t \end{bmatrix}$$

5.2 Casos particulares:

$$5.2.1. \quad g(h_t^2) = h_t^2 \Rightarrow \frac{\partial g(h_t^2)}{\partial h_t^2} = 1$$

$$\frac{\partial L}{\partial \theta} = \sum_1^T \frac{1}{h_t^2} \left[\begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix} \begin{bmatrix} u_t^2 \\ 1 - \frac{u_t^2}{h_t^2} \end{bmatrix} + \beta u_t \right] \begin{bmatrix} -2\delta_1 \sum_1^P w_i u_{t-i} \\ -2\delta_1 \sum_1^P w_i u_{t-i} h_{t-i}^2 \\ 1 \\ S_t \end{bmatrix} + \begin{bmatrix} 1 \\ h_t^2 \\ 0 \\ 0 \end{bmatrix} u_t$$

$$5.2.2. \quad g(h_t^2) = h_t \Rightarrow \frac{\partial g(h_t^2)}{\partial h_t^2} = \frac{1}{2\sqrt{h_t^2}}$$

$$\frac{\partial L}{\partial \theta} = \sum_1^T \frac{1}{h_t^2} \left[\begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix} \begin{bmatrix} u_t^2 \\ 1 - \frac{u_t^2}{h_t^2} \end{bmatrix} + \beta \frac{u_t}{2\sqrt{h_t^2}} \right] \begin{bmatrix} -2\delta_1 \sum_1^P w_i u_{t-i} \\ -2\delta_1 \sum_1^P w_i u_{t-i} h_{t-i}^2 \\ 1 \\ S_t \end{bmatrix} + \begin{bmatrix} 1 \\ h_t^2 \\ 0 \\ 0 \end{bmatrix} u_t$$

$$5.2.3. \quad g(h_t^2) = \ln h_t^2 \Rightarrow \frac{\partial g(h_t^2)}{\partial h_t^2} = \frac{1}{h_t^2}$$

$$\frac{\partial L}{\partial \theta} = \sum_1^T \frac{1}{h_t^2} \left[-\frac{1}{2} \left[1 - \frac{u_t^2}{h_t^2} \right] + \beta \frac{u_t^2}{h_t^2} \right] \left[\begin{array}{c} -2\delta_1 \sum_1^P w_i u_{t-i} \\ -2\delta_1 \sum_1^P w_i u_{t-i} \ln h_{t-i}^2 \\ 1 \\ s_t \end{array} \right] + \left[\begin{array}{c} 1 \\ 2 \\ 0 \\ 0 \end{array} \right] u_t$$

$$5.2.4. \quad g(h_t^2) = \ln \sqrt{h_t^2} \Rightarrow \frac{\partial g(h_t^2)}{\partial h_t^2} = \frac{1}{2h_t^2}$$

$$\frac{\partial L}{\partial \theta} = \sum_1^T \frac{1}{h_t^2} \left[-\frac{1}{2} \left[1 - \frac{u_t^2}{h_t^2} \right] + \beta \frac{u_t^2}{2h_t^2} \right] \left[\begin{array}{c} -2\delta_1 \sum_1^P w_i u_{t-i} \\ -2\delta_1 \sum_1^P w_i u_{t-i} \ln \sqrt{h_{t-i}^2} \\ 1 \\ s_t \end{array} \right] + \left[\begin{array}{c} 1 \\ 2 \\ 0 \\ 0 \end{array} \right] u_t$$

Para estimar este modelo, Engle et al. (1987) recomiendan utilizar la sugerencia de Berndt et al. (1974) para el algoritmo de Scoring:

$$\hat{\theta}_{i+1} = \hat{\theta}_i + \lambda \left[\sum_1^T \left[\frac{\partial L_t}{\partial \theta} \right] \left[\frac{\partial L_t}{\partial \theta} \right]' \right]^{-1} \cdot \sum_1^T \frac{\partial L_t}{\partial \theta}$$

El estimador resultante es consistente, con matriz de covarianzas:

$$\text{Var}(\hat{\theta}) = \frac{1}{T} \left[\Sigma_1^T \begin{bmatrix} \frac{\partial L}{\partial \theta} \end{bmatrix} \begin{bmatrix} \frac{\partial L}{\partial \theta} \end{bmatrix}' \right]^{-1}$$

que, a diferencia del modelo ARCH simple, no es diagonal a bloques entre los vectores de coeficientes β y δ de las ecuaciones de y_t y h_t^2 , por lo que la estimación de los vectores β y δ por separado no sería eficiente en este modelo ARCH-M.

Toda la discusión concerniente a la contrastación de hipótesis lineales, así como a los criterios de convergencia a utilizar, puede reproducirse a partir de los resultados citados al final de la sección 4.

6. EL MODELO GARCH (Generalized ARCH)
(Bollerslev (1986))

6.1 El modelo GARCH (1,1)

El modelo GARCH especifica que, en cada periodo, la varianza condicional de y_t , h_t^2 , depende de los últimos residuos, pero también de sus propios valores previos, es decir que:

$$h_t^2 = \delta_0 + \delta_1 \epsilon_{t-1}^2 + \theta_1 h_{t-1}^2, \text{ con } \delta_0, \delta_1, \theta_1 > 0$$

representa el modelo GARCH(1,1). En este modelo GARCH (1,1) se tiene:

$$\begin{aligned} E_t h_{t+s}^2 &= \delta_0 + \delta_1 E_t \epsilon_{t+s-1}^2 + \theta_1 E_t h_{t+s-1}^2 = \\ &= \delta_0 + (\delta_1 + \theta_1) \cdot E_t h_{t+s-1}^2 \end{aligned}$$

donde puede verse que, para que el proceso de varianza condicional sea estacionario, es preciso que: $\delta_1 + \theta_1 < 1$. En tal caso, la varianza incondicional de ϵ_t^2 viene dada por:

$$\sigma^2 = E \epsilon_t^2 = \delta_0 / [1 - (\delta_1 + \theta_1)]$$

y al aumentar el horizonte de previsión s , puede probarse que la varianza condicional $E_t h_{t+s}^2$ converge hacia σ^2 . Bajo tal supuesto, el modelo GARCH equivale a un modelo ARCH de orden infinito, con coeficientes en los retardos ϵ_{t-i} que decaen hacia cero de modo exponencial.

El momento de orden $2n$ del modelo GARCH

$(1,1)$ existe si:

$$\sum_{j=0}^n \begin{bmatrix} n \\ j \end{bmatrix} a_j \delta_1^j \theta_1^{n-j} < 1$$

donde: $a_0 = 1$; $a_j = \prod_{i=1}^j (2i-1)$, $j \geq 1$.

En el caso del proceso ARCH(1), por ejemplo, $\theta_1=0$ y esta condición se reduce a $a_n \delta_1^n < 1$. Como otro caso particular, en el modelo GARCH(1,1), el momento de cuarto orden existe si:

$$3\delta_1^2 + 2\delta_1 \theta_1 + \theta_1^2 < 1, \text{ en cuyo caso:}$$

$$E \epsilon_t^2 = \delta_0 (1 - \delta_1 - \theta_1)^{-1} \quad y$$

$$E(\epsilon_t^4) = 3\delta_0^2 (1 + \delta_1 + \theta_1) \cdot [(1 - \delta_1 - \theta_1)(1 - \theta_1^2 - 2\delta_1 \theta_1 - 3\delta_1^2)]^{-1}$$

por lo que es fácil ver que el coeficiente de Kurtosis es:

$$K = [E \epsilon_t^4 - 3(E \epsilon_t^2)^2] (E \epsilon_t^2)^{-2} = 6\delta_1^2 (1 - \theta_1^2 - 2\delta_1 \theta_1 - 3\delta_1^2)^{-1}$$

que es positivo bajo nuestras hipótesis, por lo que el proceso GARCH(1,1) es leptocúrtico, al igual que ocurre con el modelo ARCH(q).

En las aplicaciones prácticas de este modelo, se tiene con frecuencia que $\delta_1 + \theta_1$ es aproximadamente igual a la unidad, con lo que el

proceso h_{t+s}^2 es no estacionario. Tal proceso se dice que es integrado en varianza, ya que la

primera diferencia de ésta sería un proceso estacionario. En tal modelo, la previsión $E_t h_{t+s}^2$, $s > 0$, no converge hacia σ^2 , sino que depende explícitamente de Ω_t .

El proceso GARCH (1,1) integrado se representa:

$$h_t^2 = \alpha \epsilon_{t-1}^2 + (1-\alpha) h_{t-1}^2$$

donde $0 < \alpha < 1$. Este proceso de varianza satisface la condición:

$$E_t h_{t+s}^2 = h_{t+1}^2 = \alpha \epsilon_t^2 + (1-\alpha) h_t^2, \quad s \geq 1$$

por lo que guarda cierta semejanza con un proceso de camino aleatorio en la varianza. En él, las innovaciones en la varianza condicional tienen un efecto permanente. Si se incorpora una tendencia w :

$$h_t^2 = w + \alpha \epsilon_{t-1}^2 + (1-\alpha) h_{t-1}^2$$

Se tiene:

$$E_t h_{t+s}^2 = sw + h_{t+1}^2 = (s+1)w + \alpha \epsilon_t^2 + (1-\alpha) h_t^2$$

similar a un paseo aleatorio con deriva. Aunque la importancia relativa de h_{t+1}^2 en la expresión de la varianza condicional h_{t+s}^2 decrece con s , su efecto es permanente.

Si definimos el proceso: $\gamma_t = \epsilon_t^2 - h_t^2$, claramente no autocorrelacionado, se tiene:

$$\epsilon_t^2 = \delta_0 + \sum_{i=1}^m (\delta_i + \theta_i) \epsilon_{t-i}^2 - \sum_{i=1}^p \theta_i \gamma_{t-i} + \gamma_t$$

donde $m = \max\{p, q\}$. Por tanto, ϵ_t^2 sigue un proceso ARMA (m, p) con un término de error sin autocorrelación. Ello ha conducido a la sugerencia (ver Bollerslev (1986)) de utilizar las funciones de autocorrelación simple y parcial de

ϵ_t^2 para identificar los órdenes del proceso GARCH, como enseguida veremos.

6.2. EL MODELO GARCH (p, q)

El modelo GARCH (p, q) para un proceso ϵ_t , queda definido por:

$$\epsilon_t / \Omega_{t-1} \sim N(0, h_t^2)$$

$$\begin{aligned} h_t^2 &= \delta_0 + \sum_{i=1}^q \delta_i \epsilon_{t-i}^2 + \sum_{j=1}^p \theta_j h_{t-j}^2 = \\ &= \delta_0 + \delta(L) \epsilon_t^2 + \theta(L) h_t^2 \end{aligned}$$

donde: $p \geq 0$, $q > 0$, $\delta_0 > 0$, $\delta_i \geq 0$, $i = 1, \dots, q$; $\theta_i \geq 0$, $i = 1, \dots, p$, reduciéndose al modelo ARCH(q) cuando $p = 0$, y a un proceso de ruido blanco cuando $p = q = 0$.

Si todas las raíces del polinomio $1 - \theta(L)$ están fuera del círculo unidad, se tiene:

$$\begin{aligned} h_t^2 &= \delta_0 (1 - \theta(1))^{-1} + \delta(L) (1 - \theta(L))^{-1} \epsilon_t^2 = \\ &= \delta_0 (1 - \sum_{i=1}^p \theta_i)^{-1} + \sum_{i=1}^{\infty} \beta_i \epsilon_{t-i}^2 \end{aligned}$$

es decir, un modelo ARCH(∞), donde los coeficientes β_i provienen del desarrollo de la expresión: $\delta(L)[1-\theta(L)]^{-1}$

El retardo medio en la ecuación de la varianza condicional h_t^2 del modelo GARCH(1,1) es:

$$r = \frac{\sum_{i=1}^{\infty} i\beta_i}{\sum_{i=1}^{\infty} \beta_i} = (1-\theta_1)^{-1}$$

y el retardo mediano:

$$s = -\log 2 / \log \theta_1$$

El proceso GARCH(p,q) es estacionario de segundo orden, con³ $E\epsilon_t=0$, $\text{Var } \epsilon_t = \delta_0(1-\delta(1)-\theta(1))^{-1}$ y $\text{Cov}(\epsilon_t, \epsilon_s) = 0$ para todo $t \neq s$, si y sólo si $\delta(1) + \theta(1) < 1$.

6.3. Identificación de modelos GARCH(p,q)

Para la identificación de una estructura del tipo GARCH (p,q) resultan muy útiles las funciones de autocorrelación simple y parcial de los residuos al cuadrado: Si denotamos por ρ_n la

función de autocorrelación de ϵ_t^2 , se tiene

(Bollerslev (1986)):

$$\rho_n = \sum_{i=1}^m \psi_i \rho_{n-i}, \quad n \geq p+1$$

donde $\psi_i = \delta_i + \theta_i$, $i=1, \dots, q$, $\theta_i = 0$ para $i > p$ y $\delta_i = 0$ para $i > q$. Estas ecuaciones son análogas a las de Yule-Walker para el proceso ARMA(m,p). Una vez más, un procedimiento similar, con $\psi_i = \delta_i$,

³ Con $\delta(1)$, $\theta(1)$ denotamos el valor numérico que resulta de sustituir L por 1 en los polinomios $\delta(L)$ y $\theta(L)$.

podría utilizarse para el modelo ARCH(q).

Si ϕ_{kk} denota la función de autocorrelación parcial se tiene:

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

mientras que $\phi_{kk}=0$ para $k>q$, al igual que ocurre con un proceso AR(q). Para llevar a cabo estos contrastes, los valores de ρ_n y ϕ_{kk} a utilizar en las ecuaciones previas deben sustituirse por sus valores estimados, utilizando su varianza asintótica (bajo la hipótesis nula de no existencia de estructura GARCH) de $1/T$. Por supuesto, una vez que se ha estimado un modelo ARCH, el contraste de las funciones de autocorrelación de los residuos se debe hacer con los residuos normalizados.

6.4. Contraste de estructura GARCH

Si descomponemos la estructura de la varianza condicional: $h_t^2 = z_{1t}' w_1 + z_{2t}' w_2$ y se quiere contrastar la hipótesis nula: $H_0: w_2 = 0$, el estadístico de multiplicadores de Lagrange correspondiente es:

$$\xi_{ML} = \frac{1}{2} f_0' Z_0 (Z_0' Z_0)^{-1} Z_0' f_0 \text{ donde:}$$

$$f_0' = \left[(\epsilon_1^2/h_1^2) - 1, \dots, (\epsilon_T^2/h_T^2) - 1 \right] \text{ es } 1 \times T$$

y: $Z_0' = (h_1^2 \cdot \partial h_1^2 / \partial w, \dots, h_T^2 \cdot \partial h_T^2 / \partial w)$ de modo que Z_0 es $T \times s$, donde s es el número de variables en h_t^2 , evaluados ambos vectores bajo H_0 . El estadístico

ξ_{ML} se distribuye χ_r^2 donde r es el número de parámetros en w_2 . La mayor diferencia con el caso ARCH (q) es que $\partial h_t^2 / \partial w$ no tiene una expresión sencilla si la ecuación de h_t^2 incluye algún retardo de ϵ_t^2 . Como es habitual, este estadístico coincide asintóticamente con el producto $T.R^2$, donde R^2 es el coeficiente de la regresión MCO en la primera iteración del procedimiento de Berndt, Hall, Hall y Hausman (1974) a partir de la estimación obtenida bajo H_0 .

En este tipo de contraste se tiene un resultado similar al de los contrastes de modelos ARMA(p, q): Si la hipótesis nula es ARCH(q), la matriz $Z_0' Z_0$ es singular para cualquier alternativa GARCH ($r_1, q+r_2$) con $r_1 > 0, r_2 > 0$. Ello ocurre, en particular, para el contraste de ruido blanco frente a GARCH(p, q). Otra propiedad es que, frente a $H_0: \text{ARCH}(q)$, los estadísticos para las alternativas ARCH($q+r$) y GARCH(r, q) coinciden.

6.5 Estimación de máxima verosimilitud del modelo GARCH de regresión

Si denotamos:

$$z_t' = (1, \epsilon_{t-1}^2, \dots, \epsilon_{t-q}^2, h_{t-1}^2, \dots, h_{t-p}^2) \text{ y}$$

$$w' = (\delta_0, \delta_1, \dots, \delta_q, \theta_1, \dots, \theta_p) \text{ y}$$

$\psi' = (\beta', w')$, se tiene el logaritmo de la función de verosimilitud:

$$L_T(\psi) = T^{-1} \sum_{t=1}^T \left[-\frac{1}{2} \log h_t^2 - \frac{1}{2} \epsilon_t^2 / h_t^2 \right] =$$

$$= T^{-1} \sum_{t=1}^T l_t(\psi)$$

con:

$$\partial l_t / \partial w = \frac{1}{2} h_t^{-2} \frac{\partial h_t^2}{\partial w} \begin{bmatrix} \frac{\epsilon_t^2}{h_t^2} & -1 \end{bmatrix}$$

$$(1) \quad \partial^2 l_t / \partial w \partial w' = \begin{bmatrix} \frac{\epsilon_t^2}{h_t^2} & -1 \end{bmatrix} \frac{\partial}{\partial w'} \begin{bmatrix} \frac{1}{2} h_t^{-2} \frac{\partial h_t^2}{\partial w} \end{bmatrix}$$

$$= \frac{1}{2} h_t^{-4} \frac{\partial h_t^2}{\partial w} \begin{bmatrix} \frac{\partial h_t^2}{\partial w} \end{bmatrix} \frac{\epsilon_t^2}{h_t^2}$$

donde:

$$(2) \quad \partial h_t^2 / \partial w = z_t + \sum_{i=1}^p \theta_i \frac{\partial h_{t-i}^2}{\partial w}$$

siendo el segundo término de $\partial h_t^2 / \partial w$ la diferencia con respecto al modelo ARCH de regresión. Nótese que (2) será estable siempre que $\theta(1) < 1$. Por otra parte, la esperanza condicional del primer término en (1) es cero, por lo que la submatriz de la matriz de información correspondiente al vector w puede estimarse consistentemente a partir tan sólo del segundo término en (2), que sólo requiere primeras derivadas.

Por otra parte:

$$\partial l_t / \partial \beta = \epsilon_t x_t / h_t^2 + (1/2 h_t^2) \cdot (\partial h_t^2 / \partial \beta) \cdot \left[(\epsilon_t^2 / h_t^2) - 1 \right]$$

$$(3) \quad \partial^2 l_t / \partial \beta \partial \beta' =$$

$$= -x_t x_t' / h_t^2 - 1/(2h_t^4) (\partial h_t^2 / \partial \beta) \cdot \left[\partial h_t^2 / \partial \beta \right]^{-1} (\epsilon_t^2 / h_t^2) -$$

$$-2(\epsilon_t x_t' / h_t^4) (\partial h_t^2 / \partial \beta) + ((\epsilon_t^2 / h_t^2) - 1) \cdot \partial \left[1/(2h_t^2) \cdot \partial h_t^2 / \partial \beta \right] / \partial \beta'$$

donde:

$$(4) \quad \partial h_t^2 / \partial \beta = -2 \sum_{i=1}^q \delta_i x_{t-i} \epsilon_{t-i} + \sum_{j=1}^p \theta_j \cdot \partial h_{t-j} / \partial \beta$$

donde, de nuevo, el segundo sumando es la única diferencia con el modelo ARCH de regresión. Una estimación consistente de la sumatriz de información correspondiente a β puede obtenerse a partir de los dos primeros términos de (3),

sustituyendo ϵ_t^2 / h_t^2 por su valor esperado de 1. Esta estimación también utiliza únicamente derivadas primeras. Finalmente, al igual que en el modelo ARCH de regresión, el bloque fuera de la diagonal en la matriz de información resulta ser cero, por lo que w puede estimarse por separado de β sin perder eficiencia asintótica.

La presencia de términos recursivos en (2) y (4) hacen que no podamos representar el algoritmo de scoring como una regresión. Por ello, es aconsejable utilizar el algoritmo de Berndt, Hall,

Hall y Hausman (1974), incorporando un parámetro de longitud de salto. En dicho algoritmo, la dirección de salto puede obtenerse a partir de una regresión de un vector de unos sobre el vector $\partial l_t / \partial \psi$; además, por las razones ya vistas, el algoritmo puede aplicarse por separado a w y β . Si no se parte del supuesto de normalidad condicional de ϵ_t , entonces la distribución asintótica del estimador que resulta de este algoritmo es Normal, pero con matriz de covarianzas:

$$I(\psi)^{-1} \cdot B(\psi) \cdot I(\psi)^{-1},$$

donde:

$$I(\psi)^{-1} = -E [\partial^2 l_t / \partial \psi \partial \psi']^{-1} \quad y$$

$$B(\psi) = E [(\partial l_t / \partial \psi) (\partial l_t / \partial \psi)'],$$

estimándose ambas por sus respectivos momentos muestrales.

REFERENCIAS

Berndt E., B.Hall, R.Hall y J.Hausman (1974)
 "Estimation and inference in non linear structural
 models" Annals of Economic and Social
 Measurement, 4, 653-665.

Bollerslev, T. (1988) "On the correlation
 structure for the generalized autoregressive
 conditional heteroskedastic process", Journal of
 Time Series Analysis, 9, 121-131.

Bollerslev, T. (1986) "Generalized
 autoregressive conditional heteroskedasticity",
Journal of Econometrics, 31f, p. 307-327.

Engle, R. (1982) "Autoregressive Conditional
 Heteroscedasticity with estimates of the variance
 of United Kingdom inflation", Econometrica, 50, 4,
 987-1007.

Engle, R., D.Lilien, R.Robbins (1987)
 "Estimating time varying risk premia in the term
 structure: The ARCH-M model", Econometrica, 55,2,
 p.391-407.

Engle, R. y T. Bollerslev (1986) "Modelling
 the persistence of conditional variances",
Econometric Reviews, 5, p.1-50.

LIBROS

COLECCION ESTUDIOS

- 1: "Excedente, inversión y empleo en la empresa española". Carlos Sebastián y Luis Servén.
- 2: "Estructura fiscal e incentivos a la inversión" Carlos Cuervo-Arango y José A. Trujillo.
- 3: "Análisis de simulación en la industria española". Carlos Sebastián y Luis Servén.
- 4: "La empresa pública industrial en España". Alfonso Novales, Carlos Sebastián, Luis Servén y José A. Trujillo.
- 5: "Análisis y predicción de la población española: 1910-2000". Juan del Hoyo y Antonio García Ferrer.
- 6: "El paro en España: Características, causas y medidas". Alfonso Novales, Carlos Sebastián y Luis Servén.
- 7: "Estudios sobre participación, empleo y paro en España". Varios autores.

COLECCION DEBATES

- 1: "La política monetaria en España".
- 2: "La fiscalidad de la empresa".
- 3: "Política monetaria e inestabilidad financiera".

COLECCION CUADERNOS DE ECONOMIA Y FINANZAS

- 1: "Riesgo de interés y riesgo de crédito en el contrato de swap". Xavier Freixas.
- 2: "El Sistema Monetario Europeo y el ECU". Juan Pérez-Campanero.