



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Análisis bayesiano de la volatilidad del tipo de
cambio peso mexicano - dólar estadounidense.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Matemático

PRESENTA:

Jesús Antonio Joel Martínez Cuatlayol

DIRECTORA DE TESIS

Dra. Eliane Regina Rodrigues

Ciudad Universitaria, Cd.Mx., 2021





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno

Martínez
Cuatlayol
Jesús Antonio Joel
221 236 88 63
Universidad Nacional Autónoma de México
Facultad de Ciencias
Matemáticas
416090798

2. Datos del tutor

Dra
Eliane Regina
Rodrigues

3. Datos del sinodal 1

Dra
María Asunción Begoña
Fernández
Fernández

4. Datos del sinodal 2

Dra
Ana
Meda
Guardiola

5. Datos del sinodal 3

M. en C.
Jaime
Vázquez
Alamilla

6. Datos del sinodal 4

Dra
María del Pilar
Alonso
Reyes

7. Datos del trabajo escrito

Análisis bayesiano de la volatilidad del tipo de cambio peso mexicano – dólar estadounidense.
112 p
2021

Agradecimientos

Primeramente quiero dar el mayor de mis agradecimientos a mi familia, principalmente a mi mamá y a mi tía, por todas sus enseñanzas y sobre todo por su apoyo en mis sueños. ¡Las amo!

Agradezco infinitamente a la Dra. Eliane Regina Rodrigues por confiar en mi para desarrollar este trabajo; por su tiempo, paciencia y observaciones. También por motivarme para volver a estudiar la Estadística bayesiana.

A mis sinodales, la Dra. Ana Meda, la Dra. Begoña Fernández, la Dra. Pilar Alonso y el M.C. Jaime Vazquez Alamilla, por los valiosos comentarios y observaciones que aportaron a este trabajo.

A mis amigos, principalmente a Ana Perez Cerecedo, con quien hice un gran equipo desde el inicio de la carrera. ¡Jamás olvidaré todas los momentos que vivimos juntos!

Finalmente agradezco a mi profesora Emma Lam Osnaya, quien desde primer semestre me motivó a ser un mejor estudiante, matemático y persona. ¡Nunca olvidaré tus clases, consejos e historias!

Índice general

Introducción.	4
1. Resultados Preliminares	7
1.1. Introducción a la estadística bayesiana	7
1.1.1. Función de verosimilitud	8
1.1.2. Distribuciones <i>a priori</i>	9
1.1.3. Distribución <i>a posteriori</i>	11
1.1.4. Familias conjugadas	13
1.1.5. Inferencia Bayesiana	15
1.2. Introducción a las cadenas de Markov y a los métodos Monte Carlo . . .	18
1.2.1. Cadenas de Markov	18
1.3. Integración Monte Carlo	27
1.4. Algoritmo de Metropolis-Hastings	29
2. Algunos conceptos básicos de series de tiempo	34
2.1. Definiciones y resultados básicos	34
2.2. Modelos de series de tiempo	38
2.2.1. Procesos ARMA(p, q)	39
2.3. Series de tiempo financieras	48
2.3.1. Hechos adicionales	50
2.3.2. Procesos ARCH(q)	52
2.3.3. Procesos GARCH(p, q)	59
3. Inferencia bayesiana en modelos GARCH($1,1$)	67
3.1. GARCH($1,1$) con innovaciones t-Student	67
3.2. Función de verosimilitud	69

3.3. Distribuciones <i>a priori</i> y <i>a posteriori</i>	71
4. Aplicación	75
4.1. Análisis exploratorio	75
4.2. Ajuste de modelos	78
4.3. Verificación del ajuste del modelo	83
4.4. Aplicación en riesgos financieros	86
Conclusión.	89
A. Definiciones y resultados preliminares	91
B. Pruebas estadísticas en el análisis de series de tiempo	94
B.1. Pruebas para autocorrelación	94
B.2. Prueba de Kolmogorov-Smirnov	96
B.3. Selección de modelos	96
C. Análisis de Convergencia	98
C.1. Criterio de Gelman-Rubin	99
C.2. Criterio de Geweke	100
D. Código de la aplicación	102

Introducción

En el ámbito económico los mercados financieros proporcionan una gran variedad de aplicaciones de la teoría de series de tiempo. Entre ellas destaca la modelación del comportamiento de activos financieros. Esto, debido a su impacto en inversiones y administración de riesgos financieros. En el análisis cuantitativo de esta clase de datos, uno de los conceptos de mayor interés es el de volatilidad, que se refiere a la variabilidad de la rentabilidad de un activo financiero. La importancia de su estudio subyace en la necesidad de perfeccionar diversas metodologías en finanzas matemáticas, por ejemplo, en la cuantificación del riesgo en inversiones mediante lo que es conocido como el *valor en riesgo* (VaR, por sus siglas en inglés).

En virtud del impacto de la volatilidad en el análisis de los mercados financieros, a finales del siglo XX se observó la presencia de ciertas características en rendimientos de activos financieros, como lo son el agrupamiento de la volatilidad y el efecto de leptocurtosis en algunas series de retornos. Esto motivó al desarrollo de modelos de series de tiempo que logaran capturar estos hechos. De esta manera en 1982, con el objetivo de estudiar la volatilidad de la inflación del Reino Unido, (Engle 1982) introdujo los modelos *autoregresivos condicionalmente heterocedásticos* (ARCH, por sus siglas en inglés). Posteriormente, con la finalidad de perfeccionar los procesos de Engle, surgieron variantes del modelo. Una de las extensiones más conocidas son los procesos *autoregresivos condicionalmente heterocedásticos generalizados* (GARCH, por sus siglas en inglés), introducidos por Bollerslev (1986), los cuales continúan vigentes para estudiar la volatilidad en activos financieros y portafolios de inversión.

Durante los últimos años han surgido diversas variantes del modelo GARCH para atender diferentes problemáticas que se encuentran en diversas aplicaciones de esta clase de datos.

En este trabajo se realizará un análisis de la volatilidad de una de las series con mayor relevancia en el sistema financiero mexicano, el tipo de cambio USD/MXN, así como su aplicación en medidas utilizadas en la administración de riesgos financieros. Se utilizará estadística bayesiana en la estimación de parámetros del modelo $\text{GARCH}(1,1)$, bajo la teoría desarrollada por Ardia (2008).

Más allá de las oportunidades que ofrece la estadística bayesiana, tal como el anexo de información extra-muestral en la estimación de parámetros, el uso de esta metodología es una alternativa para evitar algunos problemas existentes en la estimación clásica de parámetros para procesos condicionalmente heterocedásticos, los cuales se tornan aun más complejos en las diversas extensiones del modelo $\text{GARCH}(p, q)$ que se han desarrollado en los últimos años.

El presente trabajo se estructura de la siguiente forma. En el Capítulo 1 se presenta un breve resumen de la estadística bayesiana, incluyendo herramientas que se utilizarán para el desarrollo de los objetivos de esta tesis. Posteriormente, se presentará una introducción a la teoría de cadenas de Markov, así como la teoría referente al algoritmo de Metropolis-Hastings, uno de los métodos Monte Carlo vía cadenas de Markov que auxilia en la estimación bayesiana de parámetros. Una vez establecidos los conceptos y resultados a utilizar desde un enfoque bayesiano, en el Capítulo 2 se presentarán algunos conceptos básicos de la teoría de series de tiempo, donde nos enfocaremos en la teoría de los modelos $\text{GARCH}(p, q)$. En el Capítulo 3 tendremos las herramientas necesarias para desarrollar la teoría implícita en la inferencia bayesiana en un proceso $\text{GARCH}(1,1)$ con innovaciones t -Student, mediante el algoritmo de Metropolis-Hastings. En el Capítulo 4 se desarrollará un análisis exploratorio de los datos del tipo de cambio USD/MXN, esto para poder verificar la asignación de un modelo volatilidad y así analizar las consecuencias de su ajuste. Finalmente se dan las conclusiones del trabajo y un apéndice con algunos resultados utilizados a lo largo de esta tesis, así como el código del programa en el software R, utilizado para estimar los parámetros del modelo estudiado.

Algunas notaciones

$X \sim F$: Una variable aleatoria tiene distribución F .

$\mathbb{E}[X]$: La esperanza de una variable aleatoria X .

$\text{Var}[X]$: La varianza de una variable aleatoria X .

$\mathbb{E}[X^n]$: El n -ésimo momento de una variable aleatoria X .

$\text{Cov}(X, Y)$: La covarianza entre dos variables aleatorias X y Y .

$\text{Corr}(X, Y)$: La correlación entre dos variables aleatorias X y Y .

$\mathbb{E}[X|Y]$: La esperanza condicional de X dada la variable Y .

$\text{Var}[X|Y]$: La varianza condicional de X dada la variable Y .

$\mathbb{P}[A]$: La probabilidad de un evento A .

A^t : La matriz transpuesta de una matriz A .

$\det(A)$: El determinante de una matriz A .

$\frac{\partial f(x)}{\partial x}$: La derivada de $f(x)$ con respecto a x .

$\int_A f(x)dx$: Integral de $f(x)$ sobre un conjunto A .

$\lim_{x \rightarrow \infty} f(x)$: Límite cuando x tiende a infinito de $f(x)$.

$\max_{x \in A} f(x)$: Máximo de la función $f(x)$ sobre un conjunto A .

$\inf(A)$: Ínfimo de un conjunto A .

$\min(A)$: Mínimo de un conjunto A .

VaR_α : VaR diario para un nivel de significancia α .

Capítulo 1

Resultados Preliminares

En este primer capítulo se presenta un breve panorama de la estadística bayesiana, la cual se utilizará para el desarrollo de este trabajo. Primeramente se muestra una descripción de este enfoque, sus objetivos y diferencias con respecto a la estadística clásica. También se presentarán algunos resultados básicos de cadenas de Markov que serán de utilidad cuando los métodos Monte Carlo vía cadenas de Markov sean estudiados. Finalmente, se describe el algoritmo de Metropolis-Hastings que es utilizado en la estimación de parámetros presentes en el modelo considerado en esta tesis. El desarrollo de este capítulo utilizará como principal referencia los textos de Prado y West (2010) y Robert (2007).

1.1. Introducción a la estadística bayesiana

Uno de los principales objetivos de la estadística es explicar el comportamiento de un experimento aleatorio que afecta a una determinada población. Esto se realiza mediante el uso de diversos métodos y herramientas. En el caso de la estadística paramétrica, se utilizan modelos matemáticos que dependen de ciertos parámetros, los cuales se desean estimar para poder establecer conclusiones del fenómeno aleatorio en estudio.

Desde un enfoque clásico se considera un vector aleatorio \mathbf{X} , compuesto de variables aleatorias independientes e idénticamente distribuidas X_1, \dots, X_n , las cuales describen a una muestra aleatoria $\mathbf{x} = \{x_1, \dots, x_n\}$, extraída de la población que rige al experimento aleatorio en estudio. De esta forma, se asocia a \mathbf{X} una función de densidad conjunta $f(\mathbf{x} | \boldsymbol{\theta})$, donde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ es un vector de parámetros desconocidos, los cuales se

estiman con base en la muestra aleatoria \mathbf{x} considerada. A diferencia de este enfoque, en la estadística bayesiana, además de estudiar a una función $f(\mathbf{x} | \boldsymbol{\theta})$, se considera a $\boldsymbol{\theta}$ como un vector aleatorio, cuyo comportamiento puede ser descrito por una distribución de probabilidad. Este hecho permite que los expertos en el experimento aleatorio puedan proporcionar información extra-muestral acerca del comportamiento de los parámetros en el modelo propuesto para analizar al fenómeno de interés.

El hecho por el cual este enfoque se denomina bayesiano se deriva del objetivo de estudiar el vector de parámetros $\boldsymbol{\theta}$, utilizando su distribución y la información proporcionada por la muestra \mathbf{x} . Esto se hace a través del análisis de la distribución condicional del vector aleatorio $\boldsymbol{\theta}$ dada la muestra \mathbf{x} , cuya función de densidad se vincula con el Teorema de Bayes.

1.1.1. Función de verosimilitud

En un análisis bayesiano también se considera una distribución que representa la probabilidad que los datos provengan de un modelo con un vector de parámetros $\boldsymbol{\theta}$. Esta función se define como sigue.

Definición 1.1 *Dado un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ y $\mathbf{x} = (x_1, \dots, x_n)$ una realización de \mathbf{X} , la función de verosimilitud del modelo que describe \mathbf{X} y depende de un vector de parámetros $\boldsymbol{\theta}$, se define como:*

$$f(\mathbf{x} | \boldsymbol{\theta}) \propto f_{X_1, \dots, X_n | \boldsymbol{\theta}}(x_1, \dots, x_n | \boldsymbol{\theta}), \quad (1.1)$$

es decir, es la función de densidad de \mathbf{X} con un vector de parámetros $\boldsymbol{\theta}$, evaluada en \mathbf{x}

Observación. Si las variables X_1, \dots, X_n son independientes, se tiene que:

$$f(\mathbf{x} | \boldsymbol{\theta}) \propto \prod_{i=1}^n f_{X_i}(x_i | \boldsymbol{\theta}),$$

donde $f_{X_i}(\cdot | \boldsymbol{\theta})$ es la distribución condicional de X_i dado $\boldsymbol{\theta}$, $i = 1, \dots, n$. Si además las variables son idénticamente distribuidas, es decir, $f_{X_i}(\cdot | \boldsymbol{\theta}) = f(\cdot | \boldsymbol{\theta})$, entonces,

$$f(\mathbf{x} | \boldsymbol{\theta}) \propto \prod_{i=1}^n f(x_i | \boldsymbol{\theta}). \quad (1.2)$$

1.1.2. Distribuciones *a priori*

Como se ha mencionado, al considerar como variables aleatorias a los parámetros del modelo que se desea ajustar a una muestra \mathbf{x} , es posible asignar distintas distribuciones de probabilidad a $\boldsymbol{\theta}$.

Definición 1.2 Sea $\boldsymbol{\theta}$ el vector de parámetros desconocido del modelo que describe a una muestra aleatoria \mathbf{x} , la distribución asignada a $\boldsymbol{\theta}$, sin tomar en cuenta la información proporcionada por \mathbf{x} , se conoce como *distribución a priori* de $\boldsymbol{\theta}$ y se le denota por $p_{\text{priori}}(\boldsymbol{\theta})$.

La distribución *a priori* $p_{\text{priori}}(\boldsymbol{\theta})$ busca capturar información sobre el posible comportamiento del vector $\boldsymbol{\theta}$. Considerando que la selección de esta función depende del conocimiento extra-muestral del investigador sobre los parámetros presentes en el modelo propuesto para estudiar a \mathbf{x} , esta distribución no es única. Incluso, en ausencia de información inicial existen algunas metodologías para su desarrollo. Esto lleva a la siguiente clasificación de distribuciones *a priori*.

Distribuciones *a priori* informativas.

En caso de contar con información extra-muestral del fenómeno en estudio: ¿Cómo podemos asignar una distribución al vector de parámetros de tal forma que se capture este conocimiento? A continuación se presentan algunos métodos para incorporar información en una distribución *a priori* de un vector de parámetros $\boldsymbol{\theta}$.

1. **Elección de una distribución de probabilidad.** Se elige una distribución *a priori* de acuerdo a los valores que un parámetro puede asumir en el modelo propuesto para describir a \mathbf{x} . El conjunto donde un parámetro asume valores se denota por Θ y se conoce como espacio parametral; además será el soporte de la función de densidad *a priori*, o el conjunto donde esta tiene mayor peso. La distribución elegida también dependerá de ciertos parámetros conocidos como hiperparámetros. Además se desea que la elección de estos sea consistente con la información extra-muestral disponible.
2. **Probabilidad subjetiva.** Consiste en establecer una función de distribución acumulativa bajo cuantiles θ_α , es decir $\mathbb{P}[\theta < \theta_\alpha] = \alpha$, para $\alpha \in [0, 1]$.

La colección de los cuantiles recolectados determinaría la estructura para $\boldsymbol{\theta}$ de manera subjetiva. Si el parámetro a estimar asume valores en un intervalo de \mathbb{R} , también es posible aproximar un histograma, determinado por las probabilidades subjetivas asignadas por el investigador y relacionadas con el fenómeno de interés

Distribuciones *a priori* no informativas

En caso de no contar información de algún experto, se dice que la distribución *a priori* es **no informativa**. A pesar de esto, existen algunas metodologías para determinar el comportamiento de una distribución bajo estas condiciones. Las más comunes se presentan a continuación.

1. **Criterio de Laplace.** Considera que todos los posibles valores del parámetro tienen la misma probabilidad de ocurrir, es decir, se consideran distribuciones uniformes de acuerdo al espacio parametral Θ . Por ejemplo, si $Y \sim \text{Bin}(n, \theta)$, es decir, Y tiene distribución binomial de parámetros n y θ , donde θ es el parámetro desconocido que se desea estimar. Como los valores que asume θ se encuentran en el intervalo $\Theta = [0, 1]$, entonces se puede considerar $\theta \sim U(0, 1)$, es decir, θ se distribuye uniformemente en el intervalo $[0, 1]$.
2. **Método de Jeffreys.** Considera una distribución *a priori* proporcional a la raíz del determinante de la Información de Fisher, es decir,

$$p_{\text{priori}}(\boldsymbol{\theta}) \propto \sqrt{\det(I(\boldsymbol{\theta}))},$$

donde para un vector de parámetros $\boldsymbol{\theta}$ de tamaño m , la Información de Fisher se define como una matriz cuadrada de tamaño m , $I(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))_{m \times m}$, con entradas:

$$I_{ij}(\boldsymbol{\theta}) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(\mathbf{x} \mid \boldsymbol{\theta}) \right],$$

es decir, $-I_{ij}(\boldsymbol{\theta})$ corresponde a la esperanza de la segunda derivada parcial con respecto a θ_i y θ_j del logaritmo de la función de densidad de \mathbf{X} dado $\boldsymbol{\theta}$.

Observación. Si un parámetro θ asume valores en un subconjunto no acotado de los reales, el método Laplace genera distribuciones $p_{\text{priori}}(\theta)$ tal que:

$$\int_{\Theta} p_{\text{priori}}(\theta) d\theta = \infty,$$

si Θ es continuo, es decir, $\Theta \subset \mathbb{R}$, o

$$\sum_{\theta \in \Theta} p_{\text{priori}}(\theta) = \infty,$$

si Θ es discreto, es decir $\Theta \subset \mathbb{Z}$. Cabe resaltar que el uso de estas distribuciones es válido siempre que las distribuciones *a posteriori*, que se introducirán más adelante, no cuenten con esta característica.

1.1.3. Distribución *a posteriori*

Recordemos que se tiene como objetivo conocer el comportamiento de $\boldsymbol{\theta}$ dada la información proporcionada por los datos \mathbf{x} , es decir, buscamos determinar la distribución de $\boldsymbol{\theta}$ dada la muestra aleatoria \mathbf{x} . Por lo tanto, se considera la función definida a seguir.

Definición 1.3 *Sea $\boldsymbol{\theta}$ el vector de parámetros desconocido del modelo que describe a la muestra aleatoria \mathbf{x} . La distribución condicional de $\boldsymbol{\theta}$ dada la muestra aleatoria \mathbf{x} se conoce como distribución *a posteriori* y se denota por $p_{\text{post}}(\boldsymbol{\theta}) := p(\boldsymbol{\theta} | \mathbf{x})$.*

Para dar una expresión de la distribución *a posteriori* en términos de la función de verosimilitud y la distribución *a priori*, se utilizan las propiedades de las funciones de probabilidad/densidad conjuntas, así como la definición de probabilidad condicional.

Teorema 1.1 *La distribución *a posteriori* de un vector de parámetros $\boldsymbol{\theta}$ para una distribución continua de $\boldsymbol{\theta}$ satisface la siguiente proporción.*

$$p_{\text{post}}(\boldsymbol{\theta}) \propto f(\mathbf{x} | \boldsymbol{\theta})p_{\text{priori}}(\boldsymbol{\theta}). \quad (1.3)$$

Demostración. Por la definición de probabilidad condicional, aplicada a la distribución *a posteriori* de un vector de parámetros $\boldsymbol{\theta}$, se tiene que,

$$p_{\text{post}}(\boldsymbol{\theta}) = p(\boldsymbol{\theta} | \mathbf{x}) = \frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})}.$$

Considerando que $p(\mathbf{x})$ no depende de $\boldsymbol{\theta}$, se tiene que,

$$p_{\text{post}}(\boldsymbol{\theta}) \propto p(\mathbf{x}, \boldsymbol{\theta}). \quad (1.4)$$

De la definición de probabilidad condicional se sigue que,

$$p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x} | \boldsymbol{\theta})p(\boldsymbol{\theta}). \quad (1.5)$$

Sustituyendo (1.5) en (1.4),

$$p_{post}(\boldsymbol{\theta}) \propto p(\mathbf{x} | \boldsymbol{\theta})p(\boldsymbol{\theta}).$$

Considerando que una distribución *a priori* se define como una distribución del vector aleatorio $\boldsymbol{\theta}$ sin tomar en cuenta la información proporcionada por \mathbf{x} y que la función de verosimilitud es proporcional a la función de densidad del vector aleatorio \mathbf{X} condicionada al vector de parámetros $\boldsymbol{\theta}$, se sigue que,

$$p_{post}(\boldsymbol{\theta}) \propto f(\mathbf{x} | \boldsymbol{\theta})p_{priori}(\boldsymbol{\theta}).$$

■

Esta manera de tratar la información que proporcionan tanto el vector de parámetros $\boldsymbol{\theta}$, como nuestros datos \mathbf{x} , permite realizar inferencia sobre el fenómeno estudiado. Para ejemplificar estos conceptos bajo el supuesto que una muestra aleatoria \mathbf{X} sigue una distribución específica, se muestra el siguiente ejemplo.

Ejemplo 1.1. Suponga que se desea analizar el grado de aceptación de un gobernador en cierta entidad. Se extrae una muestra aleatoria de n ciudadanos, la cual se representará mediante variables aleatorias X_1, \dots, X_n . Podemos definir $X_i = 1$ si el i -ésimo individuo de la muestra aleatoria \mathbf{x} aprueba al gobernador y $X_i = 0$ en otro caso. También suponga que los individuos tienen opiniones independientes uno del otro.

De esta forma, se asume que las variables son idénticamente distribuidas con distribución Bernoulli(θ), donde θ es la proporción de aceptación del gobernador. Por ser θ un valor entre 0 y 1 se podría pensar en una distribución *a priori* uniforme en $[0, 1]$ o una distribución truncada sobre dicho intervalo. Suponga que con base en la experiencia del investigador se decide tomar $\theta \sim \text{Beta}(a, b)$.

Entonces, la función de densidad *a priori* de θ está dada como sigue,

$$p_{priori}(\theta) = \frac{1}{B(a, b)}\theta^{a-1}(1 - \theta)^{b-1}1_{[0,1]}(\theta),$$

donde la función $B(a, b)$ se define como:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)},$$

con $\Gamma(\cdot)$ la función gamma, dada por:

$$\Gamma(u) = \int_0^{\infty} t^{u-1} \exp(-t) dt.$$

Para construir la función de verosimilitud de la muestra aleatoria \mathbf{X} , se considera la función de densidad para cada X_i , $i = 1, \dots, n$, es decir,

$$f_{X_i}(x_i | \theta) \propto \theta^x (1 - \theta)^{1-x} 1_{\{0,1\}}(x_i).$$

Entonces, la función de verosimilitud esta dada como,

$$\begin{aligned} f(\mathbf{x} | \theta) &\propto \prod_{i=1}^n f(x_i | \theta) \\ &\propto \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} 1_{\{0,1\}}(x_i) \\ &= \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{\left(n - \sum_{i=1}^n x_i\right)} 1_{\{0,1\}}(x_i). \end{aligned}$$

Suponga que existen m ciudadanos, dentro de los n muestreados, que aprueban al gobernador, entonces, $\sum_{i=1}^n x_i = m$ y podemos escribir a la función de verosimilitud de la siguiente forma,

$$f(\mathbf{x} | \theta) \propto \theta^m (1 - \theta)^{n-m} 1_{\{0,1\}}(x_i).$$

De esta forma, por (1.3) la función de densidad *a posteriori*, está dada por,

$$\begin{aligned} p_{post}(\theta) &\propto f(\mathbf{x} | \theta) p_{priori}(\theta) \\ &= \theta^m (1 - \theta)^{n-m} \theta^{a-1} (1 - \theta)^{b-1} \\ &= \theta^{a+m-1} (1 - \theta)^{b+n-m-1}. \end{aligned}$$

De esta manera, la distribución *a posteriori* de θ es proporcional a una distribución Beta($a + m, b + n - m$).

1.1.4. Familias conjugadas

Con el fin de facilitar la obtención de la distribución *a posteriori* de un vector de parámetros $\boldsymbol{\theta}$, se busca establecer distribuciones de tal manera que tanto $p_{priori}(\boldsymbol{\theta})$ como $p_{post}(\boldsymbol{\theta})$ pertenezcan a la misma familia de distribuciones, tal como en el **Ejemplo 1.1**. De esta forma conoceremos de una más directa el comportamiento de la distribución *a posteriori* mediante una transformación en los hiperparámetros de $p_{priori}(\boldsymbol{\theta})$.

Definición 1.4 Una familia $F = \{f(\cdot | \boldsymbol{\theta}) | \boldsymbol{\theta} \in \Theta\}$ de distribuciones de probabilidad es conjugada para una función verosimilitud $f(\mathbf{x} | \boldsymbol{\theta})$, si para cada distribución a priori $p_{\text{priori}}(\boldsymbol{\theta}) \in F$, la distribución a posteriori $p_{\text{post}}(\boldsymbol{\theta}) \in F$.

Observación. A partir del **Ejemplo 1.1**, se observa que la familia de distribuciones beta es conjugada para una verosimilitud de una muestra aleatoria con distribución Bernoulli.

Una familia de distribuciones de probabilidad que nos puede ayudar a realizar un análisis de distribuciones a *posteriori* mediante la **Definición 1.4** es la familia exponencial.

Definición 1.5 Sea $\boldsymbol{\theta}$ un vector de m parámetros. Se dice que la función de densidad de una variable aleatoria X pertenece a la familia exponencial de dimensión m , si puede escribirse de la siguiente forma,

$$f(x | \boldsymbol{\theta}) = h(x) \exp [\boldsymbol{\eta}(\boldsymbol{\theta})^t T(x) - A(\boldsymbol{\theta})], \quad (1.6)$$

donde $h(x)$ y $A(\boldsymbol{\theta})$ son funciones con imágenes en \mathbb{R} y que dependen únicamente de x y $\boldsymbol{\theta}$ respectivamente, $T(x)$ es una transformación que depende de x , con imagen en un subconjunto de \mathbb{R}^m y $\boldsymbol{\eta}(\boldsymbol{\theta})^t = (\eta_1(\boldsymbol{\theta}), \dots, \eta_m(\boldsymbol{\theta}))$, donde $\eta_i(\boldsymbol{\theta})$ es una función del vector $\boldsymbol{\theta}$, cuya imagen asume valores en \mathbb{R} para $i = 1, \dots, m$.

Observación. Para facilitar el estudio de ciertos resultados de la familia exponencial, como el que se mostrará a continuación, la ecuación (1.6) puede verse como:

$$f(x | \boldsymbol{\theta}) = h(x)g(\boldsymbol{\theta}) \exp [\boldsymbol{\eta}(\boldsymbol{\theta})^t T(x)], \quad (1.7)$$

donde $g(\boldsymbol{\theta})$ es función del vector de parámetros y cuya imagen asume valores en \mathbb{R} .

Proposición 1.1 La familia exponencial es conjugada para una función de verosimilitud de la misma familia.

Demostración. Sea $\mathbf{x} = \{x_1, \dots, x_n\}$ una muestra aleatoria de n variables aleatorias independientes e idénticamente distribuidas en la familia exponencial. La función de verosimilitud estará dada como sigue:

$$\begin{aligned} f(\mathbf{x} | \boldsymbol{\theta}) &\propto f_{X_1, \dots, X_n | \boldsymbol{\theta}}(x_1, \dots, x_n | \boldsymbol{\theta}) \\ &= \prod_{i=1}^n f(x_i | \boldsymbol{\theta}). \end{aligned}$$

$$\begin{aligned}
f(\mathbf{x} \mid \boldsymbol{\theta}) &= \prod_{i=1}^n h(x_i) g(\boldsymbol{\theta}) \exp(\boldsymbol{\eta}(\boldsymbol{\theta})^t T(x_i)) \\
&= g(\boldsymbol{\theta})^n \left[\prod_{i=1}^n h(x_i) \right] \exp \left(\boldsymbol{\eta}(\boldsymbol{\theta})^t \sum_{i=1}^n T(x_i) \right).
\end{aligned}$$

Considerando a $\boldsymbol{\theta}$ como un vector aleatorio de hiperparámetros conocidos $\boldsymbol{\delta} = (\delta_1, \dots, \delta_l)$. Se elige una distribución a priori cuya forma sea parte de la familia exponencial. Particularmente, se considera la siguiente estructura,

$$\begin{aligned}
p_{\text{priori}}(\boldsymbol{\theta}) &= g(\boldsymbol{\theta}) m(\boldsymbol{\delta}) \exp[\boldsymbol{\eta}(\boldsymbol{\theta})^t \boldsymbol{\delta}] \\
&\propto g(\boldsymbol{\theta}) \exp[\boldsymbol{\eta}(\boldsymbol{\theta})^t \boldsymbol{\delta}],
\end{aligned}$$

donde se utiliza las mismas funciones $g(\boldsymbol{\theta})$ y $\boldsymbol{\eta}(\boldsymbol{\theta})^t$ tanto para la distribución inicial, como la verosimilitud del modelo. Además $m(\boldsymbol{\delta})$ es una función que sólo depende del vector de hiperparámetros.

Por lo tanto, utilizando (1.3), la distribución *a posteriori* del vector de parámetros $\boldsymbol{\theta}$ estará dada por:

$$\begin{aligned}
p_{\text{post}}(\boldsymbol{\theta}) &\propto f(\mathbf{x} \mid \boldsymbol{\theta}) p_{\text{priori}}(\boldsymbol{\theta}) \\
&\propto g(\boldsymbol{\theta})^n \left[\prod_{i=1}^n h(x_i) \right] \exp \left(\boldsymbol{\eta}(\boldsymbol{\theta})^t \sum_{i=1}^n T(x_i) \right) g(\boldsymbol{\theta}) \exp(\boldsymbol{\eta}(\boldsymbol{\theta})^t \boldsymbol{\delta}) \\
&= g(\boldsymbol{\theta})^{n+1} \left[\prod_{i=1}^n h(x_i) \right] \exp \left[\boldsymbol{\eta}(\boldsymbol{\theta})^t \left(\sum_{i=1}^n T(x_i) + \boldsymbol{\delta} \right) \right].
\end{aligned}$$

Se observa que la función de densidad de probabilidad de la distribución a posteriori también posee la forma de la ecuación (1.7). Esto implica que $p_{\text{post}}(\boldsymbol{\theta})$ también pertenece a la familia exponencial. ■

1.1.5. Inferencia Bayesiana

Una vez establecido el comportamiento de la distribución *a posteriori*, se posee información relevante de $\boldsymbol{\theta}$ dada la muestra \mathbf{x} . En el caso univariado, para la elección de un estimador puntual de un parámetro θ , se elige alguna característica de $p_{\text{post}}(\theta)$ tal como la media, la mediana o la moda.

Definición 1.6 *El estimador maximum a posteriori de θ es aquél que maximiza la distribución a posteriori, es decir,*

$$\hat{\theta}_{MAP} = \underset{\theta \in \Theta}{\text{máx}} p_{post}(\theta).$$

La elección de la moda de la distribución *a posteriori* como estimador de θ correspondería al método de máxima verosimilitud en la estadística clásica. Se introducen las siguientes definiciones a partir de la distribución *a posteriori* de un parámetro θ .

Definición 1.7 *Sea θ el parámetro de un modelo con distribución a posteriori $p_{post}(\theta)$.*

(a) *La media a posteriori se define como la siguiente esperanza,*

$$\mathbb{E}_{post}[\theta] = \int_{\Theta} \theta p_{post}(\theta) d\theta. \quad (1.8)$$

(b) *La varianza a posteriori se define como la siguiente varianza,*

$$Var_{post}[\theta] = \int_{\Theta} (\theta - \mathbb{E}_{post}[\theta])^2 p_{post}(\theta) d\theta. \quad (1.9)$$

En la estimación puntual de un parámetro, surge el interés de conocer su precisión. Para el caso univariado, utilizaremos el valor esperado de la diferencia entre la estimación y el valor real del parámetro. Se desea que ésta sea mínima.

Definición 1.8 *Dado un modelo que depende de un parámetro θ , si $\hat{\theta}(\mathbf{x})$ es un estimador para θ , obtenido usando la información por una muestra \mathbf{x} , el error cuadrático medio a posteriori con respecto al estimador $\hat{\theta}(\mathbf{x})$ está dado por:*

$$ECM_{post}[\hat{\theta}(\mathbf{x})] = \mathbb{E}_{post}[(\theta - \hat{\theta}(\mathbf{x}))^2 | \mathbf{X} = \mathbf{x}], \quad (1.10)$$

Al considerar un estimador igual a la media a posteriori del parámetro θ , este recibe el nombre de estimador por mínimo error cuadrático medio, (MMSE, por sus siglas en inglés). Esto debido al siguiente resultado.

Teorema 1.2 *Sea θ un parámetro de un modelo que describe a una muestra \mathbf{x} . El estimador $\hat{\theta}(\mathbf{x})$ que minimiza al error cuadrático medio a posteriori corresponde a la media a posteriori de θ .*

Demostración. De (1.10) usando la notación $\mu = \mathbb{E}_{post}[\theta]$, se tiene.

$$\begin{aligned} ECM_{post}[\hat{\theta}(\mathbf{x})] &= \mathbb{E}_{post}[(\theta - \mu + \mu - \hat{\theta}(\mathbf{x}))^2 | \mathbf{x}] \\ &= \mathbb{E}_{post}[(\theta - \mu)^2 + 2(\theta - \mu)(\mu - \hat{\theta}(\mathbf{x})) + (\mu - \hat{\theta}(\mathbf{x}))^2 | \mathbf{x}]. \end{aligned}$$

Aplicando las propiedades de esperanza, tenemos,

$$\begin{aligned}\text{ECM}_{post}[\hat{\theta}(\mathbf{x})] &= \mathbb{E}_{post}[(\theta - \mu)^2 | \mathbf{x}] + 2(\mu - \mu)(\mu - \hat{\theta}(\mathbf{x})) + (\mu - \hat{\theta}(\mathbf{x}))^2 \\ &= \mathbb{E}_{post}[(\theta - \mu)^2 | \mathbf{x}] + (\mu - \hat{\theta}(\mathbf{x}))^2.\end{aligned}$$

Por definición de varianza *a posteriori*, obtenemos,

$$\text{ECM}_{post}[\hat{\theta}(\mathbf{x})] = \text{Var}_{post}[\theta] + (\mathbb{E}_{post}[\theta] - \hat{\theta}(\mathbf{x}))^2. \quad (1.11)$$

Para encontrar el estimador $\hat{\theta}(\mathbf{x})$ que minimiza al error cuadrático medio, derivamos (1.11) con respecto a $\hat{\theta}(\mathbf{x})$, es decir,

$$\frac{d}{d\hat{\theta}(\mathbf{x})} \text{ECM}_{post}[(\theta - \hat{\theta}(\mathbf{x}))^2 | \mathbf{x}] = -2(\mathbb{E}_{post}[\theta] - \hat{\theta}(\mathbf{x})).$$

Igualando la derivada cero para buscar los puntos críticos se obtiene que $\hat{\theta}(\mathbf{x}) = \mathbb{E}_{post}[\theta]$.

Además se verifica que este valor alcanza un mínimo porque:

$$\frac{d^2}{d\hat{\theta}(\mathbf{x})^2} \text{ECM}_{post}[(\theta - \hat{\theta}(\mathbf{x}))^2 | \mathbf{x}] = 2 > 0.$$

De esta manera, $\hat{\theta}(\mathbf{x}) = \mathbb{E}_{post}[\theta]$ minimiza al error cuadrático medio, el cual toma el valor de la varianza *a posteriori*.

$$\mathbb{E}_{post}[(\theta - \hat{\theta}(\mathbf{x}))^2 | \mathbf{x}] = \text{Var}_{post}[\theta].$$

■

Observaciones.

1. En el caso de un vector aleatorio de parámetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$, si cada uno de los parámetros posee media *a posteriori* finita, entonces la media a posterior de θ será un vector de tamaño m con entradas reales.

$$\mathbb{E}_{post}[\boldsymbol{\theta}] = (\mathbb{E}_{post}[\theta_1], \dots, \mathbb{E}_{post}[\theta_n]),$$

donde la precisión puede determinarse a través de la matriz de varianzas y covarianzas *a posteriori*, dada por,

$$\mathbb{E}[(\boldsymbol{\theta} - \mathbb{E}_{post}[\boldsymbol{\theta}])(\boldsymbol{\theta} - \mathbb{E}_{post}[\boldsymbol{\theta}])^t].$$

Análogamente al caso unidimensional, se verifica que la media a posteriori del vector de parámetros $\boldsymbol{\theta}$, minimiza la matriz de varianzas y covarianzas *a posteriori*.

2. Como se ha mencionado, se puede usar la media de la distribución posteriori como su estimador. Sin embargo, hay distribuciones *a posteriori* con estructuras complicadas o que no poseen una expresión cerrada, entonces la tarea de estimar la media de este tipo de distribuciones puede ser fácilmente realizada a través de los métodos Monte Carlo vía cadenas de Markov, tema que se abordará en la siguiente sección.

1.2. Introducción a las cadenas de Markov y a los métodos Monte Carlo

En esta sección se realiza una breve exposición de algunos aspectos teóricos de las cadenas de Markov y de los métodos Monte Carlo vía cadenas de Markov (MCMC). Estos son una herramienta que auxilian en la estimación de un vector de parámetros a partir de su distribución *a posteriori*. Primeramente se presentan algunos resultados de interés acerca de la teoría de cadenas de Markov. Después se desarrolla una breve descripción de la integración Monte Carlo en la estimación bayesiana de parámetros. Finalmente se presenta el algoritmo de Metropolis-Hastings, el cual se utilizará en el desarrollo de este trabajo. Para la sección de cadenas de Markov se utilizará como principal referencia a Grimmett y Stirzaker (2001), el resto se basará en los textos de Robert y Casella (2010) y Rodrigues (2013).

1.2.1. Cadenas de Markov

Una de las principales definiciones que se utilizarán en los próximos capítulos es la de proceso estocástico. Estos conforman estructuras de variables aleatorias que evolucionan en función de un conjunto de índices, generalmente representando el tiempo.

Definición 1.9 *Sea T un conjunto de índices. Un **proceso estocástico** es una colección de variables aleatorias $X = \{X_t : t \in T\}$, definidas sobre un espacio muestral común y que además asumen valores en un mismo conjunto S llamado espacio de estados.*

Las cadenas de Markov conforman uno de los ejemplos más conocidos en la teoría de procesos estocásticos debido a sus diversas aplicaciones. Este tipo de procesos fue desarrollado por Markov (1906). La definición formal de este tipo de proceso estocástico se presenta a seguir.

Definición 1.10 Sea $X = \{X_n : n \in T\}$ un proceso estocástico con espacio de estados S . El proceso X es una cadena de Markov discreta si $T \subset \mathbb{Z}$ y para instantes $t_0 < t_1 < \dots < t_{n-2} < s < t \in T$ satiface la propiedad de Markov, es decir,

$$\mathbb{P}[X_t = j \mid X_{t_0} = x_0, \dots, X_{t_{n-2}} = x_{n-2}, X_s = i] = \mathbb{P}[X_t = j \mid X_s = i], \quad (1.12)$$

donde $x_0, \dots, x_{n-2}, i, j \in S$.

Observación. La propiedad de Markov (1.12) es equivalente a las siguientes igualdades. (ver Grimmett y Stirzaker, 2001)

(a) Para $n \in \mathbb{N}$ y $x_k, k = 0, \dots, n-2, i, j \in S$,

$$\mathbb{P}[X_n = j \mid X_0 = x_0, \dots, X_{n-2} = x_{n-2}, X_{n-1} = i] = \mathbb{P}[X_n = j \mid X_{n-1} = i]. \quad (1.13)$$

(b) Para $m, n \in \mathbb{N}$ y $x_k, k = 0, \dots, n-1, i, j \in S$,

$$\mathbb{P}[X_{n+m} = j \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = i] = \mathbb{P}[X_{n+m} = j \mid X_n = i]. \quad (1.14)$$

Probabilidades de transición

En lo que sigue, se utilizará la equivalencia en (1.14) para referirnos a la propiedad de Markov. De esta igualdad se derivan las siguientes definiciones.

Definición 1.11 Sea $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S . La probabilidad de transición del estado $i \in S$ en un instante n , al estado $j \in S$ al tiempo $n+m$, $n, m \in T$ se define como,

$$p_{i,j}^{(n,n+m)} = \mathbb{P}[X_{n+m} = j \mid X_n = i]. \quad (1.15)$$

Observación. La probabilidad de transición $p_{i,j}^{(n,n+m)}$ también es llamada probabilidad de transición en m pasos, $n, m \geq 0, i, j \in S$.

Definición 1.12 Sea cadena de Markov $X = \{X_n : n \in T\}$ con espacio de estados S . Se dice que X es homogénea en el tiempo si la probabilidad de transición $p_{i,j}^{(n,n+m)}$ no depende de n .

Notación. Para una cadena de Markov homogénea en el tiempo, la probabilidad de transición en m pasos se denotará por $p_{i,j}^{(m)}$, $i, j \in S, m \geq 0$. Cuando $m = 0$, se tiene,

$$p_{i,j}^{(0)} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Para $m = 1$, $p_{i,j}^{(1)} = p_{i,j}$ y simplemente se le llamará como la probabilidad de transición del estado $i \in S$ al estado $j \in S$.

En lo que sigue consideramos solamente cadenas de Markov homogéneas en el tiempo.

Definición 1.13 Sea $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S . La matriz de transición en $m \geq 0$ pasos se define como:

$$P^{(m)} := \left(p_{i,j}^{(m)} \right)_{i,j \in S}.$$

Para $m = 1$ se indica $P := P^{(1)}$ y se le conocerá como matriz de transición de la cadena.

A continuación se presentan algunas propiedades que satisfacen las probabilidades (1.15) y que serán de nuestro interés en resultados posteriores.

Proposición 1.2 Sea X una cadena de Markov con espacio de estados S y matriz de transición en $m \geq 0$ pasos $P^{(m)} = \left(p_{i,j}^{(m)} \right)_{i,j \in S}$. Entonces para $m \in \mathbb{N}$, la matriz $P^{(m)}$ satisface las siguientes condiciones:

- (i) Para cada $i, j \in S$, se tiene $p_{i,j}^{(m)} \geq 0$.
- (ii) Para cada $i \in S$ se cumple:

$$\sum_{j \in S} p_{i,j}^{(m)} = 1.$$

Demostración. El punto (i) es inmediato por definición, las entradas $p_{i,j}^{(m)}$ son probabilidades condicionales, la cuales son mayores o iguales a cero.

Para el punto (ii) se considera que para $m \in T$, las variables aleatorias X_m , que representan el instante m del proceso, están definidas sobre el mismo espacio muestral Ω . La familia de eventos $\mathfrak{S} = \{A_j : j \in S\}$ donde $A_j := \{X_m = j\}$, descompone a Ω en eventos disjuntos dos a dos, es decir,

$$\Omega = \bigcup_{j \in S} A_j,$$

donde $A_i \cap A_j = \emptyset$ para $i \neq j$. De esta forma, \mathfrak{S} forma una partición del espacio muestral.

De esta forma, considerando que la probabilidad condicional cumple los axiomas de

Kolmogorov de una función de probabilidad (ver Ross, 2009), para cada $i \in S$, el espacio muestral Ω satisface.

$$1 = \mathbb{P}[\Omega \mid X_0 = i] = \mathbb{P}\left[\bigcup_{j \in S} \{X_m = j\} \mid X_0 = i\right].$$

Dado que los eventos son disjuntos dos a dos, es decir $A_i \cap A_j = \emptyset$ si $i \neq j$, entonces,

$$\begin{aligned} 1 &= \sum_{j \in S} \mathbb{P}[X_m = j \mid X_0 = i] \\ &= \sum_{j \in S} p_{i,j}^{(m)}. \end{aligned}$$

■

El siguiente resultado permite expresar una probabilidad de transición en $m \geq 0$ pasos en términos de otras para un instante intermedio $k < m$. Este teorema facilitará la demostración de resultados posteriores.

Teorema 1.3 (Ecuación de Chapman-Kolmogorov) Sea $X = \{X_n, n \in T\}$ una cadena de Markov con espacio de estados S y matriz de transición en m pasos $P^{(m)}$. Para $m, k \in \mathbb{N}$ tal que $k < m$, $i, j \in S$, $m \geq 0$ se cumple:

$$p_{i,j}^{(m)} = \sum_{s \in S} p_{i,s}^{(k)} p_{s,j}^{(m-k)}. \quad (1.16)$$

Demostración. Por (1.15) y la definición de probabilidad condicional, se puede escribir,

$$\begin{aligned} p_{i,j}^{(m)} &= \mathbb{P}[X_m = j \mid X_0 = i] \\ &= \frac{P[X_m = j, X_0 = i]}{\mathbb{P}[X_0 = i]}. \end{aligned}$$

Por el teorema de probabilidad total (ver Ross, 2009), se tiene

$$p_{i,j}^{(m)} = \sum_{s \in S} \frac{P[X_m = j, X_k = s, X_0 = i]}{\mathbb{P}[X_0 = i]}.$$

Multiplicamos y dividimos entre $\mathbb{P}[X_k = s, X_0 = i] > 0$,

$$\begin{aligned} p_{i,j}^{(m)} &= \sum_{s \in S} \frac{P[X_m = j, X_k = s, X_0 = i]}{\mathbb{P}[X_k = s, X_0 = i]} \frac{\mathbb{P}[X_k = s, X_0 = i]}{\mathbb{P}[X_0 = i]} \\ &= \sum_{s \in S} P[X_m = j \mid X_k = s, X_0 = i] \mathbb{P}[X_k = s \mid X_0 = i]. \end{aligned}$$

Aplicando la propiedad de Markov a $P[X_m = j \mid X_k = s, X_0 = i]$ se sigue:

$$\begin{aligned} p_{i,j}^{(m)} &= \sum_{s \in S} \mathbb{P}[X_k = s \mid X_0 = i] P[X_m = j \mid X_k = s] \\ &= \sum_{s \in S} p_{i,s}^{(k)} p_{s,j}^{(m-k)}. \end{aligned}$$

■

Clasificación de estados

Los estados de una cadena de Markov pueden ser divididos de acuerdo a sus propiedades. Para poder establecer esta clasificación se introducen las siguientes definiciones.

Definición 1.14 Sea $X = \{X_n : n \geq 0\}$ una cadena de Markov con espacio de estados S y matriz de transición en m pasos $P^{(m)} = (p_{i,j}^{(m)})_{i,j \in S}$. Se dice que un estado $j \in S$ es accesible desde un estado $i \in S$, indicado por $i \rightarrow j$, si existe un $n \in \mathbb{N}$ tal que:

$$p_{i,j}^{(n)} > 0.$$

Definición 1.15 Sea S el espacio de una cadena de Markov. Los estados $i, j \in S$ se comunican si $i \rightarrow j$ y $j \rightarrow i$. Esta propiedad se denota por $i \leftrightarrow j$.

Definición 1.16 Una cadena de Markov $X = \{X_n : n \geq 0\}$ con espacio de estados S es irreducible si para cualesquiera $i, j \in S$, los estados i y j se comunican.

Definición 1.17 Sea $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S y matriz de transición en m pasos $P^{(m)} = (p_{i,j}^{(m)})_{i,j \in S}$. Para $i \in S$, el periodo de i se define como:

$$d(i) = \text{m.c.d} \{n > 0 : p_{i,i}^{(n)} > 0\},$$

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

Observación. Si $i \in S$, es tal que $p_{i,i}^{(n)} = 0$ para toda $n > 0$, entonces $d(i) = 0$.

Definición 1.18 Para una cadena de Markov $X = \{X_n : n \in T\}$ con espacio de estados S , un estado $i \in S$ es aperiódico si $d(i) = 1$. La cadena X es aperiódica si todos sus estados son aperiódicos.

Definición 1.19 Sea $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S y matriz de transición en m pasos $P^{(m)} = (p_{i,j}^{(m)})_{i,j \in S}$. El tiempo de la primera visita al estado $j \in S$ desde el estado $i \in S$ se describe mediante la siguiente variable aleatoria,

$$T_{i,j} = \begin{cases} \min_{n \in \mathbb{N}} \{X_n = j \mid X_0 = i\} & X_n = j, n > 0 \\ \infty & X_n \neq j, n > 0. \end{cases} \quad (1.17)$$

Observación. Para T_{ij} tal que $i = j$ se denotará $T_{ii} = T_i$. A la función de probabilidad de esta variable aleatoria se denota por: $f_{i,j}(n) = \mathbb{P}[T_{i,j} = n]$.

Definición 1.20 Sea $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S . La probabilidad de una eventual visita al estado $j \in S$ a partir del estado $i \in S$ es:

$$f_{i,j} = \sum_{n \in \mathbb{N}} f_{i,j}(n). \quad (1.18)$$

Definición 1.21 Sean $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S y un estado $i \in S$:

- (i) El estado i es recurrente si y sólo si $f_{ii} = 1$.
- (ii) El estado i es transitorio si y sólo si $f_{ii} < 1$.

De manera intuitiva, un estado es recurrente si existe la seguridad de regresar a él en algún periodo de tiempo finito. Un estado es transitorio si existe una probabilidad positiva de no regresar a él.

Definición 1.22 Sean $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S y $i, j \in S$. El tiempo esperado de recurrencia $\mu_{i,j}$ de un estado recurrente j a partir del estado i se define como:

$$\mu_{i,j} = \mathbb{E}[T_{i,j}] = \sum_{n \in \mathbb{N}} n f_{i,j}(n).$$

Particularmente, para $i = j$, si $f_{ii} = 1$, se tiene,

$$\mu_i = \mathbb{E}[T_i] = \sum_{n \in \mathbb{N}} n f_{ii}(n). \quad (1.19)$$

Definición 1.23 Sea $i \in S$ un estado recurrente. Se dice que:

- (i) i es recurrente nulo si $\mu_i = \infty$,
- (ii) i es recurrente positivo si $\mu_i < \infty$.

Definición 1.24 Una cadena de Markov $X = \{X_n : n \in T\}$ es recurrente positiva si para cualquier $i \in S$, el estado i es recurrente positivo.

Cadenas ergódicas

Una de las definiciones con mayor relevancia en los métodos Monte Carlo vía cadenas de Markov es la siguiente.

Definición 1.25 Sea $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S . X es ergódica si es irreducible, aperiódica y es recurrente positiva.

Si S es un espacio de estados finito, la **Definición 1.25** puede simplificarse. Para ello presentaremos los siguientes resultados.

Proposición 1.3 Sea S el espacio de estados de una cadena de Markov. Sean $i, j \in S$ tal que $i \longleftrightarrow j$, entonces,

- (i) Si i es recurrente positivo, entonces j también lo es.
- (ii) Si i es aperiódico, entonces j también lo es

Demostración. Ver el desarrollo en Hoel *et. al* (1972).

Proposición 1.4 Sea $X = \{X_t : t \in T\}$ una cadena de Markov con espacio de estados S . Si S es finito entonces existe al menos un estado recurrente positivo.

Demostración. Ver el desarrollo en Ross (1996).

Observación. Si $X = \{X_n : n \in T\}$ es una cadena con espacio de estados finito S , en la **Definición 1.25** para que X sea ergódica sólo se requiere que esta sea irreducible y aperiódica, ya que por la **Proposición 1.4** existe un estado recurrente positivo $i \in S$ y dado que la cadena es irreducible, para cualquier $j \in S$ se cumple que $i \longleftrightarrow j$. Por lo tanto, por la **Proposición 1.3**, la cadena también es recurrente positiva.

Definición 1.26 La distribución estacionaria de una cadena de Markov $X = \{X_n : n \in T\}$ con matriz de transición $P = (p_{i,j})_{i,j \in S}$ y espacio de estados S es un conjunto de números reales positivos $\{\pi(j) : j \in S\}$ que satisfacen las siguientes condiciones:

- (a) $\sum_{j \in S} \pi(j) = 1$,
- (b) Para toda $j \in S$ se cumple

$$\pi(j) = \sum_{i \in S} \pi(i)p_{i,j}.$$

Las igualdades (a) y (b) de la **Definición 1.26** se llaman ecuaciones de *balance total*. Uno de los resultados que auxilia en determinar el comportamiento de una solución al sistema de ecuación que generan, se presenta a continuación.

Proposición 1.5 Sea $X = \{X_n : n \in T\}$ una cadena de Markov con espacio de estados S y matriz de transición en m pasos $P^{(m)} = \left(p_{i,j}^{(m)}\right)_{i,j \in S}$. Si $\mathbb{P}[X_0 = j] = \pi(j)$ entonces para toda $n \in \mathbb{N}$ se satisface que $\mathbb{P}[X_n = j] = \pi(j)$.

Demostración. Se realiza inducción matemática para verificar este hecho. Para $n = 1$, se toma $j \in S$ y por el teorema de probabilidad total (ver Ross, 2009) y la **Definición 1.11**, se tiene

$$\begin{aligned}\mathbb{P}[X_1 = j] &= \sum_{i \in S} \mathbb{P}[X_1 = j | X_0 = i] \mathbb{P}[X_0 = i] \\ &= \sum_{i \in S} p_{ij} \mathbb{P}[X_0 = i].\end{aligned}$$

Por hipótesis $\mathbb{P}[X_0 = j] = \pi(j)$, donde $\pi(j)$ es una distribución estacionaria. Así,

$$\begin{aligned}\mathbb{P}[X_1 = j] &= \sum_{i \in S} p_{ij} \pi(i) \\ &= \pi(j).\end{aligned}$$

Supongamos que el resultado es correcto para $n = m$. Por demostrar que el caso $n = m+1$ es válido. Por el teorema de probabilidad total y la **Definición 1.11** de probabilidad de transición en m pasos, tenemos que,

$$\begin{aligned}\mathbb{P}[X_{m+1} = j] &= \sum_{i \in S} \mathbb{P}[X_{m+1} = j | X_m = i] \mathbb{P}[X_m = i] \\ &= \sum_{i \in S} p_{ij} \mathbb{P}[X_m = i].\end{aligned}$$

Por hipótesis de inducción $\mathbb{P}[X_m = i]$ es igual a $\pi(i)$. Por lo tanto:

$$\begin{aligned}\mathbb{P}[X_{m+1} = j] &= \sum_{i \in S} p_{ij} \pi(i) \\ &= \pi(j).\end{aligned}$$

Por lo tanto, para toda $n \in \mathbb{N}$ se satisface que $\mathbb{P}[X_n = j] = \pi(j)$. ■

Definición 1.27 La distribución límite de una cadena de Markov $X = \{X_n : n \in T\}$ con espacio de estados S y matriz de transición en m pasos $P^{(m)} = \left(p_{i,j}^{(m)}\right)_{i,j \in S}$, se define como el siguiente límite:

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)}, \quad (1.20)$$

cuando este existe y es independiente de $i \in S$.

Es de interés saber cuando (1.20) cumple las condiciones para ser considerado como la distribución límite de una cadena de Markov. Ross (1996) nos ofrece el siguiente resultado.

Proposición 1.6 Sean $i, j \in S$ dos estados de una cadena de Markov con matriz de transición $P^{(m)} = \left(p_{i,j}^{(m)} \right)_{i,j \in S}$, si $i \longleftrightarrow j$ y j es aperiódico, entonces.

$$\lim_{n \rightarrow \infty} p_{i,j}^{(n)} = \frac{1}{\mu_j}. \quad (1.21)$$

Demostración. Ver, por ejemplo, Ross (1996).

Observación. De la **Proposición 1.6**, se puede concluir que si una cadena de Markov $X = \{X_n : n \in T\}$ es irreducible y aperiódica entonces la distribución límite del proceso X coincide con (1.21) para cualesquiera $i, j \in S$.

Uno de los resultados en los que se inspiran los métodos Monte Carlo vía cadenas de Markov es el siguiente y que relaciona los conceptos de distribución límite y estacionaria es el siguiente.

Teorema 1.4 Si $X = \{X_n : n \in T\}$ una cadena de Markov **ergódica** con espacio de estados S y matriz de transición en m pasos $P^{(m)} = \left(p_{i,j}^{(m)} \right)_{i,j \in S}$, entonces para cualesquiera $i, j \in S$, la distribución límite existe y corresponde a la única distribución estacionaria.

Demostración. La demostración de este hecho puede encontrarse en Ross (1996) o en Grimmett y Strizaker (2001).

El siguiente resultado nos permite establecer de manera más detallada la distribución estacionaria de una cadena de Markov y que además auxiliará en la construcción de métodos Monte Carlo vía Cadenas de Markov.

Proposición 1.7 Sea $X = \{X_n : n \in \mathbb{N}\}$ una cadena de Markov tal que $\mathbb{P}[X_0 = j]$ es igual a su distribución estacionaria $\pi(j)$. Si $P = (p_{i,j})_{i,j \in S}$ es la matriz de transición de X y $f(\cdot)$ es una función de probabilidad con soporte S que satisface:

$$f(i)p_{ij} = f(j)p_{ji},$$

entonces $f(\cdot)$ es la distribución estacionaria de la cadena X .

Demostración. Se deben verificar las condiciones de la **Definición 1.26**. Por la definición de función de probabilidad, el conjunto $\{f(j) : j \in S\}$ está compuesto de valores positivos tales que:

$$\sum_{j \in S} f(j) = 1.$$

Para verificar el punto (b) de la definición, se considera $j \in S$ fijo y se realiza una suma sobre $i \in S$ en ambas partes de la igualdad $f(i)p_{i,j} = f(j)p_{i,j}$, es decir,

$$\begin{aligned} \sum_{i \in S} f(i)p_{i,j} &= \sum_{i \in S} f(j)p_{ji} \\ &= f(j) \sum_{i \in S} p_{ji}. \end{aligned}$$

Bajo el punto (ii) de la **Proposición 1.2** se concluye que:

$$f(j) = \sum_{i \in S} f(i)p_{ji}.$$

■

1.3. Integración Monte Carlo

En la **Sección 1.1.5** se realizó una breve descripción de la estimación de un vector de parámetros $\boldsymbol{\theta}$ desde la estadística bayesiana, donde se introdujo la distribución *a posteriori* de $\boldsymbol{\theta}$. Esta reúne toda la información del vector de parámetros, proporcionada por la creencia *a priori* del investigador, así como la brindada por la muestra. Se verificó que la esperanza de dicha distribución es el estimador que minimiza el error cuadrático medio *a posteriori*.

Por lo tanto, para un vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ se busca determinar:

$$\mathbb{E}_{post}[\theta_i] = \int_S \theta_i p_{post}(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

si θ_i es una variable aleatoria continua, y

$$\mathbb{E}_{post}[\theta_i] = \sum_{\boldsymbol{\theta} \in S} \theta_i p_{post}(\boldsymbol{\theta}),$$

si θ_i es una variable aleatoria discreta, donde S es el soporte de $p_{post}(\boldsymbol{\theta})$.

En la práctica existen casos donde la distribución *a posteriori* posee una expresión compleja y por ende el desarrollo analítico de su esperanza se torna complicado. De manera

general, en el caso continuo, si deseamos calcular la integral del producto de dos funciones con soporte S , es decir,

$$\int_S h(x)f(x)dx, \quad (1.22)$$

donde $f, h : S \rightarrow \mathbb{R}$ y $f(\cdot)$ satisface las propiedades de una función de densidad, entonces el problema de calcular (1.22) puede traducirse a determinar la esperanza de una variable aleatoria. Esto dado que existiría una variable aleatoria X que asumiría valores en S y cuya función de densidad sería $f(\cdot)$ (Billingsley, 1996).

Por lo tanto, si $h(\cdot)$ es tal que $h(X)$ es una variable aleatoria con esperanza finita y X tiene función de densidad $f(\cdot)$, entonces:

$$\mathbb{E}[h(X)] = \int_S h(x)f(x)dx.$$

Al ver a (1.22) en términos de la esperanza de una variable aleatoria, podemos aproximarla usando el siguiente teorema (ver Feller, 1968).

Teorema 1.5 (Ley fuerte de los grandes números) *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas tal que $\mathbb{E}|X_i| < \infty$ entonces:*

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}[X_i] \right] = 1. \quad (1.23)$$

Observaciones.

1. Suponga que $f(\cdot)$ es la función de densidad de una variable aleatoria X , si generamos a partir de dicha función valores independientes x_1, \dots, x_n , aplicando (1.23) cuando $n \rightarrow \infty$:

$$\frac{1}{n} \sum_{i=1}^n h(x_i) \rightarrow \mathbb{E}[h(X)]. \quad (1.24)$$

2. En Robert y Casella (2010) se puede encontrar un resultado análogo al **Teorema 1.5**, pero para el caso de cadenas de Markov.

Si en (1.22) ninguna de las funciones $f(\cdot)$ y $h(\cdot)$ satisface las condiciones de función de densidad de alguna distribución conocida, es posible aplicar la siguiente técnica, conocida como muestreo por importancia. En este caso se considera una función de densidad $g(\cdot)$ con $g(x) > 0$, $x \in S$, que pueda ser vinculada a una distribución de probabilidad conocida. De esta forma (1.22) puede escribirse como:

$$\int_S \frac{f(x)h(x)g(x)}{g(x)} dx,$$

Por lo tanto, podemos considerar a $g(\cdot)$ como la función de densidad de una variable aleatoria X de la cual se generan los valores x_1, x_2, \dots, x_n . Si la transformación $f(X)h(X)/g(X)$ satisface ser una variable aleatoria con esperanza finita, entonces puede aplicarse la ley de los grandes números para obtener,

$$\frac{1}{n} \sum_{i=1}^n \frac{f(x_i)h(x_i)}{g(x_i)} \longrightarrow \mathbb{E} \left[\frac{f(X)h(X)}{g(X)} \right], \quad (1.25)$$

cuando $n \rightarrow \infty$. En la práctica al ser imposible generar una cantidad infinita de valores simulados, se debe tomar una muestra *suficientemente* grande.

Observación. A diferencia de (1.24), los valores simulados x_1, \dots, x_n para determinar la esperanza (1.25) se generan a partir de la distribución con función de densidad $g(\cdot)$.

Como se ha observado, todas estas aproximaciones se basan en la generación de valores a partir de cierta distribución. Existen algunos métodos clásicos para realizar dicha tarea, como por ejemplo el de aceptación y rechazo o el uso de la inversa de la distribución de la variable de interés (ver Rodrigues, 2013). Sin embargo, en la práctica existen distribuciones en las que no es posible usar estos algoritmos. Esto debido a las hipótesis que estos algoritmos requieren, o porque no son eficientes desde una perspectiva computacional.

Una posible solución es utilizar los métodos Monte Carlo vía cadenas de Markov. Estos algoritmos consisten en construir una cadena de Markov ergódica tal que su distribución estacionaria sea la distribución de la cuál se quiere generar valores. Uno de los algoritmos más utilizados es el de Metropolis-Hastings, el cual se estudiará a continuación.

1.4. Algoritmo de Metropolis-Hastings

La versión original de este algoritmo fue dado por Metropolis *et. al* (1953) donde se genera valores de la distribución estacionaria de una cadena Markov a partir de probabilidades condicionales simétricas. Dicho algoritmo se generalizó en Hastings (1970). Algunos resultados de esta generalización se presentan a continuación. Se describirá el caso donde S es discreto y finito. El caso donde S es el soporte de una distribución continua puede aproximarse mediante el caso discreto con las consideraciones pertinentes (ver Robert y Casella, 2010).

Algoritmo de Metropolis-Hastings (1970).

Dada $f(\cdot)$, la función de probabilidad con soporte S de la cual se busca muestrear valores. Sea $q(j | i) > 0$ la probabilidad de generar un valor j dado otro valor i , de tal forma que muestrear valores a partir de $q(\cdot | i)$ sea relativamente fácil. Se construye una cadena de Markov $X = \{X_k : k \in \mathbb{N}\}$ como sigue.

1. Se determina un estado inicial $X_0 = x_0$, $x_0 \in S$.
2. Para $n = 1, 2, \dots$, sea $X_n = x_n$.
 - a) Se genera un valor x^* que provenga de la distribución $q(\cdot | x_n)$.
 - b) Para X_{n+1} se considera lo siguiente.

$$X_{n+1} = \begin{cases} x^* & \text{con probabilidad } \alpha(x_n, x^*), \\ x_n & \text{con probabilidad } 1 - \alpha(x_n, x^*). \end{cases}$$

donde la probabilidad $\alpha(x_n, x^*)$ está dada por:

$$\alpha(x_n, x^*) = \min \left\{ 1, \frac{f(x^*)q(x_n | x^*)}{f(x_n)q(x^* | x_n)} \right\}.$$

Observación. El algoritmo original de Metropolis *et. al.*, (1953), requiere que la matriz de probabilidades $q(j | i)$ sea simétrica, es decir, $q(j | i) = q(i | j)$, $i, j \in S$. De esta forma, la probabilidad $\alpha(x_n, x^*)$ quedaría definida como:

$$\min \left\{ 1, \frac{f(x^*)q(x_n | x^*)}{f(x_n)q(x^* | x_n)} \right\} = \min \left\{ 1, \frac{f(x^*)}{f(x_n)} \right\}.$$

Actualmente existen variantes de este algoritmo, las cuales buscan atender casos particulares. Para un desarrollo más profundo de estas extensiones y su ejemplificación se recomienda Robert y Casella (2010).

Proposición 1.8 *La cadena de Markov $X = \{X_k : k \in T\}$ generada por el algoritmo de Metropolis-Hastings posee matriz de transición $P = (p_{i,j})_{i,j \in S}$.*

$$p_{i,j} = \begin{cases} q(j|i)\alpha(i, j) & j \neq i \\ q(i|i) + \sum_{k \neq i} q(k|i)(1 - \alpha(i, k)) & j = i. \end{cases}$$

Demostración. Supongamos que $i \neq j$, por definición de probabilidad de transición se observa que:

$$p_{i,j} = \mathbb{P}[X_{n+1} = j \mid X_n = i].$$

Para que $X_{n+1} = j$, en el punto 2 del algoritmo se tuvo que generar al estado j a partir de la distribución $q(\cdot|i)$ y posteriormente haber sido aceptado, lo que ocurre con probabilidad $\alpha(i, j)$. Por lo tanto, éstos al ser eventos independientes, producen,

$$p_{i,j} = q(j|i)\alpha(i, j), \quad i \neq j.$$

Suponga que $i = j$, se busca la probabilidad que el estado a tiempo $n + 1$ sea i dado que en el instante anterior también fue i . Fijando $X_n = i$, existen dos maneras para establecer $X_{n+1} = i$.

Caso 1. Generar al estado i a partir de la distribución condicional $q(\cdot|i)$. Una vez establecido se acepta con probabilidad $\alpha(i, i)$, que es dada por,

$$\begin{aligned} \alpha(i, i) &= \min \left\{ 1, \frac{f(i)q(i|i)}{f(i)q(i|i)} \right\} \\ &= \min \{1, 1\} \\ &= 1. \end{aligned}$$

Caso 2. Si al simular la variable con distribución $q(\cdot|i)$ se obtuviera $k \in S$, tal que $k \neq i$. Para poder fijar $X_{n+1} = i$, se necesitaría rechazar $X_{n+1} = k$ con probabilidad $1 - \alpha(i, k)$.

Por lo tanto, si $i = j$, la probabilidad de transición $p_{i,j}$ será igual a la suma de los casos anteriores:

$$\begin{aligned} p_{i,j} &= q(i|i)\alpha(i, i) + \sum_{k \neq i} q(k|i)(1 - \alpha(i, k)) \\ &= q(i|i) + \sum_{k \neq i} q(k|i)(1 - \alpha(i, k)). \end{aligned}$$

■

Teorema 1.6 *La cadena de Markov $X = \{X_k : k \in T\}$ generada por el algoritmo de Metropolis-Hastings posee distribución estacionaria $\{f(x) : x \in S\}$.*

Demostración. Se probará que la función de densidad $f(\cdot)$ cumpla con la **Proposición 1.7** para cualesquiera $i, j \in S$. El caso $i = j$ es inmediato. Se verifica el caso $i \neq j$.

Considere $f(i)p_{i,j}$. De las probabilidades de transición desarrolladas en la **Proposición 1.8** para el caso $i \neq j$ se tiene que,

$$\begin{aligned} f(i)p_{i,j} &= f(i)q(j|i)\alpha(i,j) \\ &= f(i)q(j|i) \min \left\{ 1, \frac{f(j)q(i|j)}{f(i)q(j|i)} \right\} \\ &= \min \{ f(i)q(j|i), f(j)q(i|j) \}. \end{aligned}$$

De esta manera,

$$\begin{aligned} f(i)p_{i,j} &= \min \left\{ \frac{f(i)q(j|i)}{f(j)q(i|j)} f(j)q(i|j), f(j)q(i|j) \right\} \\ &= f(j)q(i|j) \min \left\{ \frac{f(i)q(j|i)}{f(j)q(i|j)}, 1 \right\} \\ &= f(j)q(i|j)\alpha(j,i) \\ &= f(j)p_{j,i}. \end{aligned}$$

De esta forma, por la **Proposición 1.7**, la función $f(\cdot)$ es la distribución estacionaria de la cadena X . ■

Teorema 1.7 *Sea $X = \{X_k : k \in T\}$ la cadena de Markov con espacio de estados S finito, generada por el algoritmo de Metropolis-Hastings. La cadena X es ergódica.*

Demostración. Para verificar que la cadena X es ergódica, por la **Definición 1.25**, se debe verificar que es irreducible, aperiódica y recurrente positiva.

1. Para verificar que la cadena es aperiódica se debe cumplir que $d(i) = 1$, para cualquier $i \in S$. Sea $i \in S$, por la **Proposición 1.8** se tiene,

$$p_{i,i} = q(i|i) + \sum_{k \neq i} q(k|i)(1 - \alpha(i,k)). \quad (1.26)$$

Por la definición del algoritmo, las probabilidades de transición $q(\cdot|i) > 0$. Además, para $j \neq i$ se cumple $\alpha(i,j) \neq 0$ ya que $f(j) > 0$ para $j \in S$. De esta manera, de (1.26), se observa que $p_{i,i} > 0$. De esta manera,

$$1 \in \left\{ n > 0 : p_{i,i}^{(n)} > 0 \right\}.$$

Por lo tanto, $d(i) = 1$ para cualquier $i \in S$ y se concluye que la cadena es aperiódica.

2. Para que X sea irreducible se debe probar que $i \longleftrightarrow j$, para cualesquiera $i, j \in S$.
Por la **Proposición 1.8** se verifica que para cualesquiera $i, j \in S$ se cumple que $p_{i,j} > 0$, incluso en el caso $i = j$.
3. Si el espacio S es finito, entonces por **Proposición 1.3** y **Proposición 1.4**, la cadena es recurrente positiva. Para un caso general donde S es continuo se recomienda revisar Robert y Casella (2010).

■

Capítulo 2

Algunos conceptos básicos de series de tiempo

En este capítulo se presentará una breve descripción de conceptos de la teoría de series de tiempo, requeridos para los objetivos del presente trabajo. Se inicia introduciendo el concepto de serie de tiempo y algunas de las razones por las cuales es importante su estudio. Después se presentarán algunas herramientas básicas que ayudarán en el análisis de este trabajo. Se concluye este capítulo con un breve panorama al estudio de algunos modelos de series de tiempo que se verán involucrados en los siguientes capítulos. Para el desarrollo del capítulo se usarán principalmente como referencia los textos de McNeil, *et. al* (2006) y Brockwell y Davis (2016).

2.1. Definiciones y resultados básicos

En la vida real es común encontrar distintos fenómenos cuyos resultados varían con respecto al tiempo. Al registrar dichos resultados es posible establecer información histórica que ayude al ajuste de un modelo que describa al fenómeno de estudio. Con ello es posible predecir los valores futuros de dicho experimento, además de proporcionar otras conclusiones del fenómeno a describir. De esta forma, el concepto de serie de tiempo es considerado como una sucesión de observaciones obtenidas a través del tiempo a partir de algún fenómeno que se busque estudiar.

Desde un punto de vista formal, la teoría de series de tiempo se encuentra ligada a los procesos estocásticos, cuya definición impulsa a concebir con mayor formalidad el concepto de serie de tiempo.

Como se ha mencionado, se busca ajustar un *modelo de series de tiempo* a un conjunto de datos, este se puede definir como un proceso estocástico. Por ende, las observaciones son una realización del modelo subyacente asignado. Para iniciar un análisis estadístico de un modelo de series de tiempo que describa las propiedades y el comportamiento del mismo, a seguir se introducen algunas nociones básicas. En lo que sigue se considera como conjunto de índices $T = \mathbb{Z}$.

Definición 2.1 Sea $X = \{X_t : t \in T\}$ un proceso estocástico. Suponga que $\mathbb{E}[X_t]$, la esperanza de la variable aleatoria X_t , es tal que para cada $t \in T$ se cumple $\mathbb{E}[X_t^n] < \infty$, $n = 1, 2$. Para la varianza $Var[X]$ de una variable aleatoria X y $Cov(X, Y)$ la covarianza entre dos variables X y Y , definimos las siguientes funciones:

(a) La función de autocovarianza de X , indicada por $\gamma : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$, está dada por:

$$\gamma(t, s) = Cov(X_t, X_s) = \mathbb{E}[X_t X_s] - \mathbb{E}[X_t]\mathbb{E}[X_s].$$

(b) La función de autocorrelación de X , indicada por $\rho : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$, está dada por:

$$\rho(t, s) = \frac{Cov(X_t, X_s)}{\sqrt{Var[X_t]}\sqrt{Var[X_s]}}$$

Observación. Las funciones $\gamma(\cdot, \cdot)$ y $\rho(\cdot, \cdot)$ están bien definidas como consecuencia de la hipótesis sobre $\mathbb{E}[X_t^2]$, que permite el uso de la desigualdad Cauchy-Schwarz (ver Ross, 2009), la cual está dada por,

$$\mathbb{E}^2[X_t X_s] \leq \mathbb{E}[X_t^2]\mathbb{E}[X_s^2]. \quad (2.1)$$

A partir de la desigualdad (2.1) se verifica que las esperanzas, varianzas y covarianzas, presentes en las funciones $\gamma(\cdot, \cdot)$ y $\rho(\cdot, \cdot)$, son finitas.

Algunos procesos poseen algunas propiedades que auxilian en el análisis del modelo que describe el fenómeno que se quiere estudiar. Por ejemplo, se ha observado que existen fenómenos cuyo comportamiento no depende del instante en el que se deseen estudiar. Esta característica ha motivado a introducir el concepto de estacionariedad, la cual simplifica la estructura de las funciones de la **Definición 2.1**. De manera formal se introducen las siguientes definiciones.

Definición 2.2 Un proceso $X = \{X_t : t \in T\}$ es estrictamente estacionario si para cualesquiera $t_1, \dots, t_n, k \in \mathbb{Z}$ y $n \in \mathbb{N}$, los vectores aleatorios $(X_{t_1}, \dots, X_{t_n})$ y $(X_{t_1+k}, \dots, X_{t_n+k})$ tienen la misma distribución.

En la práctica verificar si un proceso cumple esta propiedad se torna complicado, es por ello que se introduce el siguiente concepto de estacionariedad.

Definición 2.3 Sea $X = \{X_t : t \in T\}$ un proceso con $\mathbb{E}[X_t^2] < \infty$. Se dice que X es debilmente estacionario si para cualesquiera $t, s, h \in \mathbb{Z}$:

- (i) $\mathbb{E}[X_t]$ es independiente de t ,
- (ii) $\gamma(t, s) = \gamma(t + h, s + h)$.

Por razones de simplicidad, a un proceso que cumple la **Definición 2.3** se le llamará estacionario. El siguiente teorema relaciona los dos tipos de estacionariedad.

Teorema 2.1 Un proceso estrictamente estacionario $X = \{X_t : t \in T\}$, que satisface $\mathbb{E}[X_t^2] < \infty$ para $t \in T$ es debilmente estacionario.

Demostración. Si X es estrictamente estacionario, entonces por la **Definición 2.2**, X_t y X_{t+h} tienen la misma distribución para cualquier $h \in \mathbb{Z}$, así como (X_s, X_t) y (X_{s+h}, X_{t+h}) . Además, por hipótesis $\mathbb{E}[X_t^2] < \infty$, así que por la **Definición 2.1**, las funciones $\gamma(s, t)$ y $\gamma(s + h, t + h)$ están bien definidas. Por lo tanto, para cualesquiera $t, s, h \in T$,

$$\begin{aligned} \gamma(s, t) &= \mathbb{E}[X_t X_s] - \mathbb{E}[X_t] \mathbb{E}[X_s]. \\ &= \mathbb{E}[X_{t+h} X_{s+h}] - \mathbb{E}[X_{t+h}] \mathbb{E}[X_{s+h}] \\ &= \gamma(s + h, t + h). \end{aligned}$$

■

En general, el hecho que un proceso sea debilmente estacionario no implica que sea fuertemente estacionario. Más adelante se observará que existen procesos que verifican este hecho y otros que reafirman la importancia de la hipótesis en la finitud de los segundos momentos en el **Teorema 2.1**.

Si un proceso es estacionario, se puede verificar a través de la **Definición 2.3** que su función de autocovarianza se puede ver en términos de una sola variable, que será la

distancia entre las variable aleatorias de interés. De esta manera, la función de autocovarianza de un proceso estacionario se redefine como sigue.

Definición 2.4 Para un proceso $X = \{X_t : t \in T\}$ estacionario, su función de autocovarianza se define como $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$, y está dada de la siguiente forma,

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}).$$

Observación. Recordando que la varianza de una variable aleatoria X_t puede ser vista como la covarianza $\text{Cov}(X_t, X_t)$, podemos obtener $\text{Var}[X_t]$ mediante la función de autocovarianza, donde tomamos el incremento $h = 0$.

$$\text{Var}[X_t] = \gamma(0). \tag{2.2}$$

De la igualdad (2.2) se concluye que la varianza de un proceso estacionario es constante a través del tiempo.

Proposición 2.1 Si $X = \{X_t : t \in T\}$ es un proceso estacionario, entonces su función de autocorrelación $\rho : \mathbb{Z} \rightarrow \mathbb{R}$, estará dada de la siguiente forma,

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}.$$

Demostración. Partiendo del hecho que el proceso X es estacionario, entonces por la **Definición 2.4**, la función $\gamma(\cdot)$ puede expresarse en términos de una variable k . Además, sustituyendo (2.2), en el punto (b) de la **Definición 2.1** se tiene,

$$\begin{aligned} \rho(t, t+k) &= \frac{\text{Cov}(X_t, X_{t+k})}{\sqrt{\text{Var}[X_t]} \sqrt{\text{Var}[X_{t+k}]}} \\ &= \frac{\gamma(k)}{\sqrt{\gamma(0)} \sqrt{\gamma(0)}} \\ &= \frac{\gamma(k)}{\gamma(0)}. \end{aligned}$$

■

Ejemplo 2.1 El ruido blanco es un proceso estacionario. Este formará parte esencial de los modelos de series de tiempo que consideramos adelante. El ruido blanco sirve usualmente para representar los posibles errores en algún modelo. De manera formal, se define a continuación.

Definición 2.5 Se dice que un proceso $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ es un ruido blanco si las variables aleatorias ε_t , $t \in \mathbb{Z}$, no son correlacionadas, es decir, para toda $t \neq s$ se cumple que $\rho(t, s) = 0$ y además se satisface:

(i) $\mathbb{E}[\varepsilon_t] = 0$,

(ii) $\text{Var}[\varepsilon_t] = \sigma_\varepsilon^2 < \infty$.

Observación. Por la **Definición 2.5** se puede ver que el ruido blanco $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ es un proceso estacionario.

Este hecho se debe a que los segundos momentos del proceso son finitos, pues por el punto (ii) de la **Definición 2.5**, su varianza es finita. Adicionalmente, por el punto (i), su esperanza es constante.

Finalmente, la función de autocovarianza es constante igual a cero cuando $h \neq 0$ debido a que las variables de un ruido blanco no están correlacionadas. Cuando $h = 0$, por la igualdad (2.2) se cumple $\gamma(0) = \text{Var}[\varepsilon_t] = \sigma_\varepsilon^2$. En resumen, para un ruido blanco, la función de autocovarianza es independiente de h y está dada como:

$$\gamma(h) = \begin{cases} \sigma_\varepsilon^2, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

Por lo tanto, bajo la **Definición 2.3** un ruido blanco es un proceso estacionario.

Notación. A un proceso con las características dadas por la **Definición 2.5** lo denotaremos como $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. En ciertos modelos de series de tiempo se necesita que las variables ε_t , además de no presentar correlación, sean independientes, en este caso se dice que el ruido blanco es estricto y se denota $\varepsilon_t \sim SWN(0, \sigma_\varepsilon^2)$.

2.2. Modelos de series de tiempo

En esta sección se estudiarán algunos de los modelos clásicos de series de tiempo que por sus propiedades pueden ajustarse a ciertos datos en la vida real. Se inicia introduciendo los modelos $ARMA(p, q)$ por su papel en el estudio de series estacionarias. Después se presentarán algunos de los modelos condicionalmente heterocedásticos de los cuales se hará un estudio más profundo debido a sus aplicaciones en el desarrollo de este trabajo.

Los resultados presentados aquí fueron tomados principalmente de Brockwell y Davis (2006), Parado y West (2010) y McNeil *et. al.*(2005).

2.2.1. Procesos ARMA(p, q)

Existen diversos fenómenos que bajo un análisis estadístico se observa que cumplen propiedades que sugieren el ajuste de un proceso estacionario. Un tipo de procesos para analizar datos con comportamiento estacionario son los ARMA(p, q) (Brockwell y Davis, 2016). A continuación se expondrá un breve estudio de este modelo, introduciendo los resultados de mayor interés en los objetivos de este trabajo.

Definición 2.6 *Un proceso $X = \{X_t : t \in \mathbb{Z}\}$ es llamado **proceso autoregresivo de media móvil** con parámetros p y q , denotado por ARMA(p, q), si es estacionario y puede escribirse como:*

$$X_t = \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j}, \quad (2.3)$$

donde $\epsilon_t \sim WN(0, \sigma^2)$.

Observación. Considere el operador $B^k X_t := X_{t-k}$ y los polinomios:

$$\begin{aligned} \phi(z) &= 1 - \sum_{i=1}^p \phi_i z^i, \\ \theta(z) &= 1 + \sum_{i=1}^q \theta_i z^i. \end{aligned}$$

Un modelo ARMA(p, q) se puede definir como aquel proceso estacionario $X = \{X_t : t \in \mathbb{Z}\}$ que satisface:

$$\phi(B)X_t = \theta(B)\epsilon_t. \quad (2.4)$$

De la **Definición 2.6** se desprenden otros dos tipos de modelos.

Definición 2.7 *Un proceso $X = \{X_t : t \in \mathbb{Z}\}$ se le conoce como **proceso autoregresivo de orden p** , denotado por AR(p), si $\theta(z) = 1$, es decir, se puede escribir como:*

$$X_t = \epsilon_t + \sum_{i=1}^p \phi_i X_{t-i}.$$

Definición 2.8 Un proceso $X = \{X_t : t \in \mathbb{Z}\}$ es llamado proceso de medias móviles de orden q , indicado por $MA(q)$, si $\phi(z) = 1$, es decir, se puede escribir como:

$$X_t = \varepsilon_t + \sum_{i=1}^q \varepsilon_{t-i} \theta_i.$$

De manera intuitiva, un modelo autoregresivo representa una dependencia lineal de un fenómeno con respecto a su comportamiento en el pasado, mientras que un modelo de medias móviles depende de uno o varios factores estocásticos.

Otra propiedad útil en el estudio de series de tiempo es cuando un proceso es función causal de otro proceso. El concepto de causalidad se define a seguir.

Definición 2.9 Un proceso $ARMA(p, q)$, $X = \{X_t : t \in \mathbb{Z}\}$ es función causal de un proceso $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$, si para $\psi_j \in \mathbb{R}$ y $j \in \mathbb{N}$ se tiene:

$$(i) \quad X_t = \sum_{j \in \mathbb{N}} \psi_j \varepsilon_{t-j},$$

$$(ii) \quad \sum_{j \in \mathbb{N}} |\psi_j| \text{ es convergente.}$$

Observación. El concepto de causalidad no contempla observaciones futuras del proceso $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$. De manera general, un proceso de la forma (2.3) no siempre es estacionario. Para el análisis de estos procesos se introduce el concepto de *causalidad*. Además con dicha propiedad se podrá establecer la función de autocorrelación de manera generalizada.

Teorema 2.2 Un proceso $X = \{X_t : t \in \mathbb{Z}\}$ causal de $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$, es estacionario, con función de autocorrelación:

$$\rho(h) = \frac{\sum_{i \in \mathbb{N}} \psi_i \psi_{i+h}}{\sum_{i \in \mathbb{N}} \psi_i^2}, \quad h \geq 0. \quad (2.5)$$

Demostración. Sea $X = \{X_t : t \in \mathbb{Z}\}$ un proceso causal de un ruido blanco $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$, entonces podemos escribir a $X_t = \sum_{j \in \mathbb{N}} \psi_j \varepsilon_{t-j}$, para $\psi_j, j \in \mathbb{N}$ una sucesión que satisface los puntos de la **Definición 2.9**.

Primeramente, para probar el punto (i) de la **Definición 2.3**, consideremos la sucesión de variables aleatorias $\{\psi_j \varepsilon_{t-j} : j \in \mathbb{N}\}$. Esta cumple la hipótesis (a) del **Teorema A.1** en el **Apéndice A**, tal como se muestra a continuación. Partimos de la siguiente serie,

$$\sum_{j \in \mathbb{N}} \mathbb{E}|\psi_j \varepsilon_{t-j}| = \sum_{j \in \mathbb{N}} |\psi_j| \mathbb{E}|\varepsilon_{t-j}|.$$

donde para cualquier $t \in \mathbb{N}$, se tiene que $\mathbb{E}|\varepsilon_t| < \sigma$, esto debido a la desigualdad de Hölder (ver Protter, 2004). Esto se muestra a continuación.

$$\begin{aligned} \mathbb{E}|\varepsilon_t| &\leq (\mathbb{E}[|\varepsilon_t|^2])^{1/2} \\ &= (\mathbb{E}[\varepsilon_t^2])^{1/2}. \end{aligned}$$

Adicionalmente, dado que $\varepsilon_t \sim WN(0, \sigma^2)$,

$$\begin{aligned} (\mathbb{E}[\varepsilon_t^2])^{1/2} &= (Var[\varepsilon_t])^{1/2} \\ &= \sigma. \end{aligned}$$

Por lo tanto, como $\mathbb{E}|\varepsilon_t| < \sigma$ para cualquier $t \in T$, se sigue que,

$$\begin{aligned} \sum_{j \in \mathbb{N}} |\psi_j| \mathbb{E}|\varepsilon_{t-j}| &\leq \sum_{j \in \mathbb{N}} |\psi_j| \sigma \\ &= \sigma \sum_{j \in \mathbb{N}} |\psi_j| < \infty, \end{aligned}$$

donde la convergencia de la serie se garantiza debido al punto (ii) de la **Definición 2.9**. Por lo tanto, la serie $\sum_{j \in \mathbb{N}} \mathbb{E}|\psi_j \varepsilon_{t-j}|$ es convergente, y por (a) del **Teorema A.1** y el punto (i) de la **Definición 2.5** se cumple que la media del proceso X es cero, dado que

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E} \left[\sum_{j \in \mathbb{N}} \psi_j \varepsilon_{t-j} \right] \\ &= \sum_{j \in \mathbb{N}} \mathbb{E}[\psi_j \varepsilon_{t-j}] \\ &= 0, \end{aligned}$$

y por lo tanto, también es independiente de t .

Para demostrar el punto (ii) de de **Definición 2.3**, considerando que la media del proceso X es cero se verifica lo siguiente, para $h > 0$.

Nuevamente bajo el **Teorema A.1** se tiene:

$$\begin{aligned}
\gamma(h) &= \mathbb{E}[X_t X_{t+h}] - \mathbb{E}[X_t] \mathbb{E}[X_{t+h}] \\
&= \mathbb{E}[X_t X_{t+h}] \\
&= \mathbb{E} \left[\sum_{i \in \mathbb{N}} \psi_i \varepsilon_{t-i} \sum_{j \in \mathbb{N}} \psi_j \varepsilon_{t+h-j} \right] \\
&= \mathbb{E} \left[\sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \psi_i \varepsilon_{t-i} \psi_j \varepsilon_{t+h-j} \right] \\
&= \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \psi_i \psi_j \mathbb{E}[\varepsilon_{t-i} \varepsilon_{t+h-j}].
\end{aligned}$$

Considerando que $\varepsilon_t \sim WN(0, \sigma^2)$, entonces por el punto (ii) de la **Definición 2.5** $\mathbb{E}[\varepsilon_{t-i} \varepsilon_{t+h-j}] = \sigma^2$ si $j = i + h$ y cero en otro caso, por lo tanto:

$$\gamma(h) = \sigma^2 \sum_{i \in \mathbb{N}} \psi_i \psi_{i+|h|}, \quad (2.6)$$

donde (2.6) sólo depende de h , la distancia entre los instantes donde están las variables X_t y X_{t+h} . Particularmente, en el caso $h = 0$ se tiene,

$$\text{Var}[X_t] = \gamma(0) = \sigma^2 \sum_{j \in \mathbb{N}} \psi_j^2. \quad (2.7)$$

La serie (2.7) es convergente debido al punto (ii) de la **Definición 2.9**. Por lo tanto, el proceso X posee segundos momentos finitos debido a que $\text{Var}[X_t] = \mathbb{E}[X_t^2]$, con lo cual se satisface la hipótesis de la **Definición 2.3**. Por lo tanto, el proceso X es estacionario. Ahora veamos como es su función de correlación $\rho(h)$.

Por la **Proposición 2.1** en (2.6) y (2.7), se tiene que la función de autocorrelación está dada por,

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\sum_{i \in \mathbb{N}} \psi_i \psi_{i+|h|}}{\sum_{i \in \mathbb{N}} \psi_i^2}.$$

De esta manera se concluye lo que buscamos demostrar. ■

Para ejemplificar el **Teorema 2.2**, se introduce la siguiente proposición para el análisis de la estacionariedad de un proceso $AR(1)$.

Proposición 2.2 *Un proceso AR(1) es estacionario si $|\phi| < 1$ y su función de autocorrelación es $\rho(h) = \phi^h, h > 0$.*

Demostración. De la **Definición 2.7** un proceso AR(1) se puede escribir como:

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (2.8)$$

donde $\varepsilon_t \sim WN(0, \sigma^2), t \in \mathbb{Z}$.

De (2.8), si iteramos el proceso en si mismo, tenemos que,

$$\begin{aligned} X_t &= \phi X_{t-1} + \varepsilon_t \\ &= \phi [X_{t-2} + \varepsilon_{t-1}] + \varepsilon_t \\ &= \phi^2 X_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\ &= \phi^2 [\phi X_{t-3} + \varepsilon_{t-2}] + \phi \varepsilon_{t-1} + \varepsilon_t \\ &= \phi^3 X_{t-3} + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\ &= \phi^3 X_{t-3} + \sum_{k=0}^2 \phi^k \varepsilon_{t-k}. \end{aligned}$$

De esta forma, iterando recursivamente se obtiene la siguiente expresión general.

$$X_t = \sum_{k \in \mathbb{N}} \phi^k \varepsilon_{t-k}. \quad (2.9)$$

Por lo tanto, considerando la hipótesis $|\phi| < 1$, el proceso (2.9) es una función causal de $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ y por el **Teorema 2.2** es estacionario.

Para el desarrollo de la función de autocorrelación se utiliza el **Teorema 2.2**. A partir de (2.9) consideramos $\psi_j = \phi^j$ para $|\phi| < 1$, de esta forma,

$$\begin{aligned} \gamma(h) &= \sigma^2 \sum_{j \in \mathbb{N}} \psi_j \psi_{j+h} \\ &= \sigma^2 \sum_{j \in \mathbb{N}} \phi^j \phi^{j+h} \\ &= \sigma^2 \phi^h \sum_{j \in \mathbb{N}} \phi^{2j} \\ &= \sigma^2 \phi^h \sum_{j \in \mathbb{N}} (\phi^2)^j \\ &= \frac{\sigma^2 \phi^h}{1 - \phi^2}. \end{aligned}$$

Aplicando la **Proposición 2.1**, tenemos que para un proceso estacionario $\rho(h) = \gamma(h)/\gamma(0)$. A partir del desarrollo anterior $\gamma(0) = \frac{\sigma^2}{1 - \phi^2}$. Por lo tanto, $\rho(h) = \gamma(h)/\gamma(0) = \phi^h$, tal como se quería demostrar. ■

Dado un proceso $ARMA(p, q)$, la tarea de verificar si este es un proceso causal, partiendo únicamente de la **Definición 2.9** puede tornarse complicado. El siguiente resultado introducido en Brockwell (2016) facilita esta tarea.

Teorema 2.3 *Un proceso $ARMA(p, q)$ $X = \{X_t : t \in \mathbb{Z}\}$ para el cual los polinomios $\phi(\cdot)$ y $\theta(\cdot)$ no comparten raíces, es causal si y sólo si $\phi(\cdot)$ no posee raíces en el disco unitario $|z| \leq 1$. Además los coeficientes ψ_j se obtienen mediante la siguiente relación:*

$$\sum_{k \in \mathbb{N}} \psi_k z^k = \frac{\theta(z)}{\phi(z)}, \quad |z| < 1. \quad (2.10)$$

Demostración. Ver Brockwell y Davis (2016). ■

Observación. En la relación (2.10) se encuentran los polinomios $\phi(\cdot)$ y $\theta(\cdot)$ introducidos en (2.4). Debido a su estructura, se toma por convención que $\theta_0 = \phi_0 = 1$. En (Prado y West, 2010) se muestra que los valores ψ_j se pueden obtener de la siguiente forma, a partir de (2.10).

$$\begin{cases} \psi_j - \sum_{h=1}^j \psi_{j-h} = \theta_j & 0 < j < \text{máx}\{p, q + 1\}. \\ \psi_j - \sum_{h=1}^p \psi_{j-h} = 0 & j \geq \text{máx}\{p, q + 1\}. \end{cases} \quad (2.11)$$

El **Teorema 2.2** no solo ofrece la característica de estacionariedad, también proporciona otra característica relevante, bajo la ecuación (2.5) que es la función de autocorrelación para un proceso $ARMA(p, q)$ causal. La importancia de esta función recae en la identificación del tipo de modelo que podemos ajustar a nuestros datos.

Además del uso del **Teorema 2.2**, existe una segunda metodología para obtener la función de autocorrelación para un proceso $ARMA(p, q)$ causal.

Proposición 2.3 La función de autocovarianza $\gamma(\cdot)$ de un proceso ARMA(p, q) causal de $\varepsilon = \{\varepsilon_t : t \in \mathbf{Z}\}$ satisface la siguiente relación.

$$\gamma(h) = \begin{cases} \sum_{i=1}^p \phi_i \gamma(h-i) & h \geq \max(p, q+1) \\ \sum_{j=1}^p \phi_j \gamma(h-j) + \sigma^2 \sum_{i=h}^q \theta_i \psi_{i-h} & 0 \leq h < \max(p, q+1). \end{cases} \quad (2.12)$$

Demostración. Ver demostración en Prado y West (2010).

Ejemplo 2.1 La función de autocorrelación de un proceso ARMA(1,1) causal del ruido blanco $\varepsilon = \{\varepsilon_t : t \in \mathbb{N}\}$ cumple.

$$\rho(k) = \phi_1^{k-1} \rho(1), \quad k > 1.$$

Demostración. A partir de la **Proposición 2.3** se observa que para $h > 1$ se cumplirá en el caso $p = 1$, que

$$\gamma(k) = \phi_1 \gamma(k-1).$$

Desarrollando de manera recursiva esta expresión se obtiene que

$$\gamma(k) = \phi_1^{k-1} \gamma(1).$$

Por el **Teorema 2.2** un proceso ARMA(1,1) causal de $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ es estacionario. Entonces a partir de la **Proposición 2.1** se tiene que, para $k > 1$:

$$\begin{aligned} \rho(k) &= \frac{\gamma(k)}{\gamma(0)} \\ &= \phi_1^{k-1} \frac{\gamma(1)}{\gamma(0)} \\ &= \phi_1^{k-1} \rho(1). \end{aligned}$$

■

Aplicando la **Proposición 2.3** a los modelos AR(p), las igualdades resultantes reciben el nombre de ecuaciones de Yulle-Walker, estas tomarán importancia en el resto del análisis de estos modelos.

Definición 2.10 Las ecuaciones de Yulle-Walker se definen como:

$$\boldsymbol{\gamma}_p = \mathbf{\Gamma}_p \boldsymbol{\phi}. \quad (2.13)$$

$$\sigma^2 = \gamma(0) - \boldsymbol{\phi}^t \boldsymbol{\gamma}_p. \quad (2.14)$$

donde $\boldsymbol{\gamma}_p = (\gamma(1), \dots, \gamma(p))^t$, $\mathbf{\Gamma}_p$ es una matriz $p \times p$ cuya entrada (ij) corresponde a $\gamma(i-j)$ y $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$. La notación \boldsymbol{v}^t indica el vector transpuesto de \boldsymbol{v} .

Otra herramienta que se utiliza para el análisis de series de tiempo para los modelos ARMA(p, q) y sirve como ayuda para identificar que proceso se puede ajustar a una serie de datos, es la función de autocorrelación parcial (PACF por sus siglas en inglés).

Definición 2.11 La función de autocorrelación parcial de un proceso ARMA(p, q) es una función $\alpha(h) = \phi_{h,h}$ donde $\phi_{h,h}$ es el último término del vector $\boldsymbol{\phi}_h = (\phi_{1,h}, \dots, \phi_{h,h})$, dado por.

$$\boldsymbol{\phi}_h = \boldsymbol{\Gamma}_h^{-1} \boldsymbol{\gamma}_h, \quad (2.15)$$

donde $\boldsymbol{\Gamma}_h^{-1}$ es la inversa de la matriz $\boldsymbol{\Gamma}_h$ de la **Definición 2.10**, así como $\boldsymbol{\gamma}_h = (\gamma(1), \dots, \gamma(h))^t$.

En general para procesos estacionarios podemos encontrar $\phi_{h,h}$ con la metodología propuesta por Durbin y Levinson, presentada a continuación.

Método de Durbin-Levinson (Durbin y Levinson, 1947).

(i) Empiece en $\phi_{1,1} = \rho(1)$

(ii) Para $n > 1$, defina

$$\phi_{n,n} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(k)},$$

donde para $k = 1, 2, \dots, n-1$

$$\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n} \phi_{n-1,n-k}.$$

Observación. Las funciones de autocorrelación simple y parcial auxilian en la identificación de modelos que pueden ajustarse a una serie de datos. Por ejemplo, en el caso de los modelos ARMA(1,1) causales de un ruido blanco $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$, se observa a partir del **Ejemplo 2.1** que esta decrece a cero conforme el rezago k aumenta.

Se han mencionado varios aspectos acerca de los modelos ARMA(p, q) para una serie de tiempo. Sin embargo, en las aplicaciones a datos reales no se conoce el proceso que modela a nuestros datos. Por ende: ¿Cómo podemos verificar que un conjunto de datos puede ajustarse a alguno de los modelos descritos? y dado un modelo ¿Cómo lo podemos expresar de manera explícita?.

Si tenemos una serie de datos $\mathbf{x} = \{x_1, \dots, x_n\}$ que deseamos ajustar a un modelo como los $\text{ARMA}(p, q)$, debemos verificar que se cumplan los supuestos que este proceso requiere, por ejemplo, la estacionariedad. Con el objetivo de verificar si una serie de datos puede ajustarse a un modelo estacionario, Dickey-Fuller (1979) desarrollaron una prueba de hipótesis para verificar este hecho (ver Tsay, 2010).

Si es posible ajustar un modelo estacionario a un conjunto de datos \mathbf{x} , ahora nos interesa saber que modelo podemos utilizar. Bajo este hecho, se utilizan las extensiones muestrales de la funciones de autocorrelación simple y parcial, cuyo objetivo es establecer una estimación de las mismas con cierto error estadístico.

Definición 2.12 Sean $\{x_t : t = 1, \dots, n\}$ observaciones y $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ su media muestral, se definen las siguientes funciones.

(a) La función de autocovarianza muestral se define como:

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-k} (x_{i+h} - \bar{x})(x_i - \bar{x}), \quad 0 \leq h < n.$$

(b) La función de autocorrelación muestral se define como:

$$\hat{\rho}(k) = \frac{\sum_{i=1}^{n-k} (x_{i+h} - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad 0 \leq h < n.$$

(c) La función de autocorrelación parcial muestral se define como:

$$\hat{\alpha}(h) = \begin{cases} 1 & h = 0, \\ \hat{\phi}_{h,h} & h > 0. \end{cases}$$

donde $\hat{\phi}_{h,h}$ es el último término del vector ϕ_h dado en la **Definición 2.11**.

Para visualizar de manera gráfica estas funciones, se utilizan los diagramas llamados correlogramas $\{(h, \hat{f}(h)) : h \in \mathbb{N}\}$, donde $f(\cdot)$ corresponde a las funciones de autocorrelación muestrales. En la Figura 2.1 mostramos un ejemplo de un correlograma para la función de autocorrelación muestral de un conjunto de datos.

Tomando en cuenta los correlogramas y el comportamiento de las funciones de autocorrelación de los modelos $ARMA(p, q)$, es posible asignar los órdenes p y q convenientes.

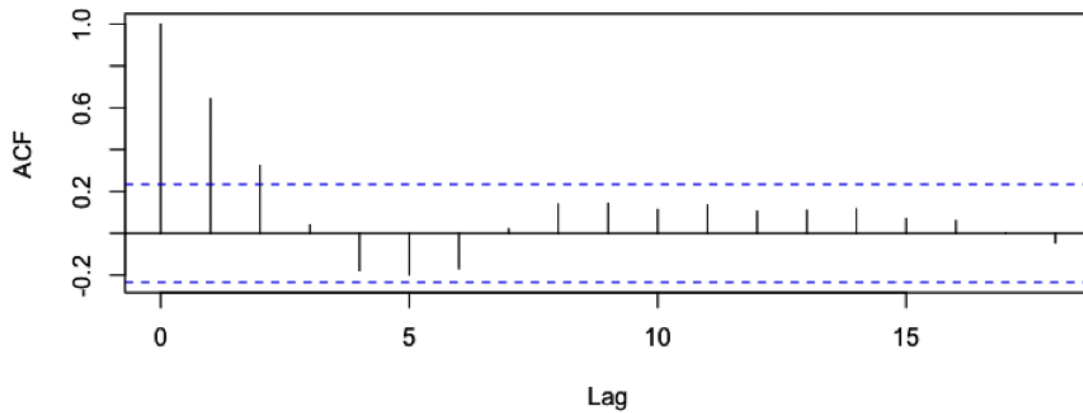


Figura 2.1: Correlograma de la función de autocorrelación (ACF) de una muestra simulada a través de un proceso $ARMA(1,1)$

2.3. Series de tiempo financieras

En datos del sector económico se ha observado que los rendimientos de activos financieros generalmente presentan algunas características que dificultan su modelación. De esta forma, a partir de los años 80 ha surgido el desarrollo de modelos de series de tiempo, tal como los procesos ARCH (procesos condicionalmente heterocedásticos), que serán abordados más adelante y fueron propuestos por Engle (1982), así como su generalización introducida por Bollerslev (1986), los cuales tienen el objetivo de capturar los principales hechos empíricos subyacentes en los datos financieros y así obtener conclusiones más certeras sobre sus aplicaciones. Para el desarrollo teórico de estos modelos se utilizarán algunos conceptos que pueden ser encontrados en los textos de Tsay (2010) y McNeil, *et. al* (2005).

En un activo financiero se busca estudiar el comportamiento de su valor en el mercado. Este se cuantifica mediante su *precio*, denotado por p_t para un instante $t \in T \subset \mathbb{N}$. Como un ejemplo de una serie de tiempo de precios, en la Figura 2.2 se presentan precios diarios de cierre del tipo de cambio Yen Japonés y Peso Mexicano (JPY/MXN).

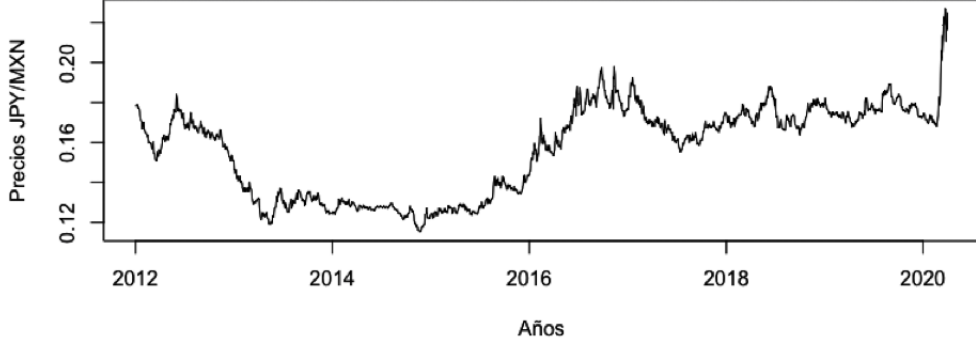


Figura 2.2: Precios diarios del tipo de cambio JPY/MXN de enero 2012 a marzo 2020.

De acuerdo a Francq, *et. al* (2010) las series históricas de precios de activos financieros no siguen un comportamiento compatible con una serie de tiempo estacionaria y son más cercanas a un proceso de caminata aleatoria. Para resolver este problema, se transforma la serie de precios a una que sea compatible con las características de un modelo estacionario, como los $ARMA(p, q)$. A seguir se presentan algunas de estas transformaciones.

Definición 2.13 Para $X = \{X_t : t \in \mathbb{Z}\}$ se define el operador ∇^n de la siguiente forma:

- (i) $\nabla X_t := X_t - X_{t-1}$.
- (ii) $\nabla^n X_t := \nabla^{n-1}(X_t - X_{t-1}) \quad , n > 1$.

El análisis de los precios recae en el beneficio generado por un activo de un instante a otro, tal magnitud se mide mediante los rendimientos. En el desarrollo de este trabajo utilizaremos como rendimientos a los log-retornos de una serie de precios, tal como se define a seguir.

Definición 2.14 Sea $P = \{p_t : t \in T \subset \mathbb{N}\}$ la serie de precios de un activo financiero. Los log-retornos de P están indicados por $R = \{r_t : t \in T \subset \mathbb{N}\}$ donde.

$$r_t = \ln \left(\frac{p_t}{p_{t-1}} \right), \quad t \in T. \quad (2.16)$$

Observación. Esta definición coincide con la aplicación del operador diferencia ∇ a los logaritmos de p_t , transformaciones utilizadas para obtener series de tiempo estacionarias. En la Figura 2.3 se muestran los rendimientos asociados a los precios diarios de cierre del tipo de cambio JPY/MXN.

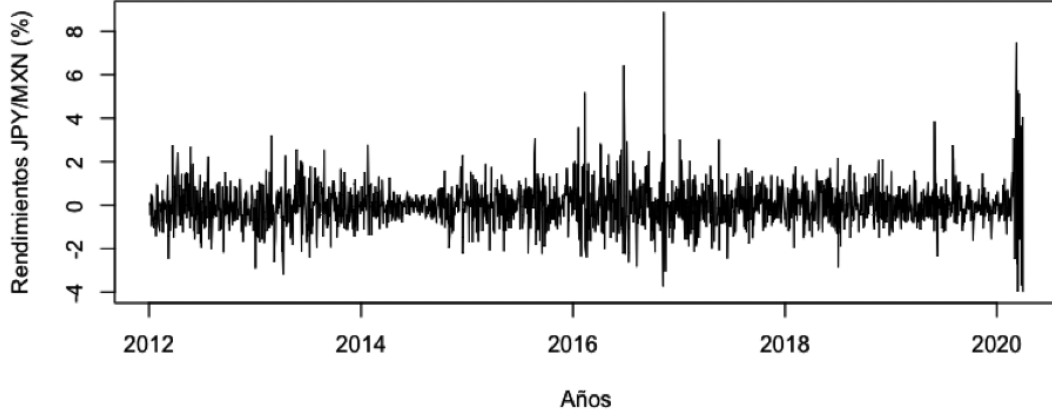


Figura 2.3: Log-retornos diarios del tipo de cambio JPY/MXN de enero 2012 a marzo 2020.

Una característica sustancial en el estudio del comportamiento de los rendimientos es su dispersión, cuya importancia recae en el análisis del riesgos financieros.

Definición 2.15 Sea $R = \{r_t : t \in T \subset \mathbb{N}\}$ la serie de log retornos de un activo financiero, la volatilidad indicada por σ_t se define como la desviación estandar condicional de los rendimientos del activo financiero en estudio, es decir.

$$\sigma_t = \text{Var}(r_t | F_{t-1})^{1/2}, \quad (2.17)$$

donde F_t es la sigma álgebra generada por las observaciones pasada del proceso X , es decir, $\sigma(\{X_s : s \leq t\})$ y representa la información histórica conocida.

El comportamiento de rendimientos de activos financieros generalmente posee varias características empíricas que dificultan su modelación con procesos $\text{ARMA}(p, q)$. La mayoría de éstos recaen en el comportamiento de la volatilidad y se presentan a seguir.

2.3.1. Hechos adicionales

Los siguientes hechos presentados en Francq, *et. al* (2010) inspiraron el desarrollo de los modelos de series de tiempo condicionalmente heterocedásticos que se introducirán más adelante.

1. Volatility Clustering. Se refiere a la presencia de rendimientos extremos en grupos. De esta manera, rendimientos extremos tienden a ser seguidos por otros de la misma magnitud. Por ejemplo, los presentados a inicios del 2020 para el tipo de cambios JPY/MXN. (ver Figura 2.3).

2. Comportamiento de la correlación. Para establecer las hipótesis de los modelos de series de tiempo aplicadas a datos financieros, es importante analizar su función de autocorrelación. Se observa que la correlación de los rendimientos de la Figura 2.3 presentan autocorrelaciones insignificantes, tal como se muestra en la Figura 2.4.

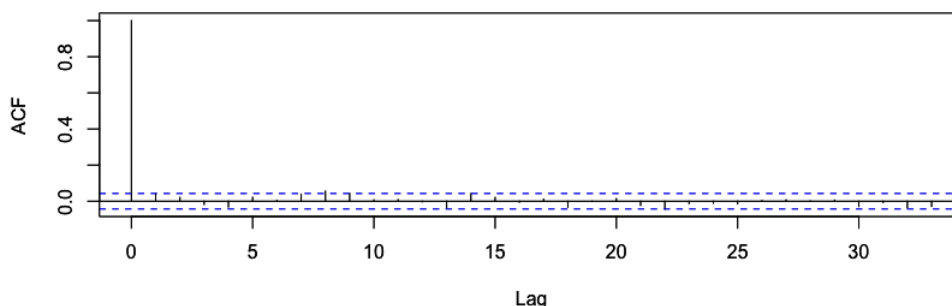


Figura 2.4: Función de autocorrelación muestral de los rendimientos diarios de JPY/MXN

En la Figura 2.5 se observa la función de autocorrelación muestral de los rendimientos cuadrados $R = \{r_t^2 : t \in T\}$ del tipo de cambio JPY/MXN.

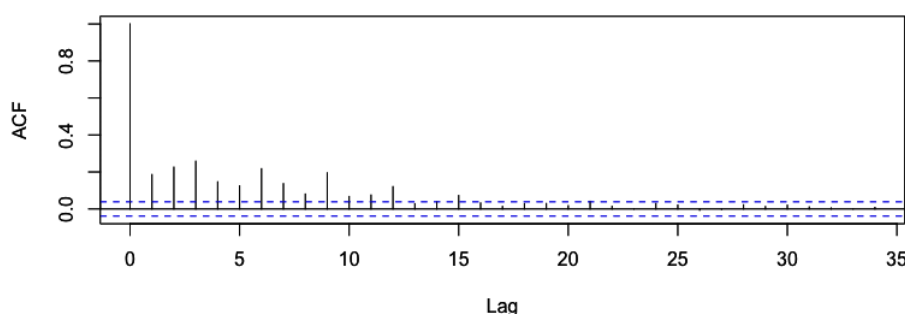


Figura 2.5: Autocorrelación muestral de los rendimientos cuadrados del tipo de cambio JPY/MXN

Se observa que la función de autocorrelación muestral de los datos $R^2 = \{r_t^2 : t \in T\}$ presenta una fuerte correlación, esto sugiere que a pesar de la baja correlación que poseen los rendimientos de activos financieros, éstos mantienen cierta dependencia en su magnitud absoluta. A este hecho se le conoce como efecto ARCH.

Observación. Al no existir correlación en la serie de rendimientos, no sería conveniente ajustar directamente un modelo de series de tiempo $ARMA(p, q)$ a rendimientos de activos financieros para su pronóstico. Sin embargo, si podría modelarse el comportamiento de las magnitudes absolutas de los rendimientos, es decir, la volatilidad.

Algunos hechos adicionales que también satisfacen los rendimientos de activos financieros se presentan en el **Apéndice A**.

2.3.2. Procesos ARCH(q)

Algunas propiedades que satisfacen los rendimientos de activos financieros, tal como el efecto de la volatilidad en grupos de los rendimientos de activos, así como la presencia de efectos ARCH, implica que la varianza sea condicionalmente heterocedástica, es decir, (2.17) no permanece constante a través del tiempo. Esto ha motivado a buscar un proceso que incluya en su estructura una forma de capturar este hecho. Entre ellos se encuentra el modelo propuesto por Engle (1982), que considera una clase de procesos que buscan incorporar de manera lineal los cuadrados de datos históricos como factores del comportamiento de la volatilidad.

El modelo propuesto por Engle (1982) se le conoce como modelo autoregresivo condicionalmente heterocedástico, cuya definición se presenta a seguir.

Definición 2.16 Sea $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ un proceso tal que $\varepsilon_t \sim SWN(0, 1)$. Decimos que $X = \{X_t : t \in \mathbb{Z}\}$ es un **proceso autoregresivo condicionalmente heterocedástico de orden q** , denotado por $ARCH(q)$, si es estrictamente estacionario y se escribe como,

$$X_t = \sigma_t \varepsilon_t, \quad (2.18)$$

donde ε_t es independiente de X_s para $s < t$ y σ_t^2 es una función lineal de los cuadrados de las observaciones X_t , es decir, existen constantes $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \dots, q$ tales que

$$\sigma_t^2 = \omega + \sum_{s=1}^q \alpha_i X_{t-s}^2. \quad (2.19)$$

A continuación mostramos la representación de un modelo $ARCH(1)$ en términos del proceso ε . Esta representación nos ayudará a establecer condiciones de estacionariedad, así como otros resultados. Para ello utilizaremos un resultado introducido por McNeil *et.al* (2005).

Proposición 2.4 Sean $A = \{A_t : t \in \mathbb{Z}\}$ y $B = \{B_t : t \in \mathbb{Z}\}$ dos procesos estocásticos de variables aleatorias independientes e idénticamente distribuidas. Si un proceso estocástico $X = \{X_t : t \in \mathbb{Z}\}$ puede expresarse como:

$$X_t = A_t X_{t-1} + B_t, \quad (2.20)$$

y se satisfacen las condiciones:

$$\mathbb{E}[\max\{0, \ln |B_t|\}] < \infty, \quad \mathbb{E}[\ln |A_t|] < 0,$$

existe una única expresión de X en términos de los procesos A y B , dada por:

$$X_t = B_t + \sum_{i=1}^{\infty} B_{t-i} \prod_{j=0}^{i-1} A_{t-j}, \quad (2.21)$$

y además es estrictamente estacionaria.

Demostración. Ver demostración en McNeil *et, al.* (2005). ■

Aplicaremos este teorema a las ecuaciones (2.18) y (2.19), que por definición constituyen un proceso estrictamente estacionario.

Proposición 2.5 *Un proceso ARCH(1) es estacionario si y sólo si $\alpha_1 < 1$.*

Demostración. Por la **Definición 2.16**, un proceso ARCH(1) satisface las siguientes ecuaciones, para $t \in \mathbb{Z}$,

$$X_t = \sigma_t \varepsilon_t, \quad \text{con} \quad \sigma_t^2 = \omega + \alpha_1 X_{t-1}^2.$$

Supongamos que el proceso es estacionario, entonces como ε y X son independientes y σ_t^2 es una función de X , se tiene,

$$\begin{aligned} \mathbb{E}[X_t^2] &= \mathbb{E}[\varepsilon_t^2 \sigma_t^2] \\ &= \mathbb{E}[\varepsilon_t^2] \mathbb{E}[\sigma_t^2] \end{aligned}$$

Como $\varepsilon_t \sim SWN(0, 1)$ se sigue que,

$$\begin{aligned} \mathbb{E}[X_t^2] &= \mathbb{E}[\sigma_t^2] \\ &= \mathbb{E}[\omega + \alpha_1 X_{t-1}^2] \\ &= \omega + \alpha_1 \mathbb{E}[X_{t-1}^2]. \end{aligned}$$

Por **Definición 2.16**, tanto la varianza como la esperanza de X son constantes en el tiempo, por lo tanto sus segundos momentos también lo son. De esta manera,

$$\mathbb{E}[X_t^2] = \omega + \alpha_1 \mathbb{E}[X_t^2].$$

Por lo tanto:

$$\mathbb{E}[X_t^2] = \frac{\omega}{1 - \alpha_1}.$$

Como X_t^2 es una variable aleatoria positiva y la **Definición 2.3** requiere que los segundos momentos sean finitos, se necesita que $1 - \alpha_1 > 0$, por lo tanto $\alpha_1 < 1$.

Por otra parte, si se supone que $\alpha_1 < 1$, entonces por la **Proposición 2.4** se demostrará que el proceso X es estacionario. Se observa que $\{X_t^2 : t \in \mathbb{Z}\}$ es de la forma (2.20) pues:

$$\begin{aligned} X_t^2 &= \sigma_t^2 \varepsilon_t^2 \\ &= (\omega + \alpha_1 X_{t-1}^2) \varepsilon_t^2 \\ &= \omega \varepsilon_t^2 + \alpha_1 \varepsilon_t^2 X_{t-1}^2, \end{aligned}$$

donde para cada $t \in \mathbb{Z}$ se tiene $A_t = \alpha_1 \varepsilon_t^2$ y $B_t = \omega \varepsilon_t^2$. Adicionalmente, note que,

$$\mathbb{E}[\max\{0, \ln |B_t|\}] = \mathbb{E}[\max\{0, \ln |\omega \varepsilon_t^2|\}].$$

Considerando que $\omega \varepsilon_t^2 > 0$, entonces por la desigualdad de Jensen:

$$\begin{aligned} \mathbb{E}[\max\{0, \ln |B_t|\}] &= \mathbb{E}[\ln(\omega \varepsilon_t^2)] \\ &\leq \ln \mathbb{E}[\omega \varepsilon_t^2] \\ &= \ln \omega. \end{aligned}$$

Por la **Definición 2.16** tenemos que $\omega > 0$, entonces $\mathbb{E}[\max\{0, \ln |B_t|\}] < \infty$.

Para verificar la segunda condición de la **Proposición 2.4** se utiliza nuevamente la desigualdad de Jensen y la hipótesis $\alpha_1 < 1$.

$$\begin{aligned} \mathbb{E}[\ln |A_t|] &= \mathbb{E}[\ln |\alpha_1 \varepsilon_t^2|] \\ &= \mathbb{E}[\ln \alpha_1 \varepsilon_t^2] \\ &\leq \ln \mathbb{E}[\alpha_1 \varepsilon_t^2]. \end{aligned}$$

Considerando que $\varepsilon_t \sim SWN(0, 1)$, y el hecho que α_1 es una constante, se tiene:

$$\mathbb{E}[\ln |A_t|] \leq \ln \alpha_1 \leq \ln 1 = 0.$$

Entonces por la **Proposición 2.4** existen las condiciones para que exista una solución (2.21) estrictamente estacionaria al proceso X_t^2 y está dada como sigue.

$$\begin{aligned} X_t^2 &= \omega \varepsilon_t^2 + \sum_{i=1}^{\infty} \omega \varepsilon_{t-i}^2 \prod_{j=0}^{i-1} \alpha_1 \varepsilon_{t-j}^2 \\ &= \omega \varepsilon_t^2 + \omega \sum_{i=1}^{\infty} \alpha_1^i \varepsilon_{t-i}^2 \prod_{j=0}^{i-1} \varepsilon_{t-j}^2. \end{aligned}$$

Ajustando los índices de la suma y el producto,

$$\begin{aligned} X_t^2 &= \omega \varepsilon_t^2 + \omega \sum_{i=1}^{\infty} \alpha^i \prod_{j=0}^i \varepsilon_{t-j}^2 \\ &= \omega \sum_{i=0}^{\infty} \alpha^i \prod_{j=0}^i \varepsilon_{t-j}^2. \end{aligned}$$

Por la definición un proceso ARCH(1) ya es un proceso estrictamente estacionario. Además el proceso es estacionario puesto que $\mathbb{E}[X_t^2] < \infty$, dado que,

$$\mathbb{E}[X_t^2] = \mathbb{E} \left[\alpha_0 \sum_{i \in \mathbb{N}} \alpha_1^i \prod_{j=0}^{i-1} \varepsilon_{t-j}^2 \right].$$

Aplicando el **Teorema A.1** del **Apéndice A** a las variables aleatorias no negativas ε_{t-j}^2 , se tiene,

$$\mathbb{E}[X_t^2] = \alpha_0 \sum_{i \in \mathbb{N}} \mathbb{E} \left[\alpha_1^i \prod_{j=0}^{i-1} \varepsilon_{t-j}^2 \right].$$

Al ser $\varepsilon_t \sim SWN(0, 1)$, las variables ε_t^2 son independientes, entonces:

$$\begin{aligned} \mathbb{E}[X_t^2] &= \alpha_0 \sum_{i \in \mathbb{N}} \alpha_1^i \prod_{j=0}^{i-1} \mathbb{E}[\varepsilon_{t-j}^2] \\ &= \alpha_0 \sum_{i \in \mathbb{N}} \alpha_1^i. \end{aligned}$$

Por hipótesis $\alpha_1 < 1$.

$$\mathbb{E}[X_t^2] = \frac{\alpha_0}{1 - \alpha_1} < \infty.$$

Por el **Teorema 2.1**, el proceso X es estacionario.

Proposición 2.6 Sea $X = \{X_t : t \in T\}$ un proceso ARCH(q), entonces la distribución condicional de X_t dada F_t posee media cero y varianza σ_t^2 .

Demostración. Debido a la independencia entre ε_t con F_t , ε_t también será independiente de σ_t , al ser una función lineal de los cuadrados de las q observaciones pasadas, entonces

$$\begin{aligned} \mathbb{E}[X_t | F_{t-1}] &= \mathbb{E}[\sigma_t \varepsilon_t | F_{t-1}] \\ &= \mathbb{E}[\sigma_t | F_{t-1}] \mathbb{E}[\varepsilon_t | F_{t-1}] \\ &= \mathbb{E}[\sigma_t | F_{t-1}] \mathbb{E}[\varepsilon_t]. \end{aligned}$$

Por definición, $\varepsilon_t \sim SWN(0, 1)$, entonces $\mathbb{E}[X_t | F_{t-1}] = 0$.

Por otro lado, por la definición de varianza condicional se sigue que,

$$\text{Var}[X_t|F_{t-1}] = \mathbb{E}[X_t^2|F_{t-1}] - \mathbb{E}[X_t|F_{t-1}]^2. \quad (2.22)$$

Sustituyendo en (2.22) el hecho que la distribución condicional de X_t dada F_{t-1} posee media cero y que $\varepsilon_t \sim SWN(0, 1)$.

$$\begin{aligned} \text{Var}[X_t|F_{t-1}] &= \mathbb{E}[X_t^2 | F_{t-1}] \\ &= \mathbb{E}[\sigma_t^2 \varepsilon_t^2 | F_{t-1}] \\ &= \sigma_t^2 \mathbb{E}[\varepsilon_t^2 | F_{t-1}] \\ &= \sigma_t^2. \end{aligned}$$

dado que σ_t^2 es una función de valores de X para $s < t$ y F_{t-1} medible. Por ende, σ_t^2 es la varianza condicional de X_t , la cuál depende del tiempo t . ■

De manera general, en McNeilet, *al.* (2005) demuestra que la condición necesaria y suficiente para la estacionariedad de un proceso ARCH(q) es:

$$\sum_{i=1}^q \alpha_i < 1. \quad (2.23)$$

Proposición 2.7 *La varianza incondicional de un proceso ARCH(q) estacionario está dada por:*

$$\text{Var}[X_t] = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i}, \quad t \in \mathbb{Z}.$$

Demostración. Por la definición de varianza obtenemos:

$$\text{Var}[X_t] = \mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2.$$

Se observa que la esperanza de X_t es cero pues $\varepsilon_t \sim SWN(0, 1)$ y es independiente de σ_t

$$\mathbb{E}[X_t] = \mathbb{E}[\sigma_t \varepsilon_t] = \mathbb{E}[\sigma_t] \mathbb{E}[\varepsilon_t] = 0.$$

Por lo tanto,

$$\begin{aligned} \text{Var}[X_t] &= \mathbb{E}[X_t^2] \\ &= \mathbb{E}[\sigma_t^2] \mathbb{E}[\varepsilon_t^2] \end{aligned}$$

$$\begin{aligned}
\text{Var}[X_t] &= \mathbb{E}[\sigma_t^2] \\
&= \mathbb{E}\left[\omega + \sum_{i=1}^q \alpha_i X_{t-i}^2\right] \\
&= \omega + \sum_{i=1}^q \alpha_i \mathbb{E}[X_{t-i}^2].
\end{aligned}$$

Pero como el proceso X es estacionario y de media cero, la varianza satisface.

$$\begin{aligned}
\text{Var}[X_t] &= \omega + \sum_{i=1}^q \alpha_i \text{Var}[X_{t-i}] \\
&= \omega + \sum_{i=1}^q \alpha_i \text{Var}[X_t].
\end{aligned}$$

De esta forma,

$$\text{Var}[X_t] - \text{Var}[X_t] \sum_{i=1}^q \alpha_i = \omega$$

y de tal forma,

$$\text{Var}[X_t] = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i},$$

■

Observación. En la Sección A.1 del Apéndice A se habla del efecto de la leptocurtosis de los datos financieros. Dado que $\mathbb{E}[X_t] = 0$, la ecuación (A.1) puede verse como,

$$K_X = \frac{\mathbb{E}[X_t^4]}{\text{Var}[X_t]^2}. \quad (2.24)$$

De la igualdad (2.24) se observa que además de la condición de estacionariedad, se requiere que X_t tenga cuarto momento finito.

Proposición 2.8 *Si un proceso ARCH(q) es estacionario con momentos de orden 4 finitos, entonces se cumple:*

$$K_X = \frac{\mathbb{E}[X_t^4]}{\text{Var}[X_t]^2} \geq \mathbb{E}[\varepsilon_t^4] = K_\varepsilon$$

Demostración. Sea $X = \{X_t : t \in \mathbb{Z}\}$ un proceso ARCH(q), al ser X estacionario, este posee segundos momentos finitos y podemos ver su varianza como (ver Ross, 2009)

$$\text{Var}[X_t] = \mathbb{E}[\text{Var}[X_t | F_{t-1}]] + \text{Var}[\mathbb{E}[X_t | F_{t-1}]]. \quad (2.25)$$

Considerando la **Proposición 2.6** se sigue que:

$$\text{Var}[X_t] = \mathbb{E}[\sigma_t^2] + \text{Var}[0] = \mathbb{E}[\sigma_t^2]. \quad (2.26)$$

Como X es estacionario, se cumple que su varianza es finita y constante. Por lo cuál podemos aplicar la desigualdad de Jensen (ver Protter, 2004) a σ_t^2 donde consideramos la función convexa $f(x) = x^2$:

$$\mathbb{E}[\sigma_t^4] \geq \mathbb{E}[\sigma_t^2]^2. \quad (2.27)$$

De (2.24) se observa que:

$$\begin{aligned} \frac{\mathbb{E}[X_t^4]}{\text{Var}[X_t]^2} &= \frac{\mathbb{E}[X_t^4]}{\mathbb{E}[\sigma_t^2]^2} \\ &= \frac{\mathbb{E}[\sigma_t^4]\mathbb{E}[\varepsilon_t^4]}{\mathbb{E}[\sigma_t^2]^2}. \end{aligned}$$

Considerando (2.27)

$$\begin{aligned} \frac{\mathbb{E}[X_t^4]}{\text{Var}[X_t]^2} &\geq \frac{\mathbb{E}[\sigma_t^2]^2\mathbb{E}[\varepsilon_t^4]}{\mathbb{E}[\sigma_t^2]^2} \\ &= \mathbb{E}[\varepsilon_t^4]. \end{aligned}$$

■

Por lo tanto, bajo los supuestos de la **Proposición 2.8**, la curtosis de un proceso ARCH(q) será mayor a la curtosis de la distribución del ruido blanco. Sin embargo, los supuestos en el modelo ARCH requeridos para que un proceso tenga cuartos momentos finitos hacen que el modelo sea más restrictivo, para ejemplificar este hecho mostramos el caso de un proceso ARCH(1).

McNeil ,*et.al.*(2005) demuestra que un proceso ARCH(1) posee momentos de orden $2m$ finitos si y sólo si $\mathbb{E}[\varepsilon_t^{2m}] < \infty$ y además el parámetro α_1 satisface:

$$\alpha_1 < \mathbb{E}[\varepsilon_t^{2m}]^{-m^{-1}}. \quad (2.28)$$

Para el caso $m = 2$, este resultado sugiere que en el ruido blanco de la **Definición 2.16** se suponga una distribución con cuartos momentos finitos, tal como una normal estándar o una $t - Student$ con más de 5 grados de libertad. Para el caso de una normal estándar se tiene que $\mathbb{E}[\varepsilon_t^4] = 3$, por lo tanto bajo (2.28) se requiere que:

$$\alpha_1 < \frac{1}{\sqrt{3}}.$$

Para el caso de ARCH(q) con ordenes superiores, las condiciones para los parámetros se vuelven más restrictivas, lo cual es una desventaja del modelo.

Observación. Considerando que el proceso ARCH(q) se vuelve restrictivo para ordenes q grandes y el hecho que algunos datos requieren de las hipótesis necesarias para estudiar un proceso estacionario que capture el efecto de leptocurtosis, Bollerslev (1986) desarrolló una generalización de estos modelos, llamados procesos autoregresivos condicionalmente heterocedásticos generalizados (GARCH, por sus siglas en inglés), en los cuales para modelar la volatilidad a un cierto tiempo no sólo se consideran los datos cuadrados, sino las volatilidades pasadas. Por lo cuál los procesos GARCH serán una alternativa a modelos ARCH(q) para ordenes de q grandes. En la siguiente sección se introduce la definición de estos procesos así como algunas de sus propiedades.

2.3.3. Procesos GARCH(p, q)

Definición 2.17 Sea $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ tal que $\varepsilon_t \sim \text{SWN}(0, 1)$. Un **proceso autoregresivo condicionalmente heterocedástico generalizado** de ordenes $p, q \in \mathbb{N}$, denotado por GARCH(p, q), es un proceso $X = \{X_t : t \in \mathbb{Z}\}$ que satisface:

$$X_t = \sigma_t \varepsilon_t,$$

donde para $\omega > 0$ finito y $\alpha_i, \beta_j > 0, i = 1, \dots, q, j = 1, \dots, p, \sigma_t^2$ puede expresarse como:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2, \quad t \in \mathbb{Z}.$$

Observación. Por la estructura del modelo GARCH(p, q), la **Proposición 2.6** también es válida para ellos. Además por la manera en la que se define $\sigma_t^2, t \in \mathbb{Z}$ se puede capturar el efecto de volatilidad en grupos (*volatility clustering*).

Proposición 2.9 Sea $X = \{X_t : t \in T\}$ un proceso GARCH(p, q), entonces la distribución condicional de X_t dada F_t posee media cero y varianza σ_t^2 .

Demostración. Es análoga a la demostración de la **Proposición 2.6**. ■

Particularmente, en este trabajo se hará uso de procesos $GARCH(1, 1)$. Analizaremos este tipo de procesos y posteriormente haremos una descripción de sus propiedades generalizadas.

Por la **Definición 2.17** un proceso $GARCH(1, 1)$ queda definido por:

$$X_t = \sigma_t \varepsilon_t, \quad t \in \mathbb{Z}$$

donde la varianza σ_t^2 está dada por:

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad t \in \mathbb{Z}$$

Proposición 2.10 *Un proceso $GARCH(1, 1)$ es estacionario si $\alpha_1 + \beta_1 < 1$*

Demostración. Por la definición de σ_t , se observa que:

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \omega + \alpha_1 \sigma_{t-1}^2 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ &= \omega + (\alpha_1 \varepsilon_{t-1}^2 + \beta_1) \sigma_{t-1}^2. \end{aligned}$$

Observamos que $S = \{\sigma_t^2 : t \in \mathbb{Z}\}$ es un ejemplo de un proceso del tipo (2.20), donde $A_t = \alpha_1 \varepsilon_{t-1}^2 + \beta_1$ y $B_t = \omega$. Para demostrar que existe una única solución estacionaria de la forma (2.21), se deben verificar las condiciones de la **Proposición 2.4**, es decir:

$$\mathbb{E}[\max\{0, \ln |\omega|\}] < \infty, \quad \mathbb{E}[\ln |\alpha_1 \varepsilon_{t-1}^2 + \beta_1|] < 0.$$

La condición $\mathbb{E}[\max\{0, \ln |\omega|\}] < \infty$ es válida al ser $\omega > 0$.

Para la segunda condición, note que por la hipótesis se tiene $\alpha_1 + \beta_1 < 1$. Adicionalmente, por la desigualdad de Jensen se obtiene,

$$\begin{aligned} \mathbb{E}[\ln |\alpha_1 \varepsilon_{t-1}^2 + \beta_1|] &= \mathbb{E}[\ln(\alpha_1 \varepsilon_{t-1}^2 + \beta_1)] \\ &\leq \ln \mathbb{E}[\alpha_1 \varepsilon_{t-1}^2 + \beta_1]. \end{aligned}$$

Como $\varepsilon_t \sim SWN(0, 1)$, entonces,

$$\begin{aligned} \mathbb{E}[\ln |\alpha_1 \varepsilon_{t-1}^2 + \beta_1|] &\leq \ln(\alpha_1 \mathbb{E}[\varepsilon_{t-1}^2] + \beta_1) \\ &= \ln(\alpha_1 + \beta_1) \\ &< \ln(1) = 0. \end{aligned}$$

Como las condiciones se satisfacen, entonces por la **Proposición 2.4** la única solución estrictamente estacionaria está dada por,

$$\begin{aligned}\sigma_t^2 &= \omega + \sum_{i=1}^{\infty} \omega \prod_{j=0}^{i-1} (\omega \varepsilon_{t-j-1}^2 + \beta_1). \\ &= \omega \left(1 + \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} (\omega \varepsilon_{t-j-1}^2 + \beta_1) \right) \\ &= \omega \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^i (\omega \varepsilon_{t-j}^2 + \beta_1) \right).\end{aligned}$$

Por lo tanto, el proceso X puede reescribirse como:

$$X_t = \varepsilon_t \left[\omega \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^i (\omega \varepsilon_{t-j}^2 + \beta_1) \right) \right]^{1/2}. \quad (2.29)$$

Considerando que X_t es el producto de dos procesos estacionarios estrictos, X_t también será estrictamente estacionario. Además se observa que los segundos momentos son finitos.

$$\begin{aligned}\mathbb{E}[X_t^2] &= \mathbb{E}[\varepsilon_t^2 \sigma_t^2] \\ &= \mathbb{E} \left[\omega \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^i (\omega \varepsilon_{t-j}^2 + \beta_1) \right) \right] \\ &= \omega \mathbb{E} \left[\sum_{i \in \mathbb{N}} \prod_{j=1}^i (\omega \varepsilon_{t-j}^2 + \beta_1) \right].\end{aligned}$$

Por el **Teorema A.1** del **Apéndice A**,

$$\mathbb{E}[X_t^2] = \omega \sum_{i \in \mathbb{N}} \mathbb{E} \left[\prod_{j=1}^i (\omega \varepsilon_{t-j}^2 + \beta_1) \right].$$

Como $\varepsilon_t \sim SWN(0, 1)$, se tiene,

$$\begin{aligned}\mathbb{E}[X_t^2] &= \omega \sum_{i \in \mathbb{N}} \prod_{j=1}^i \mathbb{E}[\omega \varepsilon_{t-j}^2 + \beta_1] \\ &= \omega \sum_{i \in \mathbb{N}} \prod_{j=1}^i (\alpha_1 + \beta_1).\end{aligned}$$

Bajo el supuesto que $\alpha_1 + \beta_1 < 1$,

$$\begin{aligned}\mathbb{E}[X_t^2] &= \omega \sum_{i \in \mathbb{N}} (\alpha_1 + \beta_1)^i \\ &= \frac{\omega}{1 - \alpha_1 - \beta_1} < \infty.\end{aligned}$$

Por el **Teorema 2.1**, el proceso X es estacionario.

■

El siguiente resultado presenta las condiciones generales para que un proceso como el de la **Definición 2.17** sea estacionario.

Teorema 2.4 *Un proceso GARCH(p, q) es estacionario si:*

$$\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1. \quad (2.30)$$

Demostración. Ver Francq y Zakoïan (2010).

Observación. Francq y Zakoïan (2010) también indican que la única solución de X es un ruido blanco fuerte con varianza:

$$Var[X_t] = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j}. \quad (2.31)$$

Para el caso de la existencia de cuartos momentos finitos en un GARCH(1,1), en McNeil, *et, al.* (2005) demuestra que se requiere de la condición:

$$(\alpha_1 + \beta_1)^2 < 1 - 2\alpha_1^2. \quad (2.32)$$

Asumiendo que (2.32) es satisfecha, se obtiene el siguiente resultado.

Proposición 2.11 *La curtosis K_X de un proceso GARCH(1,1) con cuartos momentos finitos esta dada por:*

$$K_X = \frac{K_\varepsilon(1 - (\alpha_1 + \beta_1)^2)}{1 - (\alpha_1 + \beta_1)^2 - (K_\varepsilon - 1)\alpha_1^2}.$$

Demostración. Ver Tsay (2010).

■

Observación. Se observa que $K_X > K_\varepsilon$ siempre que $K_\varepsilon > 1$. Por lo tanto, un proceso GARCH(1,1) bajo un proceso ε_t de curtosis mayor a uno cumple el supuesto de leptocurtosis de las series de tiempo financieras (Tsay, 2010).

Para analizar algunas características de los procesos GARCH(1,1) se presenta los siguientes resultados.

Proposición 2.12 *Si $X = \{X_t : t \in T\}$ es un proceso GARCH(p, q) estacionario de cuartos momentos finitos, entonces $X^2 := \{X_t^2 : t \in T\}$ es un proceso ARMA(máx(p, q), p).*

Demostración. Sea $X = \{X_t : t \in T\}$ un proceso GARCH(p, q), consideramos el proceso $\eta = \{\eta_t : t \in \mathbb{Z}\}$ definido como sigue,

$$\eta_t = X_t^2 - \sigma_t^2. \quad (2.33)$$

Por (2.33), se observa,

$$\begin{aligned} X_t^2 - \eta_t &= \sigma_t^2 \\ &= \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \\ &= \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j (X_{t-j}^2 - \eta_{t-j}) \\ &= \omega + \sum_{i=1}^q \alpha_i X_{t-i}^2 + \sum_{j=1}^p \beta_j X_{t-j}^2 - \sum_{j=1}^p \beta_j \eta_{t-j}. \end{aligned}$$

Por lo tanto:

$$X_t^2 - \sum_{i=1}^q \alpha_i X_{t-i}^2 - \sum_{j=1}^p \beta_j X_{t-j}^2 = \omega + \eta_t - \sum_{j=1}^p \beta_j \eta_{t-j}.$$

Si $q > p$ se conviene $\beta_j = 0$ para $j = p + 1, \dots, q$, o si $p > q$ se toma $\alpha_i = 0$ para $i = q + 1, \dots, p$. Ambos casos permiten considerar,

$$X_t^2 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) X_{t-i}^2 = \omega + \eta_t - \sum_{j=1}^p \beta_j \eta_{t-j}. \quad (2.34)$$

Finalmente:

$$X_t^2 = \omega + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) X_{t-i}^2 + \eta_t - \sum_{j=1}^p \beta_j \eta_{t-j}. \quad (2.35)$$

donde $\eta = \{\eta_t : t \in \mathbb{Z}\}$ es un proceso de ruido blanco, tal como se muestra a seguir.

Primeramente la esperanza del proceso η está dada por:

$$\begin{aligned} \mathbb{E}[\eta_t] &= \mathbb{E}[X_t^2 - \sigma_t^2] \\ &= \mathbb{E}[X_t^2] - \mathbb{E}[\sigma_t^2]. \end{aligned}$$

Por definición de X_t y el hecho que $\varepsilon_t \sim SWN(0, 1)$

$$\begin{aligned} \mathbb{E}[\eta_t] &= \mathbb{E}[\sigma_t^2 \varepsilon_t^2] - \mathbb{E}[\sigma_t^2] \\ &= \mathbb{E}[\sigma_t^2] \mathbb{E}[\varepsilon_t^2] - \mathbb{E}[\sigma_t^2] \\ &= \mathbb{E}[\sigma_t^2] - \mathbb{E}[\sigma_t^2]. \end{aligned}$$

Por lo tanto la media del proceso $\eta = \{\eta_t : t \in \mathbb{Z}\}$ es igual a cero.

Además, la covarianza $t \neq s$ de las variables cumple:

$$\begin{aligned}\gamma(t, s) &= \mathbb{E}[\eta_t \eta_s] - \mathbb{E}[\eta_t] \mathbb{E}[\eta_s] \\ &= \mathbb{E}[\eta_t \eta_s] \\ &= \mathbb{E}[(\sigma_t^2 \varepsilon_t^2 - \sigma_t^2)(\sigma_s^2 \varepsilon_s^2 - \sigma_s^2)] \\ &= \mathbb{E}[\sigma_t^2 (\varepsilon_t^2 - 1) \sigma_s^2 (\varepsilon_s^2 - 1)].\end{aligned}$$

Sin pérdida de generalidad, suponga $s > t$. Como ε es un proceso de variables aleatorias independientes de observaciones pasadas, entonces,

$$\begin{aligned}\gamma(t, s) &= \mathbb{E}[\sigma_t^2 (\varepsilon_t^2 - 1) \sigma_s^2] \mathbb{E}[\varepsilon_s^2 - 1] \\ &= \mathbb{E}[\sigma_t^2 (\varepsilon_t^2 - 1) \sigma_s^2] (\mathbb{E}[\varepsilon_s^2] - 1) \\ &= \mathbb{E}[\sigma_t^2 (\varepsilon_t^2 - 1) \sigma_s^2] (1 - 1) = 0.\end{aligned}$$

Finalmente veamos el comportamiento de la varianza, la cual debe ser finita y constante.

$$\begin{aligned}Var[\eta_t] &= \mathbb{E}[\eta_t^2] - \mathbb{E}^2[\eta] \\ &= \mathbb{E}[\eta_t^2] \\ &= \mathbb{E}[(X_t^2 - \sigma_t^2)^2] \\ &= \mathbb{E}[X_t^4] - 2\mathbb{E}[X_t^2] \mathbb{E}[\sigma_t^2] + \mathbb{E}[\sigma_t^4].\end{aligned}$$

Asumiendo por hipótesis que X_t es un proceso estacionario, entonces los momentos $\mathbb{E}[X_t^2]$ y $\mathbb{E}[\sigma_t^2]$ existen. Además como X es un proceso de cuartos momentos finitos, entonces σ_t^4 también lo es. De esta manera, la varianza del proceso η es finita y por la **Definición 2.5**, η constituye un proceso de ruido blanco.

Al ser η_t un ruido blanco en (2.35), el proceso $X^2 = \{X_t^2 : t \in T\}$ conforma un proceso ARMA(máx(p, q), p).

■

Observación. Por lo tanto, si X_t sigue un proceso GARCH(p, q), podemos estudiar los datos cuadráticos mediante un modelo ARMA(máx(p, q), p). De este hecho se pueden elegir órdenes candidatos para el modelo GARCH(p, q) a través de las funciones de autocorrelación simple y parcial de los términos cuadráticos.

Una vez establecidos los órdenes del proceso GARCH(p, q), se procede a realizar la

estimación de parámetros, tema que se abordará en la siguiente capítulo. Si los residuales del proceso cumplen las hipótesis del proceso en la **Definición 2.17**, entonces es posible realizar pronósticos de los términos cuadráticos, dada la correlación que existe entre ellos en datos de activos financieros. Se observa que esto equivale a predecir el comportamiento de la varianza de X_t^2 condicionada a F_{t-1} , dado que para $h > 0$ se tiene.

$$\mathbb{E}[X_{t+h}^2 | F_t] = \mathbb{E}[\sigma_{t+h}^2 \varepsilon_{t+h}^2 | F_t].$$

Por la independencia entre X y ε y el hecho que $\varepsilon_t \sim SWN(0, 1)$, se tiene,

$$\begin{aligned} \mathbb{E}[X_{t+h}^2 | F_t] &= \mathbb{E}[\sigma_{t+h}^2 | F_t] \mathbb{E}[\varepsilon_{t+h}^2 | F_t] \\ &= \mathbb{E}[\sigma_{t+h}^2 | F_t]. \end{aligned}$$

Por lo tanto

$$\mathbb{E}[X_{t+h}^2 | F_t] = \mathbb{E}[\sigma_{t+h}^2 | F_t]. \quad (2.36)$$

Definición 2.18 *La h -ésima predicción de la varianza condicional de un proceso $GARCH(1,1)$, dada la información histórica F_t , se denotará por $\sigma_t^2(h)$ y se define por:*

$$\sigma_t^2(h) = \mathbb{E}[\sigma_{t+h}^2 | F_t] \quad (2.37)$$

Proposición 2.13 *Para un proceso $GARCH(1,1)$, el h -ésimo pronóstico de la varianza condicional σ_t^2 está dado por.*

$$\sigma_t^2(h) = \omega \sum_{i=0}^{h-1} (\alpha + \beta)^i + (\alpha + \beta)^{h-1} (\alpha X_t^2 + \beta \sigma_t^2).$$

Demostración. Sea $h > 0$, por la **Definición 2.18**, se tiene,

$$\sigma_t^2(h) = \mathbb{E}[\sigma_{t+h}^2 | F_t].$$

Por **Definición 2.17** se observa,

$$\begin{aligned} \mathbb{E}[\sigma_{t+h}^2 | F_t] &= \mathbb{E}[\omega + \alpha_1 X_{t+h-1}^2 + \beta_1 \sigma_{t+h-1}^2 | F_t] \\ &= \omega + \alpha_1 \mathbb{E}[X_{t+h-1}^2 | F_t] + \beta_1 \mathbb{E}[\sigma_{t+h-1}^2 | F_t]. \end{aligned}$$

De la igualdad (2.36) se sigue que,

$$\begin{aligned} \mathbb{E}[\sigma_{t+h}^2 | F_t] &= \omega + \alpha_1 \mathbb{E}[\sigma_{t+h-1}^2 | F_t] + \beta_1 \mathbb{E}[\sigma_{t+h-1}^2 | F_t] \\ &= \omega + (\alpha_1 + \beta_1) \mathbb{E}[\sigma_{t+h-1}^2 | F_t] \\ &= \omega + (\alpha_1 + \beta_1) \sigma_t^2(h-1). \end{aligned}$$

Nuevamente, dado que $\sigma_t^2(h-1) = \mathbb{E}[\sigma_{t+h-1}^2 | F_t]$, tenemos:

$$\sigma_t^2(h-1) = \omega + (\alpha_1 + \beta_1)\sigma_t^2(h-2).$$

Por lo tanto:

$$\begin{aligned}\sigma_t^2(h) &= \mathbb{E}[\sigma_{t+h}^2|F_t] = \omega + (\alpha_1 + \beta_1)\sigma_t^2(h-1) \\ &= \omega + (\alpha_1 + \beta_1) [\omega + (\alpha_1 + \beta_1)\sigma_t^2(h-2)] \\ &= \omega + \omega(\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2\sigma_t^2(h-2).\end{aligned}$$

De esta manera tenemos una relación recursiva para $\sigma_t^2(h)$, tal que para $j < h$ se tiene:

$$\sigma_t^2(h) = \mathbb{E}[\sigma_{t+h}^2|F_t] = \omega + \omega \sum_{k=1}^{j-1} (\alpha + \beta)^k + (\alpha + \beta)^j \sigma_t^2(h-j).$$

Considerando $j = h - 1$ resulta:

$$\begin{aligned}\sigma_t^2(h) &= \omega + \omega \sum_{k=1}^{h-2} (\alpha + \beta)^k + (\alpha + \beta)^{h-1} \sigma_t^2(1) \\ &= \omega \sum_{i=0}^{h-2} (\alpha + \beta)^i + (\alpha + \beta)^{h-1} \sigma_t^2(1).\end{aligned}$$

Aplicando la **Definición 2.18** y la igualdad (2.36), se tiene que,

$$\begin{aligned}\sigma_t^2(h) &= \omega \sum_{i=0}^{h-2} (\alpha + \beta)^i + (\alpha + \beta)^{h-1} \mathbb{E}[\sigma_{t+1}^2|F_t] \\ &= \omega \sum_{i=0}^{h-2} (\alpha + \beta)^i + (\alpha + \beta)^{h-1} \mathbb{E}[X_{t+1}^2|F_t].\end{aligned}$$

Por la **Proposición 2.9** se sigue que,

$$\begin{aligned}\sigma_t^2(h) &= \omega \sum_{i=0}^{h-2} (\alpha + \beta)^i + (\alpha + \beta)^{h-1} \sigma_{t+1} \\ &= \omega \sum_{i=0}^{h-2} (\alpha + \beta)^i + (\alpha + \beta)^{h-1} (\omega + \alpha X_t + \beta \sigma_t) \\ &= \omega \sum_{i=0}^{h-1} (\alpha + \beta)^i + (\alpha + \beta)^{h-1} (\alpha X_t^2 + \beta \sigma_t^2).\end{aligned}$$

■

En el **Apéndice D**, se desarrolló la función predicción sigma para obtener los pronósticos de un proceso GARCH(1, 1).

Capítulo 3

Inferencia bayesiana en modelos

GARCH(1,1)

La estimación de parámetros del modelo que describe a una muestra \mathbf{x} es sustancial para ajustar un proceso que concuerde con la serie de datos considerados. Desde un punto de vista clásico, el método más utilizado es mediante la maximización de la función de verosimilitud, que está dado en la **Definición A.2.** del **Apéndice A.**

En este capítulo, se considera una alternativa para esta metodología usando el enfoque dado por la estadística bayesiana. Nos basaremos en Ardia (2008) y su implementación usando los métodos Monte Carlo basados en cadenas de Markov bajo la paquetería BayesGARCH del software R, que puede obtenerse a través de <https://www.r-project.org/>

3.1. GARCH(1,1) con innovaciones t-Student

Nos enfocaremos en el caso más sencillo de los modelos GARCH, que sustituye el uso de modelos ARCH(q) para órdenes de q grandes. El modelo considerado se ajusta a una gran variedad de activos financieros y portafolio de inversión. Se considera el caso donde el proceso de ruido blanco estricto $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ posee variables con distribución t -Student estandarizada. Este hecho permite capturar de manera más eficiente el efecto de leptocurtosis que presentan los activos financieros (Ardia, 2010).

De esta manera el proceso GARCH(1, 1) con innovaciones t se escribe como sigue:

$$X_t = \sigma_t \varepsilon_t, \quad t \geq 0.$$

Como la **Definición 2.17** requiere que tenga media cero y varianza uno, no es posible suponer directamente que las variables del proceso ε siguen una distribución t -Student. Por esta razón se utilizarán distribuciones t -Student estandarizadas, es decir,

$$\varepsilon_t \sim \frac{t(\nu)}{\sqrt{\nu/(\nu-2)}}, \quad \nu > 2.$$

Para realizar inferencia bayesiana se requieren de las distribuciones *a priori* de los parámetros y la función de verosimilitud del proceso. El uso directo de innovaciones con distribución t estandarizada, complica el desarrollo de estas funciones. Para facilitar el análisis de la inferencia con este tipo de inovaciones, se utiliza el **Teorema A.3** del **Apéndice A** para redefinir este tipo de procesos.

Por lo tanto, se puede considerar la siguiente definición (Ardia, 2010).

Definición 3.1 *Un proceso $X = \{X_t : t \in \mathbb{Z}\}$ se llama GARCH(1,1) con inovaciones $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$ distribuidas t -Student estandarizada de ν grados de libertad, si puede escribirse como sigue.*

$$X_t = \sigma_t \sqrt{\frac{(\nu-2)\tau_t}{\nu}} z_t, \quad (3.1)$$

donde para $t \in \mathbb{Z}$, $z_t \sim N(0, 1)$, $\tau_t \sim \text{InvGamma}(\nu/2, \nu/2)$, y además,

$$\sigma_t^2 = \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad (3.2)$$

con $\omega > 0$, $\alpha_1, \beta_1 \geq 0$.

Observación. De las ecuaciones (3.1) y (3.2) se puede notar que se debe estimar el siguiente vector de parámetros.

$$\boldsymbol{\theta} = (\omega, \alpha_1, \beta_1, \nu).$$

Además de considerar una serie de rendimientos históricos $\mathbf{x} = \{x_1, \dots, x_n\}$, en el desarrollo de la estimación también se deben considerar las variables $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)$, generadas por una distribución inversa gamma.

3.2. Función de verosimilitud

De acuerdo a McNeil, *et al.* (2005), en el caso de procesos ARCH(1) y GARCH(1,1), la función de verosimilitud puede obtenerse a través del **Teorema A.4** del **Apéndice A**, para establecer la siguiente proposición.

Proposición 3.1 *Para un proceso GARCH(1,1), la función de verosimilitud condicionada a las observaciones x_1 y σ_1 está dada por:*

$$f(\mathbf{x} \mid \theta, x_1, \sigma_1) = \prod_{i=2}^n \frac{1}{\sigma_i} h\left(\frac{X_i}{\sigma_i}\right), \quad (3.3)$$

donde h es la función de densidad de ε_t y $\sigma_i = \sqrt{\omega + \alpha_1 X_{i-1}^2 + \beta_1 \sigma_{i-1}^2}$, $i = 1, \dots, n$.

Demostración. Ver McNeil, *et al.* (2005).

Observaciones.

1. De acuerdo a McNeil, *et al.* (2005) y Tsay (2010), en el caso de los procesos GARCH(1, 1) es necesario aproximar a la función de verosimilitud a través de una función de verosimilitud condicional debido a que es complicado obtener una expresión analítica de las funciones de densidad de X_1 y σ_1 .
2. En el caso del proceso GARCH(1,1) con innovaciones t - Student, puede utilizarse directamente la **Definición A.1** en su versión estandarizada para realizar el desarrollo de (3.3), sin embargo este se torna complicado. Es por ello que Ardia (2010) utiliza el ajuste (3.1) para facilitar el desarrollo de la función de verosimilitud.

Proposición 3.2 *La función de verosimilitud para un proceso GARCH(1,1) con innovaciones t - Student estandarizadas con ν grados de libertad satisface:*

$$f(\mathbf{x} \mid \theta, x_1, \sigma_1) \propto \frac{1}{(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x}\right), \quad (3.4)$$

donde $\mathbf{x} = (x_1, \dots, x_n)$ y Σ es una matriz diagonal:

$$\Sigma = \begin{pmatrix} \tau_1 \frac{\nu - 2}{\nu} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \tau_2 \frac{\nu - 2}{\nu} \sigma_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \tau_n \frac{\nu - 2}{\nu} \sigma_n^2 \end{pmatrix}, \quad (3.5)$$

Demostración. Al considerar en (3.1) al modelo GARCH(1,1) con inovaciones t - Student estandarizadas en términos de inovaciones normales z_t , retomamos el desarrollo de la función de verosimilitud utilizando (3.3), es decir,

$$f(\mathbf{x} \mid \theta, x_1, \sigma_1) = \prod_{i=1}^n \frac{1}{s_i} h\left(\frac{x_i}{s_i}\right), \quad (3.6)$$

donde $h(\cdot)$ es la función de densidad de una normal estandar y s_i está dado por:

$$s_i = \sigma_i \sqrt{\frac{(\nu - 2)\tau_i}{\nu}}, \quad i = 1, \dots, n.$$

De (3.6) y considerando que $h(\cdot)$ es la función de las variables z_i , $i = 1, \dots, n$, entonces,

$$\begin{aligned} f(\mathbf{x} \mid \theta, x_1, \sigma_1) &= \prod_{i=1}^n \frac{1}{s_i} h\left(\frac{x_i}{s_i}\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi s_i^2}} \exp\left(-\frac{x_i^2}{2s_i^2}\right). \end{aligned}$$

Entonces,

$$f(\mathbf{x} \mid \theta, x_1, \sigma_1) = \left(\frac{1}{2\pi}\right)^{n/2} \prod_{i=1}^n \frac{1}{\sqrt{s_i^2}} \times \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{x_i^2}{s_i^2}\right).$$

Se observa que s_i^2 corresponde al elemento de la columna y fila i de la matriz (3.5), la cual por ser diagonal tendrá como determinante al producto de sus entradas. Por ende,

$$f(\mathbf{x} \mid \theta, x_1, \sigma_1) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{x_i^2}{s_i^2}\right),$$

Por otro lado, se observa que, al ser Σ una matriz diagonal, las entradas Σ_{jj}^{-1} , $j = 1, \dots, n$ de su matriz inversa Σ^{-1} están dadas por:

$$\Sigma_{jj}^{-1} = \tau_j^{-1} \frac{\nu}{\nu - 2} \sigma_j^{-2}.$$

Por lo tanto, si $\mathbf{x} = (x_1, \dots, x_n)$ se verifica que:

$$\mathbf{x}' \Sigma^{-1} \mathbf{x} = \sum_{i=1}^n \frac{x_i^2}{s_i^2}. \quad (3.7)$$

Como el término $\left(\frac{1}{2\pi}\right)^{n/2}$ es constante en relación a \mathbf{x} , se concluye que:

$$f(\mathbf{x} \mid \theta, x_1, \sigma_1) \propto \frac{1}{(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x}\right). \quad (3.8)$$

■

Observación. Desde un punto de vista clásico, la función de verosimilitud se utiliza para estimar los valores de los parámetros, tal como se introduce en la **Definición A.2** del **Apéndice A**. Para desarrollar esta tarea se utiliza la paquetería `fgarch`. Sin embargo, esta metodología puede proporcionar estimadores que no sean congruentes con la propiedad de estacionariad que requieren este tipo de procesos. Por lo tanto se debe realizar una optimización de la función de verosimilitud con restricciones que proporcionen estimadores congruentes con las características de los log-retornos financieros. En procesos con un mayor número de parámetros, esta optimización suele complicarse (Ardia, 2008).

3.3. Distribuciones *a priori* y *a posteriori*

Debido a la incorporación de la generación de variables aleatorias $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)$ para el desarrollo de la función de verosimilitud de un proceso GARCH(1,1) con innovaciones t estandarizadas, del **Teorema 1.1** se sigue que:

$$p_{post}(\boldsymbol{\theta}, \boldsymbol{\tau}) \propto f(\mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\tau}) p_{priori}(\boldsymbol{\theta}, \boldsymbol{\tau}), \quad (3.9)$$

donde $\boldsymbol{\theta} = (\omega, \alpha_1, \beta_1, \nu)$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)$ y $\mathbf{x} = \{x_1, \dots, x_n\}$.

De esta manera, se deben introducir distribuciones a priori para los parámetros del proceso GARCH(1,1) con innovaciones t estandarizadas. Asumiendo independencia entre los parámetros y considerando que la generación del vector $\boldsymbol{\tau}$ depende de los grados de libertad, se tiene que:

$$\begin{aligned} p_{priori}(\boldsymbol{\theta}, \boldsymbol{\tau}) &= p_{priori}(\omega) p_{priori}(\alpha_1) p_{priori}(\beta_1) p_{priori}(\nu, \boldsymbol{\tau}) \\ &= p_{priori}(\omega) p_{priori}(\alpha_1) p_{priori}(\beta_1) p(\boldsymbol{\tau} \mid \nu) p_{priori}(\nu), \end{aligned}$$

donde $p(\boldsymbol{\tau} \mid \nu)$ es la función de densidad de $\boldsymbol{\tau}$ dado ν . A continuación se presentan las propuestas establecidas por Ardia (2010) para la elección de las distribuciones *a priori* de los parámetros. Por la **Definición 2.17** se observa que los parámetros ω , α_1 y β_1 asumen valores no negativos. Por esta razón se utilizan distribuciones normales que asuman valores estrictamente positivos.

(a) Distribución *a priori* de ω y α_1 .

A pesar de la independencia entre los parámetros α_1 y ω , por practicidad la paquetería en R `bayesGARCH` de Ardia (2010) maneja una distribución *a priori* conjunta

para estos parámetros, es decir,

$$p_{priori}(\omega, \alpha_1) = \begin{cases} f(\omega, \alpha_1), & (\omega, \alpha_1) \in \mathbb{R}_+^2 \\ 0, & (\omega, \alpha_1) \notin \mathbb{R}_+^2. \end{cases} \quad (3.10)$$

donde $f(\omega, \alpha_1)$ es la función de densidad conjunta de una distribución normal bivariada de vector de media $(\mu_\omega, \mu_{\alpha_1})$ y matriz de varianzas y covarianzas

$$\begin{pmatrix} \sigma_\omega^2 & 0 \\ 0 & \sigma_{\alpha_1}^2 \end{pmatrix}.$$

En ausencia de información inicial se recomienda $(\mu_\omega, \mu_{\alpha_1}) = (0, 0)$ y $\sigma_{\alpha_1} = \sigma_\omega = 1$.

(b) Distribución a priori de β_1 .

Para la distribución a priori del parámetro β_1 se utiliza la misma estructura para una distribución normal truncada univariada, es decir.

$$p_{priori}(\beta_1) = \begin{cases} f(\beta_1), & \beta_1 \geq 0 \\ 0, & \beta_1 < 0, \end{cases} \quad (3.11)$$

donde $f(\beta_1)$ es la función de densidad de una distribución normal de parámetros μ_β y σ_β^2 . Al igual que la distribución a priori (2.10), en ausencia de información inicial se recomienda manejar $\mu_\beta = 0$ y $\sigma_\beta^2 = 1$, es decir, una distribución normal estándar truncada.

(c) Distribución a priori de ν .

Para la distribución *a priori* de parámetro ν se utiliza una distribución exponencial de hiperparámetro λ y una traslación $\delta > 1$. De acuerdo a Ardia (2010) y Deschamps (2006), esta distribución logra capturar el hecho que entre mayor sean los grados de libertad en (3.1), las innovaciones tendrán una distribución que se aproxima a la de una normal estándar. De esta forma,

$$p_{priori}(\nu) = \begin{cases} \lambda \exp[-\lambda(\nu - \delta)] & \nu > \delta, \\ 0 & \nu \leq \delta. \end{cases} \quad (3.12)$$

(d) Distribución condicional de τ dado ν .

Finalmente se determina una expresión analítica para la función de densidad condicional del vector τ dados los grados de libertad ν , que equivale a obtener la función

de verosimilitud de un vector aleatorio de variables independientes con distribución inversa gamma de parámetros $\nu/2$.

Proposición 3.3 *La función de densidad conjunta de $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)$ condicionada de ν , donde cada τ_i tiene una distribución inversa gamma de parámetros igual a $\nu/2$, está dada por:*

$$p(\boldsymbol{\tau} | \nu) = \left(\frac{\nu}{2}\right)^{n\nu/2} \Gamma\left(\frac{\nu}{2}\right)^{-n} \left(\prod_{i=1}^n \tau_i\right)^{-\frac{\nu}{2}-1} \times \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{\nu}{\tau_i}\right), \quad (3.13)$$

Demostración. En el caso de la función de densidad condicional $p(\boldsymbol{\tau} | \nu)$, considerando que $\boldsymbol{\tau}$ es un vector aleatorio de variables independientes e idénticamente distribuidas $InvGamma(\nu/2, \nu/2)$ se sigue que:

$$\begin{aligned} p(\boldsymbol{\tau} | \nu) &= p(\tau_1, \dots, \tau_n | \nu) \\ &= \prod_{i=1}^n p(\tau_i | \nu), \end{aligned}$$

Como $\tau_i \sim InvGamma(\nu/2, \nu/2)$, entonces:

$$p(\tau_i | \nu) = \frac{(\nu/2)^{(\nu/2)}}{\Gamma(n/2)} \tau_i^{-\nu/2-1} \exp\left(-\frac{\nu}{2\tau_i}\right), \quad (3.14)$$

y por ende, sustituyendo (3.14) en el desarrollo de $p(\boldsymbol{\tau} | \nu)$:

$$\begin{aligned} p(\boldsymbol{\tau} | \nu) &= \prod_{i=1}^n \frac{(\nu/2)^{(\nu/2)}}{\Gamma(n/2)} \tau_i^{-\nu/2-1} \exp\left(-\frac{\nu}{2\tau_i}\right), \\ &= \frac{(\nu/2)^{(n\nu/2)}}{\Gamma(n/2)^n} \left(\prod_{i=1}^n \tau_i\right)^{-\frac{\nu}{2}-1} \prod_{i=1}^n \exp\left(-\frac{\nu}{2\tau_i}\right) \\ &= \frac{(\nu/2)^{(n\nu/2)}}{\Gamma(n/2)^n} \left(\prod_{i=1}^n \tau_i\right)^{-\frac{\nu}{2}-1} \times \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{\nu}{\tau_i}\right), \end{aligned}$$

Por lo tanto:

$$p(\boldsymbol{\tau} | \nu) = \left(\frac{\nu}{2}\right)^{n\nu/2} \Gamma\left(\frac{\nu}{2}\right)^{-n} \left(\prod_{i=1}^n \tau_i\right)^{-\frac{\nu}{2}-1} \times \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{\nu}{\tau_i}\right). \quad (3.15)$$

■

De esta manera, tomando en cuenta las distribuciones *a priori* de los parámetros, la expresión (3.13) de la función $p(\boldsymbol{\tau} \mid \nu)$ y el desarrollo de la función verosimilitud del proceso en la **Proposición 3.2**, se tiene una expresión proporcional a la distribución *a posteriori*, a partir de (3.9).

A través de la distribución *a posteriori*, Ardia (2008) ejecuta el algoritmo de Metropolis-Hastings mediante su función `bayesGARCH`. Este permite agregar condiciones a la generación de los valores de la distribución *a posteriori*, tal como (2.31), con el objetivo para garantizar la estacionariedad del proceso.

Observación. Para más detalles sobre las distribuciones candidatas y probabilidades de aceptación del Metropolis-Hastings, utilizadas para muestrear valores de la distribución *a posteriori*, se recomienda revisar el Capítulo 5 de Ardia (2008).

Capítulo 4

Aplicación

A continuación se presentará un análisis de la volatilidad de una de las series con mayor relevancia en el sistema financiero mexicano, el tipo de cambio peso mexicano y dólar estadounidense (USD/MXN), mediante el uso de un modelo GARCH(1, 1) con innovaciones t -Student. Se realizará un análisis exploratorio para verificar que la serie de datos puede ser una realización del modelo. También se llevará a cabo la estimación de parámetros mediante estadística bayesiana y se analizará la convergencia de la serie por el algoritmo Monte Carlo vía cadenas de Markov. Finalmente se analizará el impacto de los resultados en la administración de riesgos financieros futuros.

4.1. Análisis exploratorio

El tipo de cambio se define como una referencia para determinar el número de unidades de moneda nacional que deben pagarse para obtener una moneda extranjera y funge como indicador de competitividad de un país con respecto a otras naciones. En el caso de México, desde diciembre de 1994, debido a diversos acontecimientos económicos y sociales, se adoptó un régimen cambiario de libre flotación, es decir el precio del tipo de cambio se determina por la oferta y la demanda de divisas el mercado sin la intervención de las autoridades. Debido a esto y a la relación comercial existente entre Estados Unidos y México, se ha decidido realizar un estudio de la volatilidad del tipo de cambio USD/MXN. Esta se estudiará utilizando precios históricos diarios a partir del 2 de enero del 2008 hasta el 31 de marzo del 2020, los cuales capturan varios escenarios de apreciación y depreciación, incluyendo periodos de alta volatilidad como la crisis del 2008 y el inicio

de la cuarentena en México debido al virus SARS-CoV-2. Los datos proporcionados por Banco de México desde la página <https://www.banxico.org.mx/tipcamb> se observan en la Figura 4.1.

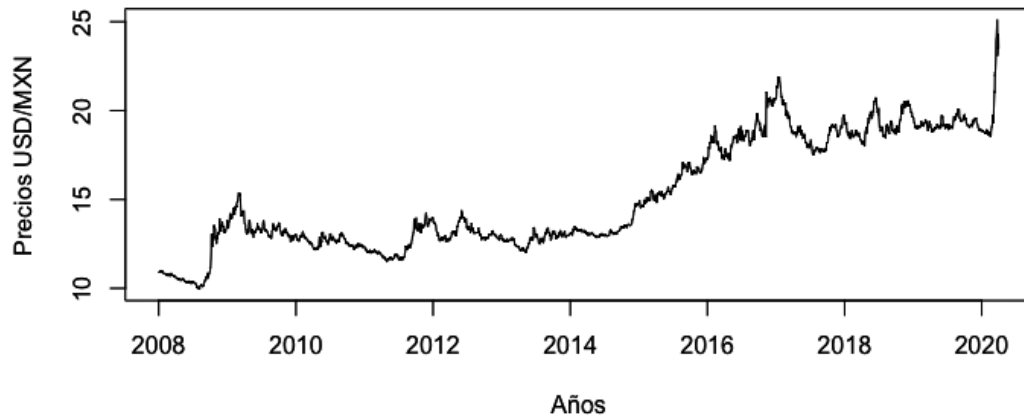


Figura 4.1: Precios de cierre del USD/MXN desde el 2 de Enero 2008 hasta 31 de Marzo 2020

Se consideran los log-retornos de los precios históricos para que la serie de datos pueda estudiarse mediante un proceso estacionario. Los datos resultantes en términos de porcentajes ($100r_t\%$) se observan en la Figura 4.2.

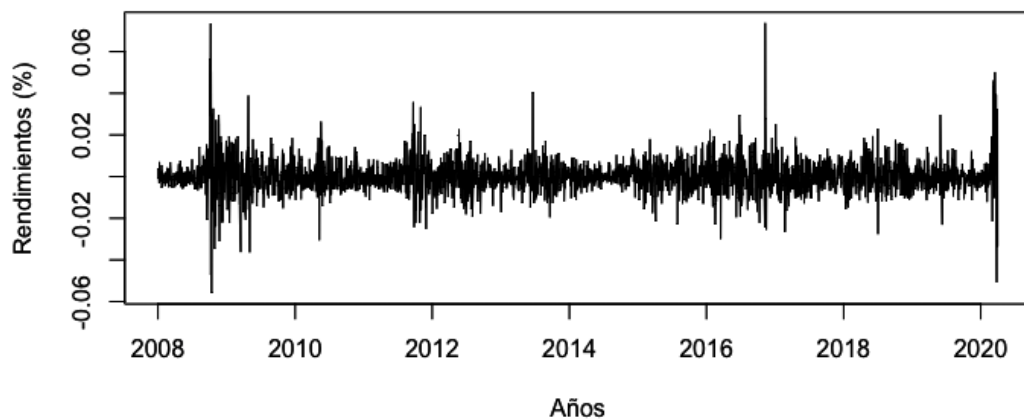


Figura 4.2: Log-retornos diarios del USD/MXN desde el 2 de Enero 2008 hasta 31 de Marzo 2020

En busca de una distribución de probabilidad que se ajuste a los log-retornos históricos del tipo de cambio USD/MXN, se observa que estos cumplen un efecto de leptocurtosis, pues cuando se compara el histograma de los rendimientos con respecto al de una distribución normal, cuyos parámetros son la media y la varianza muestrales de la serie de datos, se obtiene:

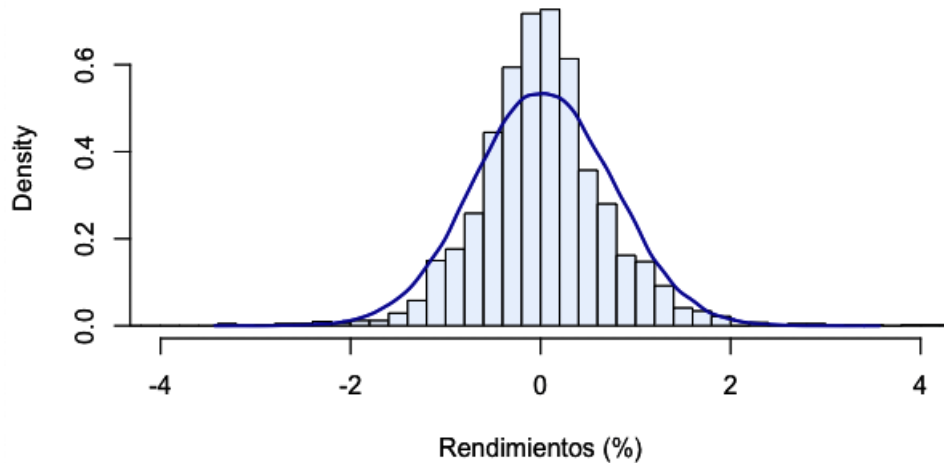


Figura 4.3: Histograma de los rendimientos con respecto una distribución normal con parámetros correspondientes a la media y varianza muestral de los datos.

El efecto de leptocurtosis de la Figura 4.3 puede verificarse a través de la curtosis de los rendimientos, la cual es 14.0309. Si nuestros datos se ajustarán a una distribución normal se desearía que la curtosis se aproximara a tres. Ante este hecho, algunos analistas de riesgos financieros asumen que los rendimientos pueden estudiarse a través de variables aleatorias independientes con distribución t -Student. Sin embargo, esto no considerará algunos supuestos implícitos en los rendimientos, tal como se muestra a continuación.

Para determinar la correlación y dependencia en los datos utilizamos un correlograma. Se analiza el comportamiento de las funciones de autocorrelación muestral simple y parcial de los log-retornos, así como la serie de los cuadrados de estos datos. Esto se presenta en la Figura 4.4, donde ACF y PACF indican función de autocorrelación simple y parcial, respectivamente.

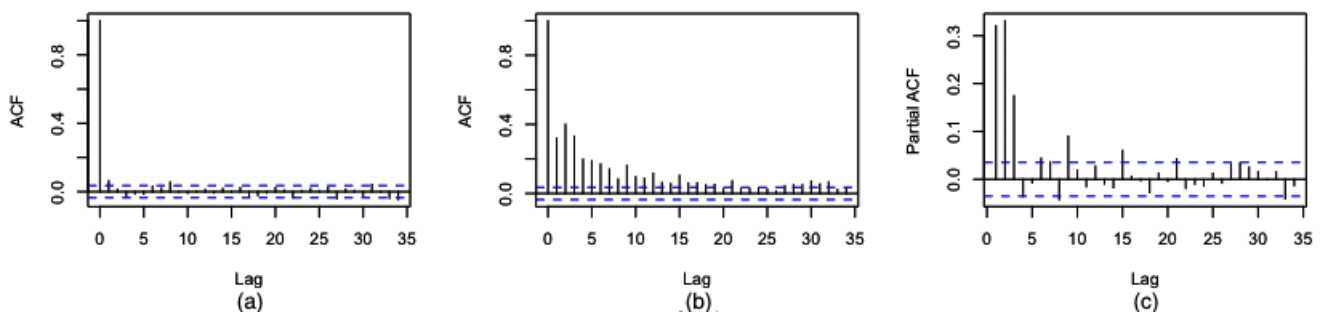


Figura 4.4: (a) ACF muestral log-retornos, (b) ACF muestral log-retornos cuadráticos, (c) PACF1 muestral rendimientos cuadráticos

De la Figura 4.4 se puede observar que los rendimientos del tipo de cambio USD/MXN no presentan correlación pero los cuadrados de esta serie si, es decir, existe un efecto ARCH en los log-retornos. Esto indica que no podemos predecir el valor de un rendimiento futuro utilizando la información histórica, sin embargo si existe dependencia entre la magnitud de los log-retornos anteriores, lo que se traduce a una volatilidad condicionalmente heterocedástica que es posible pronosticar.

De manera formal, es posible verificar los efectos ARCH a través de la prueba del multiplicador de Lagrange del **Apéndice B**, la cual proporciona un valor p prácticamente nulo para los log-retornos del tipo de cambio USD/MXN y reafirma la información proporcionada en la Figura 4.4. De estos hechos se puede concluir la presencia del efecto *volatility clustering* en los rendimientos.

4.2. Ajuste de modelos

Dadas las características resultantes, no es conveniente asignar únicamente una distribución $t - Student$ a los log-retornos del tipo de cambio USD/MXN para capturar todas las características que estos poseen. Debido a la dependencia entre los rendimientos y la heterocedasticidad de su varianza, es posible sugerir ajustar un modelo GARCH(p, q), con innovaciones $t - Student$, que capture estos hechos. Sin embargo, considerando que la media de un proceso GARCH(p, q) es cero, utilizaremos la siguiente representación.

$$X_t = \mu + a_t, t \in \mathbb{Z},$$

donde X_t : es el rendimiento en el tiempo t , μ es la media de los log-retornos históricos y $a_t, t \in \mathbb{Z}$ serán los residuos que se estudiarán a través un modelo GARCH(p, q).

Al igual que los log-retornos históricos, la serie de residuos tambien preserva el mismo valor de curtosis y el efecto ARCH en el comportamiento de sus funciones de autocorrelación. Se utilizará μ igual a la media muestral de los rendimientos, que en el presente caso es 0.02495.

Con la información proporcionada por los gráficos (b) y (c) de la Figura 4.4 y a la **Proposición 2.12**, se proponen los varios modelos GARCH(p, q) para estudiar los datos.

En el Cuadro 4.1 se presentan los modelos considerados, así como los correspondientes valores AIC y BIC. (ver Apéndice B)

Modelo		Criterios	
Distribucion ε_t	(p, q)	AIC	BIC
Normal	(1,1)	1.99525	2.00309
	(1,0)	2.21676	2.22264
	(2,1)	1.99625	2.00605
t student	(1,1)	1.94626	1.95606
	(1,0)	2.05630	2.06414
	(2,1)	1.94719	1.95895

Cuadro 4.1: Modelos GARCH(p, q) propuestos para los residuos a_t del tipo de cambio USD/MXN y los correspondientes valores AIC y BIC para la selección del proceso más adecuado a los datos.

Bajo el criterio de información de Akaike (AIC) y el de información bayesiano (BIC), elegimos el modelo con menor valor en estos rubros (ver Apéndice B). En este caso se elige un modelo GARCH(1, 1) con inovaciones $t - student$ para estudiar los residuos a_t

$$a_t = \varepsilon_t \sigma_t,$$

donde la varianza condicional.

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

Y $\varepsilon_t \sim SWN(0, 1)$ con distribución t estandarizada de ν grados de libertad, es decir.

$$\varepsilon_t \sim t(\nu) / \sqrt{\frac{\nu}{\nu - 2}},$$

Bajo el enfoque bayesiano, se generan 3 cadenas de Markov, cada una compuesta por 25000 iteraciones. Se utilizará la función `bayesGARCH` de la paquetería con el mismo nombre, presente en el software R, basada en el algoritmo de Metroplis-Hastings y que utiliza la teoría de estimación bayesiana en los parámetros de procesos GARCH(p, q), desarrollada por Ardia (2008) y descrita en el Capítulo 3.

En el Cuadro 4.2 se presentan los distintos valores iniciales que se utilizarán en la función,

Cadena	ω	α_1	β_1	ν
1	0.005	0.075	0.75	5.8
2	0.007	0.095	0.80	6.2
3	0.010	0.100	0.85	6.5

Cuadro 4.2: Valores iniciales propuestos para las 3 cadenas generadas.

Para la generación de las cadenas, se utiliza los hiperparámetros preestablecidos por la paquetería, de los cuales se recomienda hacer uso en ausencia de información *a priori* sobre los parámetros del proceso GARCH(1,1) con inovaciones $\varepsilon = \{\varepsilon_t : t \in T\}$ con distribución *t*-Student (ver Ardia, 2010).

Una vez simulados los valores de la cadenas, se analiza si sus respectivas distribuciones estacionarias coinciden con la distribución *a posteriori* que deseamos muestrear. Un análisis gráfico es mediante las trazas de los valores simulados para cada parámetro. En la Figura 4.5 se presentan las trazas para las tres cadenas simuladas.

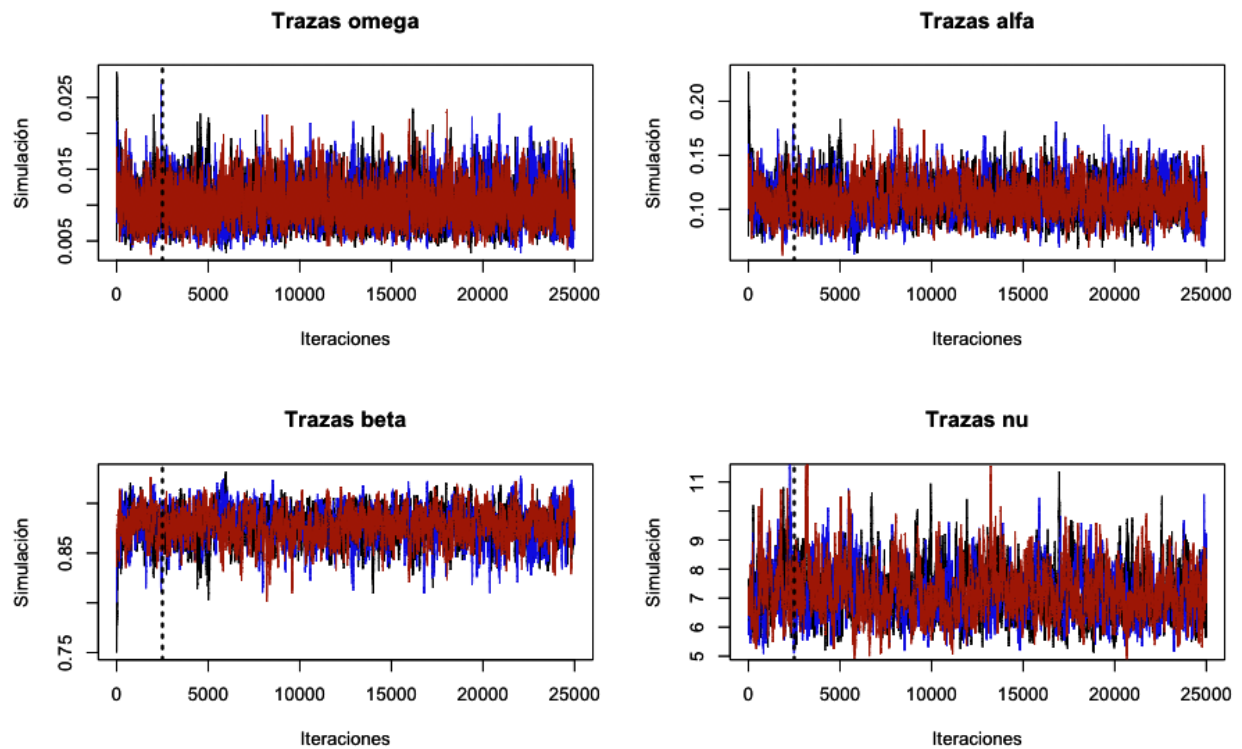


Figura 4.5: Trazas de los parámetros de la cadena 1 del modelo seleccionado. Cadena 1 en negro, cadena 2 en azul y cadena 3 en rojo.

Se observa que las simulaciones exploran correctamente los valores que los parámetros pueden asumir. De manera formal se analiza la convergencia de las cadenas simuladas a través de los criterios de convergencia presentados en el **Apéndice C**. Los resultados en el caso del criterio de Geweke se muestran el Cuadro 4.3.

Z-score de Geweke				
Cadena	ω	α_1	β_1	ν
1	0.8673	0.8863	-1.0670	0.8835
2	-1.0157	-0.7362	0.7311	0.6035
3	-0.1812	-0.2240	0.2160	0.5375

Cuadro 4.3: Estadístico Z del diagnóstico de convergencia de Geweke de las cadenas simuladas.

Del Cuadro 4.3 se observa que los valores absolutos de los estadísticos Z se encuentran por debajo de 2, por lo tanto se garantiza la convergencia de las medias en cada parámetro de las cadenas simuladas.

Por otro lado, el criterio de Gelman-Rubin nos ofrece un estadístico R de 1.00858, el cual es cercano a 1 y por lo tanto se garantiza la convergencia del método MCMC y se afirma que las simulaciones provienen de la distribución *a posteriori*.

Observaciones.

1. A partir del comportamiento de las trazas de las cadenas simuladas en la Figura 4.5 y el resultado de convergencia en el criterio de Geweke, se establece que el periodo de calentamiento (todas aquellas simulaciones que se descartan por no pertenecer a la distribución de la que se pretende muestrear) corresponde a las primeras 2500 iteraciones de cada cadena, correspondiente al 10 % de las simulaciones totales.
2. A pesar de obtener una convergencia a la distribución estacionaria, no es posible aplicar las muestras simuladas para poder aplicar Integración Monte Carlo en la estimación de parámetros. Por esta razón, con el objetivo de debilitar esta dependencia, se utilizará cada una de las 15 simulaciones de la distribución *a posteriori* generadas por el algoritmo de Metropolis-Hastings. Este hecho nos deja 4602 simulaciones finales para la estimación de cada uno de los parámetros.

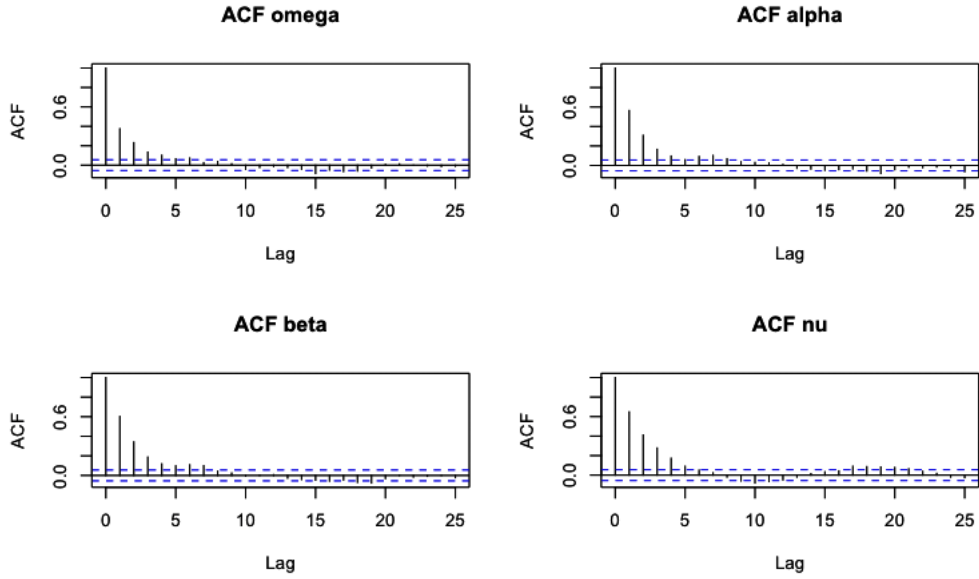


Figura 4.6: Funciones de autocorrelación muestrales para cada 15 valores de la tercera cadena.

De esta manera, para cada una de las cadenas se considerarán cada 15 valores simulados para calcular la media *a posteriori* de los parámetros. La media *a posteriori* de cada parámetro será calculada a través de la media muestral de los valores generados a partir de las tres cadenas.

En el Cuadro 4.4 se encuentran las medias a posteriori estimadas, así como la desviación estándar y los intervalos de credibilidad al 95 % de cada parámetro en el modelo seleccionado.

Parámetro	Media	Desviación	95 % credibilidad
ω	0.01007	0.00246	[0.01, 0.02]
α	0.10946	0.01529	[0.08, 0.14]
β	0.87792	0.01598	[0.84, 0.91]
ν	7.11867	0.82394	[5.73, 8.92]

Cuadro 4.4: Medias a posteriori de las cadenas simuladas generadas por Metropolis-Hastings, así como la desviación estándar de las simulaciones y el intervalo de credibilidad al 95 %.

Considerando los valores estimados, la expresión explícita del modelo bayesiano para el tipo de cambio es,

$$\hat{X}_t + 0,02495 = \hat{\sigma}_t \hat{\epsilon}_t,$$

con σ_t^2 dada por

$$\hat{\sigma}_t^2 = 0,01007 + 0,10946\hat{a}_{t-1}^2 + 0,87792\hat{\sigma}_{t-1}^2,$$

donde $\{\varepsilon_t : t \in \mathbb{N}\}$ es un proceso de ruido blanco estricto con distribución $t - Student$ estandariza y 7.11867 grados de libertad.

Observación. Se verifica que los parámetros estimados satisfacen las condiciones de estacionariedad (2.30) y existencia de cuartos momentos (2.32), esto debido a que.

$$0,10946 + 0,87792 = 0,98738 < 1$$

y además

$$1 - 2(0,10946^2) - (0,10946 + 0,87792)^2 = 0,001498989 > 0$$

De esta manera el modelo ajustado cumple con las condiciones del modelo.

4.3. Verificación del ajuste del modelo

Una vez establecidos los parámetros de los modelos, se debe verificar que los modelos ajustados cumplan con los supuestos de la **Definición 2.17**. Esta tarea se realiza a través del análisis de los residuales estandarizados, dados por,

$$\hat{\varepsilon}_t = \frac{\hat{a}_t}{\hat{\sigma}_t}. \quad t \in \mathbb{Z}. \quad (4.1)$$

Estos deben ajustarse a la condiciones de ruido blanco estricto del proceso $\varepsilon = \{\varepsilon_t : t \in \mathbb{Z}\}$, es decir, los valores (4.1) deben ser independientes con distribución t -Student estandarizada.

En primer lugar se verificará que los residuales constituyan una realización de variables aleatorias independientes. Para ello se puede analizar la ausencia de efectos ARCH mediante las pruebas presentadas en el **Apéndice B.1** a los residuales estandarizados en términos cuadráticos. Los resultados se presentan en el Cuadro 4.5

Valores- p	
Lagrange	0.278
Ljung-Box	0.996

Cuadro 4.5: Valores p para analizar la independencia de los residuales estandarizados.

Por el Cuadro 4.5 se puede verificar la independencia entre los residuos estandarizados, más aún, la prueba del multiplicador de Lagrange descarta la presencia de varianza condicionalmente heterocedástica.

Ahora se verifica que los residuales estandarizados pueden ser vistos como una realización de variables aleatorias independientes con distribución t -Student estandarizada con los grados de libertad estimados en la generación de las cadenas de Markov.

En la Figura 4.7 tenemos la comparación del histograma de los residuales con respecto a la distribución buscada.

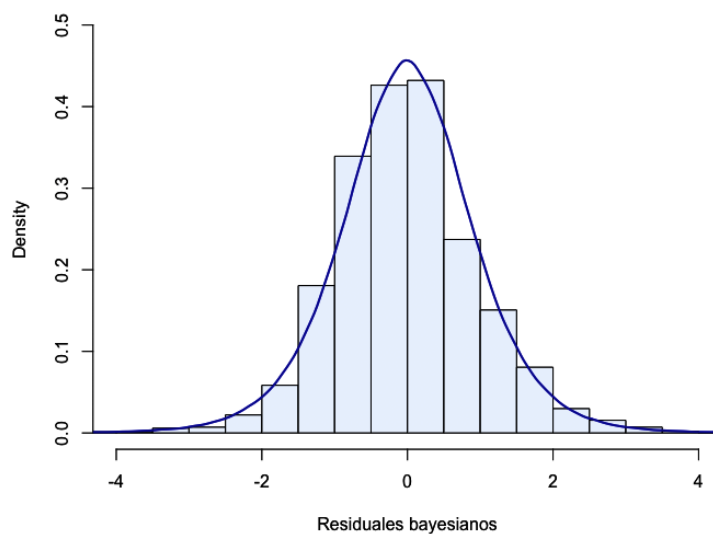


Figura 4.7: Histograma de los residuales estandarizados comparado con una distribución t

Intuitivamente se podría afirmar que los residuales siguen una distribución t - *Student* estandarizada con los grados de libertad estimados por el método bayesiano. Para verificar este hecho se realiza una prueba de Kolmogorov-Smirnov para dos muestras (ver Apéndice B). Los resultados se encuentran dados en el Cuadro 4.6.

Prueba Kolmogorov-Smirnov	
Estadístico	0.0317
p -value	0.1754

Cuadro 4.6: Resultados de la prueba de Kolmogorov-Smirnov aplicados a los residuales del modelo en contraste a una distribución t - *Student* estandarizada con los grados de libertad estimados.

A partir de los resultados del Cuadro 4.6 se observa que bajo un nivel de significancia del 5%, los residuales se ajustan a una distribución t - *Student* estandarizada, cuyos

grados de libertad son los estimados por el método bayesiano.

Una vez establecidos el, cuyos residuales cumplen hipótesis establecidas; es posible analizar la volatilidad de las observaciones históricas y realizar predicciones futuras de su comportamiento.

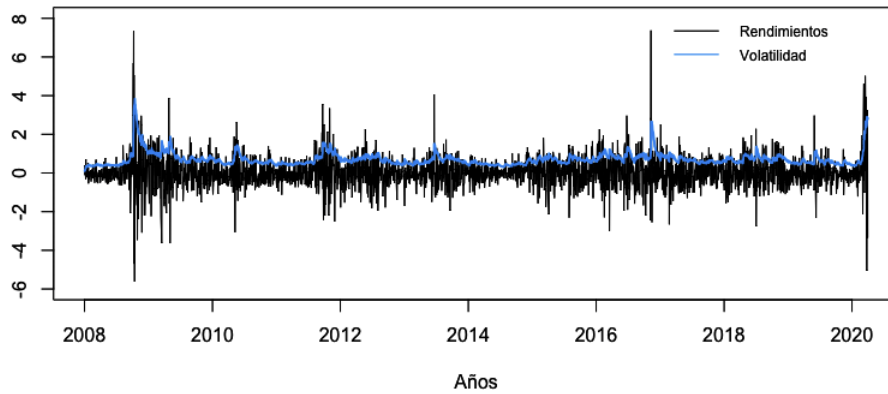


Figura 4.8: Volatilidad estimada por el modelo GARCH(1,1) con inferencia bayesiana, comparada con los rendimientos del tipo de cambio USDMXN

Por la Figura 4.8 se observa que la volatilidad generada por el modelo GARCH reacciona de acuerdo al efecto de *volatility clustering* de los rendimientos del activo.

Utilizando la función `predicciónsigma` (ver Apéndice D), desarrollada bajo la **Proposición 2.13**, se realizan los pronósticos del comportamiento de la volatilidad para el mes de Abril 2020. En la Figura 4.11 se observa el comportamiento de las predicciones bajo el modelo ajustado.

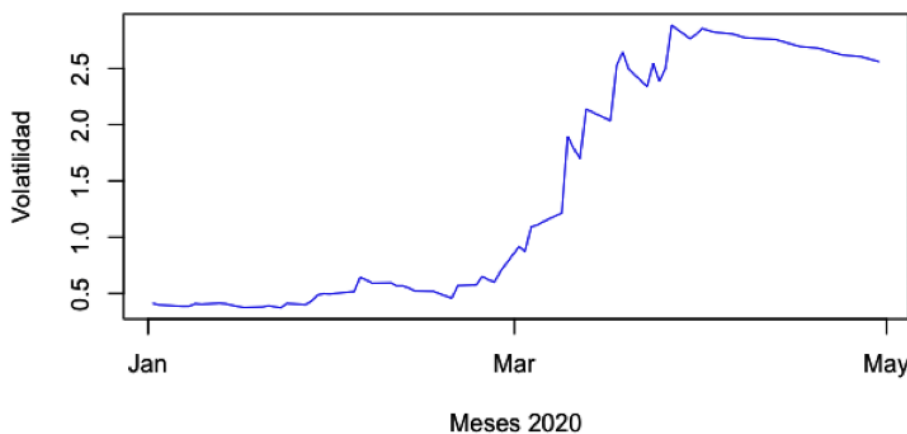


Figura 4.9: Predicciones de la volatilidad con respecto al comportamiento del inicio del año 2020

De la Figura 4.9 se observa que el modelo indica que hay una predicción en la que la dispersión de los rendimientos del mes de abril disminuirá. Esta se ajusta a la realidad de los rendimientos reales para el mes abril, tal como se muestra a continuación.

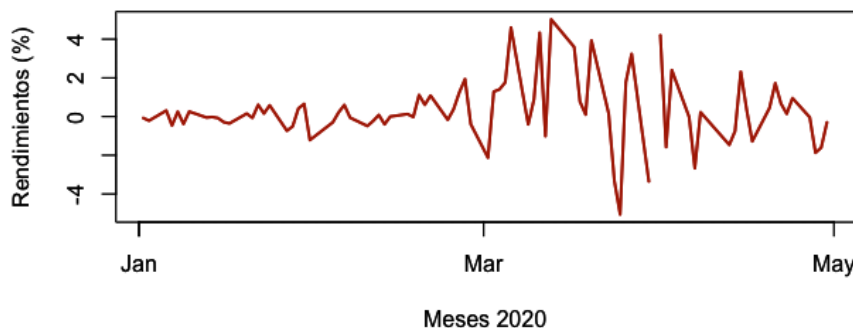


Figura 4.10: Rendimientos del USD/MXN 1 de Enero al 31 de Abril 2020

Por ende, el uso de la estadística bayesiana constituye una alternativa para la estimación de parámetros, que en procesos más complejos donde el número de parámetros aumente, la estimación sería más eficaz desde un punto de vista computacional cuando el número de restricciones en los parámetros complique la maximización de la función de verosimilitud. A continuación se presenta una aplicación de la volatilidad estimada por el modelo GARCH(1,1) con inovaciones t -Student.

4.4. Aplicación en riesgos financieros

En el análisis de riesgos financieros, el riesgo de mercado es aquel al que el valor de un activo disminuye debido a las fluctuaciones de acuerdo con las condiciones que están en el mercado. Este riesgo se busca controlar y gestionar para la toma de decisiones. Con el objetivo de cuantificar el riesgo de mercado dentro de activos financieros y portafolios de inversión se introduce el Value at Risk (VaR), el cual mide la pérdida potencial de valor dentro de un plazo determinado y un nivel de confianza.

Definición 4.1 Sean X una variable aleatoria que representa las **pérdidas y ganancias** de un activo financiero y $\alpha \in (0, 1)$ un nivel de significancia, el VaR a un día para X bajo un nivel de significancia α , está dado por:

$$VaR^\alpha = \inf\{x \in \mathbb{R} : \mathbb{P}[X \leq x] \leq \alpha\}. \quad (4.2)$$

Observación. McNeilet, *al.* (2005) y Ardia (2008) afirman que si utilizamos un proceso GARCH(1,1) para estudiar los residuos de los rendimientos de activos financieros con respecto a un valor medio μ . El VaR a un día para un instante t – *Student* a un nivel de significancia α está dado por:

$$VaR_t^\alpha = \mu + \sigma_t \sqrt{(\nu - 2)/\nu} t_\nu(\alpha) \quad (4.3)$$

donde $t_\nu(\alpha)$ es el cuantil a un nivel α de una distribución t estandarizada con ν grados de libertad y σ_t es la volatilidad ajustada por el modelo GARCH(1,1).

A continuación, en la Figura 4.11 se presenta la serie histórica de VaR diario bajo un nivel de significancia del 0.05, utilizando la volatilidad proporcionada por el modelo GARCH(1,1) con innovaciones t – *Student* y estimaciones bayesianas,.

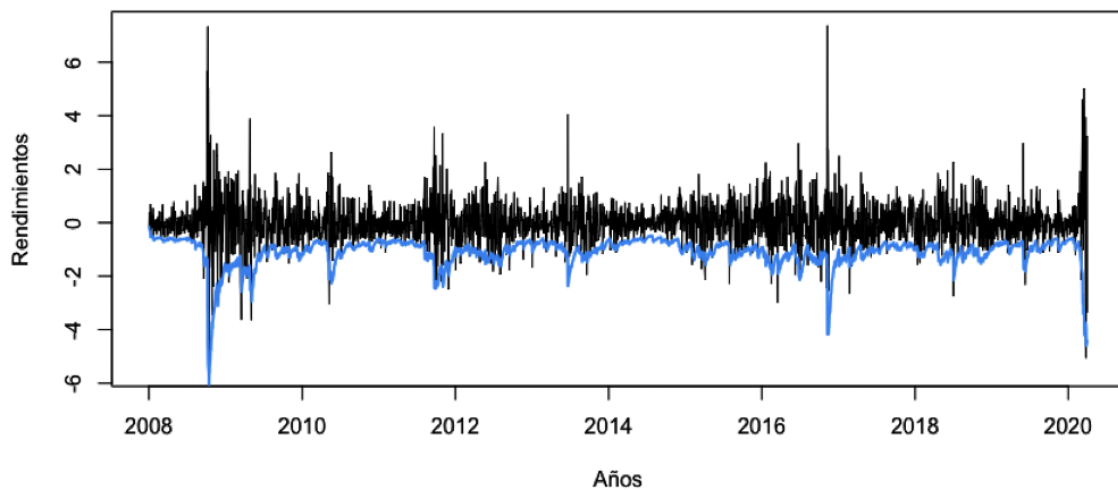


Figura 4.11: Datos históricos del VaR del tipo de cambio USD/MXN utilizando un proceso GARCH(1,1) con innovaciones t (en azul) en comparación a los log-retornos del tipo de cambio.

De igual forma, podemos utilizar los pronósticos de la volatilidad, descritos en la **Sección 4.5** para realizar pronósticos del comportamiento del VaR para el mes de abril del 2020. En la Figura 4.12 se muestran los log-retornos del 1 de enero 2020 hasta el 30 de abril del 2020, en contraste al VaR calculado con el modelo GARCH(1,1) hasta el 31 de marzo del 2020 y su pronóstico para el mes de abril.

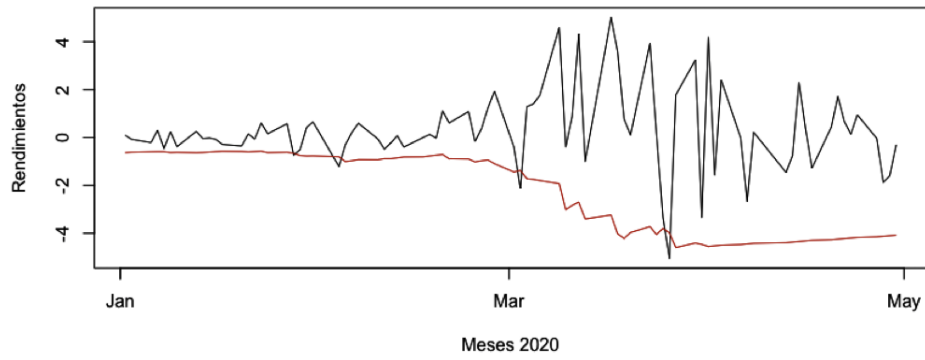


Figura 4.12: VaR calculado con el modelo GARCH(1,1) hasta el 31 de Marzo del 2020 y su pronóstico para el mes de Abril (en rojo), en comparación al comportamiento de los log-retornos del tipo de cambio.

Se observa que el cálculo del VaR a través de la volatilidad estimada por un proceso GARCH(1,1) responde ante los diversos escenarios de alta y baja volatilidad de los log-retornos.

Conclusión

Con el objetivo de estudiar el comportamiento de los rendimientos del tipo de cambio USD/MXN, una de las series con mayor impacto en el sistema financiero mexicano debido a su referencia como indicador de competitividad con respecto a otras naciones, se observó que las series de rendimientos de activos financieros cuentan con características que han motivado al desarrollo de modelos de series de tiempo que aborden estos fenómenos.

Para el caso del tipo de cambio USD/MXN, en la **Sección 4.2** se observó la inexistencia de correlación entre sus log-retornos y por ende la imposibilidad de realizar pronósticos de los beneficios y las pérdidas futuras. A pesar de ello, se observó dependencia en sus magnitudes cuadráticas, lo cual nos lleva a concluir que la volatilidad de sus rendimientos es condicionalmente heterocedástica y por ende es posible realizar pronósticos sobre su comportamiento. De esta forma, se utilizaron los procesos $\text{GARCH}(p, q)$ para el análisis de la volatilidad del tipo de cambio USD/MXN.

Debido a la correlación los rendimientos cuadráticos, así como la concentración de los datos a la media y los criterios de selección de modelos (Akaike y bayesiano), se decidió utilizar un proceso $\text{GARCH}(1,1)$ con innovaciones t -Student para analizar el comportamiento de su volatilidad. Se realizó la estimación de parámetros mediante inferencia bayesiana, donde se utilizó integración Monte Carlo, a través de valores simulados de la distribución *a posteriori* del vector de parámetros del modelo, esto bajo el algoritmo de Metropolis-Hastings.

De manera general, el proceso $\text{GARCH}(1,1)$ con innovaciones t -Student, al capturar el efecto de *volatility clustering*, nos proporciona una volatilidad condicionalmente heterocedástica, que aplicada a medidas de riesgo financiero como el VaR, nos proporcionan un

mayor reflejo del escenario en el que se encuentra el activo o portafolio en estudio, tal como se mostró en la **Sección 4.4** para el tipo de cambio USD/MXN. Se observó que uno de los beneficios que ofrece esta metodología para el cálculo del VaR es su pronóstico a corto plazo, esto derivado de la posibilidad de predecir la volatilidad del activo financiero en estudio a través de los procesos GARCH(p, q). Utilizando los datos hasta el mes de marzo de 2020 del tipo de cambio USD/MXN, se observó que el pronóstico para el mes de abril indicaría una caída en la volatilidad, lo cual se vio reflejado en los rendimientos reales del tipo de cambio, así como en el cálculo del VaR.

En virtud de la importancia de realizar un estudio de la volatilidad para el análisis de riesgos financieros, existen extensiones del modelo GARCH(p, q) que buscan abordar problemas específicos (como el análisis de valores extremos) y otros que fortalecen algunos hechos que los modelos GARCH(p, q) no logran capturar (como la asimetría de la volatilidad); en estos casos el número de parámetros aumenta, así como las restricciones que hay en ellos. Estos hechos dificultan la maximización de la función de verosimilitud y no es eficiente desde un punto de vista computacional. Al considerar los métodos Monte Carlo basados en cadenas de Markov, es posible implementar esta tarea de una manera más sencilla, aplicando las restricciones a los parámetros desde la generación de las cadenas y la definición de las distribuciones *a priori*, tal como se realizó en este trabajo para garantizar las condiciones de estacionariedad y existencia de cuartos momentos del proceso GARCH(1,1) con innovaciones t-Student.

Apéndice A

Definiciones y resultados preliminares

A continuación se presentarán algunas definiciones y resultados básicos de probabilidad y estadística, los cuales fueron utilizados durante el desarrollo de esta tesis.

A.1. Definiciones básicas

Definición A.1 *Se dice que una variable aleatoria X tiene una distribución t de Student con ν grados de libertad y se denota $X \sim t(\nu)$ si su función de densidad está dada por la siguiente expresión,*

$$f_X(x) = \frac{\Gamma(\nu + 1)/2}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

Definición A.2 *Sea $\boldsymbol{\theta}$ el vector de parámetros desconocido del modelo que describe al experimento que produjo una muestra aleatoria \mathbf{x} . El estimador máximo verosímil de $\boldsymbol{\theta}$, denotado por $\hat{\boldsymbol{\theta}}_{ML}$, es aquel que maximiza la función $f(\mathbf{x} | \boldsymbol{\theta})$, es decir.*

$$\hat{\boldsymbol{\theta}}_{ML} = \max_{\boldsymbol{\theta} \in \Theta} f(\mathbf{x} | \boldsymbol{\theta}),$$

donde Θ es el conjunto de valores donde el vector de parámetros $\boldsymbol{\theta}$ asume valores.

Observación. Para facilitar la maximización de la función de verosimilitud, se hace uso funciones monótonas, las cuales preservan el valor máximo de la función $f(\mathbf{x} | \boldsymbol{\theta})$. Por ejemplo, se puede usar $\ln f(\mathbf{x} | \boldsymbol{\theta})$, que es la función de **log-verosimilitud**.

Definición A.3 Sea X una variable aleatoria de media μ y varianza σ^2 . La curtosis de X es el cuarto momento de la variable estandarizada.

$$K_X = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right]. \quad (\text{A.1})$$

Hechos adicionales de la Sección 2.3.1

Leptocurtosis. Se refiere al comportamiento de la curtosis de los rendimientos de un activo financiero. La distribución de los rendimientos generalmente presenta una curtosis mayor a 3, correspondiente a la de una distribución normal.

Asimetría. Se ha observado que la volatilidad reacciona diferente a la orientación de los rendimientos, donde los retornos negativos generan un crecimiento en la volatilidad.

A.2. Resultados básicos

Teorema A.1 Sea $X = \{X_n : n \in \mathbb{N}\}$ una sucesión de variables aleatorias.

(a) Si $\sum_{n \in \mathbb{N}} \mathbb{E}|X_n| < \infty$, entonces $\sum_{i \in \mathbb{N}} X_i$ converge casi seguramente y además:

$$\mathbb{E} \left[\sum_{i \in \mathbb{N}} X_i \right] = \sum_{i \in \mathbb{N}} \mathbb{E}[X_i].$$

(b) Si para cada $n \in \mathbb{N}$, las variables X_n asumen valores no negativos, entonces la igualdad anterior también es válida.

Demostración. Ver Protter y Jacod (2004). ■

Teorema A.2 Sea X es una variable aleatoria $X \sim t(\nu)$, entonces su esperanza está dada por

$$\mathbb{E}[X] = 0$$

y su varianza:

$$\text{Var}[X] = \frac{\nu}{\nu - 2}, \nu > 2.$$

Demostración. Ver Ross (2009). ■

Teorema A.3 Si $X \sim N(0, 1)$ y $Y \sim \Gamma(n/2, n/2)$ son dos variables aleatorias independientes, entonces:

$$\frac{X}{\sqrt{Y}} \sim t(n).$$

Demostración. Ver Ross (2009). ■

Teorema A.4 La función de verosimilitud de un vector aleatorio \mathbf{X} de tamaño $n \in \mathbb{N}$ satisface.

$$f(\mathbf{x} | \theta) = f_{X_1, \dots, X_m}(x_1, \dots, x_m) \prod_{i=m+1}^n f_{X_i | X_1, \dots, X_{i-1}}(x_i | x_{i-1}, \dots, x_1, \theta), \quad (\text{A.2})$$

para $m \in \mathbb{N}$ tal que $m < n$.

Demostración. Partimos de la igualdad (1.1) de la **Definición 1.2**.

$$f(\mathbf{x} | \theta) = f_{X_1, \dots, X_n | \theta}(x_1, \dots, x_n | \theta),$$

Aplicamos la definición de probabilidad condicional a la función de densidad conjunta del vector aleatorio \mathbf{X} .

$$f(\mathbf{x} | \theta) = f_{X_1, \dots, X_{n-1}}(x_1, \dots, x_{n-1}) f_{X_n | X_1, \dots, X_{n-1}}(x_n | x_1, \dots, x_{n-1}, \theta),$$

Aplicando nuevamente la definición de probabilidad condicional al vector de variables aleatorias X_1, \dots, X_{n-1} resulta.

$$f(\mathbf{x} | \theta) = f_{X_1, \dots, X_{n-2}}(x_1, \dots, x_{n-2}) \prod_{i=1}^2 f_{X_{n-i+1} | X_1, \dots, X_{n-i}}(x_{n-i+1} | x_1, \dots, x_{n-i}, \theta)$$

Repetiendo este paso, para $m < n$ se obtiene que.

$$f(\mathbf{x} | \theta) = f_{X_1, \dots, X_m}(x_1, \dots, x_m) \prod_{i=m+1}^n f_{X_i | X_1, \dots, X_{i-1}}(x_i | x_{i-1}, \dots, x_1, \theta),$$
■

Apéndice B

Pruebas estadísticas en el análisis de series de tiempo

En este apéndice se describirán algunas pruebas que fueron utilizadas en el desarrollo del Capítulo 5, donde se aplicó un modelo que analizara la volatilidad del tipo de cambio USD/MXN. Primeramente mostramos algunas pruebas para analizar el comportamiento de la función de autocorrelación de una serie de datos. Posteriormente se presenta la prueba de bondad de ajuste de Kolmogorov-Smirnov para verificar que dos muestras de datos siguen una misma distribución. Además se describen los criterios que se utilizaron para la selección del modelo de volatilidad.

B.1. Pruebas para autocorrelación

Una de las características de las series de tiempo que se desea estudiar es el comportamiento de la función de autocorrelación. En algunos casos se puede observar que los datos presentan una débil correlación pero son dependientes debido a que la autocorrelación de los datos en términos cuadráticos presentan una correlación significativa.

A seguir se presentan algunas pruebas que auxilian en este análisis en las funciones de autocorrelación para verificar la presencia de estos hechos.

Ljung-Box

El trabajo desarrollado por Ljung y Box (1970) es una prueba de hipótesis que realiza el contraste para decidir si un grupo de autocorrelaciones en una serie de tiempo son

diferentes de cero. Por lo tanto la hipótesis nula de la prueba esta dada por:

H_0 : Los primeros h rezagos de la función de autocorrelación son cero.

La estadística de prueba es :

$$Q(h) = n(n+2) \sum_{i=1}^h \frac{\hat{\rho}_i^2}{n-i},$$

donde n es el tamaño de la muestra, $\hat{\rho}_k$ es la autocorrelación muestral en el retraso k y h es el número de retardos considerados. Tomando un nivel de significancia α , se rechaza la hipótesis nula si $Q(h)$ es mayor al cuantil $(1-\alpha)$ de una distribución ji-cuadrada de h grados de libertad. Esta prueba utiliza la función `Box.test(x)` de la paquetería `stats` del software R.

Prueba del multiplicador de Lagrange, (Engle, 1982)

Dadas una muestra X_1, \dots, X_n , se considera una regresión con base a los términos cuadrados de la muestra, es decir,

$$X_t^2 = \epsilon_t + \alpha_0 + \sum_{i=1}^h \alpha_i X_{t-i}^2, \quad t = h+1, \dots, n.$$

Se busca contrastar:

$$H_0 : \alpha_i = 0, i = 1, \dots, h \quad \text{y} \quad H_1 : \text{Existe } j \text{ tal que } \alpha_j \neq 0.$$

Se considera la siguiente estadística de prueba:

$$F = \frac{\left(\sum_{i=h+1}^n (X_i^2 - \hat{X})^2 - \sum_{j=h+1}^n \hat{\epsilon}_j^2 \right) / h}{\left(\sum_{j=h+1}^n \hat{\epsilon}_j^2 \right) / (n - 2h - 1)},$$

donde $\hat{\epsilon}_t^2 = \hat{X}_t - X_t$ y \hat{X} es la media muestral de $\{X_i : i = 1, \dots, n\}$.

Se verifica que F sigue una distribución ji-cuadrada de h grados de libertad y por ende se considera rechazar la hipótesis nula a un nivel α si F es mayor al percentil $(1-\alpha)100$ de una distribución ji-cuadrada de h grados de libertad. Para aplicar esta prueba se utiliza la función `ArchTest` de la paquetería `FinTS` del software R.

B.2. Prueba de Kolmogorov-Smirnov

Con el objetivo de verificar si una muestra de observaciones pueden ajustarse a una distribución, surgen las pruebas de bondad de ajuste. La prueba de Kolmogorov-Smirnov es una prueba de bondad de ajuste no paramétrica que tiene como objetivo determinar la bondad de ajuste de dos distribuciones empíricas, es decir, si dos muestras provienen de la misma distribución teórica. El método puede describirse como sigue.

Dadas dos muestras aleatorias $X_{1,1}, \dots, X_{n,1}$ y $X_{1,2}, \dots, X_{m,2}$ con funciones de distribución empíricas $F_1(\cdot)$ y $F_2(\cdot)$ respectivamente. Se desea ver si ambas muestras provienen de la misma distribución, es decir, se realiza el siguiente contraste de hipótesis:

$$H_0 : F_1(x) = F_2(x) \quad \text{y} \quad H_1 : F_1(x) \neq F_2(x).$$

Se considera la estadística de prueba $D_{n,m}$ definida como:

$$D_{n,m} = \sup_x |F_1(x) - F_2(x)|.$$

Se rechaza la hipótesis nula H_0 a un nivel de significancia α si:

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{nm}}.$$

Algunos de los valores más utilizados son:

α	0.001	0.005	0.01	0.05	0.1	0.2
$c(\alpha)$	1.949	1.731	1.628	1.358	1.224	1.073

B.3. Selección de modelos

En el análisis de un conjunto de datos pueden surgir varios modelos candidatos a ser utilizados. En el caso de las series de tiempo, las funciones de autocorrelación simple y parcial auxilian en esta tarea para los procesos ARMA(p, q) y GARCH(p, q). Sin embargo esta técnica es intuitiva y no puede considerarse un criterio formal para determinar cuando un modelo es mejor que otro.

Por lo tanto, a seguir se presentan algunos criterios planteados que auxilian en la tarea de seleccionar el modelo que mejor se ajusta a un conjunto de datos estudiados.

Criterio de información de Akaike

Akaike (1974) introduce el criterio de información de Akaike, AIC , el cuál es una medida de bondad de ajuste definida como:

$$AIC = 2k - 2 \ln(L),$$

donde k es el número de parámetros a estimar y L es el máximo de la función de verosimilitud, es decir, la función de verosimilitud evaluada en el estimador máximo verosimil. Por su estructura se busca el modelo con menor AIC .

Criterio de información Bayesiano

El criterio de información Bayesiano BIC , es una medida de bondad de ajuste implementada por Schwarz (1978) y definida como:

$$BIC = k \ln(n) - 2 \ln(L),$$

donde k y L tienen el mismo significado que en el AIC y n es el tamaño de la muestra. Su interpretación se sigue del AIC, se desea que el modelo elegido entre el conjunto de candidaos tenga BIC mínimo.

Apéndice C

Análisis de Convergencia

Teóricamente el algoritmo de Metropolis-Hastings garantiza que la cadena generada posee distribución estacionaria igual a la distribución de la cual buscamos generar valores. Sin embargo, en la práctica es necesario garantizar que los valores simulados realmente son de la distribución estacionaria, es decir, se debe verificar que los valores simulados son una muestra aleatoria de la distribución deseada. Para este análisis se verifican las siguientes características.

1. **Soporte de la distribución objetivo.** Al ejecutar un método Monte Carlo vía cadenas de Markov es importante ver que los valores que se generan en el algoritmo recorran correctamente el espacio de estados de la cadena, que equivale al soporte de la distribución objetivo. Para ello es recomendable hacer el uso de los siguientes gráficos.
 - a. **Traza.** Es un gráfico de línea que muestra los valores generados por el algoritmo a través del número de iteraciones. Se desea que este tenga una gran dispersión sobre todo el soporte de la distribución objetivo a lo largo del número de iteraciones.
 - b. **Histograma.** Realizar un histograma de los valores generados por el algoritmo y verificar que este sea consistente a la distribución objetivo.
2. **Diagnóstico de convergencia.** Se utilizan metodologías para verificar que la cadena de Markov generada por el algoritmo ha logrado converger a su distribución estacionaria. Se verifica la convergencia de las media de la muestra simulada a través del tiempo.

3. **Independencia.** La ley de los grandes números requiere que las variables aleatorias sean aproximadamente independientes, por ende se requiere verificar la dependencia de los datos mediante correlogramas, los cuales se introducirán en el siguiente capítulo.

Observación. En el caso del algoritmo de Metropolis-Hastings, una vez garantizada la convergencia sabremos que existirá un $n_0 \in \mathbb{N}$ tal que la colección de variables aleatorias $\{X_i : i > n_0\}$ tendrán una distribución con función de densidad $f(\cdot)$. Para que una muestra se asemeje a una cadena compuesta de valores independientes, aproximadamente se selecciona un valor k de tal manera que $\{X_{n_0+ik} : i \in \mathbb{N}\}$ sean variables independientes. La elección de k dependerá del correlograma de los datos.

A continuación se dará una descripción de dos de los diagnósticos de convergencia más conocidos. Para los objetivos de este trabajo, su implementación se desarrollará con ayuda de la paquetería `coda`, presente en el software R.

C.1. Criterio de Gelman-Rubin

Uno de los criterios más conocidos para diagnosticar la convergencia de un algoritmo Monte Carlo vía cadenas de Markov fue publicado en Gelman y Rubin (1992). Este diagnóstico consiste en generar $J > 1$ cadenas de Markov con N iteraciones y compararlas entre sí. Este criterio se realiza de la siguiente forma. Si denotamos por $\{x_i^{(j)} : i = 1, \dots, N\}$ a los valores generados por la cadena j , $j \in \{1, \dots, J\}$, el criterio de Gelman-Rubin consiste en analizar las varianzas entre y dentro de las cadenas. Denote a la media muestral de los valores de la cadena j como:

$$\bar{x}^{(j)} = \frac{1}{N} \sum_{i=1}^N x_i^{(j)}.$$

Por otra parte, la media de los valores $\{\bar{x}^{(j)} : j = 1, \dots, J\}$ es:

$$\bar{x}^{(*)} = \frac{1}{J} \sum_{j=1}^J \bar{x}^{(j)}.$$

Se calcula la varianza entre las medias muestrales de las cadenas,

$$B = \frac{N}{J-1} \sum_{j=1}^J (\bar{x}^{(j)} - \bar{x}^{(*)})^2.$$

Se calcula la siguiente estadística dentro de las J cadenas,

$$W = \frac{1}{J} \sum_{j=1}^J S_j^2,$$

donde,

$$S_j^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i^{(j)} - \bar{x}^{(j)})^2.$$

Gelman y Rubin (1992) utilizan B y W para estimar la varianza σ^2 de la cadena que se generaría con el algoritmo Monte Carlo vía cadenas de Markov planteado, es decir,

$$\hat{\sigma}^2 = \frac{N-1}{N} B + \frac{1}{N} W.$$

Si no existe una convergencia a la distribución objetivo, los valores generados por el algoritmo seguirán manteniendo cierta dependencia de los valores iniciales del algoritmo. Mientras esto ocurre, el valor real de σ^2 será sobreestimado por $\hat{\sigma}^2$ y subestimado por W . Gelman y Rubin (1992) analizan $\hat{\sigma}^2$ y W , deseando que estos estimadores se estabilicen en función del número de iteraciones. Para ello considere.

$$R = \frac{\hat{\sigma}^2 + B/(NJ)}{W}.$$

La metodología de Gelman-Rubin desea que R se aproxime a 1 para que la cadena mantenga a la distribución objetivo como estacionaria. En caso contrario, el algoritmo no está realizando un buen muestreo o necesitará un mayor número de iteraciones.

C.2. Criterio de Geweke

Geweke (1992) desarrolló un diagnóstico de convergencia para algoritmos Monte Carlo vía cadenas de Markov. Este busca analizar la convergencia de la media de los datos generados por el algoritmo.

Esta metodología consiste en dividir en dos conjuntos a las iteraciones de la cadena y así analizar la media y varianza de los datos correspondientes. Usualmente se consideran las primeras m iteraciones de la cadena, correspondientes al 10% de las N iteraciones totales, las cuales se buscan contrastar con respecto a las últimas k iteraciones de la cadena, que comúnmente corresponden al 50% de las N iteraciones.

Se busca establecer una prueba de hipótesis de diferencia de medias, donde la hipótesis nula es establecer la igualdad entre las medias de ambos conjuntos de iteraciones, si esta se rechaza, es un indicativo que la cadena todavía no ha alcanzado la convergencia.

El criterio de Geweke se describe a seguir. Si se considera $\bar{x}^{(m)}$ y $\bar{x}^{(k)}$ como las medias muestrales de las primeras m y las últimas k iteraciones, respectivamente. La estadística para la prueba de hipótesis es:

$$Z = \frac{\bar{x}^{(m)} - \bar{x}^{(k)}}{\sqrt{\frac{\hat{\sigma}_k^2}{k} + \frac{\hat{\sigma}_m^2}{m}}},$$

donde $\hat{\sigma}_m^2$ y $\hat{\sigma}_k^2$ corresponden a los estimadores de la varianza de los primeras m y últimas k iteraciones, respectivamente. Teóricamente estas se obtienen al analizar la densidad espectral de los datos generados (Geweke, 1992). La paquetería `coda` proporciona el valor del estadístico Z .

Observación. Bajo la condición $m + k < n$ se verifica que $Z \sim N(0, 1)$, por lo tanto bajo un nivel de confianza del 95 %, si $|Z| > 2$ entonces el algoritmo no converge a la distribución estacionaria.

Apéndice D

Código de la aplicación

```
##### Importamos la paqueteria necesaria.
library(fGarch, tseries, e1071, FinTS, bayesGARCH, readxl, coda, bayestestR
)

##### Importamos los datos el vector de los precios.
base_datos <- read_excel("Documents/USDMXN.xlsx")
preciosUSD <- base_datos$Precio[2018:5096]
fechas <- base_datos$Fecha[2018:5096]
B <- length(preciosUSD)

##### Se presenta la gráfica de la serie de tiempo de los precios.
plot(fechas, preciosUSD, type = "l", xlab = "Años", ylab = "Precios_USD/
MXN")

##### Se calculan los log-retornos.
rendimientosUSD <- log(preciosUSD[2:B]) - log(preciosUSD[1:(B-1)])
USD_r <- rendimientosUSD*100
USD <- USD_r - mean(USD_r)

##### Se presenta la gráfica de la serie de tiempo de los log-retornos.
plot(fechas[2: 3079], USD_r, type = "l", xlab = "Años", ylab = "
Rendimientos_(%)")

### Se observa el efecto de leptocurtosis
kurtosis(USD) + 3
hist(USD, freq = F, col = "Aliceblue", breaks = 70, xlim = c(-4,4), xlab
= "Rendimientos_(%)")
```

```

lines(density(rnorm(100000, mean = mean(USD), sd = sd(USD))), lwd = 2, col
      = "blue3")

#### Se analiza el comportamiento de la ACF y la PACF para verificar la
      presencia de efectos ARCH.
par(mfrow=c(1,3))
acf(USD, main = "ACF_residuos")
acf(USD^2, main = "ACF_residuos_cuadráticos")
pacf(USD^2, main = "PACF_residuos_cuadráticos")
ArchTest(USD)

##### Se proponen los siguientes modelos.
propuesta1USD <- garchFit(formula = ~garch(1,1), data = USD, cond.dist = "
      norm")
propuesta2USD <- garchFit(formula = ~garch(1,0), data = USD, cond.dist = "
      norm")
propuesta3USD <- garchFit(formula = ~garch(2,1), data = USD, cond.dist = "
      norm")
propuesta4USD <- garchFit(formula = ~garch(1,1), data = USD, cond.dist = "
      std")
propuesta5USD <- garchFit(formula = ~garch(1,0), data = USD, cond.dist = "
      std")
propuesta6USD <- garchFit(formula = ~garch(2,1), data = USD, cond.dist = "
      std")

##### Se determinan sus AIC y BIC
propuesta1USD@fit$ics
propuesta2USD@fit$ics
propuesta3USD@fit$ics
propuesta4USD@fit$ics
propuesta5USD@fit$ics
propuesta6USD@fit$ics

##### Se simulan 3 cadenas de Markov para la estimación bayesiana de pará
      metros del modelo GARCH
##### Al considerar datos estacionarios, el modelo debe cumplir
      estacionariedad.
addPriorConditions <- function(psi){psi[2] + psi[3] < 1 & 1 -2psi[2]^2 -(
      psi[2]+psi[3])^2 >0 }

```

```

##### Se generan las cadenas.
set.seed(21)
MCMC1 <- bayesGARCH(USD, control = list(n.chain = 1, l.chain = 25000,
                                       start.val = c(0.005, 0.075, 0.75,
                                                    5.8), addPriorConditions =
                                       addPriorConditions))

set.seed(22)
MCMC2 <- bayesGARCH(USD, control = list(n.chain = 1, l.chain = 25000,
                                       start.val = c(0.0075, 0.095, 0.8,
                                                    6.2), addPriorConditions =
                                       addPriorConditions))

set.seed(23)
MCMC3 <- bayesGARCH(USD, control = list(n.chain = 1, l.chain = 25000,
                                       start.val = c(0.01, 0.1, 0.85,
                                                    6.5), addPriorConditions =
                                       addPriorConditions))

##### Verificamos la convergencia de las cadenas mediante Geweke y Gelman
-Rubin.
geweke.diag(MCMC1)
geweke.diag(MCMC2)
geweke.diag(MCMC3)

MCMC <- c(MCMC1, MCMC2, MCMC3)
gelman.diag(MCMC)

##### Dibujamos las trazas de la cadena 1.
par(mfrow=c(2,2))
plot(as.vector(MCMC1$chain1[,1]), type = "l", xlab = "Iteraciones", ylab =
     "Simulación", main = "Traza_omega")
plot(as.vector(MCMC1$chain1[,2]), type = "l", xlab = "Iteraciones", ylab =
     "Simulación", main = "Traza_alfa")
plot(as.vector(MCMC1$chain1[,3]), type = "l", xlab = "Iteraciones", ylab =
     "Simulación", main = "Traza_beta")
plot(as.vector(MCMC1$chain1[,4]), type = "l", xlab = "Iteraciones", ylab =
     "Simulación", main = "Traza_nu")

```

```

##### Dibujamos las trazas de la cadena 2.
par(mfrow=c(2,2))
plot(as.vector(MCMC2$chain1[,1]), type = "l", xlab = "Iteraciones", ylab =
      "Simulación", main = "Traza_omega")
plot(as.vector(MCMC2$chain1[,2]), type = "l", xlab = "Iteraciones", ylab =
      "Simulación", main = "Traza_alfa")
plot(as.vector(MCMC2$chain1[,3]), type = "l", xlab = "Iteraciones", ylab =
      "Simulación", main = "Traza_beta")
plot(as.vector(MCMC2$chain1[,4]), type = "l", xlab = "Iteraciones", ylab =
      "Simulación", main = "Traza_nu")

##### Dibujamos las trazas de la cadena 3.
par(mfrow=c(2,2))
plot(as.vector(MCMC3$chain1[,1]), type = "l", xlab = "Iteraciones", ylab =
      "Simulación", main = "Traza_omega")
plot(as.vector(MCMC3$chain1[,2]), type = "l", xlab = "Iteraciones", ylab =
      "Simulación", main = "Traza_alfa")
plot(as.vector(MCMC3$chain1[,3]), type = "l", xlab = "Iteraciones", ylab =
      "Simulación", main = "Traza_beta")
plot(as.vector(MCMC3$chain1[,4]), type = "l", xlab = "Iteraciones", ylab =
      "Simulación", main = "Traza_nu")

##### Burnin-in
burning <- 2000
par(mfrow=c(2,2))
plot(cummean(as.vector(MCMC1$chain1[,1])), type = "l", xlab = "Iteraciones
      ", ylab = "Media", main = "Medias_omega")
lines(cummean(as.vector(MCMC2$chain1[,1])), type = "l", xlab = "
      Iteraciones", ylab = "Media", col = "blue")
lines(cummean(as.vector(MCMC3$chain1[,1])), type = "l", xlab = "
      Iteraciones", ylab = "Media", col = "red")
abline(v = burning, lty = 3, lwd = 2)
plot(cummean(as.vector(MCMC1$chain1[,2])), type = "l", xlab = "Iteraciones
      ", ylab = "Media", main = "Medias_alfa")
lines(cummean(as.vector(MCMC2$chain1[,2])), type = "l", xlab = "
      Iteraciones", ylab = "Media", col = "blue")
lines(cummean(as.vector(MCMC3$chain1[,2])), type = "l", xlab = "
      Iteraciones", ylab = "Media", col = "red")

```

```

abline(v = burning, lty = 3, lwd = 2)
plot(cummean(as.vector(MCMC1$chain1[,3])), type = "l", xlab = "Iteraciones
", ylab = "Media", main = "Medias_beta")
lines(cummean(as.vector(MCMC2$chain1[,3])), type = "l", xlab = "
Iteraciones", ylab = "Media", col = "blue")
lines(cummean(as.vector(MCMC3$chain1[,3])), type = "l", xlab = "
Iteraciones", ylab = "Media", col = "red")
abline(v = burning, lty = 3, lwd = 2)
plot(cummean(as.vector(MCMC1$chain1[,4])), type = "l", xlab = "Iteraciones
", ylab = "Media", main = "Medias_nu")
lines(cummean(as.vector(MCMC2$chain1[,4])), type = "l", xlab = "
Iteraciones", ylab = "Media", col = "blue")
lines(cummean(as.vector(MCMC3$chain1[,4])), type = "l", xlab = "
Iteraciones", ylab = "Media", col = "red")
abline(v = burning, lty = 3, lwd = 2)

##### Debilitamos la dependencia de los valores simulados.
##### Definimos una función que tome cada m valores de un vector.
m_vector <- function(x,m){
  N <- length(x)
  n <- length(x)/m
  y <- 1:(n+1)
  y[1] <- x[1]
  for (i in 2:n+1) {
    y[i] = x[(i-1)*m]
  }
  y[2] <- x[m]
  print(y)
}

##### ACF de cada m valores generados por la tercera cadena
par(mfrow=c(2,2))
acf(m_vector(as.vector(MCMC3$chain1[,1]), m = 15), main = "ACF_omega", lag
=25)
acf(m_vector(as.vector(MCMC3$chain1[,2]), m = 15), main = "ACF_alpha", lag
=25)
acf(m_vector(as.vector(MCMC3$chain1[,3]), m = 15), main = "ACF_beta", lag
=25)

```

```

acf(m_vector(as_vector(MCMC3$chain1[,4]), m = 15), main = "ACF_nu", lag
    =25)

### Medias a posteriori, desviaciones e intervalos de credibilidad.
### Muestreo de la distribución a posteriori bajo las 3 cadenas.
omega_sim <- c(m_vector(as_vector(MCMC1$chain1[,1])[burning:25000], m =
    15), m_vector(as_vector(MCMC2$chain1[,1])[burning:25000], m = 15), m_
vector(as_vector(MCMC3$chain1[,1])[burning:25000], m = 15))
alfa_sim <- c(m_vector(as_vector(MCMC1$chain1[,2])[burning:25000], m = 15)
    , m_vector(as_vector(MCMC2$chain1[,2])[burning:25000], m = 15), m_
vector(as_vector(MCMC3$chain1[,2])[burning:25000], m = 15))
beta_sim <- c(m_vector(as_vector(MCMC1$chain1[,3])[burning:25000], m = 15)
    , m_vector(as_vector(MCMC2$chain1[,3])[burning:25000], m = 15), m_
vector(as_vector(MCMC3$chain1[,3])[burning:25000], m = 15))
nu_sim <- c(m_vector(as_vector(MCMC1$chain1[,4])[burning:25000], m = 15),
    m_vector(as_vector(MCMC2$chain1[,4])[burning:25000], m = 15), m_vector(
as_vector(MCMC3$chain1[,4])[burning:25000], m = 15))

### Medias a posteriori
omega_hat <- mean(omega_sim)
alfa_hat <- mean(alfa_sim)
beta_hat <- mean(beta_sim)
nu_hat <- mean(nu_sim)
estimadores_bayesianos <- c(omega_hat, alfa_hat, beta_hat, nu_hat)

### Desviaciones
sd(omega_sim)
sd(alfa_sim)
sd(beta_sim)
sd(nu_sim)

### Intervalos de credibilidad.
ci(omega_sim, ci=0.95)
ci(alfa_sim, ci=0.95)
ci(beta_sim, ci=0.95)
ci(nu_sim, ci=0.95)

```

```

##### Desarrollamos una función que determine los valores sigma_t de la
      definición GARCH(p,q)
sigmat_MCMC <- function(data, omega, alfa, beta){

  ##### N es el tamaño de los datos aplicados a la función.
  N = length(data)
  ##### Calculamos sigma_t cuadrada, la varianza condicional de los modelos
      GARCH(1,1)
  varianza_cond_t <- 1:N
  for (i in 2:N) {
    varianza_cond_t[1] <- omega
    varianza_cond_t[i] <- omega + alfa*data[i-1]^2 + beta*varianza_cond_t[
      i-1]
  }
  ##### Sigma_t es la raiz positiva de la varianza condicional
  sigma_t <- sqrt(varianza_cond_t)
  print(sigma_t)
}

##### Por lo tanto los residuales estandarizados para la estimación
      bayesiana son:
res_bayes <- USD/sigmat_MCMC(data = USD,omega = omega_hat, alfa = alfa_hat
, beta = beta_hat)

##### Se verifica la independencia de los residuales
ArchTest(res_bayes)
Box.test(res_bayes^2)

##### Se analiza la distribución de los residuales
hist(res_bayes, breaks = 40, col = "aliceblue", freq = FALSE, xlab = "
  Residuales_bayesianos", xlim = c(-4,4), ylim = c(0,0.5))
lines(density(rstd(1000000, mean = 0, sd = 1, nu = estimadores_bayesianos
  [4])), col = "blue3", lwd = 2)

##### Pruebas de Kolmogorov-Smirnov para los residuales
ks2Test(res_bayes, rstd(2000, mean = 0, sd = 1, nu = estimadores_
  bayesianos[4]))

```

```

##### Se crea una función para los pronósticos de la volatilidad.
suma_geom <- function(a,b,h){
  vector_resultados <- 1:h
  for (i in 1:h) {
    vector_resultados[i] <- (a+b)^(i-1)
  }
  sum(vector_resultados)
}

##### Con ayuda de la función anterior , creamos una apartir de la
  proposición 1.5
prediccionsigma <- function(tamaño, rendimiento_inicial , volatilidadinicial
  , omega, alfa , be){
  h <- tamaño
  x <- 1:h
  for (i in 1:h) {
    x[i] <- omega*suma_geom(alfa ,be , i) + (alfa*rendimiento_inicial^2 + be
      *volatilidadinicial^2)*(alfa+ be)^(i-1)
  }
  print(sqrt(x))
}

##### Calculamos la volatilidad de los datos históricos.
vol_bayesiana <- sigmat_MCMC(data = USD,omega = omega_hat , alfa = alfa_hat
  , beta = beta_hat)

##### Se realizan los pronósticos
pred_bayesiana <- prediccionsigma(rendimiento_inicial = USD[3078] ,
  volatilidadinicial = vol_bayesiana[3078], omega = estimadores_
  bayesianos[1], alfa = estimadores_bayesianos[2], be = estimadores_
  bayesianos[3], tamaño = 20)

##### Cargamos los datos para abril.
preciosUSD_Ab <- base_datos$Precio[5096:5116]
fechas_Abr <- base_datos$Fecha[5097:5116]
D <- length(preciosUSD_Ab)
rendimientosUSD_A <- log(preciosUSD_Ab[2:D]) - log(preciosUSD_Ab[1:(D-1)])
USD_Abril <- rendimientosUSD_A*100

```



```

##### Se visualiza el comportamiento de la volatilidad de Abril con
      respecto al inicio del año.
plot(c(fechas[3018:3079], fechas_Abr), c(vol_bayesiana[3017:3078], pred_
      bayesiana),
      xlab = "Meses_2020", ylab = "Volatilidad", type = "l", col= "blue")

### Calculo del Valor en Riesgo.
#### Metodología con distribuciones t.
mu_t <- mean(USD_r)
sd_t <- vol_bayesiana
VaR_t <- mu_t + sd_t*qt(0.05, nu_hat)*sqrt((nu_hat-2)/nu_hat)

plot(fechas[2:3079], USD_r, xlab = "Años", ylab = "Rendimientos", type = "
      l")
lines(fechas[2:3079], VaR_t, col = "deepskyblue1", lwd = 2)

### Pronóstico del VaR.
mu_t <- mean(USD_r)
vol_t <- pred_bayesiana
VaR_pronos <- mu_t + vol_t*qt(0.05, nu_hat)*sqrt((nu_hat-2)/nu_hat)

### Comparación del VaR 2020.
plot(c(fechas[3018:3079], fechas_Abr), c(USD_r[3017:3078],USD_Abril),
      xlab = "Meses_2020", ylab = "Rendimientos", type = "l", col= "black")
lines(c(fechas[3018:3079], fechas_Abr), c(VaR_t[3017:3078],VaR_pronos),
      xlab = "Meses_2020", ylab = "Rendimientos", type = "l", col= "red")

```

Bibliografía

- [1] ARDIA, D. (2008). *Financial risk management with bayesian estimation of garch models: theory and applications*. Springer.
- [2] ARDIA, D., HOOPERHEIDE, L.F. (2010). Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations. *The R Journal*, 2(2), 41-47.
- [3] BILLINGSLEY, P. (1995) *Probability and Measure* John Wiley & Sons Inc., New York, third edition.
- [4] BOLLERSLEV, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal in Econometrics*.
- [5] BROCKWELL, P., DAVIS R. (2016), *Intoduction to time series and forecasting*. (Tercera edición), Springer.
- [6] ENGLE, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*.
- [7] FELLER, W. (1968). *An introduction to probability theory and its applications* Vol. I. 3rd.ed. John Wiley and Sons, New York.
- [8] FRANCO, C., ZAKOÏAN, J. (2010). *GARCH models: structure, statistical inference, and financial applications*. Wiley.
- [9] GELMAN, A., RUBIN D. (1992), *Inference from iterative simulation using multiple sequences*. *Statistical Science* 7: 457–472.
- [10] GEWEKE, J. (1993). *Bayesian treatment of the independent Student-t linear model*. *Journal of Applied Econometrics*.

- [11] GRIMMETT, G. STIRZAKER, D. (2001). *Probability and Random Processes*. Oxford University Press.
- [12] HOEL, P., PORT, S., STONE, C. (1972). *Introduction to Stochastic Processes*. Houghton Mifflin Co., Boston, Mass.
- [13] KOOP, G. (2003). *Bayesian Econometrics*. Wiley-Interscience, London, UK, first edition.
- [14] MCNEIL, A., FREY R. y EMBRECHS, P. (2005), *Quantitative Risk Management* Princeton University Press.
- [15] NAKATSUMA, T. (1998), A Markov-Chain Sampling Algorithm for GARCH Models *Studies in Nonlinear Dynamics and Econometrics*,
- [16] NAKATSUMA, T. (2000), Bayesian Analysis of ARMA-GARCH Models: A Markov Chain Sampling Approach. *Journal in Econometrics*.
- [17] PRADO, R., WEST, M. (2010), *Time series: modeling, computation, and inference*. CRC Press.
- [18] PROTTER, P., JACOD, J. (2004), *Probability Essentials* Second edition. Springer.
- [19] ROBERT, C. P., CASELLA, G. (2010), *Introducing Monte Carlo methods with R* Springer, New York.
- [20] ROBERT, C. P. (2007), *The bayesian choice: from a decision-theoric foundations to computational implementation* (2nd ed.) Springer Verlag.
- [21] RODRIGUES, E. R. (2013), *Introducción a los métodos de Monte Carlo*. Mini-curso EMALCA 2013. Escuela de Matemáticas de América Latina y del Caribe.
- [22] ROSS, S. (1996), *Stochastic Processes*. Second edition, Jhon Wiley & Sons Inc.
- [23] ROSS, S. (2009), *A first course in Probability*. Eight edition, Pearson.
- [24] SCHWARZ (1978), *Estimating the dimension of a model*. *Annals of Statistics* 6 (2): 461-464.
- [25] TSAY, R. S. (2010), *Analysis of financial time series*. Third edition, Wiley.