

MEZCLAS DE DISTRIBUCIONES Y VOLATILIDAD

JUAN CARLOS ABRIL¹, MARÍA DE LAS MERCEDES ABRIL² Y CARLOS ISMAEL
MARTÍNEZ³

*Universidad Nacional de Tucumán, Facultad de Ciencias Económicas y Consejo
Nacional de Investigaciones Científicas y Técnicas (CONICET). Argentina.*

¹jabril@herrera.unt.edu.ar, ²mabrilblanco@hotmail.com y

³cimartinez@herrera.unt.edu.ar

RESUMEN

Mediante métodos Monte Carlo se generan series de tiempo con las siguientes características: a) series cuyas distribuciones son mezclas de dos distribuciones normales con varianzas diferentes, b) series que satisfacen un modelo de volatilidad, c) series que satisfacen un modelo autorregresivo de primer orden (AR(1)) pero con errores contaminados que se distribuyen como las mezclas dadas en a) y d) series cuyas distribuciones son mezclas como las dadas en a) pero con heterocedasticidad condicional. Se estudian las semejanzas y diferencias entre estas series de tiempo. Del análisis se observa que en situaciones prácticas puede, en algunos casos, resultar difícil identificar el verdadero proceso generador de las series. Efectivamente, los procesos que surgen de mezclas de distribuciones tienen muchas características similares a los procesos que satisfacen el esquema de volatilidad.

El análisis es realizado por un lado mediante las consideraciones teóricas correspondientes, y por otro lado mediante el uso intensivo de las herramientas habituales en el proceso de identificación de las series de tiempo: esto es, los gráficos de las series, los histogramas, las distribuciones muestrales correspondientes, los correlogramas y los correlogramas parciales.

Palabras claves: Autorregresión, Errores contaminados, Mezclas de distribuciones, Modelos AR(1), Modelos ARCH, Volatilidad.

1. Combinación de distribuciones normales

Supongamos que tenemos el proceso $\{u_t\}$ cuya distribución es una mezcla de dos densidades normales, esto es $N(\mu, \sigma^2)$ y $N(\mu, k\sigma^2)$, donde $0 < \sigma^2 < \infty$, $k > 0$ y sin pérdida de generalidad podemos suponer que $\mu = 0$. Supongamos también que las variables aleatorias u_t son independientes entre sí. Siguiendo a Lindsay (1995), la mezcla puede

escribirse como $pN(0, k\sigma^2) + (1 - p)N(0, \sigma^2)$, $0 < p < 1$. Entonces, la densidad de u_t puede escribirse como

$$g(u_t) = \frac{p}{\sqrt{2\pi}\sqrt{k}\sigma} \exp\left(-\frac{u_t^2}{2k\sigma^2}\right) + \frac{1-p}{\sqrt{2\pi}\sigma} \exp\left(-\frac{u_t^2}{2\sigma^2}\right),$$

donde ahora tenemos los parámetros σ^2 , p y k , $0 < p < 1$. Se puede considerar a p como la proporción de “contaminación” existente entre las variables con densidad $N(0, \sigma^2)$.

Vemos que $E(u_t) = 0$ y

$$\begin{aligned} \text{var}(u_t) &= E(u_t^2) = \sigma^2\{pk + (1 - p)\} \\ &= \sigma_u^2. \end{aligned}$$

De lo dicho anteriormente, es claro que el proceso $\{u_t\}$ es estacionario (en el sentido amplio). De acuerdo a Durbin y Koopman (2001), este es un proceso cuya distribución tiene “colas pesadas” y puede ser usado para explicar algunos hechos que pueden tomar valores extremos con una probabilidad superior a la normal.

Sin pérdida de generalidad vamos a considerar en nuestro caso que $\sigma^2 = 1$.

2. La volatilidad

La volatilidad se define como la varianza de una variable aleatoria, comúnmente un retorno en aplicaciones económicas, condicional a toda la información pasada. Como la volatilidad no puede ser medida directamente, la misma puede manifestarse de varias maneras en una serie financiera.

A continuación se introduce una notación que será utilizada en el. Sea x_t la serie bajo estudio. Definimos

$$\mu_t = E(x_t | \mathcal{F}_{t-1}) = E_{t-1}(x_t), \quad (1)$$

$$\begin{aligned} h_t &= \text{var}(x_t | \mathcal{F}_{t-1}) = E\{(x_t - \mu_t)^2 | \mathcal{F}_{t-1}\} \\ &= E_{t-1}(y_t - \mu_t)^2 = \text{var}_{t-1}(x_t), \end{aligned} \quad (2)$$

como la media y la varianza condicionales de x_t dada la información hasta el instante $t - 1$ contenida en \mathcal{F}_{t-1} .

Un modelo típico para la volatilidad es de la forma

$$x_t = \mu_t + \sqrt{h_t}\varepsilon_t, \quad (3)$$

donde $E_{t-1}(\varepsilon_t) = 0$, $\text{var}_{t-1}(\varepsilon_t) = 1$ y típicamente los ε_t son independientes e idénticamente distribuidos (*IID*) con distribución F . La media y la varianza incondicional de x_t se denotarán como $\mu_x = E(x_t)$ y $\sigma_x^2 = \text{var}(x_t)$, respectivamente, y sea G la distribución de x_t . Es claro que (1), (2) y F determinan μ_x , σ_x^2 y G , pero no lo contrario. Mayores detalles pueden verse en Abril (2012).

3. Diseño de las simulaciones

Cuando se trabaja con una sola serie, sin hacer replicaciones, consideramos el tamaño muestral $T = 2000$. Cuando se hacen n replicaciones, tomamos el tamaño muestral $T = 200$ y el número de replicaciones $n = 1000$. A partir de esto generamos series de tiempo empíricas de tamaño T por medio de los siguientes pasos:

3.1. Generación de los datos de las mezclas

- a. Se generan números pseudoaleatorios independientes $N(0, 1)$, denotados por $\varepsilon_1, \dots, \varepsilon_T$. La sucesión se inicia con un valor “semilla” provisto automáticamente (“by default”) por el programa. La semilla puede también ser seleccionada por el operador.
- b. Se genera \hat{p}_t , $t = 1, \dots, T$, a partir de la distribución uniforme en el intervalo $(0, 1)$, esto es a partir de una $U(0, 1)$.
- c. Se transforma ε_t para producir la mezcla

$$u_t = \begin{cases} \sqrt{k} \varepsilon_t, & \text{si } \hat{p}_t \leq p, \\ \varepsilon_t, & \text{si } \hat{p}_t > p, \end{cases} \quad (4)$$

para $t = 1, \dots, T$, donde los valores de k y p han sido dados en el proceso de generación de la mezcla. En nuestro caso hemos tomado $k = 9, 16, 25$ y 100 ; $p = 0, 10; 0, 15; 0, 20$ y $0, 30$, habiéndose trabajado con todas las combinaciones de los valores de (k, p) dados. En situaciones prácticas podría suceder que se necesite estimar a estos valores de k y p , ya que tienen las características de parámetros desconocidos. La fórmula (4) puede ser escrita de la siguiente manera

$$u_t = \sqrt{kb(\hat{p}_t - p)} \varepsilon_t, \quad (5)$$

donde

$$b(a_t) = \begin{cases} 1, & \text{si } a_t \leq 0, \\ k^{-1}, & \text{si } a_t > 0. \end{cases} \quad (6)$$

3.2. Generación de los datos de la volatilidad

d. Se genera la serie $\{x_t\}$ que satisface el modelo

$$x_t = \sqrt{h_t}\varepsilon_t, \quad (7)$$

$$h_t = \alpha_0 + \alpha_1 x_{t-1}^2, \quad t = 1, \dots, T. \quad (8)$$

donde ε_t fue generada en el punto a. anterior, y los valores de α_0 y α_1 han sido dados en el proceso de generación de la serie. en nuestro caso tomamos $\alpha_0 = 1$ y $\alpha_1 = 0,5$. Es claro que (7) y (8) define una serie que satisface un modelo ARCH(1).

e. Se usa la serie $\{x_t\}$ para estimar los valores de α_0 y α_1 .

3.3. Generación de los datos del modelo AR(1)

f. Para un dado valor del parámetro ϕ , generar la serie $\{y_t\}$ que satisface el modelo

$$y_t = \phi y_{t-1} + u_t, \quad t = 1, \dots, T, \quad (9)$$

donde $\{u_t\}$ fue definido en (5) y (6) anteriores. En nuestro caso tomamos $\phi = 0,6$.

g. Se usa la serie y_t para obtener estimaciones $\hat{\phi}$ de ϕ . Esas estimaciones se realizan mediante el procedimiento de máxima verosimilitud bajo normalidad.

3.4. Generación de los datos de las mezclas con heterocedasticidad condicional

h. A partir de los $\varepsilon_1, \dots, \varepsilon_T$ generados en el punto a. y los \hat{p}_t , $t = 0, 1, \dots, T - 1$, generados en el punto b. se crea la serie

$$w_t = \sqrt{kc(\hat{p}_{t-1} - p)}\varepsilon_t, \quad (10)$$

donde

$$c(a_{t-1}) = \begin{cases} 1, & \text{si } a_{t-1} \leq 0, \\ k^{-1}, & \text{si } a_{t-1} > 0, \end{cases} \quad (11)$$

y los valores de k y p usados son los dados en el punto c.

4. Análisis de las series y conclusiones

La estructura de la serie definida en (5) y (6) es similar a la de un modelo típico de volatilidad definido en (3) pero el analizar los correlogramas y correlogramas parciales de

la serie $\{u_t\}$ y la serie $\{u_t^2\}$ se observan que para ambas se podría aceptar la hipótesis de falta de correlación serial, lo que concuerda con la teoría subyacente y con el proceso generador de la serie. La única particularidad que se observa es que para la serie $\{u_t^2\}$, en todas las combinaciones de k y p , la autocorrelación y la autocorrelación parcial muestral de orden 11 conducen a rechazar las hipótesis que los respectivos parámetros son cero, lo que no tiene una explicación teórica. Esto último podría llevar en casos prácticos a identificar para $\{u_t\}$ un modelo ARCH(11) con algunos de los coeficientes de ordenes menores que 11 iguales a cero y con errores distribuidos proporcionalmente a una χ_1^2 . Cuando se estima el respectivo modelo ARCH(11) con coeficientes menores que 11 iguales a cero, los residuos resultantes satisfacen la hipótesis de que provienen de un proceso constituido por variables aleatorias independientes e idénticamente distribuidas con media y varianzas constantes.

La serie definida en (10) y (11) satisface la definición de volatilidad y cumple con lo dado en (1) y (2). Por otra parte, esa serie $\{w_t\}$ es, desde un punto de vista práctico, indistinguible de la serie $\{u_t\}$ definida en (5) y (6) lo que corrobora el hecho que mezcla de distribuciones con diferentes varianzas y volatilidad están íntimamente relacionados y pueden, en algunos casos, ser confundidas unas con otras.

La serie definida en (7) y (8) fue generada siguiendo un modelo ARCH(1). Cuando se analiza el correlograma y correlograma parcial de la serie $\{x_t\}$ se observa que se podría aceptar la hipótesis de falta de correlación serial. Por otra parte, cuando se analiza a la serie $\{x_t^2\}$ se observa que se puede identificar un modelo AR(2) con errores distribuidos proporcionalmente a una χ_1^2 . Todo esto concuerda con la respectiva teoría. En cuanto a la estimación de α_0 y α_1 , a pesar que la misma se la hace por máxima verosimilitud bajo el supuesto de normalidad, se encuentra que las distribuciones empíricas de esos estimadores logradas por métodos de Monte Carlo están centradas en el verdadero valor de los respectivos parámetros.

La serie definida en (9) corresponde a un proceso AR(1) con errores cuya distribución es una mezcla de normales con diferentes varianzas. Las estimaciones del parámetro ϕ , a pesar de ser hechas por máxima verosimilitud bajo el supuesto de normalidad, conduce a que la distribución empírica del estimador lograda por métodos de Monte Carlo está centrada en el verdadero valor del parámetro. Con los residuos de esta serie se puede identificar un modelo ARCH(11), similar a lo que ocurriría con la serie $\{u_t\}$. lo que nuevamente nos conduce a pensar que en muchas situaciones prácticas sería posible confundir modelos con “colas pesadas” con modelos de volatilidad.

En todos los casos, la estimación por máxima verosimilitud bajo el supuesto de normalidad de la varianzas no condicional de las diferentes series dio estimadores cuyas dis-

tribuciones empíricas están centradas en el verdadero valor del parámetro respectivo.

5. Referencias

ABRIL, JUAN CARLOS. (1999), *Análisis de Series de Tiempo Basado en Modelos de Espacio de Estado*, EUDEBA: Buenos Aires.

ABRIL, JUAN CARLOS. (2004). *Modelos para el Análisis de las Series de Tiempo*. Ediciones Cooperativas, Buenos Aires.

ABRIL, MARÍA DE LAS MERCEDES. (2012). *El Enfoque de Espacio de Estado de las Series de Tiempo para el Estudio de los Problemas de Volatilidad*. Tesis Doctoral en ejecución. Universidad Nacional de Tucumán. Argentina.

BOLLESLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-27.

DURBIN, J. Y S. J. KOOPMAN (2001), *Time Series Analysis by State Space Methods*, Oxford University Press: Oxford.

DURBIN, J. Y S. J. KOOPMAN. (2012). *Time Series Analysis by State Space Methods* (2nd Edition). Oxford University Press: Oxford.

ENGLE, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, **50**, 987-1007.

LINDSAY, B. G. (1995). *Mixture models: theory, geometry and applications*. Regional conference series in probability and statistics, Institute of Mathematical Statistics.

SHEPHARD, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press: Oxford.