



UNIVERSIDAD NACIONAL DE COLOMBIA

Pronósticos basados en un modelo multivariado autorregresivo de umbrales (MTAR) con distribución de error *t-Student* multivariada desde el enfoque Bayesiano

Nicolás Rivera Garzón

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá, D.C., Colombia
2023

Pronósticos basados en un modelo multivariado autorregresivo de umbrales (MTAR) con distribución de error *t-Student* multivariada desde el enfoque Bayesiano

Nicolás Rivera Garzón

Trabajo de grado presentado como requisito parcial para optar al título de:
Magíster en Ciencias - Estadística

Director:

Ph.D., Sergio Alejandro Calderón Villanueva

Co-Director:

M.Sc., MBA., Óscar Andrés Espinosa Acuña

Línea de Investigación:

Series de Tiempo

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá, D.C., Colombia
2023

Dedicatoria

A mis padres por su amor, aliento y apoyo incondicional. Todo lo que soy se los debo a ellos.

Agradecimientos

Agradezco a los Profesores Sergio Calderón y Óscar Espinosa por su asesoría, tiempo y dedicación durante la realización de esta investigación.

Resumen

Pronósticos basados en un modelo multivariado autorregresivo de umbrales (MTAR) con distribución de error *t-Student* multivariada desde el enfoque Bayesiano

En este trabajo se presenta un método que permite obtener los pronósticos basados en un modelo MTAR con distribución de error *t-Student* multivariada desde el enfoque Bayesiano. Para ello, se encuentra la distribución predictiva Bayesiana que incluye la incertidumbre sobre los verdaderos valores de los parámetros del modelo MTAR. El procedimiento planteado se basa en la obtención de muestras de la distribución predictiva para obtener el pronóstico puntual e intervalos de predicción del proceso de interés. El desempeño del algoritmo planteado se verifica a través de un estudio de simulación basado en tres modelos en donde se calcula el porcentaje de veces en que los valores verdaderos del proceso de salida se encuentran dentro del intervalo de predicción del 95 % de la distribución predictiva. Posteriormente se presenta una aplicación a un conjunto de series de tiempo financieras donde se obtienen los pronósticos de los retornos de los índices Bovespa y Colcap usando como variable umbral los retornos del índice Standard and Poor's 500 y se comparan los pronósticos con los obtenidos por un modelo MTAR con distribución de error normal multivariada.

Palabras clave: modelos MTAR, estadística Bayesiana, pronósticos, distribución predictiva, distribución *t-Student*, modelos no lineales.

Abstract

Forecasts based on a multivariate autoregressive threshold model (MTAR) with a multivariate t-Student error distribution from a Bayesian approach

This paper presents a Bayesian method to obtain forecasts based on a MTAR model with a multivariate t-Student error distribution. For this, the Bayesian predictive distribution is found, which includes the uncertainty about the true values of the parameters of the MTAR model. The proposed procedure is based on drawing samples from the predictive distribution to obtain the point forecast and prediction intervals of the process of interest. The performance of the proposed algorithm is verified through a simulation study based on three models where the percentage of times in which the true values of the output process are within the prediction interval of 95% of the predictive distribution is calculated. Subsequently, an application to a set of financial time series is presented where the forecasts of the returns of the Bovespa and Colcap indexes are obtained using the returns of the Standard and Poor's 500 index as a threshold variable and the forecasts are compared with those obtained by a MTAR model with multivariate normal error distribution.

Keywords: MTAR models, Bayesian statistics, forecasting, predictive distribution, Student's t-distribution, nonlinear models.

Contenido

Agradecimientos	VII
Resumen	IX
Contenido	XI
Lista de Figuras	XIII
Lista de Tablas	XV
1. Introducción	1
2. Aspectos teóricos	4
2.1. Estadística Bayesiana	4
2.2. Métodos de simulación MCMC	6
2.2.1. Algoritmo de Metropolis-Hastings	7
2.2.2. Muestreador de Gibbs	9
2.2.3. Comprobación de convergencia	10
2.3. Precisión de los pronósticos	11
2.3.1. Validación cruzada de series de tiempo	11
2.3.2. Prueba de Diebold y Mariano	13
2.4. Modelo multivariado autorregresivo de umbrales con distribución de error <i>t</i> - <i>Student</i> multivariada	14
2.4.1. Especificación del modelo MTAR	14
2.4.2. Pruebas de no linealidad	15
2.4.3. Función de verosimilitud	17
2.4.4. Estimación de parámetros no estructurales	18
2.4.5. Estimación e identificación de parámetros estructurales	20
2.4.6. Validación del modelo	21
3. Pronósticos	22
4. Ejercicio de simulación	26
4.1. Modelo 1	26

4.2. Modelo 2	31
4.3. Modelo 3	36
4.4. Evaluación de la metodología de pronóstico	42
5. Aplicación	43
5.1. Evaluación de los pronósticos	49
6. Conclusiones	53
A. Generalidades del Modelo MTAR ajustado para las series de retornos COLCAP y BOVESPA	55
A.1. Prueba de no linealidad	55
A.2. Pruebas de raíz unitaria	56
A.3. Estimación de parámetros estructurales	56
A.4. Estimación de parámetros no estructurales	58
A.5. Residuos	67
B. Modelo multivariado autorregresivo de umbrales (MTAR) con distribución de error normal multivariada	69
C. Aplicación	72
C.1. Resultados de la validación cruzada	72
C.2. Resultados de la prueba de Diebold y Mariano modificada	74
C.3. ARIMA - Proceso de umbrales	76
C.4. Cadenas y densidades de pronósticos conjuntos	78
Referencias	80

Lista de Figuras

4-1.	Serie simulada del modelo MTAR con $k = 2$ y $l = 2$ - Modelo 1.	27
4-2.	Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Modelo 1.	29
4-3.	Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 1.	30
4-4.	Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Modelo 1.	30
4-5.	Serie simulada del modelo MTAR con $k = 2$ y $l = 3$ - Modelo 2.	32
4-6.	Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Modelo 2.	34
4-7.	Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 2.	35
4-8.	Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Modelo 2.	35
4-9.	Serie simulada del modelo MTAR con $k = 3$ y $l = 2$ - Modelo 3.	36
4-10.	Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Modelo 3.	39
4-11.	Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 3.	40
4-12.	Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Modelo 3.	41
5-1.	Retornos e Índices S&P 500, COLCAP y BOVESPA - Aplicación.	43
5-2.	Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Aplicación.	47
5-3.	Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Aplicación.	48
5-4.	Validación cruzada de los pronósticos obtenidos.	50
A-1.	Cadena y densidad del vector de umbrales r - Aplicación.	57
A-2.	Criterio de convergencia de Geweke del vector de umbrales r - Aplicación. .	57
A-3.	Cadenas y densidades de las matrices autorregresivas correspondientes al régimen $j = 1$ - Aplicación.	58
A-4.	Cadenas y densidades de las matrices autorregresivas correspondientes al régimen $j = 2$ - Aplicación.	59
A-5.	Criterio de convergencia de Geweke de las matrices autorregresivas correspondientes al régimen $j = 1$ - Aplicación.	60

A-6.	Criterio de convergencia de Geweke de las matrices autorregresivas correspondientes al régimen $j = 2$ - Aplicación.	61
A-7.	Cadenas y densidades del vector de desviaciones estándar escaladas correspondiente al régimen $j = 1$ - Aplicación.	61
A-8.	Cadenas y densidades del vector de desviaciones estándar escaladas correspondiente al régimen $j = 2$ - Aplicación.	62
A-9.	Criterio de convergencia de Geweke del vector de desviaciones estándar escaladas correspondiente al régimen $j = 1$ - Aplicación.	63
A-10.	Criterio de convergencia de Geweke del vector de desviaciones estándar escaladas correspondiente al régimen $j = 2$ - Aplicación.	63
A-11.	Cadenas y densidades del coeficiente de correlación escalado correspondiente al régimen $j = 1$ - Aplicación.	64
A-12.	Cadenas y densidades del coeficiente de correlación escalado correspondiente al régimen $j = 2$ - Aplicación.	64
A-13.	Criterio de convergencia de Geweke del coeficiente de correlación escalado correspondiente al régimen $j = 1$ - Aplicación.	65
A-14.	Criterio de convergencia de Geweke del coeficiente de correlación escalado correspondiente al régimen $j = 2$ - Aplicación.	65
A-15.	Cadena y densidad de los grados de libertad ν - Aplicación.	66
A-16.	Criterio de convergencia de Geweke de los grados de libertad ν - Aplicación.	66
A-17.	Residuales del modelo MTAR estimado - Aplicación.	67
A-18.	CUSUM y CUSUMSQ de los residuales del modelo estimado - Aplicación.	67
A-19.	Correlaciones de los residuales del modelo estimado - Aplicación.	68
C-1.	Residuos del modelo ARIMA de S&P 500 - Aplicación.	76
C-2.	Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Semana extra - Aplicación.	78
C-3.	Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Semana extra - Aplicación.	79

Lista de Tablas

4-1.	Pronósticos de 1 a 10 pasos adelante - Modelo 1.	28
4-2.	Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 1.	28
4-3.	Pronósticos de 1 a 10 pasos adelante - Modelo 2.	33
4-4.	Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 2.	33
4-5.	Pronósticos de 1 a 10 pasos adelante - Modelo 3.	37
4-6.	Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 3.	38
4-7.	Porcentaje de veces en que los valores verdaderos del proceso de salida se encuentran dentro del intervalo de predicción del 95 % y rango medio de los intervalos - Modelos 1, 2 y 3.	42
5-1.	Prueba de no linealidad de Lo y Zivot (2001) - Aplicación.	44
5-2.	Prueba de no linealidad de Tsay (1998) - Aplicación.	45
5-3.	Pronósticos de 1 a 10 pasos adelante - Aplicación.	48
5-4.	Pronósticos de 1 a 10 pasos adelante - Pronósticos conjuntos - Aplicación.	49
5-5.	Resultados de validación cruzada - Aplicación.	50
5-6.	Resultados de la prueba de Diebold y Mariano modificada - Aplicación.	51
A-1.	Prueba de no linealidad de Tsay (1998) para diferentes rezagos de la variable de umbral - Aplicación.	55
A-2.	Pruebas de raíz unitaria ADF, PP y KPSS para las series de retornos S&P 500, COLCAP y BOVESPA - Aplicación.	56
A-3.	NAIC de los 5 mejores modelos considerados - Aplicación.	56
A-4.	Estimación puntual del vector de umbrales r con su respectivo intervalo de credibilidad del 95 % - Aplicación.	57
A-5.	Estimación puntual de las matrices autorregresivas correspondientes al régimen $j = 1$ - Aplicación.	59
A-6.	Estimación puntual de las matrices autorregresivas correspondientes al régimen $j = 2$ - Aplicación.	60
A-7.	Estimación puntual del vector de desviaciones estándar escaladas correspondiente al régimen $j = 1$ - Aplicación.	62

A-8.	Estimación puntual del vector de desviaciones estándar escaladas correspondiente al régimen $j = 2$ - Aplicación.	62
A-9.	Estimación puntual del vector del coeficiente de correlación escalado correspondiente al régimen $j = 1$ - Aplicación.	64
A-10.	Estimación puntual del vector del coeficiente de correlación escalado correspondiente al régimen $j = 2$ - Aplicación.	65
A-11.	Estimación puntual de los grados de libertad ν con su respectivo intervalo de credibilidad del 95 % - Aplicación.	66
C-1.	Resultados ampliados de la validación cruzada - Aplicación.	73
C-2.	Resultados ampliados de la prueba de Diebold y Mariano modificada - Aplicación.	75
C-3.	Modelo ARIMA(2,0,3) estimado para S&P 500 - Aplicación.	76
C-4.	Prueba Ljung-Box S&P 500 - Aplicación.	77

1. Introducción

El análisis de series de tiempo a través de modelos no lineales permite trabajar con características más ajustadas a la realidad que los modelos lineales no permiten, como ciclos asimétricos, irreversibilidad temporal, multimodalidad y agrupamiento de volatilidad (Calderón y Nieto, 2017). Una clase de modelos no lineales son los modelos de cambio de régimen, donde los datos se caracterizan como pertenecientes a regímenes o estados diferentes y recurrentes. Esta clasificación permite que las características de los datos de series temporales, incluidas las medias, las varianzas y los parámetros del modelo, cambien entre regímenes. En este contexto, el enfoque Bayesiano de series de tiempo no lineales a través de modelos de cambio de régimen ha tomado mayor relevancia en los últimos años debido a los avances computacionales y la flexibilidad que brindan estos modelos.

Tong (1978; 1990) y Tong y Lim (1980) introducen los modelos autorregresivos de umbrales (TAR), mientras que, Tsay (1998) presenta los modelos TAR multivariados (MTAR) junto a una prueba para detectar la no linealidad con presencia de un umbral en un vector de series de tiempo. El trabajo de Chen y Lee (1995) es el primer ejemplo de estimación Bayesiana de un modelo de umbrales autorregresivos. La estructura de los modelos TAR y MTAR en conjunto con la metodología Bayesiana permite a los investigadores proponer aplicaciones y avances teóricos diversos. Resultado de esta metodología, los impactos de los modelos TAR y MTAR han sido importantes como lo reseña Tong (2015) a nivel general y Hansen (2011) en los campos de finanzas y economía.

La flexibilidad de esta clase de modelos permite crear variaciones para analizar problemas específicos, un ejemplo reciente son los trabajos de Calderón y Nieto (2017; 2021) donde se presenta la estimación y pronóstico de un modelo MTAR desde el enfoque Bayesiano con datos faltantes y una aplicación empírica en el campo hidrológico. Estos autores expanden al caso multivariado la metodología de Nieto (2005) quien analiza series de tiempo univariadas con datos faltantes mediante un modelo TAR. Por otra parte, Espinosa y Nieto (2016) muestran que es posible representar el efecto de apalancamiento en series financieras a través de un modelo TAR y son susceptibles a ser adaptados para medir el riesgo de carteras financieras. También es posible agregar un componente estacional a los modelos de umbrales, específicamente, González y Nieto (2020) proponen un modelo autorregresivo de umbrales estacionales multiplicativos con entrada exógena (TSARX) desde el enfoque Bayesiano y lo aplican en el análisis de la variación mensual de la tasa de desempleo en Colombia.

De igual manera que en el campo teórico, los avances computacionales han facilitado aplicaciones prácticas más amplias. Resalta el trabajo de Li y Tong (2016), donde se propone el algoritmo de búsqueda de submuestra anidada para producir estimaciones más rápidas y confiables. El algoritmo reduce el número necesario de operaciones de mínimos cuadrados para estimar un modelo TAR, lo que conlleva a menores costos computacionales. Asimismo, Nieto, Zhang y Li (2013) proponen el algoritmo de cadena de Markov de salto reversible Monte Carlo para identificar un modelo TAR y estimar sus parámetros. El algoritmo puede saltar entre modelos con diferentes espacios de parámetros dimensionales, al tiempo que conserva un equilibrio detallado que garantiza la distribución límite correcta.

La mayoría de las aproximaciones analíticas mencionadas anteriormente asumen una distribución normal del proceso de error, sin embargo, muchas veces este supuesto no se cumple en la realidad; y más aún en aplicaciones financieras y económicas. Bartkowiak (2007) y Wong, Chan y Kam (2009) demuestran que la distribución t-Student es apropiada para modelar fenómenos financieros ya que sus colas más pesadas pueden tener en cuenta rendimientos extremos observados en los mercados bursátiles. Además, Chiu, Mumtaz, y Pintér (2017) exponen que los pronósticos de modelos de series de tiempo que usan la distribución t-Student tienen una capacidad predictiva superior a los modelos que asumen normalidad.

Teniendo en cuenta este contexto, Zhang y Nieto (2015) proponen una metodología para la estimación y pronóstico de un modelo TAR cuando el proceso de ruido sigue una distribución t-Student desde el enfoque Bayesiano. La metodología de Zhang y Nieto (2015) es extendida al caso multivariado por Romero y Calderón (2021) para la estimación de los parámetros no estructurales. Mientras que Ibañez y Calderón (2020) se enfocan en la estimación de los parámetros estructurales del mismo modelo. Con el panorama anterior, el siguiente paso en el estudio de los modelos MTAR cuando el proceso de ruido sigue una distribución t-Student multivariada es el cálculo de sus pronósticos. En este ámbito resaltan los trabajos de Nieto (2008) y Calderón y Nieto (2021), donde presentan los pronósticos de los modelos TAR y MTAR respectivamente, a través de la computación de las distribuciones predictivas cuando el proceso de ruido sigue una distribución normal.

Con el panorama de investigación expuesto, el objetivo general de este documento es desarrollar un método que permita obtener los pronósticos basados en un modelo MTAR con distribución de error t-Student multivariada desde el enfoque Bayesiano. Los objetivos específicos del documento incluyen hallar la distribución predictiva para la predicción futura del proceso de salida basado en el modelo MTAR con distribución de error t-Student, proponer un algoritmo que permita obtener el pronóstico puntual y los intervalos de predicción para el vector de salida, implementar un estudio de simulación para validar la metodología propuesta, realizar una aplicación de pronóstico del modelo MTAR estimado y comparar los

resultados con los pronósticos obtenidos por modelos alternativos.

Para ello, se encuentra la distribución predictiva Bayesiana que incluye la incertidumbre sobre los verdaderos valores de los parámetros del modelo MTAR. El procedimiento planteado se basa en la obtención de muestras de la distribución predictiva para obtener el pronóstico puntual e intervalos creíbles del proceso de salida. El desempeño del algoritmo planteado se verifica a través de un estudio de simulación basado en tres modelos en donde se calcula el porcentaje de veces en que los valores verdaderos del proceso de salida se encuentran dentro del intervalo de predicción del 95 % de la distribución predictiva. Posteriormente se presenta una aplicación a un conjunto de series financieras donde se obtienen los pronósticos de los retornos de los índices Bovespa y Colcap usando como variable umbral los retornos del índice Standard and Poor's 500 y se comparan los pronósticos con los obtenidos por un modelo MTAR con distribución de error normal multivariada.

El documento se divide en cuatro capítulos, además de introducción y conclusión. En el primer capítulo se presentan los aspectos teóricos del modelo MTAR y su estimación Bayesiana a través de métodos de cadenas de Markov Monte Carlo (MCMC). En el segundo se presenta la distribución predictiva Bayesiana para pronosticar el vector de salida. En el tercer capítulo se realizan simulaciones para verificar el comportamiento de la metodología propuesta en el capítulo dos. Por último, se realiza una aplicación bursátil de la metodología propuesta.

2. Aspectos teóricos

2.1. Estadística Bayesiana

La característica principal de los métodos Bayesianos es el uso de probabilidades para cuantificar la incertidumbre de las inferencias basadas en datos estadísticos. La estadística Bayesiana ajusta un modelo de probabilidad a un conjunto de datos y resume los resultados con una distribución de probabilidad sobre los parámetros del modelo y sobre cantidades no observadas, como predicciones para nuevas observaciones. Gelman et al. (2004) muestra que el análisis Bayesiano se puede dividir en tres fases:

1. *Modelo de probabilidad completo*: se establece una distribución de probabilidad conjunta para todas las cantidades observables y no observables en un problema. El modelo debe ser consistente con el problema y datos a tratar.
2. *Distribución a posteriori*: se calcula e interpreta la distribución de probabilidad condicional de las cantidades no observadas de interés final dados los datos observados.
3. *Evaluación del ajuste del modelo*: se evalúa el ajuste del modelo a los datos, la sensibilidad a los supuestos del paso uno y se decide si las conclusiones que se obtienen a partir de las distribuciones a posteriori dan respuestas razonables al problema tratado.

Sea θ un vector de parámetros desconocidos, y los datos observados y \tilde{y} las cantidades desconocidas, pero potencialmente observables. Para realizar enunciados de probabilidad sobre θ dado y , se debe plantear un modelo con una distribución de probabilidad conjunta para θ y y . Esta distribución de probabilidad conjunta se puede escribir como el producto de la distribución a priori $p(\theta)$ y la distribución de muestreo de datos $p(y|\theta)$:

$$p(\theta|y) = p(\theta)p(y|\theta). \quad (2-1)$$

Al usar la regla de Bayes o la propiedad básica de probabilidad condicional, se obtiene la distribución a posteriori de θ que tiene en cuenta la incertidumbre de θ dado los datos:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{\int_{\Theta} p(\theta)p(y|\theta)d\theta}, \quad (2-2)$$

donde $p(y|\theta)$ es la función de verosimilitud de y . En el contexto de la estadística Bayesiana se debe elegir un criterio para generar una estimación puntual óptima de la distribución a

posteriori de θ . En esta dirección, una función de pérdida $l(\theta, a)$ mide el costo no negativo incurrido de decidir que el parámetro tiene el valor a , cuando en realidad es igual a θ (Bickel y Doksum, 2015). La función de pérdida cuadrática es la función de pérdida más común y se define como:

$$l(\theta, a) = (q(\theta) - a)^2, \quad (2-3)$$

donde $q(\theta)$ es un parámetro de valor real. Teniendo en cuenta las ecuaciones (2-2) y (2-3), la estimación Bayesiana con una función de pérdida cuadrática es la solución al siguiente problema de minimización:

$$\begin{aligned} \min_{\hat{\theta}} c(\hat{\theta}) &= E_{\theta|y}[l(\hat{\theta}, \theta)] \\ \text{con } E_{\theta|y}[l(\hat{\theta}, \theta)] &= \int_{\Theta} l(\hat{\theta}, \theta)p(\theta|y)d\theta, \end{aligned} \quad (2-4)$$

donde c es una constante conocida y $E_{\theta|y}$ es la estimación de la función a posteriori de θ que minimiza la función de pérdida cuadrática.

Con la misma lógica de la ecuación (2-2), dado que los datos y ya han sido observados, se pueden predecir los valores futuros de \tilde{y} a través de la distribución a posteriori predictiva. A posteriori porque es condicional a los valores observados de y , predictiva porque es una predicción para un \tilde{y} (Gelman et al., 2004). Ahora bien, dicha distribución predictiva está dada por:

$$\begin{aligned} p(\tilde{y}|y) &= \int_{\Theta} p(\tilde{y}, \theta|y)d\theta \\ &= \int_{\Theta} p(\tilde{y}|\theta, y)p(\theta|y)d\theta \\ &= \int_{\Theta} p(\tilde{y}|\theta)p(\theta|y)d\theta, \end{aligned} \quad (2-5)$$

donde $p(\tilde{y}|\theta, y)$ es la distribución de \tilde{y} dado θ y y . La segunda y tercera línea muestran la distribución predictiva a posteriori como un promedio de predicciones condicionales sobre la distribución a posteriori de θ . El último paso se deriva del supuesto de independencia condicional de y y \tilde{y} dado θ . En el caso de series de tiempo multivariadas, la decisión Bayesiana consiste en elegir el pronóstico que minimice la pérdida esperada condicionada a la información disponible de $Y_T = (y_1, \dots, y_T)$:

$$E[\mathcal{L}(a, y_{T+1:T+H})|Y_T] = \int \mathcal{L}(a, y_{T+1:T+H})p(y_{T+1:T+H}|Y_T)dy_{T+1:T+H}, \quad (2-6)$$

donde $\mathcal{L}(a, y_{T+1:T+H})$ es una función de pérdida dada. En el caso específico de la función de pérdida cuadrática $(a - y_{T+1:T+H})'(a - y_{T+1:T+H})$, la solución es el valor esperado condicional $a(Y_T) = E(y_{T+1:T+H}|Y_T)$. Por último, es importante exponer los dos tipos de distribuciones

a priori que se pueden aplicar en la inferencia Bayesiana. Las distribuciones a priori informativas implican que existe suficiente información para asociar una distribución al parámetro θ de interés. Sin embargo, Gelman et al. (2004) argumentan que dicha distribución a priori informativa no debe concentrarse demasiado en torno al valor real del parámetro, ya que la información sobre θ contenida en los datos del problema superarán cualquier especificación de probabilidad a priori elegida. Mientras que las distribuciones a priori no informativas representan los casos en los que no se tiene información sobre la distribución del parámetro. La justificación para usar distribuciones no informativas es evitar que las inferencias se vean afectadas por información externa a los datos del problema.

2.2. Métodos de simulación MCMC

Las cadenas de Markov Monte Carlo (MCMC) son métodos de simulación para generar muestras aleatorias de una distribución de probabilidad y así, estimar cantidades o parámetros de interés. Su uso en inferencia Bayesiana ha sido extenso, ya que en la mayoría de los casos la forma analítica de la distribución objetivo no pertenece a una familia de distribuciones conocidas, además de ser compleja y de alta dimensionalidad. La idea principal del método es construir una cadena de Markov tal que su distribución de equilibrio sea la distribución objetivo. Las realizaciones de Monte Carlo de dicha cadena de Markov se pueden usar como muestras de la distribución objetivo. Gelman et al. (2004) argumentan que el éxito de los métodos MCMC se basan en la mejora de las distribuciones aproximadas que se da en cada paso de la simulación, con lo que el método converge a la distribución objetivo.

Una cadena de Markov es un proceso estocástico X_1, \dots, X_t, \dots con densidad condicional igual a:

$$p(X_t | X_{t-1}, X_{t-2}, \dots) = p(X_t | X_{t-1}). \quad (2-7)$$

Las cadenas de Markov tienen la propiedad de que X_t sólo depende del estado inmediatamente anterior del sistema X_{t-1} . Una cadena de Markov es homogénea si está completamente especificada por la densidad de probabilidad de transición de un paso $p(\cdot | x) = p(X_{t+1} | X_t = x)$ para todo t . En otras palabras, las probabilidades son las mismas en cada paso y no dependen del tiempo en que se consideren. De acuerdo con Tsay y Chen (2019), los métodos MCMC se basan en el siguiente resultado: para una cadena de Markov ϕ -irreducible y aperiódica con densidad de transición $p(x_t | x_{t-1})$, si es invariante respecto a $\pi(x)$ entonces la cadena de Markov es ergódica, con $\pi(x)$ como su distribución de equilibrio.

Una cadena de Markov es irreducible si se puede llegar a cualquier conjunto A desde cualquier otro conjunto B con probabilidad distinta de cero. Es aperiódica si para cualquier conjunto A , el número de pasos necesarios para volver a A no siempre debe ser un múlti-

plo de algún valor k . Además, la cadena es invariante si existe una densidad $\pi(x)$ tal que $\pi(y) = \int p(y|x)\pi(x)dx$. La ergodicidad de la cadena implica el teorema ergódico, según el cual, si una cadena de Markov es ergódica, entonces la ley de los grandes números se cumple, es decir, $\bar{x} \rightarrow E_\pi(x)$ donde $\pi(\cdot)$ es la distribución de equilibrio. Además, empezando con $x_0 \sim g(x)$, la distribución de X_t converge a $\pi(x)$.

Teniendo en cuenta lo anterior, los métodos MCMC construyen una regla de transición $p(y|x)$ de modo que la cadena de Markov resultante sea ergódica, con la distribución objetivo $\pi(\cdot)$ como su distribución de equilibrio. Dada una distribución condicional $p(y|x)$ apropiada para muestrear y a partir de x , se comienza con $X_0 = x_0 \in \Omega$, e iterativamente se generan muestras de $x_t \sim p(y|X_{t-1} = x_{t-1}), t = 1, \dots$. Las muestras resultantes x_{K+1}, \dots, x_{K+m} siguen aproximadamente $\pi(x)$. Para reducir el impacto del valor inicial, los primeros valores simulados (iteraciones de *burn-in* $t = 1, \dots, K$) se eliminan. A partir del proceso anterior, $\mu_h = E_\pi(x)$ se estima como:

$$\hat{\mu}_h = \frac{1}{m} \sum_{j=K+1}^{K+m} h(x_j). \quad (2-8)$$

Además de la estimación puntual, se pueden computar los intervalos creíbles, que tienen la capacidad de describir y resumir la incertidumbre relacionada con los parámetros desconocidos que se intentan estimar. Ya que la inferencia Bayesiana devuelve una distribución a posteriori, el intervalo creíble es el rango que contiene un porcentaje particular de valores de la distribución a posteriori. En otras palabras, el intervalo creíble del 95 % es el rango de valores que representan el 95 % de la distribución a posteriori, es decir, hay un 95 % de probabilidad de que la estimación de un parámetro desconocido se encuentre dentro del intervalo, dada la evidencia proporcionada por los datos observados.

Un algoritmo MCMC rara vez se inicializa a partir de su distribución invariable, existe un periodo de calentamiento o quemado en el que el estado de la cadena de Markov depende del punto de partida inicial. Para compensar esto, se descartan las primeras B iteraciones, y se elige un N que sea lo suficientemente grande para que la cadena alcance su régimen estacionario.

2.2.1. Algoritmo de Metropolis-Hastings

El algoritmo Metropolis-Hastings es una técnica popular para construir cadenas de Markov con una distribución invariante dada (Karlsson, 2013). El algoritmo Metropolis-Hastings considera dos distribuciones: la distribución objetivo σ y la distribución condicional propuesta $q(x^*|x)$ de la que se extrae una muestra x^* para el nuevo estado de la cadena de Markov. Si el estado actual de la cadena es x , entonces, de acuerdo con el algoritmo Metropolis-Hastings,

la cadena se mueve a su nuevo estado x^* con probabilidad:

$$P(x, x^*) = \min \left[1, \frac{\sigma(x^*)q(x|x^*)}{\sigma(x)q(x^*|x)} \right]. \quad (2-9)$$

Debski (2010) expone el pseudocódigo del algoritmo Metropolis-Hastings de la siguiente forma:

1. Inicializar con x^0 .
2. Repetir para cada i .
 - Generar un número aleatorio uniforme $u \sim U(0, 1)$.
 - Generar una muestra de prueba $x^* \sim q(x^*|x^i)$.
 - Si $u < P(x^i, x^*) = \min \left[1, \frac{\sigma(x^*)q(x^i|x^*)}{\sigma(x^i)q(x^*|x^i)} \right]$, entonces $x^{i+1} = x^*$, de lo contrario $x^{i+1} = x^i$.
 - $i = i + 1$.
3. Continuar hasta que se genere un número suficiente de muestras de $\{x^i\}$.

El algoritmo Metropolis-Hastings generaliza el algoritmo básico de Metropolis, ya que la distribución condicional propuesta $q(x^*|x)$ ya no necesita ser simétrica, es decir, no es necesario que $q(x^*|x) = q(x|x^*)$ (Gelman et al., 2004). La convergencia del algoritmo Metropolis-Hastings a la distribución invariante se da en el límite de un número infinito de iteraciones. Por lo tanto, siempre existe una probabilidad diferente de cero de que la cadena alcance la etapa invariante después de un número finito de iteraciones. Debski (2010) argumenta que la duración de este período de desequilibrio (*burn-in*) depende de la distribución propuesta $q(x^*|x^i)$, especialmente en el caso de una x multidimensional. El ancho de la distribución condicional propuesta se refiere a la varianza de la distribución que determina la probabilidad de transición entre estados en la cadena. Si $q(x^*|x^i)$ es demasiado delgada, se cae en el riesgo de sobre-muestrear una parte de la cadena y, en consecuencia, solo se muestrea una sección de $\sigma(x)$. Mientras que si $q(x^*|x^i)$ es demasiado ancha, la tasa de rechazo puede ser muy alta debido a un intento de muestrear distintas partes del espacio.

La elección de la distribución propuesta q es determinante para el desempeño de la cadena de Markov y es importante que esté bien adaptada a la distribución a posteriori. Una posibilidad es extraer a $q(x^*|x)$ de una densidad fija $f(x)$, ya que es importante para el rendimiento de la cadena que la distribución propuesta se adapte bien a la distribución a posteriori en todo el espacio de parámetros (Karlsson, 2013). La segunda alternativa se da cuando los pasos de la distribución propuesta q siguen una caminata aleatoria, en este caso el algoritmo implementado se denomina como Metropolis-Hastings con caminata aleatoria. En dicho algoritmo, la transición entre los estados de la cadena se realiza utilizando una caminata aleatoria, para pasar al siguiente estado de la cadena se agrega un término de ruido aleatorio extraído de una distribución normal al estado actual.

2.2.2. Muestreador de Gibbs

El Muestreador de Gibbs se puede emplear cuando los estados de la cadena x^i son vectores de dos o más dimensiones y se pueden extraer muestras de las distribuciones condicionales completas de todos los parámetros o bloques de ellos. Sea $x \in \mathcal{X}$, $x = (x_1, x_2, \dots, x_N)$ donde N es la dimensión de \mathcal{X} y el conjunto de vectores con $k = 1, 2, \dots, N$ con el componente k -ésimo eliminado es $x_{-k} = (x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_N)$. Se asume que se conocen todas las probabilidades condicionales de $\sigma(x_j, x_{-j})$ y la probabilidad para un cambio del j -ésimo componente en un solo movimiento de la cadena desde el estado actual x_{-j}^i es:

$$q_j(x^* | x^i) = \begin{cases} \sigma(x_j^* | x_{-j}^i) & \text{si } x_j^* = x_{-j}^i \\ 0 & \text{de lo contrario.} \end{cases} \quad (2-10)$$

La distribución propuesta en el algoritmo Metropolis-Hastings se puede tomar como el producto con respecto a todas las componentes de x :

$$q(x^* | x^i) = \prod_j q_j(x^* | x^i), \quad (2-11)$$

con lo que la probabilidad de cambio de estado de Metropolis-Hastings es:

$$\begin{aligned} P(x^i, x^*) &= \min \left[1, \frac{\sigma(x^*)q(x^i | x^*)}{\sigma(x^i)q(x^* | x^i)} \right] \\ &= \min \left[1, \prod_j \frac{\sigma(x^*)\sigma(x_j^* | x_{-j}^i)}{\sigma(x^i)\sigma(x_j^* | x_{-j}^*)} \right] \\ &= \min \left[1, \prod_j \frac{\sigma(x_{-j}^*)}{\sigma(x_{-j}^i)} \right] = 1. \end{aligned} \quad (2-12)$$

Por lo tanto, todas las muestras recién generadas siempre se aceptan dado que es posible muestrear directamente todas las distribuciones condicionales. De esta manera, el muestreador de Gibbs es un caso especial del algoritmo de Metropolis-Hastings en que la relación de aceptación es siempre uno. Debski (2010) expone el pseudocódigo del muestreador de Gibbs de la siguiente forma:

1. Inicializar con x^0 .
2. Repetir para cada i .
 - Generar una muestra de $x_1^{i+1} \sim \sigma(x_1 | x_2^i, x_3^i, \dots, x_n^i)$.
 - Generar una muestra de $x_2^{i+1} \sim \sigma(x_2 | x_1^i, x_3^i, \dots, x_n^i)$.
 - ⋮
 - Generar una muestra de $x_n^{i+1} \sim \sigma(x_n | x_1^i, x_2^i, \dots, x_{n-1}^{i+1})$.

- $i = i + 1$.

3. Hasta que se genere el número requerido de muestras.

El muestreador de Gibbs es adecuado para el cálculo Bayesiano, ya que se basa en las distribuciones condicionales de subconjuntos del vector de parámetros. Resulta más sencillo generar números aleatorios a partir de las distribuciones a posteriori condicionales de un subconjunto del vector de parámetros, incluso si dicha distribución no es estándar (Karlsson, 2013). El muestreador de Gibbs es sencillo de implementar y la forma del muestreador se deriva directamente del modelo cuando las distribuciones condicionales son distribuciones conocidas.

2.2.3. Comprobación de convergencia

Se requieren herramientas de diagnóstico MCMC para decidir la convergencia de las cadenas al estado de estacionariedad (Karlsson, 2013). Aunque en general, cuanto más larga sea la cadena, mejores estimaciones produce, en la práctica es recomendable utilizar reglas para un uso eficiente de los recursos computacionales. A continuación, se exponen los dos diagnósticos de MCMC más usuales que pueden ser empleados para decidir la convergencia de la cadena de Markov a una distribución estacionaria.

Criterio de Geweke

Geweke (1992) propone un diagnóstico de convergencia para cadenas de Markov basado en una prueba de diferencia de las medias de la primera parte n_a y última parte n_b de una cadena de Markov (usualmente el primer 10 % y el último 50 %). Si las muestras se extraen de la distribución estacionaria de la cadena, las dos medias son iguales y el estadístico de Geweke tiene una distribución normal asintóticamente estándar. En otras palabras, si la cadena ha convergido después de m pasos, la distribución de las muestras $m + 1, \dots, m + n_a$ es igual a la distribución de las muestras al final de la cadena. La estadística de prueba es Z-score estándar, es decir, la diferencia entre las medias de las dos muestras dividida por su error estándar estimado:

$$Z_n = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\text{Var}(\bar{\theta}_A - \bar{\theta}_B)}}. \quad (2-13)$$

El error estándar se estima a partir de la densidad espectral en cero y, por lo tanto, tiene en cuenta cualquier autocorrelación.

Criterio de Gelman y Rubin

Gelman y Rubin (1992) proponen un enfoque general para monitorear la convergencia de los resultados de los métodos MCMC en el que se ejecutan $m > 1$ cadenas paralelas con valores iniciales que tienen sobre-dispersión en relación con la distribución a posteriori. Se

basa en una comparación de varianzas dentro de la cadena y entre cadenas, y es similar a un análisis clásico de varianza. Karlsson (2013) expone el criterio de Gelman y Rubin (1992) para la estimación Bayesiana de modelos de vectores autorregresivos. Sea $\theta_i^{(j)}$ la salida de la cadena i para m cadenas ejecutadas con $n = R - B$ pasos desde la etapa de calentamiento, la varianza entre cadenas (B) y dentro de cadenas (W) se define como:

$$B = \frac{n}{m-1} \sum_{i=1}^m (\bar{g}_i - \bar{g})^2, \quad W = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (g(\theta_i^{(j)}) - \bar{g}_i)^2$$

$$\text{con } \bar{g}_i = \frac{1}{n} \sum_{j=1}^n g(\theta_i^{(j)}), \quad \bar{g} = \frac{1}{m} \sum_{i=1}^m \bar{g}_i. \quad (2-14)$$

La falla de convergencia o la convergencia en diferentes distribuciones estacionarias después del período de calentamiento se puede identificar mediante la variación entre cadenas, que es mayor que la variación dentro de la cadena. Si las cadenas han convergido después de B muestras, se tienen dos estimaciones no sesgadas de la varianza, $V = (1 - 1/n)W + B/n$ y W . La primera tiende a sobrestimar la varianza si no se ha logrado la convergencia (la variación entre cadenas es grande) y la segunda tiende a subestimar la varianza (las cadenas no han tenido tiempo de explorar el espacio de parámetros completo). Karlsson (2013) expone que el diagnóstico de convergencia es $r = \sqrt{V/W}$, y además sugiere como regla general para aceptar la convergencia tener $r < 1.2$ para todas las cantidades monitoreadas.

2.3. Precisión de los pronósticos

En esta sección se introducen las métricas elegidas para evaluar y comparar el desempeño de pronóstico del modelo propuesto con un modelo alternativo. Para ello, se comparan los pronósticos obtenidos con base en el error cuadrático medio multivariado (MSE) y error cuadrático medio relativo multivariado (RelMSE) en cada horizonte de pronóstico a partir de la metodología de validación cruzada de series de tiempo o método *Rolling*. Además se realiza la prueba de Diebold y Mariano (1995) para determinar estadísticamente si el pronóstico de un modelo es más preciso que el de otro.

2.3.1. Validación cruzada de series de tiempo

Hyndman y Athanasopoulos (2018) argumentan que la precisión de los pronósticos se debe determinar considerando qué tan bien funciona un modelo con datos nuevos que no se usaron cuando se ajustó el modelo. Para ello, se separan los datos disponibles en datos de entrenamiento y de prueba, donde los primeros se utilizan para estimar los parámetros del modelo de interés y los segundos para evaluar su precisión.

En el procedimiento de validación cruzada de series de tiempo o método *Rolling*, hay una serie de conjuntos de prueba, del cual su correspondiente conjunto de entrenamiento consta solo de observaciones que ocurrieron antes de las observaciones que forman el conjunto de prueba (Hyndman y Athanasopoulos, 2018). La longitud de los conjuntos de prueba depende de qué tan lejos se desea pronosticar y la precisión del pronóstico se computa con una métrica basada en los errores de pronóstico del conjunto de prueba como el error cuadrático medio.

Para llevar a cabo el proceso anterior, el conjunto de datos se divide inicialmente en una muestra de entrenamiento y una muestra de prueba. Posteriormente, se ajusta el modelo utilizando la muestra de entrenamiento y se realizan predicciones de h pasos adelante para la muestra de prueba (Zivot y Wang, 2005). Dado que el conjunto de datos contiene los verdaderos valores del conjunto de prueba, se pueden formar errores de predicción de h pasos adelante. A continuación, la muestra de entrenamiento avanza un paso y el ejercicio de estimación y predicción se repite hasta que ya no es posible hacer más pasos de predicción. Luego, se resumen las propiedades de predicción de h pasos adelante y se utilizan para evaluar la precisión del modelo propuesto.

La longitud de la muestra de entrenamiento se mantiene constante en todo el proceso, es decir, a medida que se incluyan una nueva observación, se elimina la primera. De esta forma, para ventanas de muestras de estimación de longitud n con $n < T$, $\{Y_n\}$, $\{X_n\}$ y $\{Z_n\}$ son procesos estocásticos tales que $Y_n = (Y_{1n}, \dots, Y_{kn})'$, $X_n = (X_{1n}, \dots, X_{vn})'$ y $\{Z_n\}$ un proceso univariado de tamaño $(n \times 1)$. Las n observaciones de $\{Y_n\}$, $\{X_n\}$ y $\{Z_n\}$ son los n valores más recientes desde el tiempo $t - n + 1$ a t (Zivot y Wang, 2005). La suma de errores al cuadrado (SEC) vía *Rolling* para cada horizonte de pronóstico h donde Y contiene los verdaderos valores del proceso y $\hat{Y}_{t+h|t}$ contiene los pronósticos obtenidos a partir de un modelo en función de sus propios rezagos, $\{X_n\}$ y $\{Z_n\}$, se define como:

$$\text{SEC}(h) = \text{tr}((Y_{t+h} - \hat{Y}_{t+h|t})'(Y_{t+h} - \hat{Y}_{t+h|t})), \quad (2-15)$$

donde t denota el momento final de la muestra de estimación y h el horizonte de pronóstico. En el proceso *Rolling* se generan errores de predicción de h pasos adelante a medida que la muestra de prueba avanza, el conjunto de errores resultantes para cada horizonte de pronóstico (h) considerado puede ser utilizado para computar distintas métricas de precisión de los pronósticos. Teniendo en cuenta lo anterior, el error cuadrático medio (MSE(h)) multivariado vía *Rolling* del horizonte de pronóstico h se define como:

$$\text{MSE}(h) = \text{SEC}(h) \times \frac{1}{T - n - h + 1}. \quad (2-16)$$

A partir del MSE(h) se puede calcular el error cuadrático medio relativo (RelMSE(h)) que es la razón entre el error cuadrático medio del modelo propuesto y el error cuadrático medio de un modelo alternativo o de referencia con el que se desee comparar la precisión de los

pronósticos obtenidos. Su principal ventaja es la facilidad de interpretación, esta medida relativa mide la posible mejora del método de pronóstico propuesto en relación con el método de pronóstico de referencia. Cuando el RelMSE es menor a uno, la precisión de los pronósticos del método propuesto es mayor que el método de referencia. Mientras que si el RelMSE es mayor a 1, el método propuesto tiene una menor precisión que el método de referencia.

2.3.2. Prueba de Diebold y Mariano

Para complementar los resultados obtenidos por el cálculo del MSE y RelMSE se realiza la prueba de Diebold y Mariano (1995). Esta prueba utiliza los errores de pronóstico h pasos adelante para determinar estadísticamente si el pronóstico de un modelo es más preciso que el de otro. Sea $\hat{\varepsilon}_{t+h|t}^{(m1)}$ y $\hat{\varepsilon}_{t+h|t}^{(m2)}$ los errores de pronóstico de h pasos adelante de dos modelos en competencia $m1$ y $m2$, N el número de pronósticos de h pasos, d_t se define de la siguiente forma:

$$d_t = [\hat{\varepsilon}_{t+h|t}^{(m1)}]^2 - [\hat{\varepsilon}_{t+h|t}^{(m2)}]^2, \quad (2-17)$$

donde d_t se nombra como pérdida diferencial. con lo que es posible definir definir \bar{d} :

$$\bar{d} = \frac{1}{N} \sum_{t=1}^N d_t. \quad (2-18)$$

Para $N > k \geq 1$, la autocovarianza γ_k de d_t de orden k se puede estimar como:

$$\hat{\gamma}_k = \frac{1}{N} \sum_{t=k+1}^N (d_t - \bar{d})(d_{t-k} - \bar{d}). \quad (2-19)$$

Con lo anterior, el estadístico de Diebold y Mariano (1995) tiene la siguiente estructura:

$$DM = \frac{\bar{d}}{\sqrt{[\gamma_0 + 2 \sum_{k=1}^{h-1} \hat{\gamma}_k/N]}}. \quad (2-20)$$

La hipótesis nula indica un MSE igual para los dos pronósticos, es decir, igual precisión en los pronósticos, frente a la alternativa de que un modelo tiene un MSE más pequeño, es decir, mejor precisión en los pronósticos frente a su comparativo. Bajo la hipótesis nula, el estadístico de prueba DM sigue una distribución normal estándar (Zivot y Wang, 2005). Por lo tanto, hay una diferencia significativa entre los pronósticos si $|DM| > z_{crit}$ donde z_{crit} es el valor crítico de dos colas para la distribución normal estándar. Vale la pena aclarar que el supuesto clave para usar la prueba de Diebold-Mariano es que la serie de pérdida diferencial d_i es estacionaria. La prueba de Diebold y Mariano (1995) tiende a rechazar la hipótesis

nula con demasiada frecuencia para muestras pequeñas. Debido a ello, Harvey, Leybourne y Newbold (1997) proponen la siguiente modificación al estadístico DM:

$$HLN = \left[\frac{N + 1 - 2h + N^{-1}(h - 1)}{N} \right]^{1/2} \times DM, \quad (2-21)$$

donde DM es el estadístico DM original de la ecuación (2-20) y el estadístico HLN sigue una distribución t-Student con $N - 1$ grados de libertad. Harvey, Leybourne y Newbold (1997) demuestran que el rendimiento del estadístico de prueba HLN es mejor que el del estadístico DM, especialmente en muestras pequeñas. Además tiene un mejor rendimiento cuando los errores de pronóstico están autocorrelacionados o tienen una distribución no normal.

2.4. Modelo multivariado autorregresivo de umbrales con distribución de error t-Student multivariada

2.4.1. Especificación del modelo MTAR

El modelo MTAR representa la relación entre tres procesos estocásticos multidimensionales $\{Y_t\}$, $\{X_t\}$ y $\{Z_t\}$, llamados procesos de salida, covariables y umbral, respectivamente. Ahora bien, sea $\{Y_t\}$ y $\{X_t\}$ procesos estocásticos tales que $Y_t = (Y_{1t}, \dots, Y_{kt})'$ y $X_t = (X_{1t}, \dots, X_{vt})'$ y sea $\{Z_t\}$ un proceso univariado. $\{Y_t\}$ sigue un proceso MTAR con variable de umbral $\{Z_t\}$ si:

$$Y_t = \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} Y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} X_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} Z_{t-i} + \Sigma_{(j)}^{1/2} \varepsilon_t \text{ si } r_{j-1} < Z_t \leq r_j. \quad (2-22)$$

Si $\{Y_t\}$ sigue el proceso descrito en la ecuación (2-22), el modelo MTAR resultante se denota como:

$$MTAR(l, p_1, \dots, p_l, q_1, \dots, q_l, d_1, \dots, d_l), \quad (2-23)$$

donde $j = 1, \dots, l$, con $l \geq 2$ el número de regímenes, y $r_0, r_1, \dots, r_{l-1}, r_l$ los umbrales del modelo que definen los regímenes siendo tales que $-\infty = r_0 < r_1 < \dots < r_{l-1}, r_l = \infty$. $\{\varepsilon_t\}$ es un proceso independiente e idénticamente distribuido, y por cada t , ε_t sigue una distribución t-Student multivariada con vector de medias 0, matriz de covarianza I_k de orden k , v grados de libertad y es mutuamente independiente de $\{X_t\}$ y $\{Z_t\}$. Para $j = 1, \dots, l$, los parámetros no estructurales del modelo son las matrices $\phi_0^{(j)}$ de tamaño $k \times 1$, $\phi_i^{(j)}$ con $i = 0, 1, \dots, p_j$ de tamaño $k \times k$, $\beta_i^{(j)}$ con $i = 0, 1, \dots, q_j$ de tamaño $k \times v$, $\delta_i^{(j)}$ con $i = 0, 1, \dots, d_j$ de tamaño $k \times 1$ y $\Sigma_{(j)}^{1/2}$ de tamaño $k \times k$. De esta forma, el vector de parámetros no estructurales es:

$$\theta_{yms} = (\theta'_1, \dots, \theta'_l, \text{vec}(\Sigma)')', \quad (2-24)$$

con $\theta_j = \text{vec}(A_j)$, $A_j = (\phi_0^{(j)}, \phi_1^{(j)}, \dots, \phi_{p_j}^{(j)}, \beta_1^{(j)}, \dots, \beta_{q_j}^{(j)}, \delta_1^{(j)}, \dots, \delta_{d_j}^{(j)})_{k \times \eta_j}$ y $\eta_j = (1 + k.p_j + v.q_j + d_j)$ para $j = 1, \dots, l$, $\Sigma = (\Sigma_1, \dots, \Sigma_l)$. Los números enteros p_j , q_j y d_j con $j =$

$1, \dots, l$ representan los órdenes autorregresivos de cada régimen, los valores de los umbrales se denotan como $r = (r_1, \dots, r_{l-1})'$ y el número de regímenes como l . Los números enteros p_j, q_j, d_j, r y l con $j = 1, \dots, l$ se definen en conjunto como los parámetros estructurales del modelo y se representan como:

$$\theta_{ys} = (p_1, \dots, p_l, q_1, \dots, q_l, d_1, \dots, d_l, r', l)'. \quad (2-25)$$

Teniendo en cuenta lo anterior, el modelo MTAR presentado tiene el siguiente vector de parámetros completos:

$$\theta_y = (\theta_{yus}, \theta_{ys})'. \quad (2-26)$$

Ahora bien, sea $U_t = (Z_t, X_t)'$ el proceso estocástico que incluye el proceso de umbrales y de variables exógenas. Para efectos de predicción, se puede asumir que $\{U_t\}$ es una cadena de Markov homogénea de orden b -ésimo de dimensión $(v + 1)$, donde b es un número entero mayor que cero con densidad estacionaria $f_u(\cdot)$ y densidad de kernel de transición $f_u(\cdot|\cdot)$. Además, no hay retroalimentación de $\{Y_t\}$ a $\{U_t\}$, es decir, la cadena $\{U_t\}$ es exógena, con esto es posible obtener los pronósticos de $\{U_t\}$ de tal forma que no dependan de $\{Y_t\}$. Sea $\theta = (\theta_y, \theta_u)$ el vector de parámetros completos del modelo compuesto por el vector de parámetros θ_y del modelo MTAR y el vector de parámetros θ_u la cadena de Markov, se asume que el proceso probabilístico de la cadena de Markov $u = (u_1, \dots, u_T)$ no depende de los parámetros θ_y , y de manera similar, se asume que la densidad conjunta de $y = (y_1, \dots, y_T)$ condicional a u y (θ_y, θ_u) no depende de θ_u . Teniendo en cuenta lo anterior, la función de verosimilitud de θ se puede escribir como:

$$L(\theta|y, u) = f(y, u|\theta_y, \theta_u) = f(u|\theta_u)f(y|u, \theta_y). \quad (2-27)$$

El modelo MTAR expuesto es lineal a trozos en el espacio de $Z_t, \dots, Z_{t-d}, X_{t-1}, \dots, X_{t-q}, Y_{t-1}, \dots, Y_{t-p}$ con $p = \max\{p_1, \dots, p_l\}$ pero es no lineal en el tiempo. La utilidad de los modelos de umbrales se debe al hecho de que las funciones lineales definidas por regímenes proporcionan típicamente una aproximación más simple y fácil de manejar a series no lineales que un sistema no lineal con mayor cantidad de parámetros.

2.4.2. Pruebas de no linealidad

El primer paso para ajustar el modelo MTAR presentado en la sección anterior, consiste en determinar si $\{Y_t\}$ sigue un proceso no lineal con umbral. En este caso, se implementa la prueba desarrollada por Lo y Zivot (2001), que es la extensión multivariada del test de linealidad desarrollado por Hansen (1997, 1999). El método de Hansen para probar la linealidad en modelos TAR univariados basados en pruebas de hipótesis anidadas se puede extender para probar la linealidad en casos multivariados. La prueba de Lo y Zivot (2001) tiene la ventaja de basarse en la forma específica del modelo de umbrales bajo la hipótesis alternativa, a diferencia de procedimiento no paramétricos como la prueba de Tsay (1998). De esta forma, en el caso multivariado se contrastan las siguientes hipótesis:

- i H_0 : los datos siguen un modelo vectores autorregresivos (VAR) lineal.
- ii H_a : los datos siguen un modelo MTAR con l regímenes.

Para realizar el test, Lo y Zivot (2001) proponen el estadístico Sup-LR (supremo de un conjunto de estadísticas de razón de verosimilitud) computado de la siguiente forma (que es equivalente al sup-Wald o al supremo de un conjunto de estadísticas de Wald):

$$LR_{1l} = T(Ln|\bar{\Sigma}| - Ln|\bar{\Sigma}_l|), \quad (2-28)$$

donde $l > 1$ corresponde al número de regímenes y $\bar{\Sigma}$ y $\bar{\Sigma}_l$ son las matrices de covarianzas de los residuos de los modelos $VAR(P)$ y $MTAR(l)$, respectivamente. La distribución de LR_{1l} no es estándar, por lo que se usa el procedimiento Bootstrap implementado por Hansen (1997, 1999) para computar los p-valores para varias pruebas de linealidad basadas en el estadístico LR_{1l} . Para complementar los resultados de la prueba anterior, se realiza la prueba de no linealidad de Tsay (1998) cuyo objetivo es detectar la no linealidad de tipo umbral de $\{Y_t\}$ asumiendo que p, q y d son conocidos. La hipótesis nula enuncia que $\{Y_t\}$ es lineal contra la alternativa en la que $\{Y_t\}$ sigue un modelo MTAR. Sea $h = \max\{p, d, q\}$, el vector $W_t = (1, y'_{t-1}, \dots, y'_{t-p}, x'_{t-1}, \dots, x'_{t-p})$ de tamaño $pk + qv + 1$ y la matriz no conocida Φ . Si la hipótesis nula se cumple, el modelo colapsa a:

$$Y_t' = W_t' \Phi + \varepsilon_t', \quad (2-29)$$

para $t = h + 1, \dots, n$ y valores independientes de la variable de umbral Z . Sea $S = \{z_{h+1-d}, \dots, z_{n-d}\}$ el conjunto de valores de Z_{t-d} y $z_{(i)}$ el elemento i -ésimo más pequeño de la estadística de orden de S . La regresión ordenada basada en el orden creciente de la variable umbral es:

$$Y'_{t(i)+d} = W'_{t(i)+d} \Phi + \varepsilon'_{t(i)+d}, \quad (2-30)$$

donde $i = 1, \dots, n - h$ y $\hat{\Phi}_m$ es el estimador de mínimos cuadrados de Φ basado en las primeras m observaciones. Además, se considera la regresión $\hat{\eta}'_{t(l)+d} = W'_{t(l)+d} \Psi + \hat{\varepsilon}'_{t(l)+d}$ con $l = m_0 + 1, \dots, n - h$. Tsay (1998) propone el siguiente estadístico para probar la hipótesis $H_0 : \Psi = 0$ contra la alternativa $H_a : \Psi \neq 0$:

$$C(d) = [n - h - m_0 - (kp + vq + 1)][ln|S_0| - ln|S_1|],$$

$$\text{con } S_0 = \frac{1}{n - h - m_0} \sum_{l=m_0+1}^{n-h} \hat{\eta}_{t(l)+d} \hat{\eta}'_{t(l)+d} Y \quad (2-31)$$

$$S_1 = \frac{1}{n - h - m_0} \sum_{l=m_0+1}^{n-h} \hat{\varepsilon}_{t(l)+d} \hat{\varepsilon}'_{t(l)+d},$$

donde $\hat{\varepsilon}'_{t(l)+d}$ son los residuales de mínimos cuadrados de la regresión de $\hat{\eta}_{t(m_0+1)+d}$, $|\cdot|$ es el determinante de una matriz y d indica el rezago de la variable de umbral. Bajo la hipótesis

nula, $\{Y_t\}$ es lineal, y bajo la alternativa es no lineal de tipo umbral. Tsay (1998) demuestra que $C(d)$ se distribuye chi-cuadrado de forma asintótica con $k(pk + qv + q)$ grados de libertad. Por su parte, Nieto y Hoyos (2011) muestran que el estadístico $C(d)$ tiene una potencia mayor que otras pruebas estadísticas similares para probar la misma hipótesis nula.

2.4.3. Función de verosimilitud

Kibria y Joarder (2006) definen la función de densidad de probabilidad de una distribución t-Student k-variada como:

$$f(x) = \frac{\Gamma((\nu + k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(x - \mu)\Sigma^{-1}(x - \mu) \right]^{-(\nu+k)/2}, \quad (2-32)$$

donde Γ es la función Gamma, x es el valor realizado de un vector aleatorio X de tamaño $p \times 1$, Σ es una matriz definida positiva de parámetros de escala de tamaño $p \times p$ y ν son los grados de libertad. La variable aleatoria k-variante X tiene un vector medias μ y matriz de covarianzas $\frac{\nu}{\nu-2}\Sigma$ con $\nu > 2$. A partir de la definición anterior, Kibria y Joarder (2006) enuncian la siguiente propiedad para la distribución t-Student k-variada:

Propiedad 1: Sea X una variable aleatoria distribuida t-Student multivariada con vector de media μ , matriz de covarianza Σ y ν grados de libertad, si A es una matriz no singular y b un vector de constantes, entonces $AX + b$ se distribuye t-Student k-variada con vector de medias $A\mu + b$, ν grados de libertad y matriz de escala $A\Sigma A'$. Los grados de libertad para la distribución de la combinación lineal siguen siendo los mismos.

Ya que $\{\varepsilon_t\}$ es un proceso independiente e idénticamente distribuido, y por cada t , ε_t sigue una distribución t-Student multivariada con vector de medias 0, matriz de covarianza I_k de orden k y ν grados de libertad, a partir de la Propiedad 1, Romero y Calderón (2021) demuestran que:

$$p(y_t|y_1, \dots, y_{t-1}, x, z, \theta_y) \sim t_\nu(\phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} Y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} X_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} Z_{t-i}, \Sigma_{(j)}), \quad (2-33)$$

es decir,

$$p(y_t|y_1, \dots, y_{t-1}, x, z, \theta_y) = \frac{\Gamma((\nu + k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)|\Sigma_{(j)}|^{1/2}} \left[1 + \frac{1}{\nu}\varepsilon_t'\varepsilon_t \right]^{-(\nu+k)/2}, \quad (2-34)$$

Romero y Calderón (2021) demuestran que que la función de verosimilitud basada en el modelo MTAR para $t = p + 1, \dots, T$ y a partir de la factorización de la ecuación (2-33) es

igual a:

$$p(y|u, \theta_y) = \left[\frac{\Gamma((\nu + k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)} \right]^{T-p} \prod_{t=p+1}^T |\Sigma_{(j)}|^{-1/2} \prod_{t=p+1}^T \left[1 + \frac{1}{\nu} \varepsilon_t' \varepsilon_t \right]^{-(\nu+k)/2} \quad (2-35)$$

$$\text{con } \varepsilon_t = \Sigma_{(j)}^{-1/2} \left(y_t - \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} x_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} z_{t-i} \right),$$

donde ν son los grados de libertad del proceso de ruido.

2.4.4. Estimación de parámetros no estructurales

El proceso para estimar los parámetros no estructurales del modelo consiste en extraer muestras de la distribución condicional completa $p(\theta_{y_{ns}}|y_{1:T}, u_{1:T}, \theta_{y_s})$ de los parámetros no estructurales dados los datos de los procesos de salida, covariables y umbral y los parámetros estructurales. En la siguiente exposición se asume que los parámetros estructurales ya han sido estimados o identificados previamente con otros métodos, y se sigue el proceso de estimación propuesto por Romero y Calderón (2021). Los parámetros de interés son independientes a priori y a partir de sus distribuciones a priori individuales se obtienen las distribuciones condicionales completas. Romero y Calderón (2021) proponen la siguiente distribución a priori no informativa para las matrices autorregresivas $\theta_{(j)}$ en el régimen j :

$$p(\theta_j) = K, \quad K \in R^+, \quad (2-36)$$

y a partir de la distribución anterior y la función de verosimilitud del modelo de la ecuación (2-35), demuestran que la distribución condicional completa para $\theta_{(j)}|\theta_{(i)}, i \neq j$ con $j = 1, \dots, l$ no pertenece a una distribución conocida y es:

$$\begin{aligned} p(\theta_{(j)}|\theta_{(i)}, i \neq j, \Sigma, u_{1:T}, y_{1:T}, \theta_{y_{ns}}) &= p(y_{1:T}|u_{1:T}, \theta, \Sigma, \nu) p(\theta_{(j)}) \\ &= \left[\frac{\Gamma((\nu + k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)} \right]^{T-p} \prod_{t=p+1}^T |\Sigma_{(j)}|^{-1/2} \prod_{t=p+1}^T \left[1 + \frac{1}{\nu} \varepsilon_t' \varepsilon_t \right]^{-(\nu+k)/2} \\ &\propto \prod_{\{t:j_t=j\}} \left[1 + \frac{1}{\nu} \varepsilon_t' \varepsilon_t \right]^{-(\nu+k)/2} \\ &\propto \prod_{\{t:j_t=j\}} \left[1 + \frac{1}{\nu} (Y_t - (W_{t,j}' \otimes I_k) \theta_j)' \Sigma_{(j)}^{-1} (Y_t - (W_{t,j}' \otimes I_k) \theta_j) \right]^{-(\nu+k)/2}, \end{aligned} \quad (2-37)$$

donde $w_{t,j} = (1, y'_{t-1}, \dots, y'_{t-p_j}, x_{t-1}, \dots, x_{t-q_j}, z_{t-1}, \dots, z_{t-d_j})'_{\eta_j \times 1}$ y $W_j = (w_{t_{1,j}}, \dots, w_{t_{N_j,j}})_{\eta_j \times N_j}$ y $t_{1,j}, \dots, t_{N_j,j}$ son los puntos donde $r_{j-1} < Z_t \leq r_j$, N_j es el número de observaciones en régimen j y $\eta_j = (1 + k.p_j + v.q_j + d_j)$. La distribución a priori de los grados de libertad ν también es no informativa de la forma:

$$P(\nu) = K * I_{(2,+\infty)}(\nu), \quad K \in R^+, \quad (2-38)$$

y su distribución condicional completa es:

$$\begin{aligned}
 p(\nu|\Sigma, \theta, u_{1:T}, y_{1:T}, \theta_{y_{ms}}) &= p(y_{1:T}|u_{1:T}, \theta, \Sigma, \nu)p(\nu) \\
 &= \left[\frac{\Gamma((\nu+k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)} \right]^{T-p} \prod_{t=p+1}^T |\Sigma_{(j)}|^{-1/2} \prod_{t=p+1}^T \left[1 + \frac{1}{\nu} \varepsilon_t' \varepsilon_t \right]^{-(\nu+k)/2} \\
 &\propto \left[\frac{\Gamma((\nu+k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)} \right]^{T-p} \prod_{t=p+1}^T \left[1 + \frac{1}{\nu} \varepsilon_t' \varepsilon_t \right]^{-(\nu+k)/2}.
 \end{aligned} \tag{2-39}$$

Romero y Calderón (2021) descomponen la matriz de covarianzas de la siguiente forma:

$$\Sigma_{(j)} = \text{diag}(\sigma_{(j)}) \mathcal{P}_{(j)} \text{diag}(\sigma_{(j)}), \tag{2-40}$$

donde $\sigma_{(j)}$ es un vector de tamaño $k \times 1$ de las desviaciones estándar escaladas del régimen j y $\text{diag}(\sigma_{(j)})$ es una matriz diagonal cuyos elementos son $\sigma_{(j)} = (\sigma_{(1j)}, \sigma_{(2j)}, \dots, \sigma_{(kj)})$. Mientras $\mathcal{P}_{(j)}$ es una matriz de correlaciones simétrica de tamaño $k \times k$ para el régimen j cuyos elementos son los coeficientes de correlación escalados siguiendo la descomposición propuesta por Barnard, McCulloch y Meng (2000). Para el vector de desviaciones estándar escaladas, Romero y Calderón (2021) proponen la siguiente distribución a priori no informativa:

$$P(\sigma_{(j)}) = K \times I_{(0,+\infty)}(\sigma_{(1j)}) \times I_{(0,+\infty)}(\sigma_{(2j)}) \times \dots \times I_{(0,+\infty)}(\sigma_{(kj)}), \quad K \in R^+, \tag{2-41}$$

y su distribución condicional completa es:

$$\begin{aligned}
 p(\sigma_{(j)}|\sigma_{(i)}, i \neq j, \mathcal{P}, \theta, u_{1:T}, y_{1:T}, \theta_{y_{ms}}) &= p(y_{1:T}|u_{1:T}, \theta, \Sigma, \nu)p(\sigma_{(j)}) \\
 &= \left[\frac{\Gamma((\nu+k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)} \right]^{T-p} \prod_{t=p+1}^T |\Sigma_{(j)}|^{-1/2} \prod_{t=p+1}^T \left[1 + \frac{1}{\nu} \varepsilon_t' \varepsilon_t \right]^{-(\nu+k)/2} \\
 &\propto |\Sigma_{(j)}|^{-\frac{N_j}{2}} \prod_{t=p+1}^T \left[1 + \frac{1}{\nu} (\Sigma_{(j)}^{-1/2} (Y_t - (W'_{t,j} \otimes I_k) \theta_j))' (\Sigma_{(j)}^{-1/2} (Y_t - (W'_{t,j} \otimes I_k) \theta_j)) \right]^{-(\nu+k)/2}.
 \end{aligned} \tag{2-42}$$

La distribución a priori para los coeficientes de correlación escalados también es una distribución a priori no informativa de la forma:

$$P(\mathcal{P}_{(j)}) = I_{(-1,1)}(\rho_1) \times I_{(-1,1)}(\rho_2) \times \dots \times I_{(-1,1)}(\rho_{\frac{k^2-k}{2}}), \tag{2-43}$$

y su distribución condicional completa es:

$$\begin{aligned}
 p(\mathcal{P}_{(j)}|\mathcal{P}_{(i)}, i \neq j, \sigma, \theta, u_{1:T}, y_{1:T}, \theta_{y_{ms}}) &= p(y_{1:T}|u_{1:T}, \theta, \Sigma, \nu)p(\mathcal{P}_{(j)}) \\
 &= \left[\frac{\Gamma((\nu+k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)} \right]^{T-p} \prod_{t=p+1}^T |\text{diag}(\sigma_{(j)}) \mathcal{P}_{(j)} \text{diag}(\sigma_{(j)})|^{-1/2} \prod_{t=p+1}^T \left[1 + \frac{1}{\nu} \varepsilon_t' \varepsilon_t \right]^{-(\nu+k)/2} \\
 &\propto |\mathcal{P}_{(j)}|^{-\frac{N_j}{2}} \prod_{t=p+1}^T \left[1 + \frac{1}{\nu} (\Sigma_{(j)}^{-1/2} (Y_t - (W'_{t,j} \otimes I_k) \theta_j))' (\Sigma_{(j)}^{-1/2} (Y_t - (W'_{t,j} \otimes I_k) \theta_j)) \right]^{-(\nu+k)/2}.
 \end{aligned} \tag{2-44}$$

Para la estimación de los parámetros no estructurales, se implementa el algoritmo Metropolis-Hastings con caminata aleatoria para obtener muestras de las cuatro distribuciones condicionales completas anteriores, ya que no son distribuciones estándar. Se implementa el algoritmo Metropolis-Hastings con caminata aleatoria para las matrices autorregresivas, el vector de desviaciones estándar escaladas y de coeficientes de correlación escalados con una distribución de ruido normal multivariada de media cero y matriz de covarianza $k\sigma$ calculada por el método de Laplace, donde k corresponde a una constante elegida para lograr tasas de aceptación adecuadas. Posteriormente se aplica el muestreador de Gibbs para estimar de manera conjunta todos los parámetros. Para extraer muestras de las distribuciones de interés y así obtener la estimación de los parámetros no estructurales del modelo MTAR se implementa el siguiente procedimiento repitiendo n veces del paso 2 al 5 para obtener cadenas de longitud n :

1. En $\kappa = 0$, se establecen valores iniciales $\theta_j^{(0)}, \sigma_j^{(0)}, \mathcal{P}_j^{(0)}$ y $\nu^{(0)}$ para $j = 1, \dots, l$.
2. Para $\kappa + 1$, a través del algoritmo Metropolis-Hastings se simula un valor para el vector θ_j de la distribución condicional completa $\theta_j^{\kappa+1} \sim p(\theta_{(j)} | \theta_{(i)}^\kappa, i \neq j, \Sigma^\kappa, \nu^\kappa, u_{1:T}, y_{1:T}, \theta_{y_{ns}})$ dada en la ecuación 2-37.
3. Para $\kappa + 1$, a través del algoritmo Metropolis-Hastings se simula un valor para σ_j de la distribución condicional completa $\sigma_j^{\kappa+1} \sim p(\sigma_{(j)} | \sigma_{(i)}^\kappa, i \neq j, \mathcal{P}^\kappa, \theta^{\kappa+1}, \nu^\kappa, u_{1:T}, y_{1:T}, \theta_{y_{ns}})$ dada en la ecuación 2-42.
4. Para $\kappa + 1$, a través del algoritmo Metropolis-Hastings se simula un valor para \mathcal{P} de la distribución condicional completa $\mathcal{P}^{\kappa+1} \sim p(\mathcal{P}_{(j)} | \mathcal{P}_{(i)}^\kappa, i \neq j, \sigma^{\kappa+1}, \theta^{\kappa+1}, \nu^\kappa, u_{1:T}, y_{1:T}, \theta_{y_{ns}})$ dada en la ecuación 2-44.
5. Para $\kappa + 1$, a través del algoritmo Metropolis-Hastings se simula un valor para ν de la distribución condicional completa $\nu^{\kappa+1} \sim p(\nu | \Sigma^{\kappa+1}, \theta^{\kappa+1}, u_{1:T}, y_{1:T}, \theta_{y_{ns}})$ dada en la ecuación 2-39.

2.4.5. Estimación e identificación de parámetros estructurales

Calderón y Nieto (2017) proponen que el vector de umbrales sea independiente de $\Sigma_{(j)}$ y tenga una distribución a priori uniforme sobre el sólido:

$$S = \{(r_1, \dots, r_{l-1}) \in \mathbb{R}^{l-1} : z_{(a)} < r_i < z_{(b)}, i = 1, \dots, l-1; r_1 < r_2 < \dots < r_{l-1}\}$$

$$\text{con volumen } V(S) = \int_{z_{(a)}}^{z_{(b)}} \int_{r_1}^{z_{(b)}} \dots \int_{r_{l-1}}^{z_{(b)}} dr_{l-1} \dots dr_1 dr_2 = \frac{(z_{(b)} - z_{(a)})^{l-1}}{(l-1)!}, \quad (2-45)$$

donde $z_{(a)}$ y $z_{(b)}$ son los percentiles a-ésimo y b-ésimo de la muestra z_1, \dots, z_T . Con lo que la distribución a priori es:

$$p(r) = \frac{1}{V(S)} 1_s(r), r \in \mathbb{R}^{l-1}, \quad (2-46)$$

donde $1_s(r)$ es la función indicadora de S sobre \mathbb{R}^{l-1} . Si el modelo tiene solo dos regímenes, entonces $r \sim U(z_{(a)}, z_{(b)})$. Ibáñez y Calderón (2020) demuestran que la distribución condicional para el vector de umbrales es:

$$p(r|\theta_{y-r}, y, u) \propto \frac{1}{V(S)} 1_S(r) \left\{ \prod_{j=1}^l |\Sigma_{(i)}|^{N_j/2} \right\} \left[\prod_{t=t_1'}^{t_{N_j}^{(j)}} 1 + \frac{1}{\nu} \varepsilon_t' \varepsilon_t \right]^{-(\nu+k)/2}, \quad (2-47)$$

donde N_j para $j = 1, \dots, l$ depende de r . Posteriormente, se implementa el algoritmo Metropolis-Hastings con caminata aleatoria para obtener muestras de la distribución condicional completa del umbral. La identificación del número de regímenes y del orden autorregresivo en cada régimen se realiza a través del criterio de información de Akaike normalizado (NAIC):

$$NAIC = \left\{ \sum_{j=1}^l AIC_j(p_j, q_j, d_j, l) \right\} / \left\{ \sum_{j=1}^l N_j \right\}, \quad (2-48)$$

con $AIC_j(p_j, q_j, d_j, l) = N_j \ln(|S_j|/N_j) + 2kn_j$

y $S_j = \sum_{\{t: j_t=t\}} \left(Y_t - \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} Y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} X_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} Z_{t-i} \right),$

donde N_j corresponde a número de observaciones en el régimen j . De esta forma, se calcula el NAIC para cada modelo con el valor de los umbrales según el número de regímenes de la combinación elegida.

2.4.6. Validación del modelo

Para evaluar la bondad de ajuste del modelo, se definen los residuos del modelo MTAR correspondiente como:

$$\varepsilon_t = \Sigma_{(j)}^{-1/2} (Y_t - (W_{t,j}' \otimes I_k) \theta_j), \text{ si } r_{j-1} < z_T \leq r_j. \quad (2-49)$$

Los residuos anteriores se analizan a través de las gráficas CUSUM con el fin de validar la especificación del modelo y CUSUMSQ para chequear la heterocedasticidad marginal en el proceso del error (Espinosa y Nieto, 2016). Con el gráfico CUSUM se puede identificar la inestabilidad de la media condicional del modelo, mientras que el gráfico CUSUMSQ permite comprobar desviaciones no aleatorias desde su línea de valor medio. Además, con el objetivo de descartar correlación serial en el proceso de ruido, se analizan las funciones de autocorrelación.

3. Pronósticos

El objetivo de los pronósticos en la estadística Bayesiana es encontrar la distribución predictiva a posteriori, es decir, la distribución de los puntos de datos futuros de y condicional a sus datos observados. La distribución predictiva captura toda la información relevante sobre los eventos futuros desconocidos, con lo que se puede elegir una característica relevante de dicha distribución, por ejemplo, la media, la moda o la mediana de la distribución predictiva junto con un intervalo de probabilidad que indique el rango de resultados probables (Karlsson, 2013). Además, las distribuciones predictivas no dependen de ninguna cantidad desconocida sino de los datos observados, por lo tanto, y da información sobre θ , que da información sobre \tilde{y} (Vaca y Nieto, 2018).

Para el caso del modelo MTAR con distribución de error t-Student multivariada presentado en la ecuación (2-22), el objetivo es encontrar $E[Y_{T+h}|y_{1:T}, x_{1:T}, z_{1:T}, l]$ con $h \geq 1$, que es la mejor predicción basada en la función pérdida cuadrática, donde T es el tamaño de la muestra y h el horizonte de pronóstico. En este caso y de manera similar a Calderón y Nieto (2017, 2021), se requiere encontrar la distribución predictiva conjunta, la cual se puede obtener de la siguiente forma:

$$p(y_{T+1:T+h}|y_{1:T}, x_{1:T}, z_{1:T}, l) = \int_{\Theta} p(y_{T+1:T+h}|y_{1:T}, x_{1:T}, z_{1:T}, \theta_{y,l}, l) p(\theta_{y,l}|y_{1:T}, x_{1:T}, z_{1:T}, l) d\theta_{y,l}, \tag{3-1}$$

donde $p(\theta_{y,l}|y_{1:T}, x_{1:T}, z_{1:T}, l)$ es la distribución a posteriori de los parámetros de un modelo MTAR con l regímenes y $p(y_{T+1:T+h}|y_{1:T}, x_{1:T}, z_{1:T}, \theta_{y,l}, l)$ es una distribución que debe ser especificada a partir de los supuestos de las distribuciones empleadas para estimar el modelo MTAR. La ecuación (3-1) se basa en la distribución predictiva Bayesiana dada en la ecuación (2-5) del Capítulo 2.

El procedimiento presentado para obtener los pronósticos del modelo MTAR asume que los valores en $t+h$ de las variables Z_t y X_t se obtienen con métodos alternos o están disponibles en el conjunto de datos. Para ello se asume que $Y_{1:t}$ no causa en el sentido de Granger a Z_t y X_t . Teniendo en cuenta lo anterior, el elemento $p(y_{T+1:T+h}|y_{1:T}, x_{1:T}, z_{1:T}, \theta_{y,l}, l)$ de la

ecuación (3-1) es igual a:

$$p(y_{T+1}, \dots, y_{T+h} | y_{1:T}, x_{1:T}, z_{1:T}, \theta_{y,l}, l) = \prod_{h=1}^H p(y_{T+h} | x_{T+h}, z_{T+h}, y_{1:T+h-1}, x_{1:T+h-1}, z_{1:T+h-1}, \theta_{y,l}, l). \quad (3-2)$$

La distribución $p(y_{T+h} | x_{T+h}, z_{T+h}, y_{1:T+h-1}, x_{1:T+h-1}, z_{1:T+h-1}, \theta_{y,l}, l)$ debe ser especificada a partir de los supuestos del modelo MTAR presentado en la ecuación (2-22), este se puede escribir de forma condensada de la siguiente manera para $h = 1, \dots, H$:

$$Y_{t+h,j} = (W'_{T+h,j} \otimes I_k) \theta_j + \Sigma_{(j)}^{1/2} \varepsilon_t \text{ si } r_{j-1} < z_T \leq r_j, \quad (3-3)$$

donde $w_{t+h,j} = (1, y'_{t+h-1}, \dots, y'_{t+h-p_j}, x_{t+h-1}, \dots, x_{t+h-q'_j}, z_{t+h-1}, \dots, z_{t+h-d_j})'_{\eta_j \times 1}$ es un vector columna para $j = 1, \dots, l$, $\eta_j = (1 + k.p_j + v.q_j + d_j)$, $W_{T+h,j} = (w_{t_{1,j}}, \dots, w_{t_{N_j,j}})_{\eta_j \times N_j}$, $t_{1,j}, \dots, t_{N_j,j}$ son los puntos donde $r_{j-1} < Z_t \leq r_j$ y N_j es el número de observaciones en el régimen j . Además, el vector columna $\theta = \text{vec}(A_j)$, donde A_j es:

$$A_j = (\phi_0^{(j)}, \phi_1^{(j)}, \dots, \phi_{p_j}^{(j)}, \beta_1^{(j)}, \dots, \beta_{q_j}^{(j)}, \delta_1^{(j)}, \dots, \delta_{d_j}^{(j)})_{k \times \eta_j}. \quad (3-4)$$

Con las definiciones anteriores, se puede ver que para $r_{j-1} < Z_t \leq r_j$:

$$\begin{aligned} Y_{t+h,j} &= A_j W_{T+h,j} + \Sigma_{(j)}^{1/2} \varepsilon_t \\ &= \text{vec}(A_j W_{T+h,j}) + \Sigma_{(j)}^{1/2} \varepsilon_t \\ &= \text{vec}(I_k A_j W_{T+h,j}) + \Sigma_{(j)}^{1/2} \varepsilon_t \\ &= (W'_{T+h,j} \otimes I_k) \text{vec}(A_j) + \Sigma_{(j)}^{1/2} \varepsilon_t \\ &= (W'_{T+h,j} \otimes I_k) \theta_j + \Sigma_{(j)}^{1/2} \varepsilon_t. \end{aligned} \quad (3-5)$$

$\{\varepsilon_t\}$ es un proceso independiente e idénticamente distribuido, y por cada t , ε_t sigue una distribución t-Student multivariada con vector de medias 0, matriz de covarianza I_k de orden k y v grados de libertad. A partir de la ecuación (3-5) y la [Propiedad 1](#) de Kibria y Joarder (2006), es posible identificar la siguiente transformación lineal:

$$Y = AX + b,$$

donde:

$$A = \Sigma_{(j)}^{1/2}, \quad (3-6)$$

$$b = (W'_{t+h,j} \otimes I_k) \theta_j$$

$$\text{y } X = \varepsilon_t \sim t_\nu(0, I).$$

En primer lugar, Tsay (1998) define a $\Sigma_{(j)}^{1/2}$ como una matriz simétrica definida positiva, mientras que Anthony y Harvey (2012) demuestran que si L es una matriz simétrica de

tamaño $n \times n$ y es definida positiva, entonces todos sus valores propios son positivos, por lo que 0 no es un valor propio de L y L es no singular. En segundo lugar, b es un vector cuyos valores son fijos, ya que contiene observaciones de $\{Y_t\}$, $\{X_t\}$ y $\{Z_t\}$ y los valores de θ_j para $j = 1, \dots, l$. Con lo anterior, se cumplen las dos condiciones de la propiedad de Kibria y Joarder (2006) y Y_{t+i} se distribuye t-Student k-variada con vector de medias igual a:

$$\begin{aligned}\mu^* &= A\mu + b \\ &= \Sigma_{(j)}^{1/2}(0) + (W'_{t+h,j} \otimes I_k)\theta_j \\ &= (W'_{t+h,j} \otimes I_k)\theta_j,\end{aligned}\tag{3-7}$$

y matriz de covarianzas definida como:

$$\begin{aligned}\Sigma^* &= A\Sigma A' \\ &= \Sigma_{(j)}^{1/2} I_k \Sigma_{(j)}^{1/2} \\ &= \Sigma_{(j)}.\end{aligned}\tag{3-8}$$

De esta forma, la distribución $p(y_{T+h}|x_{T+h}, z_{T+h}, y_{1:T+h-1}, x_{1:T+h-1}, z_{1:T+h-1}, \theta_{y,l}, l)$ de la ecuación (3-2) es igual a:

$$\frac{\Gamma((\nu + k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)|\Sigma_{(j)}|^{1/2}} \left[1 + \frac{1}{\nu}(y_{t+h} - (W'_{t+h,j} \otimes I_k)\theta_j)' \Sigma_{(j)}^{-1}(y_{t+h} - (W'_{t+h,j} \otimes I_k)\theta_j) \right]^{-\nu+k/2},\tag{3-9}$$

con media μ^* presentada en la ecuación (3-7), matriz de covarianza Σ^* presentada en la ecuación (3-8) y ν grados de libertad. Con la demostración anterior, se conoce la forma que toma la distribución de la ecuación (3-2). Con el supuesto de que los pronósticos de la variable de umbral y las covariables han sido obtenidos previamente o están disponibles, para extraer muestras de la distribución predictiva conjunta se implementa el siguiente procedimiento hasta llegar a la iteración κ -ésima:

1. Extraer una muestra aleatoria para el vector $\theta_{y,l}^{(\kappa)}$ de la distribución a posteriori $p(\theta_{y,l}|y_{1:T}, x_{1:T}, z_{1:T}, l)$ dada en la ecuación (2-37).
2. Extraer una muestra aleatoria para $y_{T+1}^{(\kappa)}$ de la densidad $p(y_{T+1}|x_{T+1}, z_{T+1}, y_{1:T}, x_{1:T}, z_{1:T}, l, \theta_{y,l}^{(\kappa)}, l)$ dada en la ecuación (3-9).
3. Extraer una muestra aleatoria para $y_{T+2}^{(\kappa)}$ de la densidad $p(y_{T+2}|x_{T+2}, z_{T+2}, x_{T+1}, z_{T+1}, y_{T+1}^{(\kappa)}, y_{1:T}, x_{1:T}, z_{1:T}, \theta_{y,l}^{(\kappa)}, l)$ dada en la ecuación (3-9).
4. Continuar extrayendo muestras aleatorias hasta obtener $y_{T+h}^{(\kappa)}$ de $p(y_{T+h}|x_{T+h}, z_{T+h}, y_{T+1}^{(\kappa)}, \dots, y_{T+h}^{(\kappa)}, y_{1:T}, x_{1:T}, z_{1:T}, \theta_{y,l}^{(\kappa)}, l)$ dada en la ecuación (3-9).

Con este procedimiento se obtienen muestras $\{y_{T+1}^{(\kappa)}, y_{T+2}^{(\kappa)}, \dots, y_{T+h}^{(\kappa)}\}$ de la distribución predictiva definida en la ecuación (3-1) que se pueden usar para estimar $E[Y_{T+h}|y_{1:T}, x_{1:T}, z_{1:T}, l]$ con $h \geq 1$. La incertidumbre sobre los verdaderos valores de los parámetros es introducida a través de $p(\theta_{y,l}|y_{1:T}, x_{1:T}, z_{1:T}, l)$, además se tiene en cuenta la incertidumbre inherente de los pronósticos a través de $p(y_{T+1:T+h}|y_{1:T}, x_{1:T}, z_{1:T}, \theta_{y,l}, l)$. Con el algoritmo planteado se generan muestras de la distribución conjunta $p(y_{T+1:T+h}, \theta_{y,l}|y_{1:T}, x_{1:T}, z_{1:T}, l)$, ya que en cada iteración se obtienen observaciones de $\theta_{y,l}$ y y_{T+h} . Posteriormente se obtienen cuantiles de los valores muestreados de y_{T+h} , este es el paso de marginalización basado en la ecuación (3-1) donde se integra la distribución predictiva conjunta sobre los parámetros que no son de interés. Los pronósticos de y_{T+h} no dependen de un valor específico de $\theta_{y,l}$, sino que dependen colectivamente de todos los valores muestreados de $\theta_{y,l}$. De este conjunto se puede calcular la media de y_{T+h} que es una aproximación numérica al pronóstico puntual (Vaca y Nieto, 2018) y los intervalos de predicción para los valores del vector $(Y_{T+1:T+h})$.

4. Ejercicio de simulación

El objetivo de este capítulo es llevar a cabo un estudio de simulación para comprobar el comportamiento de la metodología presentada en el Capítulo 3. Para lograr este objetivo, se realizan simulaciones de series de tamaño $1000 + h$ de tres modelos MTAR con $h = 10$. Para generar muestras de la distribución predictiva y obtener los pronósticos puntuales e intervalos de predicción, se implementa el algoritmo expuesto con 5000 iteraciones y con un periodo de calentamiento de 2000 (40%) iteraciones. Ya que las series simuladas incluyen todos los valores de $\{X_t\}$ y $\{Z_t\}$, el algoritmo se implementa con los verdaderos valores de estas variables. Posteriormente, se simulan 100 realizaciones de los tres modelos MTAR considerados, cada uno de tamaño 500 y se calcula el porcentaje de veces en que los últimos 10 valores verdaderos del proceso de salida se encuentran dentro del intervalo de predicción del 95% de la distribución predictiva. En este proceso la longitud de las cadenas también es igual a 5000 con un periodo de calentamiento de 2000 (40%).

4.1. Modelo 1

El primer modelo a considerar es un MTAR bivariado ($k = 2$) con dos regímenes ($l = 2$) donde $\varepsilon_t \sim t_7(0, I)$ y el valor del umbral $\hat{Z}_{0.5}$ se refiere al percentil muestral 50 de Z :

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)}Y_{t-1} + \phi_2^{(1)}Y_{t-2} + \beta_1^{(1)}X_{t-1} + \Sigma_{(1)}^{1/2}\varepsilon_t & \text{si } Z_t \leq \hat{Z}_{0.5} \\ \phi_0^{(2)} + \phi_1^{(2)}Y_{t-1} + \delta_1^{(2)}Z_{t-1} + \Sigma_{(2)}^{1/2}\varepsilon_t & \text{si } Z_t > \hat{Z}_{0.5}, \end{cases} \quad (4-1)$$

y:

$$\begin{aligned} \phi_0^{(1)} &= \begin{bmatrix} 2.2 \\ 1.5 \end{bmatrix}, & \phi_1^{(1)} &= \begin{bmatrix} 0.3 & -0.1 \\ -0.2 & 0.4 \end{bmatrix}, & \phi_2^{(1)} &= \begin{bmatrix} -0.1 & 0.5 \\ 0.2 & 0.4 \end{bmatrix}, \\ \beta_1^{(1)} &= \begin{bmatrix} -0.5 \\ 0.2 \end{bmatrix}, & \Sigma_{(1)}^{1/2} &= \begin{bmatrix} 0.9 & 0.5 \\ 0.5 & 1.1 \end{bmatrix}, \\ \phi_0^{(2)} &= \begin{bmatrix} -1.3 \\ 1.4 \end{bmatrix}, & \phi_1^{(2)} &= \begin{bmatrix} -0.3 & 0.6 \\ 0.4 & 0.3 \end{bmatrix}, & \delta_1^{(2)} &= \begin{bmatrix} -1.0 \\ 1.0 \end{bmatrix}, \\ \Sigma_{(2)}^{1/2} &= \begin{bmatrix} 1.8 & 0.6 \\ 0.6 & 2.0 \end{bmatrix}. \end{aligned} \quad (4-2)$$

Las covariables X_t y Z_t se generaron a partir del siguiente modelo VAR:

$$\begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} 0.2 & 0.6 \\ 0.4 & 0.3 \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \quad (4-3)$$

donde:

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.7 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right). \quad (4-4)$$

A partir del modelo descrito en la ecuación (4-1) se simula Y_t con $r = \hat{Z}_{0.5}$. En la figura 4-1 se presenta el gráfico de su simulación.

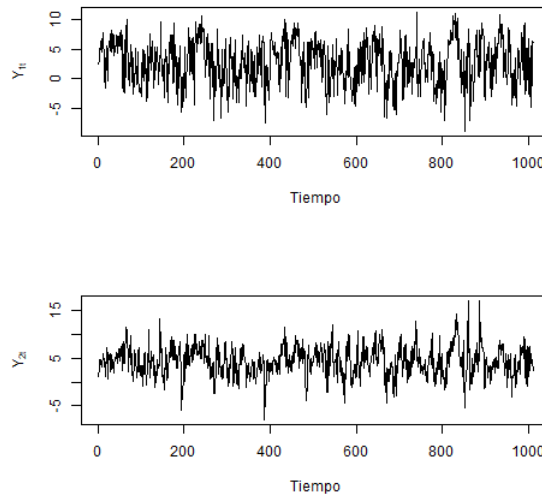


Figura 4-1.: Serie simulada del modelo MTAR con $k = 2$ y $l = 2$ - Modelo 1.

En la figura 4-2 se presentan las cadenas y densidades resultantes de una réplica con las que se obtienen los pronósticos del modelo 1. Se observa que las densidades tienen forma de campana y son simétricas y las cadenas presentan un comportamiento estacionario.

En la tabla 4-1 se presentan los verdaderos valores del proceso del salida y los pronósticos obtenidos junto con los intervalos de predicción del 95 % para una realización. Se puede observar que todos los verdaderos valores del proceso de salida se encuentran dentro de los intervalos de predicción individuales del 95 %.

Para verificar la convergencia de las cadenas, en la tabla 4-2 y en la figura 4-3 se muestra el criterio de convergencia de Gelman y Rubin basado en 25 réplicas, donde los factores de reducción y los límites superiores son cercanos a uno, por consiguiente hay convergencia de las cadenas. En cada réplica la longitud de las cadenas es igual a 5000.

En la figura 4-4 se presenta el criterio de convergencia de Geweke para las cadenas obtenidas y se observa que los puntajes Z están dentro de las bandas de confianza, lo que indica convergencia de las cadenas.

Horizonte de pronóstico	y_{1000+h}	\hat{y}_{1000+h}	I.C. 95 %
1	1.6677	-0.4762	(-4.7264;4.047)
	2.3644	4.643	(-0.2282;9.6341)
2	-2.9452	0.3676	(-4.7298;5.3094)
	7.4086	3.9103	(-1.9231;9.4559)
3	3.5055	-0.5734	(-5.6025;4.7294)
	3.0413	4.2106	(-1.8171;10.0317)
4	-4.71	1.1642	(-4.0158;6.3866)
	4.536	2.648	(-3.5063;8.5885)
5	1.0655	5.3822	(2.5741;8.169)
	5.5306	7.2726	(3.5169;10.9417)
6	5.5623	5.2904	(2.4026;8.2237)
	5.925	7.5247	(3.6852;11.1768)
7	-1.0503	-1.6719	(-6.7061;3.8665)
	3.7572	4.605	(-1.4741;10.6709)
8	0.9926	6.4835	(3.7192;9.5478)
	4.532	7.1172	(3.2765;10.9224)
9	6.3957	6.9952	(4.0046;9.9679)
	4.5536	6.9368	(3.2083;10.7391)
10	6.201	7.0016	(4.0844;10.0385)
	2.4566	6.7652	(2.8647;10.808)

Tabla 4-1.: Pronósticos de 1 a 10 pasos adelante - Modelo 1.

Horizonte de pronóstico	Y_1		Y_2	
	Factor de reducción	Límite superior I.C.	Factor de reducción	Límite superior I.C.
1	1.0004	1.0008	1.0002	1.0005
2	1.0001	1.0004	1.0006	1.0012
3	1.0006	1.0011	1.0004	1.0009
4	1.0004	1.0008	1.0008	1.0014
5	1.0029	1.0050	1.0028	1.0048
6	1.0033	1.0055	1.0033	1.0057
7	1.0008	1.0015	1.0009	1.0016
8	1.0037	1.0062	1.0048	1.0080
9	1.0027	1.0045	1.0044	1.0074
10	1.0033	1.0056	1.0043	1.0072

Tabla 4-2.: Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 1.

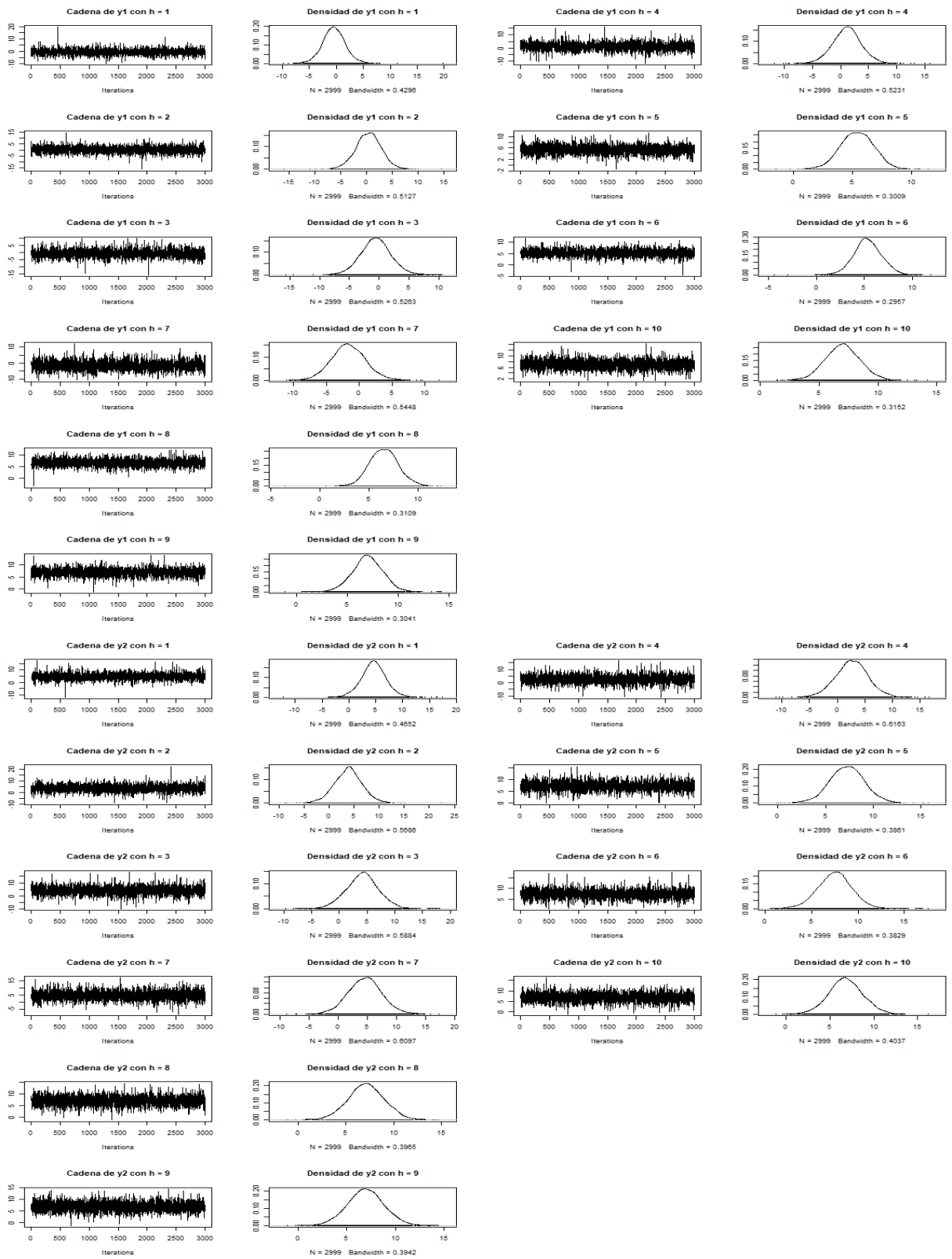


Figura 4-2.: Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Modelo 1.

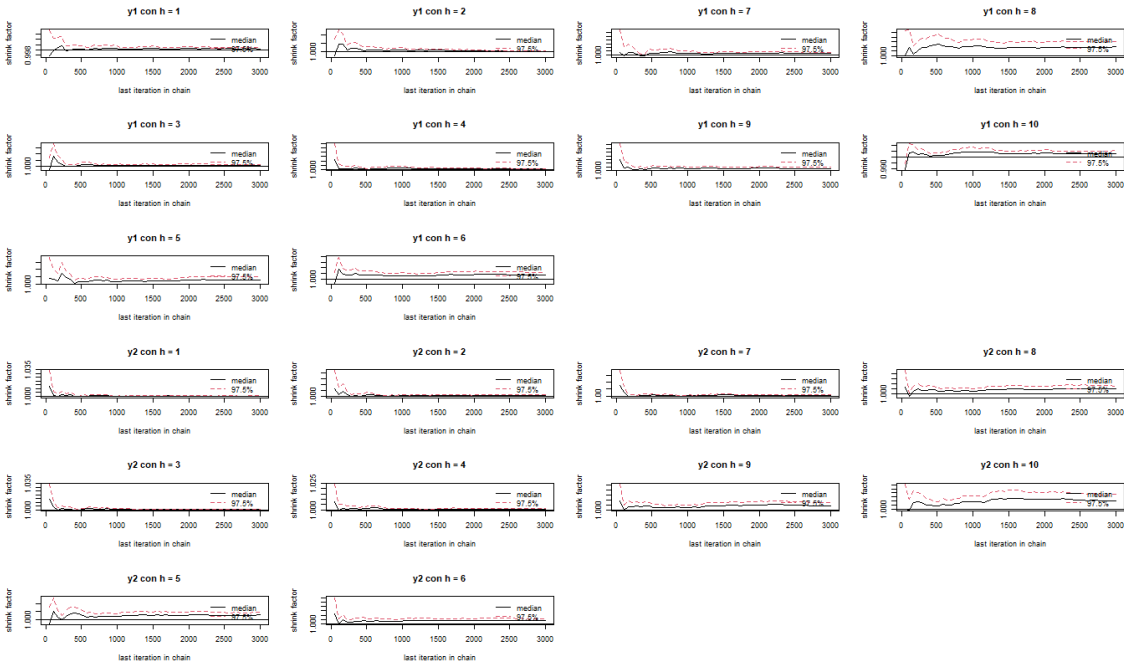


Figura 4-3.: Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 1.

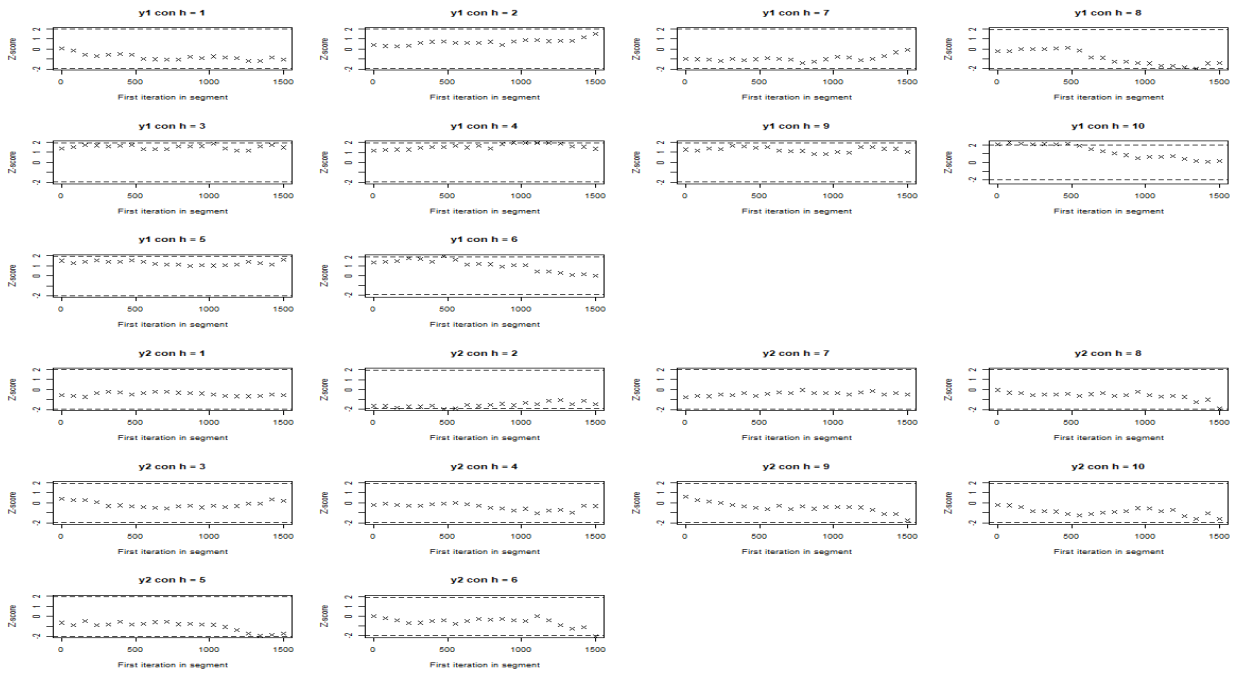


Figura 4-4.: Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Modelo 1.

4.2. Modelo 2

El segundo modelo a considerar es un MTAR bivariado ($k = 2$) con dos regímenes ($l = 2$) donde $\varepsilon_t \sim t_8(0, I)$ y el valor del umbral $\hat{Z}_{0.55}$ se refiere al percentil muestral 55 de Z :

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)}Y_{t-1} + \phi_2^{(1)}Y_{t-2} + \beta_1^{(1)}X_{t-1} + \Sigma_{(1)}^{1/2}\varepsilon_t & \text{si } Z_t \leq \hat{Z}_{0.55} \\ \phi_0^{(2)} + \phi_1^{(2)}Y_{t-1} + \Sigma_{(2)}^{1/2}\varepsilon_t & \text{si } Z_t > \hat{Z}_{0.55}, \end{cases} \quad (4-5)$$

y:

$$\begin{aligned} \phi_0^{(1)} &= \begin{bmatrix} 1.0 \\ -1.0 \end{bmatrix}, & \phi_1^{(1)} &= \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix}, & \Sigma_{(1)}^{1/2} &= \begin{bmatrix} 1.0 & 0.6 \\ 0.6 & 1.5 \end{bmatrix}, \\ \phi_2^{(1)} &= \begin{bmatrix} 0.0 & 0.0 \\ 0.25 & 0.0 \end{bmatrix}, & \beta_1^{(1)} &= \begin{bmatrix} 0.3 \\ -0.4 \end{bmatrix}, & \Sigma_{(2)}^{1/2} &= \begin{bmatrix} 2.5 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}, \\ \phi_1^{(2)} &= \begin{bmatrix} 0.3 & 0.5 \\ 0.2 & 0.7 \end{bmatrix}, \end{aligned} \quad (4-6)$$

Las covariables X_t y Z_t se generaron a partir del siguiente modelo VAR:

$$\begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \quad (4-7)$$

donde:

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 2.0 \end{pmatrix} \right) \quad (4-8)$$

A partir del modelo descrito en la ecuación (4-5) se simula Y_t con $r = \hat{Z}_{0.55}$. En la figura 4-5 se presenta el gráfico de su simulación.

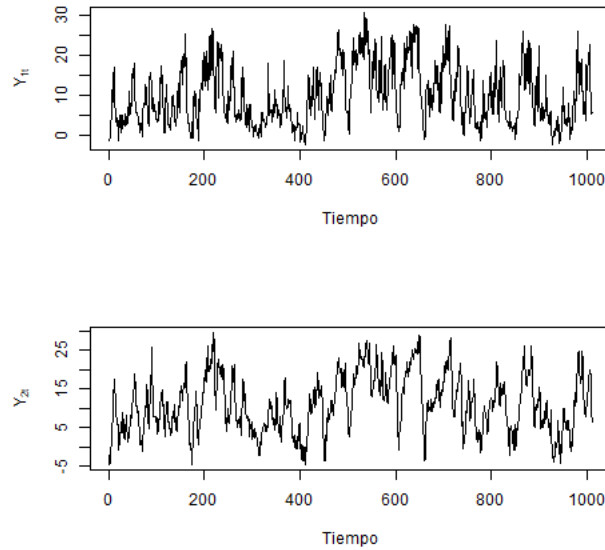


Figura 4-5.: Serie simulada del modelo MTAR con $k = 2$ y $l = 3$ - Modelo 2.

En la figura 4-6 se presentan las cadenas y densidades resultantes de una réplica con las que se obtienen los pronósticos del modelo 2. Las cadenas obtenidas se distribuyen alrededor del valor verdadero del proceso de salida y son simétricas.

En la tabla 4-3 se observan los pronósticos obtenidos. De manera similar al modelo 1, los verdaderos valores del proceso de salida se encuentran al interior de los intervalos resultantes de una realización.

Para verificar la convergencia de las cadenas, en la tabla 4-4 y en la figura 4-7 se muestra el criterio de convergencia de Gelman y Rubin basado en 25 réplicas, donde los factores de reducción y los límites superiores son cercanos a uno, por consiguiente hay convergencia de las cadenas. Ningún factor de reducción es mayor a 1.2, por lo que al seguir la regla propuesta por Karlsson (2013), se acepta que hay convergencia de las cadenas monitoreadas. El criterio de convergencia de Geweke es presentado en la gráfica 4-8, se observa que las cadenas obtenidas convergen a pesar de que algunos valores de Y_1 se salen de las bandas de confianza.

Horizonte de pronóstico	y_{1000+h}	\hat{y}_{1000+h}	I.C. 95 %
1	20.2868	18.3529	(13.6716;23.0090)
	13.6025	16.9376	(14.7089;18.9042)
2	17.6876	18.8695	(13.6805;23.8828)
	16.8939	17.5552	(14.4773;20.5009)
3	18.6480	19.4028	(13.7726;24.9395)
	17.3046	18.0693	(14.6272;21.6311)
4	22.7529	19.7750	(13.93478;25.5797)
	20.0347	18.5726	(14.8180;22.5475)
5	18.4430	20.2000	(14.5543;26.0489)
	19.7043	19.0219	(14.7454;23.3210)
6	5.5736	7.2618	(6.0868;11.7601)
	9.6779	4.7847	(-0.2685;9.8804)
7	9.1609	10.7845	(5.7809;11.56187)
	16.7236	14.1985	(10.7012;17.4536)
8	7.4756	11.0993	(6.1433;12.1075)
	7.8095	3.0791	(-2.3084;8.2468)
9	5.2812	3.1552	(0.1597;6.1764)
	9.3108	2.4879	(-3.01215;7.9375)
10	5.4838	3.3531	(0.4482;6.2938)
	6.5172	1.9908	(-3.2006;7.2958)

Tabla 4-3.: Pronósticos de 1 a 10 pasos adelante - Modelo 2.

Horizonte de pronóstico	Y_1		Y_2	
	Factor de reducción	Límite superior I.C.	Factor de reducción	Límite superior I.C.
1	1.0006	1.0012	1.0008	1.0014
2	1.0005	1.0010	1.0012	1.0022
3	1.0007	1.0013	1.0015	1.0026
4	1.0031	1.0052	1.0016	1.0028
5	1.0032	1.0054	1.0017	1.0030
6	1.0020	1.0035	1.0028	1.0047
7	1.0023	1.0039	1.0033	1.0056
8	1.0028	1.0048	1.0040	1.0067
9	1.0021	1.0036	1.0012	1.0021
10	1.0006	1.0012	1.0018	1.0032

Tabla 4-4.: Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 2.

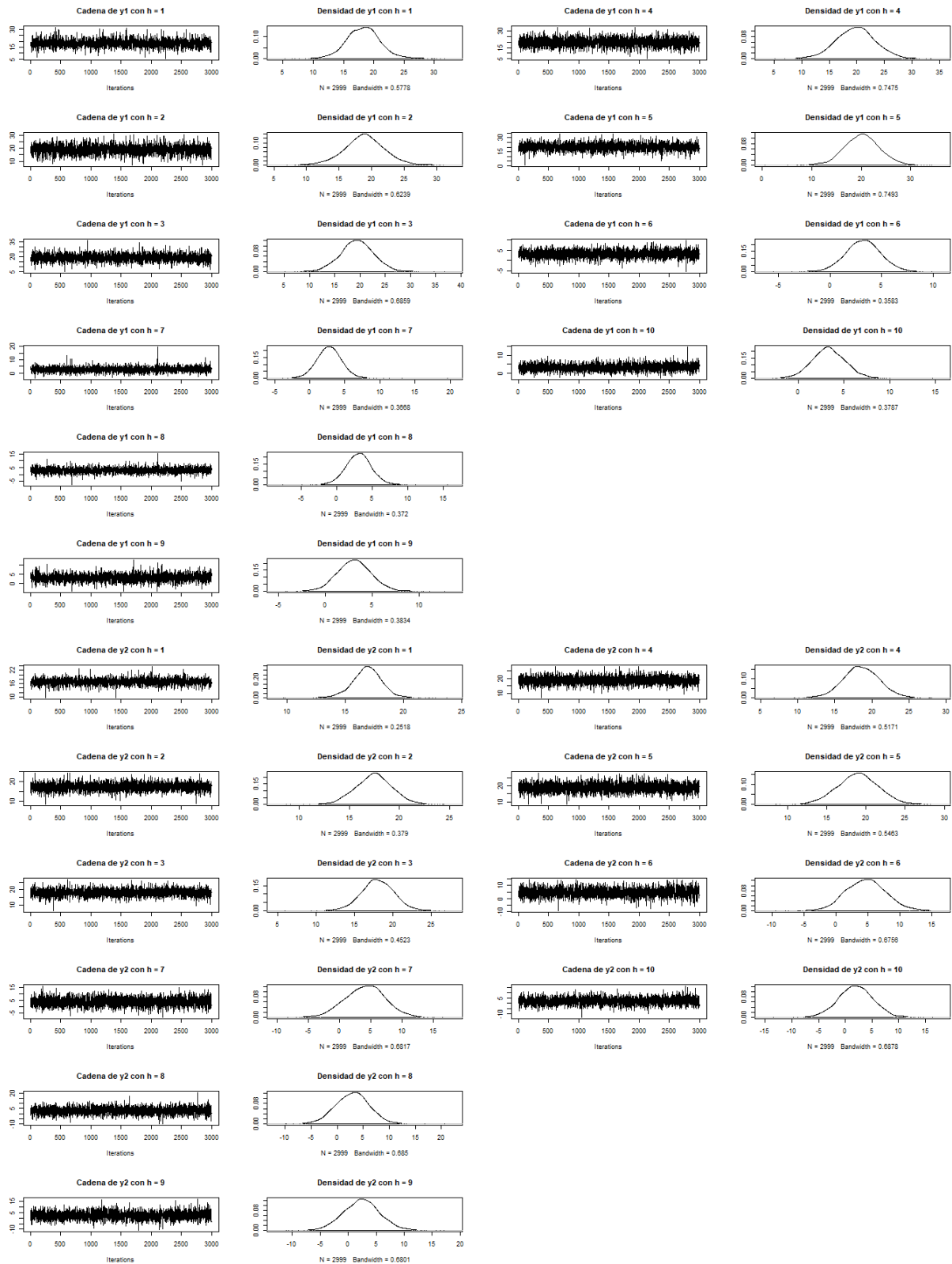


Figura 4-6.: Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Modelo

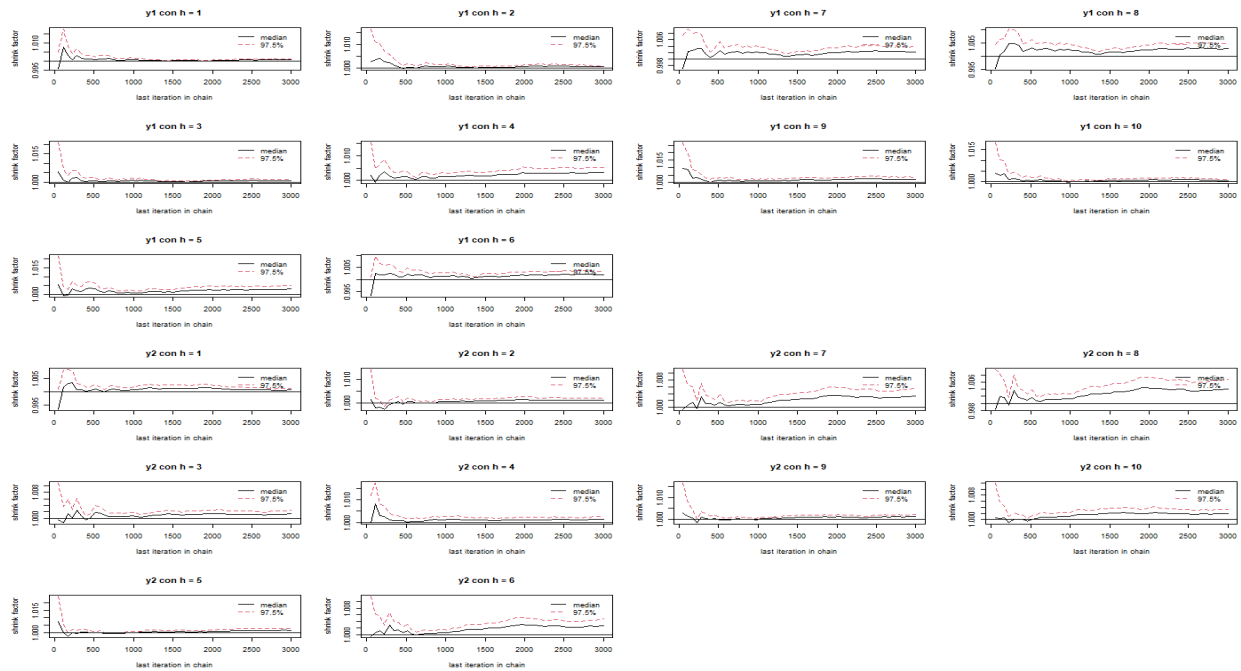


Figura 4-7.: Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 2.

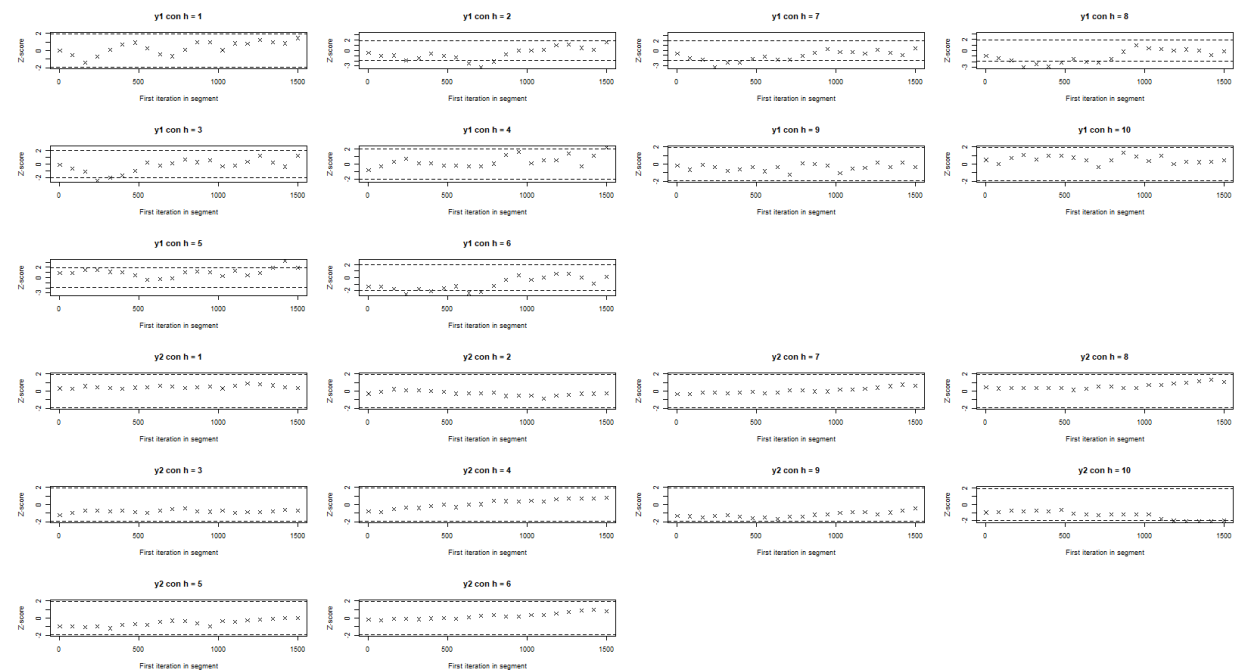


Figura 4-8.: Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Modelo 2.

4.3. Modelo 3

El tercer modelo a considerar es un MTAR trivariado ($k = 3$) con dos regímenes ($l = 2$) donde $\varepsilon_t \sim t_4(0, I)$ y el valor del umbral $\hat{Z}_{0.5}$ se refiere al percentil muestral 50 de Z :

$$Y_t = \begin{cases} \phi_1^{(1)} Y_{t-1} + \beta_1^{(1)} X_{t-1} + \Sigma_{(1)}^{1/2} \varepsilon_t & \text{si } Z_t \leq \hat{Z}_{0.5} \\ \phi_0^{(2)} + \phi_1^{(2)} Y_{t-1} + \Sigma_{(2)}^{1/2} \varepsilon_t & \text{si } Z_t > \hat{Z}_{0.5}, \end{cases} \quad (4-9)$$

y:

$$\begin{aligned} \phi_1^{(1)} &= \begin{bmatrix} 0.5 & 0.2 & -0.4 \\ 0.4 & -0.3 & 0.6 \\ -0.1 & 0.3 & 0.5 \end{bmatrix}, & \beta_1^{(1)} &= \begin{bmatrix} 0.3 \\ 0.5 \\ 0.4 \end{bmatrix}, & \Sigma_{(1)}^{1/2} &= \begin{bmatrix} 1.8 & -0.4 & 0.6 \\ -0.4 & 1.1 & -0.6 \\ 0.6 & -0.6 & 2.2 \end{bmatrix}, \\ \phi_0^{(2)} &= \begin{bmatrix} 1.2 \\ 1.5 \\ 1.8 \end{bmatrix}, & \phi_1^{(2)} &= \begin{bmatrix} 0.3 & 0.2 & -0.2 \\ 0.1 & 0.3 & 0.3 \\ -0.2 & 0.2 & -0.5 \end{bmatrix}, & \Sigma_{(2)}^{1/2} &= \begin{bmatrix} 1.6 & -0.6 & 0.7 \\ -0.6 & 1.4 & 0.4 \\ 0.7 & 0.4 & 2.0 \end{bmatrix}. \end{aligned} \quad (4-10)$$

Las covariables X_t y Z_t se generaron a partir del siguiente modelo VAR:

$$\begin{pmatrix} Z_t \\ X_t \end{pmatrix} = \begin{pmatrix} 0.2 & 0.6 \\ 0.4 & 0.3 \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \quad (4-11)$$

donde:

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.7 & 0.5 \\ 0.5 & 1.5 \end{pmatrix} \right). \quad (4-12)$$

A partir del modelo descrito en la ecuación (4-9) se simula Y_t con $r = \hat{Z}_{0.5}$. En la figura 4-1 se presenta el gráfico de su simulación.

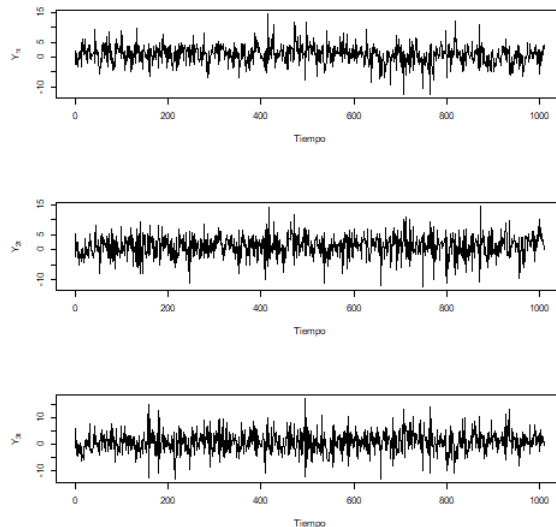


Figura 4-9.: Serie simulada del modelo MTAR con $k = 3$ y $l = 2$ - Modelo 3.

Horizonte de pronóstico	y_{1000+h}	\hat{y}_{1000+h}	I.C. 95 %
1	-0.8067	-0.8977	(-6.1085;4.5146)
	-1.0570	-1.4435	(-5.0711;2.1308)
	-0.0335	0.5388	(-5.8244;6.8493)
2	1.7396	-1.2045	(-6.7869;4.4646)
	-3.1249	-0.1869	(-7.3719;6.8268)
	2.3091	-0.5978	(-7.4067;6.4118)
3	1.5635	-0.7906	(-6.9129;5.3683)
	1.8684	-1.2727	(-8.7587;6.0793)
	-0.3284	-0.6169	(-8.0339;6.9103)
4	4.9848	-1.2306	(-7.7935;5.1829)
	-4.2504	-1.7628	(-9.2889;5.6875)
	4.2992	-1.5999	(-9.1215;5.8535)
5	4.4491	2.2010	(-3.4001;7.739)
	4.8403	2.9890	(-2.1539;7.946)
	11.4601	1.3250	(-5.8859;8.9733)
6	1.2442	2.2395	(-3.0155;7.9216)
	6.7737	2.9942	(-2.0698;7.8628)
	-4.5909	1.1746	(-6.1069;8.8573)
7	2.8574	2.1245	(-3.1393;7.4108)
	4.3841	3.0230	(-2.124;8.2761)
	5.2993	1.3442	(-6.2943;8.8955)
8	1.4663	2.2595	(-3.0912;7.696)
	2.6730	2.9502	(-2.0743;8.3527)
	-3.7898	1.3265	(-6.1055;8.3589)
9	0.5516	2.2202	(-3.2894;7.5441)
	3.7550	2.9941	(-2.2507;7.9759)
	4.8551	1.2471	(-6.2383;9.2445)
10	1.8627	2.1333	(-2.9255;7.4298)
	2.4074	3.0420	(-1.8502;8.285)
	0.5574	1.2947	(-6.8873;9.0888)

Tabla 4-5.: Pronósticos de 1 a 10 pasos adelante - Modelo 3.

Las cadenas y densidades resultantes de una réplica con las que se obtienen los pronósticos del modelo 3 se presentan en la figura 4-10. Se observa que tienen un comportamiento simétrico y son estacionarias.

Los pronósticos e intervalos obtenidos se muestran en la tabla 4-5, los verdaderos valores del proceso de salida se encuentran al interior de los intervalos de predicción individuales del

95 % para una realización.

Los resultados del criterio de convergencia de Gelman y Rubin basado en 25 réplicas se encuentran en la tabla 4-6 y en la figura 4-6. Los resultados de este criterio confirman la convergencia de las cadenas generadas. La figura 4-12 presenta los resultados del criterio de convergencia de Geweke donde también se observa la convergencia de las cadenas obtenidas.

Horizonte de pronóstico	Y_1		Y_2		Y_3	
	Factor de reducción	Límite superior I.C.	Factor de reducción	Límite superior I.C.	Factor de reducción	Límite superior I.C.
1	1.0013	1.0022	1.0006	1.0011	1.0006	1.0011
2	1.0011	1.0019	1.0009	1.0015	1.0001	1.0003
3	1.0010	1.0016	1.0004	1.0008	1.0005	1.0010
4	1.0003	1.0006	1.0007	1.0014	1.0002	1.0005
5	1.0007	1.0011	1.0005	1.0008	1.0007	1.0013
6	1.0013	1.0018	1.0007	1.0013	1.0003	1.0005
7	1.0006	1.0012	1.0008	1.0014	1.0005	1.0008
8	1.0003	1.0007	1.0006	1.0011	1.0002	1.0005
9	1.0006	1.0012	1.0011	1.0017	1.0002	1.0004
10	1.0004	1.0008	1.0011	1.0019	1.0001	1.0003

Tabla 4-6.: Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 3.

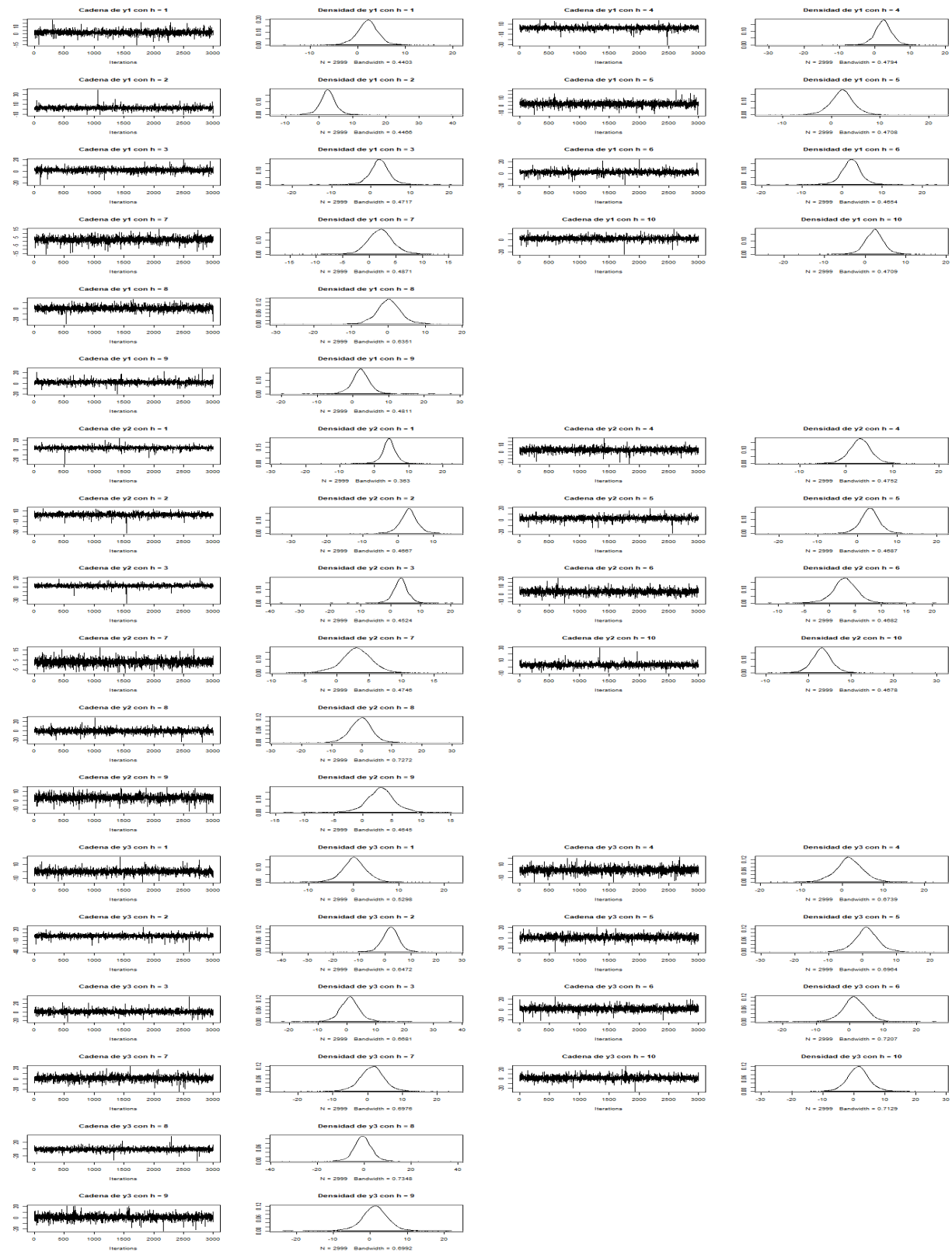


Figura 4-10.: Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Modelo

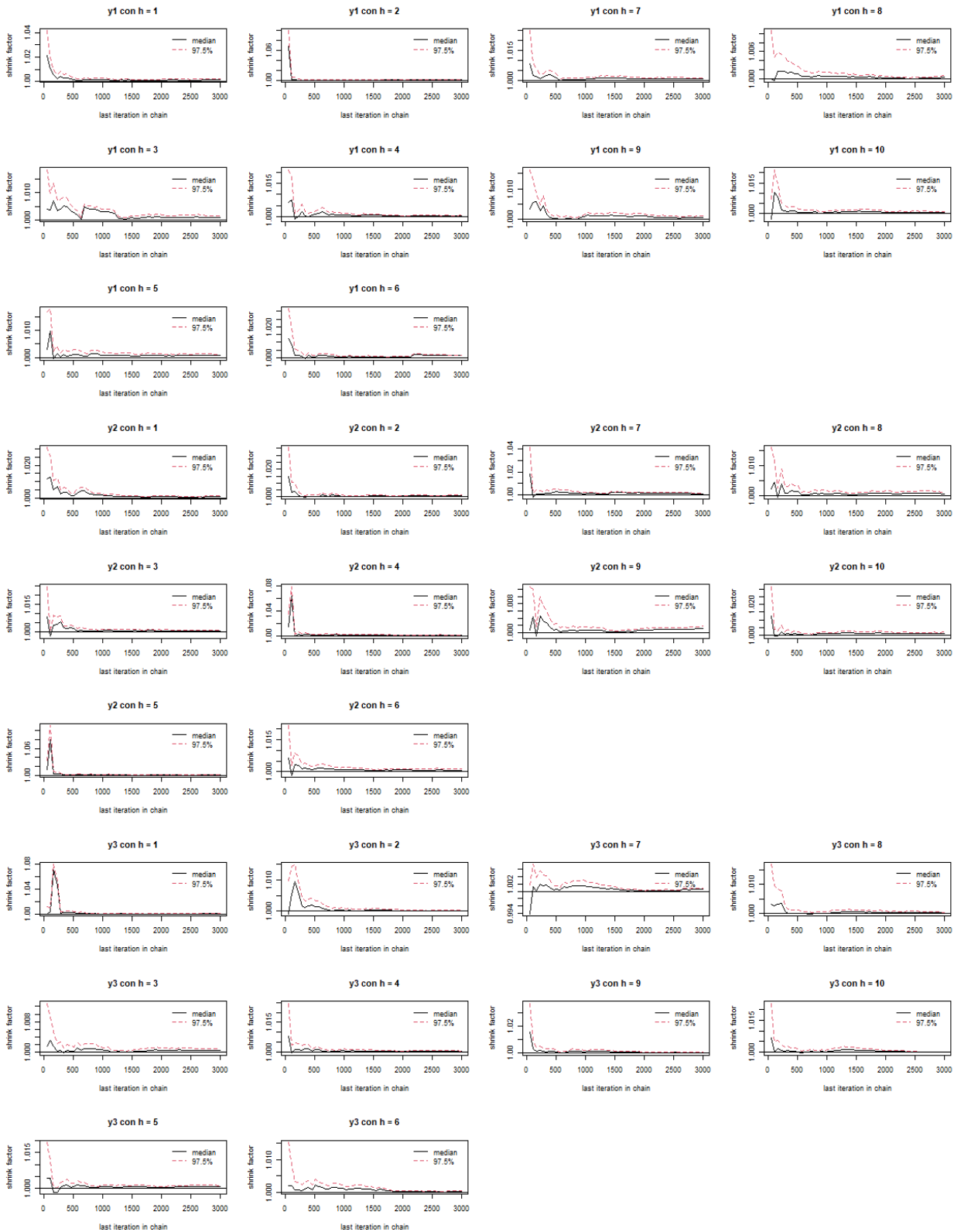


Figura 4-11.: Resultados del criterio de convergencia de Gelman y Rubin para 25 réplicas de los pronósticos de 1 a 10 pasos adelante - Modelo 3.

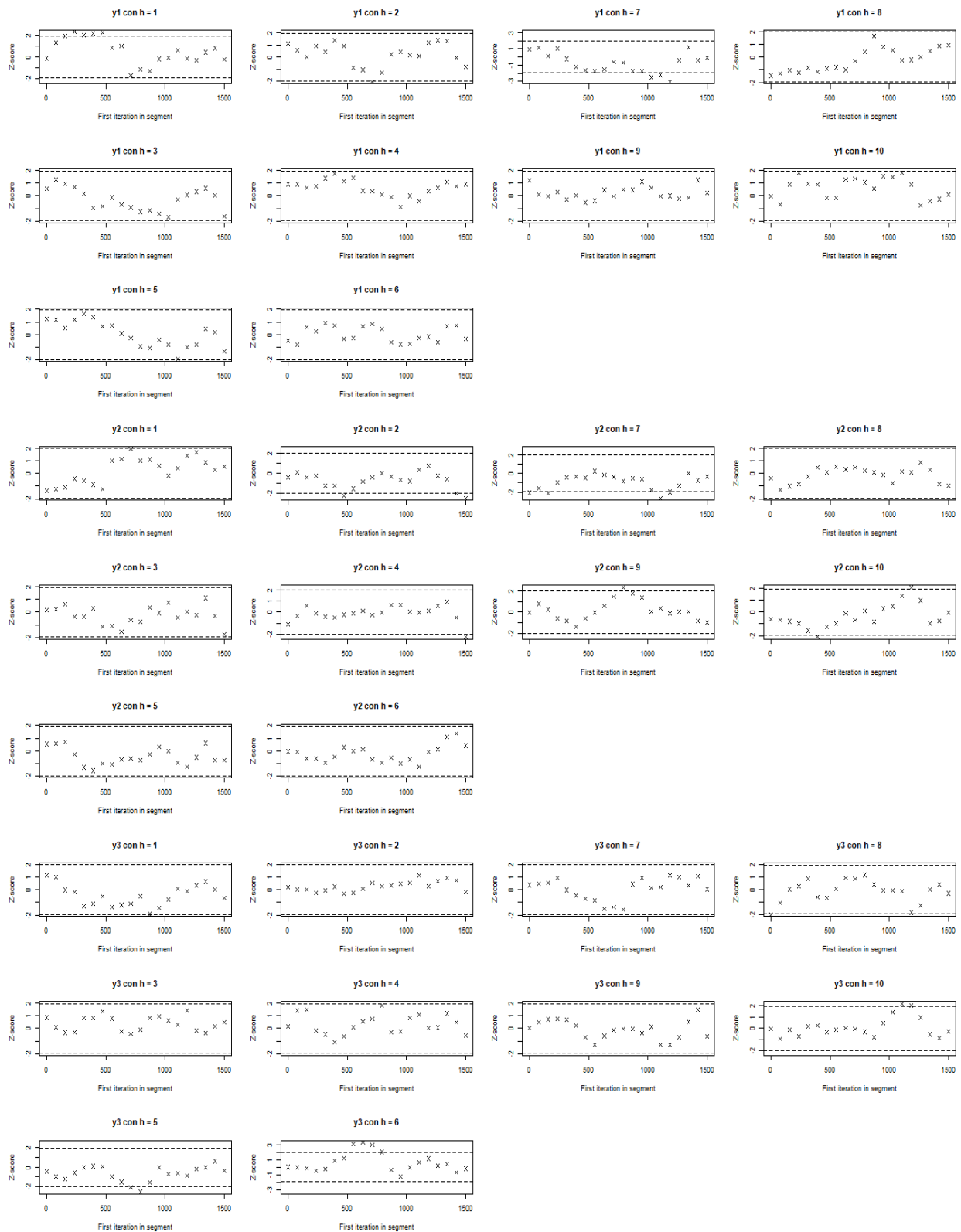


Figura 4-12.: Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Modelo 3.

4.4. Evaluación de la metodología de pronóstico

Ya que aplicar el método propuesto para obtener los pronósticos basados en un modelo MTAR con distribución de error t-Student multivariada desde el enfoque Bayesiano solamente a tres series simuladas no es suficiente para comprobar su validez, se realiza un estudio de simulación ampliado. Para ello, se simulan 100 realizaciones de los tres modelos MTAR considerados cada uno de tamaño 500 y se calcula el porcentaje de veces en que los últimos 10 valores verdaderos del proceso de salida se encuentran dentro del intervalo de predicción del 95 % de la distribución predictiva. La longitud de las cadenas obtenidas es igual a 5000 con un periodo de calentamiento de 2000 (40 %).

Los resultados para los tres modelos simulados se presentan en la tabla 4-7. Las columnas Y1, Y2 y Y3 muestran para cada uno de los tres modelos según sea bivariado o trivariado, el porcentaje de veces que los verdaderos valores del proceso de salida están incluidos dentro de los intervalos de predicción individuales del 95 % para los horizontes de pronóstico de 1 a 10 pasos adelante. El número promedio de veces que los intervalos de predicción contienen los verdaderos valores del proceso de salida para los tres modelos es de 93 %, cercano al 95 % esperado. La tabla 4-7 muestra el rango medio de los intervalos de predicción individual para cada modelo, observando que este es variable y no estrictamente creciente.

Horizonte	Modelo 1			Modelo 2			Modelo 3			
	Y1	Y2	Rango medio IC	Y1	Y2	Rango medio IC	Y1	Y2	Y3	Rango medio IC
1	87	90	7.340	94	99	10.160	94	87	95	10.353
2	90	93	7.790	88	91	11.150	92	93	92	12.347
3	96	95	8.530	91	94	12.030	95	95	90	12.960
4	96	93	9.030	90	88	12.450	97	97	90	13.360
5	95	100	9.080	95	95	12.580	94	91	91	13.433
6	96	94	8.790	97	96	12.460	89	93	94	13.580
7	88	92	9.090	92	92	12.560	92	92	93	13.260
8	94	97	9.090	96	93	12.590	96	91	96	13.527
9	96	95	9.200	99	97	12.610	91	90	95	13.300
10	95	95	9.160	90	93	12.670	97	98	93	13.467

Tabla 4-7.: Porcentaje de veces en que los valores verdaderos del proceso de salida se encuentran dentro del intervalo de predicción del 95 % y rango medio de los intervalos - Modelos 1, 2 y 3.

Los resultados obtenidos con datos simulados sugieren que el procedimiento planteado tiene un desempeño satisfactorio en términos del número de veces que los intervalos de predicción contienen el verdadero valor de $\{Y_t\}$. De igual forma, los porcentajes de cobertura se asemejan a los obtenidos por Calderón y Nieto (2017, 2021) para los pronósticos del modelo MTAR con distribución de error normal multivariada.

5. Aplicación

El objetivo de este capítulo es aplicar la metodología propuesta a un conjunto de datos financieros reales, donde la distribución de error t-Student multivariada es de utilidad ya que puede modelar el exceso de curtosis comúnmente encontrado en series bursátiles. La serie $\{Y_t\}$ está conformada por los retornos del índice BOVESPA de la Bolsa de Valores de Brasil (Y_{1t}) y del índice COLCAP de la Bolsa de Valores de Colombia (Y_{2t}). La variable de umbral $\{Z_t\}$ corresponde a los retornos del índice S&P 500 de Estados Unidos. Para las tres variables, los retornos se definen como la diferencia entre el logaritmo en los tiempos t y $t - 1$.

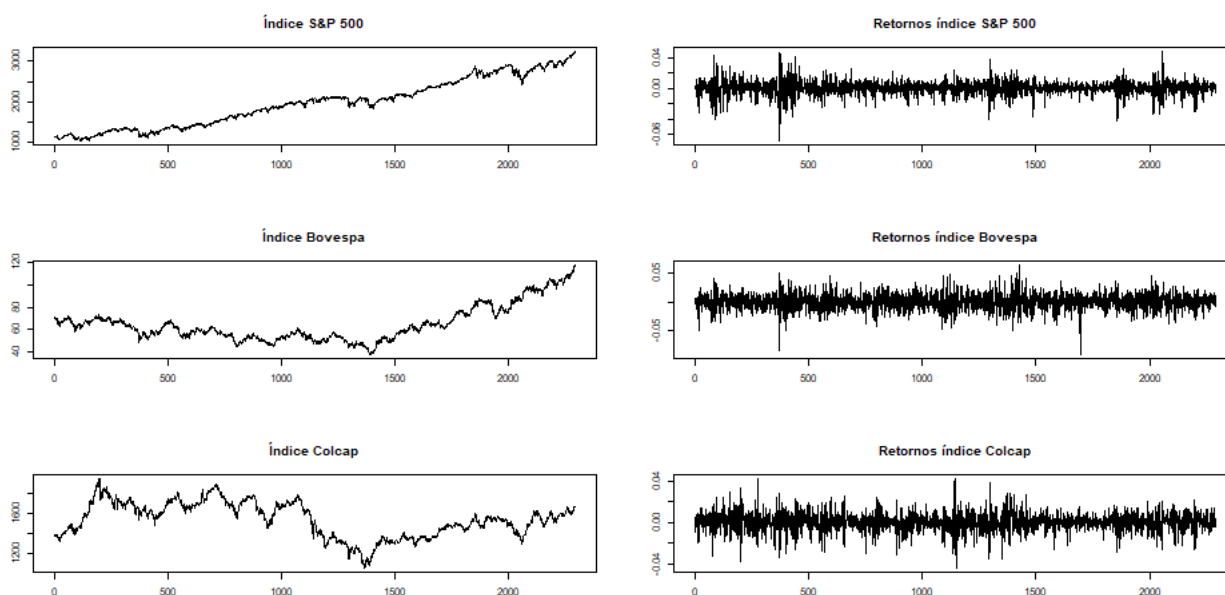


Figura 5-1.: Retornos e Índices S&P 500, COLCAP y BOVESPA - Aplicación.

El COLCAP es un índice de capitalización que refleja las variaciones de los precios de las acciones más líquidas de la Bolsa de Valores de Colombia (BVC), donde la participación de cada acción en el índice está determinada por el correspondiente valor de la capitalización bursátil ajustada. La canasta del índice COLCAP está compuesta por mínimo 20 acciones de 20 emisores diferentes. De manera similar, el BOVESPA es un índice bursátil compuesto por las 50 empresas más líquidas que cotizan en la Bolsa de Valores de Sao Paulo. El índice

está compuesto por los títulos de las empresas que suponen el 80 % del volumen transado en los últimos 12 meses y que fueron negociados por lo menos el 80 % de los días de cotización. Es revisado trimestralmente, para mantener el grado de representación de todas las acciones negociadas en el mercado. Mientras que, el índice S&P 500 se basa en la capitalización bursátil de las 500 más grandes empresas que poseen acciones que cotizan en la Bolsa de Valores de Nueva York o en la Asociación Nacional de Cotizaciones Automatizadas de Comerciantes de Valores (NASDAQ), y es una referencia del estado de la economía de Estados Unidos. La influencia que tiene la economía de Estados Unidos junto a la Reserva Federal sobre demás economías y mercados bursátiles a nivel mundial, hacen que el índice S&P 500 sea una variable apropiada de umbrales para las series elegidas. Colombia y Brasil se pueden entender como *tomadores de precios* en los mercados financieros internacionales, por lo que el índice S&P 500 tiene una influencia significativa sobre sus comportamientos.

Los datos fueron tomados de la plataforma *Investing* y corresponden a las observaciones registradas desde el 5 de enero de 2010 al 30 de diciembre de 2019, para un total de 2290 datos con periodicidad diaria. El periodo de tiempo elegido se caracteriza por una nueva arquitectura del sistema financiero internacional dirigida a reducir el riesgo sistémico, luego de crisis de las hipotecas *subprime* (Espinosa y Nieto, 2016). La muestra de entrenamiento consiste en las primeras 2290 observaciones, mientras que la muestra de prueba está compuesta por las últimas 10 observaciones. En la figura 5-1 se presentan la gráficas de las series elegidas. La estacionariedad de las tres series temporales se verifica a través de las pruebas de raíz unitaria Dickey-Fuller Aumentada (ADF), Phillip-Perron (PP) y Kwiatkowski-Phillips-Schmidt-Shin (KPSS) y sus resultados se encuentran en la tabla A-2 del Anexo A.2. Las tres series son estacionarias según las pruebas realizadas a nivel de significancia de 0.05.

Como se mencionó en el capítulo 2, el primer paso para ajustar un modelo MTAR consiste en determinar si $\{Y_t\}$ sigue un proceso no lineal con umbral. En este caso, se implementa la prueba desarrollada por Lo y Zivot (2001), que es la extensión multivariada del test de no linealidad desarrollado por Hansen (1997, 1999). En la tabla 5-1 se muestran los resultados de la prueba realizada en mención.

Dos regímenes - $m = 2$										
Rezagos en cada régimen										
	1	2	3	4	5	6	7	8	9	10
LR Test	37.38361	45.93660	57.32864	76.50008	90.22438	90.09702	105.97392	119.03742	132.00562	148.07415
P - valor	0.00000	0.00000	0.00000	0.00000	0.10000	0.30000	0.30000	0.00000	0.00000	0.00000
Tres regímenes - $m = 3$										
Rezagos en cada régimen										
	1	2	3	4	5	6	7	8	9	10
LR Test	37.38361	45.93660	57.32864	76.50008	90.22438	90.09702	105.97392	119.03742	132.00562	148.07415
P - valor	0.00000	0.00000	0.00000	0.00000	0.10000	0.30000	0.30000	0.00000	0.00000	0.00000

Tabla 5-1.: Prueba de no linealidad de Lo y Zivot (2001) - Aplicación.

La prueba se realiza tomando 2 y 3 regímenes y 1 a 10 rezagos autorregresivos en cada régimen. En la mayoría de los casos, a excepción de la combinación de 2 y 3 regímenes con 5, 6 y 7 rezagos, se rechaza la hipótesis nula y por lo tanto, hay presencia de una no linealidad en los datos que puede ser explicada mediante un modelo MTAR. Para complementar los resultados de la prueba anterior también se realiza la prueba de no linealidad de Tsay (1998) cuyos resultados se encuentran en la tabla 5-2 con un rezago autorregresivo. El valor máximo del estadístico de prueba se da para $d = 0$, es decir, la variable de umbrales apropiada es Z_t , entendida como los retornos de S&P 500 sin rezagar. En la tabla A-1 del Anexo A.1 se presentan los resultados de la prueba para los ordenes autorregresivos de 1 a 10, donde la interpretación sobre la variable de umbrales es la misma.

Variable	Rezago de la variable de umbral							
	0	1	2	3	4	5	6	7
P-valor	0.00000	0.00061	0.94411	0.46843	0.03719	0.05273	0.02803	0.25736
$\hat{C}(d)$	760.61	23.63	1.71	5.61	13.39	12.45	14.15	7.75

Tabla 5-2.: Prueba de no linealidad de Tsay (1998) - Aplicación.

En el Anexo A se encuentran todos los detalles del procedimiento implementado para estimar un modelo MTAR con distribución de error t-Student multivariada desde el enfoque Bayesiano a partir de las series de retornos de BOVESPA (Y_{1t}), COLCAP (Y_{2t}) y S&P 500 (Z_t) como variable de umbral. El modelo estimado es de la forma:

$$Y_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)}Y_{t-1} + \phi_2^{(1)}Y_{t-2} + \Sigma_{(1)}^{1/2}\varepsilon_t & \text{si } Z_t \leq r \\ \phi_0^{(2)} + \phi_1^{(2)}Y_{t-1} + \Sigma_{(2)}^{1/2} + \varepsilon_t & \text{si } Z_t > r, \end{cases} \quad (5-1)$$

donde $\hat{r} = 0.00139$, $\hat{\nu} = 6.7794$ y la estimación de los demás parámetros es:

$$\begin{aligned} \hat{\phi}_0^{(1)} &= \begin{bmatrix} -0.00150 \\ -0.0050 \end{bmatrix}, & \hat{\phi}_1^{(1)} &= \begin{bmatrix} 0.05170 & 0 \\ 0.05250 & 0.04050 \end{bmatrix}, \\ \hat{\phi}_2^{(1)} &= \begin{bmatrix} 0 & 0 \\ 0 & 0.07950 \end{bmatrix}, & \hat{\Sigma}_{(1)}^{1/2} &= \begin{bmatrix} 0.00730 & 0.00153 \\ 0.00153 & 0.01120 \end{bmatrix}, \\ \hat{\phi}_0^{(2)} &= \begin{bmatrix} 0.0015 \\ 0.0044 \end{bmatrix}, \\ \hat{\phi}_1^{(2)} &= \begin{bmatrix} 0.0794 & 0 \\ 0.0347 & 0 \end{bmatrix}, & \hat{\Sigma}_{(2)}^{1/2} &= \begin{bmatrix} 0.00670 & 0.0010 \\ 0.0010 & 0.0110 \end{bmatrix}. \end{aligned} \quad (5-2)$$

En el primer régimen, los retornos del índice S&P 500 son en su mayoría negativos. Se observa que un periodo atrás, los retornos del índice BOVESPA afectan a los retornos del índice COLCAP, pero no al contrario. Además, para uno y dos periodos atrás, los retornos

de índice COLCAP se ven afectados por rezagos de sí mismo, mientras que los retornos del índice BOVESPA solo se ven afectados por sus mismos rezagos un periodo atrás.

A diferencia del primer régimen, en el segundo los retornos del índice S&P 500 son en su mayoría positivos. Se observa que un periodo atrás, los retornos del índice BOVESPA afectan a los retornos del índice COLCAP, pero no al contrario. Los resultados anteriores indican que el tamaño de economía brasileña hace que su mercado accionario tenga una influencia mayor sobre el mercado accionario colombiano y son similares a los resultados obtenidos por Romero y Calderón (2021).

Una vez estimado el modelo, se procede a obtener los pronósticos de las series elegidas de 1 a 10 pasos adelante. Para generar muestras de las distribuciones predictivas, el algoritmo expuesto se ejecuta con 10000 iteraciones y un periodo de calentamiento de 4000 (40 %) iteraciones. Las cadenas y densidades obtenidas para la obtención de los pronósticos del ejercicio de aplicación se presentan en la figura 5-2. En la tabla 5-3 se presentan los pronósticos e intervalos obtenidos cuando los valores futuros de la variable de umbrales son conocidos. Los verdaderos valores del proceso de salida se encuentran al interior de los intervalos de predicción individuales del 95 % para una realización. Los retornos se mueven alrededor de cero y tienen un comportamiento similar al de la serie original. La figura 5-3 presenta los resultados del criterio de convergencia de Geweke para las cadenas obtenidas. En la figura se observa la convergencia de las cadenas estimadas.

El procedimiento presentado para obtener los pronósticos del modelo MTAR con distribución de error t-Student multivariada asume que los valores en $t + h$ de las variables Z_t y X_t se obtienen con métodos alternos o están disponibles en el conjunto de datos. Para ello se plantea el supuesto de que $Y_{1:t}$ no causa en el sentido de Granger a Z_t y X_t . Sin embargo, es de utilidad ilustrar el proceso de pronóstico en el caso de que los valores de la variable de umbrales no estén disponibles en el conjunto de datos. En este caso, los valores futuros de los retornos del índice S&P 500 se obtienen a partir de un modelo autorregresivo de media móvil (ARIMA) con la variable en niveles, ya que es estacionaria. En el Anexo C.3 se encuentran los detalles del modelo ARIMA estimado donde los residuos no tienen presencia de autocorrelación. Posteriormente se implementa el procedimiento planteado con cadenas de longitud 10000, 4000 (40 %) iteraciones de periodo de calentamiento y con las predicciones de $\{Z_t\}$. Las cadenas empleadas para calcular los pronósticos puntuales e intervalos de predicción del proceso de salida se encuentran en el Anexo C.4. La tabla 5-4 muestra los resultados de predecir 10 días adelante tanto del proceso de salida como de la variable de umbrales $\{Z_t\}$, es decir cuando los valores futuros de la variable de umbrales no son conocidos.

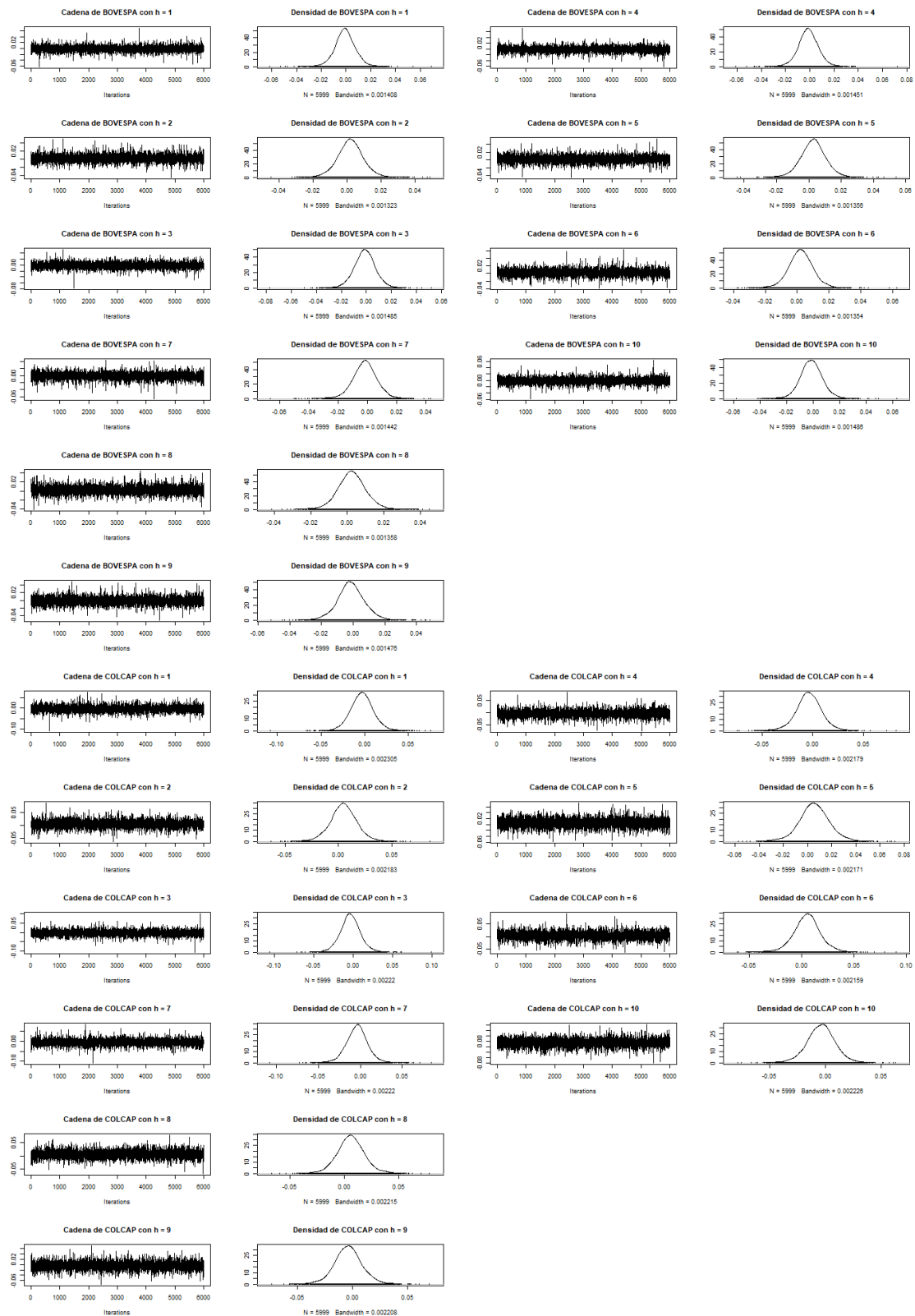


Figura 5-2.: Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Aplicación.

Horizonte de pronóstico	y_{2280+h}	\hat{y}_{2280+h}	I.C. 95 %
1	0.0075	-0.0008	(-0.0181;0.0168)
	0.0032	-0.0031	(-0.0295;0.0247)
2	0.0007	0.0018	(-0.0141;0.018)
	-0.006	0.0047	(-0.0213;0.031)
3	-0.0042	-0.0013	(-0.0189;0.0165)
	0.0064	-0.0038	(-0.0298;0.0232)
4	0.0065	-0.0013	(-0.0191;0.0159)
	0.015	-0.0039	(-0.0311;0.0221)
5	0.0067	0.0028	(-0.0137;0.0184)
	0.0071	0.0056	(-0.0217;0.032)
6	-0.0023	0.0022	(-0.0137;0.0182)
	-0.0001	0.0053	(-0.022;0.0317)
7	0.0136	-0.0016	(-0.0194;0.0153)
	0.0064	-0.0038	(-0.0326;0.0231)
8	-0.004	0.0024	(-0.0132;0.0187)
	0.0115	0.0061	(-0.0198;0.0339)
9	0.0008	-0.0015	(-0.0189;0.0164)
	-0.0057	-0.0039	(-0.031;0.023)
10	0.0007	-0.0014	(-0.0185;0.0162)
	-0.0077	-0.0038	(-0.0318;0.0222)

Tabla 5-3.: Pronósticos de 1 a 10 pasos adelante - Aplicación.

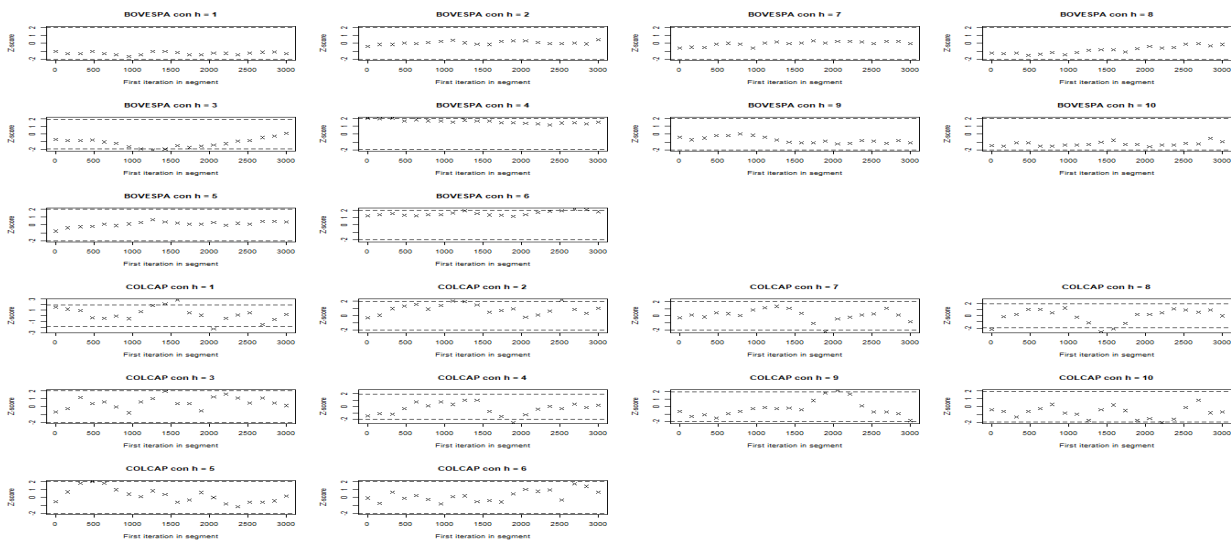


Figura 5-3.: Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Aplicación.

Horizonte de pronóstico	\hat{y}_{2290+h}	I.C. 95 %	\hat{z}_{2290+h}	I.C. 95 %
1	-0.0019	(-0.0189;0.0154)	-0.0002	(-0.0163;0.0159)
	-0.0059	(-0.0325;0.0212)		
2	-0.001	(-0.0197;0.0152)	-0.0000	(-0.0161;0.0161)
	-0.0041	(-0.0329;0.0212)		
3	-0.0019	(-0.02;0.0148)	-0.0003	(-0.0164;0.0158)
	-0.0055	(-0.032;0.0212)		
4	-0.0011	(-0.0192;0.0154)	-0.0001	(-0.0163;0.016)
	-0.0036	(-0.0323;0.0209)		
5	-0.0023	(-0.0198;0.0153)	-0.0002	(-0.0163;0.016)
	-0.0057	(-0.0322;0.021)		
6	-0.0011	(-0.0198;0.0154)	-0.0001	(-0.0163;0.016)
	-0.0043	(-0.0334;0.0214)		
7	-0.0022	(-0.0199;0.0157)	-0.0001	(-0.0163;0.016)
	-0.0055	(-0.0324;0.0211)		
8	-0.0012	(-0.0199;0.0154)	-0.0001	(-0.0162;0.0161)
	-0.004	(-0.0324;0.0212)		
9	-0.0022	(-0.019;0.0159)	-0.0001	(-0.0162;0.0161)
	-0.0055	(-0.0317;0.0204)		
10	-0.0014	(-0.02;0.0156)	-0.0000	(-0.0162;0.0161)
	-0.004	(-0.0331;0.0211)		

Tabla 5-4.: Pronósticos de 1 a 10 pasos adelante - Pronósticos conjuntos - Aplicación.

5.1. Evaluación de los pronósticos

El objetivo de esta sección es comparar la capacidad predictiva de un modelo MTAR con distribución de error t-Student multivariada con la de un modelo MTAR con distribución de error normal multivariada. Este ejercicio es relevante ya que al comparar distintos modelos, es posible evaluar la capacidad predictiva de las predicciones obtenidas y minimizar el riesgo de sesgos o errores en los resultados. Además, se puede observar si los resultados son consistentes entre ellos y si se mantienen incluso cuando se utilizan diferentes enfoques del mismo modelo.

Para ello, se comparan los pronósticos de 1 a 60 pasos adelante con base en el error cuadrático medio (MSE(h)) y error cuadrático medio relativo (RelMSE(h)) en cada horizonte de pronóstico a partir de la metodología de validación cruzada de series de tiempo o método *Rolling*. Además, se realiza la prueba de Diebold y Mariano (1995) para determinar estadísticamente si el pronóstico de un modelo es más preciso que el de otro.

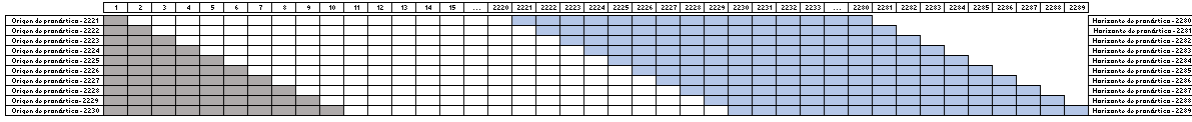


Figura 5-4.: Validación cruzada de los pronósticos obtenidos.

En la figura 5-4 se encuentra el plan de ejecución de validación cruzada de series de tiempo con método *Rolling*. En total, el conjunto de datos está conformado por 2290 observaciones con periodicidad diaria que se dividen en conjuntos de entrenamiento y de prueba. Se consideran 10 muestras de entrenamiento, cuyos puntos iniciales van de la observación número 1 a la 10 y puntos finales de la observación 2220 a la 2229. De esta forma, las 10 muestras de entrenamiento tienen un tamaño igual a 2220 observaciones. Mientras que los puntos iniciales de las 10 muestras de prueba van de la observación 2221 a la 2230 y puntos finales de la observación 2280 a la 2289. El tamaño de las muestras de prueba es constante y es igual a 60 observaciones con lo que es posible evaluar los pronósticos de 1 a 60 pasos adelante. Para ambos modelos MTAR considerados, tanto la etapa de estimación como la etapa de pronóstico se realiza con cadenas de longitud 10000 con 4000 (40%) iteraciones de calentamiento con los parámetros fijos. En el Anexo B se encuentran todos los detalles teóricos del modelo MTAR considerado con distribución normal multivariada.

Horizonte de pronóstico	MSE		RelMSE
	MTAR t-Student	MTAR Normal	
1	0.000097	0.000126	0.771072
5	0.000111	0.000135	0.821427
10	0.000126	0.000147	0.858804
15	0.000131	0.000159	0.821427
20	0.000151	0.000176	0.858804
25	0.000160	0.000194	0.821427
30	0.000189	0.000220	0.858804
35	0.000205	0.000250	0.821427
40	0.000252	0.000294	0.858804
45	0.000287	0.000350	0.821427
50	0.000379	0.000441	0.858804
55	0.000479	0.000583	0.821427
60	0.000757	0.000882	0.858804

Tabla 5-5.: Resultados de validación cruzada - Aplicación.

La tabla 5-5 muestra los resultados resumidos del ejercicio llevado a cabo, en el Anexo C.1 se encuentran los resultados ampliados con todos los horizontes de pronósticos considerados.

Se observa que para el conjunto de datos elegido, la capacidad predictiva del modelo MTAR con distribución de error t-Student es mayor que la del modelo MTAR con distribución de error normal multivariada. La razón principal se debe a que la distribución t-Student es apropiada para modelar fenómenos financieros ya que sus colas más pesadas incluyen rendimientos extremos observados en los mercados financieros, como lo demuestran Bartkowiak (2007) y Wong, Chan y Kam (2009). También se observa que el MSE en ambos modelos es creciente a medida que el horizonte de pronóstico aumenta. A pesar de que la capacidad predictiva del modelo MTAR con distribución de error t-Student es mayor que la del modelo MTAR con distribución de error normal multivariada, ambos modelos tienen una capacidad de pronóstico considerable, ya que permiten que las características de los datos de series temporales, incluidas las medias, las varianzas y los parámetros del modelo, cambien entre regímenes.

Para complementar los resultados anteriores, se obtiene el estadístico propuesto por Harvey, Leybourne y Newbold (1997) (HLN) presentado en la ecuación (2-21) para los horizontes de pronóstico considerados. Se evalúa la hipótesis nula de igual precisión predictiva frente a la hipótesis alternativa cuando los pronósticos obtenidos a partir del modelo MTAR con distribución de error t-Student multivariada son más precisos que los obtenidos a partir del modelo MTAR con distribución de error normal multivariada.

Horizonte de pronóstico	HLN	
	COLCAP	BOVESPA
1	53.17034	67.55761
5	43.71142	46.20000
10	113.1255	105.15121
15	59.50796	60.72916
20	90.72199	84.32690
25	57.65419	57.21297
30	68.31753	63.50175
35	47.26222	45.83838
40	45.91085	42.67454
45	31.48733	29.96072
50	23.49562	21.83939
55	11.92577	11.16529

Tabla 5-6.: Resultados de la prueba de Diebold y Mariano modificada - Aplicación.

La tabla 5-6 muestra los resultados resumidos de la prueba de Diebold y Mariano modificada, en el Anexo C.2 se encuentran los resultados ampliados con todos los horizontes de

pronósticos considerados. El valor de los estadísticos obtenidos permite rechazar la hipótesis nula con un p-valor menor a 0.05 en todos los horizontes considerados, por lo que se sugiere que los pronósticos de COLCAP y BOVESPA obtenidos a partir del modelo MTAR con distribución de error t-Student multivariada son más precisos que los obtenidos a partir del modelo MTAR con distribución de error normal multivariada.

6. Conclusiones

Este documento desarrolló un método analítico para obtener los pronósticos basados en un modelo MTAR con distribución de error t-Student multivariada desde el enfoque Bayesiano. Para ello, se encontró la distribución predictiva Bayesiana que incluye la incertidumbre sobre los verdaderos valores de los parámetros del modelo MTAR. El procedimiento planteado se basa en la obtención de muestras de la distribución predictiva para obtener los pronósticos puntuales e intervalos de predicción del proceso de salida.

Con el objetivo de comprobar el comportamiento de la metodología presentada se realizó un estudio de simulación a partir de tres modelos MTAR. Se ilustró la metodología propuesta para una réplica de cada modelo establecido y se obtuvieron los pronósticos puntuales e intervalos de predicción del proceso de salida que contenían los verdaderos valores de las series simuladas. Posteriormente, se simularon 100 realizaciones de los tres modelos y se calculó el porcentaje de veces que los últimos 10 verdaderos valores del proceso de salida estaban dentro del intervalo de predicción del 95 % de la distribución predictiva. Para los 10 horizontes de pronóstico considerados, se observó que la metodología propuesta tenía un porcentaje de cobertura cercano al esperado de 95 %, por lo que es una alternativa válida para obtener los pronósticos del modelo MTAR con error t-Student multivariados en lo que respecta a escenarios de corto plazo. Además, se pudo verificar que las cadenas obtenidas convergían satisfactoriamente.

Adicionalmente, se aplicó la metodología propuesta a un conjunto de datos financieros. Los verdaderos valores del proceso de salida se encontraron al interior de los intervalos obtenidos. Además, se observó que para el conjunto de datos elegido, la capacidad predictiva del modelo MTAR con distribución de error t-Student es mayor que el modelo MTAR con distribución de error normal multivariada.

Como limitantes de esta investigación, el costo computacional del algoritmo y su implementación en R es bastante alto, lo cual demanda un tiempo considerable para su ejecución. En investigaciones futuras relacionadas, se debe mejorar este aspecto. Se propone como trabajo futuro desarrollar un método que también permita obtener los pronósticos de las covariables $\{X_t\}$ y el umbral $\{Z_t\}$ de manera conjunta. Además, se propone realizar un trabajo de comparación de la precisión de los pronósticos que un modelo MTAR con distribución de error t-Student multivariada puede brindar frente a otros modelos lineales y no lineales

como vectores autorregresivos o redes neuronales.

A. Generalidades del Modelo MTAR ajustado para las series de retornos COLCAP y BOVESPA

A.1. Prueba de no linealidad

p	Variable	Rezago de la variable de umbral							
		0	1	2	3	4	5	6	7
1	P-valor	0.00000	0.00061	0.94411	0.46843	0.03719	0.05273	0.02803	0.25736
	$\hat{C}(d)$	760.61	23.63	1.71	5.61	13.39	12.45	14.15	7.75
2	P-valor	0.00000	0.00144	0.60818	0.73759	0.02242	0.01349	0.08054	0.36095
	$\hat{C}(d)$	775.67	28.61	8.21	6.87	20.81	22.33	16.73	10.96
3	P-valor	0.00000	0.00045	0.60916	0.04460	0.07267	0.00022	0.05599	0.22485
	$\hat{C}(d)$	790.23	38.42	11.96	24.10	22.30	40.41	23.27	17.62
4	P-valor	0.00000	0.00010	0.41888	0.07253	0.28499	0.00145	0.06911	0.34703
	$\hat{C}(d)$	791.68	49.07	18.57	27.36	20.89	41.15	27.56	19.75
5	P-valor	0.00000	0.00014	0.30214	0.11051	0.34875	0.00155	0.11275	0.01921
	$\hat{C}(d)$	791.07	54.56	24.89	30.34	23.97	46.83	30.24	37.82
6	P-valor	0.00000	0.00022	0.28434	0.18077	0.29130	0.00500	0.21300	0.04667
	$\hat{C}(d)$	797.54	59.11	29.60	32.38	29.44	48.29	31.42	39.20
7	P-valor	0.00000	0.00038	0.00784	0.13166	0.21798	0.00595	0.30434	0.04299
	$\hat{C}(d)$	797.12	63.11	51.88	38.74	35.70	52.98	33.43	44.49
8	P-valor	0.00000	0.00037	0.00686	0.17661	0.33303	0.00540	0.31606	0.06820
	$\hat{C}(d)$	798.29	68.92	57.66	41.49	36.98	58.65	37.39	47.00
9	P-valor	0.00000	0.00031	0.00276	0.25823	0.21186	0.00539	0.31545	0.03147
	$\hat{C}(d)$	807.94	75.13	66.65	43.22	44.67	63.86	41.64	55.76
10	P-valor	0.00000	0.00021	0.00755	0.22448	0.02467	0.00168	0.43636	0.01628
	$\hat{C}(d)$	805.33	82.21	67.49	48.60	61.84	73.98	42.81	63.90

Tabla A-1.: Prueba de no linealidad de Tsay (1998) para diferentes rezagos de la variable de umbral - Aplicación.

A.2. Pruebas de raíz unitaria

Serie	Prueba	Hipótesis nula	Estadístico de prueba	Valor crítico al 5%	Intepretación
Retornos S&P 500	ADF	No estacionariedad	-30,857	-1,955	Estacionaria en niveles
	PP		-49,652	-2,86	
	KPSS con constante	Estacionariedad	0,0377	0,463	
Retornos Bovespa	ADF	No estacionariedad	-32,238	-1,955	Estacionaria en niveles
	PP		-47,392	-2,86	
	KPSS con constante	Estacionariedad	0,31	0,463	
Retornos Colcap	ADF	No estacionariedad	-24,719	-1,955	Estacionaria en niveles
	PP		-47,392	-2,86	
	KPSS con constante	Estacionariedad	0,086	0,463	

Tabla A-2.: Pruebas de raíz unitaria ADF, PP y KPSS para las series de retornos S&P 500, COLCAP y BOVESPA - Aplicación.

A.3. Estimación de parámetros estructurales

Número de regímenes y órdenes autorregresivos

Para la elección del número de regímenes y el orden autorregresivo en cada régimen se emplea el criterio de información NAIC definido en la ecuación (2-38).

MTAR	NAIC
MTAR(2,2,3)	-18.30375
MTAR(2,3,3)	-18.30347
MTAR(2,2,2)	-18.30274
MTAR(2,2,1)	-18.30262
MTAR(2,3,1)	-18.30159

Tabla A-3.: NAIC de los 5 mejores modelos considerados - Aplicación.

Se consideran distintas combinaciones de 2 a 3 regímenes y de 1 a 4 rezagos en cada régimen, en la tabla A-3 se presenta el NAIC de los 5 mejores modelos considerados. Al usar este criterio, el modelo a estimar es un MTAR(2,2,3) que se describe en la ecuación (A.1) a continuación:

$$Y_t = \begin{cases} \begin{pmatrix} \phi_{0,1}^{(1)} \\ \phi_{0,2}^{(1)} \end{pmatrix} + \begin{pmatrix} \phi_{1,11}^{(1)} & \phi_{1,12}^{(1)} \\ \phi_{1,21}^{(1)} & \phi_{1,22}^{(1)} \end{pmatrix} Y_{t-1} + \begin{pmatrix} \phi_{2,11}^{(1)} & \phi_{2,12}^{(1)} \\ \phi_{2,21}^{(1)} & \phi_{2,22}^{(1)} \end{pmatrix} Y_{t-2} + \Sigma_{(1)}^{1/2} \varepsilon_t & \text{si } Z_t \leq r \\ \begin{pmatrix} \phi_{0,1}^{(2)} \\ \phi_{0,2}^{(2)} \end{pmatrix} + \begin{pmatrix} \phi_{1,11}^{(2)} & \phi_{1,12}^{(2)} \\ \phi_{1,21}^{(2)} & \phi_{1,22}^{(2)} \end{pmatrix} Y_{t-1} + \begin{pmatrix} \phi_{2,11}^{(2)} & \phi_{2,12}^{(2)} \\ \phi_{2,21}^{(2)} & \phi_{2,22}^{(2)} \end{pmatrix} Y_{t-2} + \begin{pmatrix} \phi_{3,11}^{(2)} & \phi_{3,12}^{(2)} \\ \phi_{3,21}^{(2)} & \phi_{3,22}^{(2)} \end{pmatrix} Y_{t-3} + \Sigma_{(2)}^{1/2} \varepsilon_t & \text{si } Z_t > r. \end{cases} \quad (\text{A.1})$$

Vector de umbrales

Con el número de umbrales y ordenes autorregresivos conocidos, se aplica el algoritmo de estimación expuesto en el Capítulo 2 con 10.000 iteraciones y un periodo de calentamiento del 40 %. En la figura A-1 se muestra la cadena y densidad resultante para r . En la tabla A-4 se presenta la estimación puntual junto con el intervalo de credibilidad del 95 %. Con ello, el valor estimado del umbral \hat{r} es igual a 0.00139.

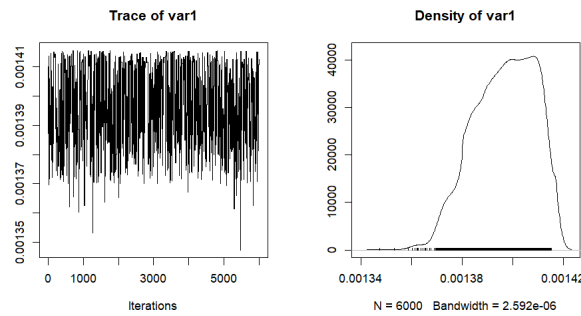


Figura A-1.: Cadena y densidad del vector de umbrales r - Aplicación.

Parámetro	Estimación	$L_i(2.5\%)$	$L_s(97.5\%)$
r	0.00139	0.001372	0.001415

Tabla A-4.: Estimación puntual del vector de umbrales r con su respectivo intervalo de credibilidad del 95 % - Aplicación.

En la figura 5-3 se presenta el criterio de convergencia de Geweke para la cadena obtenida y se observa que los puntajes Z están dentro de las bandas de confianza, lo que indica convergencia de la cadena.

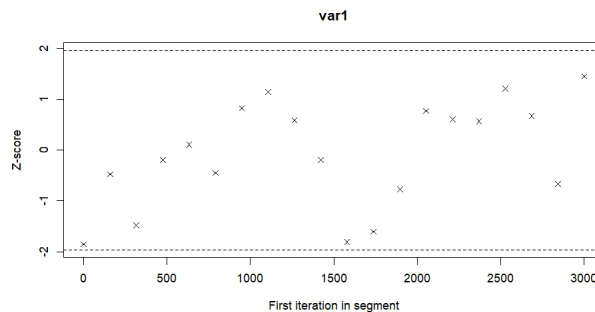


Figura A-2.: Criterio de convergencia de Geweke del vector de umbrales r - Aplicación.

A.4. Estimación de parámetros no estructurales

Matrices autorregresivas

Las cadenas y densidades obtenidas para las matrices autorregresivas de ambos regímenes se presentan en las figuras A-3 y A-4. Se observa que tienen un comportamiento simétrico y las cadenas son estacionarias.

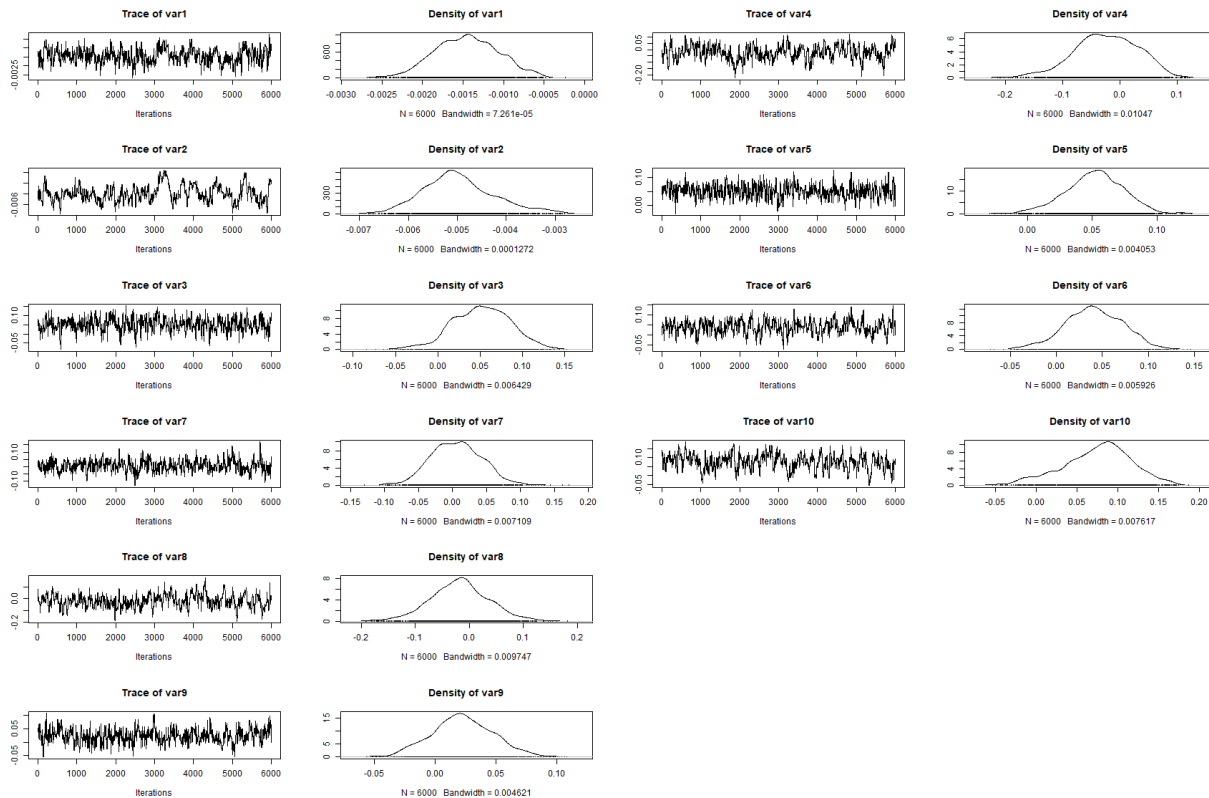


Figura A-3.: Cadenas y densidades de las matrices autorregresivas correspondientes al régimen $j = 1$ - Aplicación.

En las tablas A-5 y A-6 se observan las estimaciones puntuales e intervalos de credibilidad del 95 % para las matrices autorregresivas de ambos regímenes. Los parámetros significativos del primer régimen, es decir, aquellos cuyo intervalo de credibilidad no toma el valor de cero, son $\phi_{0,1}^{(1)}$, $\phi_{0,2}^{(1)}$, $\phi_{1,11}^{(1)}$, $\phi_{1,21}^{(1)}$, $\phi_{1,22}^{(1)}$ y $\phi_{1,22}^{(1)}$. Mientras que los parámetros significativos para el segundo régimen son $\phi_{0,1}^{(2)}$, $\phi_{0,2}^{(2)}$, $\phi_{1,11}^{(2)}$ y $\phi_{1,21}^{(2)}$. En las figuras A-5 y A-6 se presenta el criterio de convergencia de Geweke para las cadena obtenidas y se observa que los puntajes Z están dentro de las bandas de confianza, lo que indica convergencia de la cadenas.

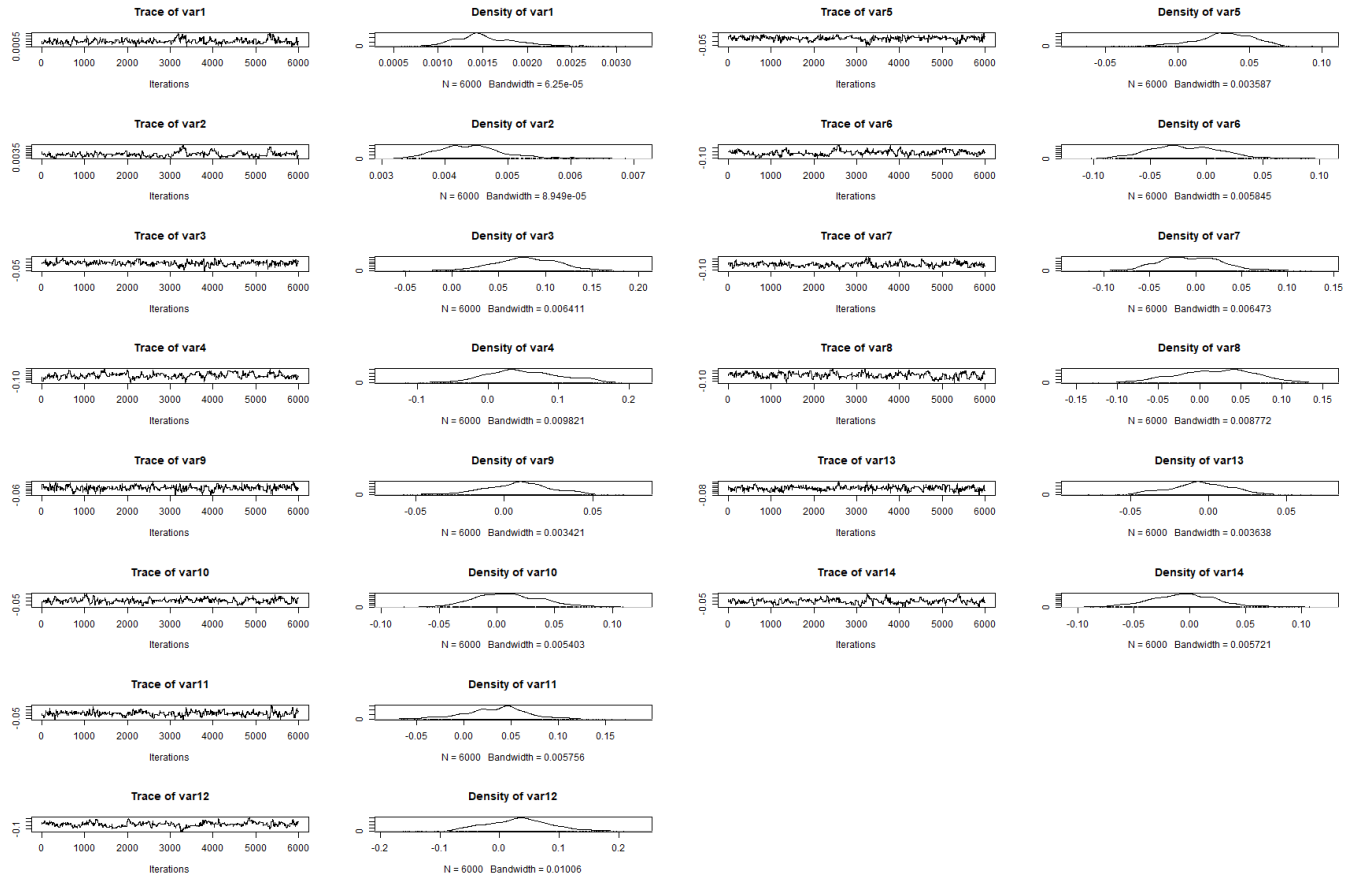


Figura A-4.: Cadenas y densidades de las matrices autorregresivas correspondientes al régimen $j = 2$ - Aplicación.

Parámetro	Estimación	$L_i(2.5\%)$	$L_s(97.5\%)$
$\phi_{0,1}^{(1)}$	-0.001500	-0.002000	-0.000900
$\phi_{0,2}^{(1)}$	-0.005000	-0.005800	-0.004000
$\phi_{1,11}^{(1)}$	0.051700	0.010300	0.093900
$\phi_{1,12}^{(1)}$	-0.022800	-0.094900	0.048600
$\phi_{1,21}^{(1)}$	0.052500	0.022800	0.079700
$\phi_{1,22}^{(1)}$	0.040500	0.001200	0.081400
$\phi_{2,11}^{(1)}$	0.006100	-0.040500	0.055400
$\phi_{2,12}^{(1)}$	-0.017700	-0.085800	0.052400
$\phi_{2,21}^{(1)}$	0.021900	-0.010400	0.055000
$\phi_{2,22}^{(1)}$	0.079500	0.015300	0.125800

Tabla A-5.: Estimación puntual de las matrices autorregresivas correspondientes al régimen $j = 1$ - Aplicación.

Parámetro	Estimación	$L_i(2.5\%)$	$L_s(97.5\%)$
$\phi_{0,1}^{(2)}$	0.0015	0.0011	0.0021
$\phi_{0,2}^{(2)}$	0.0044	0.0038	0.0051
$\phi_{1,11}^{(2)}$	0.0794	0.0353	0.1217
$\phi_{1,12}^{(2)}$	0.0468	-0.0129	0.1289
$\phi_{1,21}^{(2)}$	0.0347	0.0046	0.0570
$\phi_{1,22}^{(2)}$	-0.0227	-0.0569	0.0208
$\phi_{2,11}^{(2)}$	-0.0060	-0.0484	0.0394
$\phi_{2,12}^{(2)}$	0.0224	-0.0468	0.0783
$\phi_{2,21}^{(2)}$	0.0089	-0.0175	0.0340
$\phi_{2,22}^{(2)}$	0.0086	-0.0251	0.0490
$\phi_{3,11}^{(2)}$	0.0374	-0.0085	0.0746
$\phi_{3,12}^{(2)}$	0.0325	-0.0403	0.1033
$\phi_{3,21}^{(2)}$	-0.0054	-0.0342	0.0199
$\phi_{3,22}^{(2)}$	-0.0049	-0.0421	0.0312

Tabla A-6.: Estimación puntual de las matrices autorregresivas correspondientes al régimen $j = 2$ - Aplicación.

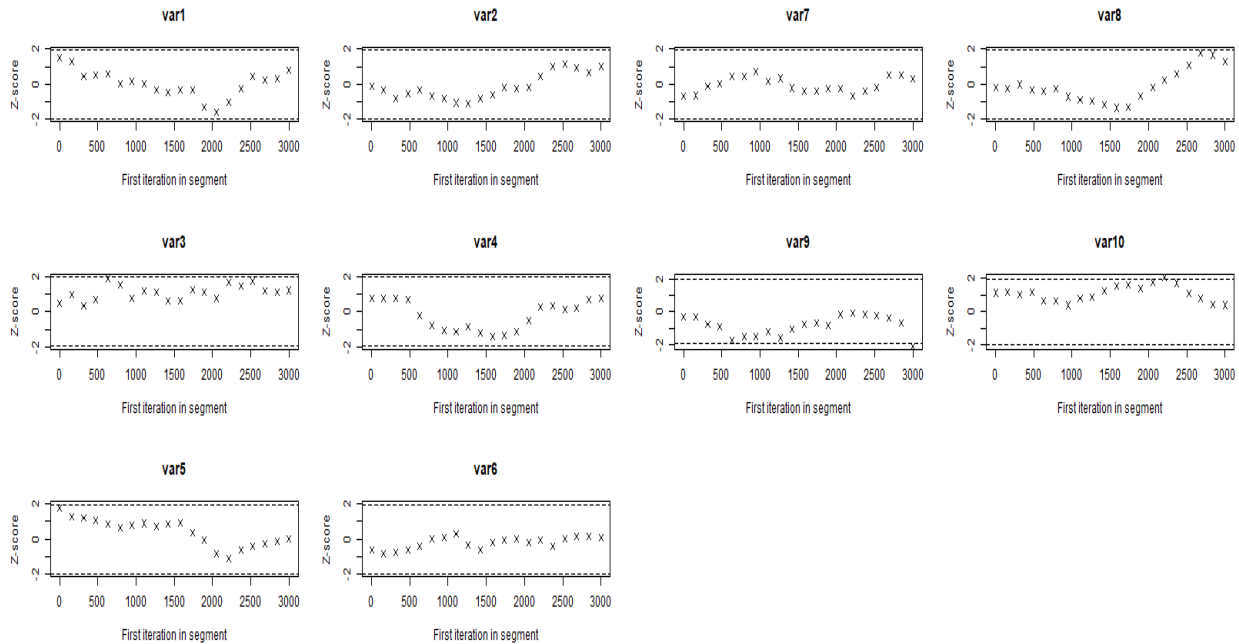


Figura A-5.: Criterio de convergencia de Geweke de las matrices autorregresivas correspondientes al régimen $j = 1$ - Aplicación.

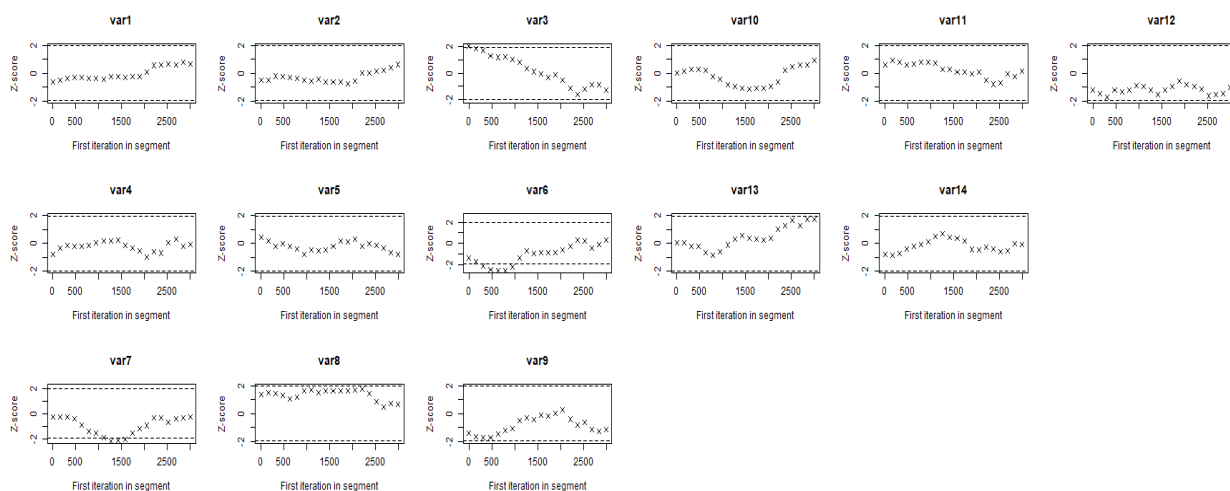


Figura A-6.: Criterio de convergencia de Geweke de las matrices autorregresivas correspondientes al régimen $j = 2$ - Aplicación.

Desviaciones estándar escaladas

En las figuras [A-7](#) y [A-8](#) se presentan las cadenas y densidades para estimar los vectores de desviaciones estándar escaladas de ambos regímenes. Se observa que las densidades son relativamente simétricas y las cadenas no toman el valor cero.

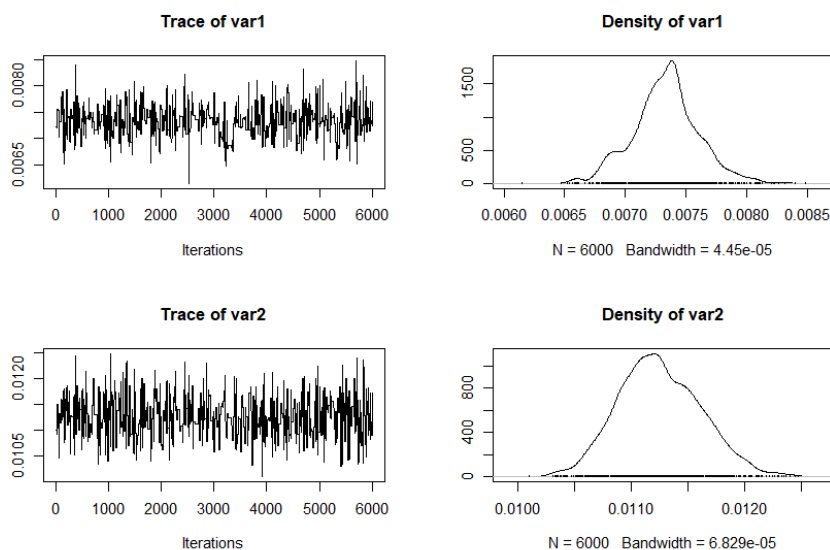


Figura A-7.: Cadenas y densidades del vector de desviaciones estándar escaladas correspondiente al régimen $j = 1$ - Aplicación.

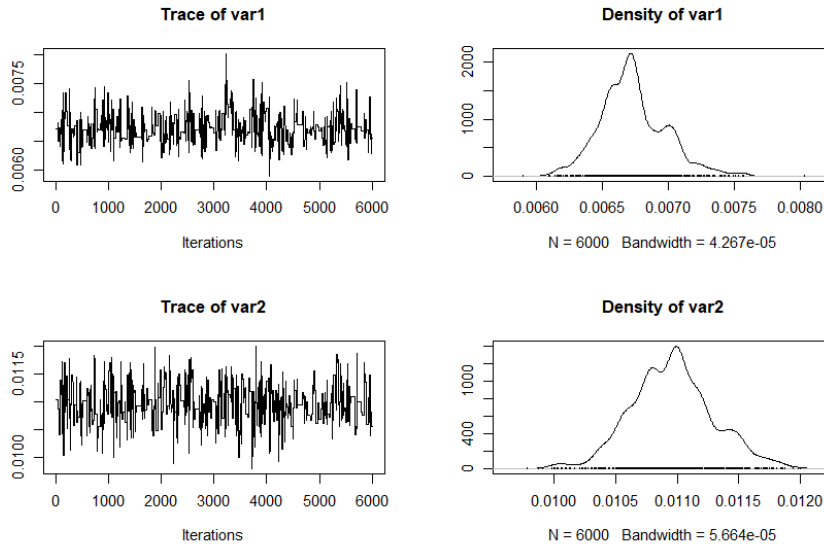


Figura A-8.: Cadenas y densidades del vector de desviaciones estándar escaladas correspondiente al régimen $j = 2$ - Aplicación.

Parámetro	Estimación	$L_i(2.5\%)$	$L_s(97.5\%)$
σ_1	0.0073	0.0068	0.0079
σ_2	0.0112	0.0106	0.0120

Tabla A-7.: Estimación puntual del vector de desviaciones estándar escaladas correspondiente al régimen $j = 1$ - Aplicación.

En las tablas [A-7](#) y [A-8](#) se presenta la estimación puntual de los vectores de desviaciones estándar escaladas para ambos regímenes junto con el intervalo de credibilidad del 95%. En ambos regímenes el primer elemento de desviaciones estándar escaladas es menor que el segundo.

Parámetro	Estimación	$L_i(2.5\%)$	$L_s(97.5\%)$
σ_1	0.0067	0.0063	0.0073
σ_2	0.0110	0.0103	0.0117

Tabla A-8.: Estimación puntual del vector de desviaciones estándar escaladas correspondiente al régimen $j = 2$ - Aplicación.

En las figuras [A-9](#) y [A-10](#) se presenta el criterio de convergencia de Geweke para las cadenas obtenidas y se observa que los valores se encuentran en las bandas de confianza de 2 desviaciones estándar, por consiguiente hay convergencia de las cadenas.

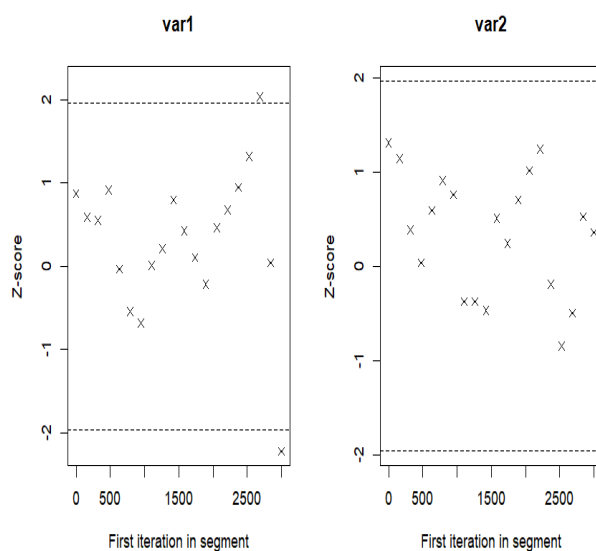


Figura A-9.: Criterio de convergencia de Geweke del vector de desviaciones estándar escaladas correspondiente al régimen $j = 1$ - Aplicación.

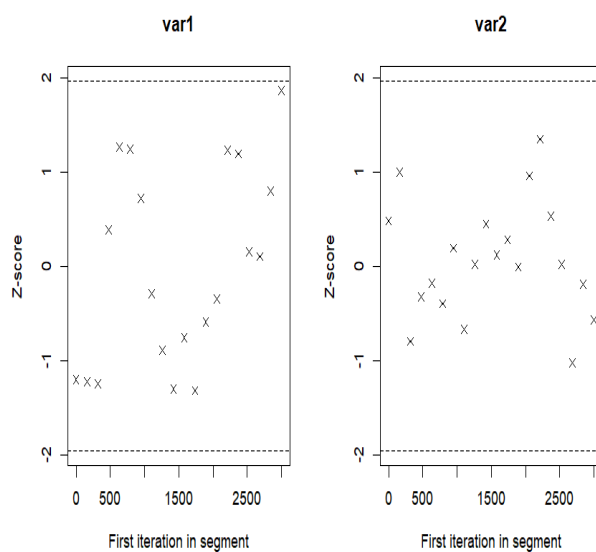


Figura A-10.: Criterio de convergencia de Geweke del vector de desviaciones estándar escaladas correspondiente al régimen $j = 2$ - Aplicación.

Coefficientes de correlación escalados

Las cadenas y densidades obtenidas para la estimación de los coeficientes de correlación escalados de ambos regímenes se presentan en las figuras [A-11](#) y [A-12](#). Se observa que

tienen un comportamiento simétrico y son estacionarias. De manera similar a las cadenas del vector de desviaciones estándar escaladas, las cadenas de los coeficientes de correlación escalados no toman el valor de cero.

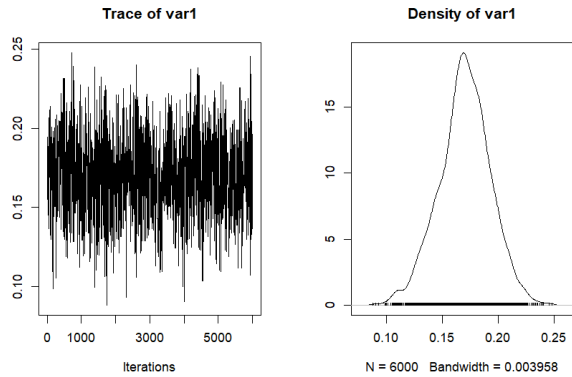


Figura A-11.: Cadenas y densidades del coeficiente de correlación escalado correspondiente al régimen $j = 1$ - Aplicación.

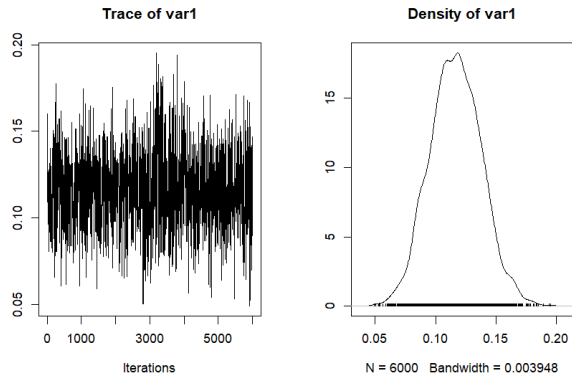


Figura A-12.: Cadenas y densidades del coeficiente de correlación escalado correspondiente al régimen $j = 2$ - Aplicación.

En las tablas A-9 y A-10 se encuentran las estimaciones de los coeficientes de correlación. El parámetro toma el valor de 0.1703 en el primer régimen y 0.1165 en el segundo.

Parámetro	Estimación	$L_i(2.5\%)$	$L_s(97.5\%)$
ρ	0.1703	0.1225	0.2138

Tabla A-9.: Estimación puntual del vector del coeficiente de correlación escalado correspondiente al régimen $j = 1$ - Aplicación.

Parámetro	Estimación	$L_i(2.5\%)$	$L_s(97.5\%)$
ρ	0.1165	0.0758	0.1596

Tabla A-10.: Estimación puntual del vector del coeficiente de correlación escalado correspondiente al régimen $j = 2$ - Aplicación.

En las figuras [A-13](#) y [A-14](#) se presenta el criterio de convergencia de Geweke para las cadenas obtenidas y se observa que los puntajes Z están dentro de las bandas de confianza, lo que indica convergencia de la cadenas.

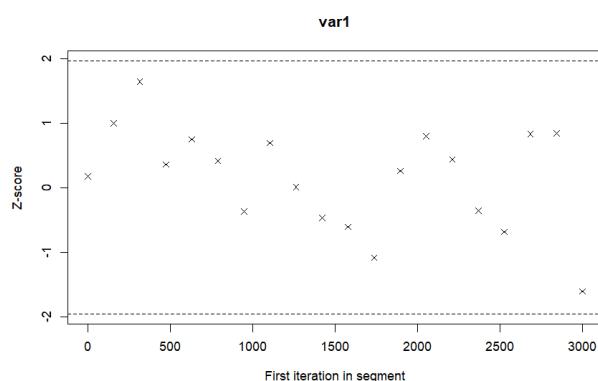


Figura A-13.: Criterio de convergencia de Geweke del coeficiente de correlación escalado correspondiente al régimen $j = 1$ - Aplicación.

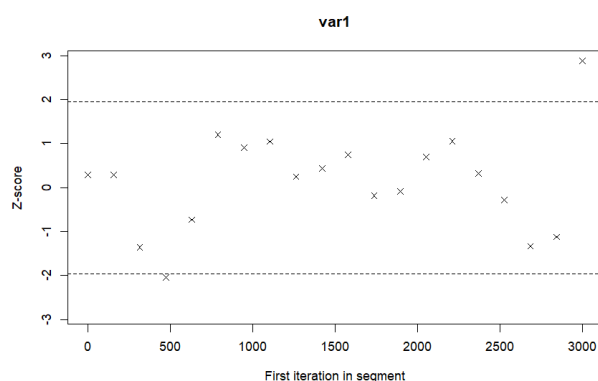


Figura A-14.: Criterio de convergencia de Geweke del coeficiente de correlación escalado correspondiente al régimen $j = 2$ - Aplicación.

Grados de libertad

La figura [A-15](#) muestra la cadena y densidad empleada para estimar los grados de libertad. Se observa que la densidad es simétrica y la cadena es estacionaria. En la tabla [A-11](#) se

presenta la estimación puntual junto con el intervalo de credibilidad del 95%. Con ello, el valor estimado de los grados de libertad $\hat{\nu}$ es igual a 6.7794.

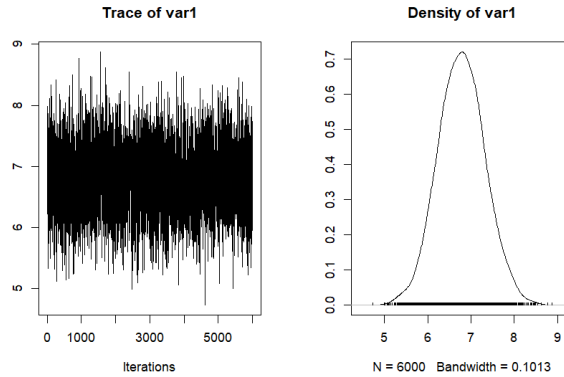


Figura A-15.: Cadena y densidad de los grados de libertad ν - Aplicación.

Parámetro	Estimación	$L_i(2.5\%)$	$L_s(97.5\%)$
ν	6.7794	5.7379	7.8659

Tabla A-11.: Estimación puntual de los grados de libertad ν con su respectivo intervalo de credibilidad del 95% - Aplicación.

En la figura A-16 se presenta el criterio de convergencia de Geweke para la cadena obtenida y se observa que los puntajes Z están dentro de las bandas de confianza, lo que indica convergencia de la cadena.

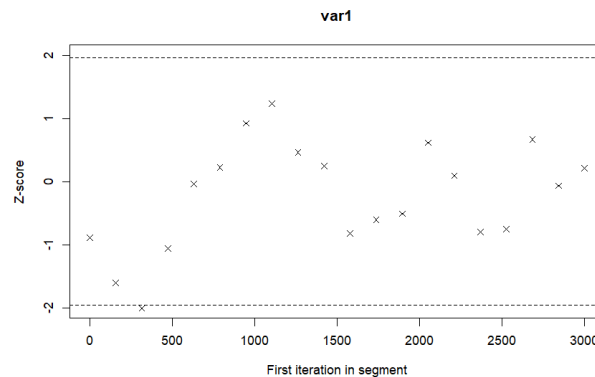


Figura A-16.: Criterio de convergencia de Geweke de los grados de libertad ν - Aplicación.

A.5. Residuos

A partir del modelo estimado se calculan los residuales y se presentan en la figura A-17. En la figura A-18 se presentan los gráficos de CUSUM y CUSUMSQ donde se observa que el ajuste del modelo es bueno. Mientras que en la figura A-19 se observa que los residuales no presentan alta correlación.

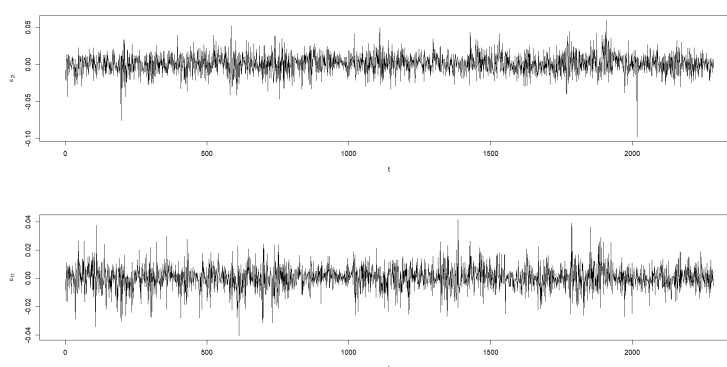


Figura A-17.: Residuales del modelo MTAR estimado - Aplicación.

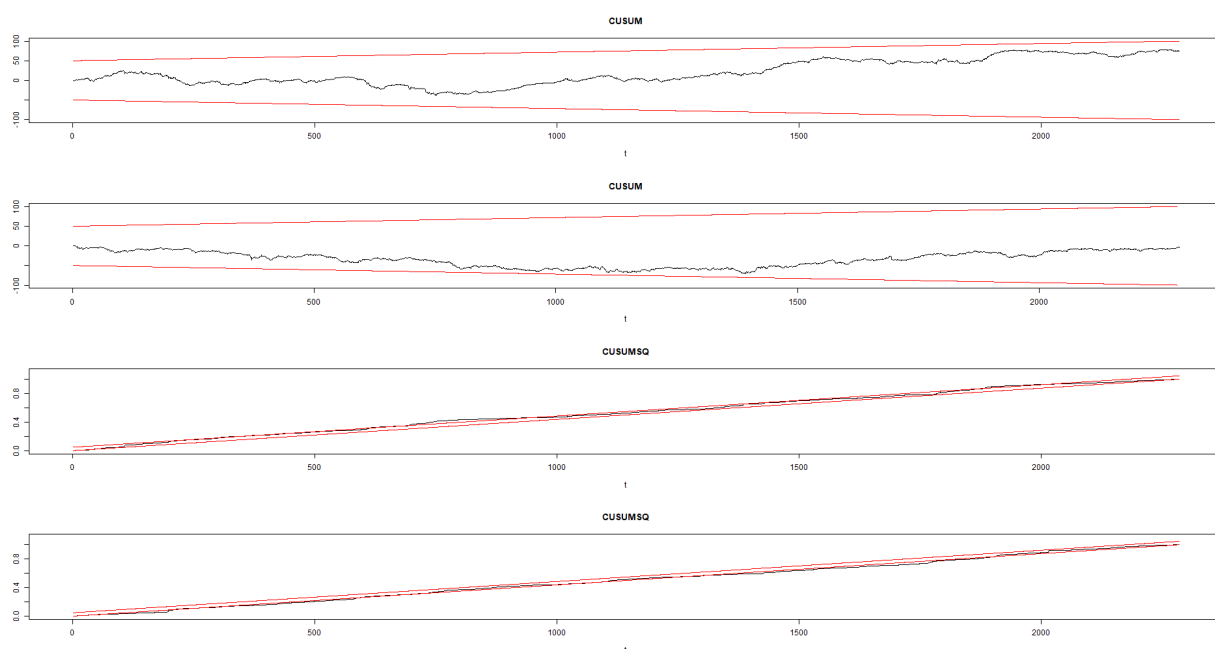


Figura A-18.: CUSUM y CUSUMSQ de los residuales del modelo estimado - Aplicación.

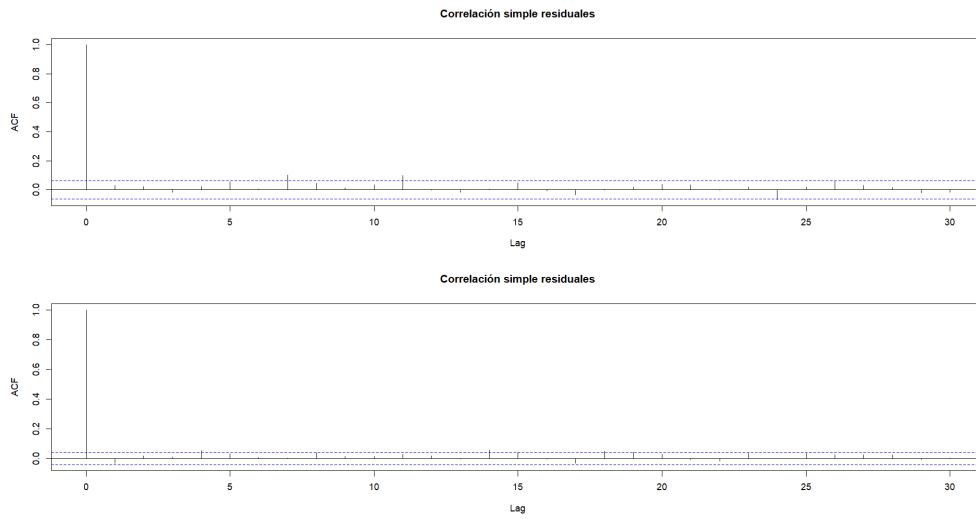


Figura A-19.: Correlaciones de los residuales del modelo estimado - Aplicación.

B. Modelo multivariado autorregresivo de umbrales (MTAR) con distribución de error normal multivariada

El modelo representa la relación entre tres procesos estocásticos $\{Y_t\}$, $\{X_t\}$ y $\{Z_t\}$ de salida, covariables y umbral, respectivamente. Ahora bien, sea $\{Y_t\}$ y $\{X_t\}$ procesos estocásticos tales que $Y_t = (Y_{1t}, \dots, Y_{kt})'$ y $X_t = (X_{1t}, \dots, X_{vt})'$ y sea $\{Z_t\}$ un proceso univariado. $\{Y_t\}$ sigue un proceso MTAR con variable de umbral $\{Z_t\}$ si:

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

donde $j = 1, \dots, l$ con $l \geq 2$ el número de regímenes, y $r_0, r_1, \dots, r_{l-1}, r_l$ son los umbrales del modelo que definen los regímenes tales que $-\infty = r_0 < r_1 < \dots < r_{l-1}, r_l = \infty$. $\{\epsilon_t\}$ es un proceso independiente e idénticamente distribuido, y por cada t , ϵ_t sigue una distribución normal con vector de medias 0 y matriz de covarianza I_k de orden k y es mutuamente independiente de $\{X_t\}$ y $\{Z_t\}$. Esta independencia se refiere a que para cualquier n , los vectores aleatorios $(X_{t_1}, \dots, X_{v_n})$ y $(\epsilon_{t_1}, \dots, \epsilon_{v_n})$ son independientes (Calderón y Nieto, 2017). Para $j = 1, \dots, l$, los parámetros no estructurales del modelo son los coeficientes $\phi_i^{(j)}$ con $i = 0, 1, \dots, p_j$, $\beta_i^{(j)}$ para $i = 0, 1, \dots, q_j$, $\delta_i^{(j)}$ para $i = 0, 1, \dots, d_j$ y $\Sigma_{(j)}^{1/2}$. Teniendo en cuenta lo anterior, el vector de parámetros no estructurales es:

$$\theta_{y_{ns}} = (\theta_1', \dots, \theta_l', \text{vec}(\Sigma)')', \quad (\text{B.2})$$

con $\theta_j = \text{vec}(A_j)$, $A_j = (\phi_0^{(j)}, \phi_1^{(j)}, \dots, \phi_{p_j}^{(j)}, \beta_1^{(j)}, \dots, \beta_{q_j}^{(j)}, \delta_1^{(j)}, \dots, \delta_{d_j}^{(j)})_{k \times \eta_j}$ y $\eta_j = (1 + k \cdot p_j + v \cdot q_j + d_j)$ para $j = 1, \dots, l$ y $\Sigma = (\Sigma_1, \dots, \Sigma_l)$. Los números enteros p_j , q_j y d_j con $j = 1, \dots, l$ representan los ordenes autorregresivos de cada régimen, los valores de los umbrales se denotan como $r = (r_1, \dots, r_{l-1})'$ y el número de regímenes como l , en conjunto se les define como los parámetros estructurales del modelo y se representan como:

$$\theta_{ys} = (p_1, \dots, p_l, q_1, \dots, q_l, d_1, \dots, d_l, r', l)'. \quad (\text{B.3})$$

Teniendo en cuenta lo anterior, el modelo MTAR presentado tiene el siguiente vector de parámetros completos:

$$\theta_y = (\theta_{y_{ns}}, \theta_{ys})'. \quad (\text{B.4})$$

Ahora bien, sea $U_t = (Z_t, X_t)'$ una cadena de Markov homogénea de orden b de tamaño $(v + 1)$, donde b es un número entero mayor que cero con densidad estacionaria $f_u(\cdot)$ y densidad de kernel de transición $f_u(\cdot|\cdot)$. Además, no hay retroalimentación de $\{Y_t\}$ a $\{U_t\}$. Sea $\theta = (\theta_y, \theta_u)$ el vector de parámetros completos del modelo compuesto por el vector de parámetros θ_y del modelo MTAR y el vector de parámetros θ_u la cadena de Markov. Se asume que el proceso probabilístico de la cadena de Markov $u = (u_1, \dots, u_T)$ no depende de los parámetros θ_y , y de manera similar, se asume que la densidad conjunta de $y = (y_1, \dots, y_T)$ condicional a u y (θ_y, θ_u) no depende de θ_u .

El proceso para estimar los parámetros no estructurales del modelo consiste en extraer muestras de la distribución condicional completa $p(\theta_{y_{ns}}|y_{1:T}, u_{1:T}, \theta_{ys})$ de los parámetros no estructurales dado los datos de los procesos de salida y umbral y los parámetros estructurales. En la siguiente exposición se asume que los parámetros estructurales ya han sido estimados previamente con otros métodos. En el caso del modelo MTAR con errores normales, las distribuciones condicionales completas para $j = 1, \dots, l$ son:

$$\begin{aligned} p(\theta_j|\theta_i, i \neq j, \Sigma, u_{1:T}, y_{1:T}), \\ p(\Sigma(j)|\Sigma_{(i)}, i \neq j, \theta_1, \dots, \theta_l, u_{1:T}, y_{1:T}), \end{aligned} \tag{B.5}$$

donde los parámetros entre cada régimen son independientes entre sí y θ_j y $\Sigma_{(j)}$ también son independientes. Para el modelo con la distribución normal en el error, la distribución a priori para θ_j es normal multivariada de media θ_{0j} y matriz de covarianza Σ_{0j} . Calderón y Nieto (2017) demuestran que la distribución condicional completa para $\theta_j|\theta_i, i \neq j$ con $j = 1, \dots, l$ es una distribución normal multivariada de la forma:

$$\begin{aligned} p(\theta_j|\theta_i, i \neq j, \Sigma, u_{1:T}, y_{1:T}, \theta_{y_{ns}}) &\propto \exp\{(\theta_j - \theta_j^*)'V_j^{-1}(\theta_j - \theta_j^*)\} \\ \text{con matriz de covarianza } V_j &= [W_j W_j' \otimes \Sigma_{(j)}^{-1} + \Sigma_{(0j)}^{-1}]^{-1} \\ \text{y media } \theta_j' &= V_j([W_j \otimes \Sigma_{(j)}^{-1}]y_j + \Sigma_{(0j)}^{-1}\sigma_{0j}), \end{aligned} \tag{B.6}$$

donde $w_{t,j} = (1, y'_{t-1}, \dots, y'_{t-p_j}, x_{t-1}, \dots, x_{t-q_j}, z_{t-1}, \dots, z_{t-d_j})'_{\eta_j \times 1}$, $W_j = (w_{t_{1,j}}, \dots, w_{t_{N_j,j}})_{\eta_j \times N_j}$, $t_{1,j}, \dots, t_{N_j,j}$ son los puntos donde $r_{j-1} < Z_t \leq r_j$ y N_j es el número de observaciones en el régimen j . Mientras que la distribución a priori de $\Sigma_{(j)}$ es una inversa de Wishart con matriz de covarianza S_{0j} y v_{0j} grados de libertad. Calderón y Nieto (2017) demuestran que la distribución condicional completa para $\Sigma_{(j)}|\Sigma_{(i)}, i \neq j$ es una inversa de Wishart de la forma:

$$\begin{aligned} p(u_{1:T}, y_{1:T}, \theta_{y_{ns}}) &= |\Sigma_{(j)}|^{-(N_j+v_{0j}+k+1)/2} \exp\left[-\frac{1}{2} \text{tr}\left\{\Sigma_{(j)}^{-1}(S_j + S_{0j})\right\}\right] \\ \text{donde } S_j &= \sum_{\{t:j_t=j\}} \left(y_t - \phi_0^{(j)} \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} x_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} z_{t-i}\right)' \\ &\left(y_t - \phi_0^{(j)} \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \beta_i^{(j)} x_{t-i} + \sum_{i=1}^{d_j} \delta_i^{(j)} z_{t-i}\right), \end{aligned} \tag{B.7}$$

con matriz de covarianza $(S_j + S_{0j})$ y $N_j + v_{0j} + k$ grados de libertad. De esta forma se implementa el muestreador de Gibbs a las distribuciones condicionales anteriores.

Para obtener los pronósticos del modelo MTAR con distribución de normal multivariada presentado en la ecuación (B.1), Calderón y Nieto (2017, 2021) demuestran que la distribución predictiva conjunta, siendo T es el tamaño de la muestra y h el horizonte de pronóstico, es igual a:

$$p(y_{T+1:T+h}, u_{T+1:T+h} | y_{1:T}, u_{1:T}, l) = \int_{\Theta} p(y_{T+1:T+h}, u_{T+1:T+h} | y_{1:T}, u_{1:T}, \theta_{y,l}, l) p(\theta_{y,l} | y_{1:T}, u_{1:T}, l) d\theta_{y,l}, \quad (\text{B.8})$$

donde $p(\theta_{y,l} | y_{1:T}, u_{1:T}, l)$ es la distribución a posteriori de los parámetros de un modelo MTAR con l regímenes y $p(y_{T+1:T+h}, u_{T+1:T+h} | y_{1:T}, u_{1:T}, \theta_{y,l}, l)$ es una distribución que debe ser especificada a partir de los supuestos de las distribuciones empleadas para estimar el modelo MTAR. Para ello se asume que $Y_{1:t}$ no causa en el sentido de Granger a Z_t y X_t . Teniendo en cuenta lo anterior, el primer elemento en la ecuación (B.8) es igual a:

$$p(y_{T+1}, \dots, y_{T+h}, u_{T+1}, \dots, u_{T+h} | y_{1:T}, u_{1:T}, \theta_{y,l}, l) = \prod_{i=1}^h p(u_{T+i} | u_{1:T+i-1}) p(y_{T+i} | u_{T+i}, y_{1:T+i-1}, u_{1:T+i-1}, \theta_{y,l}, l). \quad (\text{B.9})$$

Calderón y Nieto (2017, 2021) demuestran que $p(y_{T+i} | u_{T+i}, y_{1:T+i-1}, u_{1:T+i-1}, \theta_{y,l}, l)$ es una distribución normal multivariada con media $(I_k \otimes w'_{T+i,j})\theta_j$ y matriz de covarianza $\Sigma_{(j)}$ si $r_{j-1} < z_{T+i} \leq r_j$. Mientras que $p(u_{T+i} | u_{1:T+i-1})$ es la densidad Kernel de la cadena de Markov $\{U_t\}$. Para extraer muestras de la distribución predictiva conjunta y así obtener el conjunto $\{y_{T+h}^{(i)}\}$ se implementa el siguiente procedimiento hasta llegar a la iteración i -ésima:

1. Extraer una muestra aleatoria $\theta_{y,l}^{(i)}$ de la distribución posterior $p(\theta_{y,l} | y_{1:T}, u_{1:T}, l)$.
2. Extraer una muestra aleatoria $u_{T+1}^{(i)}$ de la densidad de kernel $p(u_{T+1} | u_{1:T})$.
3. Extraer una muestra aleatoria $y_{T+1}^{(i)}$ de la densidad $p(y_{T+1} | u_{T+1}^{(i)}, y_{1:T}, u_{1:T}, \theta_{y,l}^{(i)}, l)$.
4. Extraer una muestra aleatoria $u_{T+2}^{(i)}$ de la densidad de kernel $p(u_{T+2} | u_{T+1}^{(i)}, u_{1:T})$.
5. Extraer una muestra aleatoria $y_{T+2}^{(i)}$ de la densidad $p(y_{T+2} | u_{T+2}^{(i)}, u_{T+1}^{(i)}, y_{T+1}^{(i)}, y_{1:T}, u_{1:T}, \theta_{y,l}^{(i)}, l)$.
6. Continuar extrayendo muestras aleatorias hasta obtener $u_{T+h}^{(i)}$ de $p(u_{T+h} | u_{T+1}, \dots, u_{T+h-1}, u_T)$ y $y_{T+h}^{(i)}$ de $p(y_{T+h} | u_{T+1}^{(i)}, \dots, u_{T+h}^{(i)}, y_{T+1}^{(i)}, \dots, y_{T+h}^{(i)}, y_{1:T}, u_{1:T}, \theta_{y,l}^{(i)}, l)$.

Con lo anterior se obtiene $\{y_{T+h}^{(i)}, u_{T+h}^{(i)}\}_{i,h}$. De este conjunto se puede calcular una medida de tendencia central como la media o mediana de y_{T+h} y u_{T+h} que es una aproximación numérica al pronóstico puntual (Vaca y Nieto, 2018) y los intervalos creíbles para los valores futuros del vector $(U_{T+1:T+h}, Y_{T+1:T+h})$.

C. Aplicación

C.1. Resultados de la validación cruzada

Horizonte de pronóstico	MSE		RelMSE
	MTAR t-Student	MTAR Normal	
1	0.000097	0.000126	0.771072
2	0.000101	0.000129	0.783271
3	0.000104	0.000129	0.804588
4	0.000104	0.000131	0.793870
5	0.000111	0.000135	0.821427
6	0.000112	0.000138	0.811218
7	0.000114	0.000137	0.833621
8	0.000121	0.000142	0.848709
9	0.000124	0.000145	0.857148
10	0.000126	0.000147	0.858804
11	0.000113	0.000147	0.771072
12	0.000118	0.000151	0.783271
13	0.000122	0.000151	0.804588
14	0.000122	0.000154	0.793870
15	0.000131	0.000159	0.821427
16	0.000132	0.000163	0.811218
17	0.000136	0.000163	0.833621
18	0.000144	0.000169	0.848709
19	0.000149	0.000174	0.857148
20	0.000151	0.000176	0.858804
21	0.000137	0.000177	0.771072
22	0.000143	0.000182	0.783271
23	0.000148	0.000184	0.804588
24	0.000149	0.000188	0.793870
25	0.000160	0.000194	0.821427
26	0.000162	0.000200	0.811218
27	0.000167	0.000200	0.833621

28	0.000178	0.000210	0.848709
29	0.000185	0.000216	0.857148
30	0.000189	0.000220	0.858804
31	0.000172	0.000222	0.771072
32	0.000181	0.000230	0.783271
33	0.000188	0.000233	0.804588
34	0.000190	0.000240	0.793870
35	0.000205	0.000250	0.821427
36	0.000210	0.000259	0.811218
37	0.000218	0.000261	0.833621
38	0.000234	0.000275	0.848709
39	0.000245	0.000286	0.857148
40	0.000252	0.000294	0.858804
41	0.000231	0.000299	0.771072
42	0.000245	0.000313	0.783271
43	0.000257	0.000319	0.804588
44	0.000264	0.000332	0.793870
45	0.000287	0.000350	0.821427
46	0.000298	0.000367	0.811218
47	0.000312	0.000375	0.833621
48	0.000340	0.000400	0.848709
49	0.000361	0.000421	0.857148
50	0.000379	0.000441	0.858804
51	0.000352	0.000457	0.771072
52	0.000381	0.000487	0.783271
53	0.000408	0.000507	0.804588
54	0.000428	0.000540	0.793870
55	0.000479	0.000583	0.821427
56	0.000510	0.000629	0.811218
57	0.000553	0.000663	0.833621
58	0.000623	0.000734	0.848709
59	0.000690	0.000805	0.857148
60	0.000757	0.000882	0.858804

Tabla C-1.: Resultados ampliados de la validación cruzada - Aplicación.

C.2. Resultados de la prueba de Diebold y Mariano modificada

Horizonte de pronóstico	HLN	
	COLCAP	BOVESPA
1	53.17034	67.55761
2	45.00728	50.96195
3	41.92945	45.82230
4	41.35270	44.74404
5	43.71142	46.20000
6	47.60183	50.13958
7	55.42916	57.47422
8	68.11971	70.20708
9	88.66496	89.90897
10	113.12556	105.15121
11	94.21573	95.53762
12	76.57515	79.15239
13	66.11449	68.45382
14	60.58388	62.82234
15	59.50796	60.72916
16	60.53782	61.48470
17	65.10283	65.05445
18	72.49165	71.59028
19	82.38938	79.86378
20	90.72199	84.32690
21	82.44225	79.91502
22	72.23822	71.51648
23	64.48137	64.41378
24	59.51101	59.72800
25	57.65419	57.21297
26	57.29728	56.55932
27	59.15594	57.58627
28	62.28930	60.00288
29	66.00178	62.74740
30	68.31753	63.50175
31	63.90595	60.75491
32	58.15666	56.14009
33	53.10706	51.71283

34	49.33438	48.24876
35	47.26222	45.83838
36	46.00798	44.38630
37	45.82756	43.75023
38	46.08023	43.67046
39	46.34791	43.58785
40	45.91085	42.67454
41	43.09835	40.53180
42	39.69232	37.68540
43	36.43135	34.80126
44	33.62271	32.19745
45	31.48733	29.96072
46	29.65512	28.07529
47	28.15222	26.47316
48	26.71748	25.03010
49	25.21675	23.55176
50	23.49562	21.83939
51	21.24273	19.84014
52	18.83664	17.67540
53	16.43969	15.47145
54	14.12556	13.29231
55	11.92577	11.16529
56	9.75715	9.09252
57	7.59390	7.05542
58	5.40216	5.01774
59	3.14164	2.92018

Tabla C-2.: Resultados ampliados de la prueba de Diebold y Mariano modificada - Aplicación.

C.3. ARIMA - Proceso de umbrales

Coefficiente	Valor estimado	Error estándar
ar1	0.46667328	0.30942888
ar2	0.47724569	0.29414831
ma1	-0.48234619	0.31025028
ma2	-0.54093292	0.29053403
ma3	0.04334495	0.02852557
Intercepto	0.00045883	0.00007498

Tabla C-3.: Modelo ARIMA(2,0,3) estimado para S&P 500 - Aplicación.

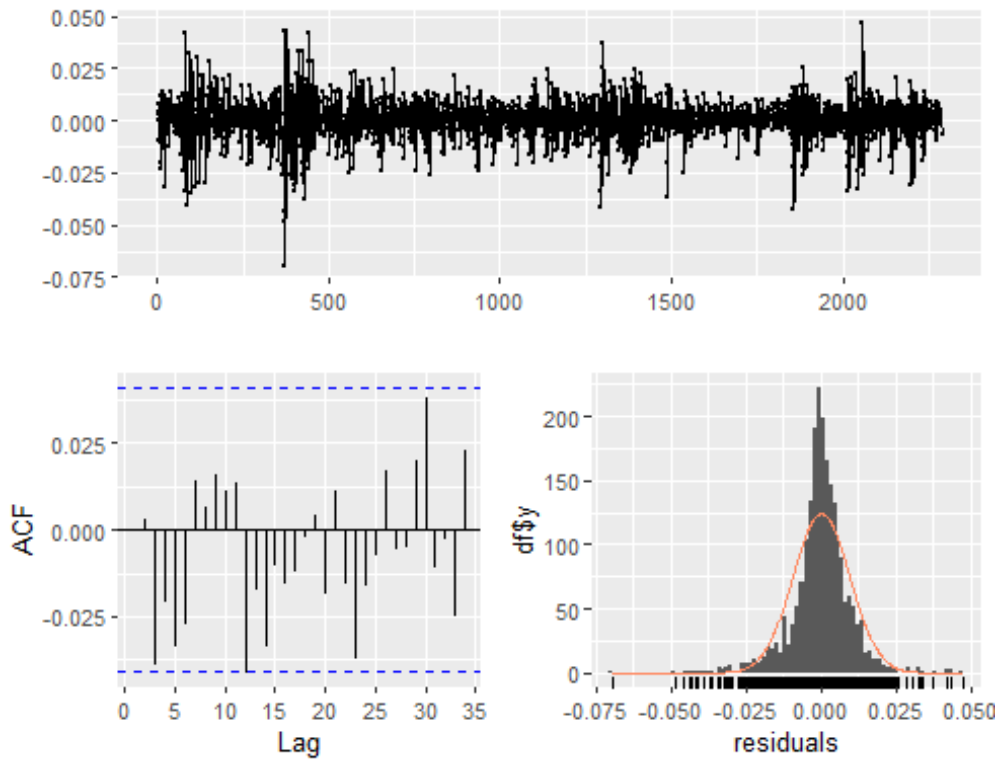


Figura C-1.: Residuos del modelo ARIMA de S&P 500 - Aplicación.

Rezagos	Estadístico	P-Valor
1	0.00000	0.99983
2	0.00055	0.99972
3	0.20649	0.97654
4	1.33415	0.85555
5	1.33535	0.93125
6	1.63223	0.95023
7	1.95780	0.96214
8	1.96408	0.98209
9	2.17808	0.98833
10	4.31899	0.93181
11	6.01596	0.87230
12	6.77347	0.87221
13	9.65143	0.72216
14	9.82066	0.77517
15	9.82717	0.83047
16	10.15141	0.85860
17	10.16236	0.89666
18	10.55325	0.91234
19	11.37102	0.91073
20	11.39556	0.93531

Tabla C-4.: Prueba Ljung-Box S&P 500 - Aplicación.

C.4. Cadenas y densidades de pronósticos conjuntos

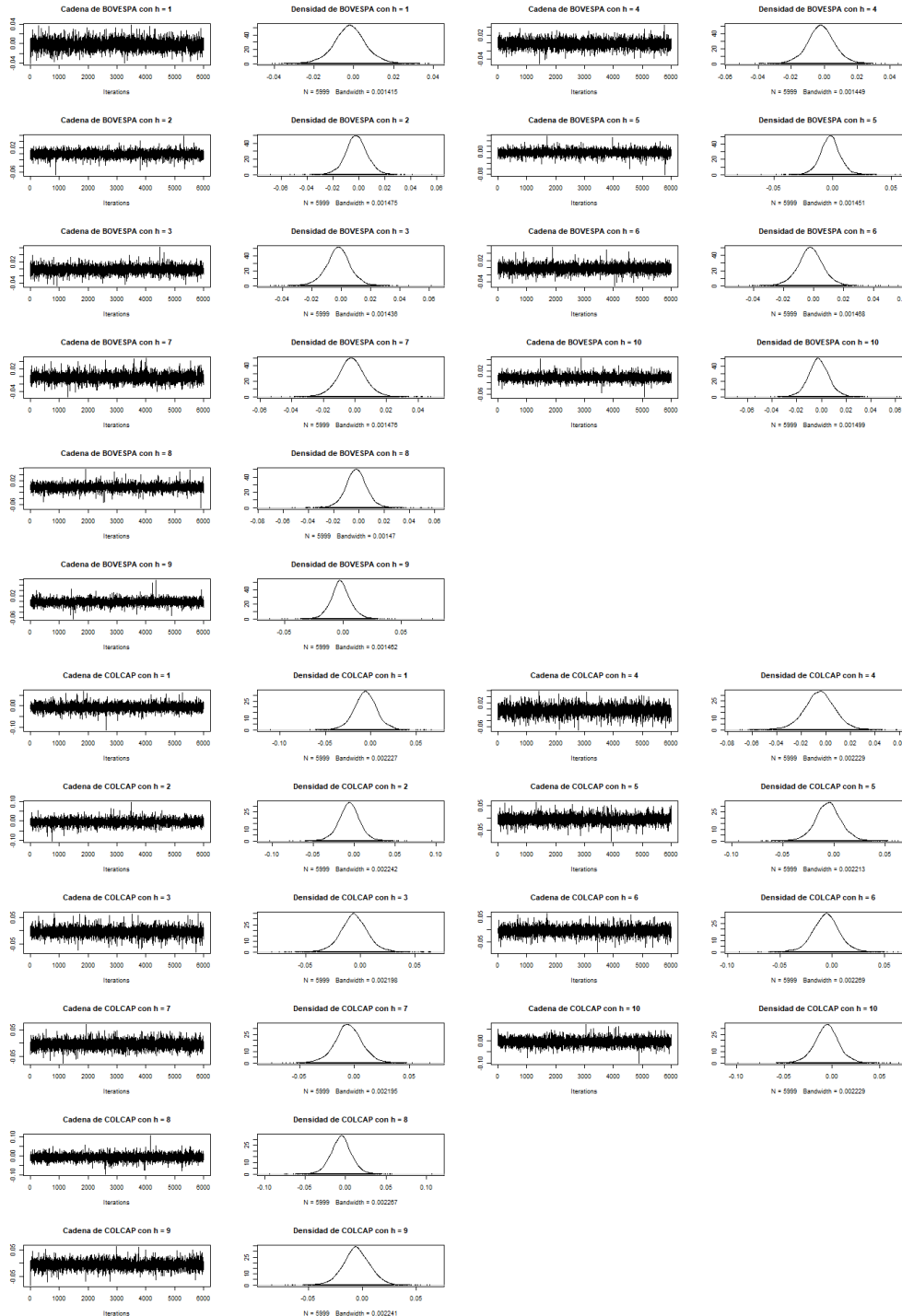


Figura C-2.: Cadenas y densidades para los pronósticos de 1 a 10 pasos adelante - Semana extra - Aplicación.

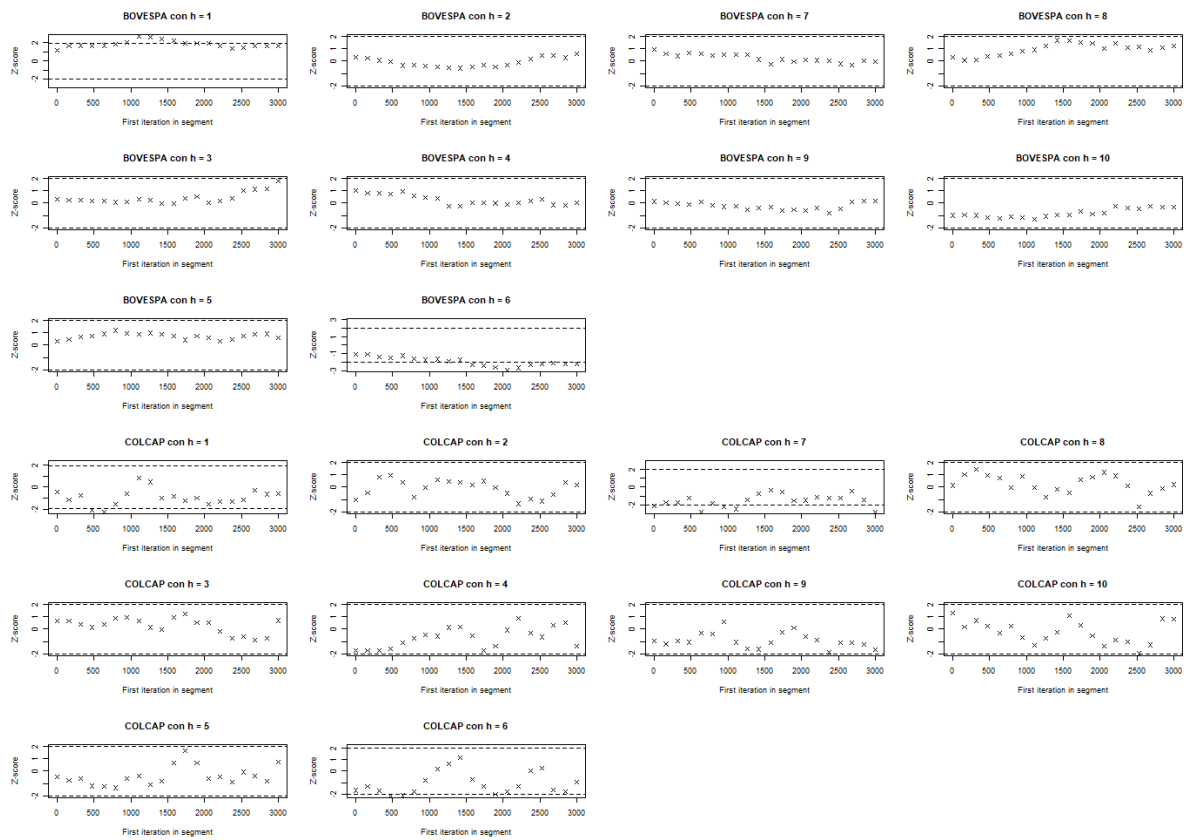


Figura C-3.: Criterio de convergencia de Geweke para los pronósticos de 1 a 10 pasos adelante - Semana extra - Aplicación.

Referencias

- Anthony, M., y Harvey, M. (2012). *Linear algebra: concepts and methods*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511844249>
- Barnard, J., McCulloch, R., y Meng, X. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4), 1281-1311.
- Bartkowiak, A. (2007). Should normal distribution be normal? The student's t alternative. En *6th international conference on computer information systems and industrial management applications (cisim'07)* (p. 3-8). DOI: <https://doi.org/10.1109/CISIM.2007.59>
- Bickel, P., y Doksum, K. (2015). *Basic ideas and selected topics, volumes i-ii* (2nd ed.). Chapman and Hall/CRC.
- Calderón, S., y Nieto, F. (2017). Bayesian analysis of multivariate threshold autoregressive models with missing data. *Communications in Statistics - Theory and Methods*, 46(1), 296-318. DOI: <https://doi.org/10.1080/03610926.2014.990758>
- Calderón, S., y Nieto, F. (2021). Forecasting with multivariate threshold autoregressive models. *Revista Colombiana de Estadística*, 44(2), 369-383. DOI: <https://doi.org/10.15446/rce.v44n2.91356>
- Chen, C., y Lee, J. (1995). Bayesian inference of threshold autoregressive. *Journal of Time Series Analysis*, 16(5), 483-492. DOI: <https://doi.org/10.1111/j.1467-9892.1995.tb00248.x>
- Chiu, C., Mumtaz, H., y Pintér, G. (2017). Forecasting with VAR models: Fat tails and stochastic volatility. *International Journal of Forecasting*, 33(4), 1124-1143. DOI: <https://doi.org/10.1016/j.ijforecast.2017.03.001>
- Debski, W. (2010). Chapter 1: Probabilistic inverse theory. En R. Dmowska (Ed.), *Advances in geophysics* (Vol. 52, p. 1-102). Elsevier. DOI: [https://doi.org/10.1016/S0065-2687\(10\)52001-6](https://doi.org/10.1016/S0065-2687(10)52001-6)
- Diebold, F., y Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business Economic Statistics*, 13(3), 253-63.
- Espinosa, O., y Nieto, F. (2016). Estudio del efecto de apalancamiento en series financieras usando un modelo TAR. *Universidad Nacional de Colombia, Bogotá, D.C.*
- Gelman, A., Carlin, J., Stern, H., y Rubin, D. (2004). *Bayesian data analysis* (2nd ed.). Chapman and Hall/CRC.
- Gelman, A., y Rubin, D. (1992). Inference from iterative simulation using multiple sequences.

- Statistical Science*, 7(4), 457 - 472. DOI: <https://doi.org/10.1214/ss/1177011136>
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. En *In bayesian statistics* (pp. 169–193). University Press.
- González, J., y Nieto, F. (2020). Bayesian analysis of multiplicative seasonal threshold autoregressive processes. *Revista Colombiana de Estadística*, 43(2), 251–284. DOI: <https://doi.org/10.15446/rce.v43n2.81261>
- Hansen, B. (1997). Inference in TAR models. *Studies in nonlinear dynamics and econometrics*, 2, 1-14.
- Hansen, B. (1999). Testing for linearity. *Journal of Economic Surveys*, 13(5), 551-576. DOI: <https://doi.org/10.1111/1467-6419.00098>
- Hansen, B. (2011). Threshold autoregression in economics. *Statistics and Its Interface Volume*, 4, 123-127. DOI: <https://doi.org/10.4310/SII.2011.v4.n2.a4>
- Harvey, D., Leybourne, S., y Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281-291. DOI: [https://doi.org/10.1016/S0169-2070\(96\)00719-4](https://doi.org/10.1016/S0169-2070(96)00719-4)
- Hyndman, R., y Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). Australia: OTexts.
- Ibáñez, L., y Calderón, S. (2020). Estimación bayesiana de los parámetros estructurales de los modelos multivariados autoregresivos de umbrales con ruido t-student multivariado. *Universidad Nacional de Colombia, Bogotá, D.C.*
- Karlsson, S. (2013). Forecasting with bayesian vector autoregression. En G. Elliott y A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 2, p. 791-897). Elsevier. DOI: <https://doi.org/10.1016/B978-0-444-62731-5.00015-4>
- Kibria, G., y Joarder, A. (2006). A short review of multivariate t-distribution. *Journal of Statistical Research ISSN*, 40, 256-422.
- Li, D., y Tong, H. (2016). Nested sub-sample search algorithm for estimation of threshold models. *Statistica Sinica*, 26(4), 1543-1554.
- Lo, M., y Zivot, E. (2001). Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics*, 5(4), 533–576. DOI: <https://doi.org/10.1017/S1365100501023057>
- Nieto, F. (2005). Modeling bivariate threshold autoregressive processes in the presence of missing data. *Communications in Statistics - Theory and Methods*, 34(4), 905-930. DOI: <https://doi.org/10.1081/STA-200054435>
- Nieto, F. (2008). Forecasting with univariate TAR models. *Statistical Methodology*, 5(3), 263-276. DOI: <https://doi.org/10.1016/j.stamet.2007.09.002>
- Nieto, F., y Hoyos, M. (2011). Testing linearity against a univariate TAR specification in time series with missing data. *Revista Colombiana de Estadística*, 34(1), 73-94.
- Nieto, F., Zhang, H., y Li, W. (2013). Using the reversible jump MCMC procedure for identifying and estimating univariate TAR models. *Communications in Statistics - Simulation and Computation*, 42(4), 814-840. DOI:

- <https://doi.org/10.1080/03610918.2012.655827>
- Romero, L., y Calderón, S. (2021). Bayesian estimation of a multivariate tar model when the noise process follows a student-t distribution. *Communications in Statistics - Theory and Methods*, 50(11), 2508-2530. DOI: <https://doi.org/10.1080/03610926.2019.1669807>
- Tong, H. (1978). On a threshold model. In: C. Chen (Ed.), *Pattern recognition and signal processing*. Amsterdam: Sijhoff Noordhoff., 42, 575-586. DOI: https://doi.org/10.1007/978-94-009-9941-1_24
- Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford: Oxford Statistical Science, Oxford University Press.
- Tong, H. (2015). Threshold models in time series analysis - some reflections. *Journal of Econometrics*, 189(2), 485-491. DOI: <https://doi.org/10.1016/j.jeconom.2015.03.039>
- Tong, H., y Lim, K. (1980). Threshold autoregression, limit cycles. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(3), 245-268. DOI: <https://doi.org/10.1111/j.2517-6161.1980.tb01126.x>
- Tsay, R. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, 93(443), 1188-1202. DOI: <https://doi.org/10.1080/01621459.1998.10473779>
- Tsay, R., y Chen, R. (2019). *Nonlinear time series analysis*. John Wiley Sons, Inc.
- Vaca, P., y Nieto, F. (2018). Analysis of the forecasting performance of the threshold autoregressive model. *Universidad Nacional de Colombia, Bogotá, D.C.*
- Wong, S., Chan, W., y Kam, P. (2009). A Student t-mixture autoregressive model with applications to heavy-tailed financial data. *Biometrika*, 96(3), 751-760. DOI: <https://doi.org/10.1093/biomet/asp031>
- Zhang, H., y Nieto, F. (2015). TAR modeling with missing data when the white noise process follows a student's t-distribution. *Revista Colombiana de Estadística*, 38(1), 239-265. DOI: <https://doi.org/10.15446/rce.v38n1.48813>
- Zivot, E., y Wang, J. (2005). Modeling financial time series with S-Plus. DOI: <https://doi.org/10.1007/978-0-387-32348-0>