

*Un Enfoque Bayesiano para la Estimación de Modelos
TAR Multivariados con Distribución de Error
Generalizada*

DAVINSON STEV ABRIL SALCEDO
ECONOMISTA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
DICIEMBRE DE 2018

*Un Enfoque Bayesiano para la Estimación de Modelos
TAR Multivariados con Distribución de Error
Generalizada*

DAVINSON STEV ABRIL SALCEDO
ECONOMISTA

DISERTACIÓN PRESENTADA PARA OPTAR AL TÍTULO DE
MSC EN ESTADÍSTICA

DIRECTOR
SERGIO ALEJANDRO CALDERÓN, PH.D.
DOCTOR EN ESTADÍSTICA

LÍNEA DE INVESTIGACIÓN
SERIES DE TIEMPO



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
DICIEMBRE DE 2018

Título en español

Un Enfoque Bayesiano para la Estimación de Modelos TAR Multivariados con Distribución de Error Generalizada.

Title in English

A Bayesian Approach for the estimation of Multivariate Threshold Autoregressive models with Generalized Error Distribution.

Resumen: En este documento se utiliza un enfoque Bayesiano para la estimación de los parámetros de un modelo TAR multivariado con Distribución de Error Generalizada Multivariada (MGED). Esta distribución nos permite modelar una gran variedad de procesos sin fijarnos, preliminarmente, en la distribución multivariada asociada. A partir de la implementación de algoritmos adaptativos se extraen muestras de las densidades condicionales completas de los parámetros. Se evalúa la metodología propuesta a través de simulaciones. Finalmente, se modeló la relación de dos índices bursátiles latinoamericanos (COLCAP y BOVESPA) donde la variable de umbral está dada por el índice norteamericano S&P500.

Abstract: This document uses a Bayesian approach for estimating the parameters of a multivariate TAR model with Multivariate Generalized Error Distribution (MGED). This distribution allows us to model a wide variety of processes without worrying us, preliminary, in the associated multivariate distribution. Based on the implementation of adaptive algorithm, samples of the complete conditional densities of the parameters are extracted. The proposed methodology is evaluated through simulations. Finally, we modeled the relationship of two Latin American stock indexes (COLCAP and BOVESPA) where the threshold variable is given by the North American index S&P500.

Palabras clave: Modelos TAR multivariados, Distribución Generalizada de Error Multivariada, Métodos MCMC adaptativos, Estimación Bayesiana.

Keywords: Multivariate TAR models, Multivariate Generalized Error Distribution, Adaptive MCMC methods, Bayesian Estimation.

Agradecimientos

A mis padres por sus enseñanzas, consejo y amor durante cada etapa de mi vida. A mis hermanos quienes aún en momentos difíciles, logran sacarme una sonrisa. A mi familia a quienes debo todo. A Luis Fernando por impregnarme su amor por la estadística y ser un apoyo fundamental en mi desarrollo personal y profesional. Al profesor Sergio Alejandro Calderón y a la profesora Hanwen Zhang por su consejo y colaboración durante el desarrollo de este trabajo. Y finalmente a mis amigos/as y compañeros/as de vida.

Índice general

Índice general	II
Índice de tablas	III
Índice de figuras	IV
Introducción	V
1. El Modelo TAR Multivariado con Distribución de Error Generalizada Multivariada	1
1.1. Distribución de Error Generalizada Multivariada (MGED)	1
1.2. Descripción del proceso MTAR	6
1.3. Función de verosimilitud	7
2. Generalidades de la estimación Bayesiana e implementación de los métodos MCMC	9
2.1. Generalidades	9
2.2. Métodos tradicionales	10
2.2.1. Algoritmo Metropolis-Hastings	10
2.2.2. Muestreador de Gibbs	11
2.3. Métodos Adaptativos	12
2.3.1. Principios	13
2.3.2. Metropolis Adaptativo (MA)	13
2.3.3. Metropolis Adaptativo dentro de Gibbs	14
3. Estimación de los parámetros	16
3.1. Distribuciones previas	16

3.2. Estimación de los parámetros no estructurales: distribuciones condicionales completas	17
3.2.1. Coeficientes autoregresivos (θ_j)	17
3.2.2. Matriz de covarianza ($\Sigma_{(j)}$)	18
3.2.3. Parámetro de no normalidad (κ)	18
3.3. Estimación de los parámetros estructurales	19
3.3.1. Parámetro de umbral \mathbf{r}	19
3.3.2. Elección del mejor modelo: estimación de los parámetros l, p, q y d	19
3.4. Algoritmo para la estimación de los parámetros	21
4. Simulaciones	23
4.1. Modelo MTAR con MGED con $\kappa = 0.1$	24
4.1.1. Resultados de la estimación	24
4.1.1.1. Estimación de los parámetros no estructurales	25
4.1.1.2. Estimación de los parámetros estructurales	25
4.1.2. Diagnóstico del modelo	26
4.2. Modelo MTAR con MGED con $\kappa = 1$	28
4.2.1. Resultados de la estimación	29
4.2.1.1. Estimación de los parámetros no estructurales	29
4.2.1.2. Estimación de los parámetros estructurales	31
4.2.2. Diagnóstico del modelo	32
5. Aplicación Empírica	34
5.1. Hechos estilizados de las series financieras	34
5.2. Datos	35
5.3. Estimación del modelo	37
5.3.1. Estimación de los parámetros estructurales	37
5.3.2. Estimación de los parámetros no estructurales	38
5.4. Diagnóstico del modelo	39
Conclusiones	42
Trabajo futuro	43
Bibliografía	44
A. Construcción de muestras MGED	1

A.1. Obtención muestras de \mathbf{R}	1
A.2. Algoritmo para la obtención de muestras	2
B. Otras simulaciones	3
B.1. Modelo 1: MTAR trivariado de dos regímenes con $\kappa = 0.5$	3
B.2. Modelo 2: MTAR bivariado de tres regímenes con $\kappa = 0.5$	5
B.3. Modelo 3: MTAR bivariado de dos regímenes con $\kappa = 0.5$	7
B.4. Modelo 4: MTAR bivariado de dos regímenes con $\kappa = 8$	8
C. Diagnóstico de cadenas	11
C.1. Tasas de cobertura	11
C.1.1. Modelo MTAR con MGED con $\kappa = 0.1$	11
C.1.2. Modelo MTAR con MGED con $\kappa = 1$	12
C.2. Pruebas de Convergencia	13
C.2.1. Modelo MTAR con MGED con $\kappa = 0.1$	14
C.2.2. Modelo MTAR con MGED con $\kappa = 1$	16

Índice de tablas

4.1. Valores propuestos de τ y α para el algoritmo de obtención de muestras según valores de κ y n	23
4.2. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo 4.1 con $\kappa = 0.1$ y $r = \hat{z}_{0.5} = 0$. Los valores de los parámetros se presentan entre paréntesis en las columnas de estimación.	26
4.3. Log verosimilitud marginal estimada para el modelo 4.1	26
4.4. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo 4.2 con $\kappa = 1$ y $r = \hat{z}_{0.5} = 0$. Los valores de los parámetros se presentan entre paréntesis en las columnas de estimación.	31
4.5. Log verosimilitud marginal estimada para el modelo 4.2	31
5.1. Log verosimilitud marginal estimada para el modelo	38
5.2. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo	40
B.1. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo B.1 con $\kappa = 0.5$ y $r = \hat{z}_{0.5}$	4
B.2. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo B.2 con $\kappa = 0.5$, $r_1 = \hat{z}_{0.44} = -0.15$ y $r_2 = \hat{z}_{0.54} = 0.1$	6
B.3. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo B.3 con $\kappa = 0.5$ y $r = \hat{z}_{0.5} = 0$	8
B.4. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo B.4 con $\kappa = 8$ y $r = \hat{z}_{0.5} = 0$	9
C.1. Tasa de cobertura y sesgo estandarizado de las estimaciones del modelo 4.1	12
C.2. Tasa de cobertura y sesgo estandarizado de las estimaciones del modelo 4.2	13
C.3. Pruebas de convergencia de Gelman & Rubin (1992) y Heidelberger & Welch (1983). $*H_0$: La cadena es estacionaria.	15
C.4. Pruebas de convergencia de Gelman & Rubin (1992) y Heidelberger & Welch (1983). $*H_0$: La cadena es estacionaria.	17

Índice de figuras

1.1. Exceso de curtosis $\gamma_{2,n}$ para $\kappa > \frac{1}{2}$	2
1.2. Funciones de densidad para cuatro distribuciones generalizadas de error bivariadas y esféricas con parámetro $\kappa = 0.1, 0.5, 1, 3$	3
2.1. Cadenas obtenidas por un algoritmo tradicional para un valor de la varianza 1) pequeño, 2) grande y 3) mediano	12
2.2. Primera coordenada de la cadena de Márkov AM para $n = 100$ y $n = 200$	14
4.1. Serie simulada del proceso descrito en 4.1	24
4.2. Valores de las cadenas para algunos parámetros del modelo 4.1	25
4.3. Pruebas CUSUM y CUSUMSQ para los residuos marginales del modelo 4.1	27
4.4. Residuales marginales del modelo 4.1	28
4.5. Serie simulada del proceso 4.2	29
4.6. Valores de las cadenas para algunos parámetros del modelo 4.2	30
4.7. Pruebas CUSUM y CUSUMSQ para los residuos marginales del modelo 4.2	32
4.8. Residuales del modelo 4.2	33
5.1. Índices accionarios COLCAP, BOVESPA Y S&P500	35
5.2. Retornos de los índices accionarios COLCAP, BOVESPA y S&P500	36
5.3. Valores de las cadenas para algunos parámetros del modelo	39
5.4. Pruebas CUSUM y CUSUMSQ para los residuos marginales del modelo	40
5.5. Residuales marginales del modelo	41
B.1. Serie simulada del proceso descrito en B.1	4
B.2. Serie simulada del proceso descrito en B.2	5
B.3. Serie simulada del proceso descrito en B.3	7
B.4. Serie simulada del proceso descrito en B.4	9
B.5. Densidades a posteriori de algunos parámetros del modelo B.4	10

C.1. Prueba de convergencia de Geweke (1992) para algunos parámetros	16
C.2. Prueba de convergencia de Geweke (1992) para algunos parámetros	18

Introducción

El modelo autoregresivo de umbrales, TAR por sus siglas en inglés, propuesto por Tong (1978) es un sistema dinámico respecto al proceso estocástico de interés, donde el proceso de entrada $\{Z_t\}$ determina las dinámicas del proceso de $\{Y_t\}$. En su forma más básica, es un proceso lineal por tramos entre la variable de interés Y_t , el conjunto $Y_t, Y_{t-1}, \dots, Y_{t-s}$ y Z_t (variable de umbral), donde $s = \max\{p_1, \dots, p_l\}$ (Nieto, 2008). Algunas características de los modelos TAR son los patrones de crecimiento o decrecimiento asimétricos y la irreversibilidad en el tiempo.

El uso y aplicación de los modelos TAR se ha extendido a todo tipo de campos como en meteorología, donde se utiliza para medir el impacto de variables climáticas sobre variables ambientales o socio-económicas (Al-Awadhi & Jolliffe, 1998; Calderón & Nieto, 2017). En psicología, donde el comportamiento afectivo de los cónyuges durante una discusión puede ser modelado (De Haan-Rietdijk et al., 2016). También en finanzas, donde se ha modelado la relación entre variables financieras y de actividad económica real (Domian & Louton, 1997; Barnes, 1999; Li et al., 2015). Y muchas más.

Desde su concepción, diferentes autores han alimentado la literatura existente sobre modelos TAR, así por ejemplo, Tsay (1989) brinda un procedimiento de modelamiento simple y ampliamente aplicable en estos modelos que consiste en la construcción de una prueba para no linealidad de tipo TAR. Luego, en Tsay (1998) se propone una prueba de no linealidad basada en los residuos y un método de estimación de los parámetros de un modelo TAR multivariado (MTAR). Además, pruebas de no linealidad de procesos no estacionarios y cointegrados han sido propuestas por un amplio número de autores como Enders & Granger (1998), Caner & Hansen (2001), Enders & Ludlow (2002) y Balke & Fomby (1997), entre otros. Más recientemente, Nieto (2008) dedujo que la varianza condicional de los modelos TAR no es constante, por lo cual propuso hacer uso de ellos como alternativa a los modelos GARCH de Engle (1982) y Bollerslev (1986) ampliamente usados en el modelamiento de series financieras.

En los trabajos anteriormente mencionados se asume que el proceso de ruido sigue una distribución Gaussiana; sin embargo, en la práctica, el supuesto de normalidad sobre los errores no siempre es apropiado. Por ejemplo, en series de alta frecuencia como los retornos de una acción, las distribuciones de error suelen tener colas más pesadas. En particular en los modelos TAR donde se tienen patrones asimétricos, cambios de estado e irreversibilidad en el tiempo puede no cumplirse dicho supuesto en una gran cantidad de aplicaciones. Es por esto que muchos autores han adaptado distintas distribuciones al error de acuerdo a sus objetivos. Ante la falta de normalidad en el proceso de ruido, se

ha hecho uso de otras distribuciones para su modelamiento, por ejemplo, la distribución t-student.

En ese sentido Zhang & Nieto (2015) considera los modelos TAR univariados cuando el proceso de ruido blanco para cada t sigue una distribución t-student en donde se desarrolla un procedimiento Bayesiano para la identificación, estimación y pronóstico. Luego, como una extensión al caso multivariado, Romero-Orjuela (2018) propone la estimación Bayesiana de los parámetros no estructurales de un modelo TAR multivariado (MTAR) cuando los errores para cada t siguen una distribución t-student multivariada.

Pero, ¿qué tal si, en principio, no tuviéramos que preocuparnos por la distribución de dicho error? ¿Y si asumiéramos que su función de densidad puede tomar infinitas formas, bajo una media y varianza fijas en el tiempo, y sólo necesitáramos estimar un parámetro adicional? La Distribución Generalizada de Error (GED) es una alternativa. Así pues, como una aproximación al caso univariado en el artículo de Castro (2016) se estima los parámetros no estructurales de un modelo TAR cuando el proceso de ruido blanco se sigue a una GED, donde a través del Muestreador de Gibbs se recogen muestras de las densidades condicionales completas asumiendo que los parámetros estructurales son conocidos.

En este documento se utiliza un enfoque Bayesiano para la estimación de los parámetros de un modelo TAR multivariado con Distribución de Error Generalizada Multivariada (MGED). A través de la implementación de métodos MCMC adaptativos se obtienen muestras de las densidades condicionales completas de los parámetros. De esta manera, primero se presenta una descripción del modelo TAR multivariado con Distribución de Error Generalizada. Segundo, se procede con la descripción de algunas generalidades de la estimación Bayesiana y el uso de algunos métodos MCMC, entre ellos los algoritmos adaptativos. Tercero, se presenta el algoritmo de estimación propuesto donde destaca la propuesta de las distribuciones previas y la obtención de las densidades condicionales completas. Cuarto, la metodología propuesta es evaluada a través de simulaciones de 6 diferentes procesos cada uno con 100 realizaciones (600 ejecuciones en total). Quinto, se realiza una aplicación de la metodología propuesta para dos series bursátiles de Colombia y Brasil. Finalmente, se presentan algunas conclusiones y recomendaciones.

CAPÍTULO 1

El Modelo TAR Multivariado con Distribución de Error Generalizada Multivariada

En este capítulo se presenta el modelo TAR multivariado (MTAR) cuando la distribución del error, para cada t , sigue una Distribución de Error Generalizada Multivariada (MGED). Inicialmente, debido a que la distribución MGED es un caso particular de las distribuciones de Contorno Elíptico (EC), se enuncian las propiedades que serán útiles en la obtención de muestras del error en el procedimiento de simulación. Luego, se describe el proceso MTAR, la notación utilizada y supuestos del modelo. Por último, se obtiene la función de verosimilitud condicional.

1.1. Distribución de Error Generalizada Multivariada (MGED)

El vector n -dimensional ϵ se distribuye $MGED_n(\mu, \Sigma, \kappa)$ si su función de densidad puede ser expresada como:

$$f(\epsilon|\mu, \Sigma, \kappa) = \frac{1}{\sqrt{\pi^n |\Sigma|}} \frac{\Gamma(1 + \frac{n}{2})}{\Gamma(1 + n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{n}{2}} \exp \left\{ - \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} (\epsilon - \mu)' \Sigma^{-1} (\epsilon - \mu) \right]^{\frac{1}{2\kappa}} \right\} \quad (1.1)$$

donde $\mu \in \mathbb{R}^n$, $\kappa > 0$ y Σ es una matriz $n \times n$ definida positiva. La anterior parametrización fue propuesta por Guiller (2005), quien además encuentra los momentos de la distribución. La distribución es unimodal con moda μ que es igual a su media. La matriz de covarianza V es igual a la matriz Σ multiplicada por el factor de escala

$$\frac{\Gamma\{(n+2)\kappa\}\Gamma(1+\kappa)}{\Gamma(3\kappa)\Gamma(1+n\kappa)}.$$

Sustituyendo μ y Σ respectivamente en 1.1 por el vector de ceros y la matriz identidad, la distribución MGED estandarizada es obtenida. Note que en el límite cuando $\kappa \rightarrow 0$ el factor de escala es $\frac{3}{n+2}$ lo que muestra la fuerte dependencia de este factor sobre κ . Por su parte, κ es conocido como parámetro de no normalidad y está estrechamente relacionado a la curtosis. La simetría de la distribución, $\beta_{1,n}$, es cero y el parámetro de curtosis, $\beta_{2,n}$,

está dado por

$$\beta_{2,n} = n^2 \frac{\Gamma\{(n+4)\kappa\}\Gamma(n\kappa)}{\Gamma^2\{(n+2)\kappa\}}.$$

La Figura 1.1 muestra el exceso de curtosis $\gamma_{2,n}$ de la distribución respecto a una normal, para distintos valores de κ y n .

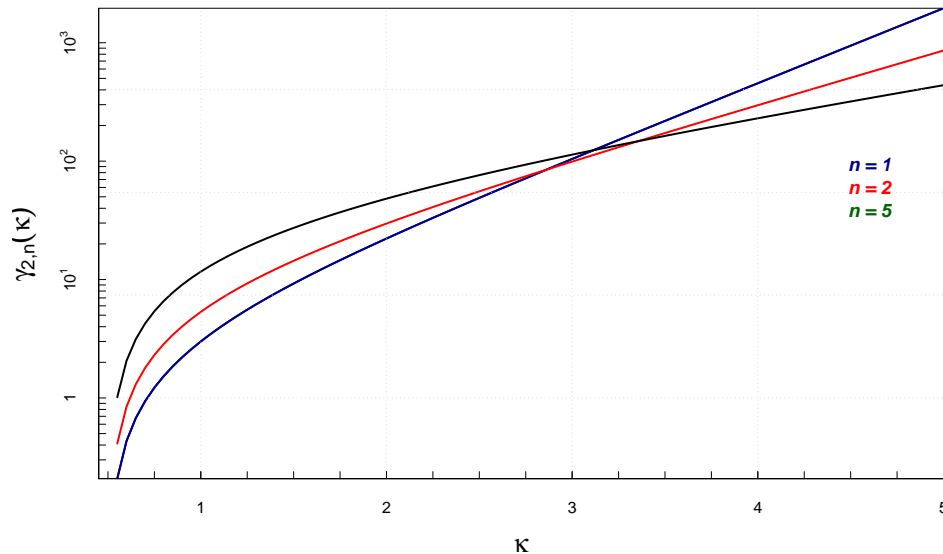


FIGURA 1.1. Exceso de curtosis $\gamma_{2,n}$ para $\kappa > \frac{1}{2}$

La MGED brinda una amplia variedad de familias de distribuciones a medida que el parámetro κ varía. Como casos especiales, para $\kappa = 1$ la distribución multivariada de Laplace (o doble exponencial multivariada) es obtenida, mientras que si $\kappa = 0.5$ se alcanza la distribución normal multivariada y en el límite $\kappa \rightarrow 0$ la densidad tiende a ser una uniforme $U^{(n)}$ sobre el hipercubo $(\mu - \sigma, \mu + \sigma)^n \in \mathbb{R}^n$ (Guiller, 2005). La figura 1.2 muestra funciones de densidad de cuatro distribuciones generalizadas de error bivariadas y esféricas para algunos valores de κ .

Al igual que la mayoría de distribuciones, la distribución MGED posee algunas propiedades que son útiles en la construcción de un algoritmo para la obtención de muestras. La distribución MGED pertenece a la familia multivariada de distribuciones Kotz¹, que es un caso particular de distribuciones elípticas y simétricas (Fang et al., 1990). Entre las distintas definiciones de distribuciones elípticas se considera la siguiente:

Definición 1.1.1 (McNeil et al. (2015)). El vector aleatorio n -dimensional \mathbf{W} se dice que tiene distribución de contorno elíptico y se denotará $\mathbf{W} \sim EC_n(\mu, A, g)$, si

$$\mathbf{W} \stackrel{d}{=} \mu + A\mathbf{R}$$

¹La familia de distribuciones multivariadas de Kotz se utiliza frecuentemente en la construcción de modelos matemáticos donde el usual supuesto de normalidad no siempre aplica (por ejemplo, en experimentos de siembra de nubes para describir la precipitación) (Nadarajah, 2003).

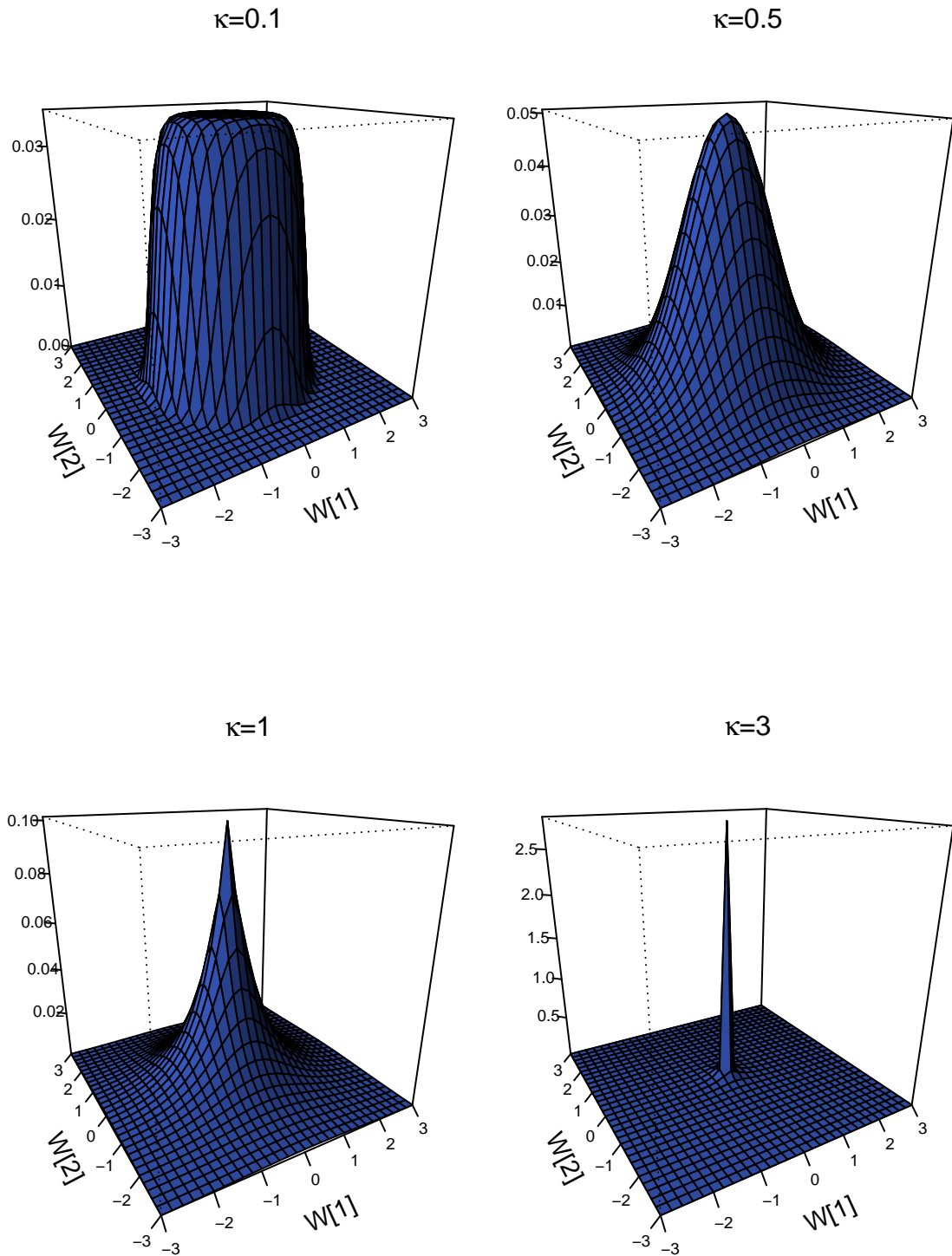


FIGURA 1.2. Funciones de densidad para cuatro distribuciones generalizadas de error bivariadas y esféricas con parámetro $\kappa = 0.1, 0.5, 1, 3$.

donde $\mathbf{R} \sim S_n(g)$ es de contorno esférico²; $g : \mathbb{R}^n \rightarrow \mathbb{R}$ es función de la variable de interés y es conocida como *generadora de densidad*; y $A \in \mathbb{R}^{n \times n}$ y $\mu \in \mathbb{R}^n$ son una matriz y vector de constantes, respectivamente. En otras palabras, las distribuciones elípticas son obtenidas por transformaciones afines multivariadas de distribuciones esféricas. De manera que su función característica, ϕ , está dada por

$$\phi_{\mathbf{W}}(t) = E(e^{it'\mathbf{W}}) = E(e^{it'(\mu + A\mathbf{R})}) = e^{it'\mu} E(e^{i(A't)'\mathbf{R}}) = e^{it'\mu} \phi(t'\Sigma t)$$

con $\Sigma = AA'$ y t real.

Además, la relación entre las distribuciones de contorno elíptico y esférico está dada por

$$\mathbf{W} \sim EC_n(\mu, \Sigma, g) \iff \Sigma^{-1/2}(\mathbf{W} - \mu) \sim S_n(g)$$

donde μ es un vector de localización, Σ una matriz de dispersión y g la función generadora de densidad. En la práctica, el caso más interesante es aquel en donde Σ es definida positiva.

De esta manera, si el vector esférico \mathbf{R} tiene generadora de densidad g , entonces $\mathbf{W} = \mu + \Sigma^{1/2}\mathbf{R}$ tiene densidad

$$f_{\mathbf{W}}(w|\mu, \Sigma) = |\Sigma|^{-1/2} g((w - \mu)'\Sigma^{-1}(w - \mu)). \quad (1.2)$$

Para el caso de la familia de distribuciones de Kotz, la función generadora de densidad $g(\cdot)$ es de la forma:

$$g(r) = c_n r^{H-1} \exp(-hr^m) \quad (1.3)$$

donde $h > 0$, $m > 0$, $2H + n > 2$, c_n es una constante de normalización y r denota la distancia de Malahanobis $\Delta_{\Sigma}^2(w, \mu) = (w - \mu)'\Sigma^{-1}(w - \mu)$. Así pues, la distribución MGED tiene contorno elíptico dado que su función de densidad (1.1) tiene la forma 1.2 y su función generadora de densidad puede ser escrita como en 1.3 con $H = 1$, $h = \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{1}{2\kappa}}$, $m = \frac{1}{2\kappa}$ y la constante de normalización $c_n = \frac{1}{\sqrt{\pi^n |\Sigma|}} \frac{\Gamma(1 + \frac{n}{2})}{\Gamma(1 + n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{n}{2}}$.

Algunas propiedades de las distribuciones de contorno elíptico son enunciadas a continuación y son tratadas en detalle por Fang et al. (1990) y McNeil et al. (2015).

1. Transformaciones lineales.

Si se toma una transformación lineal de vectores aleatorios elípticos, entonces estos tienen la misma función generadora g . Sea $\mathbf{W} \sim EC_n(\mu, \Sigma, g)$ y sea $A \in \mathbb{R}^{n \times n}$ y $b \in \mathbb{R}^n$. Entonces es sencillo ver que

$$A\mathbf{W} + b \sim EC_n(A\mu + b, A\Sigma A', g), \quad (1.4)$$

como un caso especial

$$b\mathbf{W} \sim EC_1(b'\mu, b'\Sigma b, g). \quad (1.5)$$

2. Distribuciones marginales.

Se sigue de 1.5 que las distribuciones marginales de W deben ser elípticas con la

²Un vector aleatorio n -dimensional X se dice tiene una distribución esférica, radial o de contorno esférico si y solo si AX tiene la misma distribución para todas las matrices ortogonales $n \times n$ (Kelker, 1970).

misma función generadora. Denotando $\mathbf{W} = (W_1, W_2)'$ y extendiendo esta notación naturalmente a μ y Σ :

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

se tiene que $W_1 \sim EC_k(\mu_1, \Sigma_{11}, g)$ y $W_2 \sim EC_{n-k}(\mu_2, \Sigma_{22}, g)$.

3. *Distribuciones condicionales.*

La distribución condicional de W_2 dado W_1 puede también ser elíptica, aunque generalmente con una función generadora distinta \tilde{g} . Los detalles de cómo la función generadora cambia puede verse en Fang et al. (1990).

En el caso especial de la normal multivariada, la función generadora se mantiene igual.

4. *Formas cuadráticas.*

Si $\mathbf{W} \sim EC_n(\mu, \Sigma, g)$ con Σ no singular, entonces se observa que

$$Q = (\mathbf{W} - \mu)' \Sigma^{-1} (\mathbf{W} - \mu) \stackrel{d}{=} R^2 \tag{1.6}$$

donde R es una variable aleatoria esférica. Para algunos casos particulares la distribución de R^2 es bien conocida. Así, si $\mathbf{W} \sim N_n(\mu, \Sigma)$, entonces $R^2 \sim \chi_n^2$; si $\mathbf{W} \sim t_n(\nu, \mu, \Sigma)$, entonces $R^2/n \sim F(n, \nu)$.

Para todas las distribuciones elípticas Q debe ser independiente de $\Sigma^{-1/2}(\mathbf{W} - \mu)/\sqrt{Q}$.

5. *Convoluciones.*

La convolución de dos vectores elípticos independientes con la misma matriz de dispersión Σ es también elíptica. Si \mathbf{W} y \mathbf{V} son vectores aleatorios independientes n -dimensionales satisfaciendo $\mathbf{W} \sim EC_n(\mu, \Sigma, g)$ y $\mathbf{V} \sim EC_n(\tilde{\mu}, \Sigma, \tilde{g})$, entonces a través del producto de funciones características se puede mostrar que

$$\mathbf{W} + \mathbf{V} \sim EC_n(\mu + \tilde{\mu}, \Sigma, \bar{g})$$

donde $\bar{g} = g(u)\tilde{g}(u)$.

Una definición un poco más conocida y útil en la obtención de muestras es aquella donde \mathbf{W} tiene distribución de contorno elíptico si y sólo si existe A , $\mathbf{U}^{(n)}$ y \mathbf{R} tal que

$$\mathbf{W} \stackrel{d}{=} \mu + \mathbf{R}A\mathbf{U}^{(n)} \tag{1.7}$$

con

- (i) $\mathbf{U}^{(n)}$ uniformemente distribuida sobre una superficie unitaria en \mathbb{R}^n .
- (ii) $\mathbf{R} \geq 0$ una variable aleatoria independiente de $\mathbf{U}^{(n)}$.
- (iii) $A \in \mathbb{R}^{n \times n}$ con $AA' = \Sigma$.

De esta manera, por 1.7 es posible construir \mathbf{W} y tan solo se requerirá extraer muestras de $\mathbf{U}^{(n)}$ y \mathbf{R} . La siguiente Definición presenta la función de densidad de \mathbf{R} para cualquier función generadora g .

Definición 1.1.2 (McNeil et al. (2015)). Si $\mathbf{W} \sim EC_n(0, I, g)$ entonces la función de densidad de \mathbf{R} es conocida y es

$$f_{\mathbf{R}}(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} g(r^2). \quad (1.8)$$

Más aún, \mathbf{R}^2 tiene la misma distribución que la forma cuadrática $\mathbf{W}'\mathbf{W} = \|\mathbf{W}\|^2$, donde $\|\cdot\|$ denota la norma euclidiana.

Dado lo anterior, extraer muestras de distribuciones de Contorno Elíptico, y en particular de la Distribución de Error Generalizado MGED es sencillo. Por ejemplo, si $\mathbf{W} \sim MGED_n(0, I, \kappa)$, entonces bajo 1.7 se tiene $\mathbf{W} \stackrel{d}{=} \mathbf{R}\mathbf{U}^{(n)}$ con \mathbf{R} distribuida de acuerdo a 1.8 como la 2κ -ésima raíz de una distribución Gamma con parámetro de forma $\alpha = n\kappa$ y parámetro de escala $\beta = \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{n}{2}}$ como se demuestra en el Apéndice A. Así,

$$\mathbf{R} \stackrel{d}{=} \sqrt[2\kappa]{\text{Gamma}\left(n\kappa, \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{n/2}\right)}. \quad (1.9)$$

Por su parte y siguiendo a Solaro (2004), el vector aleatorio $\mathbf{U}^{(n)}$, uniformemente distribuido sobre una superficie unitaria en \mathbb{R}^n puede ser obtenido a través de una distribución normal multivariada Z con componentes estandarizados, así

$$\mathbf{U}^{(n)} \stackrel{d}{=} \frac{Z}{\|Z\|}. \quad (1.10)$$

Generar estas muestras será importante para los procedimientos de simulación presentados más adelante. El algoritmo de extracción de muestras para la distribución MGED se encuentra en el Apéndice A.

1.2. Descripción del proceso MTAR

Siguiendo a Calderón & Nieto (2017), suponga que $\{Y_t\}$ y $\{X_t\}$ son procesos estocásticos tal que $Y_t = (Y_{1t}, \dots, Y_{nt})'$ y $X_t = (X_{1t}, \dots, X_{vt})'$, y sea $\{Z_t\}$ un proceso univariado. $\{Y_t\}$ sigue un modelo *MTAR* con variable de umbral Z_t si

$$Y_t = \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} Y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} X_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} Z_{t-i} + \Sigma_{(j)}^{1/2} \varepsilon_t, \quad r_{j-1} < Z_t \leq r_j \quad (1.11)$$

para algún $j = 1, \dots, l$, donde $l \geq 2$ es el número de regímenes, y $r_0, r_1, \dots, r_{l-1}, r_l$ son números reales tales que $-\infty = r_0 < r_1 < \dots < r_{l-1} < r_l = \infty$; estos son conocidos como umbrales y definen los regímenes del modelo. Para $j = 1, \dots, l$, los coeficientes $\phi_i^{(j)}$ para $i = 1, \dots, p_j$, $\beta_i^{(j)}$ para $i = 1, \dots, q_j$, $\delta_i^{(j)}$ para $i = 1, \dots, d_j$ y $\Sigma_{(j)}^{1/2}$ son matrices reales, conocidas como parámetros no estructurales. $p = \{p_1, \dots, p_l\}$, $q = \{q_1, \dots, q_l\}$ y $d = \{d_1, \dots, d_l\}$ son conjuntos de los números enteros no negativos p_j , q_j y d_j que denotan el máximo rezago asociado a los vectores Y_t , X_t y Z_t , respectivamente en cada régimen y son conocidos como ordenes autoregresivos. Los valores de umbral $\mathbf{r} = (r_1, \dots, r_{l-1})'$ junto

con el número de regímenes l y los ordenes autoregresivos son conocidos como parámetros estructurales del modelo MTAR.

Para cada $j = 1, \dots, l$ se define el vector de parámetros autoregresivos no estructurales como $\theta_j = \left(\phi_0^{(j)}, \phi_1^{(j)}, \dots, \phi_{p_j}^{(j)}, \beta_1^{(j)}, \dots, \beta_{q_j}^{(j)}, \delta_1^{(j)}, \dots, \delta_{d_j}^{(j)} \right)_{n \times \eta_j}$, donde $\eta_j = n \cdot p_j + v \cdot q_j + d_j + 1$.

Adicionalmente, el proceso de ruido *iid* n -dimensional $\{\varepsilon_t\}$ es tal que para cada t , ε_t sigue una Distribución de Error Generalizada Multivariada Estandarizada ($\varepsilon_t \sim MGED_n(\mathbf{0}, \mathbf{I}, \kappa)$) cuya función de densidad está dada por 1.1.

1.3. Función de verosimilitud

Dados el vector de parámetros $\Theta = (l, r, p, q, d, \theta, \Sigma, \kappa)$, con $\theta = (\theta_1, \dots, \theta_l)$ y $\Sigma = (\Sigma_{(1)}, \dots, \Sigma_{(l)})$; el vector $\mathbf{y} = (y_1, \dots, y_T)$; las variables exógenas $\mathbf{x} = (x_1, \dots, x_T)$ y la variable de umbral $\mathbf{z} = (z_1, \dots, z_T)$ se construye la función de verosimilitud condicional asociada. Así,

$$L(\Theta) = f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \Theta) = f(y_{s+1}|y_s, \dots, y_1, \mathbf{x}, \mathbf{z}, \Theta) \dots f(y_T|y_{T-1}, \dots, y_1, \mathbf{x}, \mathbf{z}, \Theta)$$

donde $s = \max(p, q, d)$.

Si $\{\varepsilon_t\} \sim iid$ y para t el vector $\varepsilon_t \sim MGED_n(\mathbf{0}, \mathbf{I}, \kappa)$ entonces $f_{\varepsilon_{s+1}, \varepsilon_{s+2}, \dots, \varepsilon_T} = \prod_{t=s+1}^T f_{\varepsilon_t}(\varepsilon_t)$. Se tiene entonces,

$$L(\Theta) = f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \Theta) = \prod_{j=1}^l \prod_{t=s+1}^T f_{\varepsilon_t}(e_t) \times |\mathbf{J}_j|$$

donde $e_t = \Sigma_{(j_t)}^{-1/2} \left[y_t - W_t \theta_j^{(t)} \right] = \Sigma_{(j_t)}^{-1/2} \left[y_t - \phi_0^{(j_t)} - \sum_{i=1}^{p_j} \phi_i^{(j_t)} y_{t-i} - \sum_{i=1}^{q_j} \beta_i^{(j_t)} x_{t-i} - \sum_{i=1}^{d_j} \delta_i^{(j_t)} z_{t-i} \right]$ para algún $j_t = j$ con $W_t = (1, y_{t-1}, \dots, y_{t-p_j}, x_{t-1}, \dots, x_{t-q_j}, z_{t-1}, \dots, z_{t-d_j})$ y $|\mathbf{J}_j|$ para $j = 1, \dots, l$ es el jacobiano asociado que está dado por

$$|\mathbf{J}_j| = \begin{vmatrix} \frac{\partial e_{s+1}}{\partial y_{s+1}} & \frac{\partial e_{s+1}}{\partial y_{s+2}} & \dots & \frac{\partial e_{s+1}}{\partial y_T} \\ \frac{\partial e_{s+2}}{\partial y_{s+1}} & \frac{\partial e_{s+2}}{\partial y_{s+2}} & \dots & \frac{\partial e_{s+2}}{\partial y_T} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_T}{\partial y_{s+1}} & \frac{\partial e_T}{\partial y_{s+2}} & \dots & \frac{\partial e_T}{\partial y_T} \end{vmatrix} = \prod_{t=s+1}^T \left| \Sigma_{(j_t)}^{-1/2} \right| = \left| \Sigma_{(j_t)}^{-\frac{1}{2}} \right|^{T-s}$$

donde $s = \max(p, q, d)$.

Finalmente, se tiene que la función de verosimilitud condicional es

$$\begin{aligned}
 f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \Theta) &= \prod_{j=1}^l \prod_{t=s+1}^T \frac{1}{\sqrt{\pi^n}} \frac{\Gamma(1 + \frac{n}{2})}{\Gamma(1 + n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{n}{2}} \exp \left\{ - \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{1}{2\kappa}} [e_t e_t']^{\frac{1}{2\kappa}} \right\} \\
 &\quad \times \left| \Sigma_{(j_t)}^{-1/2} \right|^{T-s} \\
 &= \prod_{j=1}^l \left[\frac{\left| \Sigma_{(j_t)}^{-\frac{1}{2}} \right|}{\sqrt{\pi^n}} \right]^{T-s} \left[\frac{\Gamma(1 + \frac{n}{2})}{\Gamma(1 + n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{n}{2}} \right]^{T-s} \times \\
 &\quad \exp \left\{ - \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{1}{2\kappa}} \sum_{t:j_t=j} [e_t e_t']^{\frac{1}{2\kappa}} \right\}. \tag{1.12}
 \end{aligned}$$

Generalidades de la estimación Bayesiana e implementación de los métodos MCMC

En este capítulo se presentan las generalidades de la estimación Bayesiana, las características de los algoritmos de obtención de muestras tradicionales, y las ventajas y propiedades que poseen los algoritmos adaptativos.

En la mayoría de los casos tratar la distribución a posteriori de los parámetros no es sencillo desde el punto de vista analítico. Esto implica que se debe hacer otro tipo de aproximación para trabajar con ella, por ejemplo, a través de la obtención de las funciones de densidad completas de los parámetros para que puedan ser estimados mediante algoritmos como el Muestreador de Gibbs. Además, cuando las distribuciones condicionales completas no corresponden a familias de distribuciones conocidas, se deberá implementar un algoritmo Metropolis-Hastings con el fin de simular muestras de dichas funciones.

Algoritmos como Metropolis-Hastings (Metropolis & Ulam, 1953; Hastings, 1970) requieren la elección de distribuciones propuestas y es bien conocido que algunas funcionan mejor que otras. Así pues, determinar cuál es la mejor distribución propuesta para una distribución objetivo particular es muy importante y muy difícil. Comúnmente este problema es tratado de una forma *ad hoc* que se refiere al mecanismo de intento y error (Rosenthal, 2008). Sin embargo, es posible hacer uso de algoritmos adaptativos para encontrar distribuciones propuestas adecuadas de forma automática.

2.1. Generalidades

A través de la regla de Bayes es posible obtener la distribución a posteriori, por ejemplo, para el parámetro θ dado los datos $Y = y$, así,

$$p(\theta|Y = y) = \frac{f(y|\theta)\pi(\theta)}{\int_{\Theta} f(y|\theta)\pi(\theta)d\theta}$$

donde el producto entre la función de verosimilitud condicional y la distribución previa, mientras el denominador denota la densidad marginal de Y y es considerada como la constante que permite que $p(\theta|Y = y)$ integre a uno. Por ello la a posteriori puede ser

reescrita como

$$p(\theta|Y = y) \propto f(y|\theta)\pi(\theta). \quad (2.1)$$

De esta manera, la distribución del parámetro será influenciada por una distribución previa $\pi(\theta)$ que resume todo el conocimiento previo que se tiene sobre el parámetro y la verosimilitud $f(y|\theta)$ obtenida para el modelo a través de los datos y .

Luego, para la estimación se define una función de penalización, que por ejemplo puede ser $c(\hat{\theta}) = E_{\theta|y}[\hat{\theta} - \theta]^2$ y que al minimizarse deja como resultado el estimador de θ que minimiza el error cuadrático medio. Además, éste es equivalente al valor esperado de la distribución a posteriori,

$$\hat{\theta} = E_{\theta|y}[\theta|y] = \int \theta p(\theta|Y = y) d\theta. \quad (2.2)$$

2.2. Métodos tradicionales

Existe una gran variedad de métodos MCMC disponibles. Algunos han surgido como mejoras o variaciones de otros. Dentro de los métodos MCMC encontramos el algoritmo de Metropolis-Hastings que explota la idea de que la condición de reversibilidad garantiza la existencia de una distribución invariante. Otro método es el Muestreador de Gibbs donde se considera la posibilidad de tomar muestras de las distribuciones condicionales completas.

Los elementos básicos de estos algoritmos se presentan a continuación.

2.2.1. Algoritmo Metropolis-Hastings

Suponga que se tiene una función de densidad de una distribución objetivo p con respecto a una medida de referencia (por ejemplo, una n -dimensional medida de Lebesgue). Entonces dado Θ_n , un valor propuesto Θ'_{n+1} es generado de una densidad previamente especificada $q(\Theta_n, \theta')$, y tiene probabilidad de ser aceptada dada por

$$\alpha(\theta, \theta') = \begin{cases} \min \left\{ \frac{p(\theta')}{p(\theta)} \frac{q(\theta, \theta')}{q(\theta', \theta)}, 1 \right\}, & \text{si } p(\theta)q(\theta, \theta') > 0 \\ 1, & \text{si } p(\theta)q(\theta, \theta') = 0 \end{cases}. \quad (2.3)$$

Si el valor propuesto es aceptado, se fija $\Theta_{n+1} = \Theta'_{n+1}$, en otro caso, $\Theta_{n+1} = \Theta_n$. La función $\alpha(\theta, \theta')$ es elegida para asegurar que la cadena de Márkov $\Theta_0, \Theta_1, \dots$ sea reversible respecto a la distribución objetivo $p(\theta')$, de manera que la distribución objetivo es estacionaria para la cadena. Si la distribución propuesta es simétrica, es decir, $q(\theta, \theta') = q(\theta', \theta)$, entonces 2.3 se reduce a

$$\alpha(\theta, \theta') = \begin{cases} \min \left\{ \frac{p(\theta')}{p(\theta)}, 1 \right\} & \text{si } p(\theta)q(\theta, \theta') > 0 \\ 1, & \text{si } p(\theta)q(\theta, \theta') = 0. \end{cases}$$

El algoritmo Metropolis-Hastings Caminata Aleatoria (RWM) es utilizado la mayor parte del tiempo. En este caso el valor propuesto está dado por $\Theta'_{n+1} = \Theta_n + \zeta_{n+1}$ donde los incrementos $\{\zeta_n\}$ son *i.i.d.* para alguna distribución dada (por ejemplo, $N(0, \sigma^2 I_n)$).

Su uso ha sido extendido a muchas aplicaciones, sin embargo, como se mencionó antes requiere de una “calibración” de σ haciéndolo un método de prueba y error.

Algoritmo M-H. El algoritmo genérico para la construcción de cadenas vía Metropolis-Hastings es el siguiente:

Sea $\Theta = (\Theta_n)_{n \geq 0}$ una cadena de Márkov de Metrópolis-Hastings con $q(\theta', \theta)$, $p(\cdot)$ y $\alpha(\theta', \theta)$ tales como se definieron previamente. El *algoritmo de Metrópolis-Hastings* está dado por:

1. Se escoge un punto de partida arbitrario $\theta_0 \in \mathbb{R}^n$ tal que $p(\theta_0) > 0$.
2. Se inicia en $j = 1$ y se repiten los siguientes pasos hasta que $j = N$, para N grande.
 - Se simula un valor θ' desde $q(\theta_{j-1}, \theta')$ y un número aleatorio u de $U(0, 1)$.
 - Si $u \leq \alpha(\theta_{j-1}, \theta')$, se fija $\theta_j = \theta'$. En caso contrario, $\theta_j = \theta_{j-1}$.
 - Se actualiza $j \leftarrow j + 1$.
3. Se seleccionan las últimas $M < N$ simulaciones que permiten aproximar $p(\cdot)$.

2.2.2. Muestreador de Gibbs

El Muestreador de Gibbs es un caso particular del algoritmo Metropolis-Hastings donde las distribuciones propuestas son las distribuciones condicionales completas $p(\theta|\cdot)$. De esta manera todas las muestras propuestas serán aceptadas en Gibbs, que implica que la probabilidad de aceptación es siempre uno ($\alpha(\theta, \theta') = 1$).

La principal ventaja del algoritmo es su fácil implementación. Sin embargo, existen varias limitantes. Primero, aún cuando se tiene la distribución $p(\theta)$ quizá no sea posible o práctico obtener las distribuciones condicionales. Segundo, aún si las distribuciones condicionales son encontradas analíticamente, ellas pueden no tener una forma conocida y por tanto no será sencillo obtener muestras de ellas. Por último, hay casos en los cuales el muestreador de Gibbs es ineficiente, por ejemplo, el “mixing” de las cadenas obtenidas puede ser muy lento y requerir un mayor número de iteraciones.

Algoritmo Gibbs. Sea $\Theta = (\Theta_n)_{n \geq 0}$ una cadena de Márkov de Metrópolis-Hastings con $p(\cdot)$ tal como se definió previamente cuyas densidades marginales y condicionales son $p(\theta)$ y $p(\theta|\Theta'_{-i} = \theta'_{-i})$ para $i = 1, \dots, n$ y $\Theta' \sim p$. Sea q igual a

$$q(\theta, \theta') = \begin{cases} \frac{1}{n} p_i(\theta'_i | \Theta'_{-i} = \theta_{-i}) & \text{si } \theta'_{-i} = \theta_{-i} \\ 0, & \text{si en otro caso} \end{cases}$$

1. Se escoge un punto de partida arbitrario $\theta_0 \in \mathbb{R}^n$ tal que $p(\theta_0) > 0$.
2. Se inicia en $j = 1$ y se repiten los siguientes pasos hasta que $j = N$, para N grande.
 - Se simula θ' de la densidad $p(\theta' | \Theta' = \theta_j)$.
 - Se fija $\theta_j = \theta'$ y $\theta_{-j} = \theta_{-(j-1)}$, para obtener $\Theta_j \in \mathbb{R}^n$.
 - Se actualiza $j \leftarrow j + 1$.
3. Se seleccionan las últimas $M < N$ simulaciones para eliminar posibles efectos de los valores iniciales. Además, se seleccionan secuencialmente los valores muestreados para eliminar posibles correlaciones.

2.3. Métodos Adaptativos

Como se mencionó, los métodos anteriormente descritos hacen uso de un mecanismo de prueba y error para ajustar la varianza en la selección de muestras. Este mecanismo es generalmente exitoso, pero es costoso e ineficiente, requiere de recursos humanos permanentemente y es bastante difícil de calibrar para altas dimensiones. La Figura 2.1 muestra tres cadenas distintas; la primera considera un valor de varianza pequeño, que permite obtener altas tasas de aceptación aunque mezclas deficientes; la segunda contempla un valor grande para la varianza cuyo resultado son bajas tasas de aceptación y una mezcla deficiente; y la tercera cadena utiliza una varianza mediana que permite conseguir tasas de aceptación óptimas y una mezcla adecuada¹.

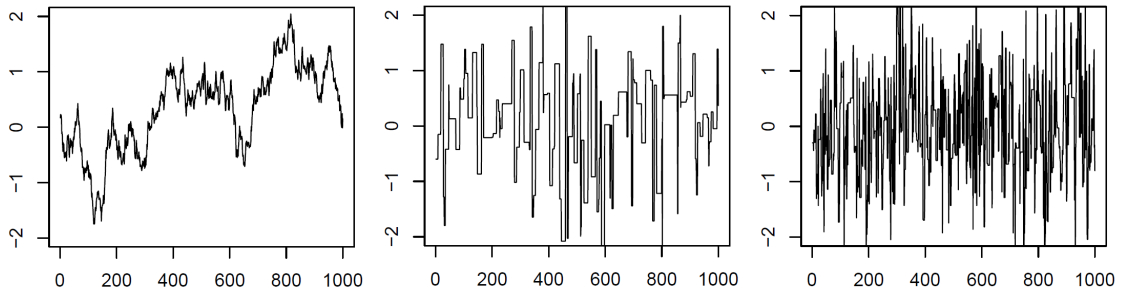


FIGURA 2.1. Cadenas obtenidas por un algoritmo tradicional para un valor de la varianza 1) pequeño, 2) grande y 3) mediano

Con el fin de facilitar la construcción de las cadenas y corregir los anteriores problemas, Haario et al. (2001), Rosenthal (2008) y Roberts & Rosenthal (2009) desarrollaron una alternativa conocida como métodos adaptativos, que le permiten al computador, durante la ejecución del algoritmo, “aprender” automáticamente acerca de los valores correctos de las varianzas y así construir cadenas de Márkov para los parámetros, reduciendo el tiempo requerido para la convergencia a su distribución estacionaria, con tasas de aceptación adecuadas y buenas mezclas.

¿Pero cómo saber cuándo la cadena consigue una mezcla adecuada? Una alternativa es mediante la visualización de una gráfica como antes. Otra alternativa es cuantificar las tasas de aceptación de las muestras propuestas. Así por ejemplo, una gran aproximación fue realizada en Roberts et al. (1997) en donde haciendo uso de un algoritmo Metropolis Caminata Aleatoria (RWM por sus siglas en inglés) con distribución de incremento propuesta $N(0, \sigma^2 I_n)$, demostraron que bajo densidades objetivo de la forma

$$p(\theta_1, \theta_2, \dots, \theta_n) = f(\theta_1)f(\theta_2) \dots f(\theta_n) \quad (2.4)$$

donde n representa la dimensión de la distribución objetivo, la tasa de aceptación óptima cuando $n \rightarrow \infty$ es de 0.234 y para $n = 1$ la tasa adecuada sería de 0.44. Adicionalmente, ellos observan que el algoritmo es relativamente eficiente cuando la tasa de aceptación asintótica se encuentra entre 0.1 y 0.6. Ahora, aun cuando la distribución objetivo no sea de la forma 2.4 esta clase es bastante robusta y las simulaciones parecen sugerir que estas tasas se mantienen para otros casos también (Rosenthal, 2008).

¹Una mezcla adecuada indica que la cadena recorrió el espacio muestral de posibles estados.

2.3.1. Principios

Como es conocido los algoritmos MCMC adaptativos no siempre preservan la estacionariedad de la distribución objetivo p . Sin embargo, si la adaptación es diseñada para satisfacer ciertas condiciones, entonces la estacionariedad es garantizada y alcanzada rápidamente.

Roberts & Rosenthal (2007) probaron la ergodicidad de los métodos adaptativos bajo condiciones simples y que no requieren que los parámetros adaptativos converjan. Suponga entonces que el algoritmo se actualiza de Θ_m a Θ_{m+1} usando la distribución propuesta o kernel P_{Γ_m} . Sea $\{P_\gamma\}_{\gamma \in \mathcal{Y}}$ una familia de cadenas de Márkov, cada una con la misma distribución estacionaria $p(\cdot)$, sea Γ_m el kernel elegido en la m -ésima iteración y A es el espacio paramétrico del Θ_m , entonces

$$P(\Theta_{m+1} \in A | \Theta_m = \theta, \Gamma_m = \gamma, \Theta_{m-1}, \dots, \Theta_0, \Gamma_{m-1}, \dots, \Gamma_0) = P_\gamma(\theta, A)$$

para $m = 1, 2, \dots$. Aquí las $\{\Gamma_m\}$ son actualizadas de acuerdo a algún algoritmo adaptativo. En principio la elección de Γ_m puede depender completamente de la historia $\Theta_{m-1}, \dots, \Theta_0, \Gamma_{m-1}, \dots, \Gamma_0$. En general el algoritmo es sencillo de implementar y no requiere mucha programación extra.

Ahora, se hace relevante fijar condiciones suficientes que permitan que $\{\Theta_m\}$ converja a p . Roberts & Rosenthal (2007) demostraron convergencia asintótica, es decir, que $\lim_{m \rightarrow \infty} \sup_{A \subseteq \mathcal{X}} \|P(\Theta_m \in A) - p(A)\| = 0$, así como la ley débil de los grandes números donde $\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m g(\Theta_i) = p(g)$ para todas las funciones acotadas $g : \mathcal{X} \rightarrow \mathcal{R}$, asumiendo solo la condición de adaptación decreciente

$$\lim_{m \rightarrow \infty} \sup_{\theta \in \mathcal{X}} \|P_{\Gamma_{m+1}}(\theta, \cdot) - P_{\Gamma_m}(\theta, \cdot)\| = 0 \quad \text{en probabilidad,} \quad (2.5)$$

y también la condición de contención o convergencia acotada

$$\{M_\epsilon(\Theta_m, \Gamma_m)\}_{m=0}^\infty \quad \text{está acotado en probabilidad con } \epsilon > 0, \quad (2.6)$$

donde $M_\epsilon(\theta, \gamma) = \inf\{m \geq 1 : \|P_\gamma^m(\theta, \cdot) - p(\cdot)\| \leq \epsilon\}$ es el tiempo de convergencia del kernel P_γ cuando se inicia en el estado $\theta \in \mathcal{X}$.

En la práctica la condición 2.6 es satisfecha para la mayoría de esquemas adaptativos y por tanto este requerimiento puede ser ampliamente ignorado (Rosenthal, 2008). Por ejemplo, para los algoritmos Metropolis Adaptativo la condición se mantiene para todas las densidades objetivo que sean cóncavas. En contraste, la condición 2.5 es más relevante. Esta requiere que la cantidad de adaptaciones en la m -ésima iteración vaya a cero cuando $m \rightarrow \infty$.

2.3.2. Metropolis Adaptativo (MA)

Considérese la versión del algoritmo Metropolis Adaptativo (MA) de Haario et al. (2001). Sea $p(\cdot)$ una distribución objetivo n -dimensional. La distribución propuesta

$Q_m(\theta, \cdot)$ bajo m iteraciones está dada por

$$Q_m(\theta, \cdot) = \begin{cases} N(\theta, (0.1)^2 I_n/n) & \text{para } 2n \geq m \\ (1 - \beta)N(\theta, (2.38)^2 \Sigma_m/n) + \beta N(\theta, \Sigma_0) & \text{para } 2n < m \end{cases} \quad (2.7)$$

para algún $0 < \beta < 1$ y alguna matriz dada no singular Σ_0 , donde Σ_m es la estimación empírica de la estructura de covarianza de la distribución objetivo, es decir, la covarianza de $\theta_0, \theta_1, \dots, \theta_m$. El objetivo de $\beta N(\theta, \Sigma_0)$ en la segunda componente de la mixtura de la ecuación 2.7 es asegurar que las covarianzas propuestas no colapsen a cero. De Roberts et al. (1997) se conoce que la distribución propuesta $N(\theta, (2.38)^2 \Sigma/n)$ es óptima en el contexto multidimensional. De este modo, $N(\theta, (2.38)^2 \Sigma_m/n)$ es un intento de aproximarse a esta.

El algoritmo MA es un algoritmo de selección de muestras válido. En primer lugar, los cambios en la estimación empírica en la m -ésima iteración están dados por $O(1/n)$ y por tanto se sigue que la condición de adaptación decreciente 2.5 es satisfecha. Segundo, la condición de contención (acotamiento) para $\beta > 0$ se cumple para una gran familia de densidades dentro de las cuales están todas aquellas que son log-cóncavas fuera de alguna región acotada (Rosenthal, 2008).

A través de un ejercicio de simulación, Roberts & Rosenthal (2009) validan el uso de este algoritmo. Inicialmente definieron la distribución objetivo $p(\cdot) = N(0, AA')$, donde $A_{n \times n}$ es una matriz generada aleatoriamente con $\{A_{ij}\}_{i,j=1}^n$ *i.i.d.* y distribuidas $N(0, 1)$. Así pues, la varianza será altamente errática y por tanto obtener muestras de $p(\cdot)$ se hará más difícil. Se obtuvieron las cadenas para $n = 100$ y $n = 200$ (Figura 2.2). En ambos casos le tomó a las cadenas bastante tiempo para adaptarse apropiadamente y conseguir una adecuada mezcla. Sin embargo, se debe considerar que fueron 5.050 parámetros para $n = 100$ y 20.100 parámetros cuando $n = 200$.

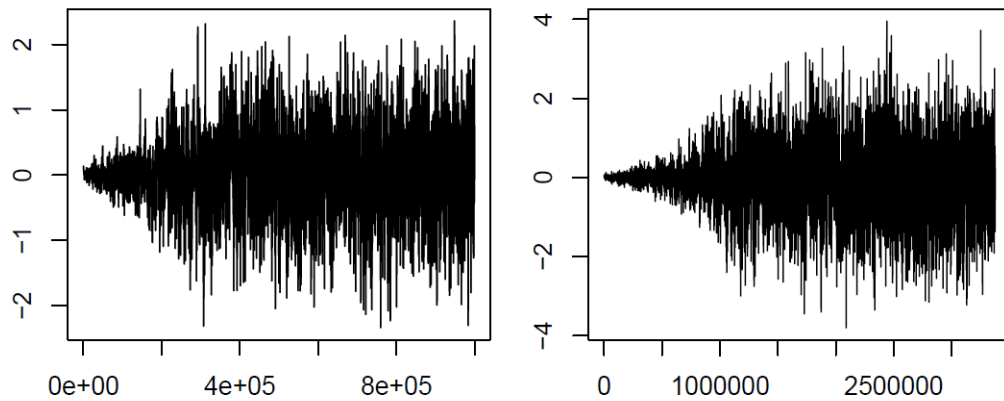


FIGURA 2.2. Primera coordenada de la cadena de Márkov AM para $n = 100$ y $n = 200$.

En las primeras iteraciones el algoritmo subestima la verdadera varianza estacionaria, mientras en las últimas iteraciones el algoritmo ha “aprendido” como obtener muestras de $p(\cdot)$.

2.3.3. Metropolis Adaptativo dentro de Gibbs

Es posible usar algoritmos adaptativos para obtener muestras de parámetros bajo cualquier partición. Sea $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_B)$ el vector de B parámetros o conjunto de ellos a estimar. Las distribuciones condicionales completas de cada grupo de parámetros vienen dadas por

$$\begin{aligned} p_1(\Theta^{(1)}|\Theta^{(2)}, \Theta^{(3)}, \dots, \Theta^{(B)}), \\ p_2(\Theta^{(2)}|\Theta^{(1)}, \Theta^{(3)}, \dots, \Theta^{(B)}), \\ \vdots \\ p_B(\Theta^{(B)}|\Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(B-1)}) \end{aligned}$$

Luego, se ejecuta un algoritmo tradicional de Gibbs en donde los valores propuestos para los parámetros provendrán de un algoritmo adaptativo, en particular un algoritmo MA. Por último, para asegurar la rápida convergencia de las cadenas se actualizará la varianza en baches de 500 iteraciones, decir, $\Sigma_m^{(i)} = V\left(\Theta_{m-1}^{(i)}, \Theta_{m-2}^{(i)}, \dots, \Theta_{m-500}^{(i)}\right)$ para un parámetro o grupo de parámetros i (Rosenthal, 2008).

Estimación de los parámetros

En este capítulo se expone la metodología Bayesiana propuesta para la estimación de los parámetros del modelo. Para ello, se presentan las distribuciones previas, las densidades condicionales completas requeridas para la estimación mediante algoritmos MCMC adaptativos, el mecanismo para la selección del mejor modelo (estimación de los parámetros estructurales) y el algoritmo propuesto para la estimación de los parámetros del modelo MTAR con MGED.

3.1. Distribuciones previas

A continuación, se presentan las distribuciones previas que fueron elegidas para la estimación Bayesiana de los parámetros del modelo MTAR.

Sea θ_j como se definió en la sección 1.2 para $j = 1, \dots, l$. La distribución previa de θ_j es una distribución normal multivariada, es decir $\theta_j \sim N(\theta_{0j}, V_{0j})$; con lo cual, la función de densidad es

$$\pi(\theta_j) = \frac{1}{(2\pi)^{\eta_j/2} |V_{0j}|^{1/2}} \exp \left\{ -\frac{1}{2} (\theta_j - \theta_{0j})' V_{0j}^{-1} (\theta_j - \theta_{0j}) \right\} \quad (3.1)$$

$$\propto \exp \left\{ -\frac{1}{2} (\theta_j - \theta_{0j})' V_{0j}^{-1} (\theta_j - \theta_{0j}) \right\} \quad (3.2)$$

donde $\eta_j = (p_j \times n + q_j \times v + d_j + 1) \times n$. La previa elegida concentra el conocimiento de los coeficientes alrededor de la media θ_{0j} con cierta incertidumbre cuantificada por su matriz de covarianza V_{0j} que para este trabajo será muy grande.

A su vez, la distribución previa de $\Sigma_{(j)}$ es una inversa Wishart con matriz de covarianza S_{0j} y ν grados de libertad. Su función de densidad será

$$\pi(\Sigma_{(j)}) = \frac{|S_{0j}|^{\nu/2}}{2^{(\nu n)/2} \Gamma_n(\frac{\nu}{2})} |\Sigma_{(j)}|^{-(n+\nu+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left(\Sigma_{(j)}^{-1} S_{0j} \right) \right\} \quad (3.3)$$

$$\propto |\Sigma_{(j)}|^{-(n+\nu+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left(\Sigma_{(j)}^{-1} S_{0j} \right) \right\}. \quad (3.4)$$

Adicionalmente, κ se asociará con una distribución previa $Gamma(\alpha, \beta)$, cuya forma funcional está dada por

$$\pi(\kappa) = \frac{1}{\Gamma(\alpha)\gamma^\alpha} \kappa^{\alpha-1} \exp\left\{-\frac{\kappa}{\gamma}\right\} \quad (3.5)$$

$$\propto \kappa^{\alpha-1} \exp\left\{-\frac{\kappa}{\gamma}\right\}. \quad (3.6)$$

Finalmente, \mathbf{r} tendrá una distribución previa uniforme en el intervalo $(P_{0.2}^z, P_{0.8}^z)$,

$$\pi(\mathbf{r}) = I_{\{\mathbf{r} \geq P_{0.2}^z, \mathbf{r} \leq P_{0.8}^z\}} \quad (3.7)$$

donde $P_{0.2}^z$ y $P_{0.8}^z$ es el percentil 20 y 80 de la variable \mathbf{z} , respectivamente.

3.2. Estimación de los parámetros no estructurales: distribuciones condicionales completas

Asumiendo que a priori existe independencia de los parámetros dentro de cada régimen y entre regímenes, y que además los parámetros estructurales son conocidos. La distribución a posteriori condicional a p, q, d y l , construida a partir de la equivalencia de Bayes, está dada por

$$\begin{aligned} p(\theta, \Sigma, \kappa, \mathbf{r} | \mathbf{y}, \mathbf{x}, \mathbf{z}, p, q, d, l) &= f(\mathbf{y} | \mathbf{x}, \mathbf{z}, \Theta) \pi(\theta, \Sigma, \kappa, \mathbf{r}) \\ &\propto f(\mathbf{y} | \mathbf{x}, \mathbf{z}, \Theta) \pi(\theta) \pi(\Sigma) \pi(\kappa) \pi(\mathbf{r}) \end{aligned}$$

con p, q, d, θ y Σ como fueron definidos en las secciones 1.2 y 1.3. La complejidad de la función a posteriori condicional resultante impide que se extraigan muestras de ella con facilidad. Por ello, se procede a obtener las funciones de densidad condicionales completas de cada parámetro, para luego, conseguir muestras de los parámetros mediante los métodos MCMC adaptativos elegidos en el capítulo 2.

3.2.1. Coeficientes autoregresivos (θ_j)

Con base en la función de verosimilitud 1.12 y la distribución previa 3.3 es posible encontrar la distribución condicional de θ_j para $j = 1, \dots, l$. Condicionando a los parámetros estructurales y a los demás parámetros no estructurales, así como las realizaciones de los procesos $\{Y_t\}$, $\{X_t\}$ y $\{Z_t\}$ se tiene

$$p(\theta_j | \theta_i, i \neq j, \Sigma, \kappa, p, q, d, l, \mathbf{r}, \mathbf{y}, \mathbf{x}, \mathbf{z}) \propto f(\mathbf{y} | \mathbf{x}, \mathbf{z}, \Theta) \pi(\theta_j)$$

De manera que eliminando las cantidades constantes de la función de verosimilitud se tiene que la función de densidad condicional completa para θ_j es

$$p(\theta_j|\theta_i, i \neq j, \Sigma, \kappa, p, q, d, l, \mathbf{r}, \mathbf{y}, \mathbf{x}, \mathbf{z}) \propto \exp \left\{ - \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{1}{2\kappa}} \sum_{t:j_t=j} [e_t e_t']^{\frac{1}{2\kappa}} \right\} \times \exp \left\{ -\frac{1}{2}(\theta_j)' V_{0j}^{-1}(\theta_j) \right\} \quad (3.8)$$

donde $e_t = \Sigma_{(j_t)}^{-1/2} [y_t - W_t \theta_j'] = \Sigma_{(j_t)}^{-1/2} [y_t - \phi_0^{(j_t)} - \sum_{i=1}^{p_j} \phi_i^{(j_t)} y_{t-i} - \sum_{i=1}^{q_j} \beta_i^{(j_t)} x_{t-i} - \sum_{i=1}^{d_j} \delta_i^{(j_t)} z_{t-i}]$ para algún $j_t = j$ con $W_t = (1, y_{t-1}, \dots, y_{t-p_j}, x_{t-1}, \dots, x_{t-q_j}, z_{t-1}, \dots, z_{t-d_j})$, $\theta_{0j} = 0_n$ y $V_{0j} = 100I_n$.

3.2.2. Matriz de covarianza ($\Sigma_{(j)}$)

Al igual que antes, al eliminar las cantidades constantes que no afectan a $\Sigma_{(j)}$ en la función de verosimilitud, y dados los parámetros estructurales y demás parámetros no estructurales, así como las realizaciones de los procesos $\{Y_t\}$, $\{X_t\}$ y $\{Z_t\}$ se tiene que la función de densidad condicional completa para $\Sigma_{(j)}$ es

$$p(\Sigma_{(j)}|\Sigma_{(i)}, i \neq j, \theta, \kappa, p, q, d, l, \mathbf{r}, \mathbf{y}, \mathbf{x}, \mathbf{z}) \propto f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \Theta) \pi(\Sigma_{(j)}) \propto \frac{1}{|\Sigma_{(j)}^{-1/2}|^{T-s}} \exp \left\{ - \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{1}{2\kappa}} \sum_{t:j_t=j} [e_t e_t']^{\frac{1}{2\kappa}} \right\} \times |\Sigma_{(j)}|^{(n+\nu+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} \left(\Sigma_{(j)}^{-1} S_{0j} \right) \right\} \quad (3.9)$$

donde $e_t = \Sigma_{(j_t)}^{-1/2} [y_t - \phi_0^{(j_t)} - \sum_{i=1}^{p_j} \phi_i^{(j_t)} y_{t-i} - \sum_{i=1}^{q_j} \beta_i^{(j_t)} x_{t-i} - \sum_{i=1}^{d_j} \delta_i^{(j_t)} z_{t-i}]$ para algún $j_t = j$, S_{0j} es la matriz de covarianzas de la previa y ν son los grados de libertad de la distribución.

3.2.3. Parámetro de no normalidad (κ)

La distribución condicional completa de κ dados los demás parámetros no estructurales y estructurales, así como las realizaciones de los procesos $\{Y_t\}$, $\{X_t\}$ y $\{Z_t\}$ está dada por

$$p(\kappa|\theta, \Sigma, p, q, d, l, \mathbf{r}, \mathbf{y}, \mathbf{x}, \mathbf{z}) \propto f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \Theta) \pi(\kappa) \propto \left[\frac{\Gamma(1 + \frac{n}{2})}{\Gamma(1 + n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{n}{2}} \right]^{T-s} \exp \left\{ - \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{1}{2\kappa}} \sum_{t:j_t=j} [e_t e_t']^{\frac{1}{2\kappa}} \right\} \times \kappa^{\alpha-1} \exp \left\{ -\frac{\kappa}{\gamma} \right\} \quad (3.10)$$

donde $e_t = \Sigma_{(j_t)}^{-1/2} [y_t - \phi_0^{(j_t)} - \sum_{i=1}^{p_j} \phi_i^{(j_t)} y_{t-i} - \sum_{i=1}^{q_j} \beta_i^{(j_t)} x_{t-i} - \sum_{i=1}^{d_j} \delta_i^{(j_t)} z_{t-i}]$ para algún $j_t = j$, α y γ son los parámetros de forma y escala asociados a la distribución previa, respectivamente.

3.3. Estimación de los parámetros estructurales

3.3.1. Parámetro de umbral \mathbf{r}

Dado el dominio continuo del parámetro \mathbf{r} , su estimación se hará conjuntamente con los parámetros no estructurales. La distribución condicional completa del vector de umbral \mathbf{r} dados los demás parámetros estructurales y no estructurales, así como las realizaciones de los procesos $\{Y_t\}$, $\{X_t\}$ y $\{Z_t\}$ está dada por

$$\begin{aligned}
 p(\mathbf{r}|\theta, \Sigma, \kappa, p, q, d, l, \mathbf{y}, \mathbf{x}, \mathbf{z}) &\propto f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \Theta)\pi(\mathbf{r}) \\
 &\propto \left[\frac{\Gamma(1 + \frac{n}{2})}{\Gamma(1 + n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{n}{2}} \right]^{T-s} \exp \left\{ - \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)} \right]^{\frac{1}{2\kappa}} \sum_{t:j_t=j} [e_t e_t']^{\frac{1}{2\kappa}} \right\} \\
 &\times I_{\{r \geq P_{0.2}^z, r \leq P_{0.8}^z\}} \tag{3.11}
 \end{aligned}$$

donde $e_t = \Sigma_{(j_t)}^{-1/2} \left[y_t - \phi_0^{(j_t)} - \sum_{i=1}^{p_j} \phi_i^{(j_t)} y_{t-i} - \sum_{i=1}^{q_j} \beta_i^{(j_t)} x_{t-i} - \sum_{i=1}^{d_j} \delta_i^{(j_t)} z_{t-i} \right]$ para algún $j_t = j$ y $I_{\{r \geq P_{0.2}^z, r \leq P_{0.8}^z\}}$ es la distribución previa de \mathbf{r} definida en la sección 3.1.

3.3.2. Elección del mejor modelo: estimación de los parámetros l, p, q y d

Seguendo a Chib (1995), cuyo método mantiene la característica inherente de flexibilidad ante problemas de baja y alta dimensionalidad, es posible estimar los parámetros l, p, q y d a través de una comparación de modelos bajo la construcción de la función de verosimilitud marginal. Dado un conjunto finito de modelos $\{\mathcal{M}_1, \dots, \mathcal{M}_k\}$, suponga que cada modelo es caracterizado por un vector de parámetros específico $\Theta_i \in S_i \subseteq \mathbb{R}^{k_i}$ de dimensión k_i y densidad $f(\mathbf{y}|\mathcal{M}_i, \Theta_i)$. La función de verosimilitud marginal, equivalente a la constante de normalización de la distribución a posteriori, puede ser escrita como

$$\begin{aligned}
 m(\mathbf{y}|\mathcal{M}_i) &= \int f(\mathbf{y}|\mathcal{M}_i, \Theta_i)\pi(\Theta_i|\mathcal{M}_i)d\Theta_i \\
 &= \frac{f(\mathbf{y}|\mathcal{M}_i, \Theta_i)\pi(\Theta_i|\mathcal{M}_i)}{p(\Theta_i|\mathbf{y}, \mathcal{M}_i)}.
 \end{aligned}$$

Evaluando el lado derecho de la anterior identidad en algún punto apropiado Θ^* y tomando logaritmos se obtiene

$$\log m(\mathbf{y}|\mathcal{M}_i) = \log f(\mathbf{y}|\mathcal{M}_i, \Theta^*) + \log \pi(\Theta^*|\mathcal{M}_i) - \log p(\Theta^*|\mathbf{y}, \mathcal{M}_i) \tag{3.12}$$

de donde la verosimilitud marginal puede ser calculada encontrando una estimación de la ordenada a posteriori, es decir, su cálculo se reduce a una estimación de la densidad a posteriori en un punto Θ^* , esto dado que $f(\mathbf{y}|\mathcal{M}_i, \Theta^*)$ y $\pi(\Theta^*|\mathcal{M}_i)$ son conocidos.

Suponga ahora que el espacio de parámetros es particionado en B bloques o grupos, es decir, $\Theta^* = (\Theta_1^*, \dots, \Theta_B^*)$, con $\Theta_i^* \in S_i \subseteq \mathbb{R}^{k_i}$ donde el indicador del modelo \mathcal{M}_i será suprimido por conveniencia. Luego, por ley de la probabilidad total se tiene que

$$p(\Theta^*|\mathbf{y}) = p(\Theta_1^*|\mathbf{y})p(\Theta_2^*|\Theta_1^*, \mathbf{y}) \dots p(\Theta_B^*|\Theta_1^*, \dots, \Theta_{B-1}^*, \mathbf{y}) \tag{3.13}$$

donde $p(\Theta_1^*|\mathbf{y})$ es la densidad marginal de Θ_1 y $p(\Theta_B^*|\Theta_1^*, \dots, \Theta_{B-1}^*, \mathbf{y})$ es la ordenada de la densidad condicional completa. Aquí $p(\Theta_1^*|\mathbf{y})$ no puede ser estimada por el método Rao-Blackwell debido a que la constante de normalización es desconocida (Chib, 1995; Chib & Jeliazkov, 2001).

De modo que es posible escribir la ordenada a posteriori en Θ^* como $p(\Theta_1^*, \dots, \Theta_B^*|\mathbf{y}) = \prod_{i=1}^B p(\Theta_i^*|\mathbf{y}, \Theta_1^*, \dots, \Theta_{i-1}^*)$. Adicionalmente, se pueden conseguir muestras de la densidad condicional completa $p(\Theta_i|\mathbf{y}, \Theta_{-i}) \propto f(\mathbf{y}|\Theta)\pi(\Theta)$, $i = 1, \dots, B$, mediante un algoritmo M-H con densidad propuesta $q(\Theta_i, \Theta'_i|\mathbf{y}, \Psi_{i-1}, \Psi^{i+1})$ para la transición entre Θ a Θ' , y probabilidad de aceptación dada por

$$\alpha(\Theta_i, \Theta'_i|\mathbf{y}, \Psi_{i-1}, \Psi^{i+1}) = \min \left\{ 1, \frac{f(\mathbf{y}|\Theta'_i, \Psi_{i-1}, \Psi^{i+1})\pi(\Theta'_i, \Theta_{-i})}{f(\mathbf{y}|\Theta_i, \Psi_{i-1}, \Psi^{i+1})\pi(\Theta_i, \Theta_{-i})} \times \frac{q(\Theta'_i, \Theta_i|\mathbf{y}, \Psi_{i-1}, \Psi^{i+1})}{q(\Theta_i, \Theta'_i|\mathbf{y}, \Psi_{i-1}, \Psi^{i+1})} \right\}$$

donde $\Psi_{i-1} = (\Theta_1, \dots, \Theta_{i-1})$ denota los bloques anteriores a i , $\Psi^{i+1} = (\Theta_{i+1}, \dots, \Theta_B)$ representa los grupos mas allá de i y $\Theta_{-i} = (\Psi_{i-1}, \Psi^{i+1})$.

Finalmente, Chib & Jeliazkov (2001) diseñan un algoritmo para la estimación de $p(\Theta_i^*|\mathbf{y}, \Theta_1^*, \dots, \Theta_{i-1}^*)$ mediante ejecuciones MCMC con algoritmos tradicionales, con lo cual es posible calcular la verosimilitud marginal. El algoritmo consta de 4 pasos:

1. Fijar $\Psi_{i-1} = \Psi_{i-1}^*$ y obtener muestras del conjunto reducido de distribuciones condicionales completas $p(\Theta_k|\mathbf{y}, \Theta_{-k})$, $k = i, \dots, B$. Sean $\{\Theta_i^{(g)}, \dots, \Theta_B^{(g)}\}$, $g = 1, \dots, M$ las muestras generadas.
2. Incluya Θ_i^* en el conjunto condicional. Sea $\Psi_{i-1} = (\Psi_{i-1}^*, \Theta_i^*)$ y retire la distribución condicional completa de Θ_i del conjunto del paso 1. Entonces obtenga las muestras de las distribuciones restantes $p(\Theta_k|\mathbf{y}, \Theta_{-k})$, $k = i+1, \dots, B$, para producir $\{\Theta_{i+1}^{(j)}, \dots, \Theta_B^{(j)}\}$. En cada paso del proceso de muestreo también obtenga $\Theta_i^{(j)}$ de $q(\Theta_i^*, \Theta_i|\mathbf{y}, \Psi_{i-1}^*, \Psi^{i+1, (j)})$.
3. Estime $p(\Theta_i^*|\mathbf{y}, \Theta_1^*, \dots, \Theta_{i-1}^*)$ a través de la razón de promedios de Monte Carlo como

$$\hat{p}(\Theta_i^*|\mathbf{y}, \Theta_1^*, \dots, \Theta_{i-1}^*) = \frac{M^{-1} \sum_{k=1}^M \alpha(\Theta_i^{(g)}, \Theta_i^*|\mathbf{y}, \Psi_{i-1}^*, \Psi^{i+1, (g)}) q(\Theta_i^{(g)}, \Theta_i^*|\mathbf{y}, \Psi_{i-1}^*, \Psi^{i+1, (g)})}{J^{-1} \sum_{j=1}^J \alpha(\Theta_i^*, \Theta_i^{(j)}|\mathbf{y}, \Psi_{i-1}^*, \Psi^{i+1, (j)})}$$

donde el promedio del denominador puede incluir ceros si se obtienen valores de $\Theta_i^{(j)}$ que caigan fuera del dominio S de la distribución a posteriori.

4. Estime la log-verosimilitud marginal 3.12 como

$$\log \hat{m}(\mathbf{y}) = \log f(\mathbf{y}|\Theta^*) + \log \pi(\Theta^*) - \sum_{i=1}^B \log \hat{p}(\Theta_i^*|\mathbf{y}, \Theta_1^*, \dots, \Theta_{i-1}^*) \quad (3.14)$$

Para el caso del modelo MTAR descrito en la sección 1.2 se tiene que $\Theta_{y,l}^* = (\mathbf{r}_1^*, \dots, \mathbf{r}_{l-1}^*, \theta_1^*, \dots, \theta_l^*, \Sigma_{(1)}^*, \dots, \Sigma_{(l)}^*, \kappa^*)$ con θ_j , $j = 1, \dots, l$ como fue definido en el capítulo 1.2. Además, las variables exógenas \mathbf{x} y de umbral \mathbf{z} no afectan el algoritmo previo debido a que están dadas. De esta manera la log-verosimilitud marginal del modelo

MTAR propuesto evaluada en $\Theta_{y,l}^*$ es

$$\log \hat{m}(\mathbf{y}) = \log f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \Theta^*) + \sum_{i=1}^B \log \pi(\Theta_i^*) - \sum_{i=1}^B \log \hat{p}(\Theta_i^*|\mathbf{y}, \mathbf{x}, \mathbf{z}, \Theta_1^*, \dots, \Theta_{i-1}^*) \quad (3.15)$$

donde $\log f(\mathbf{y}|\Theta^*)$ es el logaritmo de la función verosimilitud condicional 1.12 evaluada en Θ^* , $\sum_{i=1}^B \log \pi(\Theta_i^*)$ es la sumatoria del logaritmo de las respectivas distribuciones previas de los parámetros de la partición Θ^* y $\sum_{i=1}^B \log \hat{p}(\Theta_i^*|\mathbf{y}, \mathbf{x}, \mathbf{z}, \Theta_1^*, \dots, \Theta_{i-1}^*)$ está dado por

$$\begin{aligned} & \sum_{i=1}^B \log \hat{p}(\Theta_i^*|\mathbf{y}, \mathbf{x}, \mathbf{z}, \Theta_1^*, \dots, \Theta_{i-1}^*) = \\ & \log \hat{p}(\mathbf{r}_1^*|\mathbf{y}, \mathbf{x}, \mathbf{z}) + \log \hat{p}(\mathbf{r}_2^*|\mathbf{r}_1^*, \mathbf{y}, \mathbf{x}, \mathbf{z}) + \dots + \log \hat{p}(\mathbf{r}_l^*|\mathbf{r}_1^*, \dots, \mathbf{r}_{l-1}^*, \mathbf{y}, \mathbf{x}, \mathbf{z}) + \\ & \log \hat{p}(\theta_1^*|\mathbf{r}_1^*, \dots, \mathbf{r}_l, \mathbf{y}, \mathbf{x}, \mathbf{z}) + \dots + \log \hat{p}(\theta_l^*|\mathbf{r}_1^*, \dots, \mathbf{r}_l, \theta_1^*, \dots, \theta_{l-1}^*, \mathbf{y}, \mathbf{x}, \mathbf{z}) + \\ & \log \hat{p}(\Sigma_{(1)}^*|\mathbf{r}_1^*, \dots, \mathbf{r}_l, \theta_1^*, \dots, \theta_l, \mathbf{y}, \mathbf{x}, \mathbf{z}) + \dots + \\ & \log \hat{p}(\Sigma_{(l)}^*|\mathbf{r}_1^*, \dots, \mathbf{r}_l, \theta_1^*, \dots, \theta_l, \Sigma_{(1)}, \dots, \Sigma_{(l-1)}, \mathbf{y}, \mathbf{x}, \mathbf{z}) + \\ & \log \hat{p}(\kappa|\mathbf{r}_1^*, \dots, \mathbf{r}_l, \theta_1^*, \dots, \theta_l, \Sigma_{(1)}, \dots, \Sigma_{(l)}, \mathbf{y}, \mathbf{x}, \mathbf{z}) \end{aligned} \quad (3.16)$$

3.4. Algoritmo para la estimación de los parámetros

Para $p = (p_1, \dots, p_l)$, $q = (q_1, \dots, q_l)$, $d = (d_1, \dots, d_l)$ y l fijos se repite el siguiente procedimiento.

1. En $t = 0$ se asignan valores iniciales θ_j^0 , $\Sigma_{(j)}^0$, κ^0 y \mathbf{r}^0 a los parámetros θ_j , $\Sigma_{(j)}$, κ y \mathbf{r} para cada $j = 1, \dots, l$.
2. Se repite el siguiente procedimiento para cada $t = 1, \dots, T$ y $j = 1, \dots, l$ el número de veces deseado.
 - Se simula un valor para el coeficiente autoregresivo θ_j^t desde la función de densidad condicional completa $p(\theta_j|\theta_i, i \neq j, \Sigma, \kappa, p, q, d, l, \mathbf{r}, \mathbf{y}, \mathbf{x}, \mathbf{z})$ utilizando el algoritmo adaptativo propuesto.
 - Se simula un valor de la matriz de covarianza $\Sigma_{(j)}^t$ desde la función de densidad condicional completa $p(\Sigma_{(j)}|\Sigma_{(i)}, i \neq j, \theta, \kappa, p, q, d, l, \mathbf{r}, \mathbf{y}, \mathbf{x}, \mathbf{z})$ utilizando el algoritmo adaptativo propuesto.
 - Se simula un valor del parámetro de no normalidad κ^t desde la función de densidad condicional completa $p(\kappa|\theta, \Sigma, p, q, d, l, \mathbf{r}, \mathbf{y}, \mathbf{x}, \mathbf{z})$ utilizando el algoritmo adaptativo propuesto.
 - Se simula un valor del parámetro de umbral \mathbf{r}^t desde la función de densidad condicional completa $p(\mathbf{r}|\theta, \Sigma, \kappa, p, q, d, l, \mathbf{y}, \mathbf{x}, \mathbf{z})$ utilizando el algoritmo adaptativo propuesto.
3. Para eliminar el efecto de los valores iniciales se eliminan las primera T-N simulaciones (burn-in) y se reducen posibles autocorrelaciones (thinning) de las cadenas resultantes.

Cuando se haya conseguido estimaciones de los parámetros θ_j , $\Sigma_{(j)}$, κ y \mathbf{r} para $j = 1, \dots, l$; y para cada combinación p , q , d y l se aplicará la metodología para elección del mejor modelo descrita 3.3.2.

Algunos de los parámetros necesarios para inicializar el método adaptativo serán descritos en el siguiente capítulo.

CAPÍTULO 4

Simulaciones

Para comprobar que la metodología propuesta es adecuada un conjunto de modelos TAR multivariados con Distribución de Error Generalizada fueron simulados. A continuación, se enuncian los supuestos, resultados y diagnósticos para los procesos construidos cuando 1) el parámetro de no normalidad de la distribución de error es igual a 0.1 (colas menos pesadas que una normal) y 2) el parámetro de no normalidad es igual a 1 (colas más pesadas que una normal). En los apéndices se presenta un ejercicio equivalente para 4 modelos más.

Las estimaciones se realizaron a través de la construcción de cadenas de Márkov mediante algoritmos MCMC adaptativos descritos en el capítulo 2. Durante el proceso de simulación se pudieron definir valores tentativos, según el valor de κ y el número de variables n del proceso, para el número de iteraciones o tamaño de las cadenas adecuado τN (sin considerar la etapa de calentamiento o *burn-in*) y el valor del parámetro α , descrito en la Tabla 4.1, que define la cota inferior para la varianza de las distribuciones propuestas en el algoritmo. Adicionalmente, se recomienda que la etapa de calentamiento no tenga una longitud inferior a 5000 (promedio de convergencia de κ).

κ	τ			α
	$n = 2$	$n = 3$	$n = 4$	
$> 0 - \leq 1$	1	1.1	1.3	0.1
$> 1 - \leq 4$	1.5	1.7	2	0.05
$> 4 - \leq 8$	2.5	2.5	2.5	0.01
> 8	3	3	3	0.001

TABLA 4.1. Valores propuestos de τ y α para el algoritmo de obtención de muestras según valores de κ y n

4.1. Modelo MTAR con MGED con $\kappa = 0.1$

Se simuló el siguiente proceso TAR bivariado $\{Y_t\}_{t=1}^T$ de tamaño $T = 1000$

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)}Y_{t-1} + \phi_2^{(1)}Y_{t-2} + \beta_1^{(1)}X_{t-1} + \Sigma_{(1)}^{1/2}\varepsilon_t & ; Z_{t-1} \leq r \\ \phi_0^{(2)} + \phi_1^{(2)}Y_{t-1} + \delta_1^{(2)}Z_{t-1} + \Sigma_{(2)}^{1/2}\varepsilon_t & ; Z_{t-1} > r \end{cases} \quad (4.1)$$

donde

$$\begin{aligned} \phi_0^{(1)} &= \begin{pmatrix} 0 \\ -0.4 \end{pmatrix}, \quad \phi_1^{(1)} = \begin{pmatrix} -0.1 & -0.8 \\ 0.4 & -0.5 \end{pmatrix}, \quad \phi_2^{(1)} = \begin{pmatrix} -0.4 & 0.2 \\ -0.4 & -0.1 \end{pmatrix}, \quad \beta_1^{(1)} = \begin{pmatrix} 0.3 \\ 0.1 \end{pmatrix}, \\ \Sigma_{(1)} &= \begin{pmatrix} 0.16 & -0.3 \\ -0.3 & 1.44 \end{pmatrix} \\ \phi_0^{(2)} &= \begin{pmatrix} -0.1 \\ -0.9 \end{pmatrix}, \quad \phi_1^{(2)} = \begin{pmatrix} -0.4 & 0.6 \\ -0.2 & 0.7 \end{pmatrix}, \quad \delta_1^{(2)} = \begin{pmatrix} 0.2 \\ 0.5 \end{pmatrix}, \quad \Sigma_{(2)} = \begin{pmatrix} 0.64 & 0.4 \\ 0.4 & 0.64 \end{pmatrix} \end{aligned}$$

con $r = \hat{z}_{0.5}$ que es el percentil 50 de \mathbf{z} y $\varepsilon_t \sim MGED_2(\mathbf{0}, \mathbf{I}, \kappa = 0.1)$. Las variables regresoras y de umbral fueron obtenidas mediante distribuciones *iid* $N(0, 1)$. Se puede observar en la Figura 4.1 una realización del proceso MTAR arriba descrito.

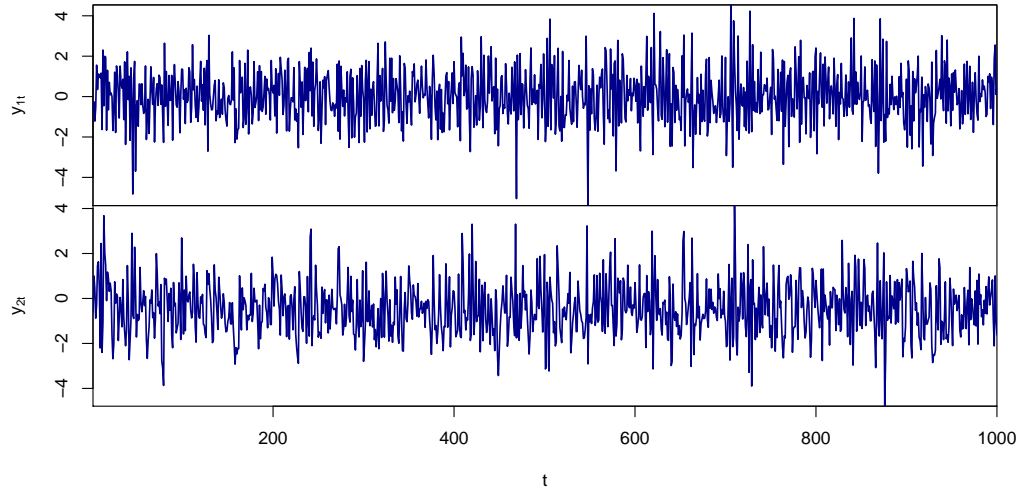


FIGURA 4.1. Serie simulada del proceso descrito en 4.1

4.1.1. Resultados de la estimación

Se procede a obtener muestras de las densidades condicionales completas encontradas en el capítulo 3. Teniendo como referencia la Tabla 4.1 se construyeron las cadenas de parámetros generando 15000 iteraciones ($1.5N$) y utilizando un valor $\alpha = 0.1$. De igual manera, estas cadenas sufrieron una etapa de calentamiento (burn-in) para finalmente conseguir una longitud de $N = 10000$. Este ejercicio se repitió 100 veces buscando alcanzar

valores de sesgos promedios bajos y altas tasas de inclusión o cobertura de los parámetros en los intervalos creíbles. Estos resultados se reportan en la Tabla C.1 del Apéndice C.

4.1.1.1. Estimación de los parámetros no estructurales

Como se mencionó antes, mediante el uso de los algoritmos MCMC adaptativos se obtuvieron muestras de los parámetros. La Figura 4.2 muestra las cadenas de un conjunto de parámetros para una realización particular. Estas cadenas presentan un comportamiento estable que preliminarmente confirma su convergencia. La convergencia hacia la distribución estacionaria es alcanzada aproximadamente en la iteración 2000 para los parámetros autoregresivos y la varianza. Para el parámetro de no normalidad κ se alcanza en la iteración 5000.

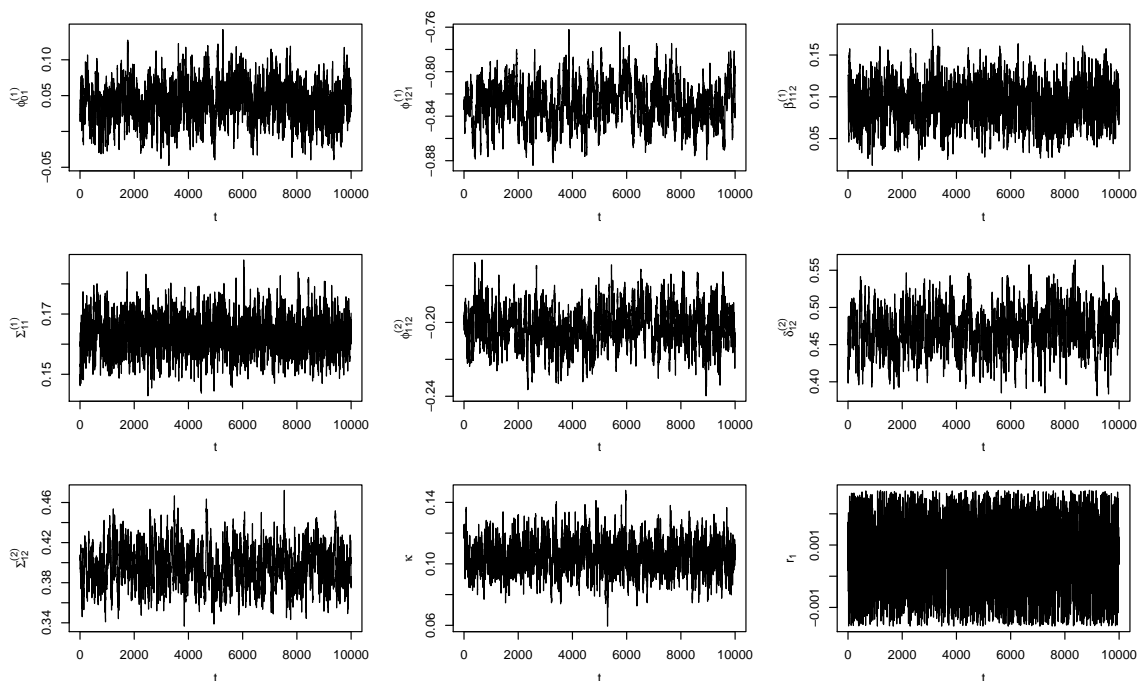


FIGURA 4.2. Valores de las cadenas para algunos parámetros del modelo 4.1

A su vez considerando la media aritmética como un estimador de los parámetros se obtiene la Tabla 4.2 donde se reportan las estimaciones puntuales, los intervalos de credibilidad asociados y las probabilidades de aceptación estimadas de las muestras propuestas. De allí puede verse que las estimaciones son cercanas a los valores verdaderos de los parámetros (que además se encuentran contenidos dentro de los intervalos de credibilidad) y que se tuvieron tasas de aceptación adecuadas.

Adicionalmente, se realizaron las pruebas de convergencia enunciadas en el Apéndice C.2 donde no se rechaza la convergencia a su distribución estacionaria.

4.1.1.2. Estimación de los parámetros estructurales

A través de la verosimilitud marginal descrita en el numeral 3.3.2 se elegirá el mejor modelo identificando los parámetros estructurales adecuados. La Tabla 4.3 muestra el

Parámetro	Régimen 1			Régimen 2		
	Estimación	IC 95 %	% acept.	Estimación	IC 95 %	% acept.
ϕ_{01}	0.0399 (0.0)	(-0.004,0.084)	38.8 %	-0.1009 (-0.1)	(-0.168,-0.035)	26.3 %
ϕ_{02}	-0.4515 (-0.4)	(-0.499,-0.406)		-0.8702 (-0.9)	(-0.917,-0.825)	
ϕ_{111}	-0.1207 (-0.1)	(-0.146,-0.095)	49.0 %	-0.4001 (-0.4)	(-0.425,-0.375)	53.1 %
ϕ_{112}	0.4333 (0.4)	(0.405,0.460)		-0.2026 (-0.2)	(-0.221,-0.185)	
ϕ_{121}	-0.8286 (-0.8)	(-0.861,-0.798)		0.5926 (0.6)	(0.561,0.623)	
ϕ_{122}	-0.4946 (-0.5)	(-0.528,-0.463)		0.7077 (0.7)	(0.685,0.731)	
ϕ_{211}	-0.4174 (-0.4)	(-0.451,-0.380)	48.6 %			
ϕ_{212}	-0.3658 (-0.4)	(-0.404,-0.331)				
ϕ_{221}	0.2259 (0.2)	(0.196,0.258)				
ϕ_{222}	-0.1168 (-0.1)	(-0.151,-0.084)				
β_{111}	0.3040 (0.3)	(0.266,0.340)	41.4 %			
β_{112}	0.0921 (0.1)	(0.053,0.135)				
δ_{11}				0.2053 (0.2)	(0.134,0.275)	45.8 %
δ_{12}				0.4692 (0.5)	(0.422,0.516)	
Σ_{11}	0.163 (0.16)	(0.153,0.173)	45.2 %	0.643 (0.64)	(0.603,0.685)	40.4 %
Σ_{12}	-0.308 (-0.3)	(-0.323,-0.295)		0.396 (0.4)	(0.364,0.430)	
Σ_{22}	1.460 (1.44)	(1.369,1.556)		0.638 (0.64)	(0.586,0.690)	
κ	0.1041 (0.1)	(0.088,0.122)	57.6 %			
r	0.0006 (0.0)	(-0.001,0.002)	32.7 %			

TABLA 4.2. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo 4.1 con $\kappa = 0.1$ y $r = \hat{z}_{0.5} = 0$. Los valores de los parámetros se presentan entre paréntesis en las columnas de estimación.

conjunto de modelos evaluados con su respectiva log verosimilitud marginal estimada. Se encuentra que el modelo 3 es el mejor y coincide con el simulado.

# Modelo	Parámetros				Rezago Z_t	$\log(\hat{m})$
	l	p	q	d		
1	2	(2,2)	(0,1)	(0,0)	0	1402.97
2	2	(2,2)	(0,1)	(0,0)	1	1320.16
3	2	(2,1)	(1,0)	(0,1)	1	1566.23
4	2	(1,1)	(1,0)	(0,1)	1	1502.12
5	3	(2,2,2)	(0,1,0)	(0,0,0)	0	101.15
6	3	(2,2,2)	(0,1,0)	(0,0,0)	1	106.45
7	3	(2,1,1)	(1,0,0)	(0,1,0)	0	325.21
8	3	(1,1,1)	(1,0,0)	(0,1,0)	1	52.69

TABLA 4.3. Log verosimilitud marginal estimada para el modelo 4.1

4.1.2. Diagnóstico del modelo

El diagnóstico es realizado a través de las pruebas CUSUM y CUSUMSQ para la bondad de ajuste y la caracterización de los residuales marginales.

1. Pruebas CUSUM y CUSUMSQ.

La Figura 4.3 muestran a un nivel del 95 % un buen ajuste del modelo. A través de los gráficos CUSUM y CUSUMSQ se puede observar que la especificación del modelo es adecuada y que los residuales no presentan heterocedasticidad. Es de destacar que los residuales presentan colas muy livianas, parecen ser estacionarios y no correlacionados.

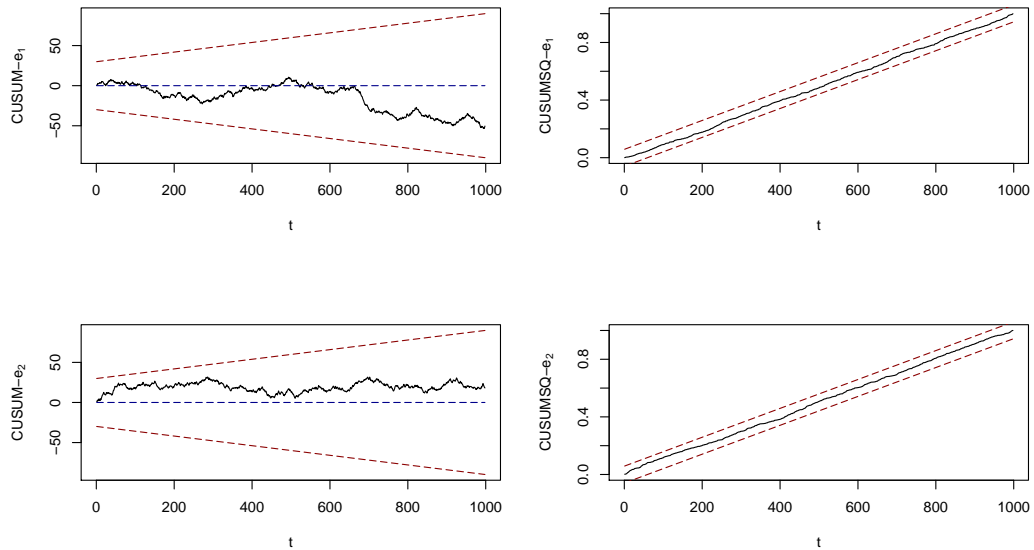


FIGURA 4.3. Pruebas CUSUM y CUSUMSQ para los residuos marginales del modelo 4.1

2. Residuales del modelo.

Por definición el kernel asociado a los residuales del modelo debería converger a una distribución uniforme. La Figura 4.4 muestra una aproximación de ello.

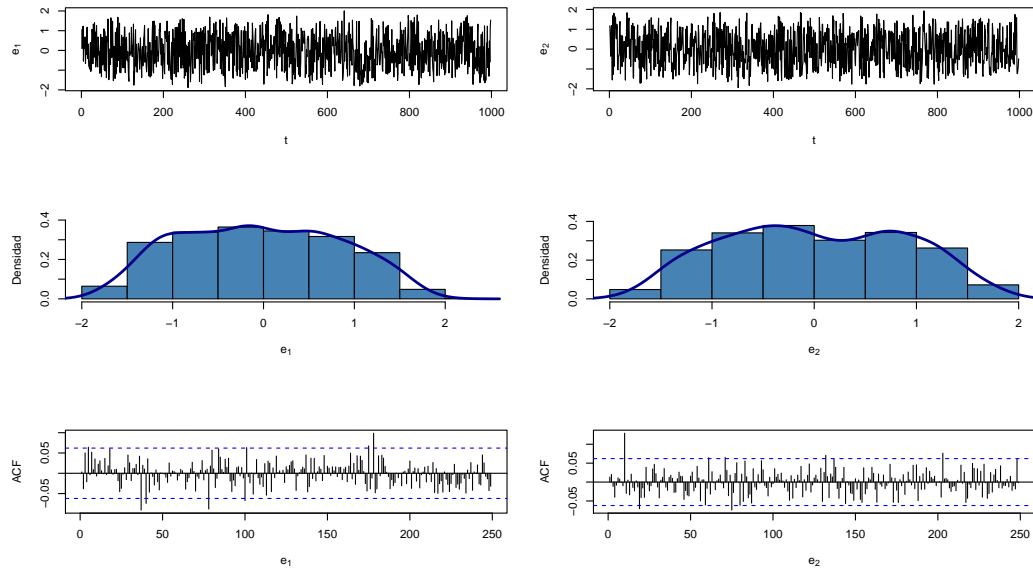


FIGURA 4.4. Residuales marginales del modelo 4.1

4.2. Modelo MTAR con MGED con $\kappa = 1$

También se simuló un proceso TAR bivariado con la misma estructura que 4.1 en donde $r = \hat{z}_{0.5}$ es el percentil 50 de \mathbf{z} y $\varepsilon_t \sim MGED_2(\mathbf{0}, \mathbf{I}, \kappa = 1)$, es decir, ε_t se distribuye Laplace bivariada. Al igual que en la simulación anterior, las variables exógenas y de umbral fueron obtenidas mediante distribuciones *iid* $N(0, 1)$. La Figura 4.5 muestra la realización de este proceso (4.2). Note que el espacio muestral de las variables y_{1t} y y_{2t} es mayor que el caso donde $\kappa = 0.1$

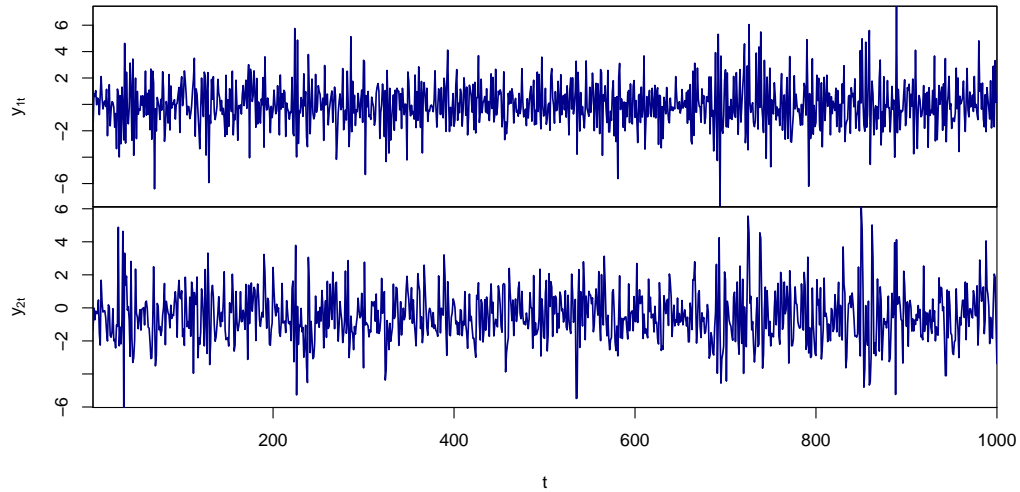


FIGURA 4.5. Serie simulada del proceso 4.2

4.2.1. Resultados de la estimación

Como antes se construyen cadenas con muestras provenientes de las densidades condicionales completas propuestas. Partiendo de la Tabla 4.1 se realizaron de 15000 iteraciones por parámetro ($1.5N$) y utilizando un valor $\alpha = 0.1$. Como antes, estas cadenas sufrieron una etapa de calentamiento (*burn-in*) para finalmente tener una longitud $N = 10000$. Se ejecutó el algoritmo 100 veces.

4.2.1.1. Estimación de los parámetros no estructurales

A través del algoritmo adaptativo se obtuvieron muestras de los parámetros. La Figura 4.6 presenta las cadenas de Márkov construidas. La estabilidad y convergencia hacia la distribución estacionaria es alcanzada aproximadamente en la iteración 3000 para los parámetros autoregresivos y la varianza. Para el parámetro de no normalidad κ se alcanza en la iteración 5000.

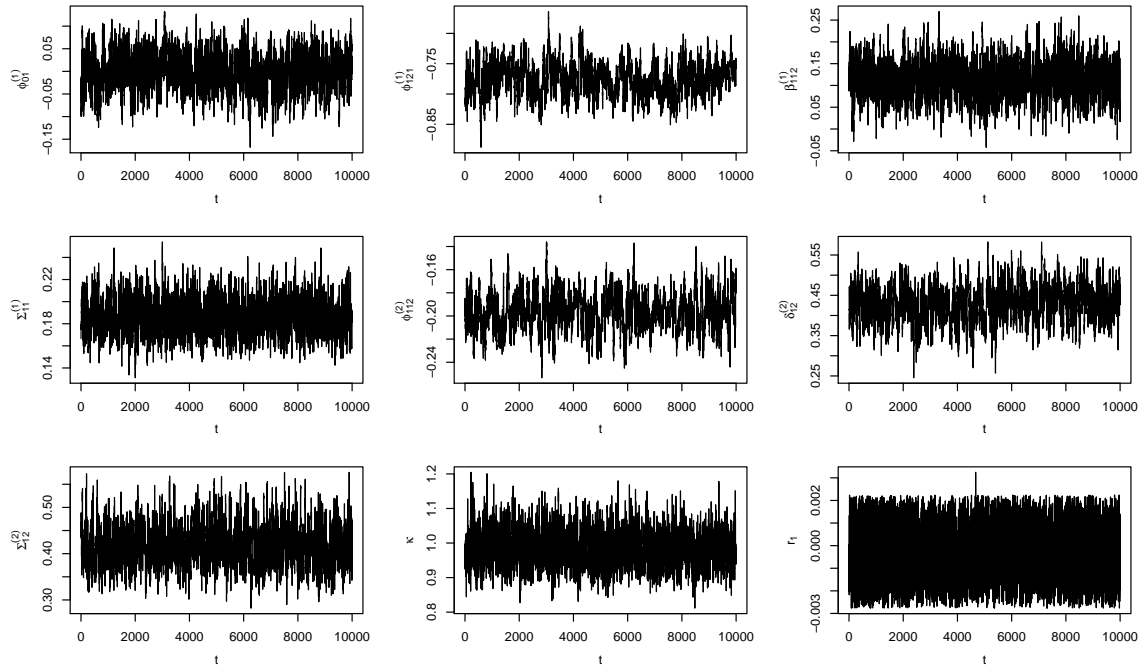


FIGURA 4.6. Valores de las cadenas para algunos parámetros del modelo 4.2

A su vez considerando la media aritmética como un estimador de los parámetros no estructurales se obtiene la Tabla 4.2 donde se reportan las estimaciones puntuales, los intervalos de credibilidad asociados y las probabilidades de aceptación de muestras propuestas. De allí puede verse que las estimaciones son cercanas a los valores verdaderos de los parámetros (además estos caen dentro los intervalos de credibilidad) y que se tuvieron tasas de aceptación adecuadas.

Parámetros	Régimen 1			Régimen 2		
	Estimación	IC 95 %	% acept.	Estimación	IC 95 %	% acept.
ϕ_{01}	-0.0002 (0.0)	(-0.074,0.067)	29.2 %	-0.0924 (-0.1)	(-0.207,0.015)	29.6 %
ϕ_{02}	-0.3843 (-0.4)	(-0.460,-0.303)		-0.8067 (-0.9)	(-0.876,-0.736)	
ϕ_{111}	-0.1469 (-0.1)	(-0.187,-0.099)	42.7 %	-0.3958 (-0.4)	(-0.435,-0.357)	49.7 %
ϕ_{112}	0.4200 (0.4)	(0.364,0.465)		-0.1960 (-0.2)	(-0.223,-0.169)	
ϕ_{121}	-0.7761 (-0.8)	(-0.820,-0.732)		0.6065 (0.6)	(0.56,0.653)	
ϕ_{122}	-0.5163 (-0.5)	(-0.567,-0.467)		0.6894 (0.7)	(0.659,0.717)	
ϕ_{211}	-0.4270 (-0.4)	(-0.471,-0.386)	39.5 %			
ϕ_{212}	-0.3471 (-0.4)	(-0.394,-0.298)				
ϕ_{221}	0.1918 (0.2)	(0.150,0.234)				
ϕ_{222}	-0.1248 (-0.1)	(-0.172,-0.078)				
β_{111}	0.2807 (0.3)	(0.218,0.341)	35.5 %			
β_{112}	0.1125 (0.1)	(0.040,0.183)				
δ_{11}				0.1453 (0.2)	(0.033,0.253)	30.2 %
δ_{12}				0.4284 (0.5)	(0.355,0.500)	
Σ_{11}	0.187 (0.16)	(0.161,0.214)	46.6 %	0.669 (0.64)	(0.581,0.765)	45.8 %
Σ_{12}	-0.328 (-0.3)	(-0.370,-0.287)		0.418 (0.4)	(0.346,0.497)	
Σ_{22}	1.536 (1.44)	(1.313,1.779)		0.655 (0.64)	(0.549,0.781)	
κ	0.9829 (1.0)	(0.899,1.082)	49.9 %			
r	-0.0003 (0.0)	(-0.002,0.002)	35.7 %			

TABLA 4.4. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo 4.2 con $\kappa = 1$ y $r = \hat{z}_{0.5} = 0$. Los valores de los parámetros se presentan entre paréntesis en las columnas de estimación.

Adicionalmente, se realizaron las pruebas de convergencia enunciadas en el Apéndice C.2 donde no se rechaza la convergencia a su distribución estacionaria.

4.2.1.2. Estimación de los parámetros estructurales

Al igual que antes mediante la verosimilitud marginal se escogerá el mejor modelo asociado al proceso. La Tabla 4.5 muestra el conjunto de modelos evaluados con su respectiva log verosimilitud marginal estimada. Se observa que los mejores modelos son el 3 y el 4, siendo el modelo 3 el simulado.

# Modelo	Parámetros				Rezago Z_t	$\log(\hat{m})$
	l	p	q	d		
1	2	(2,2)	(0,1)	(0,0)	0	-2230.23
2	2	(2,2)	(0,1)	(0,0)	1	-2215.43
3	2	(2,1)	(1,0)	(0,1)	1	-2198.25
4	2	(1,1)	(1,0)	(0,1)	1	-2200.22
5	3	(2,2,2)	(0,1,0)	(0,0,0)	0	-3132.33
6	3	(2,2,2)	(0,1,0)	(0,0,0)	1	-3115.26
7	3	(2,1,1)	(1,0,0)	(0,1,0)	0	-3098.25
8	3	(1,1,1)	(1,0,0)	(0,1,0)	1	-3095.21

TABLA 4.5. Log verosimilitud marginal estimada para el modelo 4.2

4.2.2. Diagnóstico del modelo

El diagnóstico es realizado a través de las pruebas CUSUM y CUSUMSQ para la bondad de ajuste y la caracterización de los residuales marginales.

1. Pruebas CUSUM y CUSUMSQ.

La Figura 4.7 muestran a un nivel del 95 % un buen ajuste del modelo. A través de los gráficos CUSUM y CUSUMSQ que los residuales no presentan heterocedasticidad. Es de destacar que los residuales presentan colas más pesada que la normal, son aparentemente estacionarios y no correlacionados.

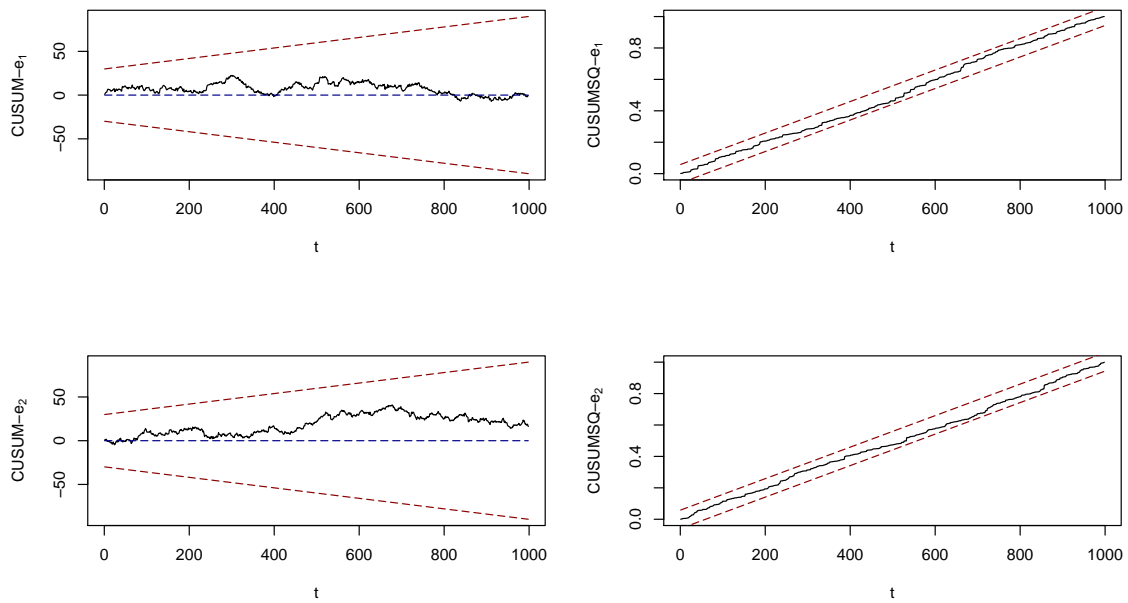


FIGURA 4.7. Pruebas CUSUM y CUSUMSQ para los residuos marginales del modelo 4.2

2. Residuales marginales del modelo.

Por definición el kernel asociado a los residuales del modelo mantienen una distribución de colas más pesadas que una normal (Laplace o doble exponencial). La Figura 4.8 da muestra de ello.

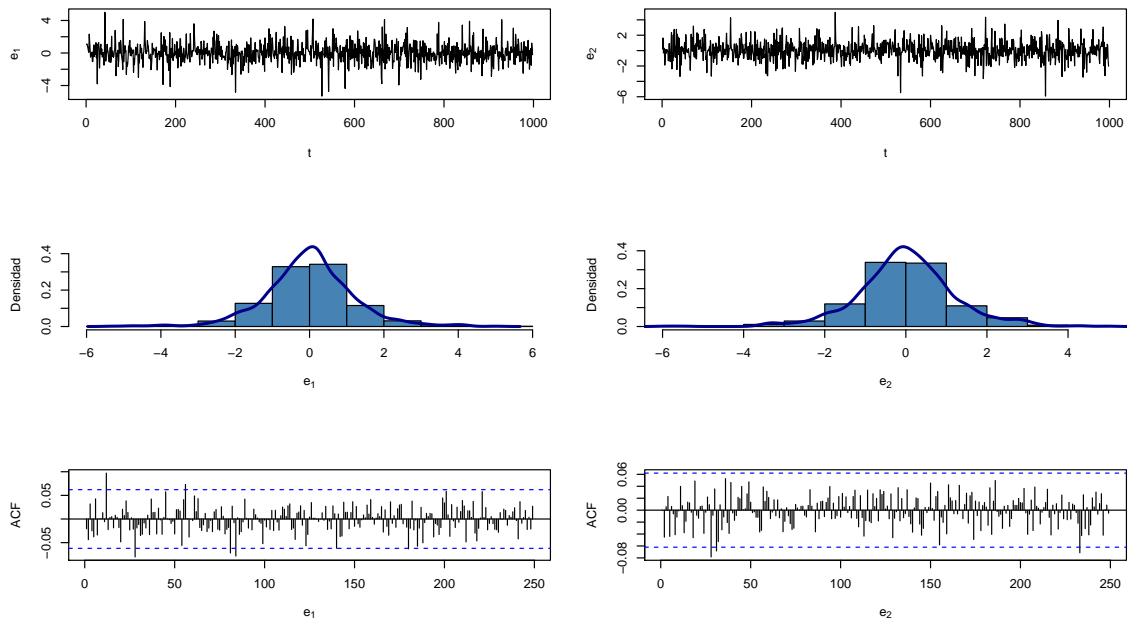


FIGURA 4.8. Residuales del modelo 4.2

Adicional a los modelos anteriormente presentados se simularon cuatro modelos más. Estos resultados se encuentran en el Apéndice B y consideran distintos números de variables, diferente cantidad de regímenes y otros valores de κ distintos a los arriba mencionados.

Aplicación Empírica

En esta sección se aplica la metodología propuesta a datos financieros, en particular, sobre índices accionarios bursátiles de los mercados de capitales de Colombia ($Y_{1t} = COLCAP$), Brasil ($Y_{2t} = BOVESPA$) y Estados Unidos ($Z_t = S\&P500$). El principal objetivo es modelar la influencia del índice S&P500 (Z_t) sobre los índices COLCAP (Y_{1t}) y BOVESPA (Y_{2t}).

5.1. Hechos estilizados de las series financieras

En la práctica las series de tiempo financieras parecen compartir cierto número de características, a menudo llamadas *hechos estilizados*. Autores como Malmsten & Teräsvirta (2004), Tsay (2005) y Melo & Becerra (2006) describen con más detalle estas características y presentan metodologías para el modelamiento de dichos procesos. Los *hechos estilizados* más comunes son:

1. Los precios tienen propiedades de caminatas aleatorias. Así pues estos son realizaciones de un proceso $P_t = P_{t-1} + \varepsilon_t$ donde $E(\varepsilon_t) = 0$. Sin supuestos adicionales, los precios pueden verse como martingalas donde se tendrá ausencia de correlaciones en los retornos, será imposible predecir precios futuros (inexistencia de arbitraje) y se tendrá una formación eficiente de precios¹.
2. Las series de retornos presentan colas pesadas (leptocurtosis) relacionadas principalmente a la alta frecuencia de los procesos. Este hecho ha sido un importante objeto de estudio en la *Teoría de Valores Extremos (EVT)*.
3. Se presentan nichos de volatilidad. Esto puede ser validado a través de las autocorrelaciones para cualquier medida de volatilidad, por ejemplo, el valor absoluto de los retornos al cuadrado ($|r_t|^2$).
4. Existe una alta correlación entre la volatilidad de los precios del activo y los volúmenes de los activos transados. Por ejemplo, activos de baja liquidez son usualmente más volátiles en precio.

¹Situación de equilibrio en la que el precio de mercado de un título o valor constituye una buena estimación de su precio teórico o intrínseco.

Para el caso multivariado aparece el concepto de contagio financiero, que no es otra cosa que el efecto que tiene un choque de un país o mercado sobre otro (Allen & Gale, 2000; Loaiza-Maya et al., 2015). En este contexto, el modelo MTAR con MGED propuesto puede aplicarse para modelar la media de los índices anteriormente mencionados vinculando el efecto que tiene la bolsa estadounidense sobre el mercado emergente latinoamericano.

5.2. Datos

Como se ha mencionado antes, el objetivo de este trabajo es modelar las relaciones entre los índices COLCAP y BOVESPA a través del modelo propuesto donde la variable de umbral es el índice S&P500. Las series diarias de índices utilizadas (Figura 5.1) fueron obtenidas del portal financiero *Investing.com* para la ventana de tiempo del 04 de enero de 2011 al 31 de agosto de 2018 y son descritas continuación.

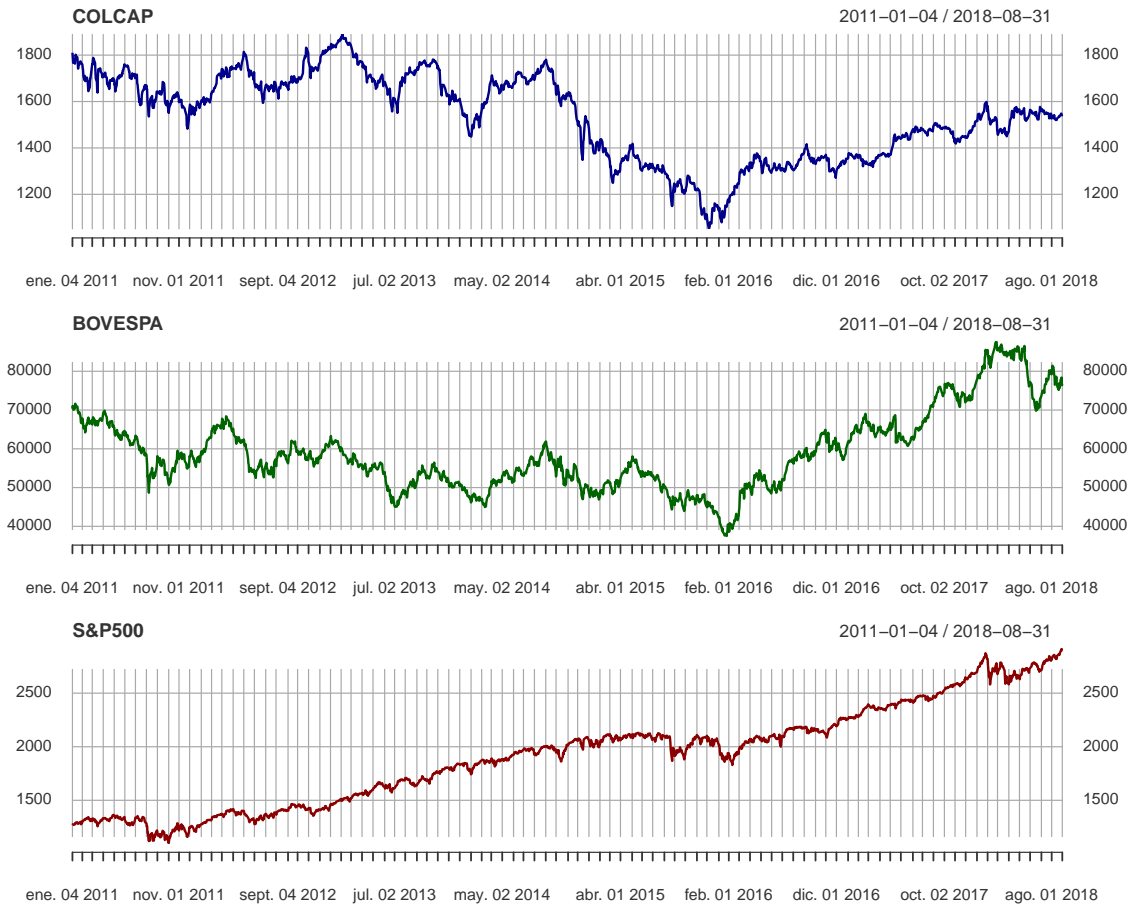


FIGURA 5.1. Índices accionarios COLCAP, BOVESPA Y S&P500

El índice COLCAP, vigente desde 2007, representa los cambios de precio de las 20 acciones más líquidas que se tranzan en el mercado de valores de Colombia, donde la capitalización bursátil ajustada determina la participación de cada valor en el índice. En la ventana de observación el índice alcanzó su máximo en febrero de 2013 y su mínimo en diciembre de 2015. Su peor desempeño se presentó entre 2014 y 2015 debido a la

desaceleración de la economía colombiana dada la caída de los precios de petróleo y el fenómeno de El Niño.

Por otro lado, el índice BOVESPA, vigente desde 1968, se compone de las 50 empresas más líquidas que se cotizan en la bolsa de San Pablo y que representan aproximadamente del 70 % de su capitalización bursátil. Es el índice más importante de Brasil y es un buen indicador del estado de su economía. En la ventana de tiempo analizada alcanza su mínimo en enero de 2016 y su máximo en febrero de 2018. Su continua subida desde 2016 se debe a la constante apreciación de su moneda, el real, y a su política monetaria contractiva sustentada en la reducción de los niveles de inflación.

El índice bursátil S&P500, se compone de las 500 empresas más grandes y líquidas de Estados Unidos ponderadas de acuerdo a su capitalización bursátil. Este índice expresa mejor la situación real del mercado accionario y de bonos, por lo que es más sensible que los índices Dow Jones y NASDAQ100.

La Figura 5.2 muestra las series de retornos de cada índice que serán utilizados en el proceso de estimación del modelo propuesto.

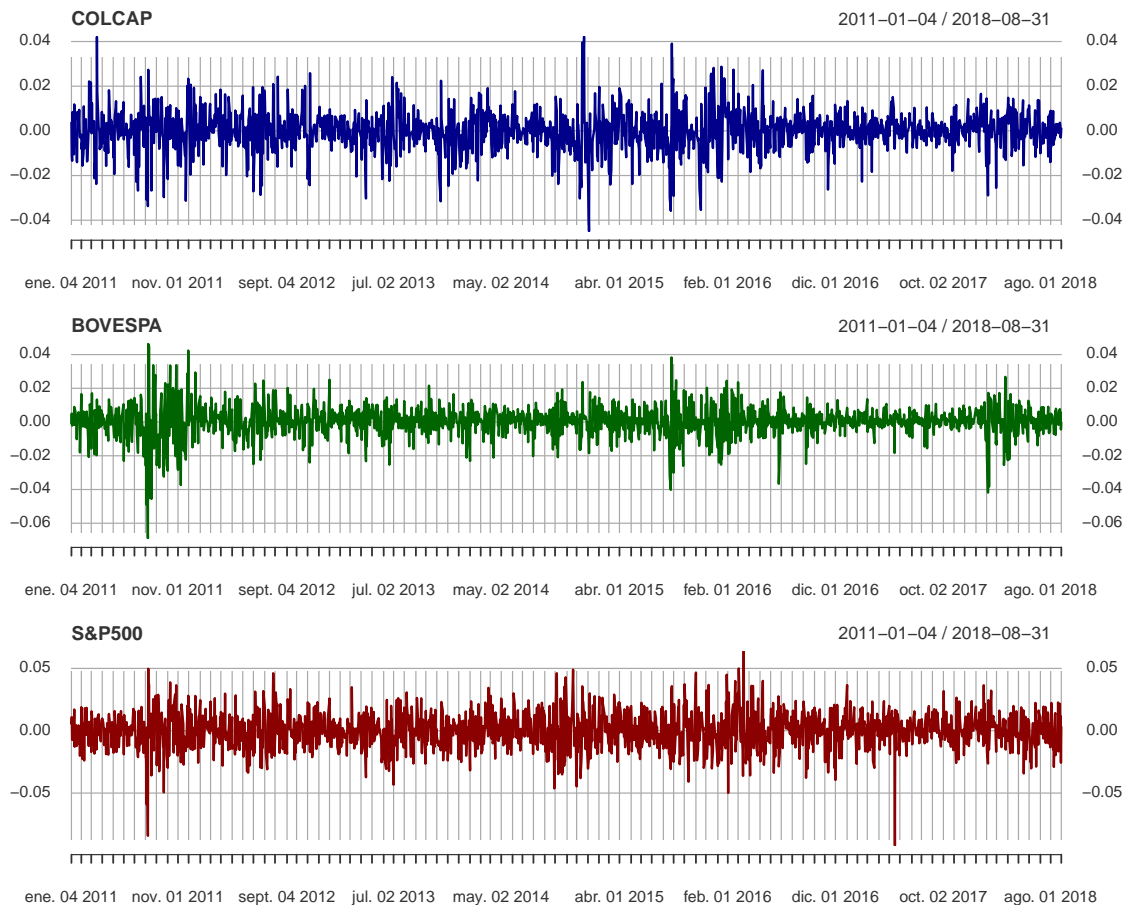


FIGURA 5.2. Retornos de los índices accionarios COLCAP, BOVESPA y S&P500

Inicialmente se requiere confirmar si el proceso es no lineal de tipo MTAR. Para esto se hace uso de la metodología propuesta por Lo & Zivot (2001), cuya prueba permite validar si las series siguen un modelo VAR (H_0) o si por el contrario se siguen a un proceso no

lineal de tipo MTAR con l regímenes. La distribución de la prueba no es estándar debido a su construcción donde depende de las matrices de covarianza de los residuales de ambos modelos (VAR y MTAR). Por ello, requiere el uso de procedimientos bootstrap en la obtención de los p-valores. La prueba es realizada para distintos rezagos considerando la variable de umbral en términos contemporáneos, de donde se tiene

Rezagos	$l = 2$		$l = 3$	
	LR Test	p valor	LR Test	p valor
1	40.11	0.00	58.70	0.00
2	45.03	0.00	71.58	0.00
3	52.87	0.00	109.28	0.00
4	75.50	0.00	120.67	0.00
5	103.77	0.00	159.26	0.00

En todos los casos se rechaza la hipótesis nula de linealidad y por tanto, las series parecen provenir de procesos no lineales de tipo MTAR.

5.3. Estimación del modelo

Se procede a obtener muestras las densidades condicionales completas propuestas. Considerando los hechos estilizados de las series financieras y teniendo como referencia la Tabla 4.1 se construyeron las cadenas de parámetros generando 30000 iteraciones ($3N$) y se utilizó un valor $\alpha = 0.001$. De igual manera, estas cadenas sufrieron una etapa de calentamiento (burn-in) para conseguir una longitud de $N = 10000$. Finalmente, el procedimiento adelgazamiento (thinning) fue aplicado hasta obtener cadenas de tamaño 1000.

5.3.1. Estimación de los parámetros estructurales

Utilizando la metodología de construcción de la verosimilitud marginal presentada en 3.3.2 se evalúan los diferentes modelos listados en la Tabla 5.1. Allí se observa que el mejor modelo es aquel que considera solamente dos regímenes, tres rezagos autoregresivos por régimen ($p_1 = p_2 = 3$) y la variable de umbral en términos contemporáneos.

# Modelo	Parámetros				Rezago Z_t	$\log(\hat{m})$
	l	p	q	d		
1	2	(3,3)	(0,0)	(0,0)	0	130,52
2	2	(3,3)	(0,0)	(0,0)	1	58,35
3	2	(3,3)	(0,0)	(0,0)	2	-15,02
4	2	(3,3)	(0,0)	(1,1)	0	81,22
5	2	(3,3)	(0,0)	(1,1)	1	-51,22
6	2	(3,3)	(0,0)	(1,1)	2	-59,88
7	3	(3,3,3)	(0,0,0)	(0,0,0)	0	-1069,35
8	3	(3,3,3)	(0,0,0)	(0,0,0)	1	-1228,14
9	3	(3,3,3)	(0,0,0)	(0,0,0)	2	-1747,32
10	3	(3,3,3)	(0,0,0)	(1,1,1)	0	-1005,95
11	3	(3,3,3)	(0,0,0)	(1,1,1)	1	-1782,88
12	3	(3,3,3)	(0,0,0)	(1,1,1)	2	-1796,56

TABLA 5.1. Log verosimilitud marginal estimada para el modelo

5.3.2. Estimación de los parámetros no estructurales

A través del modelo y algoritmo propuesto se estimaron los parámetros no estructurales y el parámetro de umbral. La Figura 5.3 muestra las cadenas de un conjunto de estos parámetros. La mayoría de las cadenas muestran un comportamiento unimodal y simétrico que apoya la idea de utilizar el promedio como un estimador de los parámetros no estructurales. Aunque incrementar el número de iteraciones o mejorar el mecanismo de selección de muestras ante distribuciones de colas pesadas permitiría mejorar la estabilidad de las cadenas, para todos los parámetros, se observó convergencia hacia su respectiva distribución estacionaria y en todo caso la estimación es adecuada. Además, por Roberts & Rosenthal (2009) se mantiene los requerimientos de ergodicidad.

A su vez considerando la media aritmética como un estimador de los parámetros no estructurales se obtiene la Tabla 5.2 donde se reportan las estimaciones puntuales, los intervalos de credibilidad asociados y las probabilidades de aceptación de las muestras propuestas. Las estimaciones significativas se resaltan en negrilla. Se observa que la mayoría de estimaciones son no significativas, por ejemplo, para el primer autoregresivo del régimen 1 el intervalo de credibilidad contiene el cero.

Los resultados muestran que el parámetro de no normalidad estimado es 8.57 por lo cual el proceso tiene colas pesadas y coincide con los hechos estilizados que caracterizan a las series trabajadas. Además, el valor de umbral se ubicó muy cercano a cero (-0.0081). Los resultados muestran que para el primer régimen los índices COLCAP y BOVESPA no se ven afectados entre sí, es decir, todo movimiento en estos recae en los rendimientos pasados obtenidos por cada uno. En contraste, en el segundo régimen el índice BOVESPA afecta en uno y hasta tres rezagos los retornos del COLCAP confirmando así la influencia del mercado de valores brasilero sobre el colombiano. De esta manera, cuando los retornos del índice S&P500 son superiores a -0.0081 los retornos del índice COLCAP también dependerán del resultado en el mercado brasilero.

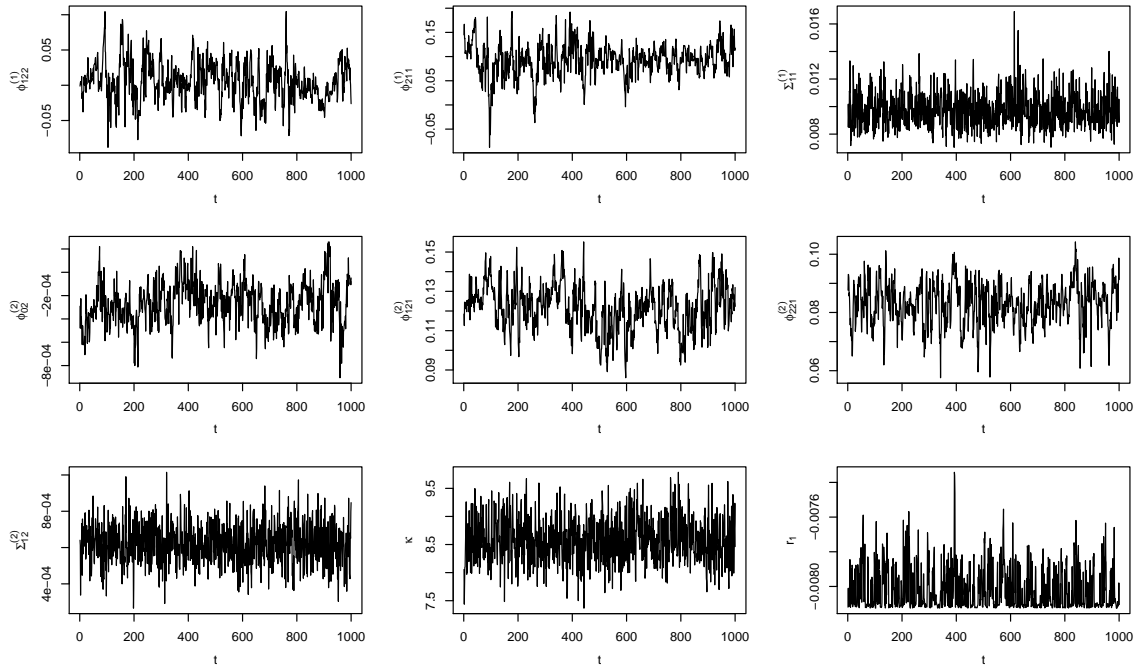


FIGURA 5.3. Valores de las cadenas para algunos parámetros del modelo

De lo anterior se confirma una leve asimetría hacia retornos negativos en donde valores del índice S&P500 inferiores a -0.0081 definen comportamientos diferenciales tanto en las estructuras de correlación (rezagos) como en la volatilidad. Se observa por ejemplo varianzas pequeñas y de distinta magnitud entre regímenes. Es así como en el régimen uno las varianzas y covarianzas estimadas son mayores, indicando mayor incertidumbre ante retornos inferiores al umbral estimado. Las covarianzas estimadas son positivas validando la relación existente entre los índices bursátiles de la región. Por último se alcanzaron las tasas de aceptación recomendadas (entre 10% y 60%).

5.4. Diagnóstico del modelo

El diagnóstico es realizado a través de las pruebas CUSUM y CUSUMSQ para la bondad de ajuste y la caracterización de los residuales marginales.

1. Pruebas CUSUM y CUSUMSQ.

La Figura 5.4 muestran a un nivel del 95% un buen ajuste del modelo. A través de los gráficos CUSUM y CUSUMSQ se observa que los residuales presentan un leve grado de heterocedasticidad.

Parámetro	Régimen 1			Régimen 2		
	Estimación	IC 95 %	% acept.	Estimación	IC 95 %	% acept.
ϕ_{01}	-0,0002	(-0.001,0.001)	30,1 %	$1.6e^{-5}$	(-0.000,0.000)	30,8 %
ϕ_{02}	0,0002	(-0.001,0.001)		-0,0003	(-0.001,0.000)	
ϕ_{111}	0,0488	(-0.024,0.124)	51,9 %	-0,0107	(-0.061,0.038)	48,3 %
ϕ_{112}	0,0386	(-0.046,0.116)		-0,0075	(-0.068,0.067)	
ϕ_{121}	0,0227	(-0.026,0.083)		0,1218	(0.098,0.146)	
ϕ_{122}	0,0058	(-0.052,0.062)		0,0039	(-0.026,0.031)	
ϕ_{211}	0,0929	(0.015,0.160)	51,1 %	-0,0732	(-0.098,-0.046)	50,3 %
ϕ_{212}	-0,0463	(-0.122,0.025)		0,0296	(-0.012,0.071)	
ϕ_{221}	0,0418	(-0.011,0.094)		0,0831	(0.067,0.096)	
ϕ_{222}	-0,0616	(-0.150,0.041)		-0,0088	(-0.031,0.025)	
ϕ_{311}	0,1350	(0.006,0.216)	50,5 %	0,0279	(0.004,0.054)	52,7 %
ϕ_{312}	0,0596	(-0.061,0.175)		-0,0127	(-0.054,0.015)	
ϕ_{321}	-0,0642	(-0.121,0.011)		0,0223	(0.004,0.041)	
ϕ_{322}	0,0059	(-0.071,0.084)		0,0514	(0.022,0.077)	
Σ_{11}	0,0098	(0.008,0.012)	48,0 %	0,0028	(0.002,0.003)	27,6 %
Σ_{12}	0,0019	(0.001,0.003)		0,0006	(0.000,0.001)	
Σ_{22}	0,0160	(0.012,0.021)		0,0054	(0.005,0.006)	
κ	8,5731	(7.871,9.391)	41,0 %			
r	-0,0081	(-0.008,-0.008)	44,9 %			

TABLA 5.2. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo

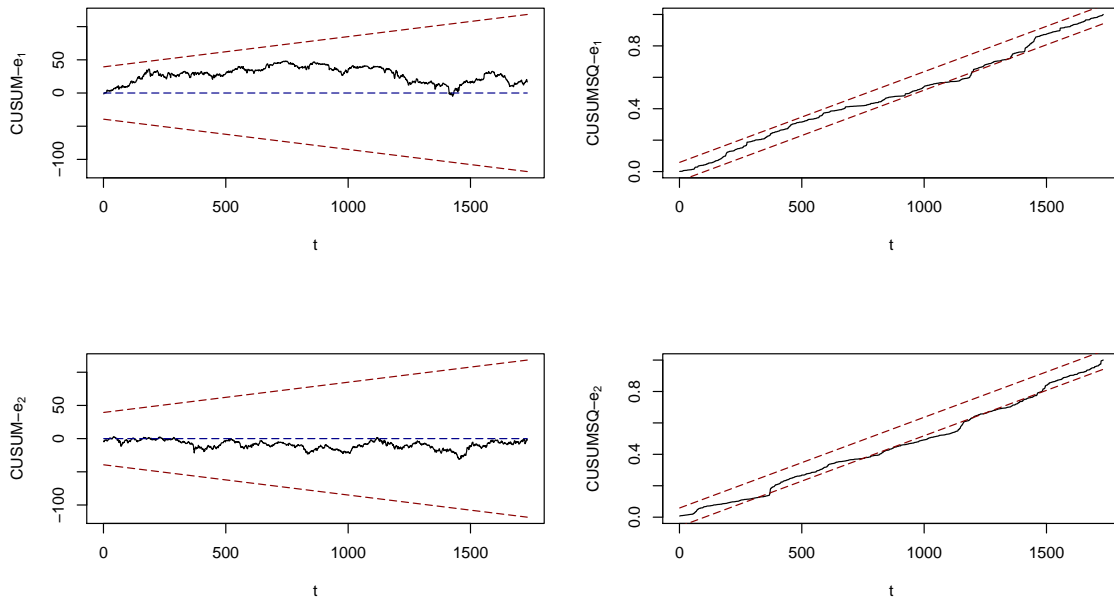


FIGURA 5.4. Pruebas CUSUM y CUSUMSQ para los residuos marginales del modelo

2. Validación de supuestos sobre los residuales.

Se observa que los residuos del modelo presenta alta curtosis, lo cual es coherente con

el valor del parámetro de no normalidad estimado. Es de destacar que los residuales presentan colas bastante pesadas y parecen ser estacionarios.

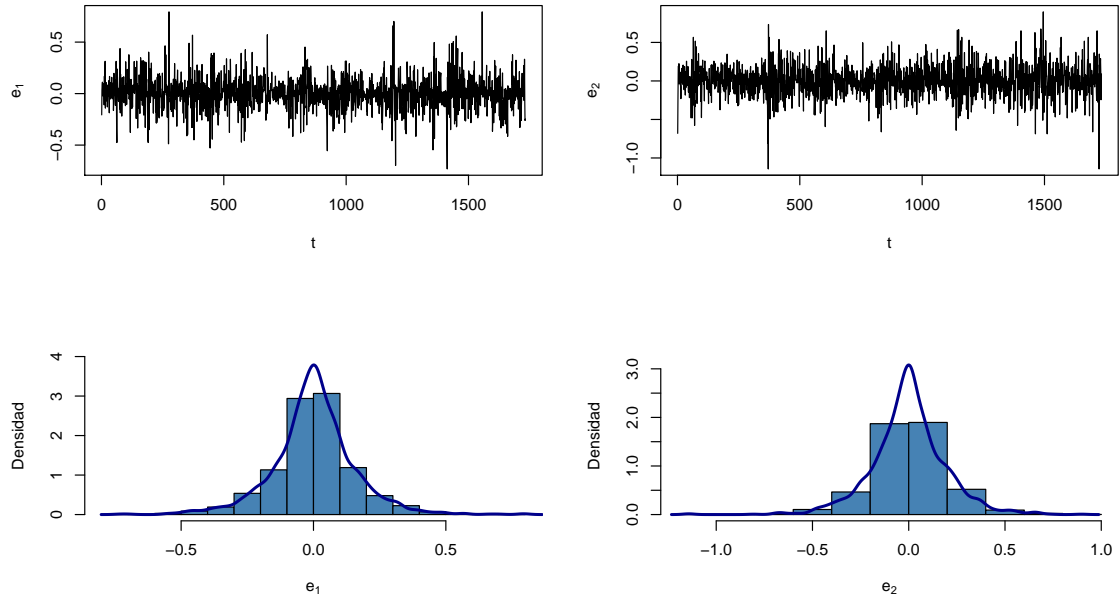


FIGURA 5.5. Residuales marginales del modelo

Conclusiones

Mediante un enfoque Bayesiano se estimaron los parámetros de un modelo MTAR cuando para cada t el error presenta una Distribución de Error Generalizada Multivariada (MGED). Esta distribución nos permite modelar una gran variedad de procesos sin fijarnos, preliminarmente, en la distribución multivariada asociada. De esta manera, el modelo puede ser aplicado en una gran variedad de campos como la meteorología, toxicología, psicología, economía, finanzas, entre otros.

Bajo los métodos MCMC tradicionales no fue posible estimar los parámetros de la metodología propuesta. Se encontró que al hacer uso de los métodos adaptativos desarrollados por Haario et al. (2001), Rosenthal (2008) y Roberts & Rosenthal (2009) se potencia la metodología propuesta al: 1) evitar que la escogencia de la distribuciones propuestas vía Metropolis se realice a prueba y error, 2) lograr estimaciones mas precisas, con un menor número de iteraciones y mejores tiempos de ejecución computacional y 3) asegurar principios de convergencia, ergodicidad y estacionariedad al satisfacer algunas condiciones comunes en la práctica.

Debido a la dificultad de extraer muestras de la distribución a posteriori de los parámetros del modelo, se construyen las distribuciones de densidad condicionales completas de los parámetros no estructurales y el parámetro de umbral. La elección de los demás parámetros estructurales se hizo a través del algoritmo propuesto por Chib & Jeliazkov (2001).

Adicionalmente, se realizó un ejercicio de simulación para 6 diferentes procesos MTAR con MGED cada uno con 100 repeticiones. Se encontró que:

1. A medida que el parámetro de no normalidad κ es más grande (colas pesadas) se requerirá un mayor número de iteraciones para que la convergencia sea alcanzada.
2. El 97% de los casos el parámetro se encuentra contenido en los intervalos de credibilidad estimados.
3. Los residuales obtenidos son ruido blanco, aparentemente estacionarios y, en los casos en que κ es grande, estos son levemente heterocedásticos.

Finalmente, se modeló el comportamiento de los índices accionarios COLCAP de Colombia y BOVESPA de Brasil cuyas transiciones están determinadas por el índice norteamericano S&P500. Se comprueba la existencia de una distribución de error de colas bastante pesadas que coincide en los hechos estilizados de las series trabajadas. Los resultados indican una influencia del mercado brasileño sobre el colombiano y un efecto importante del índice S&P500 sobre las dinámicas de los mercados latinoamericanos.

Trabajo futuro

En investigaciones futuras se tienen oportunidades de mejora como:

- Mejorar el algoritmo adaptativo de selección de muestras cuando se tienen distribuciones de colas pesadas. Esto debido a que las cadenas de los parámetros tienden a estar sesgadas, presentar inestabilidad en su varianza y no converger rápidamente. Una alternativa para el modelamiento de colas pesadas es hacer uso de la transformación logaritmo sobre cada ordenada de la distribución a posteriori donde la propiedad de ergodicidad se mantendrá si y solo si, la distribución tiene colas exponenciales o subexponenciales (Roberts & Tweedie, 1996).
- Completar la metodología Bayesiana con los componentes de identificación (prueba de no linealidad en el sentido MTAR) y de generación de pronósticos.
- Ampliar el alcance de la metodología a la modelación de la volatilidad condicional y no consicional. No solo la media (Moreno, 2006; Espinosa, 2016).
- Modelar la asimetría dadas las características de crecimientos asimétricos de los procesos MTAR (Bauwens & Laurent, 2005).

Bibliografía

- Al-Awadhi, S. & Jolliffe, I. (1998). Time series modelling of surface pressure data, *International Journal of Climatology* **18**(4): 443–455.
- Allen, F. & Gale, D. (2000). Financial contagion, *Journal of Political Economy* **108**(1): 1–33.
- Balke, N. S. & Fomby, T. B. (1997). Threshold Cointegration, *International Economic Review* **38**(3): 627–645.
- Barnes, M. L. (1999). Inflation and returns revisited: a TAR approach, *Journal of Multinational Financial Management* **9**(3): 233 – 245.
- Bauwens, L. & Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models, *Journal of Business and Economic Statistics* **23**: 346–354.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**(3): 307 – 327.
- Calderón, S. & Nieto, F. (2017). Bayesian analysis of multivariate threshold autoregressive models with missing data, *Communications in Statistics Theory and Methods* **46**(1): 296–318.
- Caner, M. & Hansen, B. E. (2001). Threshold Autoregression with a Unit Root, *Econometrica* **69**(6): 1555–1596.
- Castro, D. (2016). Estimación del modelo autorregresivo de umbrales cuando el proceso de ruido sigue una GED, Universidad Santo Tomás. Trabajo de Grado.
- Chib, S. (1995). Marginal Likelihood from the Gibbs Output, *Journal of the American Statistical Association* **90**(432): 1313–1321.
- Chib, S. & Jeliazkov, I. (2001). Marginal Likelihood from the Metropolis Hastings Output, *Journal of the American Statistical Association* **96**(453): 270–281.
- De Haan-Rietdijk, S., Gottman, J. M., Bergeman, C. S. & Hamaker, E. L. (2016). Get Over It! A Multilevel Threshold Autoregressive Model for State-Dependent Affect Regulation, *Psychometrika* **81**(1): 217–241.
- Domian, D. L. & Louton, D. A. (1997). A threshold autoregressive analysis of stock returns and real economic activity, *International Review of Economics & Finance* **6**(2): 167 – 179.

- Enders, W. & Granger, C. (1998). Unit-Root Tests and Asymmetric Adjustment With an Example Using the Term Structure of Interest Rates, *Journal of Business and Economic Statistics* **16**(3): 304–311.
- Enders, W. & Ludlow, J. (2002). Non-linear decay: tests for an attractor using a Fourier approximation, *The University of Alabama, Economics, Finance and Legal Studies, Working Paper Series WP01-02-02* (02).
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* **50**(4): 987–1007.
- Fang, K., Ng, K. & Kotz, S. (1990). *Symmetric multivariate and related distributions*, Chapman and Hall Ltd.
- Gelman, A. & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences, *Statist. Sci.* **7**(4): 457–472.
- Geman, S. & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**(6): 721–741.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments, in J. M. Bernardo, J. Berger, A. P. Dawid & J. F. M. Smith (eds), *Bayesian Statistics 4*, Oxford University Press, Oxford, pp. 169–193.
- Guiller, G. L. (2005). A Generalized Error Distribution, *Guiller Investments* pp. 1–7. 20031222/1.
- Haario, H., Saksman, E. & Tamminen, J. (2001). An Adaptive Metropolis Algorithm, *Bernoulli* **7**(2): 223–242.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**(1): 97–109.
- Heidelberger, P. & Welch, P. D. (1983). Simulation Run Length Control in the Presence of an Initial Transient, *Operations Research* **31**(6): 1109–1144.
- Kelker, D. (1970). Distribution Theory of Spherical Distributions and a Location-Scale Parameter Generalization, *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)* **32**(4): 419–430.
- Li, J. S.-H., Ng, A. C. & Chan, W.-S. (2015). Managing financial risk in Chinese stock markets: Option pricing and modeling under a multivariate threshold autoregression, *International Review of Economics & Finance* **40**: 217 – 230.
- Lo, M. & Zivot, E. (2001). Threshold Cointegration and Nonlinear Adjustment to The Law of One Price, **5**: 533–576.
- Loaiza-Maya, R. A., Gómez-González, J. E. & Melo-Velandia, L. F. (2015). Exchange rate contagion in Latin America, *Research in International Business and Finance* **34**: 355 – 367.
- Malmsten, H. & Teräsvirta, T. (2004). Stylized Facts of Financial Time Series and Three Popular Models of Volatility, *SSE/EFI Working Paper Series in Economics and Finance 563*, Stockholm School of Economics.

- McNeil, A. J., Frey, R. & Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, Princeton, NJ, USA.
- Melo, L. F. & Becerra, O. R. (2006). *Medidas de riesgo, características y técnicas de medición una aplicación del VaR y el ES a la tasa interbancaria de Colombia*, Centro Editorial Rosarista, 1 edn, Universidad del Rosario.
- Metropolis, N. & Ulam, S. (1953). A property of randomness of an arithmetical function, **60**(4): 252–253.
- Nadarajah, S. (2003). The Kotz-type distribution with applications, *Statistics* **37**(4): 341–358.
- Nieto, F. H. (2008). Forecasting with univariate TAR models, *Statistical Methodology* **5**(3): 263 – 276.
- Roberts, G. O., Gelman, A. & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms, *Ann. Appl. Probab.* **7**(1): 110–120.
- Roberts, G. O. & Rosenthal, J. S. (2007). Coupling and Ergodicity of Adaptive Markov Chain Monte Carlo Algorithms, *Journal of Applied Probability* **44**(2): 458–475.
- Roberts, G. O. & Rosenthal, J. S. (2009). Examples of Adaptive MCMC, *Journal of Computational and Graphical Statistics* **18**(2): 349–367.
- Roberts, G. O. & Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms, *Biometrika* **83**(1): 95–110.
- Romero-Orjuela, L. V. (2018). *Estimación Bayesiana de un modelo TAR multivariado cuando el proceso de ruido sigue una distribución t-student*, Master's thesis, Universidad Nacional de Colombia, Bogotá.
- Rosenthal, J. S. (2008). Optimal Proposal Distributions and Adaptive MCMC, *Handbook of Markov Chain Monte Carlo*, Chapman & Hall CRC.
- Solaro, N. (2004). Random variate generation from Multivariate Exponential Power distribution, *Statistica and Applicazioni* **2**(2).
- Tong, H. (1978). On a Threshold Model in Pattern Recognition and Signal Processing, *NATO ASI Series E: Applied Sc.* pp. 575–586.
- Tsay, R. S. (1989). Testing and Modeling Threshold Autoregressive Processes, *Journal of the American Statistical Association* **84**(405): 231–240.
- Tsay, R. S. (1998). Testing and Modeling Multivariate Threshold Models, *Journal of the American Statistical Association* **93**(443): 1188–1202.
- Tsay, R. S. (2005). *Analysis of financial time series*, Wiley series in probability and statistics, 2. ed. edn, Wiley-Interscience.
- Zhang, H. & Nieto, F. H. (2015). TAR Modeling with Missing Data when the White Noise Process Follows a Students t-Distribution, *Revista Colombiana de Estadística* **38**(1): 239–266.

APÉNDICE A

Construcción de muestras MGED

A.1. Obtención muestras de \mathbf{R}

Sea $g(r) = c(n, k)r^{H-1} \exp\{-hr^m\}$. Para $H = 1$, $h = \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{1}{2\kappa}}$, $m = \frac{1}{2\kappa}$, $c(n, \kappa) = \frac{1}{\sqrt{\pi^n}} \frac{\Gamma(1+n/2)}{\Gamma(1+n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{n}{2}}$ y r una variable aleatoria.

De acuerdo a la definición 1.1.2.

$$\begin{aligned} f_{\mathbf{R}}(r) &= \frac{2\pi^{\frac{n}{2}}}{\Gamma(n/2)} r^{n-1} g(r^2) \\ &= \frac{2\pi^{\frac{n}{2}}}{\Gamma(n/2)} r^{n-1} \frac{1}{\sqrt{\pi^n}} \frac{\Gamma(1+n/2)}{\Gamma(1+n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{n}{2}} \exp\left\{-\left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{1}{2\kappa}} r^{\frac{1}{\kappa}}\right\} \\ &= \frac{1}{\kappa\Gamma(n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{n}{2}} r^{n-1} \exp\left\{-\left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{1}{2\kappa}} r^{\frac{1}{\kappa}}\right\} \end{aligned}$$

Ahora, si $\mathbf{R}^* = \mathbf{R}^{\frac{1}{\kappa}}$, entonces,

$$\begin{aligned} f_{\mathbf{R}^*}(r^*) &= f_{\mathbf{R}}(r)|J| \\ &= \frac{1}{\Gamma(n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{n}{2}} r^{*n\kappa-1} \exp\left\{-\left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{1}{2\kappa}} r^*\right\} r^{*n\kappa-1} \\ &= \frac{1}{\Gamma(n\kappa)} \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{n}{2}} r^{*n\kappa-1} \exp\left\{-\left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{1}{2\kappa}} r^*\right\} \end{aligned}$$

Definiendo $\alpha = n\kappa$ y $\beta = \left[\frac{\Gamma(3\kappa)}{\Gamma(\kappa)}\right]^{\frac{1}{2\kappa}}$, se tiene que W se distribuye como una distribución Gamma y su función de densidad queda dada por:

$$f_{\mathbf{R}^*}(r^*) = \frac{\beta^\alpha}{\alpha} r^{*\alpha-1} \exp\{-\beta r^*\} I_{\{r^* > 0; \alpha, \beta > 0\}} \quad (\text{A.1})$$

A.2. Algoritmo para la obtención de muestras

Sea $\mathbf{W} \sim MGED_n(0, I, \kappa)$, entonces bajo 1.7 se tiene $\mathbf{W} \stackrel{d}{=} \mathbf{R}\mathbf{U}^{(n)}$.

1. Obtenga una muestra de \mathbf{R} de acuerdo a A.1.
2. Obtenga una muestra n-dimensional del vector $\mathbf{U}^{(n)}$ como $\frac{Z}{\|Z\|}$ donde Z tiene distribución normal multivariada.
3. Multiplique de forma escalar los dos conjuntos de muestras obtenidos en los pasos anteriores para construir una muestra n-dimensional de \mathbf{W} .

Otras simulaciones

Como se mencionó en el Capítulo 4 se generaron ejercicios adicionales de simulación donde se contemplaron distintos números de variables, diferente cantidad de regímenes y valores de κ distintos a los allí revisados.

B.1. Modelo 1: MTAR trivariado de dos regímenes con $\kappa = 0.5$

Sea $\{Y_t\}$ un proceso TAR trivariado de tamaño $N = 1000$ dado por

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} Y_{t-1} + \Sigma_{(1)}^{1/2} \varepsilon_t & ; Z_{t-1} \leq r \\ \phi_0^{(2)} + \phi_1^{(2)} Y_{t-1} + \delta_1^{(2)} Z_{t-1} + \Sigma_{(2)}^{1/2} \varepsilon_t & ; Z_{t-1} > r \end{cases} \quad (\text{B.1})$$

donde

$$\phi_0^{(1)} = \begin{pmatrix} -0.7 \\ 0.2 \\ 0.2 \end{pmatrix}, \quad \phi_1^{(1)} = \begin{pmatrix} 0.2 & -0.9 & 0 \\ 0.7 & -0.5 & 0.4 \\ 0.3 & 0.3 & 0.1 \end{pmatrix}, \quad \Sigma_{(1)} = \begin{pmatrix} 0.9 & -0.3 & -0.7 \\ -0.3 & 2.3 & -0.6 \\ -0.7 & -0.6 & 1 \end{pmatrix}$$

$$\phi_0^{(2)} = \begin{pmatrix} 0 \\ 0.8 \\ 0.6 \end{pmatrix}, \quad \phi_1^{(2)} = \begin{pmatrix} -0.9 & -0.4 & 0.5 \\ -0.1 & 0 & -0.6 \\ -0.4 & -0.6 & -0.5 \end{pmatrix}, \quad \delta_1^{(2)} = \begin{pmatrix} 0.9 \\ 0.6 \\ 0.1 \end{pmatrix}, \quad \Sigma_{(2)} = \begin{pmatrix} 2 & 0 & 0.7 \\ 0 & 1.3 & -0.4 \\ 0.7 & -0.4 & 0.9 \end{pmatrix}$$

con $r = \hat{z}_{0.5}$ como el percentil 50 de \mathbf{z} y $\varepsilon_t \sim MGED_3(\mathbf{0}, \mathbf{I}, \kappa = 0.5)$. Las variables regresoras y de umbral fueron obtenidas mediante distribuciones $N(0, 1)$. Se puede observar en la Figura B.1 una realización del proceso MTAR arriba descrito.

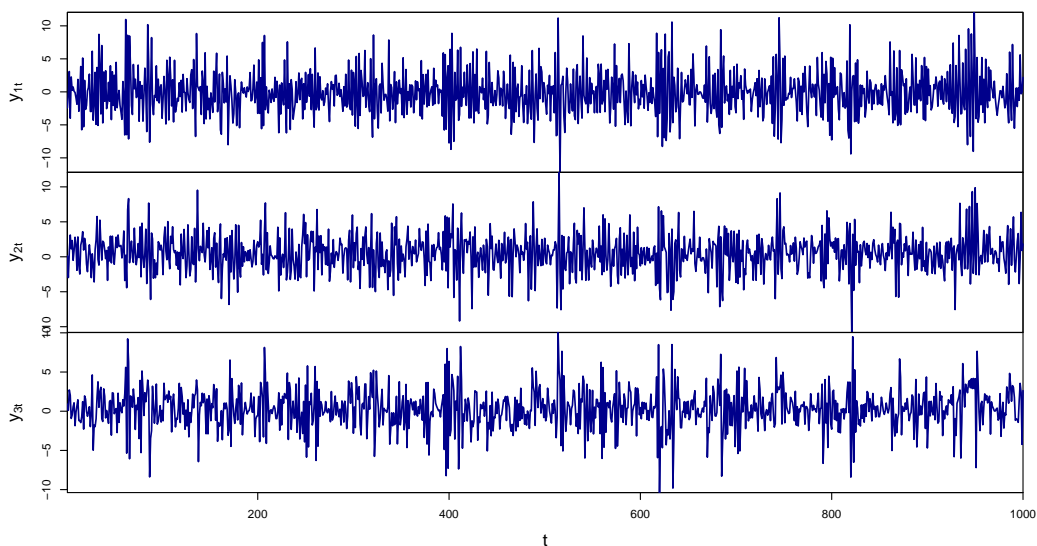


FIGURA B.1. Serie simulada del proceso descrito en B.1

Los resultados obtenidos del proceso de estimación son presentados en la Tabla B.1 y pueden ser interpretados como en el Capítulo 4.

Parámetros	Régimen 1			Régimen 2		
	Estimación	IC 95 %	% acept.	Estimación	IC 95 %	% acept.
ϕ_{01}	-0.6332	(-0.724,-0.547)	57.0 %	-0.0448	(-0.233,0.136)	39.7 %
ϕ_{02}	0.1775	(0.059,0.3)		0.8199	(0.673,0.97)	
ϕ_{03}	0.2101	(0.182,0.237)		0.5700	(0.475,0.664)	
ϕ_{111}	0.2315	(0.205,0.259)	22.7 %	-0.9172	(-0.95,-0.883)	24.5 %
ϕ_{112}	0.6856	(0.652,0.719)		-0.1025	(-0.13,-0.077)	
ϕ_{113}	0.2962	(0.288,0.304)		-0.3935	(-0.41,-0.377)	
ϕ_{121}	-0.9094	(-0.944,-0.869)		-0.4291	(-0.475,-0.384)	
ϕ_{122}	-0.5317	(-0.574,-0.486)		-0.0012	(-0.04,0.034)	
ϕ_{123}	0.3022	(0.291,0.313)		-0.6342	(-0.657,-0.611)	
ϕ_{131}	-0.0268	(-0.064,0.012)		0.5079	(0.46,0.557)	
ϕ_{132}	0.4552	(0.407,0.503)		-0.5986	(-0.64,-0.555)	
ϕ_{133}	0.1007	(0.09,0.111)		-0.4739	(-0.499,-0.448)	
δ_{11}				0.9334	(0.755,1.11)	42.5 %
δ_{12}				0.5834	(0.442,0.728)	
δ_{13}				0.1648	(0.075,0.252)	
Σ_{11}	0.9315	(0.823,1.05)	38.2 %	1.9753	(1.755,2.228)	34.6 %
Σ_{12}	-0.2896	(-0.357,-0.226)		-0.0494	(-0.202,0.105)	
Σ_{13}	-0.7703	(-0.861,-0.676)		0.6865	(0.538,0.852)	
Σ_{22}	2.2604	(1.993,2.552)		1.4242	(1.25,1.604)	
Σ_{23}	-0.5121	(-0.674,-0.366)		-0.4673	(-0.591,-0.351)	
Σ_{33}	1.0444	(0.888,1.216)		0.9710	(0.806,1.141)	
κ	0.4973	(0.4541,0.5443)	35.4 %			
r	-0.0012	(-0.0046,0.002)	28.8 %			

TABLA B.1. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo B.1 con $\kappa = 0.5$ y $r = \hat{z}_{0.5}$

Asimismo, las tasas de cobertura oscilan entre el 94 % y el 100 % con un error del 5 %. Igualmente las cadenas convergen siendo estables y estacionarias permitiendo encontrar la estimación adecuada y consiguiendo un buen ajuste.

B.2. Modelo 2: MTAR bivariado de tres regímenes con $\kappa = 0.5$

Sea $\{Y_t\}_{t=1}^N$ un proceso TAR bivariado de tamaño $N = 1000$ dado por

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} Y_{t-1} + \beta_1^{(1)} X_{t-1} + \Sigma_{(1)}^{1/2} \varepsilon_t & ; Z_{t-1} \leq r_1 \\ \phi_0^{(2)} + \delta_1^{(2)} Z_{t-1} + \Sigma_{(2)}^{1/2} \varepsilon_t & ; r_1 < Z_{t-1} \leq r_2 \\ \phi_0^{(3)} + \phi_1^{(3)} Y_{t-1} + \Sigma_{(3)}^{1/2} \varepsilon_t & ; Z_{t-1} > r_2 \end{cases} \quad (\text{B.2})$$

donde

$$\phi_0^{(1)} = \begin{pmatrix} 0.4 \\ 0.7 \end{pmatrix}, \quad \phi_1^{(1)} = \begin{pmatrix} 0.5 & -0.1 \\ 0.7 & -0.6 \end{pmatrix}, \quad \beta_2^{(1)} = \begin{pmatrix} -0.3 \\ 0 \end{pmatrix}, \quad \Sigma_{(1)} = \begin{pmatrix} 1.7 & -0.3 \\ -0.3 & 2.6 \end{pmatrix}$$

$$\phi_0^{(2)} = \begin{pmatrix} 0.1 \\ -0.5 \end{pmatrix}, \quad \delta_1^{(2)} = \begin{pmatrix} 0 \\ 0.6 \end{pmatrix}, \quad \Sigma_{(2)} = \begin{pmatrix} 1.2 & 0.2 \\ 0.2 & 1.7 \end{pmatrix}$$

$$\phi_0^{(3)} = \begin{pmatrix} 0.5 \\ -0.1 \end{pmatrix}, \quad \phi_1^{(3)} = \begin{pmatrix} 0.8 & -0.5 \\ 0.5 & -0.3 \end{pmatrix}, \quad \Sigma_{(3)} = \begin{pmatrix} 2 & 0.6 \\ 0.6 & 0.3 \end{pmatrix}$$

con $r_1 = \hat{z}_{0.44} = -0.15$, $r_2 = \hat{z}_{0.54} = 0.1$ y $\varepsilon_t \sim MGED_2(\mathbf{0}, \mathbf{I}, \kappa = 0.5)$. X_t y Z_t fueron obtenidas a través distribuciones $N(0, 1)$. Se puede observar en la Figura B.2 una realización del proceso MTAR arriba descrito.

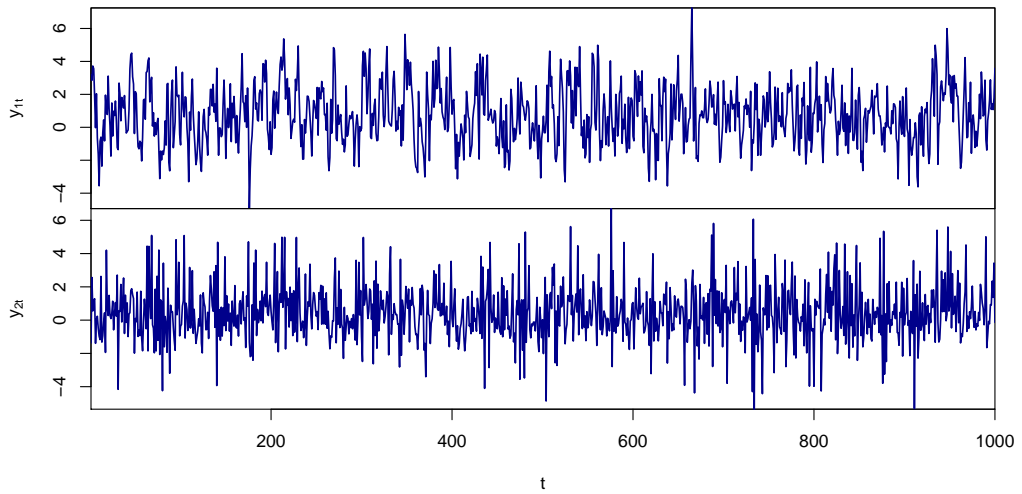


FIGURA B.2. Serie simulada del proceso descrito en B.2

Parámetros	Régimen 1			Régimen 2			Régimen 3		
	Estimación	IC 95 %	% acept.	Estimación	IC 95 %	% acept.	Estimación	IC 95 %	% acept.
ϕ_{01}	0.4242	(0.308,0.535)	48.5 %	0.2969	(0.096,0.494)	50.1 %	0.5146	(0.377,0.634)	47.8 %
ϕ_{02}	0.6295	(0.5,0.763)		-0.3297	(-0.526,-0.142)		-0.1079	(-0.137,-0.081)	
ϕ_{111}	0.5439	(0.482,0.606)	30.2 %				0.7059	(0.634,0.772)	27.1 %
ϕ_{112}	0.7603	(0.69,0.84)					0.4922	(0.477,0.506)	
ϕ_{121}	-0.1013	(-0.158,-0.04)					-0.4039	(-0.479,-0.326)	
ϕ_{122}	-0.6278	(-0.698,-0.556)					-0.2859	(-0.302,-0.27)	
β_{111}	-0.2613	(-0.366,-0.159)	52.5 %						
β_{112}	-0.0064	(-0.125,0.124)							
δ_{11}				0.0315	(-0.175,0.243)	51.9 %			30.2 %
δ_{12}				0.7222	(0.518,0.938)				
Σ_{11}	1.7767	(1.588,1.988)	29.9 %	1.1804	(0.912,1.502)	37.7 %	2.3143	(2.063,2.592)	32.4 %
Σ_{12}	-0.2426	(-0.381,-0.101)		0.3310	(0.106,0.567)		0.5570	(0.351,0.757)	
Σ_{22}	2.7467	(2.431,3.079)		1.5657	(1.186,2.062)		0.2508	(0.165,0.351)	
κ	0.4862	(0.4421,0.5386)	48.0 %						
r_1	-0.1526	(-0.1598,-0.1465)	43.1 %						
r_2	0.0993	(0.0965,0.1019)							

TABLA B.2. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo B.2 con $\kappa = 0.5$, $r_1 = \hat{z}_{0.44} = -0.15$ y $r_2 = \hat{z}_{0.54} = 0.1$

Los resultados obtenidos del proceso de estimación son presentados en la Tabla B.2 y pueden ser interpretados como en el Capítulo 4. Al igual que el modelo B.1, se encontraron tasas de cobertura que oscilan entre el 94 % y el 100 % con un nivel de confianza del 95 % y las cadenas convergen a su distribución estacionaria ofreciendo estimaciones adecuadas.

B.3. Modelo 3: MTAR bivariado de dos regímenes con $\kappa = 0.5$

Sea $\{Y_t\}$ un proceso TAR bivariado de tamaño $N = 1000$ dado por

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)}Y_{t-1} + \phi_2^{(1)}Y_{t-2} + \beta_1^{(1)}X_{t-1} + \Sigma_{(1)}^{1/2}\varepsilon_t & ; Z_{t-1} \leq r \\ \phi_0^{(2)} + \phi_1^{(2)}Y_{t-1} + \delta_1^{(2)}Z_{t-1} + \Sigma_{(2)}^{1/2}\varepsilon_t & ; Z_{t-1} > r \end{cases} \quad (\text{B.3})$$

donde

$$\begin{aligned} \phi_0^{(1)} &= \begin{pmatrix} 0 \\ -0.4 \end{pmatrix}, \quad \phi_1^{(1)} = \begin{pmatrix} -0.1 & -0.8 \\ 0.4 & -0.5 \end{pmatrix}, \quad \phi_2^{(1)} = \begin{pmatrix} -0.4 & 0.2 \\ -0.4 & -0.1 \end{pmatrix}, \quad \beta_1^{(1)} = \begin{pmatrix} 0.3 \\ 0.1 \end{pmatrix}, \\ \Sigma_{(1)} &= \begin{pmatrix} 0.16 & -0.3 \\ -0.3 & 1.44 \end{pmatrix} \\ \phi_0^{(2)} &= \begin{pmatrix} -0.1 \\ -0.9 \end{pmatrix}, \quad \phi_1^{(2)} = \begin{pmatrix} -0.4 & 0.6 \\ -0.2 & 0.7 \end{pmatrix}, \quad \delta_1^{(2)} = \begin{pmatrix} 0.2 \\ 0.5 \end{pmatrix}, \quad \Sigma_{(2)} = \begin{pmatrix} 0.64 & 0.4 \\ 0.4 & 0.64 \end{pmatrix} \end{aligned}$$

con $r = \hat{z}_{0.5} = 0$ y $\varepsilon_t \sim MGED_2(\mathbf{0}, \mathbf{I}, \kappa = 0.5)$. Las variables regresoras y de umbral fueron obtenidas mediante distribuciones $N(0, 1)$. Se puede observar en la Figura B.3 una realización del proceso MTAR arriba descrito.

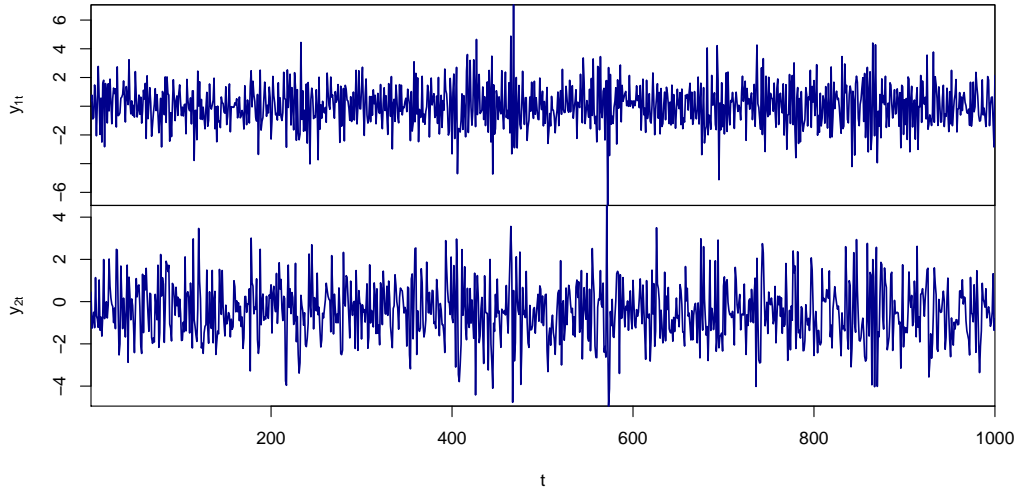


FIGURA B.3. Serie simulada del proceso descrito en B.3

Los resultados obtenidos del proceso de estimación son presentados en la Tabla B.3 y pueden ser interpretados como en el Capítulo 4.

Parámetros	Régimen 1			Régimen 2		
	Estimación	IC 95 %	% acept.	Estimación	IC 95 %	% acept.
ϕ_{01}	-0.0070	(-0.076,0.064)	38.9 %	-0.0489	(-0.167,0.057)	37.3 %
ϕ_{02}	-0.4129	(-0.489,-0.334)		-0.8812	(-0.96,-0.806)	
ϕ_{111}	-0.0969	(-0.142,-0.055)	38.4 %	-0.3990	(-0.44,-0.357)	52.4 %
ϕ_{112}	0.4041	(0.357,0.454)		-0.1695	(-0.2,-0.139)	
ϕ_{121}	-0.8267	(-0.88,-0.778)		0.5901	(0.545,0.634)	
ϕ_{122}	-0.4883	(-0.543,-0.428)		0.6934	(0.66,0.728)	
ϕ_{211}	-0.4180	(-0.459,-0.377)	43.6 %			
ϕ_{212}	-0.3840	(-0.432,-0.339)				
ϕ_{221}	0.2306	(0.19,0.269)				
ϕ_{222}	-0.1218	(-0.168,-0.074)				
β_{111}	0.3873	(0.331,0.443)	33.5 %			
β_{112}	-0.0253	(-0.092,0.04)				
δ_{11}				0.2192	(0.111,0.337)	37.3 %
δ_{12}				0.4943	(0.418,0.575)	
Σ_{11}	0.1513	(0.136,0.168)	47.4 %	0.6043	(0.538,0.676)	42.1 %
Σ_{12}	-0.2834	(-0.311,-0.259)		0.3531	(0.3,0.409)	
Σ_{22}	1.3993	(1.251,1.564)		0.6115	(0.522,0.704)	
κ	0.5023	(0.454,0.552)	45.0 %			
r	0.0009	(-0.003,0.003)	38.0 %			

TABLA B.3. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo B.3 con $\kappa = 0.5$ y $r = \hat{z}_{0.5} = 0$

En este caso las tasas de cobertura oscilan entre el 95 % y el 98 % con un nivel de confianza del 95 %. Dado que

B.4. Modelo 4: MTAR bivariado de dos regímenes con $\kappa = 8$

Sea $\{Y_t\}_{t=1}^N$ un proceso TAR bivariado de tamaño $N = 1000$ dado por

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)}Y_{t-1} + \beta_1^{(1)}X_{t-1} + \delta_1^{(1)}Z_{t-1} + \Sigma_{(1)}^{1/2}\varepsilon_t & ; Z_{t-1} \leq r \\ \phi_0^{(2)} + \phi_1^{(2)}Y_{t-1} + \Sigma_{(2)}^{1/2}\varepsilon_t & ; Z_{t-1} > r \end{cases} \quad (\text{B.4})$$

donde

$$\begin{aligned} \phi_0^{(1)} &= \begin{pmatrix} 0 \\ -0.4 \end{pmatrix}, \quad \phi_1^{(1)} = \begin{pmatrix} -0.1 & -0.8 \\ 0.4 & -0.5 \end{pmatrix}, \quad \beta_1^{(1)} = \begin{pmatrix} -0.4 \\ -0.4 \end{pmatrix}, \quad \delta_1^{(1)} = \begin{pmatrix} 0.2 \\ -0.1 \end{pmatrix}, \\ \Sigma_{(1)} &= \begin{pmatrix} 1.44 & -0.8 \\ -0.8 & 1.44 \end{pmatrix} \\ \phi_0^{(2)} &= \begin{pmatrix} 0.2 \\ -0.3 \end{pmatrix}, \quad \phi_1^{(2)} = \begin{pmatrix} -0.1 & -0.9 \\ -0.4 & -0.2 \end{pmatrix}, \quad \Sigma_{(2)} = \begin{pmatrix} 2.56 & 0.2 \\ 0.2 & 2.56 \end{pmatrix} \end{aligned}$$

con $r = \hat{z}_{0.5} = 0$ y $\varepsilon_t \sim MGED_2(\mathbf{0}, \mathbf{I}, \kappa = 8)$. Las variables regresoras y de umbral fueron obtenidas mediante distribuciones $N(0, 1)$. Se puede observar en la Figura B.4 una realización del proceso MTAR arriba descrito.

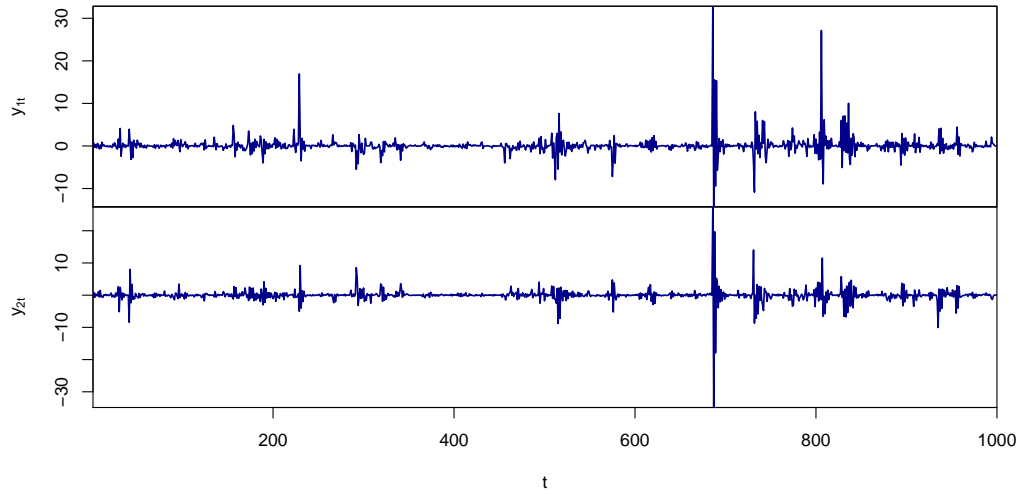


FIGURA B.4. Serie simulada del proceso descrito en B.4

Se puede observar en la series marcadas características de procesos de colas pesadas donde se presentan nichos de volatilidad y grandes valores atípicos. Esta simulación es justificada por el resultado de la aplicación empírica del Capítulo 5. Los resultados obtenidos del proceso de estimación son presentados en la Tabla B.4 y al igual que antes pueden ser interpretados de acuerdo al Capítulo 4.

Parámetros	Régimen 1			Régimen 2		
	Estimación	IC 95 %	% acept.	Estimación	IC 95 %	% acept.
ϕ_{01}	-0.0002	(-0.014,0.017)	26.1 %	0.1998	(0.193,0.207)	52.1 %
ϕ_{02}	-0.4021	(-0.414,-0.39)		-0.3060	(-0.313,-0.296)	
ϕ_{111}	-0.0991	(-0.1,-0.099)	49.8 %	-0.1000	(-0.101,-0.099)	53.8 %
ϕ_{112}	0.3996	(0.399,0.4)		-0.9005	(-0.901,-0.9)	
ϕ_{121}	-0.8006	(-0.801,-0.8)		-0.4007	(-0.401,-0.4)	
ϕ_{122}	-0.4999	(-0.5,-0.5)		-0.2005	(-0.201,-0.2)	
β_{111}	-0.3932	(-0.407,-0.38)	38.3 %			
β_{112}	-0.3969	(-0.405,-0.39)				
δ_{11}	0.2083	(0.19,0.232)	30.5 %			
δ_{12}	-0.1032	(-0.121,-0.09)				
Σ_{11}	1.3480	(0.964,1.831)	45.3 %	2.7667	(1.966,3.825)	51.0 %
Σ_{12}	-0.8813	(-1.199,-0.612)		0.3648	(0.077,0.692)	
Σ_{22}	1.5672	(1.102,2.173)		2.7815	(1.983,3.861)	
κ	8.0370	(7.4498,8.6152)	46.4 %			
r	-0.0009	(-0.0033,0.0034)	45.2 %			

TABLA B.4. Estimación puntual, intervalos de credibilidad (IC) y tasas de aceptación (% acept.) para el modelo B.4 con $\kappa = 8$ y $r = \hat{z}_{0.5} = 0$

Cabe destacar que la estimación presenta heterocedasticidad en ambos componentes de los residuales verificada a través de las gráficas CUSUMSQ. De esta manera es posible

concluir que el modelo fue ajustado correctamente en media pero no en varianza (volatilidad condicional). Asimismo se consiguieron tasas de cobertura más bajas que en los modelos anteriores oscilando entre 87% y 99% con una significancia del 5%. En cuanto a la estimación las cadenas alcanzaron su distribuciones estacionarias aunque se presenta mayor sesgo. La Figura B.5 muestra algunas cadenas de los parámetros en donde se observa que el algoritmo adaptativo ajustó la varianza de la cadena a su óptimo consiguiendo la mezcla deseada con tasas de aceptación entre el rango adecuado.

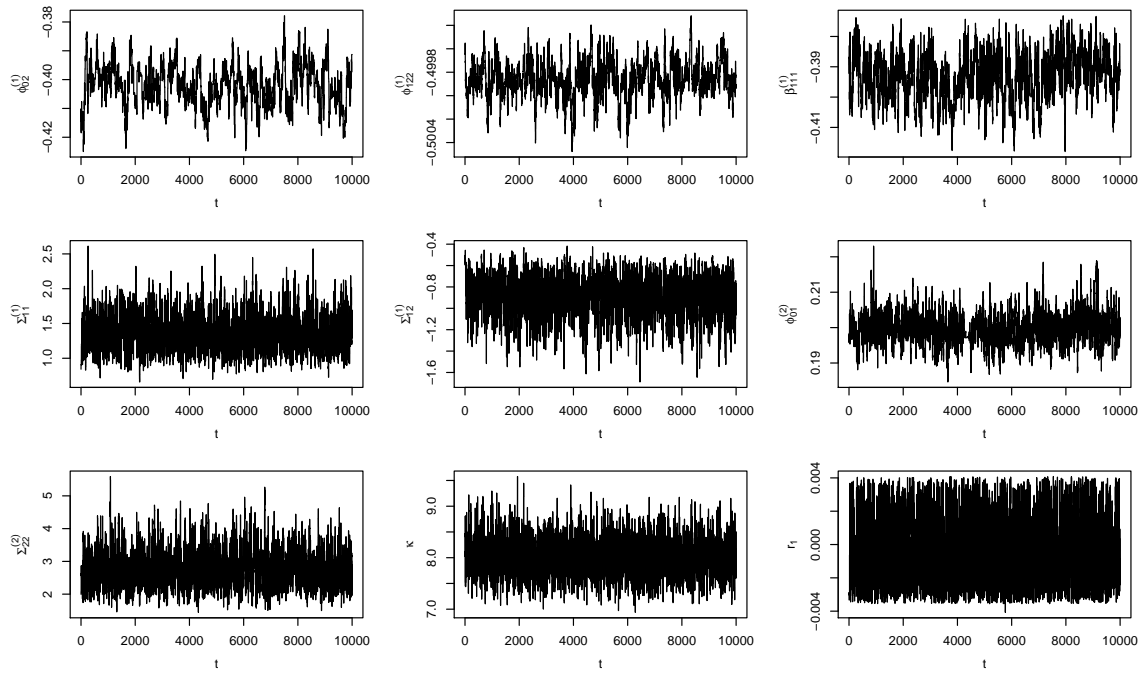


FIGURA B.5. Densidades a posteriori de algunos parámetros del modelo B.4

Diagnóstico de cadenas

Rápidamente los métodos MCMC se han convertido en herramientas muy populares en el análisis de modelos estadísticos complejos. Ese es el caso de algoritmos tan usuales como Metropolis-Hastings (Metropolis & Ulam (1953)) y Gibbs Sampling (Geman & Geman (1984)). Apropiadamente definidos e implementados, los métodos MCMC permiten al usuario conseguir exitosamente muestras de una cadena de Márkov convergente y a la vez de la distribución límite o estacionaria de la cual proviene sin importar que tan compleja ella sea.

C.1. Tasas de cobertura

A través de la simulación y estimación de 100 procesos para cada modelo presentado, se obtuvieron las estimaciones de cada parámetro y se compararon con su valor verdadero (valor simulado). Se encuentra que, en promedio, para el 97% de los casos, el verdadero parámetro cae dentro de los intervalos de credibilidad con una significancia del 5%, es decir, la tasa cobertura es del 97%. Luego, con una confianza del 99% la tasa de cobertura es del 99%.

C.1.1. Modelo MTAR con MGED con $\kappa = 0.1$

Aunque sus tasas de cobertura superan el 90%, los mayores sesgos estandarizados lo tienen las estimaciones de la matriz de varianzas y covarianzas; sin embargo, estos no superan media desviación estándar por lo cual la estimación es aceptada.

Parámetros	Régimen 1			Régimen 2		
	Parámetro	Cobertura	Sesgo Est.	Parámetro	Cobertura	Sesgo Est.
ϕ_{01}	0,00	0,98	0,08	-0,10	0,98	0,13
ϕ_{02}	-0,40	0,98	-0,17	-0,90	0,97	0,17
ϕ_{111}	-0,10	0,96	0,08	-0,40	0,96	0,01
ϕ_{112}	0,40	0,94	-0,10	-0,20	0,97	-0,05
ϕ_{121}	-0,80	0,97	0,02	0,60	0,96	0,10
ϕ_{122}	-0,50	0,96	-0,02	0,70	0,95	-0,02
ϕ_{211}	-0,40	0,97	-0,06			
ϕ_{212}	-0,40	0,97	0,13			
ϕ_{221}	0,20	0,96	0,17			
ϕ_{222}	-0,10	0,96	-0,22			
β_{111}	0,30	0,99	-0,12			
β_{112}	0,10	0,98	0,11			
δ_{11}				0,20	0,94	-0,04
δ_{12}				0,50	0,97	-0,08
Σ_{11}	0,16	0,99	0,40	0,64	0,98	0,17
Σ_{12}	-0,30	0,96	-0,25	0,40	1,00	0,05
Σ_{22}	1,44	0,98	-0,01	0,64	0,97	0,07
κ	0,10	1,00	-0,09			
r	0,00	0,98	-0,15			

TABLA C.1. Tasa de cobertura y sesgo estandarizado de las estimaciones del modelo 4.1

C.1.2. Modelo MTAR con MGED con $\kappa = 1$

Se encuentra que a medida que el valor del parámetro de no normalidad κ se hace más grande el sesgo estandarizado asociado a cada estimación aumenta. Luego, a pesar de esto los sesgos se mantienen por debajo de una desviación estándar y las tasas de cobertura son adecuadas.

Parámetros	Régimen 1			Régimen 2		
	Parámetro	Cobertura	Sesgo est.	Parámetro	Cobertura	Sesgo est.
ϕ_{01}	0,00	0,99	-0,01	-0,10	1,00	0,01
ϕ_{02}	-0,40	0,95	0,00	-0,90	0,98	0,04
ϕ_{111}	-0,10	0,96	0,10	-0,40	0,94	0,11
ϕ_{112}	0,40	0,98	-0,18	-0,20	0,97	0,10
ϕ_{121}	-0,80	0,96	0,03	0,60	0,98	-0,16
ϕ_{122}	-0,50	0,96	0,02	0,70	0,99	-0,11
ϕ_{211}	-0,40	0,95	-0,07			
ϕ_{212}	-0,40	0,91	0,03			
ϕ_{221}	0,20	0,94	0,06			
ϕ_{222}	-0,10	0,91	-0,14			
β_{111}	0,30	0,98	-0,11			
β_{112}	0,10	0,96	0,10			
δ_{11}				0,20	0,98	0,10
δ_{12}				0,50	0,98	-0,05
Σ_{11}	0,16	0,95	0,55	0,64	0,98	0,34
Σ_{12}	-0,30	0,94	-0,58	0,40	1,00	0,19
Σ_{22}	1,44	0,96	0,43	0,64	0,98	0,30
κ	1,00	0,99	-0,24			
r	0,00	0,99	-0,01			

TABLA C.2. Tasa de cobertura y sesgo estandarizado de las estimaciones del modelo 4.2

C.2. Pruebas de Convergencia

Una de las principales desventajas de los métodos MCMC es la dificultad para determinar el número de iteraciones necesario para concluir que la cadena ha convergido. La solución del problema de determinar la convergencia de un algoritmo MCMC se ha concentrado en dos áreas. La primera es teórica y es explorada por autores como Polson (1994) o Rosenthal(1993, 1995a, 1995b), entre otros. Estos trabajos han analizado el kernel de transición de la cadena buscando determinar el número de iteraciones necesario para alcanzar la convergencia, con algún grado de significancia, a su verdadera distribución estacionaria. Tal enfoque requiere cálculos poco estandarizados y muy elaborados que deben repetirse para cada modelo a evaluar.

El segundo enfoque es utilizado en la mayoría de los trabajos que hacen uso de algoritmos MCMC y se concentran en el resultado obtenido del algoritmo. En este caso se comparan las distribuciones empíricas de los valores conseguidos de las iteraciones y se concluye convergencia cuando las distancia entre las cadenas sea insignificante en algún grado.

Algunas de las pruebas desarrolladas en el marco del segundo enfoque son:

1. Gelman & Rubin (1992) propuso un enfoque general para monitorear la convergencia de las cadenas MCMC en la $m > 1$ cadenas son comparadas.
 - Usa diferentes valores iniciales que están sobre dispersos en relación a la distribución posterior.

- La convergencia se presenta cuando las todas cadenas, independientemente de sus valores iniciales, presentan diferencias indistinguibles en sus últimas iteraciones.
 - El diagnóstico es aplicado a un único parámetro por cadena y se base en una comparación de las varianzas entre e intra.
 - Valores del estadístico \hat{R} cercanos a 1 sugieren convergencia.
2. Geweke (1992) propone un diagnóstico de convergencia basado en la igualdad de las medias de la primera y la última parte de una cadena de Márkov (por defecto el primer 10 % y el ultimo 50 %). Si las muestras fueron tomadas de la distribución estacionaria de la cadena, las dos medias son iguales y el estadístico *Geweke* sigue asintóticamente una distribución normal estándar.
 3. A través del uso del estadístico Cramer-von-Mises para probar la hipótesis nula que las muestras obtenidas provienen de una distribución estacionaria (Heidelberger & Welch, 1983).
 - La prueba es aplicada sucesivamente, iniciando con la cadena completa, luego descartando primero el 10 %, 20 % de la cadena hasta que la hipótesis nula no sea rechazada o el 50 % de la cadena ha sido descartada.
 - El último resultado indica que la estacionariedad no fue alcanzada y se requiere construir una cadena más larga.
 - Si la estacionariedad es alcanzada (no se rechaza H_0), el número de iteraciones adecuado y el número a descartar (burn-in) es reportado.

C.2.1. Modelo MTAR con MGED con $\kappa = 0.1$

La Tabla C.3 y la Figura C.2 muestra los resultados de las pruebas de convergencia realizadas para el modelo MTAR con MGED cuando el parámetro de no normalidad κ es 0.1. Se observa que para todos los casos alcanza convergencia y estacionariedad.

Parámetros	Heidelberger & Welch (1983)*															
	Gelman & Rubin (1992)				Cadena 1				Cadena 2				Cadena 3			
	Régimen 1		Régimen 2		Régimen 1		Régimen 2		Régimen 1		Régimen 2		Régimen 1		Régimen 2	
	\hat{R}	IC 95 %	\hat{R}	IC 95 %	H_0^*	p-valor	H_0^*	p-valor	H_0^*	p-valor	H_0^*	p-valor	H_0^*	p-valor	H_0^*	p-valor
ϕ_{01}	1,11	1,31	1,18	1,48	✓	0,55	✓	0,34	✓	0,56	✓	0,49	✓	0,37	✓	0,89
ϕ_{02}	1,24	1,77	1,05	1,13	✓	0,69	✓	0,30	✓	0,40	✓	0,38	✓	0,55	✓	0,10
ϕ_{111}	1,02	1,08	1,26	1,87	✓	0,87	✓	0,54	✓	0,86	✓	0,17	✓	0,19	✓	0,55
ϕ_{112}	1,04	1,05	1,26	1,81	✓	0,82	✓	0,61	✓	0,16	✓	0,11	✓	0,19	✓	0,86
ϕ_{121}	1,02	1,06	1,21	1,70	✓	0,26	✓	0,69	✓	0,11	✓	0,33	✓	0,13	✓	0,48
ϕ_{122}	1,09	1,11	1,19	1,62	✓	0,20	✓	0,74	✓	0,08	✓	0,23	✓	0,36	✓	0,59
ϕ_{211}	1,22	1,66			✓	0,34				0,43				0,29		
ϕ_{212}	1,27	1,90			✓	0,42				0,16				0,37		
ϕ_{221}	1,29	2,27			✓	0,54				0,14				0,18		
ϕ_{222}	1,30	2,42			✓	0,58				0,17				0,07		
β_{111}	1,17	1,38			✓	0,23				0,08				0,55		
β_{112}	1,21	1,52			✓	0,23				0,57				0,64		
δ_{11}			1,18	1,58				0,42	✓			0,29	✓			0,06
δ_{12}			1,16	1,45				0,75	✓			0,19	✓			0,13
Σ_{11}	1,16	1,19	1,11	1,19	✓	0,82	✓	0,29	✓	0,76	✓	0,53	✓	0,60	✓	0,09
Σ_{12}	1,19	1,20	1,10	1,25	✓	0,54	✓	0,68	✓	0,50	✓	0,55	✓	0,39	✓	0,46
Σ_{22}	1,16	1,32	1,24	1,71	✓	0,81	✓	0,84	✓	0,57	✓	0,57	✓	0,72	✓	0,70
κ	1,03	1,03			✓	0,60				0,46				0,51		
r	1,10	1,11			✓	0,55				0,06				0,63		

TABLA C.3. Pruebas de convergencia de Gelman & Rubin (1992) y Heidelberger & Welch (1983).

* H_0 : La cadena es estacionaria.

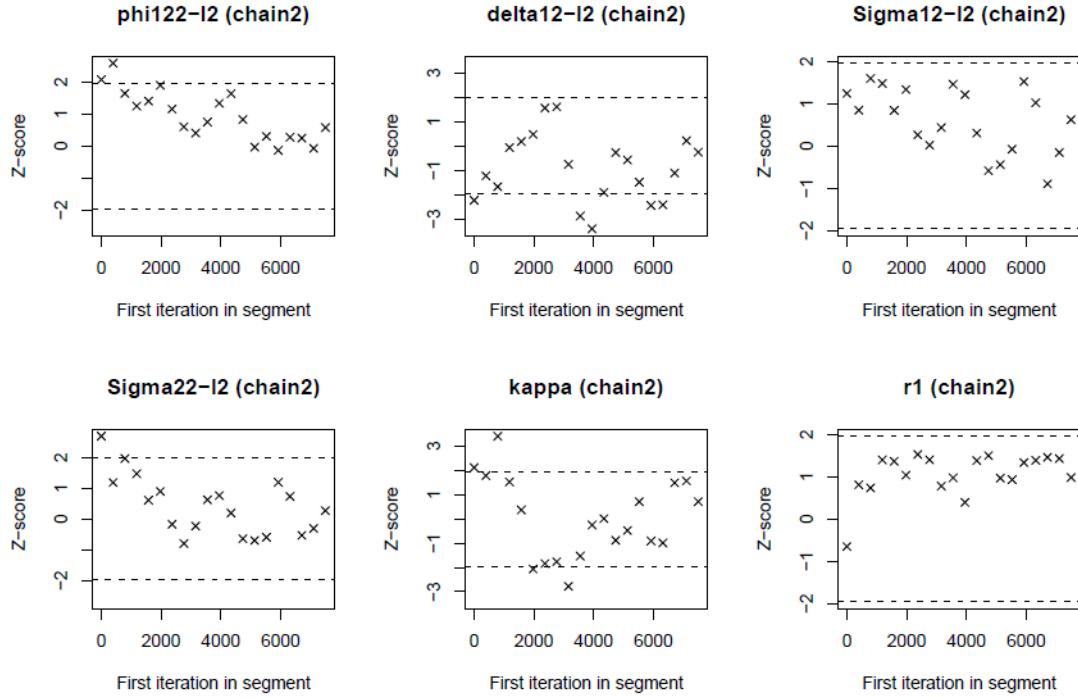


FIGURA C.1. Prueba de convergencia de Geweke (1992) para algunos parámetros

C.2.2. Modelo MTAR con MGED con $\kappa = 1$

Al igual que antes, la Tabla C.4 y la Figura muestran los resultados de las pruebas de convergencia realizadas para el modelo MTAR con MGED cuando el parámetro de no normalidad es 1. Se observa que para todos los casos alcanza convergencia y estacionariedad.

Parámetros	Gelman & Rubin (1992)						Heidelberger & Welch (1983)*								
	Régimen 1			Régimen 2			Cadena 1		Cadena 2		Cadena 3				
	\hat{R}	IC 95 %	\hat{R}	IC 95 %	H_0^*	p-valor	Régimen 1	Régimen 2	Régimen 1	Régimen 2	Régimen 1	Régimen 2			
ϕ_{01}	1,11	1,25	1,14	1,43	✓	0,17	✓	0,40	✓	0,56	0,44	✓	0,42	✓	0,06
ϕ_{02}	1,05	1,06	1,21	1,76	✓	0,06	✓	0,31	✓	0,24	0,34	✓	0,25	✓	0,67
ϕ_{111}	1,26	2,03	1,21	1,69	✓	0,08	✓	0,27	✓	0,50	0,44	✓	0,68	✓	0,58
ϕ_{112}	1,19	1,46	1,17	1,49	✓	0,06	✓	0,28	✓	0,25	0,22	✓	0,46	✓	0,47
ϕ_{121}	1,15	1,48	1,12	1,33	✓	0,09	✓	0,21	✓	0,53	0,36	✓	0,55	✓	0,36
ϕ_{122}	1,13	1,35	1,19	1,48	✓	0,16	✓	0,21	✓	0,27	0,13	✓	0,08	✓	0,39
ϕ_{211}	1,19	1,59			✓	0,21			✓	0,38		✓	0,55		
ϕ_{212}	1,13	1,40			✓	0,27			✓	0,10		✓	0,08		
ϕ_{221}	1,11	1,14			✓	0,49			✓	0,27		✓	0,19		
ϕ_{222}	1,08	1,23			✓	0,22			✓	0,20		✓	0,20		
β_{111}	1,03	1,03			✓	0,96			✓	0,11		✓	0,21		
β_{112}	1,10	1,14			✓	0,12			✓	0,06		✓	0,37		
δ_{11}			1,18	1,60	✓		✓	0,34	✓		0,42	✓		✓	0,42
δ_{12}			1,31	2,12	✓		✓	0,27	✓		0,24	✓		✓	0,27
Σ_{11}	1,12	1,22	1,23	1,34	✓	0,77	✓	0,69	✓	0,81	0,50	✓	0,88	✓	0,49
Σ_{12}	1,27	1,63	1,01	1,01	✓	0,45	✓	0,57	✓	0,62	0,62	✓	0,81	✓	0,60
Σ_{22}	1,22	1,62	1,04	1,04	✓	0,51	✓	0,55	✓	0,63	0,70	✓	0,81	✓	0,68
κ	1,05	1,14			✓	0,13			✓	0,13		✓	0,21		
r	1,12	1,23			✓	0,30			✓	0,61		✓	0,65		

TABLA C.4. Pruebas de convergencia de Gelman & Rubin (1992) y Heidelberger & Welch (1983).

* H_0 : La cadena es estacionaria.

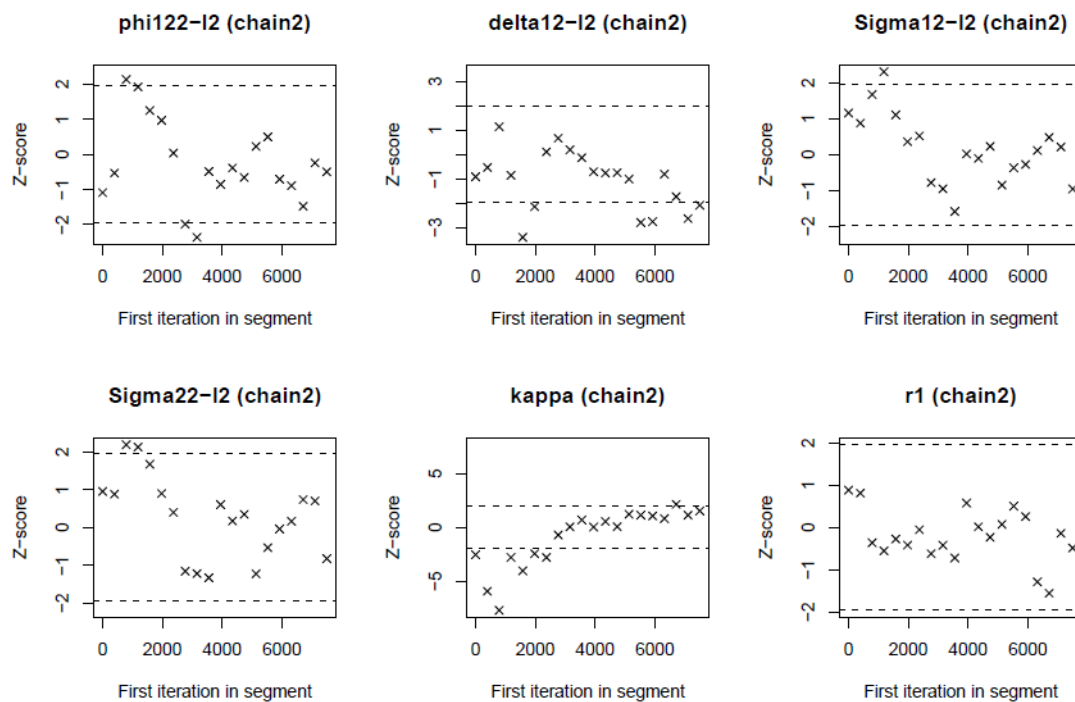


FIGURA C.2. Prueba de convergencia de Geweke (1992) para algunos parámetros