



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado

Facultad de Ciencias Matemáticas

Unidad de Posgrado

**Un enfoque bayesiano en modelos heterocedásticos de
series de tiempo y su aplicación en la volatilidad de
activos financieros**

TESIS

Para optar el Grado Académico de Magíster en Estadística
Matemática

AUTOR

Edwin Antero FLORES MONTOYA

ASESOR

Mg. Antonio BRAVO QUIROZ

Lima, Perú

2022



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Flores, E. (2022). *Un enfoque bayesiano en modelos heterocedásticos de series de tiempo y su aplicación en la volatilidad de activos financieros*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Edwin Antero Flores Montoya
Tipo de documento de identidad	DNI
Número de documento de identidad	45218056
URL de ORCID	https://orcid.org/0000-0001-9803-9912
Datos de asesor	
Nombres y apellidos	Antonio Bravo Quiroz
Tipo de documento de identidad	DNI
Número de documento de identidad	10130035
URL de ORCID	https://orcid.org/0000-0001-9051-2808
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Roger Pedro Norabuena Figueroa
Tipo de documento	DNI
Número de documento de identidad	41493243
Miembro del jurado 1	
Nombres y apellidos	Rosa María Inga Santivañez
Tipo de documento	DNI
Número de documento de identidad	06416696
Miembro del jurado 2	
Nombres y apellidos	Jorge Luis Llanos Tizado
Tipo de documento	DNI
Número de documento de identidad	17848381
Datos de investigación	
Línea de investigación	A.3.2.4 Probabilidad y Procesos Estocásticos (Espaciales)

Grupo de investigación	No aplica
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Latitud: -12.056445 Longitud: -77.085994
Año o rango de años en que se realizó la investigación	Octubre 2020 - enero 2022
URL de disciplinas OCDE	Estadísticas, Probabilidad https://purl.org/pe-repo/ocde/ford#1.01.03

Dedicatoria

A mi madre y a mi abuelo por todo su amor y formación de vida.

*A mi hijo y a mi esposa por su amor y por su comprensión
por las horas que dejamos de compartir.*

Agradecimiento

A Dios por ser mi guía y por la salud que me brinda cada día.

Al Mg. Antonio Bravo Quiroz, asesor de la investigación, por su compromiso personal y sus consejos que me impulsaron a concluir la investigación.

A mi familia y amigos por sus consejos y palabras de aliento.

Índice general

Índice de tablas	vi
Índice de figuras	vii
RESUMEN	viii
ABSTRACT	ix
CAPÍTULO1: PLANTEAMIENTO DEL PROBLEMA	3
1.1 Descripción de la realidad problemática	3
1.2 Formulación del Problema	4
1.3 Justificación de la Investigación	5
1.3.1 Justificación Teórica	5
1.3.2 Justificación Práctica	6
1.4 Objetivos	6
1.4.1 Objetivo General	6
1.4.2 Objetivos Específicos	7
CAPÍTULO2: MARCO TEÓRICO	8
2.1 Antecedentes	8
2.2 Preliminares	10
2.3 Distribuciones Asimétricas	15
2.3.1 Método Univariado	15
2.3.2 Método Multivariado	20
2.3.3 Distribución Normal	21

2.3.4	Distribución t -Student	22
2.3.5	Distribución de Error Generalizada (GED)	23
2.4	Series de Tiempo Financieras	25
2.4.1	Series de Tiempo Financieras	25
2.4.2	Retornos	26
2.5	Inferencia Bayesiana	26
2.5.1	Teorema de Bayes	27
2.5.2	Distribución a Priori	28
2.5.3	Distribución a Posteriori	29
2.6	Modelo ARCH Univariado	33
2.7	Cadenas de Markov y MCMC	35
2.7.1	Introducción Histórica	35
2.7.2	Nociones Básicas	37
2.7.3	Irreducibilidad	40
2.7.4	Átomos y Conjuntos Pequeños	42
2.7.5	Aperiocidad	43
2.7.6	Transitoriedad y Recurrencia	44
2.7.7	Harris - Recurrencia	46
2.7.8	Medidas Invariantes	47
2.7.9	Ergodicidad y Convergencia	48
2.7.10	Teoremas Límite	49
2.7.10.1	Teorema de Ergodicidad	50
CAPÍTULO3: METODOLOGÍA DE LA INVESTIGACIÓN		54
3.1	Tipo y Diseño de la Investigación	54
3.1.1	Modelo GARCH Univariado	54
3.1.2	Modelos GARCH Multivariados: DCC - GARCH	56
3.2	Estudio de la Distribución a Priori	59
3.3	Algoritmo: Metropolis - Hastings (MH)	63
3.3.1	Algoritmo MH: Caminata Aleatoria	64

3.3.1.1	Núcleo de Transición	66
3.3.2	Propiedades de Convergencia	67
3.3.3	Tasas de Aceptación	69
3.3.4	MH 1-Bloque	73
3.3.5	MH np -Bloques	74
3.4	Población y Muestra	75
3.4.1	Población	75
3.4.2	Muestra	75
3.5	Técnicas e instrumentos de recolección de datos	76
3.6	Técnicas para el procesamiento de información	76
3.6.1	Análisis Estadístico	76
3.7	Selección del mejor modelo heterocedástico	80
3.7.1	Aplicación de los algoritmos Metropolis - Hastings	81
3.7.2	Criterios de selección del mejor modelo	82
CAPÍTULO4: RESULTADOS Y DISCUSIÓN		84
CAPÍTULO5: CONCLUSIONES		88

Índice de tablas

3.1	Transformaciones adoptadas para cada parámetro para que sean mapeados en la recta real.	72
3.2	Estadísticas de las series de retornos de los índices DAX, NIK-KEI225, CAC40 y BVL	78
4.1	Criterios para la selección del mejor modelo multivariado	84
4.2	Resumen de las simulaciones MCMC para el modelo heterocedástico con errores SST	85

Índice de figuras

3.1	Autocorrelaciones de los retornos (primera columna) y de los retornos al cuadrado (segunda columna) de los índices DAX, NIKKEI225, CAC40 y BVL	79
3.2	Histogramas de los retornos (primera columna) y qqplots de los retornos (segunda columna) de los índices DAX, NIKKEI225, CAC40 y BVL.	80
5.1	Gráficos de las series (primera columna) y de los retornos (segunda columna) de los índices DAX, NIKKEI225, CAC40 y BVL.	96
5.2	Diagrama de estrategias para elegir entre el algoritmo MH 1 - Block y MH np - Blocks.	97
5.3	Volatilidad aproximada de los retornos de los índices de las series, DAX, NIKKEI225, CAC40 y BVL para el modelo heterocedástico DCC - GARCH(1,1) con errores SST	98
5.4	Traza y densidades aproximados de todos los parámetros del modelo heterocedástico DCC - GARCH(1,1) con errores SST, para los retornos de las series DAX, NIKKEI225, CAC40 y BVL.	99
5.5	Autocorrelaciones de todos los parámetros del modelo heterocedástico DCC - GARCH(1,1) con errores SST, para los retornos de las series DAX, NIKKEI225, CAC40 y BVL.	104

RESUMEN

UN ENFOQUE BAYESIANO EN MODELOS HETEROCEDÁSTICOS DE SERIES DE TIEMPO Y SU APLICACIÓN EN LA VOLATILIDAD DE ACTIVOS FINANCIEROS

Edwin Antero Flores Montoya

Enero - 2022

El estudio de la variabilidad de un activo se ha convertido en las últimas décadas en un concepto muy importante en el área financiera. Se han propuesto varios modelos en la literatura para evaluar este fenómeno. En este trabajo, se estudia la modelación de la volatilidad de activos financieros, mediante un enfoque bayesiano. Para la modelación se utilizaron modelos de heterocedasticidad condicional autorregresiva generalizada y su principal generalización multivariada, los modelos con correlación condicional dinámica DCC - GARCH. Para los errores de estos modelos se consideran distribuciones de probabilidad posiblemente asimétricas y leptocúrticas, las cuales se parametrizan en función de la asimetría y el peso de las colas, por lo que se estiman también estos parámetros adicionales a los modelos. La estimación de los parámetros del modelo heterocedástico se realizó mediante la metodología MCMC, algoritmo Metropolis - Hastings caminata aleatoria, presente en el paquete `bayesDccGarch` software R, considerando datos diarios del 1 de abril del 2015 al 31 de enero del 2020 de las series estadísticas de las bolsas de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225) y París (CAC40), y además del índice General de la Bolsa de Valores de Lima (BVL). El enfoque bayesiano para la estimación de los parámetros del modelo heterocedástico que modela la volatilidad de estos activos financieros facilita la interpretación y brinda la posibilidad de insertar información a priori para los parámetros.

Palabras Claves: Series de tiempo, modelos heterocedásticos, distribuciones asimétricas, inferencia bayesiana, MCMC.

ABSTRACT

A BAYESIAN APPROACH IN HETEROCEDASTIC MODELS OF TIME SERIES AND ITS APPLICATION IN THE VOLATILITY OF FINANCIAL ASSETS

Edwin Antero Flores Montoya

January - 2022

The study of the variability of an asset has become a very important concept in the financial area in recent decades. Several models have been proposed in the literature to evaluate this phenomenon. In this work, the modeling of the volatility of financial assets is studied, using a Bayesian approach. For the modeling, generalized autoregressive conditional heteroscedasticity models were used and their main multivariate generalization, the DCC-GARCH dynamic conditional correlation models. Possibly asymmetric and leptokurtic probability distributions are considered for the errors of these models, which are parameterized according to the asymmetry and the weight of the tails, therefore these additional parameters to the models are also estimated. The estimation of the heteroscedastic model parameters was carried out using the MCMC methodology, the Metropolis - Hastings random walk algorithm, present in the `bayesDccGarch` software R package, considering daily data from April 1, 2015 to January 31, 2020 from the statistical series of the daily closing stock exchanges of the Frankfurt (DAX), Tokyo (NIKKEI225) and Paris (CAC40) stock indices, and in addition to the General Index of the Lima Stock Exchange (BVL). The Bayesian approach for estimating the parameters of the heteroscedastic model that models the volatility of these financial assets facilitates interpretation and provides the possibility of inserting a priori information for the parameters.

Keywords: Time series, heteroscedastic models, asymmetric distributions, Bayesian inference, MCMC.

Introducción

Modelar la volatilidad (varianza condicional) es de gran interés en economía. Cuando se grafican series de tiempo financieras se observan periodos más volátiles que otros y que su distribución generalmente es en grupos, lo que sugiere una dependencia en el tiempo, otra característica que se suele observar en estos gráficos es que, en general, los shocks negativos tienen más influencia en la volatilidad que los shocks positivos, caracterizando así un cierto grado de asimetría en la volatilidad según las series. Dadas estas características modelar la dinámica de los rendimientos financieros ha sido objeto de mucha atención y se ha investigado extensamente durante décadas desde la introducción del modelo de heterocedasticidad condicional autorregresiva (ARCH) Engle (1982) y su generalización, el modelo GARCH Bollerslev (1986). Sin embargo, cuando se trata de rendimientos multivariados, se debe tener en cuenta la dependencia mutua entre ellos, la dependencia entre activos tiende a aumentar en períodos de turbulencia del mercado, lo que a su vez podría tener implicaciones para la gestión de la cartera y el riesgo.

El análisis de las covarianzas de rendimiento de los activos es fundamental para la selección de la cartera, la gestión de activos y la evaluación de riesgos, es decir las variaciones en varias series financieras pueden correlacionarse, de modo que la volatilidad de una serie pueda verse influenciada por las volatilidades de otras series. Para considerar estas correlaciones al estimar modelos GARCH, han surgido en la literatura varias extensiones multivariadas, entre los más conocidos se encuentran los modelos CCC - GARCH (Correlación Condicional Constante GARCH) Bollerslev (1990) y DCC - GARCH (Correlación Condicional Dinámica GARCH) propuestos simultáneamente en Engle (2002) y Tse and Tsui (2002).

Generalmente, el uso de la distribución de probabilidad normal estándar para los errores de los modelos heterocedásticos GARCH no es suficiente para adecuar las

características de colas pesadas y asimetría de rendimientos financieros (diferencia entre el logaritmo del precio de un activo en el instante t y el logaritmo del precio de un activo en el instante $t - 1$), por ello, tanto para los errores de los modelos heterocedásticos GARCH como para los errores de los modelos heterocedásticos DCC - GARCH, se estudia el uso de distribuciones de probabilidad con colas más pesadas que la distribución normal estándar y también se considera una forma de insertar asimetría a estas distribuciones.

La estimación se realizó bajo el enfoque bayesiano, ya que estas distribuciones se parametrizan según parámetros de peso en las colas y asimetría. De modo que con el enfoque bayesiano es posible analizar estas características mediante la distribución a posteriori de cada parámetro.

En la literatura los estudios utilizando el enfoque bayesiano son muy escasos, incluso para modelos univariados. Esto es debido a la complejidad de estos modelos y el costo computacional de usar métodos computacionales basados en la simulación de Monte Carlo a través de Cadenas de Markov (MCMC). La metodología MCMC consiste en construir una cadena de Markov con distribución límite igual a π y obtener muestras de la cadena con las que se aproximan esperanzas bajo π , si la cadena tiene ciertas propiedades y las esperanzas existen, entonces los promedios muestrales convergen a las esperanzas. La idea se debe a Metropolis et al. (1953), y fue extendida por Hastings (1970). En esta tesis se utiliza esta metodología MCMC, paquete `bayesDccGarch`, implementado en el software R.

En el capítulo 1, se presenta el problema y los objetivos de estudio, en el capítulo 2 se presenta un método para insertar asimetría en distribuciones simétricas univariadas y multivariadas, además se desarrolla parte de la teoría de cadenas de Markov, en el capítulo 3 se presentan los modelos GARCH y los algoritmos Metropolis - Hastings y la aplicación del modelo heterocedástico DCC - GARCH con datos reales y finalmente en el capítulo 4 se presentan resultados, conclusiones y recomendaciones.

CAPÍTULO 1

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la realidad problemática

Uno de los problemas más interesantes y de mayor dificultad en el estudio de series de tiempo que surge en econometría es sobre la modelación de la volatilidad (varianza condicional) producida por periodos de crisis e incertidumbre, inestabilidad o variabilidad de los precios.

Generalmente al graficar las series de tiempo de activos financieros se observa presencia de volatilidad, por ello muchos autores recomiendan usar modelos heterocedásticos condicionales (modelos que consideran que la varianza condicional de una serie de tiempo no es constante), es decir modelos de heterocedasticidad tales como el ARCH propuesto por Engle (1982), y su generalización GARCH propuesta por Bollerslev (1986), así como los modelos de volatilidad estocástica de Taylor (1982).

Se debe tener en cuenta que las variaciones en varias series financieras pueden estar correlacionadas, es decir los valores de una serie varían sistemáticamente con respecto a los valores homónimos de otras series, de modo que la volatilidad de una serie puede estar influenciada por la volatilidad de otras series.

Teniendo en cuenta estas correlaciones en el estudio de los modelos GARCH, dieron origen a varias extensiones multivariadas surgiendo, entre muchos, los modelos CCC - GARCH (correlación condicional constante GARCH) Bollerslev (1990) y DCC - GARCH (correlación condicional dinámica GARCH) propuesta simultáneamente por Engle (2002) y Tse and Tsui (2002), que son los modelos más conocidos.

Al revisar la literatura existente sobre modelación de series de tiempo de tipo financieras, se encontró muchas metodologías clásicas, pero desde el enfoque bayesiano son muy escasos, motivo por el cual se propone este trabajo de investigación dado que es un tema novedoso y facilita la interpretación de los parámetros del modelo y posibilita insertar información a priori, y además es viable pues las simulaciones se hicieron utilizando la metodología Monte Carlo a través de las cadenas de Markov (MCMC), específicamente los algoritmos Metropolis - Hastings (MH) caminata aleatoria en bloques y en un solo bloque, todas ellas desarrolladas utilizando el software R Team et al. (2013), y teniendo como base datos reales de activos financieros: datos diarios de las series estadísticas de las bolsas de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225) y París (CAC40) y del índice General de la Bolsa de Valores de Lima (BVL).

1.2. Formulación del Problema

Los modelos de series de tiempo tradicionales que suponen varianza homocedástica, no son adecuados para modelar series de tiempo financieras. Engle (1982) desarrolla una nueva clase de modelos llamados ARCH, en los cuales la varianza condicional a la información pasada no es constante y depende del cuadrado de las inovaciones pasadas, Bollerslev (1986) generaliza los modelos ARCH al proponer los modelos GARCH en los cuales la varianza condicional depende no solo de los cuadrados de las perturbaciones pasadas sino además depende de la varianza condicional de los periodos anteriores.

Es decir el $GARCH(p, q)$ define la varianza condicional en el tiempo t de y_t (rendimiento de un activo o un índice financiero en el tiempo t), denotado por h_t , como una función lineal de los cuadrados de los p rendimientos pasados y las q variaciones condicionales pasadas.

Considerando una serie de rendimientos financieros $y = \{y_t, t = 1, \dots, T\}$ el

modelo tiene la siguiente expresión:

$$y_t = \epsilon_t \sqrt{h_t}, \quad \epsilon_t \sim D(0, 1) \quad (1.1)$$

$$h_t = \omega + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} \quad (1.2)$$

siendo h_t la variación condicional (no observable) de y_t dada la información previa $I_t = \{y_{t-1}, y_{t-2}, \dots\}$, los errores ϵ_t son independientes e idénticamente distribuidos. $D(0, 1)$ denota una distribución con media 0 y una varianza 1. Las restricciones de estacionaridad y positividad son $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \dots, p$, $\beta_j \geq 0$, $j = 1, \dots, q$.

Una extensión natural de estos modelos son los modelos GARCH multivariantes, a los cuales le prestamos atención en estudiar las relaciones entre las volatilidades y covolatilidades de varios mercados; la construcción de medidas de riesgo de una cartera de activos financieros será influenciada por la estructura de dependencia entre las series que componen la cartera. Ante esta disyuntiva surgen dos extensiones multivariadas como son el CCC - GARCH Bollerslev (1990) y el DCC - GARCH Engle (2002); Tse and Tsui (2002) descritos en párrafos anteriores.

Según lo expuesto, permite formular el siguiente problema asociado con el trabajo de investigación de tesis:

¿Cómo el enfoque bayesiano en los modelos heterocedásticos de series de tiempo permite modelar la volatilidad de activos financieros?

1.3. Justificación de la Investigación

1.3.1. Justificación Teórica

El estudio de la volatilidad es de mucha importancia en econometría, en especial en series de tiempo financieras. Por ello esta tesis se realizó con el propósito

de aportar conocimiento a la literatura existente de modelos de heterocedasticidad condicional autorregresiva generalizada con correlación condicional dinámica DCC - GARCH, cuando los errores de estos modelos tienen distribuciones de probabilidad asimétricas y además son leptocúrticas. Se hizo uso de los métodos computacionales basados en simulación MCMC, cuyos resultados podrán sistematizarse en una propuesta, para ser incorporado como modelo de solución para series de tiempo financieras ya que se logra modelar la volatilidad de los activos financieros dados.

1.3.2. Justificación Práctica

Hay muchos trabajos que utilizan metodología clásica para modelar series de tiempo financieras pero desde un enfoque bayesiano son muy escasos, motivo por el cual en este trabajo se utilizó metodología bayesiana para modelar la volatilidad de activos financieros mediante modelos heterocedásticos, teniendo como base datos reales: datos diarios de las series estadísticas de las bolsas de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225) y París (CAC40) y del índice General de la Bolsa de Valores de Lima (BVL), por lo que este trabajo es un aporte para investigaciones futuras que involucren esta metodología.

1.4. Objetivos

1.4.1. Objetivo General

OG: Utilizar el enfoque bayesiano en los modelos heterocedásticos de series de tiempo para modelar la volatilidad de activos financieros.

1.4.2. Objetivos Específicos

- OE1: Estudiar el método de Fernández and Steel (1998) en los modelos heterocedásticos de series de tiempo para modelar la volatilidad de activos financieros.
- OE2: Estudiar la metodología MCMC basada en el método de Monte Carlo a través de las cadenas de Markov (algoritmos Metropolis - Hastings) en los modelos heterocedásticos de series de tiempo, para la modelación de la volatilidad de activos financieros.
- OE3: Obtención de los estimadores de los parámetros del modelo heterocedástico de series de tiempo y las correspondientes estimaciones en la modelación de la volatilidad de activos financieros mediante el enfoque bayesiano (metodología MCMC).

CAPÍTULO 2

MARCO TEÓRICO

2.1. Antecedentes

Los modelos ARCH introducidos por Engle (1982), tienen como base fundamental la volatilidad (varianza condicional), es decir, la varianza en un instante t no es constante y depende de los retornos pasados.

Bollerslev (1986), propuso una clase de modelos que considera que la volatilidad depende, no solo de los retornos pasados, sino que también depende de los valores pasados de la volatilidad, estos son los modelos GARCH, estos modelos son considerados más parcimoniosos pues describen la volatilidad de la serie con menos parámetros.

En la literatura existente hay muchas extensiones multivariadas de los modelos GARCH entre los que se destacan los modelos multivariados CCC - GARCH (correlación condicional constante GARCH) y el modelo DCC - GARCH (correlación condicional dinámica GARCH) propuestos por los mismos creadores de los modelos ARCH y GARCH Bollerslev (1990); Engle (2002); Tse and Tsui (2002). En Bauwens et al. (2006) se establecen diversas formas de modelar la matriz de covarianzas condicionales.

Seguidamente surgen algunos enfoques bayesianos para la estimación de los parámetros del modelos GARCH. David (2006) propone un algoritmo para la estimación bayesiana del modelo GARCH(1,1) con errores normales, este algoritmo obtiene los valores de las muestras de la distribución a posteriori de los parámetros.

En general se debe tener en cuenta que en los modelos GARCH las condiciones completas de la a posteriori no coinciden con funciones de densidad de pro-

babilidad conocidas en la literatura debido a la evidencia en la literatura que muchas series temporales financieras tienden a tener la curtosis observada mayor que aquella considerada en los modelos GARCH con errores normales, Ardia et al. (2008) proponen la estimación bayesiana del modelo GARCH con errores *t*-Student a través de un algoritmo MCMC. Este algoritmo MCMC fue propuesto por Metropolis et al. (1953) y luego generalizado por Hastings (1970), quedando así el algoritmo Metropolis - Hastings (MH).

Ehlers (2012), compara modelos GARCH con distribuciones asimétricas de los errores. Cabe resaltar que Ehlers utilizó la metodología propuesta por Fernández and Steel (1998) para insertar asimetría mediante los factores de escala inversa en los valores positivos y negativos de la variable, obteniendo así, una distribución asimétrica ajustada solo a un parámetro, de tal manera que éste determina el grado de asimetría para el caso univariado, Bauwens and Laurent (2005) generalizó esta idea para el caso multivariado.

Fioruci (2012); Fioruci et al. (2014a,b) estudian los modelos GARCH y su principal generalización multivariada, los modelos heterocedásticos DCC - GARCH, para los errores de estos modelos consideraron distribuciones de probabilidad asimétricas y leptocúrticas que se parametrizan de acuerdo con la asimetría y el peso en las colas, por lo que les fué necesario estimar estos parámetros adicionales para sus modelos.

Aquino Gutierrez et al. (2017), consideran dos características principales a ser modeladas: La asimetría y las colas pesadas presentes en la distribución incondicional de las series de retornos, la estimación de los parámetros de los modelos propuestos fueron hechos mediante un enfoque bayesiano con la metodología MCMC específicamente los algoritmos Metropolis - Hastings (MH).

Morettin (2017), recomienda el uso de modelos de bajo orden (por ejemplo: (1,1), (1,2), (2,1) y (2,2)) y el modelo se elige según criterios como el: Criterio de información Akaike AIC Akaike (1974) y el criterio de información Bayesiano

BIC Schwarz et al. (1978) entre otros. El autor también afirma que en la mayoría de las series financieras, un modelo GARCH(1,1) es suficiente para describir bien la volatilidad.

En base a todo lo expuesto y teniendo en cuenta el criterio de Desvío de Información (DIC) propuesto Spiegelhalter et al. (2002). La motivación del presente trabajo es modelar el índice del mercado accionario empleando la metodología propuesta desde una perspectiva bayesiana. Para la elección del modelo heterocedástico se consideró las esperanzas a posterioris de los criterios AIC, BIC y DIC.

2.2. Preliminares

Sea Ω un conjunto no vacío y $\mathcal{P}(\Omega) \equiv \{A : A \subset \Omega\}$ el conjunto potencia de Ω , es decir, la clase de todos los subconjuntos de Ω .

Definición 2.1 Una colección de conjuntos $\mathcal{F} \subset \mathcal{P}(\Omega)$ se llama **álgebra** si

- (a) $\Omega \in \mathcal{F}$
- (b) $A \in \mathcal{F}$ entonces $A^c \in \mathcal{F}$
- (c) $A, B \in \mathcal{F}$ entonces $A \cup B \in \mathcal{F}$

Por lo tanto, un álgebra es una clase de conjuntos que contienen a Ω , que están cerrados bajo complementación y uniones por pares (y por lo tanto finitas). Además se puede definir de manera equivalente un álgebra al requerir que las propiedades (a), (b) se mantengan y que la propiedad $A, B \in \mathcal{F}$ entonces $A \cap B \in \mathcal{F}$, se cumpla.

Definición 2.2 Una clase $\mathcal{F} \subset \mathcal{P}(\Omega)$ se llama σ -álgebra si se cumple:

$$A_n \in \mathcal{F} \text{ para } n \geq 1 \Rightarrow \bigcup_{n \geq 1} A_n \in \mathcal{F}.$$

Una clase particularmente útil de σ -álgebras son las generadas por conjuntos abiertos de un espacio topológico. Estos se denominan σ -álgebras de Borel. Más adelante se definen en espacios medibles.

Definición 2.3 Sea Ω un conjunto no vacío y \mathcal{F} un algebra en Ω entonces, una función fija μ en \mathcal{F} se llama **medida** si:

(a) $\mu(A) \in [0, \infty] \forall A \in \mathcal{F}$

(b) $\mu(\emptyset) = 0$

(c) Para cualquier colección disjunta de conjuntos $A_1, A_2, A_3 \dots, \in \mathcal{F}$ con

$$\bigcup_{n \geq 1} A_n \in \mathcal{F},$$

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

Definición 2.4 Una medida μ se llama finita o infinita si $\mu(\Omega) < \infty$ o $\mu(\Omega) = \infty$, respectivamente. Una medida finita con $\mu(\Omega) = 1$ se llama **medida de probabilidad**. Una medida μ en una σ -álgebra \mathcal{F} se llama **σ -finita** si existe una colección contable de conjuntos $A_1, A_2, A_3 \dots, \in \mathcal{F}$, no necesariamente disjuntos, tal que:

(a) $\bigcup_{n \geq 1} A_n = \Omega$

(b) $\mu(A_n) < \infty, \forall n \geq 1$

Definición 2.5 Sea Ω un conjunto no vacío y sea \mathcal{F} un σ -álgebra en Ω . Entonces, el par (Ω, \mathcal{F}) se llama **espacio medible**. Si μ es una medida en (Ω, \mathcal{F}) , entonces el triplete $(\Omega, \mathcal{F}, \mu)$ se llama **espacio de medida**. Si además, μ es una medida de probabilidad, entonces $(\Omega, \mathcal{F}, \mu)$ se llama **espacio de probabilidad**.

Definición 2.6 Sea (Ω, \mathcal{F}) un espacio medible. Entonces una función $f : \Omega \rightarrow \mathbb{R}$ se llama $\langle \mathcal{F}, \mathcal{B}(\mathbb{R}) \rangle$ -**medible (o \mathcal{F} -medible)** si para cada a en \mathbb{R}

$$f^{-1}((-\infty, a]) \equiv \{\omega : f(\omega) \leq a\} \in \mathcal{F}.$$

Definición 2.7 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Entonces una función $X : \Omega \rightarrow \mathbb{R}$ se llama **variable aleatoria**, si el evento

$$X^{-1}((-\infty, a]) \equiv \{\omega : f(\omega) \leq a\} \in \mathcal{F}$$

para cada a en \mathbb{R} , es decir, una variable aleatoria es una función \mathcal{F} -medible de valor real en un espacio de probabilidad (Ω, \mathcal{A}, P) .

En James (1996) puede ver más detalles de variable aleatoria así como algunos ejemplos.

Definición 2.8 La σ -álgebra de Borel en un espacio medible \mathbb{S} , se define como la σ -álgebra generada por la colección de conjuntos abiertos en \mathbb{S} . Es decir, un boreliano es un conjunto que puede ser obtenido de una cantidad numerable de intervalos aplicándoles las operaciones \cup , \cap y c una cantidad numerable de veces.

Ejemplo 2.1 Sea $\mathcal{B}(\mathbb{R}^k)$ el σ -álgebra de Borel en \mathbb{R}^k , $1 \leq k < \infty$.

Luego, $\mathcal{B}(\mathbb{R}^k) \equiv \sigma \{A : A \text{ es un subconjunto abierto de } \mathbb{R}^k\}$. Para más detalles sobre teoría de la medida ver Athreya and Lahiri (2006).

Definición 2.9 Sea T un conjunto arbitrario, un **proceso estocástico**: es una familia $y = \{y_t, t \in T\}$, tal que para todo $t \in T$, y_t es una variable aleatoria.

Definición 2.10 Dado un proceso estocástico

$$\{\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, y_3, \dots, y_t, y_{t+1}, y_{t+2}, \dots\}$$

una **serie de tiempo** es un subconjunto: $\{y_1, y_2, y_3, \dots, y_t\}$

Definición 2.11 Un proceso estocástico $y = \{y_t, t \in T\}$ se dice **estrictamente estacionario** si, luego de hacer transformaciones en el tiempo, la distribución conjunta

$$F(y_1, y_2, \dots, y_n; t_1 + h, t_2 + h, \dots, t_n + h) = F(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n)$$

para cualquier t_1, t_2, \dots, t_n, h de T . Esto significa, en particular, que todas las distribuciones unidimensionales son invariantes sobre traslaciones en el tiempo,

luego la $\mu(t)$ media y la varianza $V(t)$ son constantes, es decir:

$$\mu(t) = \mu, V(t) = \sigma^2$$

para todo $t \in T$, sin pérdida de generalidad se considera $\mu(t) = 0$; caso contrario se considera el proceso $\{y_t - \mu\}$

Definición 2.12 Un proceso estocástico $y = \{y_t, t \in T\}$ se dice **debilmente estacionario** si, y solo si:

- (i) $E(y_t) = \mu(t) = \mu$, constante $\forall t \in T$
- (ii) $E(y_t^2) < \infty$, $\forall t \in T$
- (iii) $\gamma(T_1; T_2) = cov\{y_{t_1}, y_{t_2}\}$ es una función de $|t_1 - t_2|$

En este trabajo a este proceso llamaremos simplemente **proceso estacionario**.

Definición 2.13 Un proceso $y = \{y_t, t \in T\}$ en el cual (ii) se satisface es llamado **proceso de segundo orden**.

Definición 2.14 Sea $y = \{y_t, t \in \mathbb{Z}\}$ un proceso estacionario real discreto de media cero, la **función de autocovarianza** (*facv*) de y es dado por $\gamma_\tau = E(y_t, y_{t+\tau})$

Definición 2.15 Una matriz hermitiana A es **definida positiva** si, todos los menores principales son positivos (criterio de Sylvester).

Definición 2.16 Una distribución de probabilidad unimodal $p(x)$ definida en \mathbb{R}^m , con $E(X) = 0$ y $V(X) = I_m$ es **simétrica** si, y sólo si, para cualquier x , $p(x) = p(Qx)$, donde Q es una matriz diagonal cuyos elementos de la diagonal son iguales a 1 o -1.

Definición 2.17 Un **modelo estadístico bayesiano** se compone de un modelo estadístico paramétrico, $f(x|\theta)$, y una distribución a priori $\pi(\theta)$ de los parámetros.

Proposición 2.1 La *facv*, γ_τ satisface las siguientes propiedades

- (i) $\gamma_0 > 0$
- (ii) $\gamma_{-\tau} = \gamma_\tau$
- (iii) $|\gamma_\tau| \leq \gamma_0$

(iv) γ_τ es definido no negativo, es decir:

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} \geq 0,$$

para cualquier número real a_1, \dots, a_n y τ_1, \dots, τ_n de \mathbb{Z}

La **función de autocorrelación** de un proceso es dado por:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}; \quad \tau \in \mathbb{Z}$$

y tiene las mismas propiedades que γ_τ , excepto que ahora $\gamma_0 = 1$

Teorema 2.1 (Teorema de cambio de variable) Sea X una variable aleatoria con valores en (Ω, \mathcal{A}, P) y sea la función Borel medible $h : \mathbb{R} \rightarrow \mathbb{R}$. Sea $Y = h(X)$ entonces

(i)

$$\int_{\Omega} |Y| dP = \int_{\mathbb{R}} |h(x)| P_X(dx) = \int_{\mathbb{R}} |y| P_Y(dy)$$

(ii) Si

$$\int_{\Omega} |Y| dP < \infty$$

entonces

$$\int_{\Omega} Y dP = \int_{\mathbb{R}} h(x) P_X(dx) = \int_{\mathbb{R}} y P_Y(dy)$$

Teorema 2.2 (Factorización de Cholesky) Si A es una matriz $n \times n$ simétrica definida positiva, entonces existe al menos una matriz $n \times n$ triangular inferior $B = (b_{ij})$ tal que $A = BB^T$. Además, se puede imponer que $b_{ii} > 0$ para todo $i = 1, \dots, n$, y en tal caso la factorización anterior es única. El recíproco del Teorema también es válido.

Proposición 2.2. Si B es una matriz $n \times n$ triangular inferior regular, y definimos $A = BB^T$, entonces A es simétrica y definida positiva.

2.3. Distribuciones Asimétricas

En esta sección se presenta el método de Fernández and Steel (1998) el cual se utilizó para insertar asimetría en cualquier función de distribución de probabilidad continua, simétrica, unimodal y definida en los reales. En seguida se presenta el método de Bauwens and Laurent (2005), el cual generaliza el método de Fernández and Steel (1998) para distribuciones multivariadas.

2.3.1. Método Univariado

En la literatura, hay varias propuestas (métodos) para introducir asimetría en distribuciones simétricas, entre las que podemos citar, Azzalini (1985), Fernández and Steel (1998) y Branco and Dey (2001). Pero debido a la simplicidad y generalidad se opta por la propuesta de Fernández and Steel (1998). En éste método los momentos son más fáciles de calcular y no es necesario obtener la función de distribución acumulada.

En el contexto Bayesiano, el método facilita la especificación de las distribuciones a priori separando los efectos de los parámetros de asimetría y de cola. Consideremos $p(x)$ una función de densidad unimodal, definida en la recta y simétrica alrededor de cero. Es decir $p(x) = p(-x)$, $\forall x \in \mathbb{R}$. El método propuesto por Fernández and Steel (1998) es conocido como “método de factores de escala inversa” y consiste en obtener una función de densidad $s(x|\gamma)$ a partir de $p(x)$, la cual es asimétrica y tiene el grado de asimetría indexado por el parámetro $\gamma > 0$.

La función de densidad de probabilidad $s(x|\gamma)$ se define como sigue:

$$s(x|\gamma) = \frac{2}{\gamma + \gamma^{-1}} \{p(x\gamma)I_{<-\infty;0>}(x) + p(\frac{x}{\gamma})I_{[0;+\infty>}(x)\}$$

Algunas de las principales características de $s(x|\gamma)$ son:

1. Si $\gamma = 1$, se cumple $s(x|\gamma = 1) = p(x)$.
2. $s(x|\gamma)$ mantiene la misma moda de $p(x)$. Como $p(x)$ es unimodal y simétrica entorno a cero, la moda siempre es en el cero.
3. La probabilidad a la derecha y a la izquierda es independiente de $p(x)$:

Probabilidad a la derecha:

$$P(X \geq 0) = \frac{2}{\gamma + \gamma^{-1}} \int_0^{\infty} p(x/\gamma) dx$$

haciendo un cambio de variable

$$y = \frac{x}{\gamma}$$

entonces

$$dy = \frac{dx}{\gamma}$$

además si $x \rightarrow 0$ entonces $y \rightarrow 0$ y si $x \rightarrow \infty$ entonces $y \rightarrow \infty$; luego

$$\begin{aligned} P(X \geq 0) &= \frac{2}{\gamma + \gamma^{-1}} \int_0^{\infty} p(x/\gamma) dx \\ &= \frac{2\gamma^2}{1 + \gamma^2} \int_0^{\infty} p(y) dy \end{aligned}$$

y como

$$\int_0^{\infty} p(y) dy = 0,5$$

tenemos

$$P(X \geq 0) = \frac{\gamma^2}{1 + \gamma^2}$$

Probabilidad a la izquierda:

$$P(X \leq 0) = 1 - P(X \geq 0) = \frac{1}{1 + \gamma^2}$$

De los dos resultados se tiene:

$$\frac{P(X \geq 0)}{P(X \leq 0)} = \gamma^2$$

4. La asimetría a la derecha (izquierda) corresponde a $\gamma > 1$ ($\gamma < 1$) respectivamente.
5. La existencia de los momentos de $s(x|\gamma)$ depende únicamente de los momentos absolutos de $p(x)$.

El r -ésimo momento es dado por:

$$E(X^r|\gamma) = \frac{\gamma^{r+1} + \frac{(-1)^r}{\gamma^{r+1}}}{\gamma + \gamma^{-1}} M_r \quad (2.1)$$

Por ser $p(x)$ simétrico respecto a cero se tiene:

$$M_r = 2 \int_0^{\infty} x^r p(x) dx$$

el r -ésimo momento absoluto de $p(x)$.

6. La media y la varianza son dados por:

$$\mu = E(X^1|\gamma) = \frac{\gamma^{1+1} + \frac{(-1)^1}{\gamma^{1+1}}}{\gamma + \gamma^{-1}} M_1 = (\gamma - \gamma^{-1}) M_1 \quad (2.2)$$

$$\sigma^2 = E(X^2|\gamma) - E^2(X^1|\gamma) = \frac{\gamma^{2+1} + \frac{(-1)^2}{\gamma^{2+1}}}{\gamma + \gamma^{-1}} M_2 - [(\gamma - \gamma^{-1}) M_1]^2$$

$$\sigma^2 = (M_2 - M_1^2)(\gamma^2 + \gamma^{-2}) + 2M_1^2 - M_2 \quad (2.3)$$

7. Para hallar la versión estandarizada de la función de densidad $s(x|\gamma)$ se hace un cambio de variable

$$z = \frac{x - \mu}{\sigma}$$

entonces

$$\frac{dx}{dz} = \sigma$$

además si $x < 0 \Rightarrow z < -\mu/\sigma$ y si $x \geq 0, \Rightarrow z \geq -\mu/\sigma$.

Luego la versión estandarizada de la función de densidad $s(x|\gamma)$ es la distribución de probabilidad de la variable aleatoria z dada por:

$$\begin{aligned} p(z|\gamma) &= s(z\sigma + \mu|\gamma) \frac{dx}{dz} \\ &= s(z\sigma + \mu|\gamma) \sigma \\ &= \frac{2}{\gamma + \gamma^{-1}} p(z^*) \end{aligned} \quad (2.4)$$

donde

$$z^* = \begin{cases} (z\sigma + \mu)\gamma, & \text{si } z < -\mu/\sigma \\ (z\sigma + \mu)/\gamma, & \text{si } z \geq -\mu/\sigma \end{cases} \quad (2.5)$$

y $p(z^*)$ es la función de densidad simétrica calculada en z^* .

Como ejemplo, se aplica el método presentado en (2.4) en la distribución de probabilidad Normal estándar, de donde se obtiene:

$$p(x|\gamma) = \frac{\sqrt{\frac{2}{\pi}}}{\gamma + \gamma^{-1}} \exp^{-\frac{(x^*)^2}{2}} \quad (2.6)$$

siendo, x^* dado como en (2.5).

Se estudia el primer y segundo momento de la función de distribución simétrica y estandar $p(x)$

El primer momento:

$$\begin{aligned} M_1 &= 2 \int_0^{\infty} xp(x)dx \\ &= 2 \int_0^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \end{aligned}$$

Haciendo el cambio de variable $y = \frac{x^2}{2}$, $dy = xdx$, y como los límites de integración no cambian, tenemos:

$$M_1 = \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-y} dy$$

y como

$$\int_0^{\infty} e^{-y} dy = 1$$

por ser la distribución exponencial con parámetro 1, se obtiene $M_1 = \sqrt{\frac{2}{\pi}}$.

El segundo momento:

$$\begin{aligned} M_2 &= 2 \int_0^{\infty} x^2 p(x) dx \\ &= 2 \int_0^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \end{aligned}$$

Haciendo el cambio de variable $y = \frac{x^2}{2}$, $dy = x dx$, y como los límites de integración no cambian, tenemos:

$$\begin{aligned} M_2 &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} y^{3/2-1} e^{-y} dy \\ &= \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) \\ &= \frac{2}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \\ &= 1 \end{aligned}$$

De esta forma, de las expresiones (2.2) y (2.3) se obtiene la media y la varianza de la versión asimétrica de la distribución normal estándar:

$$\begin{aligned} \mu &= \sqrt{\frac{2}{\pi}} (\gamma - \gamma^{-1}) \\ \sigma^2 &= (\gamma^2 + \gamma^{-2} - 1) - \mu^2 \end{aligned}$$

La función de densidad (2.6) será denotado como $SSN(0, 1, \gamma)$ (Normal estándar asimétrica).

2.3.2. Método Multivariado

Sea $p(x)$ como en la **Definición 2.16**. El método para insertar asimetría de Bauwens and Laurent (2005) generaliza para el caso multivariado el método de Fernández and Steel (1998) presentado en la sección (2.3.1). Este método está dado por:

$$s(x|\gamma) = 2^m \left(\prod_{i=1}^m \frac{\gamma_i}{1 + \gamma_i^2} \right) p(x^*) \quad (2.7)$$

siendo,

$$x^* = (x_1^*; x_2^*; \dots; x_m^*)^t = \begin{cases} \frac{x_i}{\gamma_i}, & \text{si } x_i \geq 0 \\ x_i \gamma_i, & \text{si } x_i < 0 \end{cases}$$

Los parámetros de asimetría son dados por $\gamma = (\gamma_1, \dots, \gamma_m)^t$, con $\gamma_i > 0$. Si $\gamma_i = 1$ entonces la marginal correspondiente es el caso simétrico.

Los momentos de $p(x|\gamma)$ se obtienen en función de los momentos absolutos de las distribuciones marginales de $p(x)$ como en Fernández and Steel (1998). Esto es,

$$E(X^r|\gamma) = \frac{\gamma^{r+1} + \frac{(-1)^r}{\gamma^{r+1}}}{\gamma + \gamma^{-1}} M_r \quad (2.8)$$

Siendo

$$M_r = 2 \int_0^\infty x_i^r p(x_i) dx_i$$

para cualquier r en N .

En general, incluso $p(x)$ siendo estandarizada la distribución resultante $s(x|\gamma)$ puede no ser estandarizada. Pero desde que el primer momento absoluto de la distribución marginal $p(x_i)$ sea conocida se puede utilizar la expresión (2.8) para obtener el vector de medias $\mu = (\mu_1, \dots, \mu_m)^t$ y el vector de varianzas $\sigma_i^2 = (\sigma_1^2, \dots, \sigma_m^2)^t$ de $s(x|\gamma)$.

Por ser $p(x)$ estandar, $M_2 = 1$, luego para el caso multivariado de la expresión (2.8) tenemos las medias y las varianzas.

$$\mu_i = E(X^1|\gamma_i) = (\gamma_i - \gamma_i^{-1}) M_1 \quad (2.9)$$

$$\sigma_i^2 = (\gamma_i^2 + \gamma_i^{-2} - 1) - \mu_i^2 \quad (2.10)$$

para $i = 1, \dots, m$. Así, la versión estandarizada de la función de densidad (2.7) es la distribución de probabilidad del vector aleatorio $z = (z_1, \dots, z_m)^t$ en el cual $z_i = (x_i - \mu_i)/\sigma_i$, la cual es dada por:

$$p(z|\gamma) = 2^m \left\{ \prod_{i=1}^m \frac{\gamma_i}{1 + \gamma_i^2} \sigma_i \right\} p(z^*) \quad (2.11)$$

siendo,

$$z_i^* = \begin{cases} (z_i \sigma_i + \mu_i) \gamma_i, & \text{si } z_i < -\mu_i/\sigma_i \\ (z_i \sigma_i + \mu_i)/\gamma_i, & \text{si } z_i \geq -\mu_i/\sigma_i \end{cases} \quad (2.12)$$

A continuación se aplica el método para las distribuciones de probabilidad Normal, t -Student y GED multivariadas. Para obtener el caso univariado de las distribuciones siguientes, hacer $m = 1$.

2.3.3. Distribución Normal

La distribución Normal m -Variada denotada por $N(0, I_m)$ se define como el producto de m distribuciones normales estándar $N(0, 1)$, por ende posee distribuciones marginales $N(0, 1)$. La distribución $N(0, I_m)$ es dada por:

$$p(x) = \frac{1}{(2\pi)^{\frac{m}{2}}} \exp\left\{-\frac{1}{2} \sum_{i=1}^m x_i^2\right\} \quad (2.13)$$

Calculando el momento absoluto de $N(0, 1)$, se tiene:

$$M_1 = \frac{2}{(2\pi)^{\frac{1}{2}}} \int_0^{\infty} x_i \exp\left\{-\frac{x_i^2}{2}\right\} dx_i = \frac{2}{(2\pi)^{\frac{1}{2}}} \quad (2.14)$$

lo anterior se obtiene haciendo $\frac{x_i^2}{2} = y_i$.

Luego sustituyendo (2.14) en (2.9) se obtienen las medias de las marginales de la versión asimétrica de la función de densidad (2.13):

$$\mu_i = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} (\gamma_i - \gamma_i^{-1})$$

Para $i = 1, \dots, m$. Las varianzas marginales se obtienen directamente de la expresión (2.10). En la expresión (2.11) se obtiene la distribución de probabilidad Normal asimétrica y estandarizada $SSN(0, I_m, \gamma)$,

$$p(z|\gamma) = \left(\frac{2}{\pi}\right)^{\frac{m}{2}} \left(\prod_{i=1}^m \frac{\gamma_i \sigma_i}{1 + \gamma_i^2}\right) \exp\left\{-\frac{1}{2} \sum_{i=1}^m z_i^{*2}\right\},$$

siendo, z_i^* dado por las expresiones (2.12)

2.3.4. Distribución t -Student

Una generalización multivariada de la distribución estandar t -Student se define como:

$$p(x|v) = \frac{\Gamma((v+m)/2)}{\Gamma(v/2)[\pi(v-2)]^{m/2}} \left(1 + \frac{x^t x}{v-2}\right)^{-\frac{(v+m)}{2}} \quad (2.15)$$

Esta distribución de probabilidad será denotada por $ST(0, I_m, v)$, con distribuciones marginales $ST(0, 1, v)$. La distribución $ST(0, I_m, v)$ satisface la **Definición 2.16** y por lo tanto, se puede aplicar el método de Bauwens and Laurent (2005).

El primer momento absoluto de la distribución $ST(0, 1, v)$:

$$M_1 = \frac{2\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})(\pi(v-2))^{\frac{1}{2}}} \int_0^\infty x_i \left(1 + \frac{x_i^2}{v-2}\right)^{-\frac{v+1}{2}} dx_i$$

haciendo:

$$1 + \frac{x_i^2}{v-2} = z_i$$

derivando

$$\frac{2x_i dx_i}{v-2} = dz_i$$

si $x_i \rightarrow 0$ entonces $z_i \rightarrow 1$ y si $x_i \rightarrow \infty$ entonces $z_i \rightarrow \infty$; luego

$$\begin{aligned} \int_0^\infty x_i \left(1 + \frac{x_i^2}{v-2}\right)^{-\frac{v+1}{2}} dx_i &= \left(\frac{v-2}{2}\right) \times \int_1^\infty z_i^{-\frac{v+1}{2}} dz_i \\ &= \left(\frac{v-2}{2}\right) \times \left\{ \frac{z_i^{-\frac{v+1}{2}+1}}{-\frac{v+1}{2}+1} \right\}_1^\infty \\ &= \frac{v-2}{v-1} \end{aligned}$$

de donde se tiene:

$$M_1 = \frac{\Gamma((v-1)/2)\sqrt{v-2}}{\Gamma(v/2)\sqrt{\pi}} \quad (2.16)$$

Sustituyendo (2.16) en la expresión (2.9) se obtiene que el vector de medias de la versión asimétrica de la función de densidad (2.15) esta dada por:

$$\mu_i = (\gamma_i - \gamma_i^{-1}) \frac{\Gamma((v-1)/2)\sqrt{v-2}}{\Gamma(v/2)\sqrt{\pi}}$$

Por último, de la expresión (2.12) se tiene la distribución de probabilidad t -Student asimétrica y estandarizada $SST(0, I_m, \gamma, v)$,

$$p(z|\gamma, v) = \left(\frac{2}{\sqrt{\pi}}\right)^m \left(\prod_{i=1}^m \frac{\gamma_i \sigma_i}{1 + \gamma_i^2}\right) \times \frac{\Gamma(\frac{v+m}{2})}{\Gamma(\frac{v}{2})(v-2)^{\frac{m}{2}}} \left(1 + \frac{(z^*)^t z^*}{v-2}\right)^{-\frac{(v+m)}{2}} \quad (2.17)$$

siendo, σ_i y z_i^* dado por las expresiones (2.10) y (2.12), respectivamente.

2.3.5. Distribución de Error Generalizada (GED)

Algunas generalizaciones multivariadas de la distribución GED fueron propuestas en Gómez et al. (1998) y Giller (2005), pero las marginales de esas distribuciones son difíciles de obtener, así como los momentos absolutos de las marginales. Por este motivo, en Fioruci et al. (2014b) se opta por utilizar la distribución conjunta de m variables aleatorias independientes, garantizando así que las marginales serán las distribuciones presentadas a continuación.

$$p(x|k) = \left[\frac{\Gamma(3/k)}{\Gamma(1/k)}\right]^{1/2} \frac{\exp\{-[\frac{\Gamma(3/k)}{\Gamma(1/k)}]x^2\}^{k/2}}{2\Gamma((k+1)/k)} \quad (2.18)$$

Esta distribución generalizada de la distribución Normal puede tener colas más ligeras ($k > 2$) o más pesadas ($k < 2$) que la distribución Normal estándar $N(0, 1)$, y si ($k = 2$) se obtiene la distribución Normal. De esta forma, la distribución de probabilidad conjunta del vector aleatorio $x = (x_1, \dots, x_m)^t$ es dada por:

$$p(x|k) = \left[\frac{\Gamma(3/k)}{\Gamma(1/k)}\right]^{m/2} \frac{\exp\{-[\frac{\Gamma(3/k)}{\Gamma(1/k)}]^{k/2} \sum_{i=1}^m |x_i|^k\}}{[2\Gamma((k+1)/k)]^m} \quad (2.19)$$

Como la distribución (2.18) es estandarizada, se tiene que $E(X) = 0$ y $Var(X) = I_m$ y así se puede utilizar el método de Bauwens and Laurent (2005) para insertar asimetría. Para facilitar la notación, nos referiremos a la distribución GED Multivariada como $GED(0, I_m, k)$ y a la distribución asimétrica estandarizada resultante de la aplicación del método como $SSGED(0, I_m, \gamma, k)$.

Se calcula primero el primer momento absoluto de la distribución $GED(0, 1, k)$:

$$M_1 = 2 \left[\frac{\Gamma(3/k)}{\Gamma(1/k)} \right]^{1/2} \frac{1}{2\Gamma((k+1)/k)} \times \int_0^\infty x \exp\left\{-\left[\frac{\Gamma(3/k)}{\Gamma(1/k)} x^2\right]^{k/2}\right\} dx$$

si:

$$\frac{\Gamma(3/k)}{\Gamma(1/k)} x^2 = z$$

derivando

$$\frac{\Gamma(3/k)}{\Gamma(1/k)} 2x dx = dz$$

si $x \rightarrow 0$, $z \rightarrow 0$ y si $x \rightarrow \infty$, $z \rightarrow \infty$; luego

$$\begin{aligned} \int_0^\infty x \exp\left\{-\left[\frac{\Gamma(3/k)}{\Gamma(1/k)} x^2\right]^{k/2}\right\} dx &= \frac{\Gamma(1/k)}{2\Gamma(3/k)} \times \int_0^\infty \exp\{-z^{k/2}\} dz \\ &= \frac{\Gamma(1/k)}{2\Gamma(3/k)} \times \Gamma\left(\frac{2}{k} + 1\right) \end{aligned}$$

y aplicando la propiedad $\Gamma(n+1) = n\Gamma(n)$ se tiene la siguiente igualdad

$$M_1 = \frac{\Gamma(2/k)}{[\Gamma(1/k)\Gamma(3/k)]^{1/2}} \quad (2.20)$$

Por lo tanto, sustituyendo (2.20) en la expresión (2.8) se obtiene que el vector de medias de la versión asimétrica de la función de densidad (2.19) está dado por:

$$\mu_i = (\gamma_i - \gamma_i^{-1}) \frac{\Gamma(2/k)}{[\Gamma(1/k)\Gamma(3/k)]^{1/2}} \quad (2.21)$$

para $i = 1, \dots, m$.

Por último, sustituyendo (2.19) en (2.11) se obtiene la distribución de probabilidad GED asimétrica y estandarizada $SSGED(0, I_m, \gamma, k)$:

$$p(z|\gamma) = 2^m \left[\prod_{i=1}^m \frac{\gamma_i \sigma_i}{1 + \gamma_i^2} \right] \left[\frac{\Gamma(3/k)}{\Gamma(1/k)} \right]^{m/2} \times \frac{\exp\left\{-\left[\frac{\Gamma(3/k)}{\Gamma(1/k)}\right]^{k/2} \sum_{i=1}^m |z_i^*|^k\right\}}{(2/k)^m [\Gamma(1/k)]^m} \quad (2.22)$$

siendo μ_i , σ_i^2 y z_i^* dados por (2.21), (2.10) y (2.12), respectivamente.

Así, como la distribución *SSN* consiste en un caso particular de *SST*, también consiste en un caso particular de *SSGED*, con $k = 2$.

Las distribuciones de probabilidad presentadas en esta sección se aplicarán en los modelos GARCH univariados y multivariados.

2.4. Series de Tiempo Financieras

En esta sección se presentan algunos conceptos básicos que se utilizarán en las siguientes secciones.

2.4.1. Series de Tiempo Financieras

El análisis de series temporales financieras se centra en la valoración de los activos a lo largo del tiempo. Una característica clave que distingue el análisis de series temporales financieras de otros análisis de series temporales es la volatilidad, un elemento de incertidumbre que no es directamente observable. Como resultado de esta incertidumbre adicional, se utilizan los llamados modelos de heteroscedasticidad condicional autorregresiva.

Se analizan los conceptos básicos de los rendimientos de activos, ya que la mayoría de los estudios financieros involucran retornos, en lugar de precios de los activos. Según Tsay (2005), hay dos razones principales para usar retornos: El primero, es que para los inversores promedio, el retorno de un activo es un resumen completo y no a escala de la oportunidad de inversión. La segunda razón, es porque la serie de retorno es más fácil de analizar que la serie de precios, puesto que la serie de retornos tiene mejores propiedades estadísticas.

2.4.2. Retornos

Uno de los objetivos en finanzas es la validación de riesgos de una cartera de activos financieros. El riesgo es frecuentemente medido en términos de variación de los precios de los activos. Hay dos métodos que se utilizan para calcular los riesgos a partir de una serie de precios (P_t), y estos implican la formación de rendimientos simples y rendimientos compuestos de forma continua (log-retornos). Denotamos por (P_t) el precio de un activo en el instante t , en un día normal de negocios, se debe suponer que no haya dividendos pagados en el periodo. Los retornos simples para la variación del precio entre los instantes $t - 1$ y t de este activo es definido por:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{\Delta P_t}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1.$$

y los rendimientos compuestos de forma continua (log - retornos):

$$y_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log(1 + R_t) = p_t - p_{t-1}$$

donde $p_t = \log P_t$,

los log - retornos son llamados también **retornos o rendimientos**.

2.5. Inferencia Bayesiana

La información que se tiene sobre un parámetro es esencial en cualquier serie que se desee estudiar, sin embargo el verdadero valor de este parámetro es desconocido, la idea es entonces tratar de reducir este error, además de la intensidad de esta incertidumbre que puede asumir diferentes niveles. Desde el punto de vista bayesiano, estos diferentes niveles de incertidumbre se representan a través de modelos probabilísticos para el parámetro en estudio. En este contexto, es natural que diferentes investigadores puedan tener diferentes niveles de incertidumbre sobre el parámetro, dando diferentes modelos. Por lo tanto, no hay distinción

entre cantidades observables y los parámetros de un modelo estadístico, todos se consideran cantidades aleatorias.

2.5.1. Teorema de Bayes

Supóngase que $y = \{y_1, \dots, y_n\}$ es un vector de n observaciones cuya distribución de probabilidad $p(y|\theta)$ depende de los valores de k parámetros $\theta = \{\theta_1, \dots, \theta_k\}$. Supóngase también que θ tiene por sí mismo una función de distribución $p(\theta)$.

Entonces

$$p(y|\theta)p(\theta) = p(y; \theta) = p(\theta|y)p(y)$$

dados los datos observados y , la distribución condicional de θ es:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (2.23)$$

Se puede escribir alternativamente la ecuación (2.23) como

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

$$p(\theta|y) = cp(y|\theta)p(\theta)$$

En la última ecuación, $p(\theta)$ es la distribución a priori de θ ($\pi(\theta)$), $p(\theta|y)$ es la distribución a posteriori de θ dado y ($\pi(\theta|y)$), la constante c se utiliza para que la distribución integre (o sume en caso de ser discreto) uno, y $p(y|\theta)$ es una función de θ llamada función de verosimilitud, es decir la función de verosimilitud queda expresada como $L(\theta|y) = p(y|\theta)$.

La función de verosimilitud juega un papel muy importante en la inferencia bayesiana ya que es la función a través de la cual los datos “y” modifican el conocimiento a priori de θ .

Con esta definición, teniendo la distribución a priori $\pi(\theta)$, y la distribución a posteriori $\pi(\theta|y)$ para θ dado y , se puede expresar el teorema de Bayes como:

Distribución a posteriori \propto Distribución a priori \times verosimilitud

Es decir, el análisis bayesiano es desarrollado por combinación de la distribución a priori $\pi(\theta)$ y la información muestral $y = \{y_1, \dots, y_n\}$ dentro de lo que es llamado distribución a posteriori para θ dado y ($\pi(\theta|y)$), de la cual toda decisión e información son hechas.

Bajo el enfoque Bayesiano es necesario entonces determinar la distribución a priori de los parámetros del modelo.

2.5.2. Distribución a Priori

El uso de información previa en Inferencia Bayesiana requiere la especificación de una distribución a priori $\pi(\theta)$ para el parámetro de interés. Esta distribución debe representar (probabilísticamente) el conocimiento que uno tiene sobre el parámetro antes de llevar a cabo el experimento.

La selección de la distribución a priori es un tema importante en la estadística bayesiana. Si se dispone de información previa sobre los datos o el modelo, puede (y debe) utilizarse en la construcción de la a priori. En muchas situaciones, sin embargo, la selección de la distribución a priori es bastante delicada, debido a la ausencia de información previa confiable, y en su lugar se deben elegir soluciones predeterminadas. Esto se debe a que la elección de la distribución a priori tiene una influencia considerable en la inferencia resultante, este paso inferencial debe realizarse con el mayor cuidado posible. Desde un punto de vista computacional, la elección más conveniente de distribuciones a priori es imitar la estructura de probabilidad dentro de la a priori. En los casos más ventajosos, las a priori y a posteriori permanecen dentro de la misma familia parametrizada. Estos antecedentes se denominan conjugados Robert (2007), tales antecedentes existen para la mayoría de las familias habituales, incluida la distribución normal.

Dado que las a priori conjugadas son tales que las densidades a priori y a posteriori pertenecen a la misma familia paramétrica, el uso de las observaciones se reduce a una actualización de los parámetros de las a prioris. Para evitar confusiones, los parámetros involucrados en la distribución a priori de los parámetros del modelo generalmente se denominan hiperparámetros. (Pueden asociarse a sí mismos con distribuciones a priori, que luego se denominan hiperprioris).

Definición 2.18 Se dice que una familia \mathcal{F} de distribuciones de probabilidad en Θ (espacio paramétrico) es **conjugada (o cerrada bajo muestreo)** para una función de probabilidad $f(x|\theta)$ si, para todo $\pi \in \mathcal{F}$, la distribución a posteriori $\pi(\theta|x)$ también pertenece a \mathcal{F} .

2.5.3. Distribución a Posteriori

Si, además de la distribución muestral $f(x|\theta)$, se dispone de una distribución previa de θ , $\pi(\theta)$, es decir, que se trata de un modelo bayesiano completo. Dadas estas dos distribuciones, se construyen varias distribuciones, como son:

(a) La distribución conjunta:

$$\varphi(\theta, x) = f(x|\theta)\pi(\theta)$$

(b) La distribución marginal de x ,

$$\begin{aligned} m(x) &= \int \varphi(\theta, x) d\theta \\ &= \int f(x|\theta)\pi(\theta) d\theta; \end{aligned}$$

(c) La distribución a posteriori de θ , obtenida por la fórmula de Bayes.

$$\begin{aligned} \pi(x|\theta) &= \frac{f(x|\theta)\pi(\theta)}{\int \varphi(\theta, x) d\theta} \\ &= \frac{f(x|\theta)\pi(\theta)}{m(x)}; \end{aligned}$$

(d) La distribución predictiva de y donde $y \sim g(y|\theta, x)$, obtenida por

$$g(y|x) = \int g(y|\theta, x)\pi(\theta|x)d\theta$$

Es decir la distribución a posteriori combina la credibilidad a priori acerca de θ con la información acerca del contenido en la muestra x dando la credibilidad final sobre θ .

Ejemplo 2.2 Sea $x \sim N(\theta, \sigma^2)$, donde θ es desconocido y σ^2 conocido, $\pi(\theta) \sim N(\mu, \tau^2)$ donde μ, τ^2 son conocidos. Hallar la distribución a posteriori de θ .

Solución:

$$\begin{aligned} x &\sim N(\theta, \sigma^2) \\ \pi(x|\theta) &= \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2} \frac{(x-\theta)^2}{\sigma^2}\right\} \\ \pi(\theta) &= \frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{1}{2} \frac{(\theta-\mu)^2}{\tau^2}\right\} \\ \pi(x; \theta) &= \pi(\theta)\pi(x|\theta) = \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2} \left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2} \right]\right\} \end{aligned}$$

La distribución conjunta es:

$$\pi(x; \theta) = \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2}\rho\left[\theta - \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right)\right]^2\right\} \exp\left\{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2 + \tau^2}\right\}$$

donde

$$\begin{aligned} \rho &= \frac{\sigma^2 + \tau^2}{\sigma^2\tau^2} \\ m(x) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2}\rho\left[\theta - \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right)\right]^2\right\} \exp\left\{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2 + \tau^2}\right\} d\theta \\ m(x) &= \frac{1}{\sqrt{2\pi\sigma\tau}} \exp\left\{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2 + \tau^2}\right\} \sqrt{\frac{1}{\rho}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{1}{\rho}}} \exp\left\{-\frac{1}{2} \frac{\left[\theta - \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right)\right]^2}{\frac{1}{\rho}}\right\} d\theta \end{aligned}$$

como

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{1}{\rho}}} \exp\left\{-\frac{1}{2} \frac{\left[\theta - \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right)\right]^2}{\frac{1}{\rho}}\right\} d\theta = 1$$

pues

$$\theta \sim N\left(\frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right), \frac{1}{\rho}\right)$$

$$m(x) = \frac{1}{\sqrt{2\pi}\sqrt{\rho}\sigma\tau} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2 + \tau^2}\right\}$$

La distribución marginal de x es:

$$m(x) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 + \tau^2}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2 + \tau^2}\right\}$$

de donde

$$x \sim N(\mu, \sigma^2 + \tau^2)$$

La distribución a posteriori de θ es:

$$\pi(\theta|x) = \frac{\pi(x;\theta)}{m(x)} = \frac{1}{\sqrt{2\pi}}\sqrt{\rho}\exp\left\{-\frac{1}{2}\rho\left[\theta - \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right)\right]^2\right\}$$

$$\pi(\theta|x) = \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{\rho}}}\exp\left\{-\frac{1}{2}\frac{\left[\theta - \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right)\right]^2}{\frac{1}{\rho}}\right\}$$

$$\pi(\theta|x) \equiv N\left(\frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right); \frac{1}{\rho}\right)$$

es decir

$$\theta \sim N\left(\frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right); \frac{1}{\rho}\right)$$

con media a posteriori

$$\mu(x) = \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right)$$

$$\mu(x) = \frac{\mu\sigma^2 + \tau^2x}{\sigma^2 + \tau^2}$$

Ejemplo 2.3 Sea $x \sim G(v, \theta)$, donde θ es desconocido y v conocido, $\pi(\theta) \sim G(\alpha, \beta)$ donde α, β son conocidos. Hallar la distribución a posteriori de θ .

Solución:

$$x \sim G(v, \theta)$$

$$\pi(x|\theta) = \frac{\theta^v x^{v-1} \exp\{-\theta x\}}{\Gamma(v)}$$

$$\pi(\theta) = \frac{\beta^\alpha \theta^{\alpha-1} \exp\{-\beta\theta\}}{\Gamma(\alpha)}$$

La distribución conjunta es:

$$\begin{aligned}\pi(x; \theta) &= \pi(\theta)\pi(x|\theta) = \frac{\theta^{\alpha+v-1}x^{v-1}\beta^\alpha \exp\{-\theta(x + \beta)\}}{\Gamma(\alpha)\Gamma(v)} \\ m(x) &= \int_0^\infty \frac{\theta^{\alpha+v-1}x^{v-1}\beta^\alpha \exp\{-\theta(x + \beta)\}}{\Gamma(\alpha)\Gamma(v)} d\theta \\ m(x) &= \frac{x^{v-1}\beta^\alpha}{\Gamma(\alpha)\Gamma(v)} \int_0^\infty \theta^{\alpha+v-1} \exp\{-\theta(x + \beta)\} d\theta \\ m(x) &= \frac{x^{v-1}\beta^\alpha \Gamma(\alpha + v)}{\Gamma(\alpha)\Gamma(v)(x + \beta)^{\alpha+v}} \int_0^\infty \frac{\theta^{\alpha+v-1}(x + \beta)^{\alpha+v} \exp\{-\theta(x + \beta)\}}{\Gamma(\alpha + v)} d\theta\end{aligned}$$

como

$$\int_0^\infty \frac{(x + \beta)^{\alpha+v} \theta^{v+\alpha-1} \exp\{-\theta(x + \beta)\}}{\Gamma(\alpha + v)} d\theta = 1$$

pues

$$\theta \sim G(\alpha + v, x + \beta)$$

La distribución marginal de x es:

$$m(x) = \frac{x^{v-1}\beta^\alpha \Gamma(\alpha + v)}{\Gamma(\alpha)\Gamma(v)(x + \beta)^{\alpha+v}}$$

La distribución a posteriori de θ es:

$$\begin{aligned}\pi(\theta|x) &= \frac{\pi(x; \theta)}{m(x)} = \frac{\theta^{\alpha+v-1}x^{v-1}\beta^\alpha \exp\{-\theta(x + \beta)\} / \{\Gamma(\alpha)\Gamma(v)\}}{x^{v-1}\beta^\alpha \Gamma(\alpha + v) / \{\Gamma(\alpha)\Gamma(v)(x + \beta)^{\alpha+v}\}} \\ \pi(\theta|x) &= \frac{(x + \beta)^{\alpha+v} \theta^{v+\alpha-1} \exp\{-\theta(x + \beta)\}}{\Gamma(\alpha + v)} \\ \pi(\theta|x) &\equiv G(\alpha + v, x + \beta)\end{aligned}$$

es decir

$$\theta \sim G(\alpha + v, x + \beta)$$

con media a posteriori

$$\mu(x) = \frac{\alpha + v}{x + \beta}$$

2.6. Modelo ARCH Univariado

Dada la serie de retornos $y = \{y_t, t = 1, \dots, T\}$ el modelo heterocedástico condicional autoregresivo ARCH(r) se define como:

$$y_t = \epsilon_t \sqrt{h_t}, \epsilon_t \sim D(0, 1)$$

$$h_t = \omega + \sum_{i=1}^r \alpha_i y_{t-i}^2$$

siendo h_t la variación condicional (no observable) de y_t dada la información previa $I_t = \{y_{t-1}, y_{t-2}, \dots\}$, los errores ϵ_t son independientes e idénticamente distribuidos. $D(0, 1)$ denota una distribución con media 0 y una varianza 1. Las restricciones de estacionaridad y positividad son $\omega > 0, \alpha_i \geq 0, i = 1, \dots, r$

Si $r = 1$ entonces ARCH(1)

$$y_t = \epsilon_t \sqrt{h_t}$$

$$h_t = \omega + \alpha_1 y_{t-1}^2$$

donde $\omega > 0, \alpha_1 \geq 0$.

Se calcula la media, varianza y la autocovarianza incondicional de la serie.

$$E[y_t] = E[E[y_t | I_t]] = 0$$

$$Var[y_t] = E[y_t^2] = E[E(y_t^2 | I_t)] = E(\omega + \alpha_1 y_{t-1}^2) = \omega + \alpha_1 E(y_{t-1}^2)$$

si el proceso y_t fuese estacionario de segundo orden, entonces, $\forall t$,

$$E[y_t^2] = E[y_{t-1}^2] = Var[y_t]$$

de donde

$$Var[y_t] = \frac{\omega}{1 - \alpha_1} > 0$$

entonces $0 \leq \alpha_1 < 1$.

$$cov(y_{t+k}; y_t) = E(y_{t+k} y_t), k \geq 1$$

$$cov(y_{t+k}; y_t) = E[E(y_{t+k} y_t | I_{t+K-1})] = E[y_t E(y_{t+k} | I_{t+K-1})]$$

$$cov(y_{t+k}; y_t) = E[y_t E(\sqrt{h_{t+K}} \epsilon_{t+K} | I_{t+K-1})] = 0$$

entonces

$$\text{cov}(y_{t+k}; y_t) = 0$$

Lo cual indica que y_t es una secuencia de variables aleatorias no correlacionadas con media 0 y varianza

$$\text{Var}[y_t] = \frac{\omega}{1 - \alpha_1} > 0$$

Se sabe que los retonos presentan generalmente colas largas de modo que la curtosis es mayor que 3.

Para calcular la curtosis, supongamos que y_t sigue el modelo ARCH(1), por ende es necesario calcular el momento de cuarto orden de y_t , para facilitar los cálculos supóngase que $\epsilon_t \sim D(0, 1)$ entonces

$$\begin{aligned} E[y_t^4] &= E[h_t^2 \epsilon_t^4 | I_t] = 3(\omega + \alpha_1 y_{t-1}^2)^2 \\ E[y_t^4] &= 3[\alpha_1 y_{t-1}^2] = 3E[\omega^2 + \alpha_1^2 y_{t-1}^4 + 2\omega\alpha_1 y_{t-1}^2] \end{aligned}$$

se admite que el proceso es estacionario de cuarto orden, luego se escribe el momento de cuarto orden

$$\begin{aligned} \mu_4 &= E[y_t^4] = 3[\omega^2 + \alpha_1 + 2\omega\alpha_1 \text{Var}[y_t] + \alpha_1^2 \mu_4] \\ &= 3[\omega^2 + 2\omega\alpha_1 \frac{\omega}{1 - \alpha_1} + \alpha_1^2 \mu_4] \\ &= 3\omega^2(1 + 2\frac{\omega}{1 - \alpha_1}) \frac{1}{1 - 3\alpha_1^2} \end{aligned}$$

se sabe que

$$\begin{aligned} k &= \frac{\mu_4}{\text{Var}[y_t]^2} = 3\omega^2 \frac{1 + \alpha_1}{1 - \alpha_1} \frac{1}{1 - 3\alpha_1^2} \frac{1 - \alpha_1^2}{\omega^2} \\ k &= 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \end{aligned}$$

si los momentos de cuarto orden son finitos y positivos entonces

$$1 - 3\alpha_1^2 > 0$$

es decir

$$0 \leq \alpha_1^2 < \frac{1}{3}$$

Por lo tanto, cuanto más restricciones se imponen al proceso de retornos más restricciones se tendrían para los coeficientes del modelo.

De lo anterior, si se admite que y_t sigue un modelo ARCH(1), las colas serán más pesadas que el de la normal, lo cual es una propiedad ventajosa del modelo.

Desventaja: Una desventaja del modelo es que trata retornos positivos y negativos de forma similar, ya que en la fórmula de la volatilidad ingresan retornos al cuadrado; y se sabe en la práctica que la volatilidad reacciona de forma diferente para retornos positivos que para retornos negativos.

2.7. Cadenas de Markov y MCMC

2.7.1. Introducción Histórica

Los métodos MCMC están basados en la utilización del método de Monte Carlo a través de cadenas de Markov. Los métodos de Monte Carlo (MC) son métodos estadísticos usados para aproximar relaciones complejas y costosas de evaluar. Este método fue desarrollado en los años cuarenta y su nombre le viene dado por el Casino de Montecarlo, que en esos años era “la capital de los juegos de azar”. El iniciador de este método fue Stanislaw Marcin Ulam (matemático polaco que participó en el proyecto Manhattan y que desarrolló herramientas matemáticas para teoría de números, teoría de conjuntos, teoría ergódica y topología algebraica) que tras haber pasado varios días jugando al solitario debido a una enfermedad, observó que era mucho más simple tener una idea del resultado general del juego haciendo pruebas múltiples y contando las proporciones de los resultados, que calcular todas las posibilidades. Esta idea la consiguió al extrapolar su trabajo en el que estaba teniendo problemas para calcular las ecuaciones integro - diferenciales que regían la difusión de neutrones.

Como se mencionó anteriormente, Ulam estaba trabajando en el Proyecto Manhattan, cuya base militar se encontraba en Los Álamos. Este proyecto científico fue dirigido por el físico estadounidense Robert Serber (2020d) durante la Segunda Guerra Mundial para el Gobierno de los Estados Unidos y con colaboración de Reino Unido y de Canadá. El objetivo era desarrollar la primera bomba atómica antes que Alemania, Proyecto Uranio, y la URS, Proyecto Borodino (2020c). En las primeras bombas atómicas desarrolladas no se tuvo en cuenta la idea de Ulam.

Años más tarde, una vez finalizada la guerra, hubo tiempo de aplicar este método en las investigaciones de física nuclear y física cuántica. Una vez que Ulam convenció a su colega de trabajo John von Neumann del potencial de este nuevo método, ambos matemáticos trabajaron conjuntamente para el desarrollo de éste. El verdadero potencial de este método lo obtuvieron Ulam, Nicholas Constantine Metropolis (matemático, físico y computador científico griego reclutado en el Proyecto Manhattan por Oppenheimer (2020b)) y Enrico Fermi (físico italiano que realizó contribuciones en mecánica estadística, física cuántica, nuclear y de partículas (2020a)) al obtener los valores estimados de la ecuación de Schrödinger que describe la evolución temporal de una partícula masiva no relativista, para la captura de neutrones a nivel nuclear Mackay (1986).

Las integrales de los estimadores que se resolvieron mediante el método de Monte Carlo se pueden encontrar en cualquier libro de física cuántica y nuclear. Aunque la investigación con este método comenzó en los proyectos secretos que llevó a cabo Estados Unidos, la primera publicación donde se citó este método data de 1949 en un artículo de Metropolis and Ulam (1949). La primera simulación de Monte Carlo fue llevada a cabo por un equipo encabezado por Metropolis, esta simulación se realizó en la computadora ENIAC (considerada la primera computadora electrónica digital) en 1948 en la Universidad de Pennsylvania. El MC es considerado uno de los 10 algoritmos más importantes desarrollados en el

Siglo XX.

En la literatura existente hay muchos ejemplos que utilizan los métodos de Monte Carlo y estos se pueden encontrar en diferentes áreas como: biología, química, ciencia de la computación, economía y finanzas, estadística, etc.

2.7.2. Nociones Básicas

A continuación se presentan algunas definiciones y un resumen de las propiedades de las cadenas de Markov, que son esenciales para el estudio de los métodos MCMC.

En la configuración de los algoritmos MCMC, las cadenas de Markov se construyen a partir de un núcleo de transición K , y una densidad de probabilidad tal que $X_{n+1} \sim K(X_n; X_{n+1})$. Un ejemplo típico lo proporciona el proceso de caminata aleatoria, definido a continuación.

Definición 2.19 Una secuencia de variables aleatorias (X_n) es un **caminata aleatoria** si satisface

$$X_{n+1} = X_n + \epsilon_n,$$

donde, ϵ_n se genera independientemente de X_n, X_{n-1}, \dots . Si la distribución de ϵ_n , es simétrica alrededor de cero, la secuencia se denomina **caminata aleatoria simétrica**.

Dada $\mathcal{B}(\mathcal{X})$ una σ -álgebra de subconjuntos de \mathcal{X} (espacio muestral), se tienen las siguientes definiciones.

Definición 2.20 Un **núcleo de transición** es una función K definida en $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ tal que

- (i) $\forall x \in \mathcal{X}, K(x, \cdot)$ es una medida de probabilidad;
- (ii) $\forall A \in \mathcal{B}(\mathcal{X}), K(\cdot, A)$ es medible.

Cuando K es discreto, el núcleo de transición simplemente es una matriz (de transición) K con elementos

$$P_{xy} = P(X_n = y | X_{n-1} = x) \quad x, y \in \mathcal{X}$$

Definición 2.21 En el caso continuo, el núcleo también denota la densidad condicional $K(x, x^t)$ de la transición $K(x, \cdot)$; es decir,

$$P(X \in A | x) = \int_A K(x, x^t) dx^t.$$

Una cadena de Markov es una secuencia de variables aleatorias, que se puede considerar que evolucionan con el tiempo con probabilidad de una transición dependiendo del conjunto particular en el que se encuentra la cadena, es por ello que se define a una cadena de Markov en función de K .

Definición 2.22 Dado un núcleo de transición K , una secuencia de variables aleatorias $X_0, X_1, \dots, X_n, \dots$ es una **cadena de Markov**, denotada por (X_n) , si, para cualquier t , la distribución condicional de X_t ; dado $x_{t-1}, x_{t-2}, \dots, x_0$ es lo mismo que la distribución de X_t ; dado x_{t-1} ; es decir,

$$P(X_{k+1} \in A | x_0, x_1, \dots, x_k) = P(X_{k+1} \in A | x_k) = \int_A K(x_k, dx)$$

Definición 2.23 La cadena es **homogénea en el tiempo**, o simplemente **homogénea**, si la distribución de $(X_{t_1}, \dots, X_{t_k})$ dada x_{t_0} , es la misma que la distribución de $(X_{t_1-t_0}, X_{t_2-t_0}, \dots, X_{t_k-t_0})$ dado x_0 para cada k y cada $(k+1)$ salidas $t_0 \leq t_1 \leq \dots \leq t_k$.

Entonces, en el caso de una cadena de Markov, si se conoce la distribución inicial o el estado inicial, la construcción de la cadena de Markov (X_n) está completamente determinada por su transición, es decir, por la distribución de X_n , condicionalmente en x_{n-1} .

En el caso general, el hecho de que el núcleo K determina las propiedades de la

cadena (X_n) se puede inferir de las relaciones

$$\begin{aligned}
P_x(X \in A_1) &= K(x, A_1) \\
P_x((X_1, X_2) \in A_1 \times A_2) &= \int_{A_1} K(y_1, A_2)K(x, dy_1) \\
P_x((X_1, X_2, A_3) \in A_1 \times A_2 \times A_3) &= \int_{A_1} \int_{A_2} K(y_2, A_3)K(x, dy_1)K(y_1, dy_2) \\
&\dots \\
P_x((X_1, \dots, A_n) \in A_1 \times \dots \times A_n) &= \int_{A_1} \dots \int_{A_{n-1}} K(y_{n-1}, A_n) \\
&\quad \times K(x, dy_1) \dots K(y_{n-2}, dy_{n-1}).
\end{aligned}$$

En particular, la relación $P_x(X \in A_1) = K(x, A_1)$ indica que $K(x_n, dx_{n+1})$ es una versión de la distribución condicional de X_{n+1} dado X_n .

Si se denota $K^1(x, A) = K(x, A)$, el núcleo para $n > 1$ transiciones estará dado por

$$K^n(x, A) = \int_{\mathcal{X}} K^{n-1}(y, A)K(x, dy) \quad (2.24)$$

de donde se derivan las ecuaciones de Chapman - Kolmogorov

Lema 2.1: Ecuaciones de Chapman-Kolmogorov Para toda $m, n \in \mathbb{N}$, $\forall x \in \mathcal{X}$ y $A \in \mathcal{B}(\mathcal{X})$,

$$K^{n+m}(x, A) = \int_{\mathcal{X}} K^m(y, A)K^n(x, dy)$$

La idea intuitiva es que para llegar a A en $n + m$ pasos partiendo de x es necesario pasar por algún y en el paso n .

Proposición 2.3: Propiedad de Markov Para toda distribución inicial μ y toda muestra (X_0, X_1, \dots, X_k) ,

$$\mathbb{E}_\mu(h(X_{k+1}, X_{k+2}, \dots) | x_0, x_1, \dots, x_k) = \mathbb{E}_{x_k}(h(X_1, X_2, \dots)),$$

con h cualquier función de medida positiva.

Es decir, cada tiempo k , condicionando sobre X_k , la cadena después del tiempo k inicia otra vez partiendo de x_k . Las siguientes cantidades permiten conocer sobre las trayectorias que podría seguir una cadena de Markov.

Definición 2.24 Considere $A \in \mathcal{B}(\mathcal{X})$. El primer n para el que la cadena entra en el conjunto A se denota por

$$\tau_A = \inf\{n > 1; X_n \in A\}$$

y se llama **tiempo de parada** en A . Por convención, $\tau_A = +\infty$ si $X_n \notin A$ para cada n . De manera más general, una función $\zeta(x_1, x_2, \dots)$ se llama regla de parada si el conjunto $\{\zeta = n\}$ es medible para la σ -álgebra inducida por (X_0, \dots, X_n) .

Definición 2.25 Asociado con el conjunto A , también se define

$$\eta_A = \sum_{n=1}^{\infty} \mathbb{I}_A(X_n),$$

el **número de pasajes** de (X_n) en A .

Proposición 2.3 Propiedad fuerte de Markov Para toda distribución inicial μ y todo tiempo de parada T finito, casi seguramente,

$$\mathbb{E}_{\mu}(h(X_{T+1}, X_{T+2}, \dots) | x_1, x_2, \dots, x_T) = \mathbb{E}_{x_T}(h(X_1, X_2, \dots)),$$

siempre que las esperanzas existan. Esto es, la propiedad de Markov se cumple en tiempos de parada.

Definición 2.26 Un estado i **conduce** a j ($i \longrightarrow j$) si para algún $n \geq 1$, $p_{ij}^{(n)} > 0$.

Definición 2.27 Se dice que un par de estados i, j se **comunican** si $i \longrightarrow j$ y $j \longrightarrow i$, es decir, si existen $n \geq 1$, $m \geq 1$ tal que $p_{ij}^{(n)} > 0$, $p_{ji}^{(m)} > 0$. Donde $p_{ij}^{(n)}$ se define como la probabilidad que la cadena esté en el estado j después de n pasos, dado que la cadena espezó en el estado i , es decir

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i)$$

2.7.3. Irreducibilidad

La irreducibilidad de la cadena (X_n) permite asegurar que todas las secciones del espacio \mathcal{X} sean visitadas por la cadena (X_n) sin importar el punto de partida. Esta

característica resulta claramente importante en los métodos MCMC. Considere uno de estos métodos donde se quiere simular una distribución específica, es importante que todo el soporte de la función esté representado por la muestra generada. En el caso en que \mathcal{X} es discreto, la irreducibilidad se verifica si ocurre que $P_x(\tau_y < \infty) > 0$ para todo y, x en \mathcal{X} . Según Casella and Robert (1999) frecuentemente ocurre que $P_x(\tau_y < \infty)$ es uniformemente igual a cero, por lo que para que la irreducibilidad esté bien definida se requiere introducir una medida auxiliar φ .

Definición 2.28 Dada una medida φ , la cadena de Markov (X_n) con núcleo de transición $K(x, y)$ es φ -irreducible si, para cada $A \in \mathcal{B}(\mathcal{X})$ con $\varphi(A) > 0$, existe n tal que $K^n(x, A) > 0$ para todo $x \in \mathcal{X}$ (equivalentemente, $P_x(\tau_A < \infty) > 0$).

La cadena es fuertemente irreducible en φ si $n = 1$ para todos los A medibles.

Una cadena de Markov con espacio de estados $\mathbb{S} \equiv \{1, 2, \dots, K\}$, $K \leq \infty$ y la matriz de probabilidad de transición $\mathbf{P} \equiv ((p_{ij}))$ es irreducible si para cada i, j en \mathbb{S} , i y j se comunican.

Teorema 2.3 La cadena (X_n) es φ -irreducible si, y sólo si, para todo $x \in \mathcal{X}$ y todo $A \in \mathcal{B}(\mathcal{X})$ con $\varphi(A) > 0$, se cumple alguna de las siguientes condiciones:

1. existe $n \in \mathbb{N}$ tal que $K^n(x, A) > 0$;
2. $\mathbb{E}_x(\eta_A) > 0$.
3. $K_\epsilon(x, A) > 0$ para un $0 < \epsilon < 1$.

Es importante señalar que la irreducibilidad es una propiedad intrínseca a la cadena de modo que no depende de la medida φ .

Definición 2.29 Dada $\mathbf{P} = ((p_{ij}))_{K \times K}$ una **matriz estocástica**, es decir, $p_{ij} \geq 0$, para todo i , $\sum_{j=1}^K p_{ij} = 1$, Un estado i es **absorbente** si $p_{ii} = 1$.

2.7.4. Átomos y Conjuntos Pequeños

La noción de átomo fue introducida por Nummelin (1978). El objetivo fue tener para el caso de espacios de estados generales un equivalente al caso discreto en el que existen puntos con medida positiva. Partiendo de este concepto es posible desarrollar la teoría de cadenas de Markov en espacios de estados generales en completa analogía al caso de espacios de estados contables.

Definición 2.30 La cadena de Markov (X_n) tiene un **átomo** $\alpha \in \mathcal{B}(\mathcal{X})$ si existe una medida asociada distinta de cero ν tal que

$$K(x, A) = \nu(A), \quad \forall x \in \alpha, \quad \forall A \in \mathcal{B}(\mathcal{X}).$$

Si (X_n) es φ -irreducible y $\varphi(A) > 0$, el átomo α es accesible.

Intuitivamente un átomo es un conjunto con probabilidad constante. Un punto siempre es un átomo. Si \mathcal{X} es contable y la cadena es irreducible, cada punto es un átomo accesible. En espacios de estados generales los átomos no son tan frecuentes, sin embargo, cuando (X_n) es φ -irreducible es posible construir artificialmente conjuntos con estructura atómica. Este resultado, de Meyn (1993) es posiblemente la mayor innovación en el análisis de cadenas de Markov en las últimas décadas, fue descubierto de formas diferentes y casi simultáneamente por Nummelin, Athreya y Ney. Si el lector está interesado en conocer el procedimiento de construcción de un pseudo-átomo puede revisar Meyn (1993).

La esencia del procedimiento es una partición probabilística del espacio de estados de forma que los átomos para la cadena de partida sean objetos naturales. Para esta construcción es necesario considerar conjuntos que satisfacen la condición de minorización que se enuncia a continuación.

Para algún $\delta > 0$, algún $C \in \mathcal{B}(\mathcal{X})$ y alguna medida de probabilidad ν con $\nu(C^c) = 0$ y $\nu(C) = 1$

$$K(x, A) \geq \delta \mathbb{I}_C(x) \nu(A), \quad A \in \mathcal{B}(\mathcal{X}), \quad x \in \mathcal{X}.$$

Esto asegura que la cadena tiene probabilidad acotada uniformemente por abajo para toda $x \in C$. Esta condición lleva a la siguiente noción.

Definición 2.31 Un conjunto C es **pequeño** si existen $m \in N$ y una medida distinta de cero ν_m , tal que

$$K^m(x, A) \geq \nu_m(A), \quad \forall x \in C, \quad \forall A \in \mathcal{B}(\mathcal{X}).$$

En este caso al conjunto C recibe el nombre de ν_m -pequeño.

En Meyn (1993) se prueba que para una cadena (X_n) φ -irreducible siempre existe un conjunto pequeño de medida positiva. El siguiente resultado dá la conexión entre conjuntos pequeños e irreducibilidad.

Teorema 2.4 Sea (X_n) una cadena φ -irreducible. Para todo conjunto $A \in \mathcal{B}(\mathcal{X})$ tal que $\varphi(A) > 0$, existe $m \in N$ y un conjunto pequeño $C \subset A$, tal que $\nu_m(C) > 0$. Más aún, \mathcal{X} puede ser descompuesto en una partición numerable de conjuntos pequeños.

Es claro que los conjuntos pequeños son más sencillos de obtener que los átomos. De hecho Meyn y Tweedie muestran que para cadenas de Markov suficientemente regulares (en el sentido topológico) todo conjunto compacto es un conjunto pequeño.

2.7.5. Aperiocidad

El comportamiento de (X_n) a veces puede estar restringido por restricciones deterministas en los movimientos de X_n a X_{n+1} . Aquí se formalizan estas restricciones y se muestra en lo que sigue que las cadenas producidas por los algoritmos de Monte Carlo cadenas de Markov MCMC no muestran este comportamiento y, por tanto, no sufren los inconvenientes asociados.

En el caso discreto, el **período** de un estado $w \in \mathcal{X}$ se define como

$$d(w) = m.c.d.\{m > 1, K^m(w, w) > 0\},$$

donde m.c.d. es el máximo común denominador.

Definición 2.32 Una cadena φ -irreducible (X_n) tiene un **ciclo** de longitud d si existe un conjunto pequeño C , un número entero asociado M , y una distribución de probabilidad ν_M tal que

$$d = m.c.d.\{m \geq 1; \exists \delta_m > 0 \text{ tal que } c \text{ es pequeño para } \nu_m > \delta_m \nu_M\}$$

El número d es independiente del conjunto pequeño C y este número caracteriza intrínsecamente a la cadena (X_n) .

Definición 2.33 El **periodo** de (X_n) se define entonces como el mayor entero d que satisface la definición anterior, y (X_n) es **aperiódico** si $d = 1$.

Si existe un conjunto pequeño A y una medida minorizante ν_1 tal que $\nu_1(A) > 0$ (por lo que es posible ir de A a A en un solo paso), se dice que la cadena es **fuertemente aperiódica**.

2.7.6. Transitoriedad y Recurrencia

En el caso de los métodos MCMC, una cadena de Markov (X_n) debe tener propiedades fuertes de estabilidad. La irreducibilidad garantiza que todas las secciones del espacio de estados sean visitadas, pero es importante que sean visitadas un número infinito de veces, esto es, el regreso seguro. A ésta propiedad se le llama recurrencia.

Definición 2.34 Un conjunto A se llama **recurrente** si $\mathbb{E}_x[\eta_A] = +\infty$ por cada $x \in A$. El conjunto A es **uniformemente transitorio** si existe una constante M tal que $\mathbb{E}_x[\eta_A] < M$ para cada $x \in A$. A es **transitorio** si existe una cobertura de \mathcal{X} por conjuntos uniformemente transitorios; es decir, una colección contable

de conjuntos B_i uniformemente transitorios; tal que

$$A = \bigcup_i B_i$$

Teorema 2.5 Sea (X_n) una cadena de Markov irreducible con un átomo α de medida positiva.

1. Si α es recurrente, entonces todo $A \in \mathcal{B}(\mathcal{X})$ con $\psi(A) > 0$, es recurrente.
2. Si α es transitorio, \mathcal{X} es transitorio.

Definición 2.35 Una cadena de Markov (X_n) es **recurrente** si

- (i) existe una medida φ tal que (X_n) es φ -irreducible, y
- (ii) para cada $A \in \mathcal{B}(\mathcal{X})$ tal que $\varphi(A) > 0$, $\mathbb{E}_x[\eta_A] = \infty$ para cada $x \in A$.

La cadena es transitoria si es irreducible en φ y si \mathcal{X} es transitoria.

Teorema 2.6 Una cadena φ -irreducible es recurrente o es transitoria.

Del **Teorema 2.5** y de la **Definición 2.35** si (X_n) es φ -irreducible entonces se puede determinar si es transitoria o recurrente con la inspección de un átomo α cuya medida sea positiva, es decir, no es necesario explorar todo el espacio de estados.

El siguiente resultado es un criterio más sencillo para determinar la recurrencia de una cadena de Markov pues está en términos de un conjunto pequeño C , y de la probabilidad de que el primer retorno a C ocurra en un tiempo finito, $P_x(\tau_C < \infty)$.

Proposición 2.4 Una cadena (X_n) φ -irreducible es recurrente si existe un conjunto pequeño C con $\varphi(C) > 0$, tal que $P_x(\tau_C < \infty) = 1$ para todo $x \in C$. Sea η_A el número de pasajes de (X_n) en A , y consideramos $P_x(\eta_A = \infty)$, la probabilidad de visitar A un número infinito de veces.

2.7.7. Harris - Recurrencia

La noción de Harris - recurrencia fue introducida por Harris (1956). Esta propiedad es más fuerte que la propiedad de recurrencia. En este caso no basta con que el número promedio de retornos a cada conjunto de medida positiva sea infinito, ahora es necesario que el número de retornos sea infinito con probabilidad 1.

Definición 2.36 Un conjunto A es **Harris - recurrente** si $P_x(\eta_A = \infty) = 1$ para todo $x \in A$. La cadena (X_n) es **Harris - recurrente** si existe una medida φ tal que (X_n) es φ -irreducible y para cada conjunto A con $\varphi(A) > 0$, A es Harris - recurrente.

La siguiente proposición expresa la Harris - recurrencia como una condición sobre $P_x(\tau_A < \infty)$.

Proposición 2.5 Si para cada $A \in \mathcal{B}(\mathcal{X})$, $P_x(\tau_A < \infty) = 1$ para cada $x \in A$, entonces $P_x(\eta_A = \infty) = 1$, para todo $x \in \mathcal{X}$, y (X_n) es Harris - recurrente.

Observe que la propiedad Harris - recurrencia no es necesaria cuando \mathcal{X} es finito o contable, pues en este caso es posible mostrar que $\mathbb{E}_x[\eta_x] = \infty$ si y sólo si $P_x(\tau_x < \infty) = 1$ para toda $x \in \mathcal{X}$. En este caso, un resultado similar al mencionado en la **Proposición 2.4** es el siguiente. Sólo obsérvese que ahora se requiere que $P_x(\tau_C < \infty) = 1$ para todo $x \in \mathcal{X}$, no sólo en C .

Teorema 2.6 Si (X_n) es una cadena de Markov φ -irreducible con un conjunto pequeño C tal que $P_x(\tau_C < \infty) = 1$ para todo $x \in \mathcal{X}$, entonces (X_n) es Harris - recurrente.

Tierney (1994) y Chan and Geyer (1994) analizan el papel de la recurrencia de Harris en la configuración de los algoritmos de Monte Carlo en cadena de Markov MCMC y señalan que la recurrencia de Harris es válido para la mayoría de estos algoritmos.

2.7.8. Medidas Invariantes

Definición 2.37 Una medida σ -finita π es **invariante** para el núcleo de transición $K(x, B)$, $x \in \mathcal{X}$, $B \in \mathcal{B}(\mathcal{X})$ y para la cadena asociada, si

$$\pi(B) = \int_{\mathcal{X}} K(x, B)\pi dx, \forall B \in \mathcal{B}(\mathcal{X})$$

Definición 2.38 Cuando existe una medida de probabilidad invariante para una cadena φ -irreducible (por lo tanto recurrente), la cadena es **positiva**. Las cadenas recurrentes que no permiten una medida invariante finita se denominan **nulas recurrentes**.

A la distribución invariante π se le llama estacionaria cuando es una medida de probabilidad, pues significa que si $X_0 \sim \pi$ entonces $X_n \sim \pi$, es decir, la cadena es estacionaria en distribución. Además cuando la cadena es irreducible y tiene una medida invariante σ -finita ésta es única.

Proposición 2.6 Si la cadena (X_n) es positiva, es recurrente.

De la **Proposición 2.6** vemos que ya no es necesario decir Harris - recurrente positiva, ahora es suficiente decir Harris positiva.

Observación 2.1 Las cadenas producidas por los métodos de Monte Carlo de cadena de Markov MCMC están, por naturaleza, garantizadas para poseer una distribución invariante.

Observación 2.2 Una cadena de Markov con un núcleo de transición K satisface la **ecuación de balance detallado** si existe una función π que satisfaga

$$K(y, x)\pi(y) = K(x, y)\pi(x), \forall (x, y).$$

2.7.9. Ergodicidad y Convergencia

Al considerar la cadena (X_n) evolucionando en el tiempo, es importante preguntarnos ¿a dónde converge?, es decir, cuál será la distribución de X_n para n muy grande. Una candidata natural es la distribución invariante π . A continuación se dan las condiciones que se deben cumplir en el caso numerable y en el caso general para que π coincida con la distribución límite de la cadena.

Definición 2.39 Para una cadena Harris positiva (X_n) , con distribución invariante π , un átomo α es ergódico si

$$\lim_{n \rightarrow \infty} |K^n(\alpha, \alpha) - \pi(\alpha)| = 0.$$

En el caso contable, la existencia de un **átomo ergódico** es, de hecho, suficiente para establecer la convergencia según la norma de variación total,

$$\| \mu_1 - \mu_2 \|_{TV} = \sup_A |\mu_1(A) - \mu_2(A)|.$$

Proposición 2.7 Si (X_n) es Harris positivo en \mathcal{X} y numerable, y si existe un átomo ergódico $\alpha \in \mathcal{X}$, entonces, para cada $x \in \mathcal{X}$,

$$\lim_{n \rightarrow \infty} \| K^n(x, y) - \pi(y) \|_{TV} = 0, \quad y \in \mathcal{B}(\mathcal{X}).$$

Teorema 2.7 Para una cadena de Markov recurrente positiva y aperiódica sobre un espacio contable, y para todo estado inicial x

$$\lim_{n \rightarrow \infty} \| K^n(x, y) - \pi(y) \|_{TV} = 0, \quad y \in \mathcal{B}(\mathcal{X}).$$

Es decir, K^n converge asintóticamente a π con respecto a la norma de variación total (TV por sus siglas en inglés) sin importar cuál sea el punto inicial x . Para el caso general el resultado que se tiene es la convergencia en promedio de K^n a π para cualquier distribución inicial μ .

Teorema 2.8 Si (X_n) es Harris positiva y aperiódica, entonces

$$\lim_{n \rightarrow \infty} \left\| \int K^n(x, y) \mu(dx) - \pi(y) \right\|_{TV} = 0, \quad y \in \mathcal{B}(\mathcal{X}).$$

para toda distribución inicial μ .

Teorema 2.9 Sea (X_n) positiva, recurrente, y aperiódica.

1. Si $\mathbb{E}^\pi(|h(X)|) = \infty$, entonces $\mathbb{E}_x(|h(X_n)|) \rightarrow \infty$ para toda x .
2. Si $\int |h(x)|\pi(dx) < \infty$, entonces

$$\lim_{n \rightarrow \infty} \sup_{|m(x)| \leq |h(x)|} |\mathbb{E}_y(m(X_n)) - \mathbb{E}^\pi(m(X))| = 0$$

en todo conjunto pequeño C tal que

$$\sup_{y \in C} \mathbb{E}_y \left(\sum_{t=0}^{\tau_C-1} h(X_t) \right) < \infty$$

2.7.10. Teoremas Límite

Los resultados mencionados hasta este punto permiten justificar los algoritmos de simulación, sin embargo, no dan información directa sobre la única observación disponible de P_x^n , x_n . Con estos resultados se pueden determinar las propiedades probabilísticas del comportamiento promedio de la cadena en un instante fijo, pero estas propiedades no aportan información para el control de la convergencia de una simulación. Lo que es relevante en este caso son las propiedades de la realización (X_n) de la cadena. Los problemas por los que no es posible aplicar directamente los teoremas límite clásicos a la muestra (X_1, X_2, \dots, X_n) son: el primero, es por la dependencia entre cualesquiera dos observaciones sucesivas, esto es, la propiedad Markoviana, y el segundo, es por la no estacionariedad de la sucesión (ya que la distribución de X_0 puede ser diferente de π). A continuación se da un resultado de convergencia equivalente a la **Ley de los Grandes Números** que es conocido por algunos autores como Teorema de Ergodicidad.

2.7.10.1. Teorema de Ergodicidad

Dadas las observaciones X_1, \dots, X_n de una cadena de Markov, se examina el comportamiento limitante de las sumas parciales

$$S_n(h) = \frac{1}{n} \sum_{i=1}^n h(X_i)$$

cuando $n \rightarrow \infty$.

En principio considérese la definición de función armónica. Como se verá a continuación, las funciones armónicas están estrechamente relacionadas con la Harris - recurrencia de una cadena.

Definición 2.40 Una función medible h es **armónica** para la cadena (X_n) si

$$\mathbb{E}[h(X_{n+1})|x_n] = h(x_n).$$

Estas funciones son invariantes para el núcleo de transición (en el sentido funcional) y caracterizan la recurrencia de Harris de la siguiente manera.

Proposición 2.8 Para una cadena de Markov positiva, si las únicas funciones armónicas acotadas son las funciones constantes, la cadena es Harris - recurrente.

De la **Proposición 2.8** se tiene un resultado que puede ser considerado como una propiedad de continuidad, esto es: por inducción se tiene que una función armónica h que satisface $h(x) = \mathbb{E}_x(h(X_n))$, por el **Teorema 2.9** $h(x)$ es casi seguramente igual a $\mathbb{E}^\pi(h(X))$, lo que significa que es constante casi en cualquier parte. En el caso de una cadena Harris, la **Proposición 2.8** establece que esto implica que $h(x)$ es constante en todas partes.

La **Proposición 2.8** permite asegurar la Harris - recurrencia de la cadena de Markov asociada a algunos algoritmos de MCMC, en particular al algoritmo de MH. Como se mencionó párrafos anteriores las funciones armónicas permiten

caracterizar la Harris - recurrencia, esto es claro considerando que el recíproco de la **Proposición 2.8** es cierto.

Lema 2.2 Para las cadenas de Markov Harris - recurrentes, las funciones constantes son las únicas funciones armónicas acotadas.

Una consecuencia del **Lema 2.2** es que si se tiene una cadena de Markov Harris - recurrente con distribución estacionaria π , y si ocurre que $S_n(h)$ converge μ_0 -casi seguramente a

$$\int_{\mathcal{X}} h(x)\pi(dx),$$

para μ_0 alguna distribución inicial, entonces esta convergencia ocurre para cualquier μ distribución inicial. Más aún, la probabilidad

$$P_x(S_n(h) \rightarrow \mathbb{E}^\pi(h))$$

es armónica, es decir no depende de x . Esto muestra que la Harris - recurrencia significa estabilidad fuerte, ya que la convergencia casi segura se reemplaza por convergencia en todo punto.

El resultado principal de esta sección, está dado por, la ley de los grandes números para las cadenas de Markov que habitualmente se denomina **Teorema Ergódico**, el cual garantiza la convergencia de $S_n(h)$.

Teorema 2.10: Teorema Ergódico Si (X_n) tiene una medida invariante σ -finita π , las siguientes dos afirmaciones son equivalentes:

(i) Si $f, g \in L^1(x)$ con $\int g(x)d\pi(x) \neq 0$, entonces

$$\lim_{n \rightarrow \infty} \frac{S_n(f)}{S_n(g)} = \frac{\int f(x)d\pi(x)}{\int g(x)d\pi(x)}$$

(ii) La cadena de Markov (X_n) es Harris - recurrente.

La demostración lo puede encontrar en Robert and Casella (2004).

Un aspecto importante del teorema anterior es que π no necesita ser una medida de probabilidad y, por tanto, que puede haber algún tipo de estabilidad fuerte incluso si la cadena es nula recurrente. En la configuración de un algoritmo MCMC, este resultado a veces se invoca para justificar el uso de medidas posteriores incorrectas.

Observación 2.3 Un estado que es recurrente, no nulo y aperiódico es **ergódico**. Si todos los estados son ergódicos entonces se dice que la **cadena es ergódica**.

En resumen en su mayor parte, las cadenas de Markov que se encuentran en la configuración de la cadena de Markov Monte Carlo (MCMC) satisfacen una propiedad de estabilidad muy fuerte. De hecho, existe una distribución de probabilidad estacionaria por construcción para esas cadenas; es decir, existe una distribución de probabilidad π tal que si $X_{(t)} \sim \pi$, entonces $X_{(t+1)} \sim \pi$. Es decir, el núcleo y la distribución estacionaria satisfacen la ecuación

$$\int_{\mathcal{X}} K(x, y)\pi(x)dx = \pi(y)$$

La existencia de una distribución estacionaria impone una restricción preliminar a K denominada irreductibilidad, que en la teoría de las cadenas de Markov, indica que el núcleo K permite movimientos libres en todo el espacio - estado, es decir, que, sin importar el valor inicial $X_{(0)}$, la secuencia $\{X_{(t)}\}$ tiene una probabilidad positiva de alcanzar eventualmente cualquier región del espacio de estados. (Una condición suficiente es que $K(x, \cdot) > 0$ en todas partes). La existencia de una distribución estacionaria tiene consecuencias importantes en el comportamiento de la cadena $\{X_{(t)}\}$, una de las cuales es que la mayoría de las cadenas involucradas en los algoritmos MCMC son recurrentes, es decir, regresarán a cualquier conjunto arbitrario no despreciable un número infinito de veces.

En el caso de cadenas recurrentes, la distribución estacionaria también es una distribución limitante en el sentido de que la distribución límite de $X_{(t)}$ es π para casi cualquier valor inicial $X_{(0)}$. Esta propiedad también se llama ergodi-

cidad, y obviamente tiene consecuencias importantes desde el punto de vista de la simulación en el sentido de que, si un núcleo K dado produce una cadena de Markov ergódica con distribución estacionaria π , generar una cadena a partir de este núcleo K eventualmente producirá simulaciones de π .

En particular, para funciones integrables h , el promedio estándar

$$\frac{1}{T} \sum_{t=1}^T h(X_{(t)}) \longrightarrow E_{\pi}[h(X)]$$

lo que significa que la Ley de los grandes números que se encuentra en la base de los métodos de Monte Carlo también se puede aplicar en la configuración de MCMC. (Teorema Ergódico).

Sin embargo, hay un caso en el que la convergencia nunca ocurre, esto se dá, cuando, en un entorno bayesiano, la distribución posterior no es adecuada (Roberts et al. (2001)) ya que la cadena no puede ser recurrente.

Dado que el uso de a priori impropias $\pi(x)$ es bastante común en modelos complejos, existe la posibilidad de que el producto de la verosimilitud \times a priori, $(L(x) \times \pi(x))$, no sea integrable y que este problema no se detecte debido a la complejidad inherente.

En tales casos, las cadenas de Markov se pueden simular junto con el objetivo $(L(x) \times \pi(x))$ pero no pueden converger. En los mejores casos, las cadenas de Markov resultantes mostrarán rápidamente un comportamiento divergente, lo que indica que hay un problema, desafortunadamente, en el peor de los casos, estas cadenas de Markov presentan todos los signos externos de estabilidad y, por lo tanto, no indican la dificultad.

Observación 2.4 Un método de Monte Carlo de cadena de Markov (MCMC) para la simulación de una distribución π es cualquier método que produzca una cadena de Markov ergódica (X_n) cuya distribución estacionaria sea π .

CAPÍTULO 3

METODOLOGÍA DE LA INVESTIGACIÓN

En este capítulo se utiliza el enfoque bayesiano en los modelos heterocedásticos de series de tiempo para la modelación de activos financieros, para este propósito se plantea el desarrollo de la generalización de los modelos de heterocedasticidad condicional autorregresiva generalizada GARCH y su principal generalización multivariada, los modelos de correlación condicional dinámica DCC - GARCH, se describen y analizan sus características, elementos y propiedades teóricas los cuales permiten una correcta comprensión de los resultados arrojados por el algoritmo, Metropolis - Hastings caminata aleatoria, implementado en el software R, paquete bayesDccGarch.

3.1. Tipo y Diseño de la Investigación

La presente tesis reúne las condiciones necesarias para ser denominada como una investigación estadística - matemática. Debido a la necesidad de estimar los parámetros del modelo heterocedástico condicional autorregresivo generalizado con correlación condicional dinámica de series de tiempo financieras.

3.1.1. Modelo GARCH Univariado

Dada la serie de retornos $y = \{y_t, t = 1, \dots, T\}$. El modelo GARCH(p, q) estima la volatilidad de los retornos como:

$$y_t = \epsilon_t \sqrt{h_t}, \epsilon_t \sim D(0, 1) \quad (3.1)$$

$$h_t = \omega + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} \quad (3.2)$$

siendo h_t la variación condicional (no observable) de y_t dada la información previa $I_t = \{y_{t-1}, y_{t-2}, \dots\}$, los errores ϵ_t son independientes e idénticamente distribuidos. $D(0, 1)$ denota una distribución con media cero y varianza 1. Las restricciones de estacionaridad y positividad son $\omega > 0$, $\alpha_i \geq 0, i = 1, \dots, p$, $\beta_j \geq 0, j = 1, \dots, q$ y $(\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j) < 1$.

Considerando la esperanza y la varianza de y_t dado I_t , se obtiene:

$$E[y_t|I_t] = \sqrt{h_t}E[\epsilon_t|I_t] = \sqrt{h_t}E[\epsilon_t] = 0 \quad (3.3)$$

$$Var[y_t|I_t] = E[y_t^2|I_t] - E^2[y_t|I_t] = E[y_t^2|I_t] = h_t Var[\epsilon_t|I_t] = h_t \quad (3.4)$$

Por lo tanto, podemos utilizar la ecuación (3.3) para calcular la esperanza incondicional de y_t

$$E[y_t] = E[E[y_t|I_t]] = E[0] = 0$$

Considerando ahora la ecuación (3.4) para calcular la varianza incondicional, se tiene:

$$Var[y_t] = E[Var[y_t|I_t] + Var[E[y_t|I_t]] = E[h_t] + Var[0] = E[h_t] = E[y_t^2] \quad (3.5)$$

Asumiendo que la serie y sea estacionaria de las ecuaciones (3.2) y (3.5) tenemos que:

$$E[y_t^2] = \frac{\omega}{1 - (\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j)} \quad (3.6)$$

Veamos:

Considerando que la serie y sea estacionaria de segundo orden

$$\begin{aligned} E[y_t^2] &= E[y_{t-1}^2] = Var[y_t] \\ E[y_t^2] &= E[\omega + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}] \\ E[y_t^2] &= E[\omega] + \sum_{i=1}^p \alpha_i E[y_{t-i}^2] + \sum_{j=1}^q \beta_j E[h_{t-j}] \\ E[y_t^2] &= \omega + \sum_{i=1}^p \alpha_i E[y_{t-i}^2] + \sum_{j=1}^q \beta_j E[y_{t-j}^2] \\ E[y_t^2] &= \omega + \sum_{i=1}^p \alpha_i E[y_t^2] + \sum_{j=1}^q \beta_j E[y_t^2] \end{aligned}$$

$$\begin{aligned}
E[y_t^2] - \sum_{i=1}^p \alpha_i E[y_t^2] - \sum_{j=1}^q \beta_j E[y_t^2] &= \omega \\
E[y_t^2] \left\{ 1 - \left(\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \right) \right\} &= \omega \\
E[y_t^2] &= \frac{\omega}{1 - \left(\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \right)}
\end{aligned}$$

De la ecuación (3.6), para que exista la varianza incondicional de y_t es necesario que $(\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j) < 1$, justificando así la restricción de estacionaridad del modelo.

De la expresión (3.1) se tiene que $\epsilon_t = y_t/\sqrt{h_t}$, usando el **Teorema 2.1**, se puede escribir la función de densidad de probabilidad condicional del retorno y_t en función de la función de densidad de probabilidad de ϵ_t . Es decir,

$$p(y_t|I_t) = p_\epsilon(y_t/\sqrt{h_t}) \left| \frac{d\epsilon_t}{dy_t} \right|$$

de donde:

$$p(y_t|I_t) = (h_t)^{-1/2} p_\epsilon(y_t/\sqrt{h_t}) \quad (3.7)$$

siendo p_ϵ la función de la densidad de probabilidad de ϵ_t .

Así, teniendo las distribuciones condicionales para los $y = \{y_t, t = 1, \dots, T\}$ retornos en función de la distribución de los errores ϵ_t , se obtiene la función de verosimilitud del conjunto de parámetros desconocidos $\theta = (\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)^t$

$$L(\theta) = p(\theta|y_1, \dots, y_T) = (h_1)^{-1/2} p_\epsilon(y_1/\sqrt{h_1}) \times \dots \times (h_T)^{-1/2} p_\epsilon(y_T/\sqrt{h_T})$$

Es decir,

$$L(\theta) = p(\theta|y_1, \dots, y_T) = \prod_{t=1}^T (h_t)^{-1/2} p_\epsilon(y_t/\sqrt{h_t})$$

3.1.2. Modelos GARCH Multivariados: DCC - GARCH

Sea $y_t = (y_{t1}, \dots, y_{tm})^t$ un vector de retornos en el instante t para m series temporales. Así, las extensiones multivariadas de los modelos GARCH se pueden

escribir como:

$$y_t = H_t^{1/2} \epsilon_t \quad (3.8)$$

siendo H_t la matriz de covarianzas condicionales y $H_t^{1/2}$ la matriz $m \times m$ definida positiva, obtenida por la descomposición de Cholesky de la matriz H_t .

El vector de los errores ϵ_t tiene orden $m \times 1$ y tiene media y varianza dado por:

$$E(\epsilon_t) = 0$$

$$Var(\epsilon_t) = I_m$$

siendo, I_m la matriz identidad de orden m . Así, se tiene que la media y la varianza del vector y_t condicionado a la información previa hasta el momento t (I_t), son dados por:

$$E[y_t|I_t] = E(H_t^{1/2} \epsilon_t|I_t) = H_t^{1/2} E(\epsilon_t) = 0$$

$$Var[y_t|I_t] = H_t^{1/2} Var(\epsilon_t) (H_t^{1/2})^t = H_t$$

Diversas formas de especificar la matriz H_t fueron propuestas en la literatura. En éste trabajo, se utiliza una forma no lineal de combinar GARCH univariados, es decir los modelos CC - GARCH (Correlación Condicional GARCH). Para una revisión de las diversas formas de modelar la matriz de covarianzas condicional, ver Bauwens et al. (2006).

Bollerslev (1990) propone la primera clase de los modelos CC - GARCH, los modelos CCC - GARCH (Correlación Condicional Constante GARCH), que definen la matriz H_t como:

$$H_t = D_t R D_t$$

siendo

$$D_t = \text{diag}(h_{11,t}^{1/2}, \dots, h_{mm,t}^{1/2}) \quad (3.9)$$

y $h_{ii,t}$ definido como en un GARCH univariado de cualquier orden. Así, si se especifica un GARCH(1,1) para cada varianza condicional de D_t , entonces

$$h_{ii,t} = \omega_i + \alpha_{1,i} y_{i,t-1}^2 + \beta_{1,i} h_{ii,t-1} \quad (3.10)$$

con $\omega_i > 0, \alpha_i \geq 0, \beta_i \geq 0$ y $(\alpha_i + \beta_i) < 1$ para $i = 1, \dots, m$.

La matriz de correlaciones $R = \{\rho_{ij}\}_{j=1, \dots, m}^{i=1, \dots, m}$ es simétrica y definida positiva con $\rho_{ii} = 1$. Obviamente, $\rho_{ij} = \rho_{ji}$, luego, los parámetros adicionales de este modelo serán todos ρ_{ij} con $i > j$, para $i, j = 1, \dots, m$.

Por lo tanto, podemos ver que el número total de parámetros de este modelo es $m(m + 5)/2$, de esta forma, el número de parámetros crece rápidamente de acuerdo con el número de series. Tenga en cuenta que para 3 series ($m = 3$), se tendrán 12 parámetros, mientras que para 5 series ($m = 5$), 25 parámetros.

Seguidamente, fueron propuestos algunos modelos más parsimoniosos entre los que podemos destacar al de Engle (2002) y Tse and Tsui (2002) quienes independientemente propusieron un modelo CC - GARCH que considera que la matriz de correlaciones condicional varía en el tiempo, este modelo se conoce como DCC - GARCH (Correlación Condicional Dinámica GARCH).

Adoptando el enfoque de Engle (2002) la matriz de covariancias se escribe como: $H_t = D_t R_t D_t$ donde, D_t es como en (3.9) y R_t como:

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}$$

donde Q_t son matrices simétricas definidas positivas de orden $m \times m$ dadas por

$$Q_t = (1 - a - b)R + a\mu_{t-1}\mu_{t-1}^t + bQ_{t-1} \quad (3.11)$$

Siendo $\mu_t = D_t^{-1}y_t$, R la matriz de covarianzas incondicionales de μ_t y las restricciones de estacionaridad de los parámetros adicionales son $a > 0, b > 0$ y $(a + b) < 1$.

Además de modelar la correlación condicional con variación en el tiempo, el modelo DCC - GARCH también tiene la ventaja de añadir sólo dos parámetros independientemente del número de series elegidas. Si se considera el modelo GARCH(1,1) para cada varianza condicional en D_t el número de parámetros del

modelo será $3m + 2$. Luego, para 2 series ($m = 2$) se tendrá 8 parámetros y si se consideran 4 series ($m = 4$), 14 parámetros.

Como en los modelos univariados, la función de densidad de probabilidad de los retornos en los modelos CC - GARCH se pueden obtener en función de la densidad de probabilidad elegida para los errores ϵ_t .

Sea θ el conjunto de todos los parámetros del modelo, la distribución conjunta de los retornos se escribe como el producto de las distribuciones condicionales, es decir:

$$p(y_1, \dots, y_T | \theta) = p(y_1 | \theta) p(y_2 | \theta, y_1) \dots p(y_T | \theta, y_1, y_2, \dots, y_{T-1}) \quad (3.12)$$

De la expresión (3.12), se obtienen las distribuciones condicionales de los retornos en función de la distribución de los errores:

$$p(y_t | \theta, y_1, y_2, \dots, y_{T-1}) = p(y_t | I_t) = |H_t|^{-1/2} p_{\epsilon_t}(H_t^{-1/2} y_t) \quad (3.13)$$

para $t = 1, \dots, T$.

Así, teniendo las distribuciones condicionales para los $y = \{y_t, t = 1, \dots, T\}$ retornos en función de la distribución de los errores, obtenemos la función de verosimilitud del conjunto de parámetros θ que son desconocidos

$$L(\theta) = p(\theta | y_1, \dots, y_T) = |H_1|^{-1/2} p_{\epsilon_1}(H_1^{-1/2} y_1) \times \dots \times |H_T|^{-1/2} p_{\epsilon_T}(H_T^{-1/2} y_T)$$

Es decir,

$$L(\theta) = p(\theta | y_1, \dots, y_T) = \prod_{t=1}^T |H_t|^{-1/2} p_{\epsilon_t}(H_t^{-1/2} y_t)$$

Para nuestro enfoque Bayesiano se necesita determinar la distribución a priori de los parámetros del modelo.

3.2. Estudio de la Distribución a Priori

La estimación de los parámetros en este trabajo se hizo mediante el enfoque Bayesiano teniendo como base el modelo GARCH para ello fueron necesarios las

a priori de los parámetros del modelo. En este estudio, para la estimación de los parámetros del modelo GARCH(1,1) se utilizaron las distribuciones a priori propuestas en David (2006). Estas son **normales truncadas en el espacio paramétrico de cada uno de los parámetros**, también se asume independencia a priori entre los parámetros, de esta forma se obtienen:

$$\begin{aligned}\omega &\sim N(\mu_\omega, \sigma_\omega^2)I_{(\omega>0)}, \\ \alpha_1 &\sim N(\mu_\alpha, \sigma_\alpha^2)I_{(0<\alpha<1)}, \\ \beta_1 &\sim N(\mu_\beta, \sigma_\beta^2)I_{(0<\beta<1)}\end{aligned}$$

siendo

$$\mu_\omega, \mu_\alpha, \mu_\beta, \sigma_\alpha^2, \sigma_\beta^2$$

hiperparámetros.

En el caso, de utilizar una distribución asimétrica, como las presentadas en la sección (2.3), es necesario estimar el parámetro de asimetría, en este caso Fernández and Steel (1998) proponen usar una distribución a priori Gamma(a, b) para γ^2 .

La idea es elegir los valores de a y b de modo que $E(\gamma) = 1$.

$$\begin{aligned}1 &= E_{\gamma^2}(\gamma) = \frac{b^a}{\Gamma(a)} \int_0^\infty \gamma(\gamma^2)^{a-1} e^{-b\gamma^2} d\gamma^2 \\ &= \frac{b^a}{\Gamma(a)} \int_0^\infty (\gamma^2)^{(a+0.5)-1} e^{-b\gamma^2} d\gamma^2\end{aligned}$$

si

$$b\gamma^2 = y$$

$$bd\gamma^2 = dy$$

los límites de integración no cambian, de donde:

$$\begin{aligned}1 &= \frac{b^a}{\Gamma(a)} \int_0^\infty \left(\frac{y}{b}\right)^{(a+0.5)-1} e^{-y} \frac{dy}{b} \\ 1 &= \frac{b^a}{\Gamma(a)} \frac{\Gamma(a + 0.5)}{b^{a+0.5}}\end{aligned}$$

por lo que

$$b = \left[\frac{\Gamma(a + \frac{1}{2})}{\Gamma(a)} \right]^2$$

del resultado anterior se elige el valor de a controlando la varianza a priori y la probabilidad a priori de $\gamma \in (0, 1)$, fijando $a = 1/2$ lleva a $V(\gamma) = \pi/2 - 1 \approx 0.57$ y $P(0 < \gamma < 1) \approx 0.58$, la cual es una elección razonable.

Además, esta particular opción es equivalente a especificar:

$$\gamma \sim N(0, 0.64^{-1})I_{(\gamma>0)}. \quad (3.14)$$

Pues si se considera π_γ y π_{γ^2} como las densidades a priori de γ y γ^2 , respectivamente, se tiene:

$$\begin{aligned} \pi_\gamma(\gamma|a = 0.5, b = 0.32) &= \pi_{\gamma^2}(\gamma^2|a = 0.5, b = 0.32) \times \left| \frac{d\gamma^2}{d\gamma} \right| \\ &= \pi_{\gamma^2}(\gamma^2|a = 0.5, b = 0.32) \times 2\gamma \\ &\propto \gamma(\gamma^2)^{1/2-1} e^{(-0.32\gamma^2)} \\ &\propto e^{(-\frac{0.64}{2}\gamma^2)} \\ &\propto e^{(\frac{-\gamma^2}{2(0.64)^{-1}})}, \gamma \in (0, \infty) \end{aligned}$$

Sin embargo, cuando se utiliza la distribución $SST(0, 1, \gamma, v)$, presentada en (2.17) o $SSGED(0, 1, \gamma, k)$ presentada el (2.22) se debe estimar el parámetro de cola v o k , respectivamente.

En estos casos, se utilizará

$$v \sim N(\mu_v, \sigma_v^2)I_{(v>2)}$$

y

$$k \sim N(\mu_k, \sigma_k^2)I_{(k>0)}$$

siendo

$$\mu_v, \mu_k, \mu_v \text{ y } \sigma_k^2$$

hiperparámetros.

Tenga en cuenta que $N(\mu, \sigma^2)I_{(a<x<b)}$, con $a < b$ representa la función de densidad de probabilidad

$$\frac{1}{\phi(b) - \phi(a)} \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}, x \in (a, b)$$

siendo $\phi(\cdot)$ la función de densidad acumulada de la distribución normal con media μ y varianza σ^2 .

Es importante recordar que utilizar la distribución de probabilidad Normal para las a prioris, facilita la inserción de información en una determinada región de interés a través de los parámetros μ y σ^2 de la distribución normal $N(\mu, \sigma^2)$, aunque en el caso del truncamiento, estos hiperparámetros no representan la media y la varianza, pero aún así controlan la región con la medida de probabilidad más alta.

Para el caso **multivariado** se extiende el enfoque presentado para los modelos univariados de los parámetros de volatilidad (ω, α_1 y β_1) y para el parámetro de asimetría (γ), presentada en la (sección 2.3).

Para ello, se asume independencia a priori entre los parámetros de los modelos. De esta forma se obtiene:

$$\begin{aligned}\omega_i &\sim N(\mu_{\omega_i}, \sigma_{\omega_i}^2)I_{(\omega_i > 0)}, \\ \alpha_{1,i} &\sim N(\mu_{\alpha_i}, \sigma_{\alpha_i}^2)I_{(0 < \alpha_i < 1)}, \\ \beta_{1,i} &\sim N(\mu_{\beta_i}, \sigma_{\beta_i}^2)I_{(0 < \beta_i < 1)}, \\ \gamma_i &\sim N(0, 0 \cdot 64^{-1})I_{\gamma_i > 0},\end{aligned}$$

para $i = 1, \dots, m$.

Las a priori $\gamma_i \sim N(0, 0 \cdot 64^{-1})I_{\gamma_i > 0}$ son una generalización de la ecuación (3.14) propuesta por Fernández and Steel (1998) y utilizada en David (2006).

Aún falta determinar las a prioris para los parámetros de correlación a y b , y en el caso de asumir la distribución de los errores del modelo $SST(0, I_m, \gamma, v)$ o $SSGED(0, I_m, \gamma, k)$, se debe determinar las a prioris para los parámetros de peso en las colas.

Siguiendo el mismo enfoque anterior se utilizan las a priori:

$$\begin{aligned}
 a &\sim N(\mu_a, \sigma_a^2)I_{(0 < a < 1)}, \\
 b &\sim N(\mu_b, \sigma_b^2)I_{(0 < b < 1)}, \\
 v &\sim N(\mu_v, \sigma_v^2)I_{(v > 2)}, \text{ si } \epsilon_t \sim SST(0, I_m, \gamma, v), \\
 k &\sim N(\mu_k, \sigma_k^2)I_{(k > 0)}, \text{ si } \epsilon_t \sim SSGED(0, I_m, \gamma, k),
 \end{aligned}$$

Vale recordar, que para $m = 1$, el modelo heterocedástico DCC - GARCH resulta ser el modelo GARCH univariado. Pues en este caso, se tendría $D_t = \{h_{11,t}^{1/2}\}$ y $R_t = \{1\}$, con lo cual $H_t = \{h_{11,t}\}$.

3.3. Algoritmo: Metropolis - Hastings (MH)

El principio de funcionamiento de los métodos de Monte Carlo cadena de Markov es bastante sencillo de describir. Dada una densidad objetivo π , se construye un núcleo de Markov K con distribución estacionaria π y luego se genera una cadena de Markov $(X_{(t)})$ usando este núcleo, de modo que la distribución límite de $(X_{(t)})$ sea π y las integrales puedan aproximarse según el **Teorema Ergódico**. Por tanto, la dificultad debería estar en la construcción de un núcleo K que esté asociado con una densidad π arbitraria. Pero, esta dificultad es relativa pues, existen métodos para determinar tales núcleos que son universales en el sentido de que son teóricamente válidos para cualquier densidad π **el algoritmo Metropolis - Hastings** es un ejemplo de estos métodos.

Dada la densidad objetivo π , se asocia con una densidad condicional de trabajo $q(y|x)$ que, en la práctica, es fácil de simular. Además, q puede ser casi arbitrario porque los únicos requisitos teóricos son que la relación $\pi(y)/q(y|x)$ se conozca hasta una constante independiente de x y que $q(\cdot|x)$ tenga suficiente dispersión para conducir a una exploración de todo el soporte de π . Una vez más, se destaca una importante característica del algoritmo Metropolis - Hastings en la que, para

cada q dado, se puede construir un núcleo de Metropolis - Hastings tal que π sea su distribución estacionaria.

El **algoritmo Metropolis - Hastings (MH)** asociado con la densidad objetivo π y la densidad condicional q produce una cadena de Markov $(X_{(t)})$ a través del siguiente núcleo de transición:

Dado $(x_{(t)})$,

1. Genere $Y_t \sim q(y|x_{(t)})$.
2. Tomar

$$X_{(t+1)} = \begin{cases} Y_t; & \text{con probabilidad } p(x_{(t)}, Y_t), \\ x_{(t)}; & \text{con probabilidad } 1 - p(x_{(t)}, Y_t), \end{cases}$$

donde

$$p(x, y) = \min\left\{\frac{\pi(y) q(x|y)}{\pi(x) q(y|x)}, 1\right\}.$$

En la práctica, el rendimiento del algoritmo obviamente dependerá en gran medida de esta elección de q . Este algoritmo siempre acepta valores y_t ; tal que la relación $\pi(y_t)/q(y_t|x_{(t)})$ aumenta, en comparación con el valor anterior $\pi(x_{(t)})/q(x_{(t)}|y_t)$. Es sólo en el caso simétrico que la aceptación es impulsada por la razón objetivo $\pi(y_t)/q(x_{(t)})$. Una característica importante del algoritmo anterior es que puede aceptar valores y_t ; de manera que la razón se reduce, de manera similar a los métodos de optimización estocástica. Al igual que el método Aceptar - Rechazar el algoritmo Metropolis - Hastings depende sólo de las relaciones $\pi(y_t)/q(x_{(t)})$ y $q(x_{(t)}|y_t)/q(y_t|x_{(t)})$.

3.3.1. Algoritmo MH: Caminata Aleatoria

En lugar de crear una propuesta desde cero o sugerir una aproximación no paramétrica basada en una ejecución preliminar, porque es poco probable que funcione para dimensiones moderadas o altas, es más realista recopilar información

sobre el objetivo por pasos, es decir, explorando la vecindad del valor actual de la cadena. Si el mecanismo de exploración tiene suficiente fuerza para llegar hasta los límites del soporte del objetivo π , el método finalmente descubrirá la complejidad del objetivo.

Un enfoque más natural para la construcción práctica de una propuesta de Metropolis - Hastings es, entonces, tener en cuenta el valor previamente simulado para generar el siguiente valor; es decir, considerar una exploración local de la vecindad del valor actual de la cadena de Markov. La implementación de esta idea es simular Y_t de acuerdo con $Y_t = X_{(t)} + \epsilon_t$, donde ϵ_t es una perturbación aleatoria con distribución g independiente de $X_{(t)}$, por ejemplo, una distribución uniforme o una distribución normal, lo que significa que $Y_t \sim U(X_{(t)} - \delta, X_{(t)} + \delta)$ o $Y_t \sim N(X_{(t)}, \sigma^2)$ en entornos unidimensionales.

En términos del algoritmo general de Metropolis - Hastings, la densidad propuesta $q(y|x)$ es ahora de la forma $g(y - x)$, la cadena de Markov asociada con q es una caminata aleatoria si la densidad g es simétrica alrededor de cero; es decir, satisface $g(-t) = g(t)$. Pero, debido al paso adicional de aceptación de Metropolis - Hastings, la cadena Metropolis Hastings Markov $X_{(t)}$ no es un paseo aleatorio. Este enfoque conduce al siguiente algoritmo de Metropolis - Hastings, que también resulta ser el original propuesto por Metropolis et al. (1953).

Algoritmo: Metropolis - Hastings: Caminata Aleatoria

Dado $(x_{(t)})$,

1. Genere $Y_t \sim g(y - x_{(t)})$.
2. Tomar

$$X_{(t+1)} = \begin{cases} Y_t; & \text{con probabilidad : } \min\{1, \pi(Y_t)/\pi(x_{(t)})\}, \\ x_{(t)}; & \text{caso contrario} \end{cases}$$

3.3.1.1. Núcleo de Transición

El núcleo de transición $K(x, y)$ de la cadena de MH esta dado por

$$K(x, y) = \begin{cases} q(y|x)p(x, y) & , \text{ si } x \neq y \\ q(y|x)p(x, y) + (1 - r(x)) & , \text{ si } y = x \end{cases}$$

con

$$r(x) = \int q(x, y)p(x, y)dy.$$

Definición 3.1 Una cadena de Markov con núcleo de transición K satisface la ecuación de balance detallado si existe una función π tal que se cumple

$$K(x, y)\pi(x) = K(y, x)\pi(y)$$

para todo (x, y) .

Además se puede ver que la cadena de Markov MH satisface la ecuación de balance detallado con π (la densidad objetivo).

Primero, para cualquiera que sea la relación entre x y y ($y \neq x$ o $y = x$) se cumple que si $p(x, y) < 1$ entonces necesariamente $p(y, x) = 1$, de donde se tiene

$$p(x, y)q(y|x)\pi(x) = p(y, x)q(x|y)\pi(y). \quad (3.15)$$

Ahora, considerando el caso $y = x$ y la relación (3.15), se sigue que

$$(1 - r(x))\pi(x) = (1 - r(y))\pi(y). \quad (3.16)$$

Con (3.15) y con la suma de (3.15) y (3.16) se establece el cumplimiento de la ecuación de balance detallado para $y \neq x$ y $y = x$, respectivamente.

Teorema 3.1 Suponga que una cadena de Markov con función de transición K satisface la ecuación de balance detallado con π , una función de densidad de probabilidad, entonces la densidad π es la densidad invariante de la cadena.

Demostración: Usando la condición de balance detallado para todo conjunto medible B ,

$$\begin{aligned}\int_y K(y, B)\pi(y)dy &= \int_y \int_B K(y, x)\pi(y)dx dy \\ &= \int_y \int_B K(x, y)\pi(x)dx dy \\ &= \int_B \pi(x)dx\end{aligned}$$

ya que $\int K(x, y)dy = 1$. Esto prueba que π es la densidad invariante. Con base en lo anterior se muestra el siguiente resultado para la cadena MH.

Dado ξ soporte de π , tenemos el siguiente teorema.

Teorema 3.2 Sea $(X_{(t)})$ la cadena producida por MH. Para cada distribución condicional q cuyo soporte incluye ξ ,

- (a) el núcleo de la cadena satisface la condición de balance detallado con π .
- (b) π es una distribución estacionaria de la cadena.

Por tanto, la estacionariedad de π se establece para casi cualquier distribución condicional q , hecho que indica la universalidad de los algoritmos Metropolis - Hastings.

3.3.2. Propiedades de Convergencia

Irreducibilidad y aperiodicidad Roberts and Smith (1994) mostraron que si q es π -irreducible y aperiódica, y si $p(x, y) > 0$, para todo x y todo y , entonces la cadena MH es π -irreducible y aperiódica.

Recurrencia positiva La condición de recurrencia se tiene por el razonamiento siguiente: si la cadena es π -irreducible y aperiódica, entonces como π es su medida invariante, la cadena es positiva, lo cual implica que la cadena sea recurrente.

Ahora sí se puede formular un resultado muy importante para la cadena MH. La prueba lo puede ver en Robert and Casella (2004).

Lema 3.1 Si la cadena Metropolis - Hastings $(X_{(t)})$ es π -irreducible, es Harris - recurrente.

Definición 3.2 Se define $L^p(\mu)$ como el espacio de las funciones $\pi : X \rightarrow \mathbb{C}$ medibles y tales que $|\pi|^p$ es integrable. Si $\pi \in L^p(\mu)$ se define

$$\|\pi\|_p = \left(\int |\pi|^p d\mu \right)^{1/p}$$

$L^p(\mu)$ es un espacio vectorial donde $\|\cdot\|_p$ es una seminorma. Con los resultados anteriores se tiene el siguiente resultado de convergencia tomado de Robert and Casella (2004).

Teorema 3.3 Suponga que la cadena Metropolis - Hastings Markov $(X_{(t)})$ es π - irreducible.

(i) Si $h \in L^1(\pi)$, entonces

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T h(X_{(t)}) = \int h(x)\pi(x)d(x) \quad a.e \pi.$$

(ii) Si, además, $(X_{(t)})$ es aperiódico, entonces

$$\lim_{n \rightarrow \infty} \left\| \int K^n(x, \cdot)\mu(dx) - \pi \right\|_{TV} = 0$$

para cada distribución inicial μ , donde $K^n(x, \cdot)$ denota el núcleo para n transiciones, como en (2.24), $\|\cdot\|_{TV}$ norma de variación total y a.e π : casi en cualquier parte de π .

La primera parte del **Teorema 3.3** asegura que es posible utilizar una cadena de Markov de forma similar que una muestra *i.i.d* cuando se quiere aproximar $\mathbb{E}_f(h(x))$.

La segunda parte del **Teorema 3.3** establece la convergencia asintótica de la distribución de (X_t) a π sin importar el valor inicial de la cadena, esto es, la ergodicidad de la cadena de MH.

3.3.3. Tasas de Aceptación

Se necesita adoptar un criterio práctico que permita la comparación de núcleos de propuesta en situaciones en las que (casi) no se sabe nada sobre π . Uno de estos criterios es la tasa de aceptación del algoritmo de Metropolis - Hastings correspondiente, ya que se puede calcular fácilmente cuando se ejecuta este algoritmo mediante la frecuencia empírica de aceptación. La optimización de la tasa de aceptación no necesariamente brindará el mejor algoritmo en términos de mezcla y convergencia.

Si bien los algoritmos independientes de Metropolis - Hastings se pueden optimizar o al menos comparar a través de su tasa de aceptación, porque esto reduce el número de réplicas en la cadena $(X_{(t)})$ y por lo tanto el nivel de correlación en la cadena, esto no es cierto para otros tipos de algoritmos Metropolis - Hastings, como la versión de caminata aleatoria.

La versión de caminata aleatoria del algoritmo Metropolis - Hastings, de hecho requiere un enfoque diferente para las tasas de aceptación, dada la dependencia de la distribución candidata en el estado actual de la cadena. De hecho, una alta tasa de aceptación no indica necesariamente que el algoritmo se esté comportando satisfactoriamente, ya que puede corresponder al hecho de que la cadena se mueve demasiado lenta en la superficie de π . Cuando $x_{(t)}$ e y_t están cerca, en el sentido de que $\pi(x_{(t)})$ y $\pi(y_t)$ son aproximadamente iguales, el algoritmo Metropolis - Hastings Caminata Aleatoria conduce a la aceptación de y_t con probabilidad

$$\min\left(\frac{\pi(y_t)}{\pi(x_t)}, 1\right) \simeq 1$$

Por lo tanto, una tasa de aceptación alta puede indicar un patrón de convergencia deficiente, ya que los movimientos en el soporte de π son más limitados. Evidentemente, este no es siempre el caso.

Por ejemplo, cuando π es casi plano (constante), las tasas de aceptación altas no son indicativas de ningún comportamiento incorrecto. Pero, a menos que π sea completamente plano (es decir, corresponda a un objetivo uniforme), hay partes del dominio a explorar donde π toma valores más pequeños y, por lo tanto, las probabilidades de aceptación deberían ser pequeñas. Una alta tasa de aceptación indica que esas partes del dominio no se exploran con frecuencia (o no se exploran en absoluto) mediante el algoritmo Metropolis - Hastings.

Por el contrario, si la tasa de aceptación promedio es baja, los valores sucesivos de $\pi(y_t)$ a menudo son pequeños en comparación con $\pi(x_{(t)})$, que corresponde al escenario en el que la caminata aleatoria se mueve rápidamente en la superficie de π y a menudo alcanza los “límites” del soporte de π (o al menos cuando la caminata aleatoria explora regiones con baja probabilidad bajo π). De nuevo, una tasa de aceptación baja no significa que la cadena explore todo el soporte de π . Incluso con una tasa de aceptación pequeña, puede pasar por alto un estado importante pero aislado de π . No obstante, una tasa de aceptación baja es un problema menor, excepto desde el punto de vista del tiempo de cálculo, porque indica explícitamente que se necesita un mayor número de simulaciones.

Tenga en cuenta que la tasa de aceptación óptima es la que nos da la mínima correlación entre las muestras de la cadena.

La cuestión es entonces decidir sobre una tasa de aceptación precisa con la cual calibrar los algoritmos Metropolis - Hastings Caminata Aleatoria para evitar tasas de aceptación “demasiado altas” y “demasiado bajas”. Bajo ciertas circunstancias, algunos autores han mostrado que, en el caso multidimensional, es óptimo escoger la tasa de aceptación igual a 0,234 Roberts et al. (2001)

En este trabajo, para la simulación de los parámetros del modelo heterocedástico condicionalmente autorregresivo generalizado multivariado con correlación condicional dinámica DCC - GARCH(1,1) se consideraron las medias de las distribuciones a posteriori de cada parámetro del modelo, considerando dos formas del algoritmo MH.

En el primer algoritmo se consideró sólo un bloque conteniendo todos los parámetros y en el segundo algoritmo se consideró que la actualización de la cadena sea con un parámetro por bloques.

Para poder simular en la recta, se consideraron transformaciones en todos los parámetros, teniendo al conjunto de los números reales como el espacio paramétrico de todos los parámetros. La distribución candidata fué la distribución Normal centrada en el último valor de la cadena, obteniéndose algoritmos Metrópolis - Hastings Caminata Alatoria.

Así, para simular la distribución a posteriori de los parámetros se consideraron las transformaciones paramétricas como en la Tabla 3.1.

La metodología MCMC requiere que la cadena sea:

- (a) Homogénea, es decir, las probabilidades de transición de un estado a otro es invariante;
- (b) Irreducible, es decir, cada estado puede alcanzarse desde cualquier otro estado en un número finito de iteraciones;
- (c) Aperiódica, es decir, no hay estados absorbentes.

Tabla 3.1: Transformaciones adoptadas para cada parámetro para que sean mapeados en la recta real.

Transformación (ϕ)	Inversa($\phi^{-1} = \theta$)	Distribución (dist. de los errores)
$\phi_1 = \log(v - 2)$	$\nu = \exp(\phi_1) + 2$	t -Student asimétrica estandarizada
$\phi_1 = \log(k)$	$k = \exp(\phi_1)$	GED asimétrica estandarizada
$\phi_{4(i-1)+2} = \log(\gamma_i)$	$\gamma_i = \exp(\phi_{4(i-1)+2})$	Parámetros de series temporales $i, i = 1, \dots, m$
$\phi_{4(i-1)+3} = \log(\omega_i)$	$\omega_i = \exp(\phi_{4(i-1)+3})$	
$\phi_{4(i-1)+4} = \log\left(\frac{\alpha_i}{1-\alpha_i}\right)$	$\alpha_i = \frac{\exp(\phi_{4(i-1)+4})}{1+\exp(\phi_{4(i-1)+4})}$	
$\phi_{4(i-1)+5} = \log\left(\frac{\beta_i}{1-\beta_i}\right)$	$\beta_i = \frac{\exp(\phi_{4(i-1)+5})}{1+\exp(\phi_{4(i-1)+5})}$	
$\phi_{4m+1} = \log\left(\frac{a}{1-a}\right)$	$a = \frac{\exp(\phi_{4m+1})}{1+\exp(\phi_{4m+1})}$	Para
$\phi_{4m+2} = \log\left(\frac{b}{1-b}\right)$	$b = \frac{\exp(\phi_{4m+2})}{1+\exp(\phi_{4m+2})}$	$m \geq 2$

Sea $\phi = (\phi_1, \dots, \phi_{np})^t$ el vector de parámetros transformados y ϕ^{-1} la transformación inversa a la escala original ($\phi^{-1} = \theta$), la función densidad posterior de ϕ es dada por:

$$\pi_\phi(\phi|y) = \pi(\phi^{-1}|y)J(\phi),$$

donde $J(\phi)$ es el jacobiano de la transformación:

si $m = 1$

$$J(\phi) = \begin{vmatrix} \frac{\partial \theta_1}{\partial \phi_1} & \frac{\partial \theta_1}{\partial \phi_2} & \dots & \frac{\partial \theta_1}{\partial \phi_5} \\ \frac{\partial \theta_2}{\partial \phi_1} & \frac{\partial \theta_2}{\partial \phi_2} & \dots & \frac{\partial \theta_2}{\partial \phi_5} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \theta_5}{\partial \phi_1} & \frac{\partial \theta_5}{\partial \phi_2} & \dots & \frac{\partial \theta_5}{\partial \phi_{np}} \end{vmatrix} = \frac{e^{\phi_1 + \dots + \phi_5}}{[(1 + e^{\phi_4})(1 + e^{\phi_5})]^2}$$

si $m \geq 2$

$$\begin{vmatrix} \frac{\partial \theta_1}{\partial \phi_1} & \frac{\partial \theta_1}{\partial \phi_2} & \dots & \frac{\partial \theta_1}{\partial \phi_{np}} \\ \frac{\partial \theta_2}{\partial \phi_1} & \frac{\partial \theta_2}{\partial \phi_2} & \dots & \frac{\partial \theta_2}{\partial \phi_{np}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \theta_{np}}{\partial \phi_1} & \frac{\partial \theta_{np}}{\partial \phi_2} & \dots & \frac{\partial \theta_{np}}{\partial \phi_{np}} \end{vmatrix} = \frac{\prod_{i=1}^{np} e^{\phi_i}}{[(1 + e^{\phi_{4m+1}})(1 + e^{\phi_{4m+2}}) \prod_{i=1}^m (1 + e^{\phi_{4(i-1)+4}})(1 + e^{\phi_{4(i-1)+5}})]^2}$$

A continuación se presentan dos algoritmos implementados en el software R, paquete bayesDccGarch propuesto por Fioruci et al. (2014b): algoritmo MH 1-

bloque y algoritmo MH np -Bloques, donde el espacio paramétrico es \mathbb{R}^{np} , y $np = 4m + 3$ es el número de parámetros del modelo.

3.3.4. MH 1-Bloque

1. Establecer valores iniciales para $\phi_i^{(0)}$, para $i = 1, \dots, np$ y hacer: $j = 0$.
2. Dejar $\phi^{(j)} = (\phi_1^{(j)}, \dots, \phi_{np}^{(j)})^t$, muestra $x \sim N_{np}(0, I_{np})$ y $u \sim U(0, 1)$.
Hacer $\phi^t = \phi^{(j)} + \lambda \sum^{1/2} x$;
3. Si $u \leq p_\phi(\phi^t, \phi^{(j)})$, entonces hacer $\phi^{(j+1)} = \phi^t$, caso contrario $\phi^{(j+1)} = \phi^{(j)}$
4. Hacer $j = j + 1$ y regresar a 2.

Como la distribución generadora de ϕ^t elegida, es la normal multivariada con media $\phi^{(j)}$ y la densidad en cuestión es la a posteriori de ϕ , la probabilidad de aceptación es dado por:

$$p_\phi(\phi^t, \phi^{(j)}) = \min\left\{1, \frac{\pi(\phi^t|y)}{\pi(\phi^{(j)}|y)}\right\}$$

siendo,

$$\pi(\phi|y) = L(\phi) \prod_{i=1}^{np} p(\phi_i).$$

y $\sum^{1/2}$ es la descomposición de Cholesky de la matriz de varianza - covarianza de ϕ . El coeficiente $\lambda > 0$ es útil para ajustar la tasa de aceptación a valores adecuados (entre 20 % y 50 % para el paquete bayesDccGarch). Observamos que, como la estimación de \sum se realiza a priori, encontrar una buena aproximación es lo más difícil involucrado en el uso del algoritmo MH 1-Bloque.

En el paquete bayesDccGarch hay dos estrategias implementadas para esta propuesta.

El primero usa la siguiente aproximación,

$$\widehat{\Sigma}^{-1/2} = H^{-1/2}$$

donde H es la matriz Hessiana de la distribución de densidad a posteriori de ϕ . Sin embargo, pueden surgir problemas computacionales al calcular esta aproximación, en cuyo caso se adopta la segunda estrategia que simula un parámetro por bloque para construir una muestra piloto. A continuación, se calcula una matriz de varianza y covarianza de muestra a partir de la salida de este algoritmo. Este algoritmo se describe a continuación.

3.3.5. MH np -Bloques

1. Establecer valores iniciales para $\phi_i^{(0)}$, para $i = 1, \dots, np$ y hacer: $j = 0$.

2. Para $i = 1$ hasta $i = np$ hacer: muestra $x \sim N(0, 1)$ y $u \sim U(0, 1)$.

Hacer $\phi_i^t = \phi_i^{(j)} + \sigma_i x$

Si $u \leq p_{\phi_i}(\phi_i^t, \phi_i^{(j)})$, entonces hacer $\phi_i^{(j+1)} = \phi_i^t$.

Caso contrario $\phi_i^{(j+1)} = \phi_i^{(j)}$

3. Hacer $j = j + 1$ y regresar a 2.

Siendo la probabilidad de aceptación $p_{\phi_i}(\phi_i^t, \phi_i^{(j)})$ dado por

$$p_{\phi_i}(\phi_i^t, \phi_i^{(j)}) = \min\left\{1, \frac{\pi(\phi_i^t | \phi_i^{(j)}, y)}{\pi(\phi_i^{(j)} | \phi_i^{(j)}, y)}\right\},$$

donde

$$\pi(\phi_i^{(t)} | \phi_i^{(j)}, y) \propto L(\phi_1^{(j+1)}, \dots, \phi_{i-1}^{(j+1)}, \phi_i^{(t)}, \phi_{i+1}^{(j)}, \dots, \phi_{np}^{(j)}) \pi(\phi_i^{(t)})$$

El esquema MH np - Bloques también implica la estimación previa de la desviación estándar de los parámetros $\sigma_1; \dots; \sigma_{np}$ para encontrar una tasa de aceptación adecuada para cada parámetro. Aunque es más costoso computacionalmente, esto suele ser más fácil de aproximar que Σ . Las estrategias para elegir el algoritmo

de simulación implementados en el paquete bayesDccGarch se esquematizan en la Figura 5.2 nótese que se favorece el algoritmo MH 1-Bloque, ya que es considerablemente más rápido y generalmente produce una mejor mezcla de cadenas de Markov.

Para más detalles de la implementación del paquete bayesDccGarch ver Fioruci et al. (2014b).

3.4. Población y Muestra

3.4.1. Población

La población está determinada por el conjunto de datos de la bolsa de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225) y París (CAC40), además del conjunto de datos de cierre diario del índice General de la Bolsa de Valores de Lima (BVL).

3.4.2. Muestra

El rango temporal de muestra que nos sirvió de insumo para la estimación de los parámetros del modelo heterocedástico usando metodología bayesiana fué de 1263 datos, analizados en el periodo del 01 de abril del 2015 al 31 de enero del 2020, extraídos de la base de datos de las bolsas de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225), París (CAC40) y Lima (BVL).

3.5. Técnicas e instrumentos de recolección de datos

Para la presente investigación se utilizó la técnica de revisión documental de páginas web, y para la recolección de datos se utilizó una ficha simple de base de datos estructurada de excel, para la cual se consideraron datos diarios de la serie estadística de las bolsas de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225), París (CAC40) y Lima (BVL), tomados de la web (Investing.com (2020a), Investing.com (2020b), Investing.com (2020d) y Investing.com (2020c)).

3.6. Técnicas para el procesamiento de información

3.6.1. Análisis Estadístico

Con el objetivo de poder entender la dinámica de los retornos de las bolsas de valores, se realizaron análisis estadísticos y gráficos de dichas series. Todos los cálculos del análisis descriptivo fueron hechos en el programa estadístico R studio utilizando el paquete bayesDccGarch, es importante mencionar que debido a observaciones missing se realizó un proceso de imputación de datos por el método de interpolación lineal con el objetivo de no distorsionar la verdadera dinámica del comportamiento de las series.

A continuación se presentan los datos de las bolsas de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225), París (CAC40) y de Lima (BVL). Analizados en el periodo del 01 de abril del 2015 al 31 de enero

del 2020 totalizando 1263 días observados.

En la Figura 5.1 se muestran los gráficos de las series en la primera columna y los rendimientos respectivos en la segunda columna. Los rendimientos considerados se multiplicaron por 100.

Es decir, se modela la serie multivariante de rendimientos $y_{it} = 100\log\left(\frac{I_{it}}{I_{i,t-1}}\right)$. Donde I_t es el valor del i - ésimo índice en el tiempo t .

En el estudio de series de tiempo, la dependencia entre observaciones consecutivas generalmente se mide mediante autocovarianza o autocorrelación Chatfield (2013). Las contrapartes de la muestra de dichas cantidades son las autovarianzas y autocorrelaciones de la muestra. Por lo tanto, analizando las series de retornos, generalmente presentan el siguiente comportamiento peculiar:

1. Muestran pequeñas autocorrelaciones de muestra, no significativas en el primer rezago.
2. Una autocorrelación de muestra significativa en los cuadrados de los retornos.
3. Las autocorrelaciones de los retornos logarítmicos absolutos para la muestra son diferentes de cero para una gran cantidad de rezagos.

En la Tabla 3.2, se presentan las estadísticas descriptivas de los rendimientos de las series DAX, NIKKEI225, CAC40 y BVL. Donde se puede ver que los retornos tienen exceso de curtosis de muestra por ser mayores a 3 (valor de referencia de la curtosis de la distribución normal), lo que indica que la distribución de la serie tiene colas pesadas. Además, presentan asimetría y media cercana a cero.

Estas características se ejemplifican en la Figura 3.1 considerando las series de retornos de los índices DAX, NIKKEI225, CAC40 y BVL. La Figura 3.1 primera columna muestra que hay pequeñas autocorrelaciones en los retornos y en la

Tabla 3.2: Estadísticas de las series de retornos de los índices DAX, NIKKEI225, CAC40 y BVL

Estadísticas	DAX	NIKKEI225	CAC40	BVL
média	0.0028	0.0065	0.00491	0.01598
máximo	2.107219	3.225145	1.763399	3.58797
mínimo	-3.069277	-3.584203	-3.641284	-2.34341
varianza	0.2229739	0.2545997	0.2050235	0.1256323
curtosis	5.914458	10.11932	8.381533	14.58706
asimetría	-0.42919	-0.2501683	-0.5908812	0.6491518
número de observaciones	1263	1263	1263	1263

segunda columna muestra que las autocorrelaciones de los retornos al cuadrado son diferentes de cero para varios rezagos.

En la Figura 3.2 (primera columna) podemos ver un pico alrededor de cero en todas las series, eso muestra que las distribuciones son leptocúrticas, pero el grosor de las colas es difícil de ver. La existencia de colas pesadas en estas series se pueden ver en la Figura 3.2 (segunda columna), en la que se presenta el gráfico cuantil-cuantil normal. Observe que los puntos caen a lo largo de la línea en el medio del gráfico, pero se curvan en los extremos para todas las series, lo cual es un indicador de que las series de retornos tienen valores más extremos de lo que se esperaría si realmente vinieran de una distribución Normal. Además, existe evidencia empírica de que la distribución incondicional de los retornos a menudo puede mostrar cierto grado de asimetría. Por ejemplo, en el caso de los datos analizados previamente, hay signos de asimetría.

Figura 3.1: Autocorrelaciones de los retornos (primera columna) y de los retornos al cuadrado (segunda columna) de los índices DAX, NIKKEI225, CAC40 y BVL

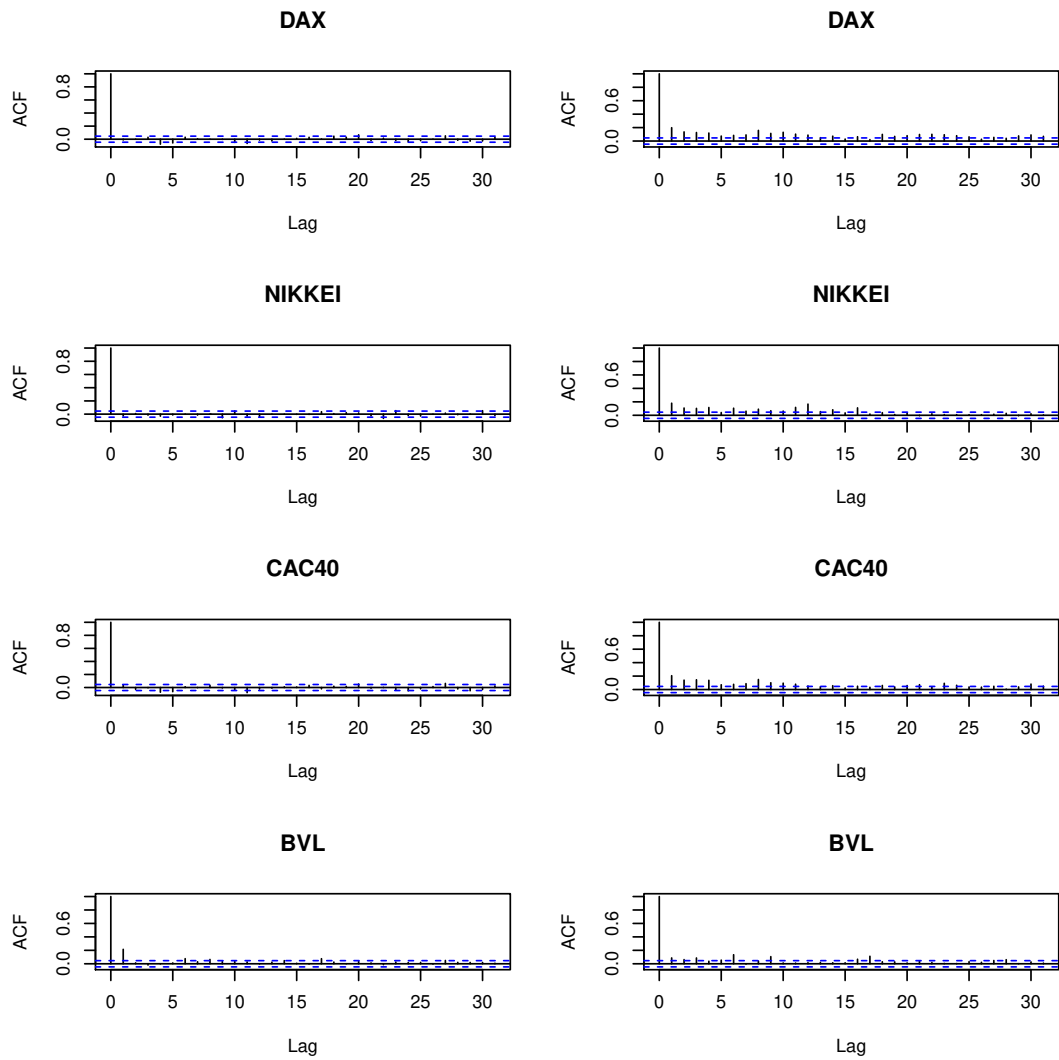
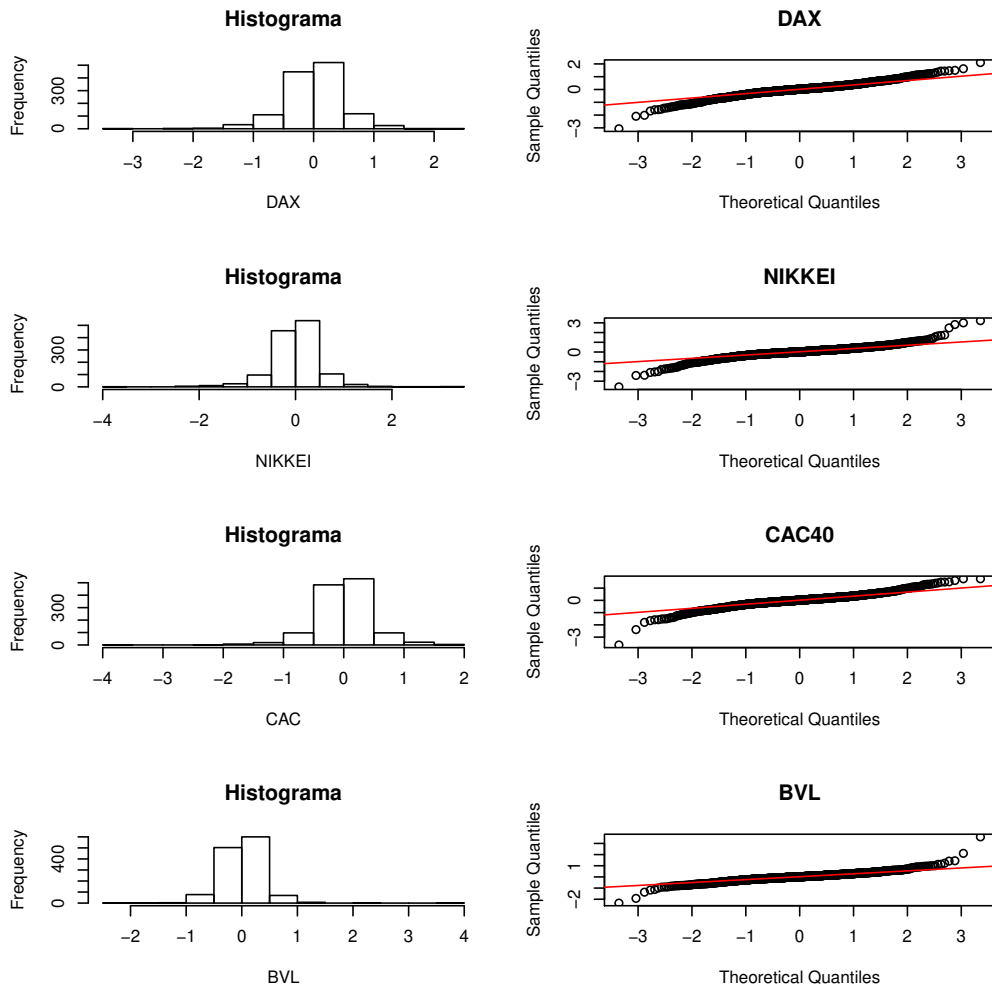


Figura 3.2: Histogramas de los retornos (primera columna) y qqplots de los retornos (segunda columna) de los índices DAX, NIKKEI225, CAC40 y BVL.



3.7. Selección del mejor modelo heterocedástico

En esta sección se presenta la aplicación del enfoque bayesiano en el modelo heterocedástico condicional autorregresivo generalizado multivariado correlación condicional dinámica DCC - GARCH(1,1) para series de activos financieros. Para el término de error del modelo, se consideran las distribuciones de probabilidad: $SSN(0, I_m, \gamma)$, $SST(0, I_m, \gamma, \nu)$ y $SSGED(0, I_m, \gamma, k)$.

3.7.1. Aplicación de los algoritmos Metropolis - Hastings

La inferencia de los modelos se realizó bajo el enfoque bayesiano utilizando la metodología MCMC. Las cadenas de Markov se generan en 2 pasos utilizando los algoritmos presentados en la sección (3.3). En el primer paso, el algoritmo **MH np - Bloques** se utiliza para generar una cadena de Markov inicial, luego esta cadena es utilizada para estimar la matriz de covarianza de los np parámetros. El segundo paso es usar el algoritmo **MH 1 - Bloque** para generar una segunda cadena de Markov. Las estimaciones de Monte Carlo se obtienen de esta segunda cadena.

Esta metodología es aplicada a un conjunto de datos reales: serie de retornos de las bolsas de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225), París (CAC40) y de Lima (BVL). Analizados en el periodo del 01 de abril del 2015 al 31 de enero del 2020 totalizando 1263 días observados. A estas series de retornos para un mejor análisis se les multiplicó por 100.

Los valores para los hiperparámetros de las distribuciones a priori son elegidos según Fioruci et al. (2014b): $\mu_v = \mu_k = \mu_a = \mu_b = \mu_{\omega_i} = \mu_{\alpha_i} = \mu_{\beta_i} = 0$ y $\sigma_v^2 = \sigma_k^2 = \sigma_a^2 = \sigma_b^2 = \sigma_{\omega_i}^2 = \sigma_{\alpha_i}^2 = \sigma_{\beta_i}^2 = 100$, para $i = 1, \dots, m$. De esta forma, se obtienen a prioris con información vaga en casi todos los espacios paramétricos, excepto para los parámetros de asimetría para los cuales se utilizó las a prioris descritas en la sección (3.2).

Para las actualizaciones de MCMC se adopta un algoritmo Metropolis - Hastings Caminata aleatoria donde todos los parámetros se actualizan como (1-bloque), en cada iteración se generan nuevos valores a partir de una distribución normal multivariante centrada alrededor del valor actual con una matriz propuesta de varianza - covarianza que se calculó a partir de un procedimiento de ajuste piloto con un total de 10 000 iteraciones.

Este ajuste piloto se llevó a cabo mediante la ejecución de actualizaciones de Metropolis - Hastings Caminata aleatoria unidimensionales (np - bloques) con distribuciones candidatas normales univariadas cuyas varianzas se calibraron para obtener buenas tasas de aceptación entre 0.20 y 0.50 (paquete bayesDccGarch), luego se ejecutaron un total de 200 000 iteraciones descartandose las primeras 30 000 realizaciones como quemado y además se hizo un adelgazamiento por cada quinta iteración. Los resultados a posterioris se basan entonces en 34 000 realizaciones de la cadena de Markov con las distribuciones a prioris descritas anteriormente.

La implementación de los algoritmos de simulación están disponibles en el paquete bayesDccGarch, software R.

3.7.2. Criterios de selección del mejor modelo

Los criterios de Akaike (AIC) y Bayesian (BIC) fueron propuestos en Akaike (1974) y Schwarz et al. (1978), respectivamente, y se definen de la siguiente manera:

$$AIC(M) = -2\log(L_M(\hat{\theta})) + 2np_M,$$

$$BIC(M) = -2\log(L_M(\hat{\theta})) + np_M\log(n),$$

donde M es el modelo en cuestión, $L_M(\hat{\theta})$ es la función de verosimilitud de M calculada sobre el vector de parámetros estimados $\hat{\theta}$, np_M el número de parámetros del modelo M y n el número de observaciones de la muestra.

Los valores más bajos obtenidos por los criterios AIC y BIC indican el mejor modelo, estos criterios AIC y BIC son ampliamente utilizados para los enfoques clásicos pero para un enfoque bayesiano, es preferible utilizar la estimación de la esperanza a posteriori de la función de verosimilitud ($E(\widehat{L_M(\theta)})$) en lugar de la función de verosimilitud calculada en el punto estimado ($L_M(\hat{\theta})$), estos criterios se conocen como EAIC y EBIC.

Para la selección del mejor modelo, se consideraron 3 criterios, los dos primeros consisten en la obtención de la esperanza a posteriori de los criterios de información Akaike (EAIC) y Bayesiano (EBIC), el tercer criterio utilizado es la Desviación de Información (DIC). Estos criterios “penalizan” la función de verosimilitud del modelo de acuerdo con su complejidad, de modo que los mejores modelos tienen un mayor valor de función de verosimilitud y menos parámetros.

El criterio de desviación de información (DIC) se propuso en Spiegelhalter et al. (2002) y se usa ampliamente en la inferencia bayesiana, especialmente cuando se consideran los métodos MCMC.

La desviación se define como

$$D(\theta) = -2\log(L_M(\hat{\theta}))$$

y \bar{D} como la esperanza a posteriori de $D(\theta)$, esta esperanza se usa como una medida de la calidad del ajuste del modelo con respecto a los datos, cuanto mayor sea el $D(\theta)$, peor será el ajuste. El número efectivo de parámetros se define como $p_D = \bar{D} - D(\hat{\theta})$. De esta manera, el criterio DIC de un modelo M se calcula como:

$$DIC(M) = p_D + \bar{D}.$$

De donde los valores DIC más bajos indican mejores modelos.

CAPÍTULO 4

RESULTADOS Y DISCUSIÓN

Para seleccionar el mejor modelo se consideraron los criterios mencionados en la sección (3.7.2). Se verificó la convergencia y la buena mezcla de las cadenas de Markov simuladas mediante inspección visual de las trazas marginales, estimaciones de densidad, autocorrelaciones y pruebas formales, estos diagnósticos indicaron buena convergencia, el tiempo de cálculo en este paso fué inferior a 30 minutos para cada modelo además las tasas de aceptación fueron de 0.24 para SST, 0.23 para SSGED y 0.27 para SSN; es decir los tiempos de ejecución son razonables y las tasas de aceptación se encontraron alrededor de 0.234 (propuesta hecha por Roberts et al. (2001)). (ver anexo figuras: 5.3, 5.4 y 5.5 para el modelo con errores SST)

La Tabla 4.1 presenta los criterios de selección del mejor modelo heterocedástico condicionalmente autorregresivo generalizado multivariado correlación condicional dinámica DCC - GARCH(1,1), para cada una de las distribuciones de probabilidad de los errores estudiados. En esta tabla se puede ver que los criterios EAIC EBIC y DIC, es decir los tres criterios seleccionaron el modelo con errores SST como mejor modelo.

Tabla 4.1: Criterios para la selección del mejor modelo multivariado

Modelos	EAIC	EBIC	DIC
SSN	2702.699	2795.242	2683.934
SST	2336.367	2434.050	2316.101
SSGED	2422.065	2519.749	2401.647

En los próximos análisis solo se utiliza este modelo.

En la Tabla 4.2 se muestra un resumen de las simulaciones de MCMC para el modelo con errores SST. Las medias posteriores, desviaciones estandar, medianas e intervalos creíbles al 95 % de los parámetros de asimetría indican una alta asimetría para los retornos de las series DAX y NIKKEI225, mientras que para los retornos de CAC40 y para BVL hay una leve inclinación hacia la izquierda.

Tabla 4.2: Resumen de las simulaciones MCMC para el modelo heterocedástico con errores SST

		Media	Desviación Estandar	2.5 %	Mediana	97.5 %
DAX	γ_1	0.939881	0.024452	0.891326	0.940202	0.98703
	ω_1	0.016850	0.004701	0.009687	0.016214	0.02781
	α_1	0.073159	0.013744	0.049254	0.072128	0.10299
	β_1	0.849125	0.030489	0.780214	0.852383	0.89898
NIKKEI	γ_2	0.900824	0.026340	0.848303	0.901216	0.95150
	ω_2	0.022563	0.006338	0.012567	0.021765	0.03704
	α_2	0.183121	0.036616	0.122100	0.179810	0.26381
	β_2	0.731489	0.051036	0.618716	0.736182	0.81644
CAC40	γ_3	0.966817	0.027869	0.912794	0.966745	1.02204
	ω_3	0.016825	0.004150	0.010343	0.016296	0.02634
	α_3	0.095057	0.017027	0.065107	0.093741	0.13242
	β_3	0.818483	0.032895	0.745736	0.821863	0.87326
BVL	γ_4	0.961111	0.026679	0.908582	0.960875	1.01454
	ω_4	0.009713	0.003943	0.004069	0.009053	0.01896
	α_4	0.079194	0.024085	0.039816	0.076742	0.13401
	β_4	0.846237	0.048256	0.737352	0.852563	0.92177
	a	0.030546	0.009782	0.013730	0.029797	0.05182
	b	0.814858	0.079481	0.624053	0.828090	0.93177
	ν	5.566379	0.350145	4.897407	5.560763	6.26530

Además, la estimación del parámetro de cola ν (última fila de la Tabla 4.2) indica que una distribución de cola gruesa es apropiada para el término de error.

Teniendo en cuenta los resultados de la Tabla 4.2 el modelo heterocedástico de series de tiempo que modela la volatilidad de activos financieros para los datos diarios de las series estadísticas de las bolsas de valores de cierre diario de los índices bursátiles de Frankfurt (DAX), Tokio (NIKKEI225) y París (CAC40) y del índice General de la Bolsa de Valores de Lima (BVL) utilizando el enfoque bayesiano es:

$H_t = D_t R_t D_t$ siendo,

$$D_t = \text{diag}(h_{11,t}^{1/2}, \dots, h_{44,t}^{1/2})$$

es decir según los valores de la Tabla 4.2:

$$h_{11,t} = 0,016850 + 0,073159y_{1,t-1}^2 + 0,849125h_{11,t-1}$$

$$h_{22,t} = 0,022563 + 0,183121y_{2,t-1}^2 + 0,731489h_{22,t-1}$$

$$h_{33,t} = 0,016825 + 0,095057y_{3,t-1}^2 + 0,818483h_{33,t-1}$$

$$h_{44,t} = 0,009713 + 0,079194y_{4,t-1}^2 + 0,846237h_{44,t-1}$$

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}$$

donde Q_t son matrices simétricas definidas positivas de orden 4×4 dadas por

$$Q_t = (1 - 0,030546 - 0,814858)R + 0,030546\mu_{t-1}\mu_{t-1}^t + 0,814858Q_{t-1}$$

Con $\mu_t = D_t^{-1}y_t$ y R la matriz de covarianzas incondicionales de μ_t .

Finalmente, es de interés práctico comprobar si las correlaciones cambian con el tiempo, lo que justifica la complejidad adicional que introduce el modelo DCC, esto se puede ver de las estimaciones en la parte inferior de la Tabla 4.2 las cuales indican que la hipótesis del modelo de correlación condicional constante (CCC) es decir $a = b = 0$ son rechazadas en base a las distribuciones marginales posteriores de a y b , además teniendo en cuenta la distribución posterior para

la suma $a + b$ vemos que hay una persistencia fuerte en la ecuación (3.11) pues ($a + b = 0,030546 + 0,814858 = 0,845404$).

En la literatura, las distribuciones de probabilidad más utilizadas para los errores del modelo GARCH, son la Normal, t -Student y GED en su forma estandarizada, para el caso univariado para insertar asimetría a la distribución del error se utilizó el método de Fernández and Steel (1998), y para los errores del modelo heterocedástico DCC - GARCH, se aplicó el método presentado en Bauwens and Laurent (2005), este método facilitó la especificación de las distribuciones a priori separando los efectos de los parámetros de asimetría y de cola, obteniendo así versiones asimétricas de las distribuciones antes mencionadas. El uso de distribuciones de probabilidad asimétricas para los errores del modelo heterocedástico DCC - GARCH es una alternativa para modelar el grado de asimetría presente en la volatilidad de los retornos para una serie dada.

En el estudio realizado por Aquino Gutierrez et al. (2017), se modelan dos características principales: la asimetría y las colas pesadas presentes en la distribución incondicional de las series de retornos, se estiman los parámetros de los modelos mediante un enfoque bayesiano con la metodología MCMC específicamente los algoritmos Metropolis - Hastings (MH), pero solo para una serie es decir para el caso univariado.

En el presente estudio, también se modelan dos características principales de las series de tiempo: la asimetría y las colas pesadas presentes en la distribución incondicional de las series, se estiman los parámetros de los modelos mediante un enfoque bayesiano con la metodología MCMC específicamente los algoritmos Metropolis - Hastings (MH) caminata aleatoria para varias series de tiempo de índices bursátiles es decir se hace un estudio multivariado, con los datos analizados se determinó que las series presentan correlación condicional dinámica y su volatilidad es heterocedástica condicionalmente autoregresiva dado que los parámetros a y b del modelo son diferentes de cero.

CAPÍTULO 5

CONCLUSIONES

1. En este trabajo de tesis se logró mediante el enfoque bayesiano, metodología MCMC, modelar la volatilidad de activos financieros. La metodología bayesiana para la estimación de los parámetros de los modelos trajo algunas ventajas, como la facilidad de interpretación y la posibilidad de insertar información a priori para los parámetros, para ello se utilizó el paquete bayesDccGarch implementado en el software R.
2. Con los datos obtenidos de la web, un total de 1263 para cada serie tomados en el periodo del 01 de abril del 2015 al 31 de enero del 2020, se obtuvo los retornos para cada serie y a éstos se les multiplicó por 100, para luego hacer la simulación de las cadenas de Markov mediante el algoritmo Metropolis - Hasting en bloques y en un solo bloque, paquete bayesDccGarch, empleando tiempos razonables (menos de 30 minutos). Para lo cual se utilizó las siguientes órdenes en el programa R:

- (a) Para la distribución de errores SSN:

```
SSN=bayesDccGarch(datosx100, nSim = 200000,  $\nu_i$  = NULL,  $\omega_i$ =  
rep(0.03, ncol(datosx100)),  $\alpha_i$ = rep(0.03, ncol(datosx100)),  $\beta_i$ = rep(0.8,  
ncol(datosx100)),  $a_i$  = 0.03,  $b_i$ = 0.8,  $\gamma_i$  = rep(1, ncol(datosx100)), error-  
Dist = 1, control = list(nPilotSim=10000))  
SSN1=window(SSN, start = 30000, end = 200000, thin = 5)  
summary(SSN1)
```

- (b) Para la distribución de errores SST:

```
SST=bayesDccGarch(datosx100, nSim = 200000,  $\nu_i$ = 8,  $\omega_i$  = rep(0.03,  
ncol(datosx100)),  $\alpha_i$  = rep(0.03, ncol(datosx100)),  $\beta_i$  = rep(0.8, ncol  
(datosx100)),  $a_i$  = 0.03,  $b_i$  = 0.8,  $\gamma_i$ = rep(1, ncol(datosx100)), errorDist
```

```

= 2, control = list(nPilotSim=10000))
SST1=window(SST, start = 30000, end = 200000, thin = 5)
summary(SST1)

```

(c) Para la distribución de errores SSGED:

```

SSGED=bayesDccGarch(datosx100, nSim = 200000, k_i = 1.5,
omega_i = rep(0.03, ncol(datosx100)), alpha_i = rep(0.03, ncol(datosx100)), beta_i=
rep(0.8, ncol(datosx100)), a_i = 0.03, b_i = 0.8, gamma_i = rep(1, ncol(datosx100)),
errorDist = 3, control = list(nPilotSim=10000))
SSGED1=window(SSGED, start = 30000, end = 200000, thin = 5)
summary(SSGED1).

```

Tomando las esperanzas a posteriori de los criterios de información (AIC, BIC) y el desvío de información DIC, y según los datos de la **Tabla 4.1**, el modelo heterocedástico condicionalmente autorregresivo generalizado multivariado con matriz de correlación condicional dinámica DCC - GARCH(1,1) seleccionado fué el que usó para los errores la distribución de probabilidad SST, ver capítulo 4.

3. En la Figura 5.4 (primera columna) se puede ver la traza de todos los parámetros, según estas gráficas, las cadenas presentan una buena exploración de estados para todos los parámetros. La Figura 5.5 muestra las autocorrelaciones de los parámetros del modelo obtenido, para los retornos aproximados de las series DAX, NIKKEI225, CAC40 y BVL. Donde se observa que las correlaciones disminuyen rápidamente a medida que aumentan los datos, es decir sigue un decaimiento geométrico.
4. El enfoque bayesiano junto con el uso de distribuciones de probabilidad asimétricas para los errores del modelo heterocedástico brindó la posibilidad de analizar la distribución de probabilidad de asimetría presente en los rendimientos de las series de activos financieros.

RECOMENDACIONES

Algunas propuestas para trabajos futuros:

1. Proponer distribuciones a priori distintas a las distribuciones normales. Por ejemplo, distribución t -Student y GED.
2. Proponer el uso de MH para otros modelos: modelos de volatilidad estocástica, familia GARCH asimétrica por ejemplo los modelos EGARCH, entre otros.
3. Para el término de error del modelo, se consideraron las distribuciones de probabilidad de $SSN(0, I_m, \gamma)$, $SST(0, I_m, \gamma, v)$ y $SSGED(0, I_m, \gamma, k)$. Se recomienda para trabajos futuros hacer un análisis bayesiano para los casos simétricos de estas distribuciones, las cuales se obtienen considerando solo $\gamma_i = 1$ para $i = 1, \dots, m$, denominadas N, ST y GED .

REFERENCIAS BIBLIOGRÁFICAS

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.
- Aquino Gutierrez, K. F. et al. (2017). Modelagem da volatilidade em séries temporais financeiras via modelos garch com abordagem bayesiana.
- Ardia, D. et al. (2008). *Financial risk management with Bayesian estimation of GARCH models*, volume 18. Springer.
- Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*. Springer Science & Business Media.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.
- Bauwens, L. and Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics*, 23(3):346–354.
- Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and statistics*, pages 498–505.
- Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79(1):99–113.
- Casella, G. and Robert, C. P. (1999). Monte carlo statistical methods.

- Chan, K. S. and Geyer, C. J. (1994). Discussion: Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1747–1758.
- Chatfield, C. (2013). *The analysis of time series: theory and practice*. Springer.
- David, A. (2006). Bayesian estimation of the garch (1, 1) model with normal innovations.
- Ehlers, R. S. (2012). Computational tools for comparing asymmetric garch models via bayes factors. *Mathematics and Computers in Simulation*, 82(5):858–867.
- Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.
- Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the american statistical association*, 93(441):359–371.
- Fioruci, J. A. (2012). *Modelagem de volatilidade via modelos GARCH com erros assimétricos: abordagem Bayesiana*. PhD thesis, Universidade de São Paulo.
- Fioruci, J. A., Ehlers, R. S., and Andrade Filho, M. G. (2014a). Bayesian multivariate garch models with dynamic correlations and asymmetric error distributions. *Journal of Applied Statistics*, 41(2):320–331.
- Fioruci, J. A., Ehlers, R. S., and Louzada, F. (2014b). Bayesdccgarch-an implementation of multivariate garch dcc models. *arXiv preprint arXiv:1412.2967*.
- Giller, G. L. (2005). A generalized error distribution.

- Gómez, E., Gomez-Viilegas, M., and Marín, J. M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics-Theory and Methods*, 27(3):589–600.
- Harris, T. (1956). The existence of stationary measures for certain markov processes. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory*, pages 113–124. University of California Press.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.
- Investing.com (2020a). Cax40 data histórica. [Web; accedido el 11-02-2020].
- Investing.com (2020b). Dax data histórica. [Web; accedido el 11-02-2020].
- Investing.com (2020c). Igbvl data histórica. [Web; accedido el 11-02-2020].
- Investing.com (2020d). Nikkei225 data histórica. [Web; accedido el 11-02-2020].
- James, B. R. (1996). *Probabilidade: um curso em nível intermediário*. Number 519.2.
- Mackay, D. (1986). *Introduction to Monte Carlo Methods*. Springer.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.
- Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, 44(247):335–341.
- Meyn, S.P. y Tweedie, R. (1993). *Markov chains and stochastic stability*. Springer-Verlag.

- Morettin, P. A. (2017). *Econometria financeira: um curso em séries temporais financeiras*. Editora Blucher.
- Nummelin, E. (1978). A splitting technique for harris recurrent markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 43(4):309–318.
- Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.
- Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods, (second edition)*. Springer-Verlag, New York.
- Roberts, G. O., Rosenthal, J. S., et al. (2001). Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, 16(4):351–367.
- Roberts, G. O. and Smith, A. F. (1994). Simple conditions for the convergence of the gibbs sampler and metropolis-hastings algorithms. *Stochastic processes and their applications*, 49(2):207–216.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639.
- Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes—a study of the daily sugar prices 1961-75. *Time series analysis: theory and practice*, 1:203–226.
- Team, R. C. et al. (2013). R: A language and environment for statistical computing.

- Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728.
- Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John wiley & sons.
- Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, 20(3):351–362.
- wikipedia (2020a). John von neumann. [Web; accedido el 14-10-2020].
- wikipedia (2020b). Nicholas metropolis. [Web; accedido el 14-10-2020].
- wikipedia (2020c). Proyecto manhattan. [Web; accedido el 14-10-2020].
- wikipedia (2020d). Robert oppenheimer. [Web; accedido el 14-10-2020].

ANEXOS

Figura 5.1: Gráficos de las series (primera columna) y de los retornos (segunda columna) de los índices DAX, NIKKEI225, CAC40 y BVL.

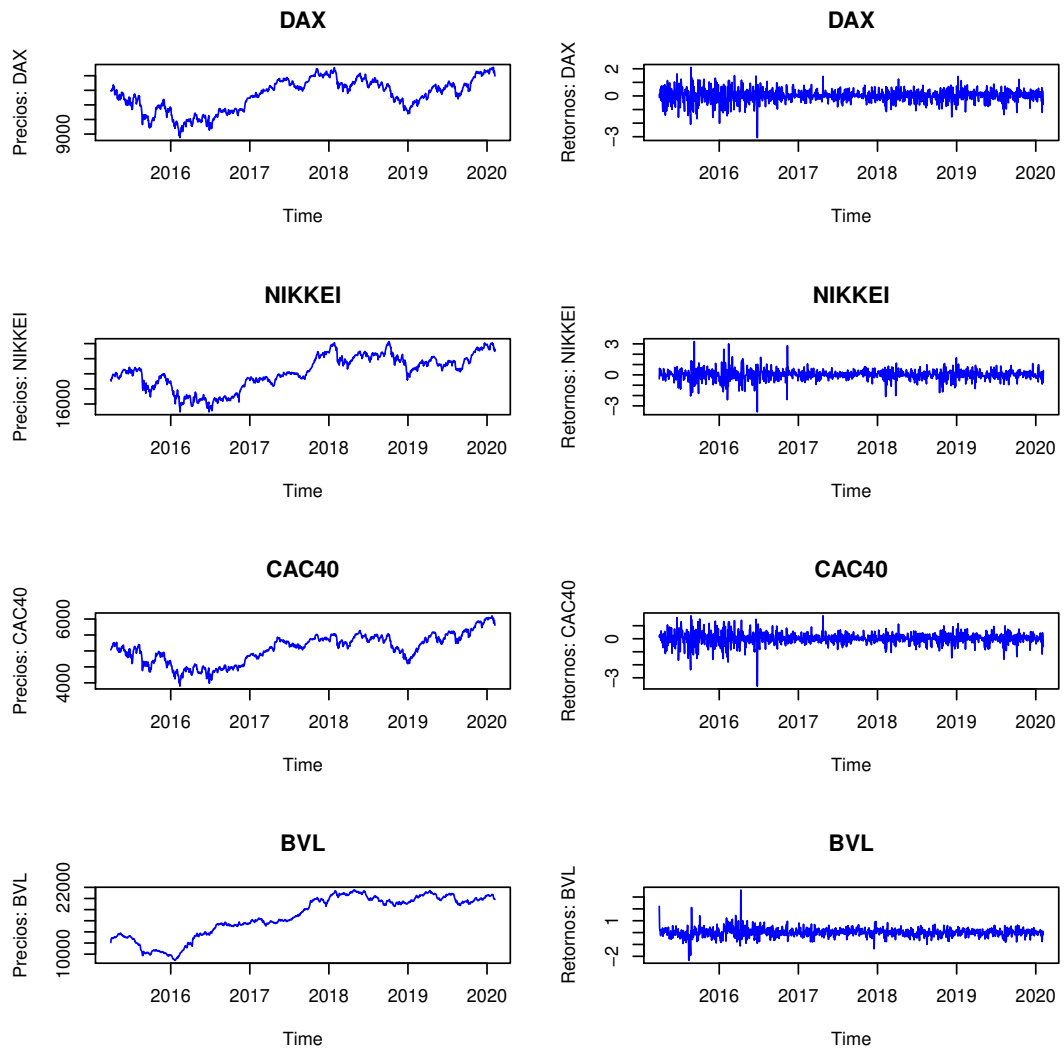
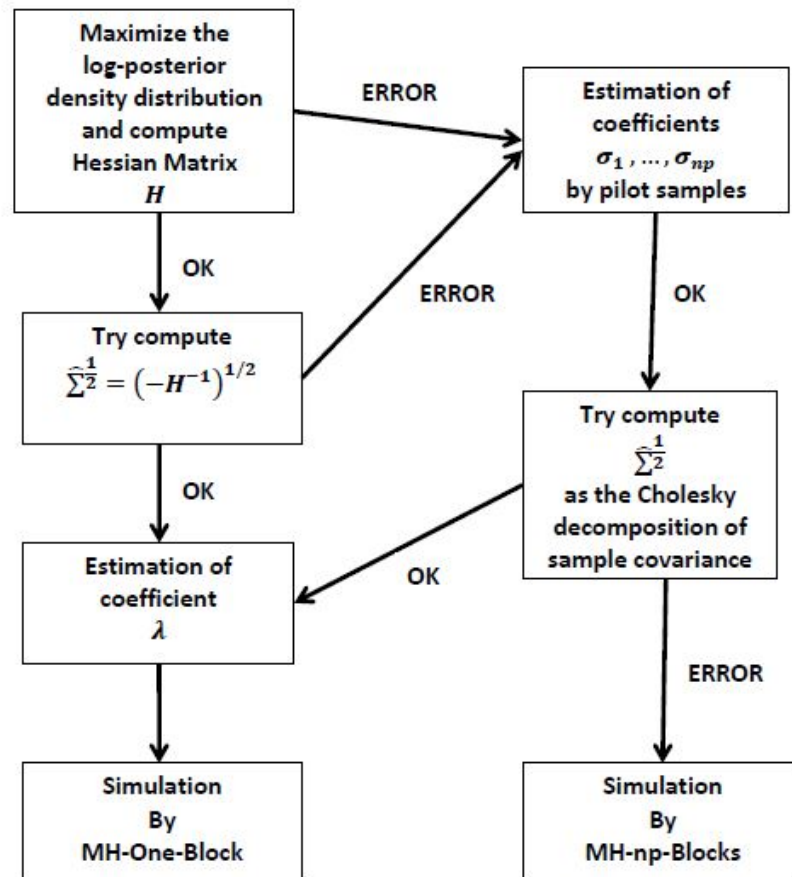


Figura 5.2: Diagrama de estrategias para elegir entre el algoritmo MH 1 - Block y MH np - Blocks.



Fuente: Diagrama tomado de Fioruci et al. (2014b)

Figura 5.3: Volatilidad aproximada de los retornos de los índices de las series, DAX, NIKKEI225, CAC40 y BVL para el modelo heterocedástico DCC - GARCH(1,1) con errores SST

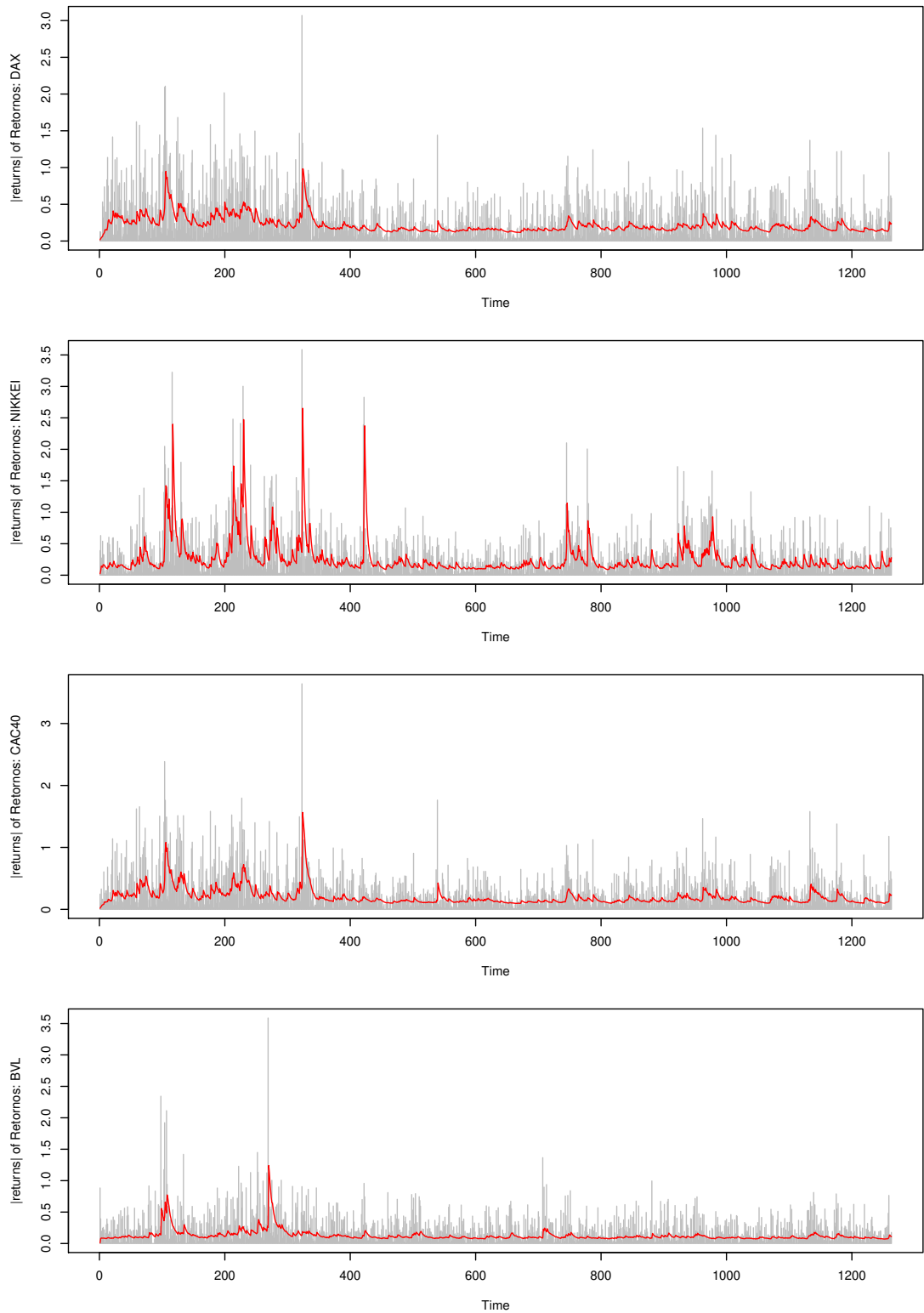
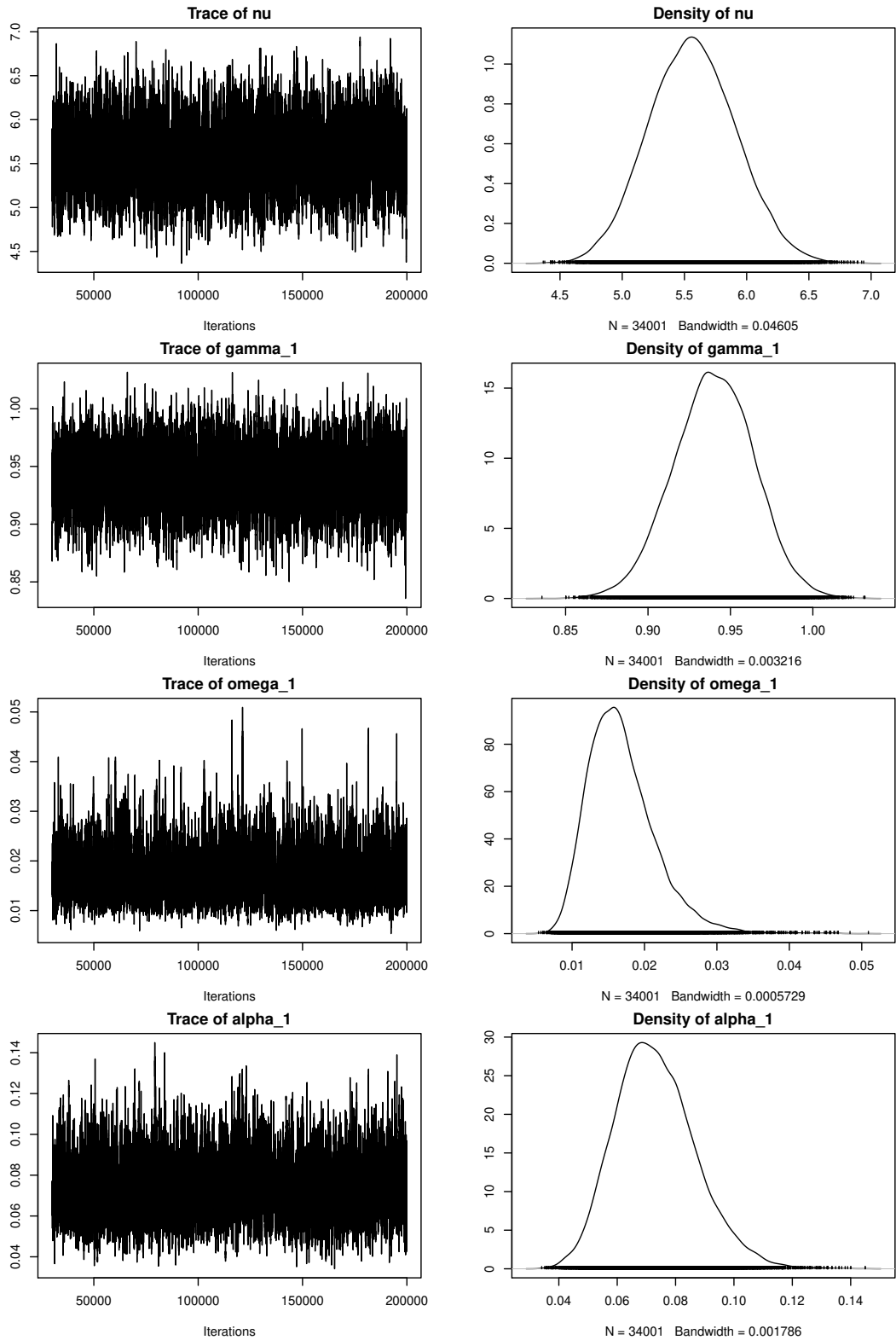
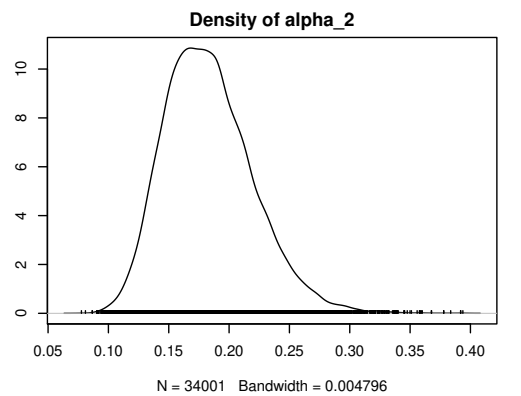
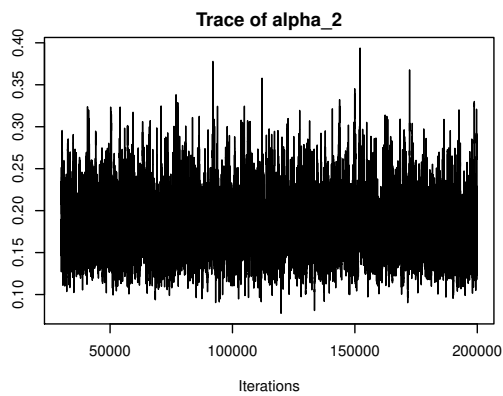
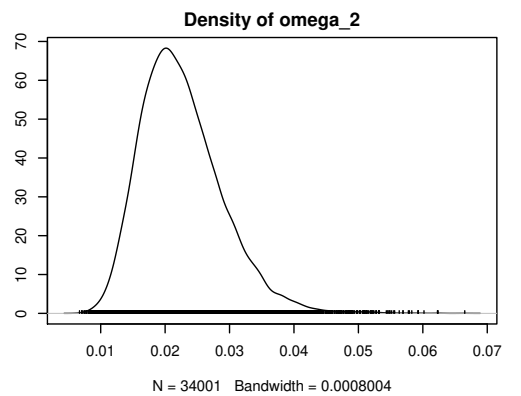
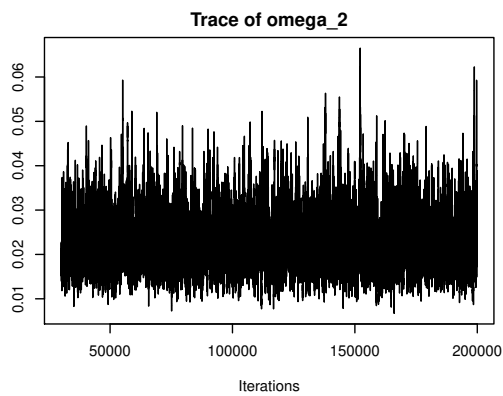
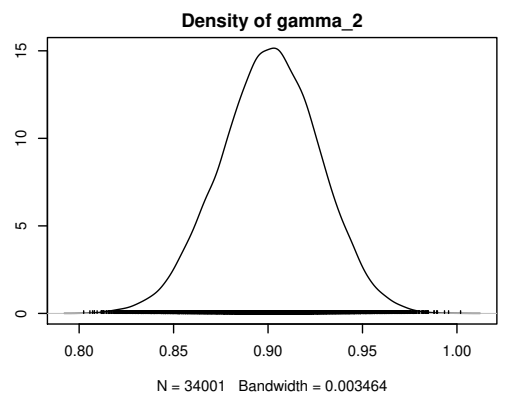
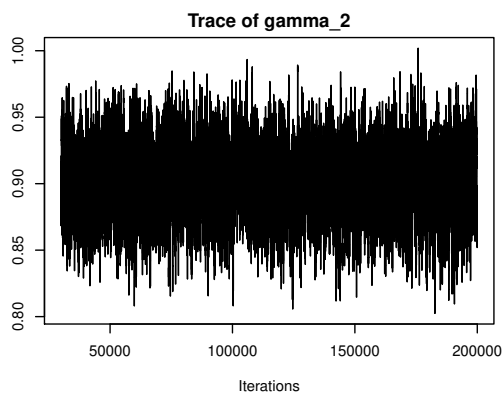
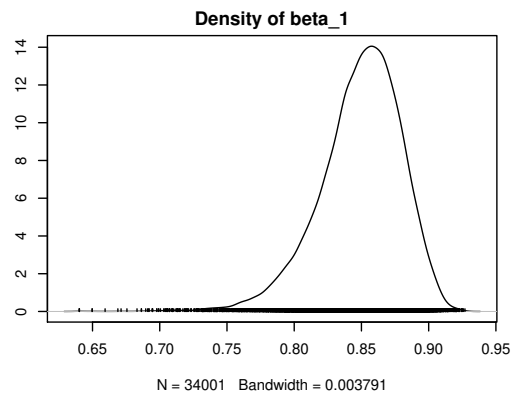
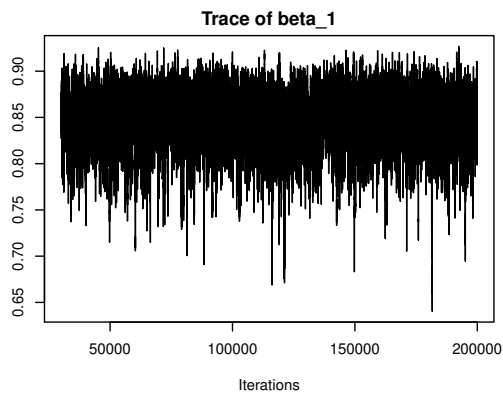
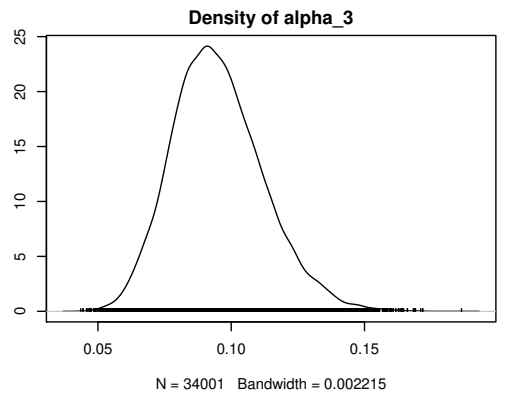
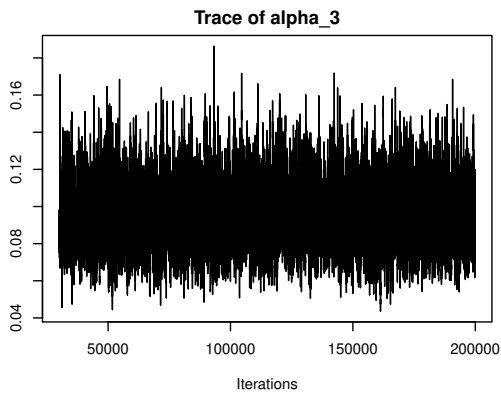
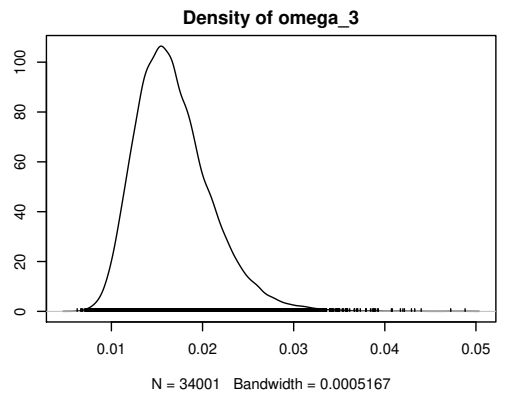
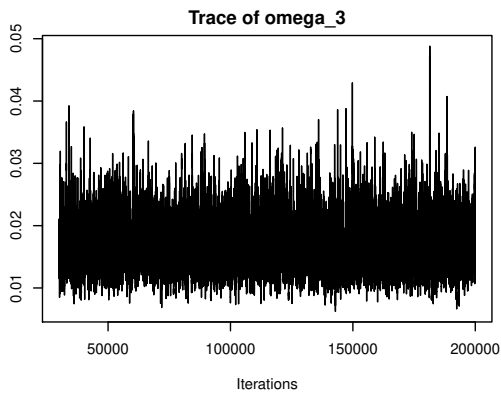
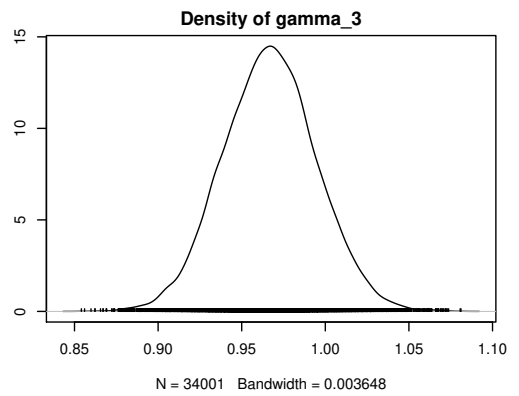
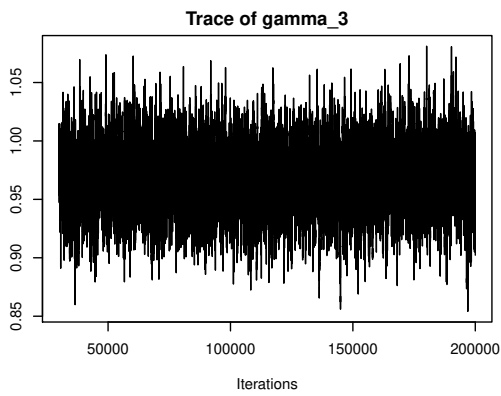
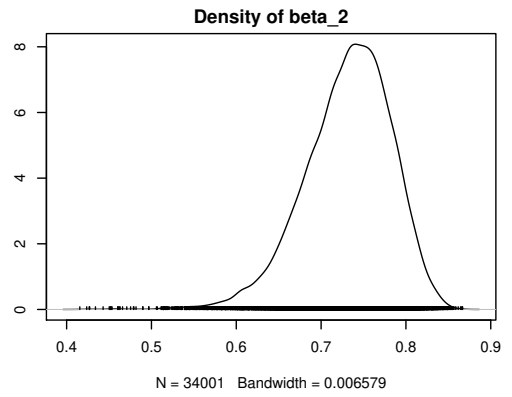
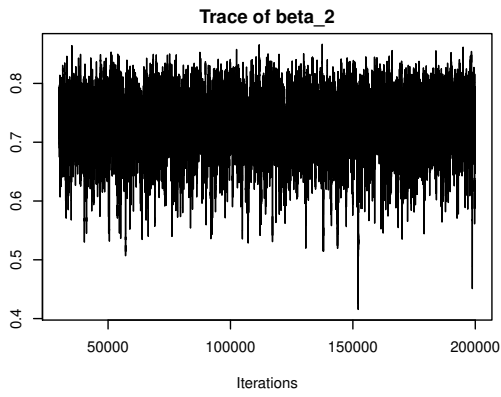
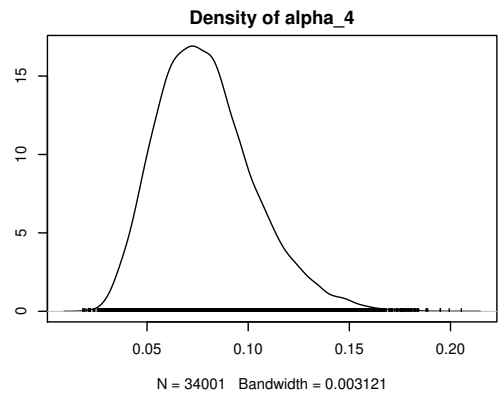
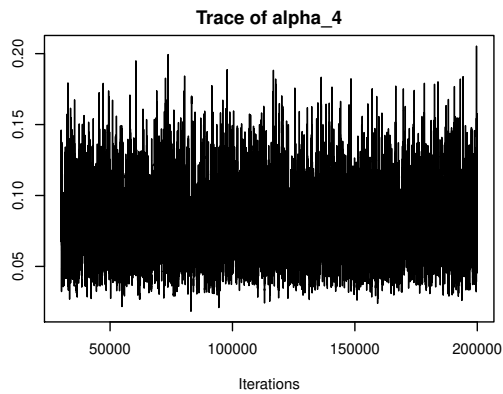
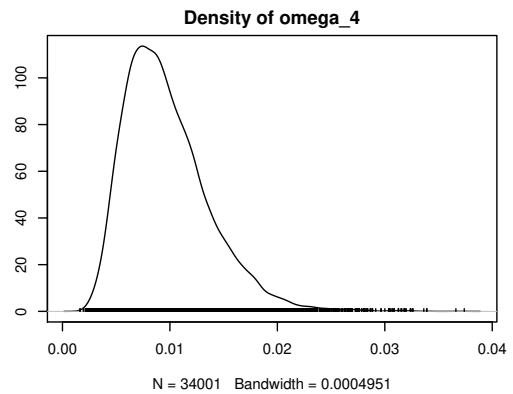
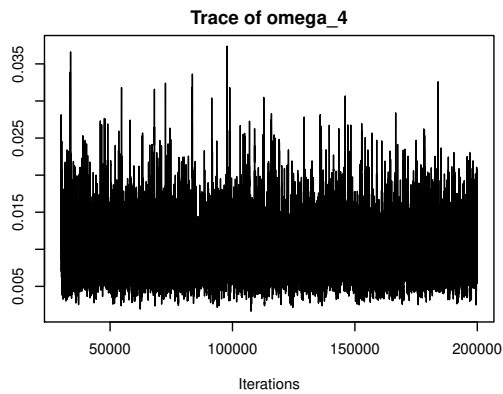
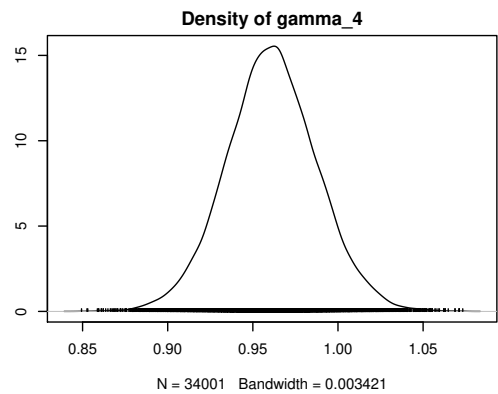
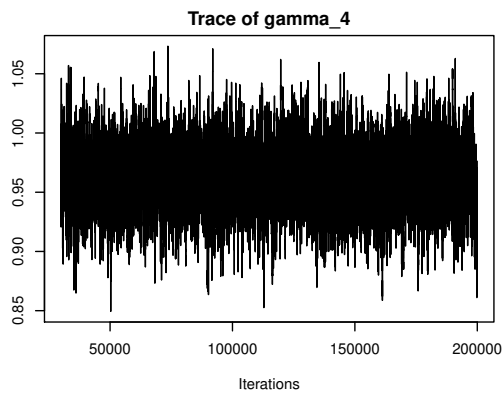
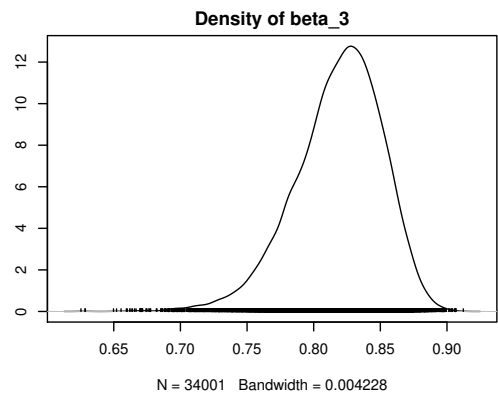
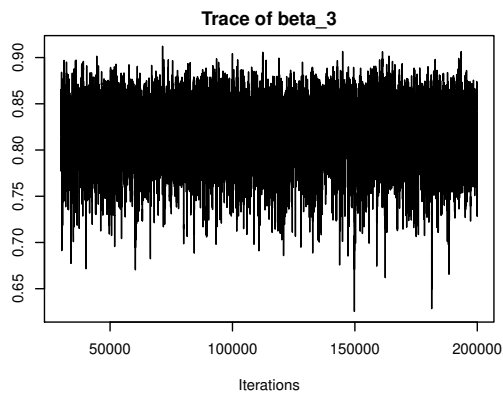


Figura 5.4: Traza y densidades aproximados de todos los parámetros del modelo heterocedástico DCC - GARCH(1,1) con errores SST, para los retornos de las series DAX, NIKKEI225, CAC40 y BVL.









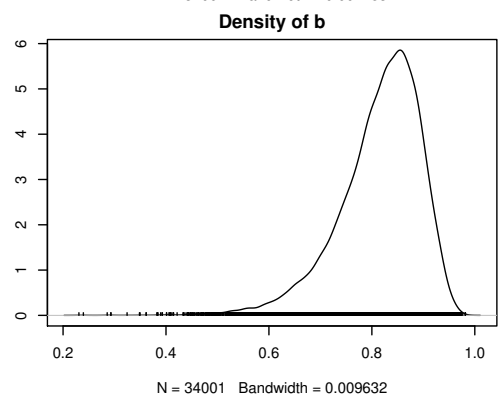
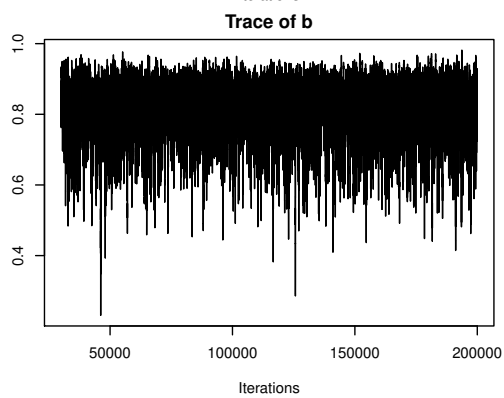
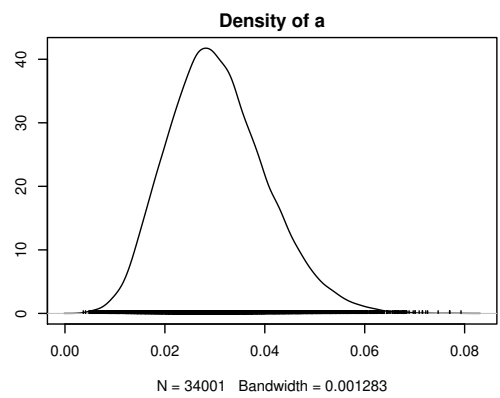
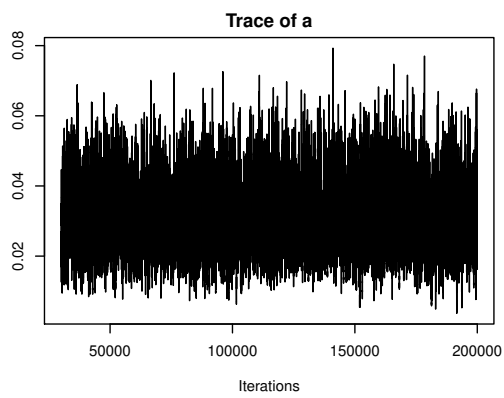
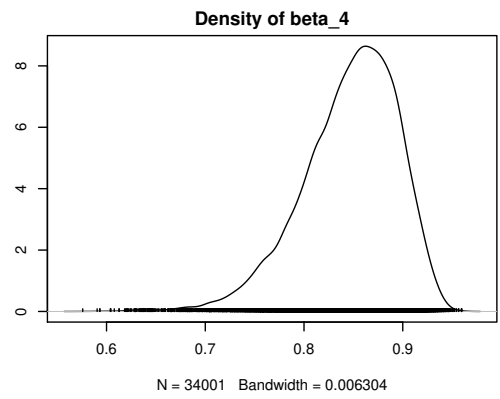
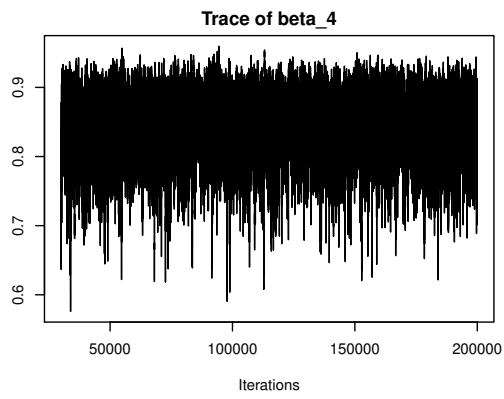
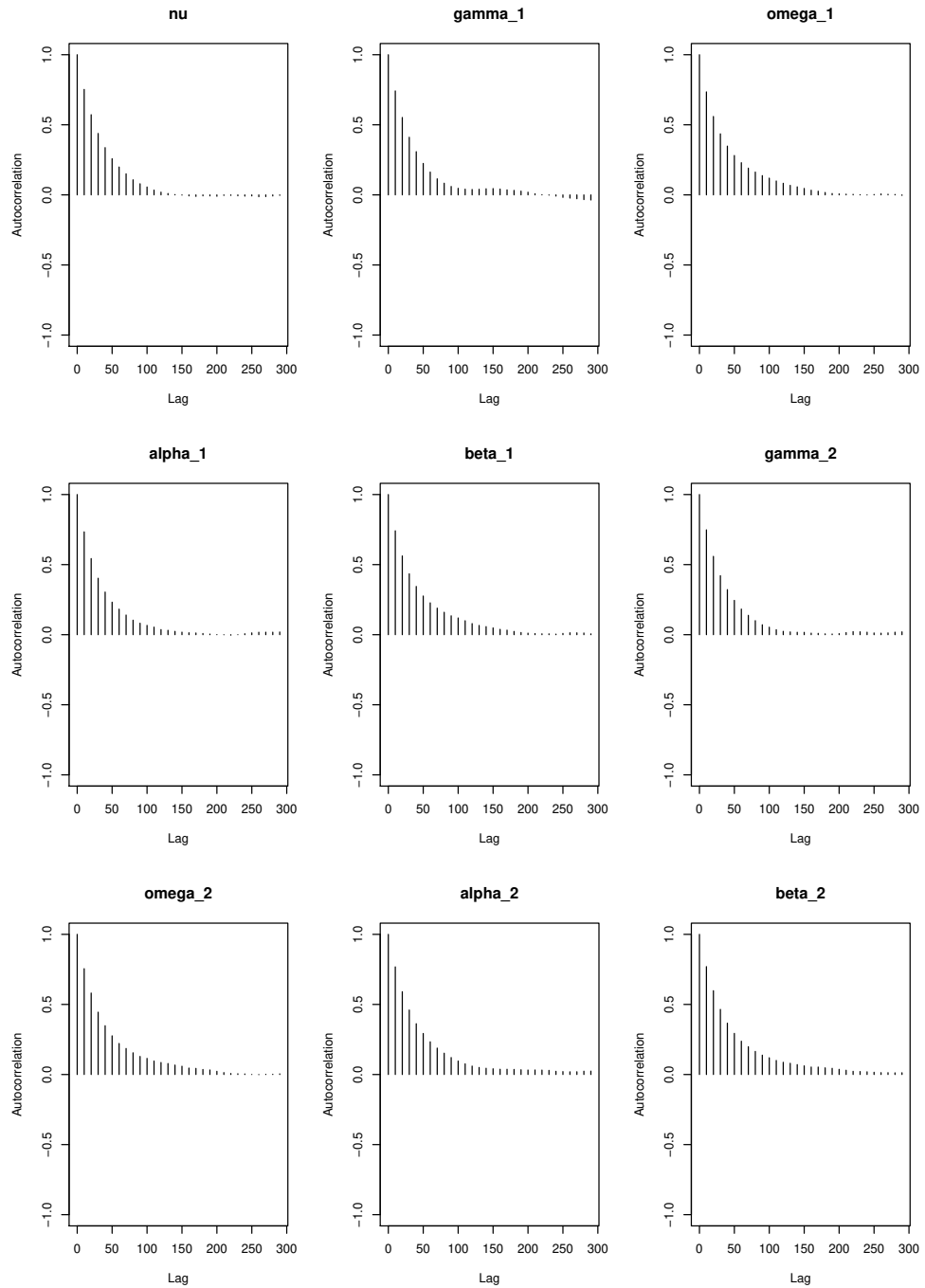
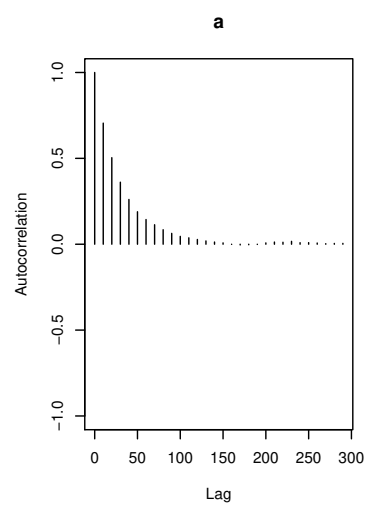
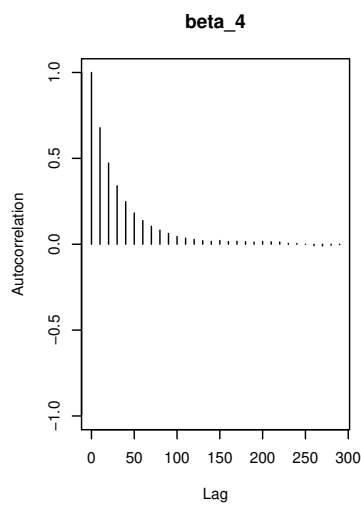
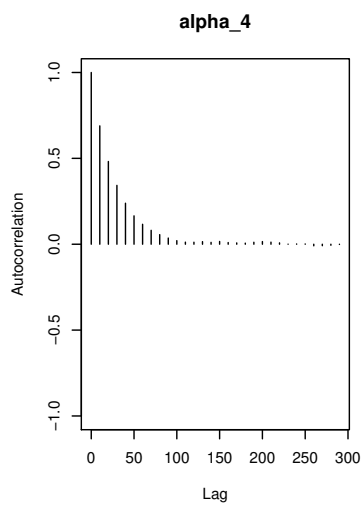
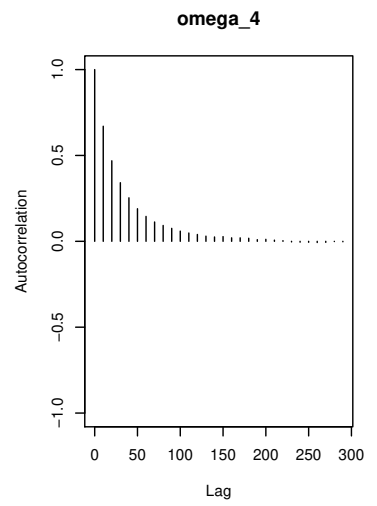
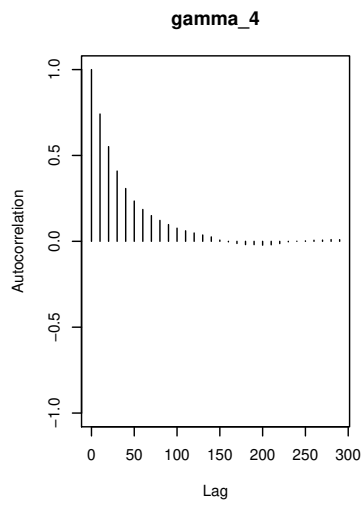
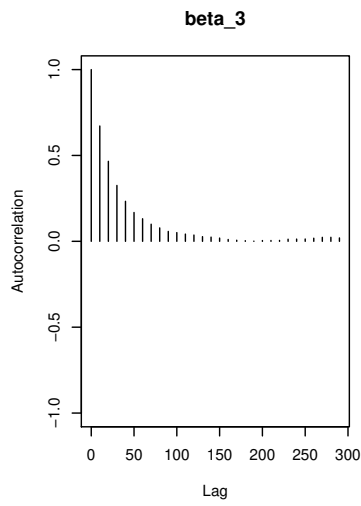
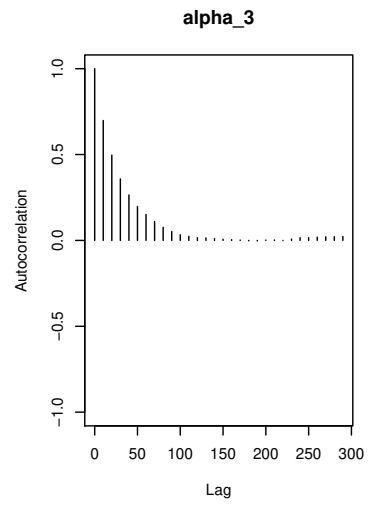
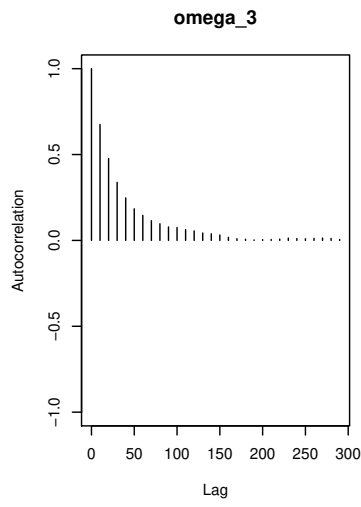
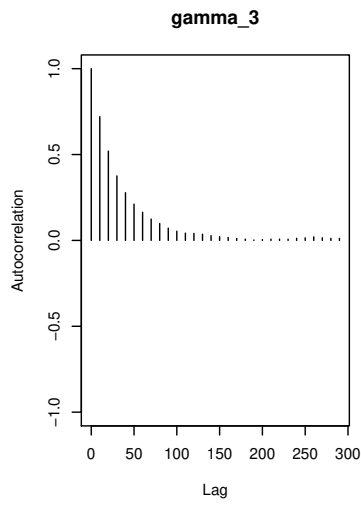
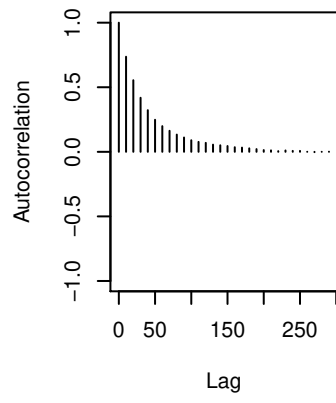


Figura 5.5: Autocorrelaciones de todos los parámetros del modelo heterocedástico DCC - GARCH(1,1) con errores SST, para los retornos de las series DAX, NIKKEI225, CAC40 y BVL.





b



La Figura 5.4, muestra las trazas y densidades de los parámetros del modelo heterocedástico condicionalmente autorregresivo generalizado multivariado con matriz de correlación condicional dinámica DCC - GARCH(1,1), según las gráficas de los parámetros se puede ver que las trazas de todos los parámetros presentan una buena exploración de estados.

La Figura 5.5 muestra las autocorrelaciones de los parámetros del modelo heterocedástico condicionalmente autorregresivo generalizado multivariado con matriz de correlación condicional dinámica DCC - GARCH(1,1), para los retornos de las series DAX, NIKKEI225, CAC40 y BVL. Donde se nota que la correlación disminuye rápidamente a medida que aumentan los datos, es decir sigue un decaimiento geométrico.