

En diversas aplicaciones los modelos $ARMA(p, q)$, o inclusive los $ARMA(p, q) \times (P, Q)_s$, no son suficientes para explicar la estructura de dependencia subyacente a los datos. Esto último nos lleva a considerar modelos más complejos que comúnmente conllevan a modelos no lineales o no Gaussianos. En las secciones subsecuentes estudiaremos los *modelos de heteroscedasticidad condicional autorregresiva* denotados por (ARCH) y algunas de sus generalizaciones. Los modelos ARCH surgen con el trabajo de Engle (1982) al tratar de modelar los rendimientos correspondientes a series financieras.

24. Motivación

Supóngase que X_t denota el valor de una acción financiera al tiempo t , entonces el retorno, rendimiento o ganancia relativa, Y_t , de la acción al tiempo t es

$$Y_t = \frac{X_t - X_{t-1}}{X_{t-1}} \quad (86)$$

lo que implica que $X_t = X_{t-1}(1 + Y_t)$, claramente Y_t representa el porcentaje de cambio del periodo $t - 1$ al periodo t . Si tomamos el logaritmo natural y diferenciamos una vez se obtiene

$$\nabla[\ln(X_t)] = \ln(X_t) - \ln(X_{t-1}) = \ln(1 + Y_t)$$

Si el porcentaje Y_t permanece relativamente pequeño, entonces $\ln(1 + Y_t) \approx Y_t$. Cualquiera de los dos, $\nabla[\ln(X_t)]$ ó (86), son validos y se les conoce como el *rendimiento*.

En la Figura 11, se muestra la ACF correspondiente a los rendimientos y sus cuadrados, como se puede observar la ACF de los rendimientos no es aparentemente significativa, mientras la correspondiente a los cuadrados es altamente significativa. Otra observación interesante (en cualquiera de los dos gráficos) es que existen periodos o grupos (clusters) de alta o baja dispersión. Es decir, desviaciones grandes (pequeñas) de la serie con respecto a su nivel van seguidas de variaciones grandes (pequeñas). Por lo tanto, la volatilidad de la serie, **condicionada** por su historia reciente, presenta algún tipo de inercia o dependencia.

Los modelos que a continuación revisaremos se ajustan bien a este tipo de datos.

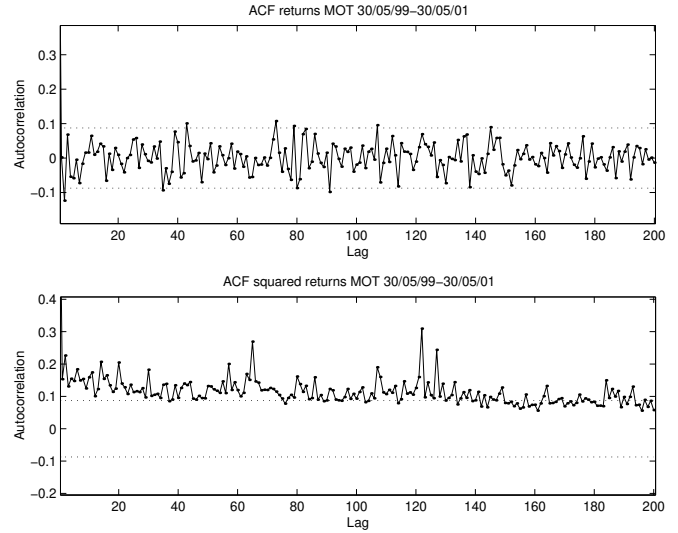


Figura 11: ACF muestral de los rendimientos, y_t , y sus cuadrados, asociados una serie observada de la acción de Motorola.

25. Modelos de heteroscedasticidad condicional autorregresiva (ARCH)

Para entender los conceptos de una forma más clara analizaremos el modelo ARCH(1) dado por

$$Y_t = \sigma_t \varepsilon_t \quad (87)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2, \quad (88)$$

donde $\varepsilon_t \stackrel{iid}{\sim} N(0, 1)$. De la misma forma que en los modelos ARMA, algunas restricciones se deben de imponer sobre los parámetros. La más obvia es que $\alpha_1 \geq 0$, de otra forma σ_t^2 podría ser negativa. De la notación en (88) se ve claramente la dependencia sobre el cuadrado de los rendimientos pasados, tal y como lo requiere los datos de la Figura 11. Condicionado en el valor de Y_{t-1} , Y_t es Gaussiano

$$Y_t | Y_{t-1} \sim N(0, \alpha_0 + \alpha_1 Y_{t-1}^2), \quad (89)$$

lo cual establece una dependencia de tipo Markoviana.

De forma alternativa, el modelo ARCH(1) se puede escribir como un modelo AR(1) en $\{Y_t^2\}$. Para ver esto, observemos que las ecuaciones (87), (88) se pueden escribir como

$$\begin{aligned} Y_t^2 &= \sigma_t^2 \varepsilon_t^2 \\ \alpha_0 + \alpha_1 Y_{t-1}^2 &= \sigma_t^2, \end{aligned}$$

por lo que

$$Y_t^2 - (\alpha_0 + \alpha_1 Y_{t-1}^2) = \sigma_t^2 \varepsilon_t^2 - \sigma_t^2,$$

lo cual se puede escribir como

$$Y_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + v_t, \quad (90)$$

donde $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$, lo cual claramente se interpreta como un ruido no Gaussiano.

Dentro de las propiedades de los procesos ARCH, se puede ver que

$$\begin{aligned} E(Y_t) &= EE(Y_t | \{Y_{t-1}, Y_{t-2}, \dots, Y_1\}) \\ &= EE(Y_t | Y_{t-1}) = 0, \end{aligned} \quad (91)$$

ya que $E(Y_t | Y_{t-1}) = 0$. A un proceso con esta propiedad se le conoce como una **diferencia martingala**.

El proceso de “error”, v_t , también es una diferencia martingala. Para ver esto,

$$\begin{aligned} EE(v_t | \{Y_{t-1}, Y_{t-2}, \dots, Y_1\}) &= EE(v_t | Y_{t-1}) \\ &= E\{v_t E(\varepsilon_t^2 - 1)\} \\ &= 0 \end{aligned} \quad (92)$$

Tarea: Usa (92) para ver que $\text{Cov}(v_{t+h}, v_t) = 0$ y $\rho(v_{t+h}, v_t) = 0$ así como también $\rho(Y_{t+h}, Y_t) = 0$.

Si la varianza de $v_t < \infty$ y $0 < \alpha_1 < 1$, entonces se puede hablar de un proceso AR(1) causal para $\{Y_t^2\}$, y por lo tanto $E(Y_t^2)$ y $\text{Var}(Y_t^2)$ deben de ser constantes con respecto a t . Esto implica que

$$\begin{aligned} E(Y_t^2) &= \text{Var}(Y_t) = \frac{\alpha_0}{1 - \alpha_1} \\ \text{Var}(Y_t^2) &= E(Y_t^4) = \frac{3\alpha_0^2}{(1 - \alpha_1)^2} \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2}, \text{ si } 3\alpha_1^2 < 1 \end{aligned}$$

Tarea: Prueba las igualdades anteriores.

Usando estos resultados se puede ver que el **coeficiente de curtosis**, κ , de Y_t es

$$\kappa = \frac{E(Y_t^4)}{[E(Y_t^2)]^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \quad (93)$$

que es más grande que 3, la curtosis correspondiente a la distribución Gaussiana. Entonces, la distribución **marginal** de los rendimientos $\{Y_t\}$, es **leptocúrtica**, es decir, tiene **colas pesadas**, al menos más pesadas que la distribución normal.

Tarea: Demuestra que la ACF correspondiente a Y_t^2 esta dada por $\rho_{Y_t^2}(h) = \alpha_1^h \geq 0$, para toda $h > 0$.

La generalización del proceso ARCH(1) al proceso ARCH(m), es decir con rezago de orden m en los cuadrados, esta dada por (87), pero con

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \dots + \alpha_m Y_{t-m}^2. \quad (94)$$

De forma análoga a los procesos ARMA, los procesos ARCH(m) también pueden ser generalizados para incluir la parte de “promedios móviles”. Esto último, resulta en los

procesos **ARCH generalizados** (GARCH(m, r)) introducidos por Bollerslev (1986), alumno de Engle, y están definidos por (87), pero con

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^m \alpha_j Y_{t-j}^2 + \sum_{j=1}^r \beta_j \sigma_{t-j}^2. \quad (95)$$

De manera análoga a los modelos ARCH, los modelos GARCH también pueden ser escritos como un modelo ARMA no Gaussiano. A manera de ilustración, consideremos el caso de un proceso GARCH(1, 1), es decir

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Bajo la condición $\alpha_1 + \beta_1 < 1$, podemos escribir

$$Y_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1) \quad (96)$$

$$\beta_1(Y_{t-1}^2 - \sigma_{t-1}^2) = \beta_1 \sigma_{t-1}^2(\varepsilon_{t-1}^2 - 1), \quad (97)$$

y substrayendo (97) de (96) se tiene

$$Y_t^2 = \alpha_0 + (\alpha_1 + \beta_1) Y_{t-1}^2 + v_t - \beta_1 v_{t-1}, \quad (98)$$

donde $v_t = \sigma_t^2(\varepsilon_t - 1)$.

Es importante señalar que en estas interpretaciones, de los modelos ARCH y GARCH como AR y ARMA respectivamente, el ruido en cuestión, $\{v_t\}$, también se puede ver como un ruido blanco ya que no esta correlacionado. Sin embargo, claramente existe una dependencia temporal en dicho ruido.

Tarea: Demuestra que el coeficiente de curtosis correspondiente a un modelo GARCH(1, 1), con $\alpha_1 + \beta_1 < 1$, esta dado por

$$\kappa = \frac{6\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2}$$

La estimación de los parámetros en modelos GARCH se puede implementar mediante **estimación máxima verosímil condicionada**. Por ejemplo, en el caso de un modelo ARCH(1), la verosimilitud de $\{y_2, \dots, y_n\}$ condicionada en y_1 esta dada por

$$L(\alpha_0, \alpha_1 | y_1) = \prod_{t=2}^n N(y_t; 0, \alpha_0 + \alpha_1 y_{t-1}^2) \quad (99)$$

Esta verosimilitud se puede maximizar mediante métodos numéricos.

Existen muchas otras generalizaciones de los modelos ARCH, por ejemplo, intercambiando ε_t por otro tipo de distribuciones que consideren colas más pesadas o con algún tipo de asimetría.

26. Un modelo ARCH(1) estrictamente estacionario

Supóngase, que dado un nivel de volatilidad al tiempo t , digamos σ_t^2 , los rendimientos, Y_t , se distribuyen Gaussianos. En notación, podríamos escribir como antes $Y_t = \sigma_t \varepsilon_t$ o alternativamente $Y_t \mid \sigma_t \stackrel{\text{ind}}{\sim} \text{N}(0, \sigma_t^2)$. Ahora, también supóngase que la volatilidad sigue una distribución **gamma inversa**, es decir, el inverso de una gamma.

Recordemos que si $Z \sim \text{Ga}(a, b)$, entonces la correspondiente función de densidad esta dada por

$$\text{Ga}(z; a, b) = \frac{b^a}{\Gamma(a)} z^{a-1} \exp\{-bz\}, \quad (100)$$

donde $a > 0$ y $b > 0$. Por lo tanto, otra variable definida como $W = 1/Z$ tiene función de densidad

$$\text{Iga}(w; a, b) = \frac{b^a}{\Gamma(a)} w^{-a-1} \exp\{-b/w\}, \quad (101)$$

donde Iga denota el inverso de una distribución gamma.

Así pues, se tiene

$$Y_t \mid \sigma_t^2 \stackrel{\text{ind}}{\sim} \text{N}(0, \sigma_t^2) \quad (102)$$

$$\sigma_t^2 \stackrel{\text{iid}}{\sim} \text{Iga}(\nu/2, \nu\beta^2/2). \quad (103)$$

Con esto se puede ver, que el nivel de volatilidad, σ_t^2 condicionado sobre el valor previo del rendimiento, Y_{t-1} se distribuye gamma, es decir,

$$\sigma_t^2 \mid Y_{t-1} \sim \text{Ga}\left(\frac{\nu+1}{2}, \frac{Y_{t-1}^2 + \nu\beta^2}{2}\right). \quad (104)$$

Con esta última ecuación se establece una relación entre Y_t y Y_{t-1} a través de σ_t^2 , sin embargo esta relación esta condicionada a un valor específico de σ_t^2 . Si integramos sobre todos los posibles valores de σ_t^2 obtenemos

$$\begin{aligned} f(y_t \mid y_{t-1}) &= \int_{\mathbb{R}_+} \text{N}(0, z^{-1}) \text{Ga}\left(z; \frac{\nu+1}{2}, \frac{y_{t-1}^2 + \nu\beta^2}{2}\right) \\ &= \text{St}\left(y_t; 0, \frac{y_{t-1}^2 + \nu\beta^2}{\nu+1}, \nu+1\right) \end{aligned} \quad (105)$$

donde $\text{St}(x; \mu, s, \nu)$ denota la densidad de una distribución t de Student con media μ , varianza s y ν grados de libertad.

El modelo (105) también se puede escribir como

$$Y_t = \sqrt{\frac{Y_{t-1}^2 + \nu\beta^2}{\nu+1}} \xi_t, \quad \xi_t \sim \text{St}(y_t; 0, 1, \nu+1). \quad (106)$$

Nótese, que este modelo se puede escribir como en (87) y (88) con la diferencia que en este caso los ruidos no son Gaussianos y $\alpha_0 = \nu\beta^2/(1+\nu)$ y $\alpha_1 = 1/(1+\nu)$.

Una de las ventajas más importantes de este modelo es que si asumimos $Y_1 \sim \text{St}(y_1; 0, 1, \nu)$ y $f(Y_t \mid Y_{t-1})$ esta dado como en (105) entonces el proceso es **estrictamente** estacionario, por lo cual no hay ninguna necesidad restringir los parámetros ν y β .

Se puede demostrar que si $\nu > 4$ se tiene $\rho_Y(h) = 0$ y $\rho_{Y_t^2}(h) = (1-\nu)^{-h}$.

Este proceso se puede simular fácilmente:

$$Y_1 \sim \text{St}(y_1; 0, 1, \nu)$$

$$\sigma_2^2 \mid Y_1 \sim \text{Ga}\left(\frac{\nu+1}{2}, \frac{Y_1^2 + \nu\beta^2}{2}\right)$$

$$Y_2 \mid \sigma_2^2 \sim \text{N}(0, \sigma_2^{-2})$$

⋮

$$\sigma_t^2 \mid Y_{t-1} \sim \text{Ga}\left(\frac{\nu+1}{2}, \frac{Y_{t-1}^2 + \nu\beta^2}{2}\right).$$

$$Y_t \mid \sigma_t^2 \sim \text{N}(0, \sigma_t^{-2})$$

⋮

Muchos otros modelos se pueden encontrar si se cambia la distribución de la $\text{Ga}(\cdot, \cdot)$ por otra más conveniente, según se requiera.

Tarea (opcional): Hacer un programa en R (o el lenguaje de preferencia) que implemente el algoritmo anteriormente descrito. Utilice esta implementación para simular una serie de datos con una elección de $\beta > 0$ y $\nu > 0$ dada. Verifica que $\rho_{Y_t^2}(h) = (1-\nu)^{-h}$. Qué sucede cuando $\nu = 1$? Simula varios escenarios variando β y con $\nu = 4$. Cuál es la labor del parámetro β ?