

*Estimación de la volatilidad de la tasa de cambio peso-dólar
a través de un modelo de volatilidad estocástica*

DANIEL PARRA AMADO

CÓDIGO: 830355



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE MATEMÁTICAS
BOGOTÁ, D.C.
ENERO DE 2014

*Estimación de la volatilidad de la tasa de cambio peso-dólar
a través de un modelo de volatilidad estocástica*

DANIEL PARRA AMADO

CÓDIGO: 830355

TRABAJO DE TESIS PARA OPTAR AL TÍTULO DE
MAESTRÍA EN MATEMÁTICA APLICADA

DIRECTOR
LILIANA BLANCO CASTAÑEDA

DR. RER. NAT

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE MATEMÁTICAS
BOGOTÁ, D.C.
ENERO DE 2014

Título en español

Estimación de la volatilidad de la tasa de cambio peso-dólar a través de un modelo de volatilidad estocástica

Title in English

Estimating the exchange rate peso-dollar volatility using a stochastic volatility model

Resumen: En el contexto económico, la volatilidad de la tasa de cambio peso colombiano-dólar afecta las decisiones de los individuos y empresas. Dicha volatilidad, genera efectos sobre las variables económicas tanto del sector financiero como del real y al ser no observable es necesario emplear técnicas econométricas que permitan realizar su estimación. Este trabajo pretende estimar la volatilidad de la tasa de cambio colombiana a través del modelo de volatilidad estocástica (SV) propuesto por Taylor (1982), para el periodo comprendido entre enero de 2000 y junio de 2013. Posteriormente, se busca contrastar estos resultados con los que arrojarían las estimaciones a través de modelos de heteroscedasticidad condicional como el GARCH (Bollerslev, 1986), los cuales son tradicionales en la literatura sobre series de tiempo. Los resultados obtenidos sugieren que la volatilidad de la tasa de cambio tiene alta persistencia y que las metodologías SV y GARCH logran modelar los excesos de curtosis y los conglomerados de volatilidad. Sin embargo, la evidencia estadística muestra que el supuesto de normalidad utilizado de las estimaciones es restrictivo, por lo que para un trabajo futuro se recomienda implementar funciones de distribución con colas más pesadas.

Abstract: In the economic context, the Colombian peso - US dollar exchange rate volatility affects both individuals and businesses decisions. This volatility impacts financial and real economic variables, and the fact it is an unobservable variable demands alternative econometric techniques for its estimation. This paper aims to estimate the Colombian exchange rate volatility by fitting a stochastic volatility model, like the one proposed originally by Taylor (1982), for the period comprised between January 2000 and June 2013. Subsequently, the results are contrasted with conditional heteroskedasticity models such as GARCH (Bollerslev, 1986), which are traditional in the time series literature. The results suggest that the exchange rate volatility is highly persistent and the methodologies SV and GARCH achieve to replicate kurtosis excess and volatility clusters. However, the statistical evidence shows that the normality assumption used in the estimates is restrictive, and therefore, implementation of heavy tailed distribution functions is recommended for future research projects.

Palabras clave: Modelos de volatilidad estocástica, Modelos de heteroscedasticidad condicional, Métodos de Monte Carlo, Tasa de cambio.

Keywords: Stochastic volatility models, Conditional heteroskedasticity models, Monte Carlo methods, Exchange rate.

Dedicado a

A mi madre, mi familia y mis seres queridos por siempre brindarme su afecto y apoyo sin importar las adversidades.

Algún día todo tendrá sentido.
Por ahora, ríete ante la confusión,
sonríe a través de las lágrimas
y síguete recordando que todo pasa por algo.
(P. COEHLLO)

Agradecimientos

Mi más profundo sentimiento de gratitud a mi directora en este trabajo, la Doctora Liliana Blanco, y a los profesores José Eduardo Gómez y Luis Fernando Melo, quienes me orientaron acertadamente cada vez que busqué su consejo. También, debo resaltar el apoyo recibido por parte del coordinador de la Maestría en Matemática Aplicada, el profesor Jorge Mauricio Ruiz. Igualmente deseo agradecer los comentarios de los jurados de este trabajo, los profesores Jaime Londoño y Mauricio Junca. Las observaciones y comentarios de cada uno de ellos se convirtieron en aportes valiosos para la culminación exitosa de este proyecto. A la Universidad Nacional de Colombia (Sede Bogotá), a los profesores, amigos, administrativos y colegas de la Facultad de Ciencias Económicas que contribuyeron de manera importante en mi desarrollo personal y profesional al obtener el grado de Economista, y luego, a la Facultad de Ciencias que me permitió acceder a una educación de altísima calidad para continuar mi proceso de formación. A mi familia, en especial a mi madre, por siempre apoyarme en cada momento y decisión que he tomado en mi vida.

Índice general

1. La tasa de cambio peso-dólar: Definición, motivación y hechos estilizados	6
1.1. La tasa de cambio nominal	7
1.1.1. Conceptos básicos	7
1.1.2. Relación de la tasa de cambio peso-dólar con otras variables de la economía	11
1.2. Hechos Estilizados de las Series Financieras: El caso de la tasa de cambio peso-dólar	14
1.2.1. Características de la tasa de cambio peso-dólar	15
1.2.2. No Normalidad de los Retornos	18
1.2.3. Autocorrelación	20
1.2.4. Hipótesis de Eficiencia de Mercado	21
2. Modelos de volatilidad: heteroscedasticidad condicional vs. volatilidad estocástica	26
2.1. Modelos de heteroscedasticidad condicional	27
2.1.1. Características y propiedades	27
2.1.2. Modelo ARCH(q)	30
2.1.3. Modelo GARCH(p,q)	33
2.2. Modelo de volatilidad estocástica	36
2.2.1. Características y propiedades	36
2.3. Pronóstico	41
3. Métodos de estimación	44
3.1. Una Breve Reseña	44
3.2. Método de Máxima Verosimilitud	46
3.3. Inferencia Bayesiana	48
3.4. Filtro de Kalman	50
3.5. Simulaciones de Monte Carlo (MCMC)	52
3.5.1. Algoritmo de muestreo Gibbs	52
3.5.2. Algoritmo de muestreo Metropolis y Metropolis-Hasting	55
3.5.3. Algoritmo FFBS (Forward Filtering Backward Sampling)	57
3.5.4. Algoritmo FFBS por aproximación de mixturas de distribuciones normales	59

3.6.	Métodos secuenciales de Montecarlo (SMC)	60
3.6.1.	Muestreo de Importancia	60
3.6.2.	Muestreo Secuencial de Importancia	62
3.6.3.	Filtro de Partículas	63
4.	Aplicación Empírica: Volatilidad diaria de la tasa de cambio peso-dólar	67
4.1.	Modelo de Heteroscedasticidad condicional: el caso del GARCH(1,1)	67
4.2.	Modelo de volatilidad estocástica (SV) de orden 1.	69
4.2.1.	Inferencia de la Posterior vía Cadenas de Markov de Monte Carlo	70
4.2.2.	Inferencia de la Posterior vía Monte Carlo Secuencial	82
4.3.	Diagnóstico de los modelos	85
4.4.	Aplicaciones en Economía	88
4.4.1.	Volatilidad del mercado: Comparación entre las diferentes esti- maciones.	89
4.4.2.	Medición del Valor en Riesgo (Value at Risk).	92
	Conclusiones	96
A.	Conceptos Básicos de probabilidad y matemáticas	98
A.1.	Introducción	98
A.2.	Conceptos básicos de matemáticas y probabilidad	99
A.2.1.	Variable aleatoria	99
A.2.2.	Funciones de distribución de probabilidad y de densidad de pro- babilidad	102
A.2.3.	Valor esperado, varianza, simetría y curtosis	103
A.2.4.	Procesos estocásticos	105
A.2.5.	Martingalas	107
A.2.6.	Movimiento browniano	108
A.2.7.	Integral estocástica y diferencial estocástica	110
B.	Funciones de distribución	114
B.1.	Funciones de distribución	114
B.1.1.	Distribución Normal	114
B.1.2.	Distribución log-normal	114
B.2.	Descripción probabilidades condicionadas del Modelo de volatilidad es- tocástica	115
C.	Conceptos Básicos de Series de Tiempo	119
C.1.	Series de tiempo	119
C.1.1.	Operador rezago y operador diferencia	119
C.1.2.	Procesos estacionarios y procesos ruido blanco	120
C.1.3.	Transformaciones aplicadas a series no estacionarias	121

D. Discretización	123
D.1. Discretización	123
D.1.1. Método de Euler	123
D.1.2. Método de Milstein	124
E. Salidas programas econométricos	125
E.1. Pruebas de raíz unitaria	125
E.2. Modelo de heteroscedasticidad condicional	128
E.3. Pruebas de Normalidad	130

Introducción

En el contexto económico, las decisiones de los inversionistas en los mercados financieros internacionales están rodeadas de incertidumbre, en cuanto más alta sea esta última, más riesgo enfrentarán los agentes. Este riesgo se asocia a la volatilidad, la cual se define como una medida de dispersión o variabilidad de la rentabilidad de un activo financiero (tales como los índices bursátiles o monedas de algún país). En particular, en este trabajo de tesis se estudia la volatilidad de la tasa de cambio de Colombia.

La variabilidad de la divisa colombiana genera preocupación sobre los efectos que puede tener sobre la actividad del sector real y financiero en una economía. Por ejemplo, sobre la actividad real en momentos en que ésta presenta gran variabilidad se pueden generar consecuencias negativas sobre los flujos de comercio internacional (Arize et al. [2000]) y las decisiones de inversión (Bailey and Tavlas [1991]). Respecto a los efectos sobre el sector financiero, la evidencia muestra que la volatilidad incentiva el desarrollo y la implementación de instrumentos de cobertura (Garber and Spencer [1996]). Asimismo, a nivel individual, un inversionista se puede beneficiar de fluctuaciones grandes de la tasa de cambio al generarse oportunidades de trading (Côte [1994]). Por estas razones dentro de la teoría económica y las finanzas ha cobrado gran importancia el estudio de la volatilidad en los mercados cambiarios y financieros.

Como la volatilidad de los activos financieros (tasas de cambio, acciones, etc¹) es una variable no observable, la economía y las finanzas se han apoyado en el desarrollo de herramientas estadísticas para estudiar su evolución y dinámica en los mercados financieros. La literatura econométrica ha propuesto dos tipos de modelos para el estudio de la volatilidad: los modelos de heteroscedasticidad condicional y los modelos de volatilidad estocástica. Las propiedades de estos modelos junto con algunos de los procedimientos utilizados para la estimación de los parámetros que los componen son estudiados en los capítulos posteriores.

Esta familia de modelos fue introducida inicialmente por Engle [1982], quien presentó un proceso estocástico para estimar volatilidad denominado modelo ARCH (*Autoregressive conditional heteroscedasticity*), en los cuales la varianza condicional depende

¹En general, no conocemos la distribución de las variables financieras dado que se tiene una sola realización en cada momento del tiempo.

del cuadrado de los rezagos de los términos aleatorios llamados innovaciones o errores (choques). De esta forma, la varianza condicional hace referencia a la volatilidad en un instante de tiempo t condicionada a la información relevante y conocida hasta ese momento. Luego, Bollerslev [1986] generalizó ese primer modelo y propuso el GARCH (*Generalized autoregressive conditional heteroscedasticity*). Allí, la varianza condicional además de depender de los cuadrados de las perturbaciones pasadas, ahora también está en función de las varianzas condicionales de períodos anteriores. Existe una gran cantidad de modelos derivados a partir de ellos, entre los más populares se encuentran: los modelos integrados IGARCH (Engle and Bollerslev [1986]) y los modelos asimétricos como EGARCH(Nelson [1991]), GJR-GARCH (Glosten et al. [1993]).

En este trabajo de tesis se compara la representación más usual y tradicional de esta familia, el GARCH, respecto al modelo de volatilidad estocástica. La estructura general del GARCH es:

Modelo $ARMA(\tilde{p}, \tilde{q}) - GARCH(p, q)$

1. Ecuación de la media: $y_t = \xi + \sum_{i=1}^{\tilde{p}} \phi_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{\tilde{q}} \theta_i \varepsilon_{t-i}$
2. Error estandarizado: $v_t = \frac{\varepsilon_t}{\sqrt{h_t}}$
3. Ecuación de la varianza: $h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \beta_i h_{t-i}$
4. $v_t \sim Normal(0, 1)$ e independiente de σ_t^2

donde $y_t = \ln(S_t) - \ln(S_{t-1})$. En el caso de este trabajo S_t es el nivel de la tasa de cambio (pesos por dólar) y y_t son los retornos de la tasa de cambio, variable sobre la cual se trabaja ya que cumple con la propiedad de estacionariedad requerida en los modelos de series de tiempo (Ver Apéndice C.3). Asimismo, ε_t y v_t son el residual o término de error de la ecuación de la media y el residual estandarizado, respectivamente. En el caso en que el mercado sea eficiente (Ver capítulo 1), la ecuación de media quedaría de la forma $y_t = \varepsilon_t$. Por último, los coeficientes $\xi, \phi_i, \theta_i, \alpha_i$ y β_i son los parámetros a estimar, y h_t es la estimación de la volatilidad ($h_t = \sigma_t^2$).

Los modelos de heteroscedasticidad condicional como los ARCH o GARCH asumen que la volatilidad puede ser observada un paso adelante con la información de las realizaciones de las variables aleatorias ε_t y h_t . Las noticias o choques se incorporan únicamente a través de las innovaciones ε_t , por lo que sólo existen perturbaciones sobre la ecuación de la media (la que describe el comportamiento del retorno del activo). De esta forma, la volatilidad se estima mediante el condicionamiento de las innovaciones al cuadrado pasadas y las varianzas condicionales pasadas; y esto se torna predecible, porque para cada momento t , ya se conoce la información en $t - 1$.

Partiendo de esto, un modelo más realista para la volatilidad se puede basar en una ecuación de comportamiento que tiene un componente predecible que depende de la

información pasada (como en los GARCH) y un componente de ruido inesperado. En este caso, la volatilidad es una variable *no observada latente*. Una interpretación de la volatilidad latente es que representa la llegada de nueva información en el mercado (Ver Clark [1973]). Este tipo de modelación corresponde a los denominados *modelos de volatilidad estocástica (SV)*.

Para garantizar la positividad de la varianza, la ecuación de la volatilidad se define para el logaritmo de la varianza ($h_t = \log(\sigma_t^2)$). En este trabajo de tesis se trabaja con la versión discreta del modelo de volatilidad estocástica en tiempo continuo. Taylor [1982] en su trabajo seminal encontró que el logaritmo de la volatilidad podía ser representado como un proceso autoregresivo de orden uno gaussiano lineal. En general, el proceso de volatilidad estocástica puede generalizarse como un proceso autoregresivo de orden p de la siguiente manera:

Modelo $ARMA(\tilde{p}, \tilde{q}) - SV(p)$

1. Ecuación de la media: $y_t = \xi + \sum_{i=1}^{\tilde{p}} \phi_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{\tilde{q}} \theta_i \varepsilon_{t-i}$
2. Error estandarizado: $v_t = \frac{\varepsilon_t}{\sqrt{h_t}}$
3. Ecuación de la varianza: $h_t = \gamma + \sum_{i=1}^p \phi_i h_{t-i} + \eta_t$
4. $v_t \sim Normal(0, 1)$ e independiente de h_t
5. $\eta_t \sim Normal(0, \sigma_\eta^2)$ e independiente de v_t

Los modelos SV encajan mejor con la teoría financiera ya que está relacionado con procesos de difusión utilizados para valoración de derivados (Danielsson [1994], Harvey et al. [1996], Kim et al. [1998]). Esto último deja entrever una de las ventajas del modelo de volatilidad estocástica: la ecuación en tiempo continuo se puede fundamentar en teorías financieras y económicas, mientras que los modelos tradicionales de series de tiempo no tienen fundamento económico. Sus propiedades dinámicas son fáciles de obtener e interpretar a partir de las del proceso estocástico subyacente, pero desafortunadamente, la estimación es más difícil que en los modelos ARCH o GARCH; al no poderse construir fácilmente la función de verosimilitud de forma exacta.

Para lograrlo, se utilizan métodos de estimación como pseudo-máxima verosimilitud, máxima verosimilitud simulada, filtros de Kalman, cadenas de markov por monte carlo (MCMC) o el método generalizado de los momentos, entre otros (Ruiz and Veiga [2008]). En general, los cuatro enfoques principales son: (i) de la cadena de Markov Monte Carlo bayesiana (MCMC) técnica sugerida por Jacquier et al. [1994], (ii) los métodos eficientes de momentos (EMM), propuesto por Gallant et al. [1997], (iii) la probabilidad de Monte Carlo (MCL), método desarrollado por Sandmann and Koopman [1998] y (iv) el muestreo de importancia eficiente (EIS) de Liesenfeld [2003]. En este

trabajo de tesis se desarrollan los primeros.

Carnero et al. [2004] encuentra que en el modelo GARCH, los parámetros que explican la persistencia y curtosis, están estrechamente vinculados. Sin embargo, Terasvirta and Zhao [2007] muestran que estas características pueden ser modeladas de forma independiente utilizando modelos tipo SV, con lo cual se puede representar mejor las regularidades empíricas observadas a menudo en series de tiempo financieras. Pulgarin [2001] realiza una aplicación para comparar el desempeño de los modelos GARCH y SV sobre algunos índices accionarios de la bolsa de Nueva York y encuentra evidencia que le permite concluir lo siguiente: el modelo SV describe mejor las formas empíricas de las series financieras reales frente a lo hecho por el modelo GARCH y encuentra que en términos de pronóstico, el modelo GARCH tiene un mejor desempeño que el modelo SV. Cabe mencionar que ambos enfoques reproducen algunas de las propiedades típicas de los rendimientos de las series financieras con periodicidad diaria, las cuales se discuten en el capítulo 1.

- * Exceso de kurtosis, por lo que su distribución tendrá colas pesadas.
- * Correlación en los cuadrados de la serie.
- * Clústers de volatilidad, o lo que es lo mismo, agrupamiento de períodos de volatilidad. Esto implica que períodos de alta volatilidad son seguidos por períodos de alta volatilidad, al igual que cuando la volatilidad es baja, se espera que los períodos siguientes tengan baja volatilidad.
- * Alta persistencia de la volatilidad debida a que las correlaciones de los cuadrados de la serie son significativas para órdenes elevados.

Con el desarrollo de nuevas plataformas de información, se han realizado avances en la obtención de datos con diferente periodicidad que permiten análisis de mayor profundidad, pero que al mismo tiempo requieren mayor complejidad en los métodos de estimación. En esta tesis se emplean datos diarios, pero debe mencionarse que en la última década también se han implementado técnicas sobre datos intradía o de alta frecuencia (hora, minutos, etc). Al respecto, el lector puede consultar Lildholdt [2002], Brandt and Jones [2006], Drost and Nijman [1993], Drost and Werker [1996], Meddahi and Renault [2004] y Meddahi and Renault [2006].

En los últimos años, la popularidad de modelos de estimación para la volatilidad ha cobrado fuerza, en especial, se encuentran estudios y aplicaciones de los modelos GARCH y SV sobre índices accionarios y monedas: Alberg et al. [2002] compara modelos GARCH simétricos y asimétricos para el índice de bolsa de Tel Aviv (TA-SE), Awartani and Corradi [2005] utiliza también modelos GARCH pero su aplicación se concentra en el índice de acciones de Estados Unidos (Standard and Poors), Kim

and Kim [2004] compara modelos de volatilidad estocástica en una aplicación sobre el índice accionario de Corea (KOSPI200), Ruiz and Veiga [2008] estima modelos de volatilidad estocástica para el IBEX (índice de la bolsa de España) y la Libra esterlina (divisa del Reino Unido), otros estudios recientes que el lector puede consultar son Micciché et al. [2002], Lam and Ng [2009] y Qu and Perron [2013]. Adicionalmente, se han utilizado las estimaciones de los modelos de volatilidad para aplicaciones en Economía y Finanzas, tales como valoración de riesgo y cálculo de opciones sobre activos financieros (Ver So and Yu [2006], Christoffersen et al. [2010]).

En el caso de Colombia algunas referencias sobre volatilidad son Casas and Cepeda [2008], Lega et al. [2007], Echavarría et al. [2009], Mesa [2010], Rincón and Toro [2010] y Echavarría et al. [2013]. Estos estudios se enfocan en analizar los determinantes y el impacto de la política cambiaria sobre la volatilidad de la tasa de cambio, y los efectos que tienen las noticias de mercado y las acciones de política sobre la volatilidad de la tasa de cambio. Sin embargo, son pocas las investigaciones dedicadas al estudio de la estimación de la volatilidad de la tasa de cambio a través de diversos métodos estadísticos, por lo que el presente proyecto de tesis aporta una aplicación a la literatura colombiana en ese sentido. Adicionalmente, esta tesis compara teórica y empíricamente algunos métodos de estimación para modelos de volatilidad estocástica y heteroscedasticidad condicional.

Para finalizar, este trabajo de tesis se desarrolla en cinco capítulos distribuidos de la siguiente manera. En el primero, se presentan al lector definiciones entorno a la tasa de cambio y se mencionan las motivaciones del trabajo, en especial, resaltando el papel que desempeña la volatilidad en las decisiones de inversión. En el segundo, se muestran las principales propiedades estadísticas de los modelos de heteroscedasticidad condicional y los de volatilidad estocástica, para luego en el tercero, plantear los diversos métodos existentes para la estimación de ambos tipos de modelos. Estos dos capítulos le permiten al lector aclarar las principales diferencias entre los dos modelos. En el cuarto, se encuentra el principal aporte de este trabajo de tesis, la implementación de ciertos algoritmos de estimación por simulación de Monte Carlo, con los cuales se ajustan los valores de los parámetros de los modelos y se estima la volatilidad de la tasa de cambio en Colombia (peso-dólar) para el período comprendido entre enero de 2002 y junio de 2013. Por último, se enuncian las principales conclusiones y resultados del documento.

Capítulo 1

La tasa de cambio peso-dólar: Definición, motivación y hechos estilizados

Las decisiones de los inversionistas siempre conllevan un alto grado de incertidumbre, lo cual es equivalente al riesgo que enfrentan los activos en los mercados, especialmente en los financieros. Este riesgo se asocia a la volatilidad, la cual se define como una medida de dispersión o variabilidad de la rentabilidad de un activo financiero (índice bursátil, precios de acciones o moneda). Con esta tesis se busca estimar la volatilidad de la tasa de cambio peso-dólar, a partir de un modelo de volatilidad estocástica para el período comprendido entre enero de 2002 hasta junio de 2013.

Para ello, es necesario presentar algunos conceptos económicos y algunas características estadísticas que tiene la tasa de cambio nominal, los cuales dan paso a la descripción teórica de los modelos econométricos y a su aplicación empírica en los capítulos posteriores. En ese sentido, este capítulo introductorio se divide en dos partes: i) definiciones económicas y motivación del trabajo y ii) características estadísticas de los datos diarios de la tasa de cambio peso-dólar entre enero de 2002 y junio de 2013.

En la primera parte, es necesario contextualizar al lector sobre la importancia y relevancia de la tasa de cambio nominal en una economía, así como establecer la relación entre ésta y el resto de variables económicas. De esta manera, comprender como su volatilidad puede afectar las decisiones de gastos de hogares y empresas. La investigación de esta tesis, la estimación de la volatilidad de la tasa de cambio peso-dólar, se convierte en un problema relevante en la práctica y por tanto un objetivo de este trabajo es contribuir a la literatura colombiana al explorar su medición y diversos métodos de estimación (Ver Capítulo 3). En la segunda parte, se muestran algunos hechos estilizados que tienen la mayoría de series de tiempo financieras, tomando como ejemplo la tasa de cambio peso-dólar. Los rasgos característicos que se muestran, ampliamente discutidos en el ámbito académico, son importantes en la medida que los modelos

econométricos que se implementan en este trabajo de tesis deben replicarlos.

1.1. La tasa de cambio nominal

1.1.1. Conceptos básicos

La tasa de cambio muestra la relación entre dos monedas de dos economías particulares, y para efectos del presente ejercicio, ésta se define como la cantidad de pesos colombianos que deben pagarse por un dólar estadounidense. En este trabajo se toma como base el dólar, porque es la divisa más utilizada en Colombia para las transacciones con el exterior. Al igual que el precio de cualquier producto, la tasa de cambio sube o baja dependiendo de la oferta y la demanda, así cuando la oferta es mayor que la demanda (hay abundancia de dólares en el mercado y pocos compradores) la tasa de cambio baja y se dice que el peso se aprecia. Por el contrario, cuando la oferta es menor que la demanda (escasez de dólares y muchos compradores), la tasa de cambio sube y se dice que el peso se deprecia¹.

La utilización de la tasa de cambio en una economía permite: i) ligar los precios internos de una economía y los de los otros países con los que comercia, ii) dar información que guíe a los diferentes agentes económicos en sus decisiones de consumo, inversión o producción y iii) es una herramienta de política, ya que a través de ella se puede incentivar las exportaciones y desestimular las importaciones con el fin de mantener un adecuado balance en las operaciones del sector externo.

Los dos tipos de regímenes cambiarios más populares que definen el comportamiento e importancia de la tasa de cambio nominal en una economía son²:

* Tasa de cambio fija:

Este sistema tiene como objetivo mantener constante la relación entre dos monedas, es decir, que la cantidad de pesos que se necesiten para comprar un dólar (u otra moneda extranjera) sea la misma a lo largo del tiempo. En este caso, el banco central, que en el caso de Colombia es el Banco de la República, se compromete a mantener esta relación y tomar las acciones necesarias para cumplir con este objetivo. Por lo tanto, cuando en el mercado existe mucha demanda por dólares o cualquier otra divisa (moneda extranjera), el Banco pone en el mercado la cantidad de dólares necesaria para mantener la tasa de cambio en el valor que se determinó. Igualmente, cuando se presentan excesos de oferta (cuando hay más dólares en el mercado de los

¹También es usual escuchar en los medios de comunicación, devaluación o revaluación.

²Tomado de Banrep [1998]

que se están demandando), el Banco compra la divisa externa para evitar que la tasa de cambio disminuya.

* Tasa de cambio flotante o flexible:

Este régimen permite que el mercado, por medio de la oferta y la demanda de divisas (monedas extranjeras), sea el que determine el comportamiento de la relación entre las monedas. El banco central no interviene para controlar el precio, por lo cual la cantidad de pesos que se necesitan para comprar una unidad de moneda extranjera (dólar, por ejemplo) puede variar a lo largo del tiempo.

Colombia a través de la historia ha contado con diversos regímenes cambiarios: tasa de cambio fija, devaluaciones gota a gota, bandas cambiarias, tasa de cambio flotante o flexible, entre otros (Ver Sánchez et al. [2005]). En la actualidad, la estrategia de política monetaria adoptada por el Banco de la República corresponde a un régimen de flexibilidad cambiaria, sujeto a unas reglas de intervención que buscan los siguientes objetivos:

1. Mantener un nivel adecuado de reservas internacionales que reduzcan la vulnerabilidad de la economía frente a choques externos, tanto de cuenta corriente como de capital.
2. Limitar la volatilidad excesiva de la tasa de cambio en horizontes cortos.
3. Moderar apreciaciones o depreciaciones excesivas que pongan en peligro el logro de las metas de inflación futuras, y la estabilidad externa y financiera de la economía.

En ese sentido, es relevante analizar la importancia del control de la volatilidad de las monedas para una economía. Una de las principales razones, es que la tasa de cambio es un precio más en la economía, cuyas fluctuaciones influyen sobre las decisiones de consumo, inversión y producción por parte de los empresarios y las familias. Todo ello tendrá efectos a nivel macroeconómico en términos de la inflación, el empleo y el crecimiento económico, variables que determinan algunas de las medidas de bienestar de la sociedad.

En el caso colombiano, el Banco de la República cuenta con diversos instrumentos para intervenir en el mercado cambiario, la mayor parte de las veces lo hace a través de subastas (discrecionales o competitivas), en las cuales hay opciones de compra o venta de divisas. En particular, tienen una regla que establece que cada vez que la tasa de cambio se desvíe en más de 5 % del promedio móvil de los últimos 20 días hábiles³, el

³El promedio móvil de los últimos 20 días días, es un promedio simple de 20 datos. La palabra móvil indica que este promedio se va actualizando diariamente al incluir nuevas observaciones disponibles, por ejemplo, cuando conocemos el dato de mañana de la tasa de cambio. Más detalles en http://www.banrep.gov.co/reglamentacion/rg_cambia2.htm

Banco debería intervenir. De esta manera, entre los objetivos de política económica, el Banco de la República pone de manifiesto su preocupación por controlar la volatilidad de la tasa de cambio puesto que su dinámica afecta las decisiones de inversión y gasto de los agentes económicos, en especial, en momentos de gran volatilidad que generan un incremento en la percepción de riesgo.

Una situación que ocurre con alta frecuencia en la realidad colombiana es la problemática que enfrentan los exportadores frente a las fuertes fluctuaciones de la moneda. Es muy común escuchar en los medios de comunicación los efectos que la volatilidad de la tasa de cambio tiene sobre la economía. Estos dos fragmentos de diarios reconocidos, son un ejemplo de ello

Diario Portafolio (Con volatilidad del dólar todos pierden-15 de julio de 2008)

" Los exportadores que vieron con alegría la subida que le permitió al dólar alcanzar 1.923 pesos a finales de junio, hoy están de nuevo tristes porque la divisa va en 1.752,38 (cotización promedio de ayer), sobre todo si tenían dinero para traer al país como pago de sus ventas al exterior y no aprovecharon para hacerlo en ese momento. La volatilidad es supremamente problemática porque no se sabe dónde se va a estabilizar el dólar. Así es muy difícil hacer cotizaciones, pues como no sé cuál será mi tasa de cambio, tampoco puedo manejar mi flujo de caja, dice Javier Díaz, presidente de Analdex...."

Diario La República (Una tasa de cambio volátil daña más que la revaluación -29 de enero de 2011)

Así lo indicó el ex ministro de Hacienda, José Antonio Ocampo: "La volatilidad es tanto o más grave que la revaluación....El riesgo es alto para las exportadoras que compiten contra las importaciones. El inversionista hace su plan de trabajo a la tasa de cambio más baja previsible. Pero si está cambiando a cada rato la incertidumbre es elevada".

Como se aprecia, la volatilidad de la tasa cambio es una variable importante en el ámbito nacional, por lo que estudiar su comportamiento manifiesta la relevancia del presente trabajo de tesis. Sin embargo, ésta presenta una particularidad que complica el análisis sobre su dinámica: *es una variable no observable*. En otras palabras, los inversionistas conocen los precios del activo (número de pesos por dólar) en cada momento del tiempo y pueden calcular la variación en los precios con el fin de calcular las ganancias o pérdidas (lo que se denomina retorno financiero). No obstante, no existe una medición que ocurra simultáneamente para ver la variabilidad de los precios. Una aproximación sencilla para su estimación es a través del cálculo de la desviación estándar de los retornos. Sin embargo, esto desconoce hechos importantes como la existencia de una estructura en la que se relaciona la volatilidad a lo largo del tiempo.

En general, la volatilidad puede ser entendida como una medida de la intensidad en que se presentan los cambios aleatorios en los retornos de un activo. Inicialmente, algunas teorías financieras tomaban ésta como un parámetro de la función de distribución de los retornos y asumían que era constante en el tiempo. Esto, como se muestra más adelante, es un error ya que no tiene en cuenta las características propias de las series financieras. Entre ellas se destacan: i) la existencia de *clusters* o aglomerados de volatilidad y ii) las correlaciones que generan una estructura dinámica en su comportamiento. En las últimas décadas, se ha abordado el estudio de la medición de la volatilidad a través de herramientas econométricas que describan mejor la realidad. En esta tesis se muestran algunos de estos avances y se realizan unas aplicaciones para estimar la volatilidad de la tasa de cambio peso-dólar.

Todo ello resulta útil en la medida que facilita el análisis de riesgo en las inversiones. De hecho, las estimaciones que se encuentran en este trabajo pueden ser utilizadas como insumo en metodologías que valoren el riesgo de mercado de acciones o portafolios que emplean diferentes instituciones financieras.

Algunos ejemplos de volatilidad a los que se enfrenta un agente económico son:

1. Volatilidad de una acción: Parámetro que mide el riesgo total de una acción, esto es, el riesgo diversificable y el riesgo de mercado. Proporciona una medida de la dispersión o variabilidad de la rentabilidad de dicha acción.
2. Volatilidad de una opción: Hace referencia a la volatilidad del activo subyacente al que se refiere la opción.
3. Volatilidad implícita: Es la volatilidad que, de acuerdo con las características de la opción, se ha de introducir en la fórmula de valoración para obtener el precio al que cotiza la opción.

Para cubrirse del riesgo, existen instrumentos financieros llamados derivados o contratos a futuro de monedas, con los cuales los inversionistas buscan suavizar las pérdidas de un portafolio o de una negociación frente a las fluctuaciones naturales de las monedas que ocurren a nivel mundial. Para implementar este tipo de herramientas es necesario tener una medición del riesgo, el cual en la teoría económica se asocia con la volatilidad. Sin embargo, como se mencionó, la volatilidad es una variable no observable por lo que se hace indispensable recurrir a técnicas estadísticas para su estimación.

Teniendo en consideración todo lo anterior, la motivación de este trabajo de tesis es estimar la volatilidad de la tasa de cambio peso-dólar mediante la exploración de herramientas econométricas. Como se muestra en los siguientes capítulos dicha medición se puede realizar a través de modelos de heteroscedasticidad condicional y de volatilidad estocástica.

1.1.2. Relación de la tasa de cambio peso-dólar con otras variables de la economía

En una economía pequeña y abierta como la colombiana, la tasa de cambio nominal es un activo financiero que juega un rol importante al medir en cierta forma el valor del dinero de otro país. En un entorno monetario como el actual, en el cual la tasa de interés es el instrumento principal y la tasa de cambio es flexible o flotante⁴, ésta última adquiere gran relevancia dentro del análisis económico ya que su comportamiento y volatilidad puede afectar las decisiones de consumo de bienes internos e importados, puede generar consecuencias negativas sobre los flujos de comercio internacional (Ver Arize et al. [2000]), influir en las decisiones de inversión (ver Bailey and Tavlas [1991]) y por último puede afectar el nivel de precios internos. En el sector financiero, la evidencia muestra que la volatilidad de la tasa de cambio incentiva el desarrollo y la implementación de instrumentos de cobertura (Garber and Spencer [1996]). Asimismo, a nivel individual, un inversionista se puede beneficiar de fluctuaciones grandes de la tasa de cambio al generarse oportunidades de trading (Côte [1994]). Es así como dentro de la teoría económica y las finanzas ha cobrado gran importancia el estudio de la volatilidad en los mercados cambiarios y financieros.

Para ilustrar los impactos de la tasa de cambio sobre la economía a continuación se describe brevemente su relación con la política monetaria, tratando de mostrar porque es relevante para el Banco de la República⁵ el seguimiento y monitoreo del mercado cambiario, toda vez que la moneda es una variable que afecta, y al mismo tiempo se ve impactada, por las decisiones del Banco Central. Adicionalmente, debe tenerse en cuenta que el efecto de los cambios en la postura de la autoridad monetaria sobre la divisa dependerán del grado de apertura de la economía y del grado de movilidad internacional de capitales. Si se supone que la economía está integrada financieramente al mundo y hay perfecta movilidad de capitales, entonces la tasa de cambio debe garantizar, en términos de rentabilidad en una misma moneda, la igualdad entre activos internos y externos⁶. A esta condición de no arbitraje se le denomina *paridad no cubierta de la tasa de interés*. Incorporando esta relación dentro del modelo tradicional IS-LM, Fleming [1962] y Mundell [1963] encontraron que un aumento de la tasa de interés interna: aumenta la rentabilidad de activos internos relativa a los denominados en moneda extranjera, genera una apreciación de la moneda local, lo cual incrementa el valor de los bienes nacionales y conlleva a una caída en las exportaciones netas y de la demanda agregada.

En términos de precios internos, la apreciación de la moneda conduce a una disminución de la inflación de la canasta de bienes transables. Este hecho es conocido en la

⁴Es decir, la divisa fluctúa libremente y no tiene un valor determinado, esto implica que la política monetaria es independiente.

⁵En general, para cualquier Banco Central.

⁶Activos internos y externos que posean las mismas características.

literatura como *pass-through* de la tasa de cambio⁷ y surge de teorías como la ley de un sólo precio (Ver Isard [1977])⁸ y la paridad del poder de compra (*PPP*)⁹. A la luz de estas teorías y bajo un régimen de tasa de cambio flotante, las variaciones de esta última se trasladan completamente a los precios internos ya que la inflación interna es la tasa de depreciación más la inflación internacional. Sin embargo, en la realidad no todos los bienes son transables internacionalmente; así que suponer que la transmisión de la tasa de cambio a los precios internos es uno a uno, no es del todo correcta¹⁰. Empíricamente los resultados muestran un coeficiente de traspaso menor a uno, por lo que dentro del ámbito académico surgen una serie de modelos en los cuales se diferencian bienes internos transables y no transables. También se establece la existencia de diferentes estructuras de mercado no competitivas y la sustitución imperfecta entre bienes nacionales e importados, todo lo cual podría causar traspasos incompletos (Ver Dornbusch [1987]).

Taylor [2000] estudia la relación entre *pass-through* del tipo de cambio y la política monetaria para finalmente ver los resultados en términos de los precios internos de un país. Este estudio encuentra que economías que gozan de una inflación baja tienen un *pass-through* más bajo ya que esto contribuye al descenso en la persistencia esperada de los costos y en los cambios en los precios, y al mismo tiempo reduce la transmisión de choques externos de precios. Cuando el coeficiente de *pass-through* es bajo el tipo de cambio no tiene efectos significativos sobre la inflación; por lo tanto, el Banco Central tiene mayor libertad para realizar una política monetaria más autónoma. Por el contrario cuando el *pass-through* es alto, los choques externos provenientes del mercado cambiario se transmiten significativamente a los precios internos, de tal forma que se hace necesario adoptar políticas cambiarias coordinadas para ayudar a solventar el impacto ésta tiene sobre la economía.

Por último, existen otros canales a través de los cuales la tasa de cambio puede afectar la economía como es el caso de las hojas de balance de las empresas. Este establece que movimientos abruptos de la divisa pueden afectar la posición deudora neta de los agentes privados. Por ejemplo, en una economía donde los residentes son deudores netos, una apreciación fuerte del tipo de cambio mejora el balance de aquellas em-

⁷Pass-through en economía es el traspaso que las variaciones en la tasa de cambio generan en la inflación interna de un país.

⁸Dentro de esta teoría los bienes transables de características similares deberían venderse al mismo precio, expresado en una misma moneda y en diferentes lugares del mundo. Esto bajo el supuesto de costos de transacción bajo y la no existencia de barreras comerciales.

⁹Purchasing power parity (*PPP*) es una condición en la cual el poder de compra de residentes domésticos relativo a los del resto del mundo es constante, así la *PPP* es una teoría simple que sostiene que el tipo de cambio nominal entre dos monedas debe ser igual a la tasa del nivel de precios entre dos países, de manera que la unidad monetaria de un país tenga el mismo poder de compra en un país extranjero. Para ver una referencia al respecto, consultar el capítulo 3 de Sarno and Taylor [2002].

¹⁰En estas teorías: $P = P^*e$; así bajo el supuesto de precios internacionales fijos (P^*), una devaluación nominal (e) se transmite completamente a los precios internos (P).

presas que tienen deudas en moneda extranjera¹¹, lo cual podría liberar recursos para estimular la demanda agregada interna. En ese entorno, el impacto sobre los precios internos es incierto y dependen de la estructura productiva de cada país, de su grado de apertura, nivel de desarrollo financiero y de la cobertura cambiaria, entre otras.

En la literatura colombiana son pocos los estudios respecto a los efectos de la tasa de cambio sobre otras variables y se observa que el análisis se ha centrado más en el *pass-through* del tipo de cambio a los precios, que sobre los efectos que sus variaciones puedan tener sobre la demanda agregada. Rincón et al. [2005] estiman el coeficiente de traspaso de la tasa de cambio a los precios importados para una muestra de sectores de la industria manufacturera y encuentran que la transmisión es incompleta tanto en el corto como en el largo plazo¹². Adicionalmente, las estimaciones realizadas por ellos no corroboran la hipótesis formulada por Taylor [2000], que en un entorno en el cual la inflación es baja y la tasa de cambio es flexible, el *pass-through* es bajo.

Algo similar se encuentra en el estudio de Parra [2010] en el cual la evidencia estadística sugiere un coeficiente de traspaso de 25 % para diferentes sectores de la economía colombiana¹³. En González et al. [2010] se encuentra que un choque (un incremento de 100 %) en la tasa de cambio, tiene un impacto de 6 % en los dos primeros trimestres sobre la inflación de la canasta de bienes importados y de un 58 % en el largo plazo, en promedio. Asimismo, los autores concluyen que el impacto de la tasa de cambio sobre los precios internos es mayor cuando cumple las siguientes condiciones: i) la economía está en auge, ii) la tasa de cambio nominal es menos volátil, iii) la tasa de cambio real está sobrevaluada y iv) la inflación se desacelera y es menos volátil.

En términos de la volatilidad, son escasos los trabajos que realizan la estimación para Colombia, puesto que la mayoría de estudios se enfocan en cuantificar la efectividad y el impacto de la política cambiaria sobre la volatilidad de la tasa de cambio (Echavarría et al. [2009], Rincón and Toro [2010]), los efectos que la nueva información o noticias tiene sobre el comportamiento de la tasa de cambio (Echavarría et al. [2013]) y los determinantes de la volatilidad y su relación con otras variables económicas (Lega et al. [2007], Mesa [2010]). Así, uno de los aportes de esta tesis a la literatura colombiana es estudiar métodos alternos para la estimación de la volatilidad y realizar un ejercicio empírico sobre la tasa de cambio peso-dólar en el período comprendido entre enero de 2002 y junio de 2013 que muestre las bondades de las diferentes alternativas.

¹¹En el caso contrario, una devaluación fuerte podría generar descalces cambiarios

¹²El coeficiente de *pass-through* estimado por ellos oscila entre 0.1 y 0.8

¹³Esto implica que un aumento de 1 % de la tasa de cambio genera un aumento de 0.25 % en la inflación interna.

1.2. Hechos Estilizados de las Series Financieras: El caso de la tasa de cambio peso-dólar

Es un consenso para la mayoría de analistas de mercado que los movimientos de los precios de los activos financieros se originan teniendo en cuenta un enfoque basado en eventos, en el cual se procura explicar y racionalizar un movimiento de mercado dado, relacionándolo con una noticia o anuncio de origen político o económico¹⁴. Partiendo de esto, la dinámica de los activos difieren entre si ya que cada activo responde a fenómenos y noticias diferentes. Numerosas investigaciones sobre las propiedades estadísticas de los retornos financieros sugieren que las variaciones que aparentemente son al azar, comparten y exhiben algunos rasgos característicos. Estas propiedades que son comunes a través de una amplia gama de instrumentos y de mercados a lo largo del tiempo se denomina hechos estilizados.

En esta sección se describen por medio de herramientas estadísticas los principales hechos estilizados de las series financieras. Para ello se toma como ejemplo la tasa de cambio peso-dólar, pero el lector puede generalizar los resultados para acciones, índices bursátiles, entre otros activos financieros. De acuerdo con Franses and van Dijk [2003], Cont [2001], Taylor [2005] y Mills and Markellos [2008], entre los principales rasgos de los retornos financieros se encuentran los siguientes:

- I **Ausencia de Autocorrelación (lineal)**: Las autocorrelaciones de los retornos financieros son casi siempre no significativas, excepto en escalas de tiempo muy pequeñas (por ejemplo datos intradía, cada 20 minutos) en donde los efectos de la micro-estructura de mercado juegan un papel importante. Este comportamiento se asemeja al de un proceso de ruido blanco, que refleja en cierta forma que el mercado del activo es eficiente, por lo tanto el precio no es pronosticable.
- II **La distribución de los retornos no es normal**: Este rasgo es conocido en la literatura financiera como colas pesadas (*Fat tails*), y parte del hecho de que los activos financieros ocasionalmente presentan grandes retornos positivos y negativos, lo cual es muy poco probable de ser observado si la distribución fuera normal.
- III **Conglomerados de volatilidad**: Esta propiedad muestra que la volatilidad condicionada del retorno de un activo cambia en el tiempo y que tiende a agruparse (*volatility clusters*). Allí se observa que la volatilidad presenta una correlación positiva, así un período en el que la volatilidad es elevada tienden a estar seguido por períodos de alta volatilidad, mientras que cuando la volatilidad es baja se espera que en los siguientes períodos permanezca de igual forma.

¹⁴Estas noticias o anuncios están asociadas dentro del término de error de los modelos (error ruido blanco ϵ_t), los cuales son innovaciones que responde a factores no explicados en cada momento del tiempo

IV **Asimetría de las pérdidas y ganancias:** En la práctica se observa que los precios de acciones e índices bursátiles presentan movimientos de mayor magnitud en las caídas que en las subidas. En las tasas de cambio esta propiedad es menos común y no se puede generalizar algo al respecto.

V **Normalidad Agregada :** A medida que se incrementa la periodicidad Δt , la distribución de los retornos tiende a parecerse más a una distribución normal. La forma de la distribución varía de acuerdo a la escala de tiempo¹⁵

VI **Efecto Apalancamiento:** Muchas medidas de volatilidad de un activo están correlacionadas negativamente con los retornos de un activo.

Lo anterior, en su mayoría, va ser modelado mediante el uso de las herramientas estadísticas y econométricas que se desarrollan en los siguientes capítulos, los cuales se aplican sobre la tasa de cambio colombiana.

1.2.1. Características de la tasa de cambio peso-dólar

Para este documento se utiliza la información de la **tasa representativa de mercado (TRM)**, que es el indicador oficial diario de la tasa de cambio en Colombia. Este corresponde al promedio ponderado de las compras y ventas efectuadas en el mercado cambiario. Las unidades de la variable están dadas por el número de pesos colombianos necesarios para obtener un dólar estadounidense, en un día determinado (Figura 1.1). La muestra utilizada comprende observaciones diarias desde enero de 2002 hasta junio de 2013, con lo que se cuenta con 2771 datos para las estimaciones¹⁶.

En la figura (1.1) se muestra el precio del activo de interés, en este caso la tasa de cambio peso-dólar, el cual indica en cuánto está valorado en cada momento del tiempo el peso colombiano respecto al dólar americano, y que en este trabajo se denota como S_t . A partir de esto se define la variable retorno como el cambio relativo en el precio de un activo financiero durante un intervalo de tiempo, la cual está expresada como un porcentaje. En tiempo discreto, esta medida se define como:

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}} = \frac{\Delta S_t}{S_{t-1}} \quad (1.1)$$

y en tiempo continuo

¹⁵sin embargo, esto no implica que necesariamente para los retornos finalmente tiendan a distribuirse normal. En la mayoría de los casos los rendimientos semestrales y anuales tienen un comportamiento más cercano al de una distribución normal.

¹⁶De la base de datos se excluyeron los días en que el mercado estaba cerrado, cuando no hubo comercio de divisas (*trading*). En otras palabras, no se incorporan días festivos ni los fines de semana



Figura 1.1: TRM

$$R(t) = \frac{S(t + \Delta t) - S(t)}{S(t)}$$

que es lo mismo que:

$$1 + R(t) = \frac{S(t + \Delta t)}{S(t)}$$

Teniendo en cuenta la propiedad que $\ln(1 + x) \approx x$ si x toma valores pequeños, entonces cuando Δt tiende a cero se tiene que $R(t) \approx \ln(1 + R(t))$. De esta manera, para un período de separación lo suficientemente pequeño ($\Delta t \rightarrow 0$), el retorno porcentual puede ser aproximado utilizando el logaritmo del retorno y éste es igual al incremento en los logaritmos de los precios del activo. Esto es,

$$R(t) \approx \ln S(t + \Delta t) - \ln S(t)$$

Para finalizar y utilizando las propiedades del retorno continuo se puede expresar (1.1) en términos de logaritmos, pero en tiempo discreto, de la siguiente manera:

$$y_t = \ln\left(\frac{S_t}{S_{t-1}}\right) = \ln S_t - \ln S_{t-1} = \Delta \ln S_t \quad (1.2)$$

Donde y_t , en este trabajo, denota el retorno diario de la tasa de cambio peso-dólar¹⁷. La Figura (1.2) muestra el comportamiento de y_t

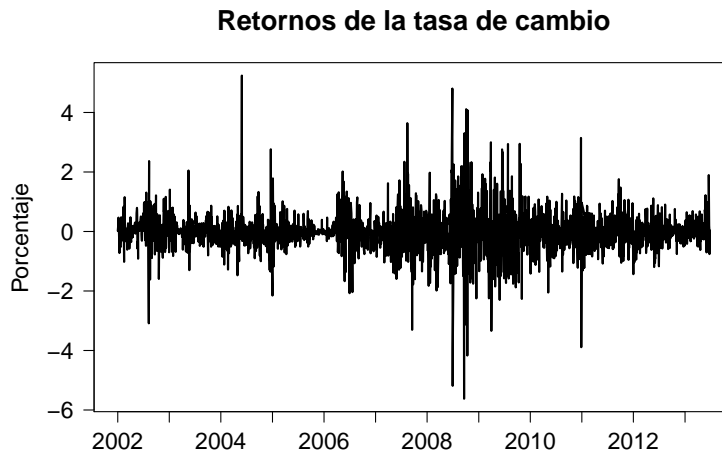


Figura 1.2: Retorno de la TRM

Se aprecia un período de gran volatilidad entre 2008 y 2009, lo que está explicado por la crisis financiera internacional y la incertidumbre a nivel mundial que se presentó durante dicho período. La turbulencia financiera que comenzó con la crisis hipotecaria en EEUU y estalló con la quiebra del Banco Estadounidense Lehman Brothers, generó problemas de liquidez, aumento en las primas de riesgo y salidas de capital en las economías emergentes como la colombiana. Todo ello ocasionó fuertes devaluaciones y un aumento de la volatilidad de la tasa de cambio en Colombia. Una descripción de los principales sucesos durante la crisis financiera y las consecuencias sobre las economías en desarrollo como la colombiana puede encontrarse en Toro et al. [2013].

La variable sobre la cual se implementan los modelos de heteroscedasticidad condicional y de volatilidad estocástica, que se presentan en el capítulo 2, es precisamente la denotada como y_t . Esto dado que y_t cumple con uno de los supuestos requeridos por ese tipo de modelos, la condición de estacionariedad (Ver definición en Apéndice C.3).

En el Apéndice C.1.3, el lector puede encontrar los procedimientos más comunes utilizados en el campo de la econometría de series de tiempo para transformar series que no son estacionarias. En particular, en este trabajo de tesis se realizó el siguiente procedimiento: primero, con el fin de estabilizar la varianza se aplicó el operador logaritmo natural al nivel de la serie (ósea, $\ln S_t$), y segundo, se realizaron pruebas de raíz unitaria para determinar el orden de integración de la serie (Dickey and Fuller [1981], Phillips

¹⁷ y_t el cual es calculado mediante la diferencia de los logaritmos del precio entre dos días (S_t y S_{t-1}), es decir, $y_t = \Delta \ln S_t$

and Perron [1988], Kwiatkowski et al. [1992] y Elliott et al. [1996]).

Como se muestra en el Apéndice C.1.3, el orden de integración, denotado por $I(d)$, corresponde al número de veces d que es necesario diferenciar la serie de tiempo original, para obtener una transformación de ella que sea estacionaria. Así, cuando la serie de tiempo es estacionaria, se dice que es integrada de orden cero ($I(0)$). En el cuadro (1.1), que se presenta a continuación, se hace un resumen de las pruebas de raíz unitaria realizadas en esta investigación.

Como resultado se encontró que el logaritmo de la TRM tiene una raíz unitaria, es decir es integrada de orden uno ($I(1)$), y que la primera diferencia de ésta era estacionaria, ósea es integrada de orden cero ($I(0)$). Las salidas del software econométrico de cada una de las pruebas pueden consultarse en Apéndice E.1):

Prueba *	log (TRM)	Conclusión log (TRM)	Δ log (TRM)	Conclusión Δ log (TRM)
Dickey and Fuller [1981]	No rechazo	I(1)	Rechazo	I(0)
Phillips and Perron [1988]	No rechazo	I(1)	Rechazo	I(0)
Elliott et al. [1996]	No rechazo	I(1)	Rechazo	I(0)
Kwiatkowski et al. [1992]	Rechazo	I(1)	No Rechazo	I(0)

*La hipótesis nula en las primeras tres pruebas es H_0 : la serie tiene una raíz unitaria (es $I(1)$), y la hipótesis alternativa es que la serie es estacionaria (es $I(0)$).

En la prueba Kwiatkowski et al. [1992], la hipótesis nula es que la serie es estacionaria, (es $I(0)$) y la hipótesis alterna es que la serie tiene raíz unitaria (es $I(1)$).

Cuadro 1.1: Resumen pruebas de raíz unitaria.

Como los modelos que se presentan en el capítulo 2, requieren que la serie de tiempo sea estacionaria, la variable sobre la cual se implementan las estimaciones es el retorno financiero de la tasa de cambio, que en este caso se define como la diferencia logarítmica de la TRM. En otras palabras, como log (TRM) tiene raíz unitaria (es lo mismo que $\ln S_t$ es $I(1)$), entonces se debe realizar una transformación para obtener una serie de tiempo que cumpla con el supuesto de estacionariedad. En este caso, se aplica el operador diferencia, Δ log (TRM), el cual de acuerdo a los resultados arrojados en el Cuadro (1.1) ya cumple con el supuesto requerido (es lo mismo que y_t es $I(0)$).

1.2.2. No Normalidad de los Retornos

Taylor [2005] y Franses and van Dijk [2003] muestran, sobre un conjunto de retornos financieros de acciones y monedas en el mundo, evidencia empírica sobre la distribución

de los retornos financieros. La principal conclusión es la no normalidad de la distribución de los retornos, de hecho, ellos observan que la distribución empírica presenta colas pesadas (kurtosis mayor a tres), apuntalamiento y en algunos casos asimetrías, en especial sobre las acciones. Para ello, utilizan el estimador de asimetría y de kurtosis como medidas muestrales para estimar ese hecho estilizado (Ver Apéndice A.2.3)

A continuación, se muestran las principales estadísticas descriptivas de los retornos de la tasa de cambio.

Obs. (T)	media	des. est.	mediana	min.	max.	asimetría	Exceso kurtosis *
2769	-0.01	0.69	-0.02	-5.62	4.86	0.19	9.09

* El programa escala la medida comparando con la normal, es decir al valor registrado originalmente le resta 3, que es la kurtosis de la distribución normal.

Cuadro 1.2: Estadísticas descriptivas de los retornos de la TRM.

En el cuadro (1.2) se observa, para las 2769 observaciones de los retornos de la tasa de cambio, que la media es cercana a cero y la desviación estándar muestral es 0.69. La asimetría es levemente positiva (0.19) y la kurtosis es elevada (>9.09). La kurtosis de la tasa de cambio peso-dólar es $(9.09 + 3) = 12.09$, porque el programa calcula el exceso de kurtosis respecto a la distribución Normal, la cual tiene una kurtosis de 3.

Asimismo, la figura (1.3) muestra una comparación entre la distribución empírica de los retornos de la TRM (azul) y una distribución Normal con media y varianza que se calculan mediante los valores muestrales de los retornos de la TRM (línea negra sólida), por medio de un histograma.

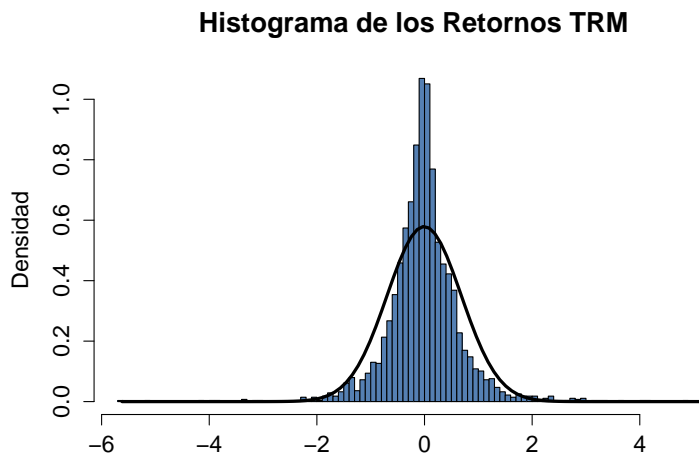


Figura 1.3: Histograma retornos de la TRM

Como se puede apreciar, la distribución de los retornos de la TRM presenta un mayor apuntalamiento comparado con la normal. Además, la distribución de los retornos de la TRM presenta colas pesadas frente a la normal, tal y como se observa en la figura (1.4) de los cuantiles empíricos (puntos azules) frente a los cuantiles teóricos (línea punteada roja).

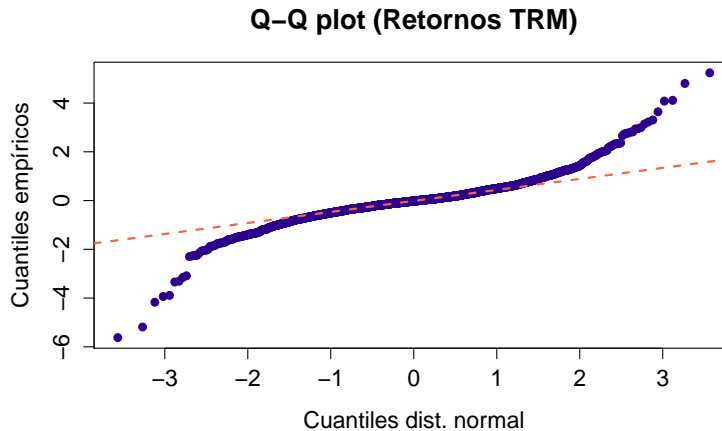


Figura 1.4: Q-Q plot retornos de la TRM

Praetz [1972], y muchos otros investigadores, han argumentado que los retornos observados (r_t) condicionados a la varianza condicionada (σ_t^2) se distribuyen como una normal, es decir que $r_t | \sigma_t^2 \sim N(\mu, f(\sigma_t^2))$, donde f es una función que permite modelar la volatilidad condicional y que es establecida *a priori* por el investigador. Clark [1973] utiliza el volumen de transacciones realizadas como aproximación de $f(\sigma_t^2)$. En el mismo documento el autor propone que la varianza condicional (σ_t^2) tiene una distribución log-normal, unos años más tarde, esta elección se convirtió en un supuesto muy popular utilizado en la literatura de los modelos de volatilidad estocástica.

En la revisión de literatura realizada se encontró que en la modelación de series de tiempo financieras también se asumen otro tipo de distribuciones para modelar los retornos, tales como la distribución log-normal, la *t-student* ó la distribución de error generalizada (*Generalized error distribution*) con colas más pesadas que permiten un mejor ajuste para los retornos. En el caso de la varianza condicional, distribuciones como la *gamma* (G) y la inversa gamma (*inverse gamma* (IG)) también son utilizadas.

1.2.3. Autocorrelación

Otro importante hecho estilizado es que, a menudo, si el mercado del activo financiero es eficiente (Ver sección 1.2.4), las correlaciones de los retornos no son significativas,

mientras que las correlaciones de los retornos al cuadrado sí lo son. Las siguientes figuras muestran el caso de la tasa de cambio peso-dólar

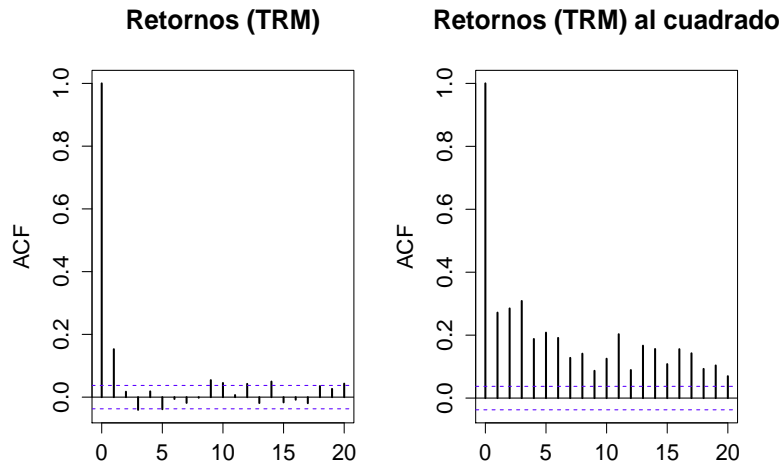


Figura 1.5: Correlogramas de los retornos de la TRM

Pese a que, en general, todos los retornos de activos financieros no presentan correlación, en el caso de la tasa de cambio de Colombia se aprecia claramente una correlación al menos de orden 1. Por su parte, los retornos al cuadrado presentan una estructura de correlación que sugiere que la varianza condicional no es constante en el tiempo, de hecho, decae exponencialmente.

Como se muestra en los siguientes capítulos la varianza condicional (σ_t^2), denotada en este trabajo como h_t , se modela a partir de la utilización de los modelos de heteroscedasticidad condicional y los modelos de volatilidad estocástica, los cuales presentan una estructura de correlación de h_t que decae exponencialmente, tal y como sucede empíricamente en los retornos al cuadrado de la TRM (figura (1.5))¹⁸. Por último, los modelos que son discutidos en el capítulo 2 replican dicho comportamiento de las correlaciones, que en este caso caracterizan a los retornos de la tasa de cambio peso-dólar elevados al cuadrado.

1.2.4. Hipótesis de Eficiencia de Mercado

Fama [1970] propuso que un mercado en el cual los precios reflejan completamente la información disponible es *mercado eficiente*. Formalmente, se dice que el mercado es efi-

¹⁸Para el retorno de la TRM, denotado en este trabajo de tesis como y_t , se puede definir su varianza de acuerdo con $V(y_t) = E(y_t^2) - [E(y_t)]^2$ (Ver Apéndice A.2.3). Si el retorno de la TRM (y_t) tiene una media que es estadísticamente cero, entonces se tiene que la varianza es $V(y_t) = E(y_t^2)$. Así, se observa que existe una relación entre los retornos al cuadrado (y_t^2) y la varianza de y_t . Este hecho se utiliza para modelar la volatilidad, en los modelos propuestos en el capítulo 2.

ciente con respecto a un conjunto de información si el precio del activo no es afectado por revelar esa información a todos los participantes. Más aún, eficiencia con respecto a un conjunto de información implica que es imposible obtener una utilidad económica por la comercialización de ese conjunto de información y el precio sólo cambia si existe nueva información disponible.

En su artículo menciona distintos niveles de eficiencia del mercado, dependiendo de tres subgrupos de información: Precios históricos, información públicamente disponible, y finalmente, toda la información incluyendo aquella privilegiada que no está disponible al público. Luego, categoriza la hipótesis de eficiencia de mercado en: eficiencia en sentido fuerte, semi-fuerte y débil. En la primera, se analiza si toda la información, tanto pública como privada está completamente reflejada en los precios de los activos. En la segunda, sólo se analiza la información disponible para el público. Por último, el concepto de eficiencia débil analiza si la información contenida en los precios históricos es completamente incorporada en los precios actuales. El autor concluye que las pruebas de la forma débil del modelo de mercado son las más voluminosas y los resultados están a favor de la hipótesis de los mercados eficientes. Para los otros dos tipos es mucho más complicado tratar de buscar evidencia al respecto.

Lo and Mackinlay [1988] define que los precios de un activo cuyo mercado sea eficiente siguen un comportamiento de caminata aleatoria o martingala, dado que en ese tipo de proceso estocástico, el mejor pronóstico que se puede dar del precio de mañana con la información disponible hasta el día de hoy, es el último precio observado. Los autores plantean que si se cumple la hipótesis de eficiencia de mercado los cambios en los precios no pueden ser predecibles y tienen la característica de un juego justo. Tanto en los modelos de tiempo continuo como los de tiempo discreto se puede modelar la anterior característica (Ver Apéndice A.8).

La literatura evalúa la eficiencia de mercado mediante diversas herramientas estadísticas. Una de ellas, la más usual en el campo de la econometría financiera, es la prueba de razón de varianzas (VR), que se basa en el trabajo pionero de Lo and Mackinlay [1988]. El test de VR verifica si los retornos siguen una caminata aleatoria evaluando si la varianzas de n períodos de los retornos es equivalente a n veces la varianzas de un período simple¹⁹.

Si el precio del activo financiero tiene una estructura de correlación serial, el retorno es predecible. Por el contrario, si el precio sigue una caminata aleatoria, el retorno no es predecible y se afirma que el mercado de dicho activo es eficiente. Para evaluar esto la prueba VR propuesta por Lo and Mackinlay [1988] plantea que para cualquier retorno de un activo financiero (y_t , en el presente trabajo es el retorno de la tasa de cambio peso-dólar) durante un período de tiempo $t = 1, 2, \dots, T$, el estimador VR está defini-

¹⁹En ese caso se dice que el mercado de ese activo es eficiente

do como (Wright [2000], Hoque and Kim [2007])

$$VR(y; n) = \left\{ \frac{1}{Tn} \sum_{t=n}^T (y_t + y_{t-1} + \dots + y_{t-n-1} - n\hat{\mu}) \right\} \div \left\{ \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2 \right\} \quad (1.3)$$

donde $\hat{\mu} = T^{-1} \sum_{t=1}^T y_t$, el estimador de VR es denotado por $VR(y; n)$ y n hace referencia al período de evaluación de la eficiencia.

Lo and Mackinlay [1988] mostró que si las variables aleatorias $\{y_t, y_{t-1}, \dots\}$ son independientes e idénticamente distribuidas (*iid*), entonces bajo la hipótesis nula de eficiencia de mercado ($H_0 : VR(y; n) = 1$), se puede evaluar la prueba VR de acuerdo con el siguiente estadístico

$$M(y, n) = (VR(y; n) - 1) \left(\frac{2(2n - 1)(n - 1)}{3nT} \right)^{-0,5} \quad (1.4)$$

el cual sigue una distribución normal estándar asintóticamente.

Utilizando el paquete *vrtest* del programa R, se ejecutó la prueba anterior. La figura (1.6) muestra el resultado para períodos de $n = 1, 2, \dots, 30$, con lo que se evalúa si el mercado de la TRM es eficiente a un día, dos días, y así sucesivamente, para treinta días.

Como se observa en la figura (1.6), el estimador puntual de VR (línea roja) está por fuera de las bandas de confianza (líneas azules punteadas), por lo que se rechaza la hipótesis nula de eficiencia de mercado para 1 día, 2 días, hasta treinta días. Esto sugiere que existe una estructura de correlación serial con la cual se puede pronosticar el retorno de la TRM.

En el entorno de series de tiempo, se puede utilizar los modelos ARIMA, *Autoregressive Integrated Moving Average* (Ver Tsay [2010]), para modelar la media de los retornos financieros. Para encontrar dicha parte predecible, se utilizó el paquete *forecast* del programa R y el comando *auto.arima*.

Como resultado, se obtuvo que la dinámica del retorno de la tasa de cambio peso-dólar puede ser modelada mediante un proceso ARIMA(1,0,2). El comando *auto.arima* elige empíricamente el mejor modelo ARIMA que se ajusta a los datos observados de los retornos de la tasa de cambio mediante criterios de información como el *Akaike* (Ver

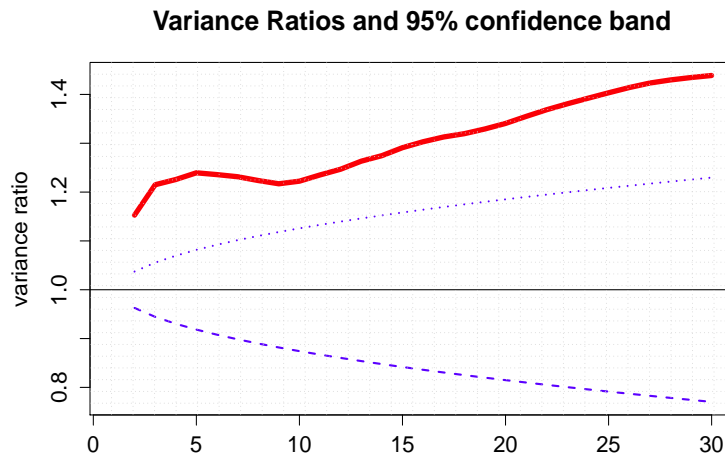


Figura 1.6: Prueba de razón de varianza sobre los retornos de la TRM

Bozdogan [1987]).

A continuación, se muestra la salida del paquete econométrico

```
>auto.arima(retorno)
Series: retorno
ARIMA(1,0,2) with zero mean

Coefficients:
ar1 ma1 ma2
-0.9356 1.0930 0.1702
s.e. 0.0356 0.0392 0.0186
```

```
sigma  $\hat{\sigma}$  estimated as 0.4609: log likelihood=-2856.64
AIC=5721.27 AICc=5721.29 BIC=5744.98
```

Teniendo en consideración esa estructura, $y_t = \beta_1 y_{t-1} + \varepsilon_t + \delta_1 \varepsilon_{t-1} + \delta_2 \varepsilon_{t-2}$, se puede limpiar la serie de los retornos eliminando la parte que puede ser explicada. En otras palabras, utilizando el ARIMA y los parámetros ($\beta_1 = -0.9356$, $\delta_1 = 1.0930$, $\delta_2 = 0.1702$), se re-escala la variable de estudio, es decir, a los retornos de la tasa de cambio se le resta el modelo estimado anterior ($y_t - \beta_1 y_{t-1} + \delta_1 \varepsilon_{t-1} + \delta_2 \varepsilon_{t-2} = \varepsilon_t$).²⁰.

²⁰Para dar un ejemplo al lector: suponga que los retornos $y_t \sim N(1, 2)$, es decir que y_t tiene media 1 y varianza 2. Asuma que el modelo ARIMA es una constante que modela la media igual a 1. Calcule una nueva variable x_t que está definida como $y_t - 1$ para cada t , entonces $x_t \sim N(0, 2)$. Ósea, el ARIMA permite re-escalar la media de y_t , sin afectar su varianza, que es el objeto de estudio de esta tesis.

En los siguientes capítulos donde se explican los modelos de heteroscedasticidad condicional y volatilidad estocástica (Ver capítulo 2), y se realiza la estimación de la volatilidad de la tasa de cambio (Ver capítulo 4), se asume que la parte predecible del retorno de la tasa de cambio peso-dólar es cero, es decir se supone que $y_t = \varepsilon_t$, entonces se implementó la siguiente estrategia: i) estimar el mejor modelo ARIMA, en este caso ARIMA(1,0,2), ii) tomar el residual del modelo estimado y iii) calcular los modelos de volatilidad sobre la nueva serie, que se obtiene como el residual del modelo ARIMA (ε_t). Lo anterior, no altera los principales resultados de este trabajo de tesis puesto que con dicho procedimiento se re-escala la media de los retornos sin modificar la varianza, que es la variable de interés de la tesis.

Tomando la serie obtenida de (ε_t) mediante el uso del ARIMA(1,0,2), se realizó nuevamente la prueba VR (Figura 1.7). Allí, se observa que el estimador puntual de VR, se encuentra dentro de las bandas de confianza por lo que no se rechaza la hipótesis nula de eficiencia de mercado. Es decir, una vez eliminada la parte explicada, el restante sigue una caminata aleatoria.

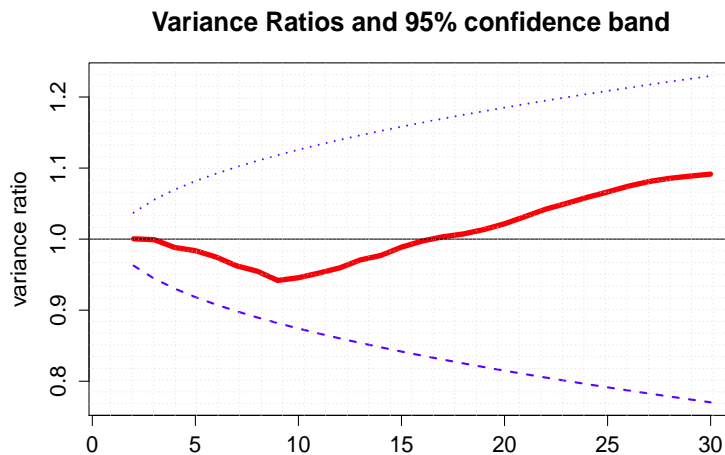


Figura 1.7: Prueba de razón de varianza sobre la parte no explicada en el ARIMA(1,0,2) de los retornos

Capítulo 2

Modelos de volatilidad: heteroscedasticidad condicional vs. volatilidad estocástica

En la modelación financiera se pueden asumir diferentes hipótesis sobre la volatilidad de un activo con el fin de simplificar el análisis de valoración, cálculos de opciones o contratos a futuro ó cualquier medición de riesgo que afronte dicho activo financiero. Cabe señalar, como se mencionó anteriormente, que la volatilidad está asociada con las variaciones que presenta una variable financiera de interés y ésta puede ser tratada como una secuencia determinista ó estocástica. Algunos investigadores suponen que la varianza es determinista y constante a lo largo del tiempo, para ello fijan como estimador puntual a la varianza muestral de los retornos del activo. Sin embargo, como se mostró, los hechos estilizados de los activos financieros (acciones, bonos, tasas de cambio) sugieren que la volatilidad, en general, no es constante. Por lo tanto, asumir homoscedasticidad¹ en el tiempo no es un supuesto muy realista.

Para modelar los rasgos de las series financieras y acercarse un poco más a la realidad, se han desarrollado modelos de volatilidad que consideran secuencias estocásticas para dicho proceso, entre ellos los modelos tradicionales de heteroscedasticidad condicional y los de volatilidad estocástica. Así, en este capítulo se pretende presentar las principales características de estos modelos, teniendo en cuenta que entre los objetivos de esta tesis se encuentra presentar algunas de las diferencias entre esas familias de modelos y contrastar los resultados que se obtienen de las estimaciones de la volatilidad de la tasa de cambio colombiana. Los desarrollos matemáticos que se muestran en las siguientes secciones están basados en gran parte de la lectura realizada de las siguientes referencias Engle [1982], Bollerslev [1986], Franses and van Dijk [2003], Enders [2004], Taylor [2005] y Casas and Cepeda [2008].

¹varianza constante

2.1. Modelos de heteroscedasticidad condicional

Esta familia de modelos fue introducida inicialmente por Engle [1982], quien utilizó un proceso estocástico para estimar volatilidad denominado modelo ARCH (*Autoregressive conditional heteroscedasticity*), en los cuales la varianza condicional depende del cuadrado de los rezagos de los términos aleatorios llamados innovaciones o errores (choques). De esta forma, la varianza condicional hace referencia a la volatilidad en un instante de tiempo t condicionada a la información relevante y conocida hasta ese momento. Luego, Bollerslev [1986] generalizó ese primer modelo y propuso el GARCH (*Generalized autoregressive conditional heteroscedasticity*). Allí, la varianza condicional, además de depender de los cuadrados de las perturbaciones pasadas, está también en función de las varianzas condicionales de períodos anteriores.

Los modelos ARCH y GARCH logran reproducir mediante la modelación econométrica los principales rasgos característicos de las series financieras, en particular, los conglomerados de volatilidad y la no normalidad de los retornos debido a asimetrías o kurtosis altas (colas pesadas). Por ejemplo, el modelo replica la siguiente situación: cuando aumenta la incertidumbre y los inversionistas están temerosos, la volatilidad se mantiene alta, mientras que por el contrario, en tiempos de calma, la volatilidad debe ser baja y continuar así por el tiempo que perduró esa tranquilidad.

Por otra parte, los pronósticos calculados a partir de estos modelos no son iguales a la última estimación que se tienen de volatilidad, por el contrario cambian a lo largo del tiempo. La volatilidad evoluciona de acuerdo a la estimación de una ecuación sugerida por el investigador, a continuación se muestra el ARCH(1), ARCH(q), GARCH(p,q) y el GARCH(1,1). La utilidad para los inversionistas radica en que los modelos capturan cambios en la volatilidad de corto plazo, lo que permite tomar decisiones de compra o venta de activos de acuerdo al riesgo que se vaya presentando en el día-día o semana-semana.

2.1.1. Características y propiedades

Para comenzar se asume que los retornos (y_t) de un activo en el momento t están en función de: i) una parte que puede ser predecible y ii) otra no predecible o aleatoria, ésta última sigue un proceso ruido blanco (Ver definición del Apéndice C.1). De esta manera, los retornos podrían escribirse como sigue:

$$y_t = \underbrace{E[y_t | \Psi_{t-1}]}_{\text{Predecible}} + \underbrace{\sigma_\varepsilon \varepsilon_t}_{\text{aleatorio}} \quad (2.1)$$

donde Ψ_{t-1} es el conjunto de información disponible² y relevante hasta el momento $t - 1$. ε_t se supone que teóricamente es independiente e idénticamente distribuido, es decir, $\varepsilon_t \stackrel{iid}{\sim} N(0, 1)$ y σ_ε es una constante asociada a la volatilidad de los retornos (también podría definirse $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon)$, simplemente multiplicando la constante σ_ε por la variable aleatoria normal estándar).

Si el mercado del activo financiero es eficiente, entonces los precios del activo no pueden ser predecibles, y la parte predecible sería igual a cero. En caso contrario, cuando el mercado no es eficiente entonces se puede encontrar una estructura de correlación temporal para hacer predicciones (Ver sección 1.2.4). En la aplicación que se realiza en el capítulo 4, se utiliza el residual estimado mediante el ARIMA(1,0,2), el cual fue previamente justificado en el capítulo 1.

Generalmente, en los modelos de series de tiempo se supone que el componente aleatorio es homoscedástico, tanto condicionado como no condicionado, es decir que, al tomar el valor esperado para hallar la varianza se obtiene una constante. En este caso, dicho supuesto está representado por el hecho de que $E[\sigma_\varepsilon \varepsilon_t^2] = E[\sigma_\varepsilon \varepsilon_t^2 | \Psi_{t-1}] = \sigma_\varepsilon^2$ para todos los momentos de tiempo t .³

En los modelos de heteroscedasticidad condicional, que se introducen a continuación, se relaja la hipótesis de normalidad y de varianza condicional constante ($E[\sigma_\varepsilon \varepsilon_t^2 | \Psi_{t-1}] = \sigma_\varepsilon^2$) que se tienen en modelos de series de tiempo como los ARIMA. Además, permiten tener procesos que mantienen propiedades como las de los procesos ruido blanco, pero que ahora están formados por variables dependientes.

Para incorporar en el análisis que la varianza condicional no es constante, se denota h_t como la volatilidad, la cual tiene una dinámica en el tiempo. De esta manera, se tiene que la varianza condicional de ε_t cambia en el tiempo de acuerdo con h_t . En este capítulo se muestran varias formas como h_t puede ser modelada, por el momento es importante resaltar al lector que $\varepsilon_t | \Psi_{t-1} \sim N(0, h_t)$, en lugar de $\varepsilon_t | \Psi_{t-1} \sim N(0, 1)$ como se supone en la ecuación (2.1). Asimismo, se tiene que $E[\sigma_\varepsilon \varepsilon_t^2 | \Psi_{t-1}] = \sigma_\varepsilon h_t$ en lugar de $E[\sigma_\varepsilon \varepsilon_t^2 | \Psi_{t-1}] = \sigma_\varepsilon^2$.

Como se había mencionado, una medida de riesgo en el ámbito financiero es la desviación estándar de los retornos de los activos, por lo que en este tipo de modelos la variable h_t se convierte en el objeto de interés. La estimación de h_t es el objetivo de esta tesis, puesto que su cuantificación refleja la volatilidad de los retornos del activo (y_t).⁴ El sub-índice t indica que la volatilidad (h_t) cambia en el tiempo. Engle [1982] y

² Ψ_{t-1} es una filtración. Ver Apéndice Conceptos Básicos de Probabilidad al final del documento.

³Como se asume en la metodología ARIMA, ε_t es ruido blanco (Ver Apéndice C.1). La varianza de ε_t está definida como $V(\varepsilon_t) = E(\varepsilon_t^2) - [E(\varepsilon_t)]$. Sabemos que ε_t tiene media igual a cero ($[E(\varepsilon_t)]=0$), por lo que la varianza está dada por $V(\varepsilon_t) = E(\varepsilon_t^2)$. Y se asumió que la varianza de ε_t es igual a uno.

⁴retorno de la tasa de cambio de Colombia

Bollerslev [1986] plantearon que h_t está en función del cuadrado de las innovaciones pasadas ($\varepsilon_t^2, \varepsilon_{t-1}^2, \dots$) y/o de las volatilidades estimadas anteriores (h_t, h_{t-1}, \dots).

Partiendo de la ecuación (2.1), la parte predecible puede ser determinista o modela mediante la metodología ARIMA⁵, mientras que la parte aleatoria se modela de la siguiente forma de acuerdo a la propuesta de Engle [1982]:

$$\sigma_\varepsilon \varepsilon_t = v_t \sqrt{h_t} \quad (2.2)$$

donde $v_t \stackrel{iid}{\sim} N(0, 1)$ y es independiente de $\sqrt{h_t}$. Dado lo anterior, Engle [1982] prueba que ε_t continúa siendo un proceso ruido blanco pese a la dependencia que h_t pueda tener sobre el pasado⁶. Así, $E[\sigma_\varepsilon \varepsilon_t] = \sigma_\varepsilon E[v_t \sqrt{h_t}] = \sigma_\varepsilon E[v_t] E[\sqrt{h_t}] = 0$ ya que la media de v_t es cero. De igual manera, la varianza no condicionada de la parte aleatoria es constante: $E[\sigma_\varepsilon^2 \varepsilon_t^2] = \sigma_\varepsilon^2 E[v_t^2 h_t] = \sigma_\varepsilon^2 E[v_t^2] E[h_t] = \sigma_\varepsilon^2 \bar{h}$.⁷

Sin embargo, la varianza no condicionada de ε_t no será contante, puesto que $\text{var}(\varepsilon_t | \Psi_{t-1}) = E[v_t^2 | \Psi_{t-1}] E[h_t | \Psi_{t-1}] = h_t$. Por último, el proceso ε_t no está correlacionado ya que $E[\varepsilon_t \varepsilon_{t-k}] = E[v_t \sqrt{h_t} v_{t-k} \sqrt{h_{t-k}}] = E[v_t v_{t-k}] E[\sqrt{h_t} \sqrt{h_{t-k}}] = 0$ para cualquier momento del tiempo $k \neq 0$ (dado que v_t es *iid*).

Como se acaba de mostrar, el componente aleatorio (ε_t) continúa siendo ruido blanco en la ecuación (2.2), la única diferencia es que ahora la varianza condicional no es constante y varía en el tiempo acorde a una función que depende de los cuadrados de las innovaciones anteriores y/o de las varianzas pasadas, la cual define h_t .⁸ A continuación se presentan los dos modelos más utilizados para estimar la volatilidad condicionada (h_t), en cada uno de ellos se especifica una función para h_t . Cabe señalar que existe una gran variedad de modelos, entre los que se destacan los que modelan asimetrías, pero para simplificar el análisis esta tesis concentra su atención en los modelos simétricos⁹.

⁵Esto no es objeto de interés en el presente trabajo de tesis.

⁶Esto dado que está en función del cuadrado de las innovaciones pasadas ($\varepsilon_t^2, \varepsilon_{t-1}^2, \dots$) y/o de las volatilidades estimadas anteriores (h_t, h_{t-1}, \dots).

⁷Esto se obtiene dado que la varianza de $v_t = 1$ ($E[v_t^2] = 1$) y $E[h_t]$ se asume constante e igual a su promedio (\bar{h})

⁸En estos modelos $\text{var}(\varepsilon_t | \Psi_{t-1}) = h_t$

⁹Aunque no se muestra en este documento, se realizaron algunos ejercicios con modelos como el EGARCH y APARCH que no fueron muy diferentes de los obtenidos por el GARCH. Pese a ello, si debe señalarse que la tasa de cambio de Colombia presenta la presencia de efectos asimétricos de acuerdo con la prueba de Engle and Ng [1993]

2.1.2. Modelo ARCH(q)

Engle [1982] propuso que el siguiente modelo para modelar series de tiempo en donde la varianza condicionada no es constante. El autor propone un sistema compuesto por tres ecuaciones, las primeras dos dadas por las ecuaciones (2.1 y 2.2), y la tercera por la siguiente forma funcional para la varianza condicionada (h_t), en donde considera que ésta puede ser expresada en función de los rezagos de las innovaciones al cuadrado:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 \quad (2.3)$$

Dentro de esta ecuación los parámetros deben cumplir las siguientes condiciones para garantizar la no negatividad de la varianza condicional: $\alpha_0 > 0$ y $\alpha_i \geq 0$ para todo $i = 1, \dots, q$. Para simplificar la notación que se utiliza en adelante, se asume que en las ecuaciones (2.1 y 2.2) se multiplicó el σ_ε por la variable aleatoria normal ε_t , por lo que ahora $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

El valor de largo plazo de la varianza condicional (σ_ε^2) puede ser calculado al tomar el valor esperado de h_t :

$$\begin{aligned} E[h_t] &= E[\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2] \\ \text{como } h_t &= E[\varepsilon_t^2 | \Psi_{t-1}] \text{ y } E[E[\varepsilon_t^2 | \Psi_{t-1}]] = E[\varepsilon_t^2] = \sigma_\varepsilon^2 \\ \text{entonces } \sigma_\varepsilon^2 &= \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i} \end{aligned} \quad (2.4)$$

Para que el modelo sea covarianza-estacionario todas las raíces del polinomio de rezagos ($1 - \alpha_1 L - \dots - \alpha_q L^q$) deben estar fuera del círculo unitario (Ver Engle [1982], Bollerslev [1986], Nelson [1990]). Esto es similar a la condición de estabilidad en una ecuación en diferencias.

Para caracterizar este tipo de modelo y evidenciar como éste replica algunos de los hechos estilizados de las series de tiempo financieras mencionados, se toma como ejemplo un q igual a uno, por lo que a continuación se expresa el modelo ARCH(1). Adicionalmente, el componente predecible en la ecuación (2.1) se asume cero por simplicidad. Así, el modelo completo queda de la siguiente forma:

$$\begin{aligned}
y_t &= \varepsilon_t \quad \text{donde } \varepsilon_t | \Psi_{t-1} \sim N(0, h_t) \\
v_t &= \frac{\varepsilon_t}{\sqrt{h_t}} \\
h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2
\end{aligned} \tag{2.5}$$

donde v_t es el residual estandarizado, y dado que condicionalmente ε_t se distribuye como una normal, entonces $v_t \sim N(0, 1)$. Una de las medidas de ajuste del modelo se efectúa mediante la evaluación de pruebas de hipótesis que verifiquen el cumplimiento de la distribución asumida para v_t . Aunque en esta parte del documento se utiliza la distribución normal para facilitar los cálculos, en la práctica se pueden implementar distribuciones que tengan colas más pesadas como por ejemplo la *t-student* ó *Generalized Error Distribution (GED)*, sin que esto afecte las principales características que se exponen enseguida.

Como en este modelo los retornos(y_t) son iguales al proceso ruido blanco ε_t , y la volatilidad condicional está dada por $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$, se puede apreciar que los conglomerados de volatilidad se dan a través de esa ecuación. Por ejemplo, cuando hay perturbación de fuerte impacto hoy (t), entonces la volatilidad condicional de mañana ($t + 1$) se ve afectada en α_1 por el tamaño del choque en t al cuadrado, ósea que el efecto se amplifica. Mientras que, cuando la perturbación es baja hoy (t), entonces la volatilidad condicional de mañana ($t + 1$) se ve influenciada en menor medida. En otras palabras, cuando la incertidumbre es elevada se espera que la volatilidad de los períodos siguientes continúe siendo alta, mientras que en momentos de calma, los inversionistas esperan que la volatilidad permanezca baja.

Otro hecho estilizado tiene que ver con las colas pesadas en los retornos, donde la kurtosis es mayor a la que se observa en una distribución normal (kurtosis =3). Para ver como el modelo replica este fenómeno, se procede a calcular los momentos incondicionales de los retornos (y_t):

$$\begin{aligned}
E[y_t^2] &= E[v_t^2 h_t] = \frac{\alpha_0}{1 - \alpha_1} \\
E[y_t^4] &= E[v_t^4 h_t^2] = 3 \left(\frac{\alpha_0^2 + 2\alpha_0 \alpha_1 E[y_t^2]}{1 - 3\alpha_1^2} \right)
\end{aligned} \tag{2.6}$$

De esto se puede inferir que $\alpha_0 > 0$ y $0 < \alpha_1 < 1$, puesto que la varianza del retorno ($E[y_t^2]$) es por definición positiva. Por su parte, la kurtosis de los retornos (y_t) está dada por:

$$\kappa = \frac{E[y_t^4]}{E[y_t^2]^2} = 3 \left(\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \right) \quad (2.7)$$

El valor de la kurtosis en (2.7) es mayor a tres si $\alpha_1^2 < \frac{1}{3}$. Lo anterior implica que el modelo ARCH(1) puede tener colas más pesadas que las que tiene la distribución normal, y con ello se comprueba que el modelo replica la no normalidad de los retornos financieros.

Por último, se puede comprobar que los retornos al cuadrado están correlacionados y que su correlación decae exponencialmente. Para ello se puede expresar el ARCH(1) de h_t como un modelo AR(1) de ε_t^2 de la siguiente forma: primero, a la expresión $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ se suma y resta a ambos lados ε_t^2 , segundo, h_t pasa a sustraer al otro lado y tercero, se define una variable u_t como la resta entre ε_t^2 y h_t . Una característica de u_t es que su media es cero y es ruido blanco, como en cualquier modelo ARIMA ($E[u_t | \Psi_{t-1}] = 0$) (Ver Franses and van Dijk [2003]). Así, la representación AR(1) de los residuales al cuadrado es:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + u_t \quad (2.8)$$

Utilizando la teoría de modelo de series de tiempo ARIMA, la función de auto-correlación simple de un AR(1) está expresada de la siguiente manera¹⁰:

$$\rho_k = \alpha_1^k \quad (2.9)$$

Dado que $\alpha_1 < 1$ entonces la función de auto-correlación simple (ρ_k) decae exponencialmente cuando k tiende a infinito.

Lo anterior, aplica para el caso $q=1$, pero de acuerdo con Franses and van Dijk [2003] y Tsay [2010], en general los modelos ARCH(q) capturan adecuadamente los patrones de la dinámica en la volatilidad condicional de una serie financiera y mencionan que es usual encontrar en las aplicaciones empíricas un valor alto para q . Esto puede convertir el proceso de estimación en algo engorroso debido a las condiciones de no negatividad y estacionariedad que necesitan ser impuesta para cada uno de ellos. Para reducir los problemas computacionales, en la práctica se imponen algunas estructuras sobre los parámetros que acompañan los rezagos de las innovaciones al cuadrado en el

¹⁰ k hace referencia a los rezagos que se tendrían en cuenta en el cálculo de la correlación

modelo ARCH(q), tal como que $\alpha_i = \alpha(q+1-i)/(q(q+1)/2)$ donde $i = 1, \dots, q$, lo cual implica que éstos decaen linealmente y suman α (Ver Engle [1982], Engle [1983]). Para solucionar este problema Bollerslev [1986] sugirió adicionar rezagos de las varianzas condicionadas estimadas al modelo ARCH(q) y surge el modelo GARCH que se explica enseguida.

2.1.3. Modelo GARCH(p, q)

Bollerslev [1986] propuso una clase de modelos más general denominado GARCH(p, q), en donde además de los rezagos de las innovaciones al cuadrado que se tenían en el modelo ARCH(q), ahora se incluyen rezagos de las varianzas condicionadas. En particular, en el modelo se tendrían q parámetros asociados al cuadrado de los errores retrasados q períodos y p estimadores asociados a las varianzas condicionadas retrasadas p períodos. Así, similar a la sección anterior, simplemente el modelo GARCH añade una nueva ecuación para estimar la varianza condicionada h_t de la siguiente forma:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \dots + \beta_p h_{t-p} \quad (2.10)$$

Dentro de esta ecuación los parámetros deben cumplir las siguientes condiciones para garantizar la no negatividad de la varianza condicional: $\alpha_0 > 0$, $\alpha_i \geq 0$ para todo $i = 1, \dots, q$ y $\beta_j \geq 0$ para todo $j = 1, \dots, p$ (Ver Bollerslev [1986] y Bollerslev et al. [1992])¹¹. Para simplificar la notación que se utiliza en adelante, se asume que en las ecuaciones (2.1 y 2.2) se multiplicó el σ_ε por la variable aleatoria normal ε_t , por lo que ahora $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

El valor de largo plazo de la varianza condicional (σ_ε^2) puede ser calculado al tomar el valor esperado de h_t , pero ahora tomando la expresión (2.10):

$$\begin{aligned} E[h_t] &= E[\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \dots + \beta_p h_{t-p}] \\ \text{como } h_t &= E[\varepsilon_t^2 | \Psi_{t-1}] \text{ y } E[E[\varepsilon_t^2 | \Psi_{t-1}]] = E[\varepsilon_t^2] = \sigma_\varepsilon^2 \\ \text{entonces } \sigma_\varepsilon^2 &= \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j} \end{aligned} \quad (2.11)$$

¹¹Esta condición podría ser relajada, siempre y cuando, se cumpla que $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$.

Para que el modelo sea covarianza-estacionario todas las raíces del polinomio de rezagos $(1 - \alpha_1 L - \dots - \alpha_q L^q)$ y $(1 - \beta_1 L - \dots - \beta_p L^p)$ deben estar fuera del círculo unitario (Ver Bollerslev [1986] y Nelson [1990]). Si el lector desea verificar las propiedades de los primeros momentos, así como las condiciones de estabilidad de esta familia de modelos se recomienda la lectura adicional de Nelson and Cao [1992], Berkes et al. [2003] y Posedel [2005].

Para caracterizar este tipo de modelo y evidenciar como éste replica algunos de los hechos estilizados de las series de tiempo financieras mencionadas en capítulos anteriores, se toma como ejemplo el modelo GARCH(1,1) y el componente predecible en la ecuación (2.1) se asume cero por simplicidad. Así, el modelo completo queda de la siguiente forma:

$$\begin{aligned} y_t &= \varepsilon_t \quad \text{donde } \varepsilon_t | \Psi_{t-1} \sim N(0, h_t) \\ v_t &= \frac{\varepsilon_t}{\sqrt{h_t}} \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned} \tag{2.12}$$

donde v_t es el residual estandarizado, y dado que condicionalmente ε_t se distribuye como una normal, entonces $v_t \sim N(0, 1)$. Una de las medidas de ajuste del modelo se efectúa mediante la evaluación de pruebas de hipótesis que verifiquen el cumplimiento de la distribución asumida para v_t .

Al igual que el caso anterior, aunque en esta parte del documento se utiliza la distribución normal para facilitar los cálculos, en la práctica se pueden implementar distribuciones que tengan colas más pesadas como por ejemplo la *t-student* ó *Generalized Error Distribution (GED)*, sin que esto afecte las principales características que se exponen enseguida.

Similar a lo expuesto en el modelo ARCH, aquí los conglomerados de volatilidad también se dan a través de la ecuación $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$, donde las noticias o sorpresas en ε_{t-1} afectan de manera positiva a la varianza condicionada h_t . Si ε_{t-1} toma valores elevados, esto incrementa la volatilidad, y cuando toma valores pequeños entonces disminuye la volatilidad.

Para verificar la presencia de colas pesadas en la distribución de los retornos, donde la kurtosis es mayor a la que se observa en una distribución normal (kurtosis =3), se procede a calcular los momentos incondicionales de los retornos (y_t):

$$\begin{aligned} E[y_t^2] &= E[v_t^2 h_t] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \\ E[y_t^4] &= E[v_t^4 h_t^2] = 3E[h_t^2] \end{aligned} \quad (2.13)$$

De esto se puede inferir que $\alpha_0 > 0$ y $\alpha_1 + \beta_1 < 1$. Por su parte, la kurtosis de los retornos (y_t) está dada por:

$$\kappa = \frac{E[y_t^4]}{E[y_t^2]^2} = 3 \left(\frac{1 - (\alpha_1 + \beta_1)^2}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} \right) \quad (2.14)$$

El valor de la kurtosis en (2.14) es mayor a tres si $(\alpha_1 + \beta_1)^2 - 2\alpha_1^2 < 1$. Lo anterior implica que el modelo GARCH(1,1) puede tener colas más pesadas que las que tiene la distribución normal, y con ello se comprueba que el modelo replica el hecho estilizado de colas pesadas y no normalidad de los retornos de la tasa de cambio peso-dólar.

Por último, se puede comprobar que los retornos al cuadrado están correlacionados y que su correlación decae exponencialmente. Para ello se puede expresar el GARCH(1,1) de h_t como un modelo ARMA(1,1) de ε_t^2 de la siguiente forma: primero, a la expresión $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$ se suma y se resta a ambos lados ε_t^2 , segundo, h_t pasa a sustraer al otro lado y tercero, se define una variable u_t como la resta entre ε_t^2 y h_t . Así, se obtiene la ecuación (2.15), la cual es una representación ARMA(1,1) de los residuales al cuadrado

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 + u_t - \beta_1 u_{t-1} \quad (2.15)$$

Utilizando la teoría de modelo de series de tiempo ARIMA, la función de auto-correlación simple de un ARMA(1,1) está expresada de la siguiente manera¹²:

$$\begin{aligned} \rho_1 &= \alpha_1 + \frac{\alpha_1^2 \beta_1}{1 - 2\alpha_1 \beta_1 - \beta_1^2} \\ \rho_k &= (\alpha_1 + \beta_1)^{k-1} \rho_1 \end{aligned} \quad (2.16)$$

Dado que $\alpha_1 + \beta_1 < 1$ entonces la función de auto-correlación simple (ρ_k) decae exponencialmente cuando k tiende a infinito. Aunque el modelo general GARCH(p,q)

¹² k hace referencia a los rezagos que se tendrían en cuenta en el cálculo de la correlación

podría ser un modelo de interés teórico, en la práctica el modelo GARCH(1,1) presenta un buen ajuste y usualmente parece ser el más adecuado para modelar las series de tiempo financieras (Ver referencias como Bollerslev et al. [1992], Hansen and Lunde [2001], Mapa [2004] y Matei [2009]).

2.2. Modelo de volatilidad estocástica

Los modelos de heteroscedasticidad condicional como los ARCH o GARCH asumen que la volatilidad puede ser observada un paso adelante con la información de las realizaciones de las variables aleatorias ε_t y h_t . En estos modelos (ARCH y GARCH) las noticias o choques se incorporan únicamente a través de las innovaciones ε_t , es decir que sólo existen perturbaciones sobre la ecuación de la media (la que describe el comportamiento del retorno del activo). De esta forma, la volatilidad se estima a través de la información pasada, y esto se torna predecible porque para cada momento t ya se conoce la información en $t - 1$.

Partiendo de esto, un modelo más realista para la volatilidad se puede basar en una ecuación de comportamiento que tiene un componente predecible que depende de la información pasada (como en los GARCH) y un componente de ruido inesperado. En este caso, la volatilidad es una variable *no observada latente*. Una interpretación de la volatilidad latente es que representa la llegada de nueva información en el mercado (Ver Clark [1973]). Este tipo de modelación corresponde a los denominados *modelos de volatilidad estocástica* (SV). A continuación se describen algunos de los principales rasgos.

2.2.1. Características y propiedades

En el modelo univariado de volatilidad estocástica (SV) la dinámica del precio del activo financiero resulta de los movimientos de un índice que mide su valor o precio S_t y su volatilidad h_t , las cuales se derivan de un proceso de difusión continuo mediante movimientos brownianos (Ver definición A.13). Así, las ecuaciones que describen el comportamiento del precio en tiempo continuo son (Algunas referencias que el lector puede consultar son Rosenberg [1972], Taylor [1982], Taylor [1986], Hull and White [1987], Ghysels et al. [1996] y Johannes and Polson [2010]):

$$\begin{aligned} d\log S_t &= \psi dt + \sqrt{\sigma_t} dB_t^s \\ d\log \sigma_t &= \kappa(\gamma - \log \sigma_t) dt + \tau dB_t^V \end{aligned} \quad (2.17)$$

donde los parámetros que gobiernan la evolución del sistema son $(\psi, \kappa, \gamma, \tau)$ y los movimientos brownianos para el precio (S) y la volatilidad (V) son (B_t^p, B_t^V) , los cuales posiblemente están correlacionados. γ es el atractor del proceso de volatilidad, κ está asociado a la velocidad de ajuste o persistencia de la volatilidad, ψ es la tasa de crecimiento del precio del activo y τ es la volatilidad del proceso de varianza del activo. Aunque esta aproximación en tiempo continuo es adecuada, los datos surgen en tiempo discreto por lo que una estrategia puede ser discretizar el proceso 2.17 mediante el uso del método de Euler (Ver Apéndice D.1).

Como resultado de este procedimiento se obtiene el comúnmente denominado *modelo de volatilidad estocástica auto-regresivo (SV)* (Taylor [1986] y West and Harrison [2007]) y este es descrito por el siguiente modelo dinámico no lineal:

$$y_t = \exp \left\{ \frac{h_t}{2} \right\} \sigma_\varepsilon \varepsilon_t \quad (2.18)$$

$$\text{donde } \varepsilon_t \stackrel{iid}{\sim} N(0, 1)$$

$$h_t = \mu + \phi h_{t-1} + \tau \eta_t \quad (2.19)$$

$$\text{donde } \eta_t \stackrel{iid}{\sim} N(0, 1)$$

donde y_t es el logaritmo de los retornos y h_t es el logaritmo de las varianzas ($h_t = \log \sigma_t$), que para esta tesis indica la medida de la volatilidad de la tasa de cambio peso-dólar. ε_t y η_t son las perturbaciones o ruidos no esperados para los retornos y para la volatilidad, respectivamente, cada uno de ellos se asume que se distribuye normal. Aunque algunos estudios presentan evidencia empírica que soporta este supuesto para tasas de cambio y acciones (Ver Andersen et al. [2001a] y Andersen et al. [2001b]), se podrían utilizar otras distribuciones que tengan colas más pesadas como es el caso de *t-student* ó *Generalized Error Distribution (GED)*. De acuerdo con Taylor [1982] la relación entre los parámetros del tiempo discreto y continuo está dada por $\mu = \kappa\gamma$ y $\phi = 1 - \kappa$.¹³

Como se expone en el siguiente capítulo este tipo de modelos se estiman por métodos más exigentes puesto que ahora la variable latente tiene otro término de ruido. Una de las técnicas de estimación más populares, pero a la vez más complejas de implementar, es la metodología de simulaciones de Monte Carlo (*MCMC: Markov Chain Monte Carlo Algorithm*).

En este entorno, es usual que al utilizar este algoritmo y mediante la estadística bayesiana, se suponga que el estado inicial de la log-volatilidad sea $h_0 \sim N(m_0, C_0)$, donde m_0 y C_0 corresponden a la media y varianza elegidos *a priori* por el investigador. Por ejemplo, partiendo del modelo representado por las ecuaciones (2.18 y 2.19), podría

¹³Como se asume que la parte predecible del retorno es cero, aquí se supone que ψ es igual a cero.

asumirse como *prior* o volatilidad inicial a $(h_0|\mu, \phi, \tau^2) \sim N(\frac{\mu}{1-\phi}, \frac{\tau^2}{1-\phi^2})$ con $|\phi| < 1$.

Teniendo en cuenta, la condición inicial asumida por el investigador para h_0 , el modelo SV se completa al incorporar un supuesto sobre las distribuciones de cada uno de los parámetros a estimar dentro del modelo con el fin de obtener la función de distribución conjunta, es decir la función de distribución de todo el proceso estocástico en conjunto, en este caso simultáneamente la distribución del proceso dado por las ecuaciones 2.18 y 2.19. Por ejemplo, suponga que $\mu = 0$ entonces la conjugada entre los dos parámetros del modelo restante esta dada por: $p(\phi|\tau^2)p(\tau^2)$ donde $\phi|\tau^2 \sim N(b_0, \tau^2 B_0)$ y $\tau^2 \sim IG(c_0, d_0)$, donde (b_0, B_0, c_0, d_0) son parámetros conocidos para la media y la varianza de cada una de esas distribuciones. Algunas veces en este tipo de problemas se utilizan algunas distribuciones específicas porque tienen propiedades que ayudan la implementación numérica, tal es el caso de las denominadas *distribuciones conjugadas*. En el ejemplo anterior, ϕ se distribuye Normal y τ^2 se distribuye inversa gamma, en general cuando se tiene lo anterior la distribución conjunta resulta ser una función de distribución conocida, en este caso la *conjugada* es una distribución Normal-Inversa gamma. En otras palabras, la conjunta de $(\phi, \tau^2) \sim NIG(b_0, \tau^2 B_0, c_0, d_0)$ (Para demostración, ver Robert and Casella [2004] y Gamerman and Lopes [2006]).

Para determinar las características del proceso SV y cómo éstas replican los hechos estilizados de las series financieras, así como algunos aspectos sobre la estacionariedad del proceso, en particular es deseable evaluar el decaimiento exponencial de las correlaciones, a continuación se calculan los primeros momentos del proceso.

$$E[y_t] = E[\exp\{\frac{h_t}{2}\}\sigma_\varepsilon E(\varepsilon_t|h_t)] = 0 \quad (2.20)$$

De allí se aprecia que los retornos tienen media cero.

$$\begin{aligned}
 E[y_t^2] &= E[\exp\{h_t\}\sigma_\varepsilon^2 E(\varepsilon_t^2|h_t)] \\
 &= \sigma_\varepsilon^2 E[\exp\{h_t\}] \\
 &= \sigma_\varepsilon^2 E[\exp\{\mu + \phi h_{t-1} + \tau\eta_t\}] \\
 &= \sigma_\varepsilon^2 E[\exp\{\mu\}\exp\{\phi h_{t-1}\}\exp\{\tau\eta_t\}] \\
 &= \sigma_\varepsilon^2 E[\exp\{\phi(\mu + \phi h_{t-2} + \tau\eta_{t-1})\}]\exp\{\mu\}\exp\{\frac{\tau^2}{2}\} \\
 &= \sigma_\varepsilon^2 E[\exp\{\phi^2 h_{t-2}\}]\exp\{\mu\phi\}\exp\{\mu\}\exp\{\frac{\tau^2}{2}\}\exp\{\frac{\phi^2\tau^2}{2}\} \\
 &= \dots \\
 &= \sigma_\varepsilon^2 E[\exp\{\phi^k h_{t-k}\}]\exp\{\mu + \mu\phi + \mu\phi^2 + \dots + \mu\phi^k\} \\
 &\quad \exp\{\frac{\tau^2}{2} + \phi^2\frac{\tau^2}{2} + \phi^4\frac{\tau^2}{2} + \dots + \phi^{2(k-1)}\frac{\tau^2}{2}\} \\
 &\quad \text{donde } E[\exp\{\phi^k h_{t-k}\}] = 1 \text{ cuando } k \rightarrow \infty \\
 &= \sigma_\varepsilon^2 \exp\left\{\frac{\mu}{1-\phi} + \frac{\tau^2}{2(1-\phi^2)}\right\}
 \end{aligned} \tag{2.21}$$

La varianza de largo plazo de los retornos es constante y converge a

$$\sigma_\varepsilon^2 \exp\left\{\frac{\mu}{1-\phi} + \frac{\tau^2}{2(1-\phi^2)}\right\}.$$

$$E[y_t^3] = E[\exp\{\frac{3h_t}{2}\}\sigma_\varepsilon^3 E(\varepsilon_t^3|h_t)] = 0 \tag{2.22}$$

$$\begin{aligned}
 E[y_t^4] &= E[\exp\{2h_t\}\sigma_\varepsilon^4 E(\varepsilon_t^4|h_t)] \\
 &= 3\sigma_\varepsilon^4 E[\exp\{2h_t\}] \\
 &= 3\sigma_\varepsilon^4 E[\exp\{2(\mu + \phi h_{t-1} + \tau\eta_t)\}] \\
 &= 3\sigma_\varepsilon^4 E[\exp\{2\phi h_{t-1}\}]\exp\{2\mu\}\exp\{\frac{2^2\tau^2}{2}\} \\
 &= 3\sigma_\varepsilon^4 E[\exp\{2\phi(\mu + \phi h_{t-2} + \tau\eta_{t-1})\}]\exp\{2\mu\}\exp\{2\tau^2\} \\
 &= 3\sigma_\varepsilon^4 E[\exp\{2\phi^2 h_{t-2}\}]\exp\{2\mu\}\exp\{2\mu\phi\}\exp\{2\tau^2\}\exp\{2\tau^2\phi^2\} \\
 &= \dots \\
 &= 3\sigma_\varepsilon^4 E[\exp\{2\phi^k h_{t-k}\}]\exp\{2\mu(1 + \phi + \phi^2 + \dots + \phi^k)\} \\
 &\quad \exp\{2\tau^2(1 + \phi^2 + \phi^4 + \dots + \phi^{2(k-1)})\} \\
 &= \dots \\
 &= 3\sigma_\varepsilon^4 \exp\left\{2\left(\frac{\mu}{1-\phi} + \frac{\tau^2}{1-\phi^2}\right)\right\}
 \end{aligned} \tag{2.23}$$

En esta expresión¹⁴ se aprecia que este momento también es constante y converge a $3\sigma_\varepsilon^4 \exp \left\{ 2 \left(\frac{\mu}{1-\phi} + \frac{\tau^2}{1-\phi^2} \right) \right\}$.

Teniendo en cuenta todo lo anterior, se puede calcular la kurtosis de los retornos de la siguiente manera:

$$\begin{aligned} \kappa &= \frac{E[y_t^4]}{(E[y_t^2])^2} \\ &= 3 \exp \left\{ \frac{\tau^2}{1-\phi^2} \right\} \end{aligned} \quad (2.24)$$

Se puede apreciar como la kurtosis es mayor a 3, lo cual evidencia que el modelo SV replica el hecho estilizado de colas pesadas en las series financieras. Por último, el modelo SV también replica la correlación de los retornos al cuadrado que se observa en cualquier activo financiero. Para mostrar esto, fijando $\sigma_t^2 = \exp\{h_t\}$ y siguiendo a Jacquier et al. [1994] se expresan las correlaciones de y_t^2 como sigue:

$$\begin{aligned} cov(y_t^2, y_{t-k}^2) &= cov(\sigma_t^2, \sigma_{t-k}^2) \\ &= \left(\exp \left\{ \frac{\tau^2 \phi^k}{1-\phi^2} \right\} - 1 \right) (E(\sigma_t^2))^2 \end{aligned} \quad (2.25)$$

Dado 2.21 y 2.23, la varianza de los retornos al cuadrado es:

$$Var(y_t^2) = (E[\sigma_t^2])^2 \left(3 \exp \left\{ \frac{\tau^2}{1-\phi^2} \right\} - 1 \right) \quad (2.26)$$

Tomando 2.25 y 2.26, la función de auto-correlación simple (ρ_k), donde k hace referencia a los rezagos que se tienen en cuenta en el cálculo de la correlación, es:

¹⁴La kurtosis de una variable se define como $\kappa = \frac{E(\varepsilon_t^4)}{(var(\varepsilon_t))^2}$ y se sabe que para una variable normal $\kappa = 3$. Como $\varepsilon_t \stackrel{iid}{\sim} N(0,1)$, entonces $E(\varepsilon_t^4) = 3$ ya que $(var(\varepsilon_t))^2 = 1$.

$$\rho_k = \frac{\left(\exp \left\{ \frac{\tau^2 \phi^k}{1 - \phi^2} \right\} - 1 \right)}{\left(3 \exp \left\{ \frac{\tau^2}{1 - \phi^2} \right\} - 1 \right)} \approx \frac{\left(\exp \left\{ \frac{\tau^2}{1 - \phi^2} \right\} - 1 \right)}{\left(3 \exp \left\{ \frac{\tau^2}{1 - \phi^2} \right\} - 1 \right)} \phi^k \quad (2.27)$$

Como $|\phi| < 1$, entonces ϕ^k tiende a cero cuando k tiende a infinito, de tal forma que la función de auto-correlación simple (ρ_k) decrece exponencialmente.

2.3. Pronóstico

Luego de realizar el proceso de estimación, el cual se explicará en el siguiente capítulo con mayor detalle, una de las principales utilidades de estas metodologías es que permite implementar pronósticos de las series financieras. En otras palabras, no sólo es importante tener una estimación de los parámetros de un modelo que se ajuste bien a los datos o la estimación histórica de variables que no son observables como la volatilidad de una serie financiera, también es relevante efectuar proyecciones de las variables a futuro. En particular, en el caso de la tasa de cambio se puede apreciar que para cualquier inversionista es de interés aproximarse a una medición de ella en los siguientes días o semanas, puesto que con base en eso podría tomar mejores decisiones de inversión.

En términos estadísticos, el pronóstico de la volatilidad es el valor esperado condicionado de la varianza de la serie financiera dada la última información disponible y relevante. Como se discutió en los apartados previos, el GARCH es la generalización del modelo ARCH, por lo que a continuación sólo se describe el pronóstico de la varianza condicional de dicho modelo. Tomando como ejemplo un GARCH(1,1), como el referenciado en 2.12, el pronóstico para la varianza condicional un paso adelante o el valor que se espera para mañana de la volatilidad dada la información que se tiene disponible hoy está dado por:

$$E[h_{t+1} | \Psi_t] = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 h_t \quad (2.28)$$

donde ε_t^2 y h_t son conocidos en el periodo t (están dentro de Ψ_t). De esta forma, la proyección se hace sencilla de calcular ya que del proceso de estimación se tienen los valores de los coeficientes del modelo (α_0, α_1 y β_1).

Hacia un horizonte de tiempo más amplio, el pronóstico para j pasos adelante se obtiene de la siguiente manera: i) se debe usar el hecho de que $\varepsilon_t^2 = h_t v_t^2$, por lo que $\varepsilon_{t+j}^2 = h_{t+j} v_{t+j}^2$ y ii) si se actualizan las ecuaciones para j períodos adelante y se toman las esperanzas condicionales a ambos lados se tiene que $E[\varepsilon_{t+j}^2 | \Psi_t] = E[h_{t+j} v_{t+j}^2 | \Psi_t] = E[h_{t+j} | \Psi_t]$ ya que v_{t+j} es independiente de h_{t+j} y $E[v_{t+j}^2 | \Psi_t] = 1$. Así, el pronóstico para la volatilidad del activo financiero para j pasos adelante se obtiene solucionando:

$$\begin{aligned} E[h_{t+j} | \Psi_t] &= \alpha_0 + \alpha_1 E[\varepsilon_{t+j-1}^2 | \Psi_t] + \beta_1 E[h_{t+j-1} | \Psi_t] \\ \text{pero } E[\varepsilon_{t+j-1}^2 | \Psi_t] &= E[h_{t+j-1} | \Psi_t] \\ E[h_{t+j} | \Psi_t] &= \alpha_0 + (\alpha_1 + \beta_1) E[h_{t+j-1} | \Psi_t] \end{aligned} \quad (2.29)$$

Como el objetivo es expresar el pronóstico en términos de la última información disponible, se puede utilizar (2.28) y (2.29) para realizar las proyecciones desde un período adelante hasta j períodos adelante, utilizando todos los valores subsecuentes de la varianza condicional ($h_{t+1}, h_{t+2}, \dots, h_{t+j-1}$). Así, mediante un proceso iterativo se obtiene:

$$E[h_{t+j} | \Psi_t] = \alpha_0 [1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2 + \dots + (\alpha_1 + \beta_1)^{j-1}] + (\alpha_1 + \beta_1)^j h_t \quad (2.30)$$

donde h_t ya es conocido. Si el modelo es estacionario (se tiene que $\alpha_1 + \beta_1 < 1$), los pronósticos de la varianza condicional (h_{t+j}) convergen a su valor de largo plazo ($E[h_t] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$).

Por su parte, para el modelo de volatilidad estocástica (SV) se tiene que el pronóstico para la log-volatilidad viene dado por la actualización de la ecuación (2.19),¹⁵ la cual se asume similar a un proceso de series de tiempo tipo AR(1). En esta tesis se asume un modelo de volatilidad estocástica de orden uno (SV(1)), tomando como base a Taylor [1982] y Lopes and Tsay [2011]. Así, para un paso adelante la proyección puntual es:

¹⁵Se asume siguiendo a Taylor [2005] que la ecuación de volatilidad es independiente de la media, y que la media no tiene componente predecible. En caso contrario, el pronóstico de acuerdo con Tsay [2010] viene dado por un proceso de simulación de Monte Carlo similar al que se explicará en los métodos de estimación.

$$\begin{aligned}
E[h_{t+1}|\Psi_t] &= E[\mu + \phi h_t + \tau\eta_{t+1}|\Psi_t] \\
E[h_{t+1}|\Psi_t] &= \mu + \phi h_t
\end{aligned}
\tag{2.31}$$

donde h_t ya es conocido y se encuentra en el conjunto de información (Ψ_t) y las perturbaciones que no se encuentran en el conjunto de información toman el valor medio de su distribución, por lo que $E[\tau\eta_{t+1}|\Psi_t] = 0$. Para más pasos adelante se toma el siguiente proceso iterativo:

$$\begin{aligned}
E[h_{t+2}|\Psi_t] &= E[\mu + \phi h_{t+1} + \tau\eta_{t+2}|\Psi_t] \\
E[h_{t+2}|\Psi_t] &= E[\mu + \phi(\mu + \phi h_t) + \tau\eta_{t+2}|\Psi_t] \\
E[h_{t+2}|\Psi_t] &= \mu + \phi\mu + \phi^2 h_t \\
&= \dots \\
E[h_{t+j}|\Psi_t] &= \mu + \phi\mu + \dots + \phi^{(j-1)}\mu + \phi^j h_t \\
&\text{cuando } j \rightarrow \infty \text{ entonces} \\
E[h_{t+j}|\Psi_t] &= \frac{\mu}{1 - \phi}
\end{aligned}
\tag{2.32}$$

Dado que el modelo se asume es estacionario ($|\phi| < 1$), el pronóstico de la log-volatilidad (h_t) converge a su valor de largo plazo $\left(\frac{\mu}{1 - \phi}\right)$.

Capítulo 3

Métodos de estimación

El estudio de las series de tiempo financieras revelan que cambios en la volatilidad ocurren a lo largo del tiempo para muchas clases de activos financieros, tales como las tasas de cambio, las acciones, los bonos, entre otros. En el capítulo 2 se mostraron modelos que son varianza-dependientes, que en la literatura son agrupados en dos tipos de modelos: unos cuya dinámica la dirigen las observaciones disponibles (*observation-driven models*) y otros que se rigen principalmente por los parámetros estimados (*parameter-driven models*). En el marco de trabajo de los primeros, se encuentra la familia de modelos de heteroscedasticidad condicional (GARCH) propuestos por Engle [1982] y Bollerslev [1986], en donde la varianza condicional es tomada como una función de los cuadrados de observaciones pasadas. Como una alternativa a esto, Taylor [1982] propuso el modelo de volatilidad estocástica (SV), en el cual la varianza condicional sigue un proceso estocástico latente. Este último se encuentra dentro del segundo grupo de modelos.

Este capítulo ofrece al lector una pequeña reseña de los métodos disponibles para la estimación de este tipo de modelos, para luego concentrarse en aquellos que se aplicaron en el desarrollo de este trabajo de tesis. En particular, se describen: i) la metodología de máxima verosimilitud que frecuentemente se utiliza en la implementación de los modelos GARCH¹, y ii) metodologías de inferencia bayesiana que permiten estimar los parámetros del modelo SV. Este apartado se basa en las lecturas de Doucet et al. [2001], Robert and Casella [2004], Gamerman and Lopes [2006] y Creal [2012].

3.1. Una Breve Reseña

El método de estimación de máxima verosimilitud (MLE) es usualmente utilizado para estimar los parámetros en los modelos tipo GARCH, pero éste no puede ser directa-

¹Esta es una técnica ampliamente utilizada debido a su sencillez y a la oportunidad en uso computacional ya que se encuentra desarrollada y programada en la mayoría de paquetes econométricos

mente aplicado al ajuste de los parámetros en los modelos tipo SV debido a la existencia de un proceso estocástico latente. Adicionalmente, es importante señalar que el modelo SV incorpora mayores no linealidades que complican aún más el procedimiento de maximización respecto a lo presentado en el GARCH. De hecho, mientras que se pueden encontrar una forma analítica para los modelos GARCH, en los modelos SV es casi imposible y se hace necesario utilizar procesos computacionales de simulación.

Para estimar eficientemente los parámetros y la volatilidad en el modelo SV, la literatura econométrica ha propuesto varios métodos. Entre las primeras metodologías propuestas se encuentran el método generalizado de momentos (GMM) introducido por Hansen [1982] (Ver también Melino and Turnbull [1990]) y el método de quasi-máxima verosimilitud (QML) (Harvey et al. [1994]), los cuales presentaron inconvenientes como la no trivialidad existente en la elección de las condiciones de los momentos en el GMM y el intensivo costo computacional presentado por el QML. Luego, se incorporaron métodos más elaborados como el de máxima verosimilitud simulada (SML) (Danielsson and Richard [1993]), el método de momentos eficiente (EMM) (Gallant et al. [1997]) y las cadenas de Markov Monte Carlo (*Markov Chain Monte Carlo (MCMC)*) (Jacquier et al. [1994]).

El método SML es una aproximación a la MLE, excepto que la verosimilitud no es directamente computada. De hecho, ésta es estimada mediante simulaciones de Monte Carlo. En otras palabras, SML busca los parámetros que maximicen la estimación de verosimilitud estimada por Monte Carlo. Para mejorar el ajuste de la verosimilitud, una forma de reducir varianza es a través de un procedimiento denominado muestreo de importancia (*Importance Sampling-(IS)*) inventado por Marshall [1956] y usado por primera vez en econometría por Kloek and van Dijk [1978]. Allí, se busca hallar dentro de las simulaciones variables aleatorias que permitan aproximar adecuadamente la distribución conjunta objetivo. Una aplicación en economía interesante se encuentra en Danielsson [1994], quien mostró empíricamente que SML era tan eficiente como MCMC para muestras finitas al realizar dichos procedimiento sobre el índice de acciones de Estados Unidos (el *standard and poor's*).

Luego de esos trabajos, la literatura existente ha abordado el problema de muchas maneras tratando de encontrar ganancias en los métodos de simulación y su aproximación a las distribuciones objetivo. Algunos de ellos son: el método de muestro de importancia eficiente (*efficient importance sampling (EIS)*) (Ver Richard and Zhang [2007], de Jong et al. [2009]), muestreo de importancia usando filtros de Kalman (Ver Durbin and Koopman [1997] y Shephard and Pitt [1997]) y muestreo de importancia secuencial (*sequential importance sampling (SIS)*) (Ver Hammersley and Morton [1954] y Handschin [1970]) con sus respectivas derivaciones tales como muestreo de importancia secuencial con reemplazamiento (*sequential importance sampling with resampling (SISR)*), entre otros. Estos últimos enmarcados en el campo de trabajo de los denominados *filtros de partículas*. En particular, en esta tesis se implementaron cuatro métodos de inferencia

bayesiana utilizando MCMC y un método de Monte Carlo Secuencial para la estimación del modelo de volatilidad estocástica, mientras que se realizó la tradicional estimación por máxima verosimilitud para el ajuste del modelo de heteroscedasticidad GARCH.

3.2. Método de Máxima Verosimilitud

El método de máxima verosimilitud consiste en elegir un estimador asintóticamente eficiente para un parámetro o conjunto de parámetros. En particular, en este trabajo de tesis se utilizó este método para estimar un modelo GARCH para la volatilidad de la tasa de cambio. Para ilustrarlo se debe formar la función de verosimilitud, lo cual se hace expresando la función de densidad de probabilidad conjunta de los datos dados en términos de un conjunto de parámetros ($f(y_1, y_2, \dots, y_T | \Theta)$). Allí, f denota la función de densidad de probabilidad conjunta, $\{y_1, y_2, \dots, y_T\}$ son los retornos del activo financiero desde el período 1 hasta T y Θ es el vector de parámetros que se desean estimar.

Si los retornos fueran independientes ó si el mercado fuera eficiente entonces la función de densidad conjunta podría ser hallada como el producto de las densidades marginales de cada una de los retornos. En caso de no cumplirse lo anterior, una aproximación a la función de densidad conjunta podría darse a través del producto de las funciones de densidad condicionales de cada uno de los retornos. Es decir, $f(y_1, y_2, \dots, y_T | \Theta) = f(y_T | y_1, y_2, \dots, y_{T-1}, \Theta) f(y_{T-1} | y_1, y_2, \dots, y_{T-2}, \Theta) \dots f(y_1 | \Theta)$.

Asumiendo que el proceso generador de datos de los retornos está dado por la ecuación (2.12) y como la perturbación es gaussiana o normal², entonces para calcular la función de verosimilitud conjunta se parte de la función de distribución de la t-ésima observación del retorno:

$$f(y_t | y_1, y_2, \dots, y_{t-1}, \Theta) = \frac{1}{\sqrt{h_t} \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{\varepsilon_t^2}{h_t} \right\} \quad (3.1)$$

Luego, tomando logaritmos, donde (l_t) es el logaritmo natural aplicado a la función de densidad de la t-ésima observación del retorno:

$$l_t = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(h_t) - \frac{1}{2} \frac{\varepsilon_t^2}{h_t} \quad (3.2)$$

²Recuerde que $y_t = \varepsilon_t$ y $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

Para encontrar la función de verosimilitud (la maximización de la distribución conjunta) de todos los retornos observados (disponibles), se toma el logaritmo de la función de densidad para la t -ésima realización, denotado por l_t , y se hace el mismo procedimiento anterior para toda la muestra de observaciones (ósea desde 1 hasta T):

$$L = \frac{1}{T} \sum_{t=1}^T l_t \quad (3.3)$$

Así, el método de máxima verosimilitud busca la maximización de L , donde L es el logaritmo de la función de densidad conjunta, respecto al vector de parámetros desconocidos Θ . El proceso de optimización se da sobre la función de densidad conjunta, cuya representación se formuló a través de L . Los estimadores de máxima verosimilitud se encuentran resolviendo el sistema de ecuaciones generado por las condiciones de primer orden (Ver demostración en Casas and Cepeda [2008]):

$$\frac{\partial l_t}{\partial \Theta} = \frac{1}{2h_t} \frac{\partial h_t}{\partial \Theta} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \quad (3.4)$$

Como la ecuación (2.12), y en general cualquier modelo GARCH, incorpora no linealidades, se hace imposible resolver las condiciones de primer orden analíticamente. En otras palabras, en (3.4) se reemplaza h_t por $\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$, de acuerdo al ejemplo del GARCH(1,1) referenciado en (2.12). Por lo tanto, el vector de parámetros está dado por $\Theta = (\alpha_0, \alpha_1, \beta_1)'$. El proceso de optimización se realiza sobre cada uno de los coeficientes desconocidos anteriores y dadas las no linealidades generadas en (3.4), se hace indispensable el uso de algoritmos de optimización en aplicaciones computacionales (Ver Hamilton [1994]).

El lector puede encontrar en Hamilton [1994] y Casas and Cepeda [2008] expresiones para el procedimiento expuesto arriba para modelos ARCH y GARCH que incorporan más parámetros en la ecuación de media y de varianza. Adicionalmente, en dichas referencias el lector puede encontrar una explicación detallada sobre los procedimientos de optimización utilizados para la estimación de los parámetros y de sus varianzas, tales como el algoritmo de Fisher-Scoring, Newton-Raphson, Davidon-Fletcher-Powell, entre otros.

3.3. Inferencia Bayesiana

Debe señalarse que en el caso de los modelos de heteroscedasticidad condicional, finalmente la varianza condicional está en función de las innovaciones pasadas y de los rezagos de la propia volatilidad. En ese caso para el momento t , la varianza está dada por valores de la perturbación y de la volatilidad en $(t - 1)$, los cuales ya son variables conocidas. Luego, utilizando los parámetros que ya han sido estimados, se puede calcular una estimación puntual para la volatilidad en t . Asimismo, nótese que en (3.4), l_t está en función de ε_t y h_t ; los cuales pueden ser expresados a su vez en términos de y_t , por lo que finalmente el proceso de optimización se realiza sobre un conjunto de variables que son observables (y_t).

Sin embargo, en diversas disciplinas se efectúan modelaciones que involucran el uso de un conjunto de observaciones que no son conocidas. Un ejemplo particular, son los modelos de volatilidad estocástica (Ver 2.18 y 2.19), en donde el análisis de un activo financiero se efectúa mediante la observación de los retornos financieros $\{y_t\}_{t=1}^T$ y de la volatilidad que ellos presenten (h_t), ésta última depende de una perturbación aleatoria en cada momento del tiempo t . Esto hace que h_t se convierta en una variable latente y no pueda ser fácilmente determinada como en el caso de los modelos GARCH. En sistemas como el expresado en el modelo de volatilidad estocástica (SV), se tienen dos tipos de variables: i) observables o conocidas (y_t) y ii) no observables o latentes (h_t).

La solución de este tipo de problemas se ha abordado en la teoría desde un punto de vista bayesiano (Ver Apéndice A.2), en donde el conocimiento disponible del sistema en estudio puede ser usado para formular distribuciones *a priori* de la evolución de las series no observables, así como las funciones de verosimilitud del sistema completo. Similar a lo que se buscaba con el método de máxima verosimilitud, el proceso de optimización de los parámetros se realiza sobre la función de densidad conjunta que ahora necesita ser planteada en términos de las funciones de densidad de probabilidad posterior de la variable *latente* (h_t). Esta última es denotada como $p(h_{0:n}|y_{1:n}, \Theta)$ donde n es un sub-índice que hace referencia a momentos del tiempo y Θ son los parámetros del modelo³.

En el contexto bayesiano, la inferencia sobre $(\{h_t\}_{t=1}^T)$ dado un conjunto de observaciones $(\{y_t\}_{t=1}^T)$ se pueden expresar a través de la distribución *posterior* para cualquier momento del tiempo t y bajo el supuesto de que los parámetros Θ son conocidos

³En el caso del modelo descrito por las ecuaciones 2.18 y 2.19, $\Theta = (\mu, \phi, \tau)'$. La notación $h_{0:n}$ y $y_{1:n}$ indica que h_t tiene observaciones desde el momento 0 hasta n y y_t tiene observaciones desde el momento 1 hasta n

$$p(h_{0:t}|y_{1:t}; \Theta) = \frac{p(y_{1:t}|h_{0:t}; \Theta)p(h_{0:t}; \Theta)}{p(y_{1:t}; \Theta)} \quad (3.5)$$

$$\text{donde } p(y_{1:t}; \Theta) = \int p(h_{0:t}, y_{1:t}; \Theta) dh_{0:t}$$

En aplicaciones, donde se dispone de observaciones en forma secuencial, se puede encontrar una representación recursiva para la densidad *posterior* (3.5) mediante el Teorema de Bayes (Ver Apéndice A.2) de la siguiente forma:

$$\begin{aligned} p(h_{0:t}|y_{1:t}; \Theta) &= p(y_{1:t}|h_{0:t}; \Theta) \frac{p(h_{0:t}; \Theta)}{p(y_{1:t}; \Theta)} \\ &= p(y_{1:(t-1)}|h_{0:t}; \Theta) \frac{p(h_{0:t}; \Theta)p(y_t|h_t; \Theta)}{p(y_{1:(t-1)}; \Theta)p(y_t|y_{1:(t-1)}; \Theta)} \\ &= p(h_{0:t}|y_{1:(t-1)}; \Theta) \frac{p(y_t|h_t; \Theta)}{p(y_t|y_{1:(t-1)}; \Theta)} \\ &= p(h_{0:(t-1)}|y_{1:(t-1)}; \Theta) \frac{p(h_t|h_{t-1}; \Theta)p(y_t|h_t; \Theta)}{p(y_t|y_{1:(t-1)}; \Theta)} \end{aligned} \quad (3.6)$$

$$\text{donde } p(y_t|y_{1:(t-1)}; \Theta) = \int p(y_t|h_t; \Theta)p(h_t|h_{t-1}; \Theta)p(h_{t-1}|y_{1:(t-1)}; \Theta) dh_{(t-1):t}$$

Como resultado, la densidad *marginal* ($p(h_t|y_{1:t}; \Theta)$) es derivada mediante dos etapas, predicción y actualización. Asimismo, se observa que en (3.6) el estado actual h_t depende del estado pasado y no de todo el camino recorrido por éste. Teniendo en cuenta esto, la primera fase consiste en estimar $p(h_{0:t}|y_{1:(t-1)}; \Theta)$ utilizando la ecuación de Chapman- Kolmogorov; dado que la densidad de la evolución del estado $p(h_t|h_{t-1}; \Theta)$ y el filtrado sobre $p(h_{0:(t-1)}|y_{1:(t-1)}; \Theta)$ se efectúan a través de un proceso generador de datos determinado⁴. Así, la predicción se representa⁵

$$p(h_t|y_{1:(t-1)}; \Theta) = \int p(h_t|h_{t-1}; \Theta)p(h_{t-1}|y_{1:(t-1)}; \Theta) dh_{t-1} \quad (3.7)$$

La segunda fase, la actualización consiste en incorporar la información recibida para mejorar la estimación de h_t realizada en la fase anterior. En otras palabras, como las

⁴Un ejemplo particular es el sistema 3.18.

⁵En esta expresión se asume que h_t está definido por un proceso de Markov de primer orden, por lo que $p(h_t|h_{0:(t-1)}, y_{1:(t-1)}; \Theta) = p(h_t|h_{t-1}; \Theta)$.

nuevas realizaciones de y sirven para agregar información relevante sobre la evolución del estado. Mediante el Teorema de Bayes se tiene:

$$\begin{aligned}
 p(h_t|y_{1:t}; \Theta) &= p(y_{1:t}|h_t; \Theta) \frac{p(h_t; \Theta)}{p(y_{1:t}; \Theta)} \\
 &= p(y_t|h_t; \Theta) p(y_{1:(t-1)}|h_t; \Theta) \frac{p(h_t; \Theta)}{p(y_t|y_{1:(t-1)}; \Theta) p(y_{1:(t-1)}; \Theta)} \quad (3.8) \\
 &= p(y_t|h_t; \Theta) \frac{p(h_t|y_{1:(t-1)}; \Theta)}{p(y_t|y_{1:(t-1)}; \Theta)}
 \end{aligned}$$

Las dificultades de la anterior aproximación surgen cuando en muchos modelos de estado-espacio la integral (3.7) y la densidad (3.8) no pueden ser calculadas analíticamente. Hay varios casos conocidos en los cuales es posible resolver las iteraciones analíticamente. El primer caso, el más usual, es cuando las funciones que rigen el comportamiento de y_t y h_t son lineales y las perturbaciones son gaussianas⁶. Bajo las anteriores condiciones el problema puede ser resuelto por medio del filtro de Kalman (Kalman [1960]). Otro caso es cuando la variable de estado h_t toma un número discreto de valores y la recursión puede resolverse por el filtro HMM (*Hidden Markov Model*), el cual soluciona sistemas de estado-espacio no gaussianos y no lineales (Baum and Petrie [1966] y Baum et al. [1970]).

Sin embargo, cuando el problema no puede ser resuelto como en los dos casos mencionados, se deben realizar aproximaciones sub-óptimas mediante la simulación de las integrales (3.7) y la densidad (3.8). Es así como surgieron métodos de simulación de Monte Carlo que permitieron implementar aproximaciones numéricas a la solución del problema (Ver Robert and Casella [2004] y Gamerman and Lopes [2006]). Más adelante, se exponen los algoritmos más usuales y los utilizados en este trabajo de tesis dentro de estos métodos.

3.4. Filtro de Kalman

El filtro de Kalman (Ver Kalman [1960]) es un enfoque recursivo usado para estimar variables de estado en un sistema que está representado por un modelo dinámico y por un conjunto de observaciones disponibles⁷. Este asume que en cada instante de tiempo, la función de densidad posterior del problema es gaussiana. En la práctica, al utilizar

⁶Por ejemplo, en el sistema (3.18) esto implica que f_t y g_t son funciones lineales. Además, ε_t y η_t se distribuyen Normal.

⁷Bajo las condiciones de perturbaciones gaussianas y linealidad en el modelo, el filtro de Kalman encuentra una solución óptima.

esta herramienta se habla a menudo de filtrado de las series a partir de la solución de un sistema de estado-espacio, pero éste también puede ser visto como un problema de inferencia bayesiana. Para mostrar lo anterior, se puede partir del siguiente sistema:

$$\begin{aligned} y_t &= H_t h_t + v_t && \text{(ecuación de observación-medida)} \\ h_t &= F_{t-1} h_{t-1} + \omega_{t-1} && \text{(ecuación de evolución-estado)} \end{aligned} \quad (3.9)$$

donde h_t es el estado del sistema en el momento t y está asociado a variables que no son observables, y_t son las observaciones conocidas en el momento t . F_{t-1} y H_t son matrices conocidas y $v_t \sim N(0, R_t)$ y $\omega_t \sim N(0, Q_t)$ son mutuamente independientes y están asociados son los términos de ruido en las ecuaciones de observación y de estado. R_t y Q_t son las varianzas de v_t y ω_t , respectivamente. En el enfoque bayesiano, (3.9) puede escribirse de la siguiente forma:

$$\begin{aligned} y_t &\sim N(H_t h_t, R_t) \\ h_t &\sim N(F_{t-1} h_{t-1}, Q_{t-1}) \end{aligned} \quad (3.10)$$

Utilizando (3.7 y 3.8) sobre el sistema (3.10) el algoritmo del filtro de Kalman puede ser formulado como sigue (Ver Kalman [1960] y Chui and Chen [1991])

$$\begin{aligned} p(h_t | y_{1:(t-1)}) &= N_{h_t}(\hat{h}_{t|t-1}, P_{t|t-1}) \\ p(h_t | y_{1:t}) &= N_{h_t}(\hat{h}_{t|t}, P_{t|t}) \end{aligned} \quad (3.11)$$

Siguiendo la notación Chui and Chen [1991], las medias y las covarianzas del sistema (3.11) están expresadas⁸

$$\begin{aligned} \hat{h}_{t|t-1} &= F_{t-1} \hat{h}_{t-1|t-1} \\ P_{t|t-1} &= Q_{t-1} + F_{t-1} P_{t-1|t-1} F_{t-1}' \\ \hat{h}_{t|t} &= \hat{h}_{t|t-1} + K_t (y_t - H_t \hat{h}_{t|t-1}) \\ P_{t|t} &= P_{t|t-1} - K_t S_t K_t' \end{aligned} \quad (3.12)$$

donde $S_t = H_t P_{t|t-1} H_t' + R_t$ corresponde a la covarianza del términos de correctivo $(y_t - H_t \hat{h}_{t|t-1})$ y $K_t = P_{t|t-1} H_t' S_t^{-1}$ es la ganancia del filtro. El filtro de Kalman

⁸También se sugiere al lector Anderson and Moore [1979].

calcula recursivamente la media y covarianza de la *posterior* ($p(h_t|y_{1:t})$), lo cual bajo el supuesto de Gaussianidad, es una condición suficiente para conocer completamente la distribución (Ver demostración Kalman [1960] y Hamilton [1994]). Con ello, el algoritmo encuentra la solución óptima para el problema de estimación del sistema (3.9). No obstante, la *posterior* no es necesariamente Gaussiana en la práctica por lo que en esos casos el filtro deja de ser óptimo. Algo similar ocurre cuando el sistema no cumple con el supuesto de linealidad (Hamilton [1994]). En el contexto de los modelos de volatilidad estocástica no se cuenta con la propiedad de linealidad, por lo que la utilización de este algoritmo dificulta encontrar una solución óptima.

3.5. Simulaciones de Monte Carlo (MCMC)

Una alternativa para solucionar el sistema (3.9), cuando no se logra contar con todos los supuestos requeridos por el filtro de Kalman (linealidad, principalmente), son los métodos desarrollados en la teoría de simulaciones de Monte Carlo. En ellos la solución, sub-óptima, se encuentra mediante algoritmos de iteración que aproximan las integrales y las densidades de probabilidad del problema en cuestión.

Teniendo en cuenta la definición (A.11), el problema de inferencia que se plantea en este apartado utiliza las cadenas de Markov para simular un *proceso de Markov* sobre un espacio generado por los parámetros Θ , el cual tiene una distribución $P(\theta|\mathbf{Y})$, la cual se asume es conocida antes de iniciar la simulación. θ es un vector de parámetros ($\theta \in \Theta$) y \mathbf{Y} son los datos.

De esta manera, la clave en las simulaciones de Monte Carlo radica en crear un *proceso de Markov*, cuya distribución de transición estacionaria se especifica a través de $P(\theta|\mathbf{Y})$, y luego, se corre un número suficientemente grande de simulaciones con las que la distribución de los valores actuales del proceso se haga lo más cercana posible a la distribución de transición estacionaria. En general, los métodos que usan este tipo de simulaciones para obtener la distribución $P(\theta|\mathbf{Y})$ son denominados métodos MCMC (*Markov Chain Monte Carlo*). A continuación se muestran algunos de los algoritmos más populares dentro de MCMC, algunos de ellos son utilizados en este trabajo de tesis.

3.5.1. Algoritmo de muestreo Gibbs

El algoritmo de muestreo de Gibbs (*Gibbs sampling*) de (Geman and Geman [1984], Gelfand and Smith [1990]) es el más popular de los métodos de MCMC. Para introducir la idea del muestreo de Gibbs enseguida se expone el funcionamiento de dicho procedimiento. Para comenzar, se supone que el problema a tratar busca la estimación de tres

parámetros $(\Theta = (\theta_1, \theta_2, \theta_3))'$. Cabe señalar que la palabra parámetro en la descripción que se va a realizar en este apartado es usada en un sentido muy general, más allá de un simple coeficiente. Por ejemplo, una observación faltante puede ser considerada como un parámetro dentro del campo de MCMC. De igual forma, un conjunto de variables no observables, tal como la volatilidad de un activo en cada momento del tiempo, también puede ser considerada como una agrupación de parámetros de tamaño T , donde T es el tamaño de muestra.

Suponga que se tienen tres parámetros no conocidos $(\Theta = (\theta_1, \theta_2, \theta_3))'$, una colección de datos disponibles \mathbf{Y} y una estructura o formulación de un modelo que relaciona los parámetros con los datos M . El objetivo es estimar los parámetros de la manera más óptima posible, lo cual puede hacerse mediante el ajuste e inferencia del modelo M y la maximización de la función de verosimilitud. Esta última se obtiene a través de la función de densidad conjunta $(f(\theta_1, \theta_2, \theta_3, \mathbf{Y}, M))$, que en adelante es referenciada en el texto como *la conjunta* de los parámetros. En la práctica, muchas veces no se cuenta con esta última, pero si se dispone de las funciones de densidad condicionadas de cada uno de los parámetros. En este ejemplo, se asume que las siguientes tres funciones de densidad condicionadas son conocidas:

$$f_1(\theta_1|\theta_2, \theta_3, \mathbf{Y}, M), \quad f_2(\theta_2|\theta_3, \theta_1, \mathbf{Y}, M), \quad f_3(\theta_3|\theta_1, \theta_2, \mathbf{Y}, M) \quad (3.13)$$

donde $f_i(\theta_i|\theta_{l \neq i}, \mathbf{Y}, M)$ denota la función de densidad condicional de parámetros θ_i dado los datos, el modelo y los otros dos parámetros. En adelante en el texto se nombra a f_i como *la condicional* del parámetro i . Empíricamente, no se necesita conocer las formas exactas de las distribuciones condicionales, de hecho, se pueden generar selecciones de números o realizaciones aleatorias para cada de las tres funciones f_i . El algoritmo funciona de la siguiente manera:

Algoritmo de Muestreo de Gibbs (Gibbs Sampling)

1. Se parte de una condición inicial para los parámetros, es decir $\theta_{1,0}, \theta_{2,0}, \theta_{3,0}$ son conocidos. Donde $\theta_{i,j}$ indican i el número del parámetro y j el número de la simulación o iteración.
2. Se genera una muestra aleatoria de realizaciones de $f_1(\theta_1|\theta_{2,0}, \theta_{3,0}, \mathbf{Y}, M)$. De allí, se obtiene una realización aleatoria para $\theta_{1,1}$, el parámetro 1 para la iteración 1.
3. Se genera una muestra aleatoria de realizaciones de $f_2(\theta_2|\theta_{3,0}, \theta_{1,1}, \mathbf{Y}, M)$. Se incorpora el parámetro $\theta_{1,1}$ que se obtuvo en el paso anterior. A partir de esto, se puede tener una realización aleatoria para $\theta_{2,1}$, el parámetro 2 para la iteración 1.
4. Se genera una muestra aleatoria de realizaciones de $f_3(\theta_3|\theta_{1,1}, \theta_{2,1}, \mathbf{Y}, M)$. Se incorporan los parámetros $\theta_{1,1}$ y $\theta_{2,1}$ que se obtuvieron en los dos pasos anteriores. A partir de esto, se puede tener una realización aleatoria para $\theta_{3,1}$, el parámetro 3 para la iteración 1.
5. Se repiten m veces los pasos 2 al 5. En cada iteración se actualizan f_1, f_2 y f_3 , condicionándolas sobre con los parámetros previamente hallados. Con ello se obtienen los parámetros $\theta_{1,m}, \theta_{2,m}, \theta_{3,m}$ de las m iteraciones de la simulación.

Como resultado del algoritmo anterior, se logra tener una secuencia de realizaciones aleatorias de los parámetros:

$$(\theta_{1,1}, \theta_{2,1}, \theta_{3,1}), (\theta_{1,2}, \theta_{2,2}, \theta_{3,2}), \dots, (\theta_{1,m}, \theta_{2,m}, \theta_{3,m}). \quad (3.14)$$

Bajo ciertas condiciones⁹, y para un número suficientemente alto de m , se puede mostrar que $(\theta_{1,m}, \theta_{2,m}, \theta_{3,m})$ es aproximadamente equivalente a una realización aleatoria proveniente de *la conjunta* de los tres parámetros. El teorema de convergencia incorporado en la teoría de cadenas de Markov (Ver Tierney [1994]) garantiza que dado un valor inicial arbitrario de los parámetros (*prior*), las iteraciones del algoritmo de Gibbs permite a los parámetros (θ) cambiar y vincularse completamente con el espacio (Θ).

Por último, es común cuando m es grande, que se descarten las primeras b observaciones con el fin de eliminar cualquier efecto que pueda generar sobre las simulaciones la adopción de la condición arbitraria inicial. A esto se le denomina *burn-in sample*. En este

⁹En general, esto hace referencia a que la función de distribución objetivo sea invariante en la cadena de Markov, es decir que la cadena sea ergódica y aperiódica. El lector puede consultar dichas propiedades en Robert and Casella [2004] y Gamerman and Lopes [2006].

trabajo de tesis se tomaron los valores de m y de b de acuerdo al criterio de investigaciones previas (Ver Lopes and Tsay [2011]), sin embargo si el lector quiere profundizar sobre el criterio de selección para b y para m puede consultar Gelman and Rubin [1992] y Raftery and Lewis [2010].

Entre las ventajas que se encuentran al utilizar este algoritmo, se pueden enunciar las siguientes:

1. Dado que se tienen secuencias de $(\theta_{1,m}, \theta_{2,m}, \theta_{3,m})$ que se aproximan a *la conjunta*, entonces éstas pueden ser utilizadas para hacer inferencia.

- a) Ejemplo 1: se pueden hallar estimadores insesgados de los parámetros y sus varianzas

$$\begin{aligned}\bar{\theta}_i &= \frac{1}{m-r} \sum_{j=r+1}^m \theta_{i,j} \\ \bar{\sigma}_i^2 &= \frac{1}{m-r-1} \sum_{j=r+1}^m (\theta_{i,j} - \bar{\theta}_i)^2\end{aligned}\tag{3.15}$$

donde $\bar{\theta}_i$ y $\bar{\sigma}_i^2$ corresponde a la media y varianza de cada uno de los parámetros θ_i .

- b) Ejemplo 2: se pueden hacer pruebas de hipótesis sobre los parámetros como probar la hipótesis nula $H_0: \theta_1 = \theta_2$ versus la alternativa $H_a: \theta_1 \neq \theta_2$.

2. Descompone un problema de estimación de altas dimensiones, en uno que esta compuesto por varios de baja dimensión. Esto mediante el uso de *las condicionales* de los parámetros. Es decir, el objetivo de estimar *la conjunta* con N parámetros (alta dimensionalidad) puede ser reducido y resuelto iterativamente mediante la utilización de N *condicionales* univariadas.

3.5.2. Algoritmo de muestreo Metropolis y Metropolis-Hasting

Estos son dos algoritmos alternativos que también permiten estimar los parámetros $\Theta = (\theta_i, i = 1, 2, 3)$ a partir de las simulaciones de *las condicionales*. El algoritmo de *Metropolis* genera una secuencia de realizaciones aleatorias de la distribución aproximada a aquellas distribuciones que convergen a $f(\Theta|\mathbf{Y})$ (Ver Metropolis et al. [1953]).

Algoritmo de Muestreo de Metropolis (Metropolis Algoritm)

1. Se genera un valor aleatorio inicial para Θ_0 tal que $f(\Theta_0|\mathbf{Y}) > 0$
2. Para $t = 1, 2, \dots$
 - a) Generar una muestra candidata para Θ , la cual se denota como Θ_* de una densidad *conocida* en la iteración t dados las previas realizaciones Θ_{t-1} . Denote a esta densidad *condicionada* como $q = J_t(\Theta_t|\Theta_{t-1})$, también denominada densidad incremental (*proposal distribution* (q)). Se asume que q es simétrica ($J_t(\Theta_i|\Theta_j) = J_t(\Theta_j|\Theta_i)$).
 - b) Calcule la proporción

$$r = \frac{f(\Theta_*|\mathbf{Y})}{f(\Theta_{t-1}|\mathbf{Y})} \quad (3.16)$$

- c) Fije

$$\Theta_t = \begin{cases} \Theta_* & \text{con probabilidad } \min(r, 1), \\ \Theta_{t-1} & \text{en otros casos.} \end{cases}$$

Bajo ciertas condiciones (Ver Gelman et al. [2003]), la secuencia Θ_t converge en distribución a $f(\Theta|\mathbf{Y})$. La implementación del algoritmo requiere: i) que sea posible calcular la proporción r para todos los Θ_* y Θ_{t-1} , ii) se puedan generar realizaciones Θ utilizando q y iii) generar una realización aleatoria proveniente de una distribución uniforme para determinar la aceptación o rechazo de Θ_* .

La regla de aceptación del algoritmo puede ser ilustrada como sigue:

1. Si el cambio entre Θ_{t-1} a Θ_* mejora el ajuste de la densidad condicional posterior, entonces se acepta Θ_* como Θ_t
2. Si el cambio entre Θ_{t-1} a Θ_* no favorece el ajuste de la densidad condicional posterior, entonces se fija $\Theta_t = \Theta_*$ con probabilidad r o $\Theta_t = \Theta_{t-1}$ con probabilidad $(1 - r)$.

Hasting [1970] generalizó lo anterior al incorporar dos mejoras: Primero, que q no tenía que ser simétrica y segundo, modificó la regla de salto (r). Ahora, se define r en lugar de la ecuación (3.16) como:

$$\begin{aligned}
r &= \frac{f(\Theta_*|\mathbf{Y})/J_t(\Theta_*|\Theta_{t-1})}{f(\Theta_{t-1}|\mathbf{Y})/J_t(\Theta_{t-1}|\Theta_*)} \\
&= \frac{f(\Theta_*|\mathbf{Y})J_t(\Theta_{t-1}|\Theta_*)}{f(\Theta_{t-1}|\mathbf{Y})J_t(\Theta_*|\Theta_{t-1})}
\end{aligned} \tag{3.17}$$

3.5.3. Algoritmo FFBS (Forward Filtering Backward Sampling)

West and Harrison [2007] propone para formular su algoritmo el siguiente modelo dinámico markoviano

$$\begin{aligned}
y_t &= f_t(h_t) + \varepsilon_t && \text{(ecuación de observación)} \\
h_t &= g_t(h_{t-1}) + \eta_t && \text{(ecuación de evolución estado)}
\end{aligned} \tag{3.18}$$

donde h_t es el estado del sistema en el momento t y está asociado a variables que no son observables, y_t son las observaciones conocidas en el momento t . f_t y g_t son funciones no lineales que se asumen son conocidas o asumidas por el investigador y rigen la evolución de las variables y_t y h_t , respectivamente. Por último, $\varepsilon_t \sim N(0, \sigma_\varepsilon)$ y $\eta_t \sim N(0, \sigma_\eta)$ son mutuamente independientes y están asociados son los términos de ruido en las ecuaciones de observación y de estado en (3.18). σ_ε y σ_η son las varianzas de las perturbaciones, ε_t y η_t , respectivamente. También, se supone que los parámetros relacionados con el sistema (3.18) son conocidos.

Con base en el grupo de datos observados $y_{1:T}$, el principal objetivo del análisis del sistema (3.18), es simular un conjunto de estados $h_{0:T}$ desde la función de densidad posterior, la cual puede ser aproximada por medio de

$$p(h_{0:T}|y_{1:T}) \propto p(h_0) \prod_{t=1}^T p(y_t|h_t)p(h_t|h_{t-1}) \tag{3.19}$$

Donde $p(h_0)$ es la densidad del estado inicial y $p(y_t|h_t)$ y $p(h_t|h_{t-1})$ son definidas por las ecuaciones (3.18). El algoritmo para realizar las simulaciones se implementa mediante la estrategia *FFBS (Forward Filtering Backward Sampling Algorithm)* de la siguiente manera:

Algoritmo FFBS (Forward Filtering Backward Sampling)

1. *FF*: Para cada momento $t = 1, 2, \dots, T$, mediante un proceso secuencial las cifras de y_t se actualizan numéricamente a partir de las densidades filtradas $p(h_t|y_{1:t})$ hasta el momento t .
2. *BS*: Se simula la distribución conjunta (3.19) mediante la forma implícita hacia atrás derivada del algoritmo que se muestra enseguida. Así, el objetivo es simular $p(h_{0:T}|y_{1:T})$.

$$p(h_{0:T}|y_{1:T}) \propto p(h_T|y_{1:T}) \prod_{t=1}^T p(h_{t-1}|h_t, y_{1:(t-1)}) \quad (3.20)$$

- a) Generar $h_T \sim p(h_T|y_{1:T})$ y se fija $t=T$
- b) Generar $h_{t-1} \sim p(h_{t-1}|h_t, y_{1:(t-1)})$
- c) Pasar de t a $t - 1$ y haga (b) de nuevo.
- d) Repita lo anterior hasta que $t = 0$

Lo anterior genera una secuencia completa conjunta de $h_{0:T}$ en orden descendente. Todos los pasos del algoritmo FFBS dependen fundamentalmente de la estructura de las densidades conjuntas:

$$p(y_t, h_t, h_{t-1}|y_{1:(t-1)}) = p(y_t|h_t)p(h_t|h_{t-1})p(h_{t-1}|y_{1:(t-1)}) \quad (3.21)$$

En particular, el filtrado (*forward filtering*) se basa en la capacidad para computar y resumir

$$\begin{aligned} p(h_{0:T}|y_{1:t}) &\propto p(y_t|h_t) \int p(h_t|h_{t-1})p(h_{t-1}|y_{1:(t-1)})dh_{t-1} \\ &= \int p(y_t, h_t, h_{t-1}|y_{1:(t-1)})dh_{t-1} \end{aligned} \quad (3.22)$$

mientras que el muestreo hacia atrás (*backward sampling*) se basa en la capacidad para simular

$$p(h_{t-1}|h_t, y_{1:t}) \propto p(y_t|h_t) \int p(h_t|h_{t-1})p(h_{t-1}|y_{1:(t-1)})dh_{t-1} \quad (3.23)$$

3.5.4. Algoritmo FFBS por aproximación de mixturas de distribuciones normales

En la práctica muchos investigadores utilizan el algoritmo FFBS incorporando una pequeña modificación en su estructura, al asumir que $p(h_{t-1}|y_{1:(t-1)})$ puede ser aproximada por la composición de un número J de distribuciones normales. A continuación se describe brevemente como funciona el algoritmo, si el lector desea revisar demostraciones de (3.25, 3.26 y 3.27) puede consultar Liu [2007].

Suponga que en el momento $t - 1$, la densidad $p(h_{t-1}|y_{1:(t-1)})$ es aproximada mediante

$$(h_{t-1}|y_{1:(t-1)}) \sim N(p_{t-1,1:J}, M_{t-1,1:J}, V_{t-1,1:J}) \quad (3.24)$$

donde $N(p_{t-1,1:J}, M_{t-1,1:J}, V_{t-1,1:J})$ representa cada una de la J componentes de la mixtura de normales gaussianas. Cada una de ellas tiene una media $M_{t-1,1:J}$ y varianza $V_{t-1,1:J}$ y están ponderadas por una probabilidad $p_{t-1,1:J}$. Entonces la densidad en la ecuación (3.21) es aproximada ahora a través de

$$p(y_t, h_t, h_{t-1}|y_{1:(t-1)}) = \sum_{j=1}^J p_{t-1,j} p(y_t|h_t) p(h_t|h_{t-1}) N(h_{t-1}|M_{t-1,j}, V_{t-1,j}) \quad (3.25)$$

En (3.25) los sumandos tiene dos términos que luego serán reemplazados por distribuciones normales, así $(h_t|h_{t-1}) \sim N(a_{t,j} + g'_t(M_{t-1,j})(h_{t-1} - M_{t-1,j}), \sigma_\eta)$ con $a_{t,j} = g_t(M_{t-1,j})$ y $(y_t|h_t) \sim N(f_{t,j} + f'_t(a_{t,j})(h_t - a_{t,j}), \sigma_\varepsilon)$ con $f_{t,j} = f_t(a_{t,j})$ donde $f'_t(\cdot)$ y $g'_t(\cdot)$ son las derivadas de las funciones de la ecuación de observación y de estado, respectivamente. Esto reduce (3.25) a una mixtura de normales trivariadas, así que todas las *marginales* y *condicionales* pueden ser computadas como una mezcla de normales.

Por último, las distribuciones clave para el filtrado y el suavizamiento que se utilizan en el algoritmo son:

- La ecuación para el filtrado (ecuación 3.22) queda así:

$$(h_t|y_{1:t}) \sim N(p_{t,1:J}, M_{t,1:J}, V_{t,1:J}) \quad (3.26)$$

para cualquier $j \in J$ se tiene que $M_{t,j} = a_{t,j} + A_{t,j}(y_t - f_{t,j})$ y $V_{t,j} = R_{t,j} - A_{t,j}^2 Q_{t,j}$ donde $A_{t,j} = \frac{R_{t,j} f'_t(a_{t,j})}{Q_{t,j}}$, $R_{t,j} = V_{t-1,j} [g'_t(M_{t-1,j})]^2 + \sigma_\eta$ y $Q_{t,j} = R_{t,j} [f'_t(a_{t,j})]^2 + \sigma_\varepsilon$. De igual manera, las probabilidades de los componentes j son actualizados a través de $p_{t,j} \propto p_{t-1,j} N(y_t | f_{t,j}, Q_{t,j})^{10}$.

- La ecuación para el suavizado (ecuación 3.23) queda así:

$$(h_{t-1} | h_t y_{1:(t-1)}) \sim N(q_{t,1:J}, x_{t,1:J}, X_{t,1:J}) \quad (3.27)$$

para cualquier $j \in J$ se tiene que $x_{t,j} = M_{t-1,j} + B_{t,j}(h_t - a_{t,j})$ y $X_{t,j} = V_{t-1,j} - B_{t,j}^2 R_{t,j}$ donde $B_{t,j} = \frac{V_{t-1,j} g'_t(M_{t-1,j})}{R_{t,j}}$. De la misma forma, las probabilidades de los componentes j son actualizados a través de $q_{t,j} \propto p_{t-1,j} N(h_t | a_{t,j}, R_{t,j})$.

3.6. Métodos secuenciales de Montecarlo (SMC)

Los métodos convencionales de MCMC asumen que la secuencia de la densidad objetivo es invariante en el tiempo, con lo cual se pierde la posibilidad de incorporar las observaciones para mejorar las estimaciones realizadas en forma recursiva. Una herramienta que permite incorporar la nueva información en la recursión son los métodos secuenciales de Monte Carlo, también denominados como *los Filtros de Partículas (FP)*.

3.6.1. Muestreo de Importancia

Considere una aproximación para distribución conjunta dada por $p(h_{0:n} | y_{1:n}; \Theta)$ y una función para la evolución de la variable estado (por ejemplo g_t en el sistema (3.18)). Un estimador de Monte Carlo estándar de la integral

$$E[g(h_{0:n})] = \int g(h_{0:n}) p(h_{0:n} | y_{1:n}; \Theta) dh_{0:n} \quad (3.28)$$

¹⁰Este procedimiento, denominado en la literatura como *adaptive mixture modeling*, genera una secuencia de mezclas gaussianas que se aproximan a la secuencia de distribuciones filtradas de estado.

consiste en generar secuencias de $h_{0:n}$ directamente de una distribución objetivo $p(h_{0:n}|y_{1:n}; \Theta)$. Por la ley fuerte de los grandes números, los promedios muestrales convergerán a los momentos poblacionales en cuanto el número de realizaciones aleatorias (*draws*) se incremente. Sin embargo, es poco probable utilizar esta estrategia en la práctica debido a que usualmente la distribución objetivo no es estándar ni conocida, y por lo tanto, se dificulta realizar selecciones aleatorias directamente de ésta (muestrear y evaluar).

Una forma de aproximarse a la implementación de dicho procedimiento es a través de una distribución *propuesta* ó distribución de *importancia* ($q_{0:n}(h_{0:n}|y_{1:n}; \psi)$)¹¹, la cual es más fácil de muestrear. Ésta es elegida por el investigador con la única restricción de que el soporte de $q(\cdot|y_{1:n})$ contenga al soporte de $p(\cdot|y_{1:n})$. La variable ψ denota un vector de parámetros de ajuste los cuales son usados para que la distribución *propuesta* se aproxime lo más posible a la distribución objetivo. La integral (3.28) es aproximada mediante:

$$E[g(h_{0:n})] = \int g(h_{0:n}) \frac{p(h_{0:n}|y_{1:n}; \Theta)}{q_{0:n}(h_{0:n}|y_{1:n}; \psi)} q_{0:n}(h_{0:n}|y_{1:n}; \psi) dh_{0:n} \quad (3.29)$$

Después de tomar N selecciones aleatorias de $\{h_{0:n}^{(i)}\}_{i=1}^N$ de la distribución *propuesta*, éstas realizaciones son reponderadas para corregir el hecho de que ellas pudieron ser elegidas de una distribución errónea. Las ponderaciones o *pesos de importancia* $\{w_n^{(i)}\}_{i=1}^N$, son definidas como el cociente entre la distribución objetivo y la distribución *propuesta*

$$w_n^{(i)} \propto \frac{p(h_{0:n}^{(i)}|y_{1:n}; \Theta)}{q_{0:n}(h_{0:n}^{(i)}|y_{1:n}; \psi)} \quad (3.30)$$

Dadas las secuencias $\{h_{0:n}^{(i)}, w_n^{(i)}\}_{i=1}^N$, el estimador de muestreo de importancia (*Importance sampling (IS)*)¹² de (3.28) es

¹¹En inglés, *proposal distribution*.

¹²Los pesos de importancia son auto-normalizados porque la constante de integración en la distribución objetivo puede ser no conocida.

$$E[g(h_{0:n})] \approx \sum_{i=1}^N g(h_{0:n}^{(i)}) \widehat{w}_n^{(i)} \quad (3.31)$$

donde $\widehat{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^N w_n^{(j)}}$

IS fue inventado por Marshall [1956] y usado por Kloek and van Dijk [1978] en el entorno econométrico, desde entonces los investigadores necesitan hallar una distribución de *importancia* $q_{0:n}(h_{0:n}|y_{1:n}; \psi)$ para facilitar el uso del algoritmo IS. En general, se han propuesto muchos procedimientos para encontrar una propuesta apropiada de la distribución conjunta implicada en sistemas de estado-espacio. Ejemplos se pueden encontrar en Shephard and Pitt [1997] y Durbin and Koopman [2001] que utilizaron aproximaciones de Laplace de las integrales como una distribución propuesta.

Por último, si bien el método IS representa un enfoque simple para aproximar integrales definidas invariantes, no es adecuado dentro del contexto de estimación recursiva. Esto obedece al elevado costo computacional que exige la alta dimensionalidad de la densidad objetivo $p(h_{0:t}|y_{1:t})$, es decir, se necesita todo el conjunto $y_{1:t}$ cada vez que se desea estimar $p(h_{0:t}|y_{1:t})$, lo cual puede resultar ineficiente. Adicionalmente, Doucet et al. [2001] muestra que los pesos de importancia tienden a degenerarse cuando se realizan estimaciones recursivas usando IS, por lo que deberían incluirse rutinas de remuestreo.

3.6.2. Muestreo Secuencial de Importancia

Para evitar el costo computacional de utilizar toda la secuencia $y_{1,t}$ cada vez que se pretende estimar $p(h_{0:t}|y_{1:t})$, Hammersley and Morton [1954] y Handschin [1970] formularon el método de muestreo secuencial de importancia (*Sequential Importance sampling (SIS)*) y su aplicación en modelo de estado-espacio. En este procedimiento es posible encontrar una relación recursiva para la estimación de la densidad *posterior* en la que, en el momento t se conservan las muestras anteriores del estado $\{h_{0:(t-1)}^{(i)}; i = 1, \dots, N\}$. Luego, únicamente se extraen las muestras $\{h_t^{(i)}; i = 1, \dots, N\}$ con base a la nueva observación recibida y_t y a la estimación realizada en el momento inmediatamente anterior de la densidad de filtrado $p(h_{0:t-1}|y_{1:t-1})$. El objetivo es aproximar (3.6) y el algoritmo SIS funciona como sigue

$$q_{0:n}(h_{0:n}|y_{1:n}; \psi) \equiv q_n(h_n|h_{0:(n-1)}, y_{1:n}; \psi) q_{0:(n-1)}(h_{0:(n-1)}|y_{1:(n-1)}; \psi) \quad (3.32)$$

La segunda distribución $q_{0:(n-1)}(h_{0:(n-1)}|y_{1:(n-1)}; \psi)$ es una “per-partícula medida de Dirac”, tomando una unidad en su masa de probabilidad sobre cada senda que ha sido simulada en las previas iteraciones hasta el momento $n - 1$. Al mismo tiempo, un nuevo conjunto de valores $\{h_n^{(i)}\}_{i=1}^N$ son seleccionados aleatoriamente en el momento n de la primera parte de la distribución de *importancia* $q_n(h_n|h_{0:(n-1)}, y_{1:n}; \psi)$. También, una nueva secuencia de sendas es obtenida al guardar las trayectorias de las viejas realizaciones, $\{h_{0:n}^{(i)}\}_{i=1}^N = \{h_{0:(n-1)}^{(i)}, h_n^{(i)}\}_{i=1}^N$ ¹³. Cabe señalar que los sub-índices de tiempo en la distribución *propuesta* ($q_{0:n}$) indican que tanto ésta, como los parámetros que la componen (ψ) pueden cambiar en el tiempo.

Para obtener los pesos de importancia se reemplaza (3.6) y (3.32) en (3.30)

$$\begin{aligned} w_n &= \frac{p(y_n|h_n; \Theta)p(h_n|h_{n-1}; \Theta)p(h_{0:(n-1)}|y_{1:(n-1)}; \Theta)}{p(y_n|y_{1:(n-1)}; \Theta)q_n(h_n|h_{0:(n-1)}, y_{1:n}; \psi)q_{0:(n-1)}(h_{0:(n-1)}|y_{1:(n-1)}; \psi)} \\ &\propto w_{n-1} \frac{p(y_n|h_n; \Theta)p(h_n|h_{n-1}; \Theta)}{q_n(h_n|h_{0:(n-1)}, y_{1:n}; \psi)} \\ &\propto w_{n-1} \tilde{w}_n \end{aligned} \quad (3.33)$$

$$\text{donde } \tilde{w}_n = \frac{p(y_n|h_n; \Theta)p(h_n|h_{n-1}; \Theta)}{q_n(h_n|h_{0:(n-1)}, y_{1:n}; \psi)}$$

Las densidades $p(y_n|h_n; \Theta)$ y $p(h_n|h_{n-1}; \Theta)$ son determinadas por el modelo planteado de estado-espacio, como por ejemplo el sistema (3.18). La proporción de densidades (\tilde{w}_n) son conocidas como *los pesos de importancia incrementales*, la cual es la única parte de la ponderación de importancia que necesita ser computada en cada iteración. Así, al final de cada iteración, el algoritmo SIS genera N sendas simuladas y una secuencia de pesos de importancia ($\{h_{0:n}^{(i)}, w_n^{(i)}\}_{i=1}^N$). El estimador de muestreo de importancia en el algoritmo SIS se calcula de la misma forma que en (3.31).

3.6.3. Filtro de Partículas

En su trabajo seminal Gordon et al. [1993] introdujo un procedimiento denominado *Filtro de Partículas*, en el cual adicionaba un paso de remuestreo dentro del algoritmo SIS con lo que mitigaba el problema de degeneramiento de los pesos o ponderaciones. Con el remuestreo se generaba una nueva población de partículas que eran replicadas tomando las ya existentes y ponderadas o normalizadas por su peso de importancia.

¹³Con el muestreo de importancia es posible obtener una muestra de $h_{0:n}^{(i)}$ sin alterar la muestra anterior correspondiente a $h_{0:(n-1)}^{(i)}$. Esto implica que $h_n^{(i)} \sim q_n(h_n|h_{0:(n-1)}^{(i)})$.

Así, se generan N variables aleatorias con reemplazamiento de una distribución multinomial con probabilidades $\{\hat{w}^i\}_{i=1}^N$. Luego, las partículas con una gran importancia son aleatoriamente duplicadas mientras que las de menor peso son eliminadas. El peso está definido dentro del algoritmo que sigue a continuación a través de $\hat{w}_n^{(i)}$.

Una vez remuestreados los pesos de las partículas, se fijan igual a una constante, por ejemplo $w_n^{(i)} = \frac{1}{N}$ para $i = 1, 2, \dots, N$. Con esto se obliga a que los pesos no se degeneren constantemente como en el algoritmo SIS. El nuevo algoritmo llamado *Muestreo secuencial de importancia con remuestreo (SISR)*¹⁴, es una combinación de el muestro de importancia con reemplazamiento (Rubin [1988]) y el algoritmo SIS expuesto arriba.

Algoritmo SISR (Sequential importance sampling with resampling)

1. Para el momento $n = 0$, para cada iteración $i = 1, 2, \dots, N$

$$\text{Generar } h_0^{(i)} \sim q_0(h_0) \text{ y fijar } w_0^{(i)} = \frac{p(h_0^{(i)})}{q_0(h_0^{(i)})}$$

2. Para $n = 1, 2, \dots, T$

- a) Para $i = 1, 2, \dots, N$ generar $h_n^{(i)} \sim q_n(h_n | h_{n-1}^{(i)}, y_n; \psi)$ y computar los pesos de importancia $w_n^{(i)} \propto w_{n-1}^{(i)} \frac{p(y_n | h_n^{(i)}; \Theta) p(h_n^{(i)} | h_{n-1}^{(i)}; \Theta)}{q_n(h_n^{(i)} | h_{n-1}^{(i)}, y_n; \psi)}$

- b) Para $i = 1, 2, \dots, N$ normalizar los pesos de importancia:

$$\hat{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^N w_n^{(j)}}$$

- c) Remuestrear N partículas con probabilidades $\{\hat{w}_n^{(i)}\}_{i=1}^N$ y para $i = 1, 2, \dots, N$ fijar $w_n^{(i)} = \frac{1}{N}$

El estimador del momento $E[g(h_{0:n})]$ se obtiene aplicando (3.31) y utilizando los resultados del algoritmo anterior.

En la literatura de métodos de Monte Carlo y filtros de partículas se han desarrollado gran variedad de algoritmos que han permitido mejorar el modelamiento de sistemas no lineales por medio de la inferencia bayesiana. Dentro del campo de los algoritmos SISR: el filtro *bootstrap*, el filtro de *partículas auxiliar* y el filtro *Liu-West* se destacan entre los más populares y usados. A continuación se presentan brevemente los dos primeros, mientras que el último se discute en el siguiente capítulo donde se realiza una aplica-

¹⁴En inglés, *sequential importance sampling with resampling (SISR)*.

ción sobre la tasa de cambio de Colombia (pesos por dólar).

El filtro de *Bootstrap* (BF) fue desarrollado por Gordon et al. [1993], quienes propusieron utilizar la densidad de transición como una distribución *propuesta*, $q_n(h_n|h_{n-1}, y_n; \psi) = p(h_n|h_{n-1}; \Theta)$. Este algoritmo remuestra en cada período de tiempo y hace que los pesos de importancia incrementales sean iguales a $\tilde{w}_n = p(y_n|h_n; \Theta)$. Creal [2012] menciona que este filtro puede tener un buen desempeño pero presenta una debilidad al no utilizar la observación actual (y_n) para proponer nuevas partículas dentro de la distribución *propuesta*. Por su parte, el filtro de *partículas auxiliar*¹⁵ propuesto por Pitt and Shephard [1999] sí incorpora lo anterior dentro del algoritmo de solución, generando distribuciones denominadas *propuestas adaptadas*.

En Pitt and Shephard [1999] se aproxima la distribución objetivo en (3.6) con la distribución de importancia (q)

$$\begin{aligned} p(y_n|h_n; \Theta)p(h_n|h_{n-1}; \Theta) &\approx q_{1,n}(y_n|h_n; \psi)q_{2,n}(h_n|h_{n-1}; \psi) \\ &= q_{1,n}(y_n|h_{n-1}; \psi)q_{2,n}(h_n|h_{n-1}; \psi, y_n) \end{aligned} \quad (3.34)$$

La distribución *propuesta* sugerida en (3.34) está compuesta por dos partes, las cuales implican que el muestreo de los nuevos valores de $\{h_n^{(i)}\}_{i=1}^N$ son generados por esta distribución a través de la ejecución de dos pasos dentro del algoritmo que se define

¹⁵En inglés, Auxiliary particle filter (APF).

Algoritmo Filtro de Partículas Auxiliar (APF)

1. Para el momento $n = 0$, para cada iteración $i = 1, 2, \dots, N$

Generar $h_0^{(i)} \sim q_0(h_0)$ y fijar $w_0^{(i)} = \frac{p(h_0^{(i)})}{q_0(h_0^{(i)})}$

2. Para $n = 1, 2, \dots, T$

a) Para $i = 1, 2, \dots, N$ calcular $\delta^{(i)} = q_n(y_n | h_{n-1}^{(i)}; \psi) w_{n-1}^{(i)}$

y normalizar $\hat{\delta}^{(i)} = \frac{\delta^{(i)}}{\sum_{j=1}^N \delta^{(j)}}$

b) Remuestrear N partículas $\{h_{n-1}^{(i)}\}_{i=1}^N$ con probabilidades $\{\hat{\delta}^{(i)}\}_{i=1}^N$

c) Para $i = 1, 2, \dots, N$ generar $h - n^{(i)} \sim q_n(h_n | h_{n-1}^{(i)}; y_n, \psi)$

d) Para $i = 1, 2, \dots, N$ calcular los pesos de importancia:

$$w_n^{(i)} = \frac{p(y_n | h_n^{(i)}; \Theta) p(h_n^{(i)} | h_{n-1}^{(i)}; \Theta)}{q_n(y_n | h_{n-1}^{(i)}; \psi) q_n(h_n^{(i)} | h_{n-1}^{(i)}; y_n, \psi)}$$

Un caso particular del algoritmo APF se da cuando se selecciona como *propuesta* a $q_{1,n}(y_n | h_{n-1}; \psi) = 1$ y $q_{2,n}(h_n | h_{n-1}; \psi, y_n) = p(h_n | h_{n-1}; \psi)$, así el algoritmo APF se reduce al filtro *Bootstrap*.

Capítulo 4

Aplicación Empírica: Volatilidad diaria de la tasa de cambio peso-dólar

En este trabajo de tesis el conjunto de variables observables denotado como $\{y_t\}_{t=1}^T$ y el conjunto de variables no observables denotado como $\{h_t\}_{t=1}^T$, corresponden a los retornos financieros de la tasa de cambio de Colombia y a su volatilidad, respectivamente. Como se mencionó antes, los retornos son variables observables, es decir son conocidos (Ver Figura 1.2) mientras que la volatilidad es una variable *no observable*. De hecho, uno de los objetivos de esta tesis, es estimar esta última mediante el uso de modelos econométricos de heteroscedasticidad condicional y los de volatilidad estocástica, que fueron mencionados en el capítulo 2 y 3. Las aplicaciones fueron realizadas en el paquete econométrico R y están basadas en programaciones desarrolladas por Gaman and Lopes [2006], Doucet et al. [2001], Tsay [2010] y Lopes and Tsay [2011], que luego fueron adaptadas al problema específico de estimar la volatilidad de la tasa de cambio de Colombia.

4.1. Modelo de Heteroscedasticidad condicional: el caso del GARCH(1,1)

Para modelar la volatilidad de los retornos de la tasa de cambio peso-dólar diaria desde enero de 2002 hasta junio de 2013, se tomó la estructura del modelo planteado en la ecuación (2.12). Esta elección se sustenta en las siguientes dos razones: i) en la práctica el modelo GARCH(1,1) presenta un buen ajuste y usualmente parece ser el más adecuado para modelar las series de tiempo financieras [Bollerslev et al. [1992], Hansen and Lunde [2001], Mapa [2004] y Matei [2009]] y ii) se estimó previamente un GARCH(p,q) desde $p = 1, \dots, 10$ y $q = 1, \dots, 10$ y se observó que mediante el uso de criterios de información (*Akaike, Bayes, Hannan-Quinn y Shibata*) el modelo GARCH(1,1) se encuentra entre los tres mejores posibles modelos para modelar la volatilidad de la

TRM¹.

De esta forma, siguiendo el modelo

$$\begin{aligned}
 y_t &= \varepsilon_t \quad \text{donde } \varepsilon_t | \Psi_{t-1} \sim N(0, h_t) \\
 v_t &= \frac{\varepsilon_t}{\sqrt{h_t}} \\
 h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}
 \end{aligned} \tag{4.1}$$

Se obtuvo la siguiente estimación de los parámetros

Parámetro	α_0	α_1	β_1
Estimador	0.003908	0.169782	0.829218
Des. est.	(0.003373)	(0.036665)	(0.049690)

Cuadro 4.1: Parámetros estimados para el GARCH(1,1)

Las salidas completas de la estimación se pueden apreciar en el Apéndice E.3. Por su parte, la estimación de la volatilidad de la TRM (h_t) de acuerdo al modelo (4.1) fue

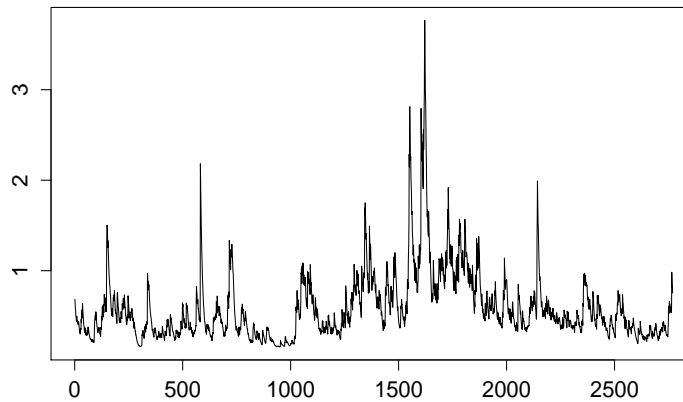


Figura 4.1: Volatilidad estimada de la TRM (Modelo GARCH(1,1))

¹Las otras dos mejores opciones son el modelo GARCH(2,3) y el GARCH(3,3).

4.2. Modelo de volatilidad estocástica (SV) de orden 1.

A partir de lo presentado en los capítulos 2 y 3, el modelo considerado es un modelo de volatilidad estocástica de orden 1 como el siguiente (Ver ecuaciones 2.17, 2.18 y 2.19)²

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\} \sigma_\varepsilon \varepsilon_t, \text{ con } \varepsilon_t \stackrel{i.i.d.}{\sim} \mathbf{N}(0, 1) \\ h_t &= \mu + \phi h_{t-1} + \tau \eta_t, \text{ con } \eta_t \stackrel{i.i.d.}{\sim} \mathbf{N}(0, 1) \end{aligned} \quad (4.2)$$

Como se apreció en el capítulo 3, los métodos de estimación bayesianos parten de una condición inicial para la variable de estado, que en este caso es la log-volatilidad (h_t) en el momento cero. La cual está definida como (también denominada en la literatura como la *prior*)

$$h_0 \sim \mathbf{N}\left(\frac{\mu}{1-\phi}, \frac{\tau^2}{1-\phi^2}\right) \quad (4.3)$$

Asimismo, siguiendo el procedimiento descrito en (3.18) y (3.19), y sabiendo que el objetivo es encontrar la función de densidad conjunta de (4.2), a continuación se muestran varias aproximaciones mediante las cuales se puede estimar el modelo de volatilidad estocástica. La estimación bayesiana del modelo SV (4.2), utiliza la *prior* (4.3) y es completado con una distribución *a priori* conjugada, denominada *conjugate prior*, para los parámetros del modelo (μ, ϕ, τ^2).

Por ejemplo, simplificando la notación, suponga que $\beta = (\mu, \phi)$ y $\theta = (\beta, \tau^2) = (\mu, \phi, \tau^2)$, entonces la *prior* conjugada del modelo es $p(\theta) = p(\beta|\tau^2)p(\tau^2)$ donde $(\beta|\tau^2) \sim N(b_0, \tau^2 B_0)$ y $\tau^2 \sim IG(c_0, d_0)$, con parámetros (b_0, B_0, c_0, d_0) asociados a medias y varianzas de dichas distribuciones que son conocidos previamente (Ver desarrollo matemático del modelo SV en Apéndice B.2). La distribución *conjugada* en este caso es una Normal-Inversa Gamma.

Dado un conjunto de información observada $y^T = (y_1, y_2, \dots, y_T)$, en este caso los retornos de la tasa de cambio nominal peso-dólar entre enero de 2002 y junio de 2013, y el modelo presentado en 4.2, la distribución *posterior* de la variable estado, en este caso la log-volatilidad (h_t), y los parámetros θ está dada por la regla de Bayes (Apéndice A.2)

²Se puede fijar $\sigma_\varepsilon^2 = \exp\{\mu\}$.

$$p(h^T, \theta | y^T) \propto p(\theta) \prod_{t=1}^T p(y_t | h_t, \theta) p(h_t | h_{t-1}, \theta) \quad (4.4)$$

la cual como se mencionó en el capítulo 3 no se puede resolver analíticamente por la no linealidad del modelo SV (ecuación 4.2). La aproximación utilizada para resolver este problema se basa en los métodos de inferencia bayesiana, señalados en el capítulo previo, que incorporan algoritmos de Cadenas de Markov de Monte Carlo (MCMC).

4.2.1. Inferencia de la Posterior vía Cadenas de Markov de Monte Carlo

El procedimiento ha seguir, teniendo en cuenta los algoritmos vistos en el capítulo 3, establece que se deben generar muestras de los parámetros y luego de la variable de estado en cada momento de tiempo t .

Siguiendo a Gamerman and Lopes [2006], dados $y^t = (y_1, y_2, \dots, y_t)$ y $h^t = (h_0, h_1, h_2, \dots, h_t)$ para $t = 1, 2, \dots, T$, la distribución condicional de los parámetros θ está dada por

$$p(\theta | y^t, h^t) = p(\theta | b_t, B_t, c_t, d_t) = f_N(\beta; b_t, \tau^2 B_t) f_{IG}(\tau^2; c_t, d_t) \quad (4.5)$$

donde $f_N(x; \bar{x}, \sigma^2)$ es la función de densidad de la distribución normal con media \bar{x} y varianza σ^2 , evaluada en x y $f_{IG}(x; \bar{x}, \sigma^2)$ es la función de densidad de la distribución Inversa Gamma con parámetros \bar{x}, σ^2 , evaluada en x . En otras palabras, b_t, B_t, c_t, d_t están asociados a la media y varianza de las distribuciones Normal e Inversa Gamma planteadas en (4.5). El lector puede encontrar en el Apéndice B.2 los parámetros para la media y la varianza de cada distribución.

Siguiendo a Lopes and Tsay [2011] los parámetros (b_t, B_t, c_t, d_t) pueden ser determinados de forma recursiva de la siguiente manera

$$\begin{aligned} b_t &= B_t^{-1}(B_{t-1}^{-1}b_{t-1} + h_t z_t) \text{ y } B_t^{-1} = B_{t-1}^{-1} + z_t z_t' \\ c_t &= c_{t-1} + 1/2 \text{ y } d_t = d_{t-1} + (h_t - b_t' z_t) h_t / 2 + (b_{t-1} - b_t)' B_{t-1}^{-1} b_{t-1} / 2 \end{aligned} \quad (4.6)$$

donde $z_t' = (1, h_{t-1})$

Con base en los algoritmos mencionados en el capítulo 3, siguiendo a Jacquier et al. [1994], el algoritmo básico de MCMC utilizado para estimar el modelo SV, el cual muestrea h_t en cada momento del tiempo t , y condiciona sobre $h_{-t} = (h_0, h_1, h_2, \dots, h_{t-1}, h_{t+1}, \dots, h_T)$ provenientes de

$$\begin{aligned} p(h_t|h_{-t}, \theta, y^{1:T}) &\propto p_N(y_t; 0, \exp\{h_t\})p_N(h_t; \mu + \phi h_{t-1}, \tau^2)p_N(h_{t+1}; \mu + \phi h_t, \tau^2) \\ &\propto p_N(y_t; 0, \exp\{h_t\})f_N(h_t; \vartheta_t, \omega^2) \end{aligned} \quad (4.7)$$

F_N se obtiene al combinar las densidades de la evolución del estado ya que éstas sólo dependen condicionalmente de h_{t-1} y h_{t+1} . Así, $\vartheta_t = (\mu(1 - \phi) + \phi(h_{t+1} + h_{t-1})) / (1 + \phi^2)$ para $t = 1, 2, \dots, T - 1$, $\vartheta_T = \mu + \phi h_{T-1}$ y $\omega^2 = \tau^2 / (1 + \phi^2)$ (Ver Lopes and Tsay [2011]). Teniendo en cuenta la descripción del algoritmo de Metropolis (sección 3.5.2) a continuación se realizan dos aproximaciones para estimar la volatilidad de la tasa de cambio peso-dólar.

Primero, se utiliza un algoritmo de Metropolis simple con caminata aleatoria y allí se calibra la varianza v_h^2 y el estado actual de $h_t^{(j)}$ (para la iteración j) de la siguiente forma. Para $t = 1, \dots, T$, genere muestras de h_t^* de $N(h_t^{(j)}, v_h^2)^3$ y acepte dicha realización con probabilidad

$$\alpha = \min \left\{ 1, \frac{f_N(h_t^*; \vartheta_t, v_t^2) f_N(y_t; 0, \exp\{h_t^*\})}{f_N(h_t^{(j)}; \vartheta_t, v_t^2) f_N(y_t; 0, \exp\{h_t^{(j)}\})} \right\} \quad (4.8)$$

Como resultado se obtuvo la siguiente estimación de los parámetros del modelo SV con el algoritmo de Metropolis simple con caminata aleatoria (SVMRW)

Parámetro	μ	ϕ	τ^2
Estimador	-0.043163	0.972371	0.093822
Des. est.	(0.010527)	(0.003817)	(0.009545)

Cuadro 4.2: Parámetros estimados para el SV(1) (SVMRW)

Se puede apreciar la primer diferencia respecto a los modelos tradicionales de heteroscedasticidad condicional, en este caso con el modelo de volatilidad estocástica estimado a través de los métodos de Monte Carlo se puede generar una distribución de los

³ $v_h^2 \in v_t^2$.

parámetros. En el cuadro se aprecia el estimador puntual de cada uno de los coeficientes (su media), y la estimación de su desviación estándar. En el modelo GARCH, sólo se contaba con un estimador puntual de los parámetros. Como resultado de la estimación por (SVMRW) se observa que la volatilidad tiene una persistencia alta, dado que el parámetro ϕ está cercano a uno. A continuación, para cada una de las simulaciones se pueden apreciar la dinámica de los parámetros y algunas de sus características como la función de correlación simple y el histograma (Figuras 4.2 y 4.3)

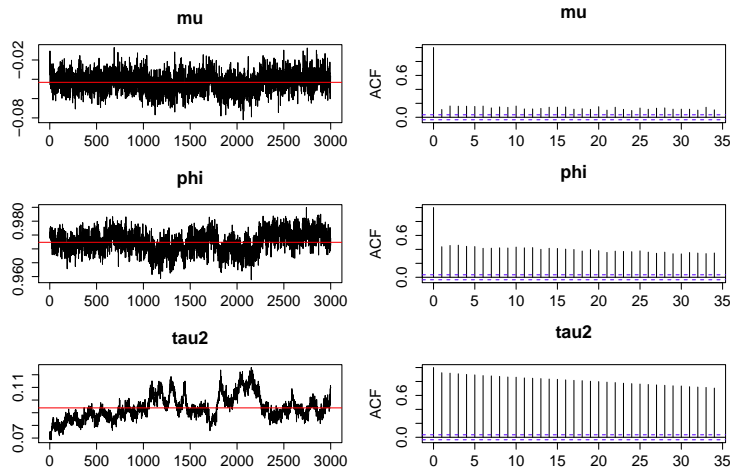


Figura 4.2: Evolución de los parámetros en cada iteración y función de correlación simple (Modelo SVMRW)

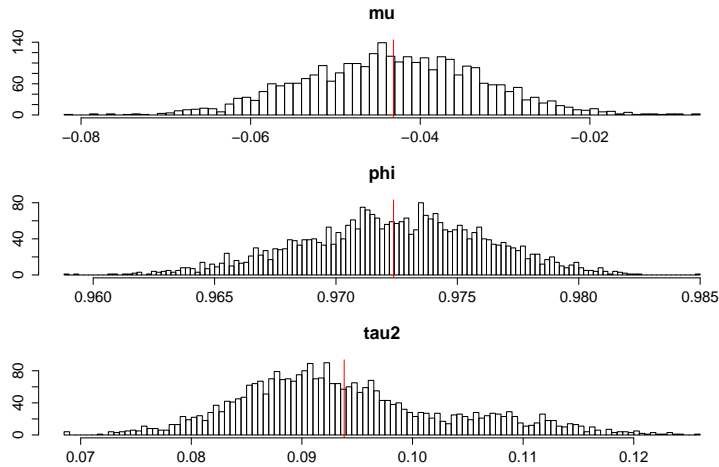


Figura 4.3: Histograma de los parámetros (Modelo SVMRW)

Por último, la volatilidad estimada (h_t) mediante el uso de este algoritmo es

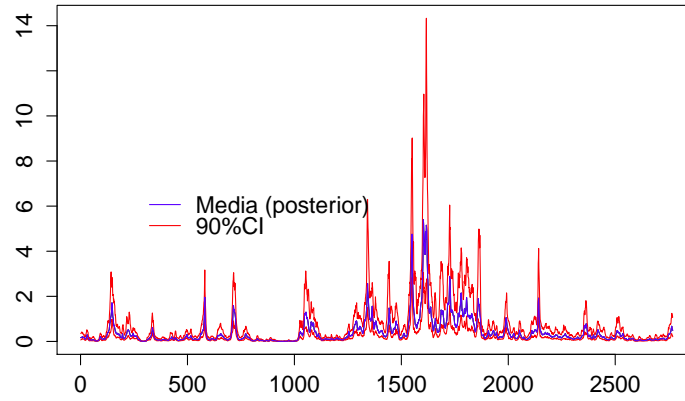


Figura 4.4: Volatilidad estimada de la TRM (Modelo SVMRW)

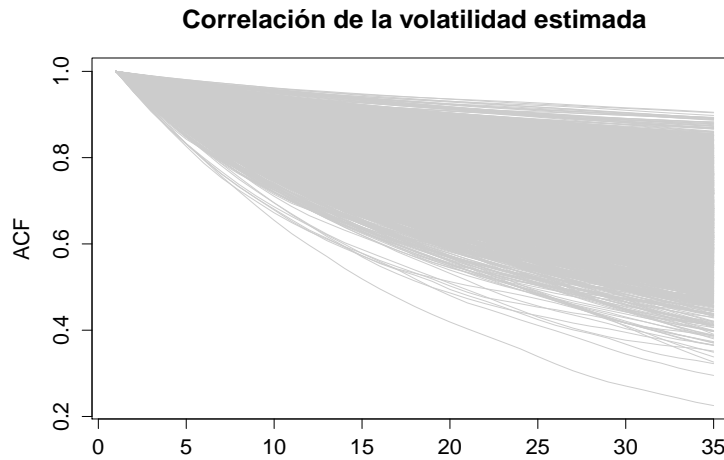


Figura 4.5: Correlación simple h_t (Modelo SVMRW)

Otra alternativa que puede considerarse, es que h_t puede ser muestreada vía Metropolis-Hasting independiente con una densidad normal *propuesta* ($q(\cdot) = \text{proposal distribution}$)

$$\begin{aligned}
 q(h_t | h_{-t}, \theta, y_T) &= f_N(h_t; \bar{\vartheta}_t, \nu_t^2) \\
 \text{donde } \bar{\vartheta}_t &= \vartheta_t + 0,5\nu_t^2(y_t^2 \exp\{-\vartheta_t\} - 1) \\
 \text{y } \nu_t^2 &= \omega^2 \text{ para } t = 1, 2, \dots, T-1 \text{ y } \nu_t^2 = \tau^2
 \end{aligned} \tag{4.9}$$

El algoritmo vía Metropolis-Hasting independiente funciona de la siguiente manera. Para $t = 1, \dots, T$ y un estado actual de $h_t^{(j)}$ en la iteración j , se muestrea h_t^* de $f_N(\bar{\vartheta}_t, \nu_t^2)$

y la realización se acepta con probabilidad

$$\alpha = \min \left\{ 1, \frac{f_N(h_t^*; \vartheta_t, \nu_t^2) f_N(y_t; 0, \exp\{h_t^*\})}{f_N(h_t^{(j)}; \vartheta_t, \nu_t^2) f_N(y_t; 0, \exp\{h_t^{(j)}\})} \times \frac{f_N(h_t^{(j)}; \bar{\vartheta}_t, \nu_t^2)}{f_N(h_t^*; \bar{\vartheta}_t, \nu_t^2)} \right\} \quad (4.10)$$

Como resultado se obtuvo la siguiente estimación de los parámetros del modelo SV con el algoritmo vía Metropolis-Hasting independiente (SVMHI)

Parámetro	μ	ϕ	τ^2
Estimador	-0.041859	0.972990	0.093374
Des. est.	(0.010682)	(0.003989)	(0.011114)

Cuadro 4.3: Parámetros estimados para el SV(1) (SVMHI)

Como resultado de la estimación por (SVMHI), también se aprecia que la volatilidad tiene una persistencia alta, dado que el parámetro ϕ está cercano a uno. Para cada una de las simulaciones se pueden apreciar la dinámica de los parámetros y algunas de sus características como la función de correlación simple y el histograma (Figuras 4.6 y 4.7)

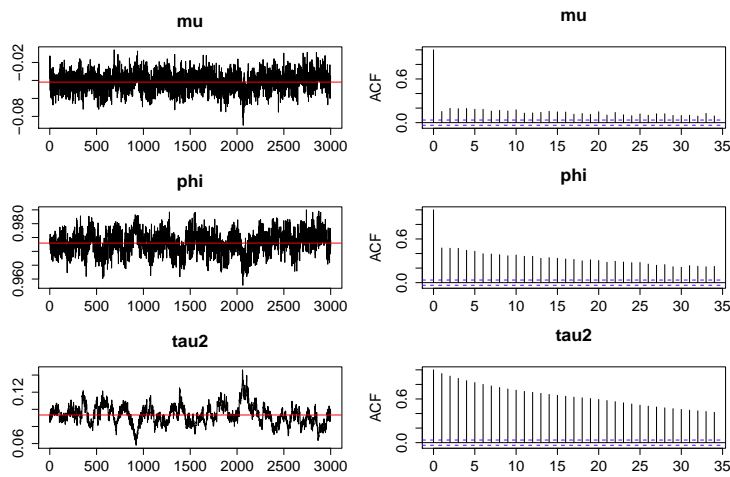


Figura 4.6: Evolución de los parámetros en cada iteración y función de correlación simple (Modelo SVMHI)

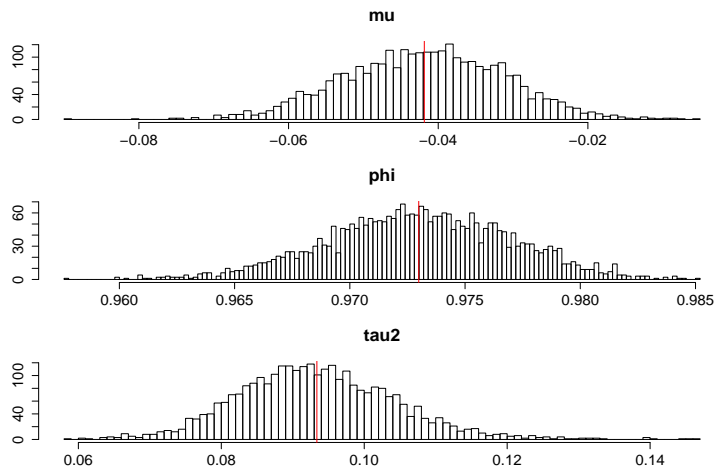


Figura 4.7: Histograma de los parámetros (Modelo SVMHI)

Por último, la volatilidad estimada (h_t) mediante el uso de este algoritmo es

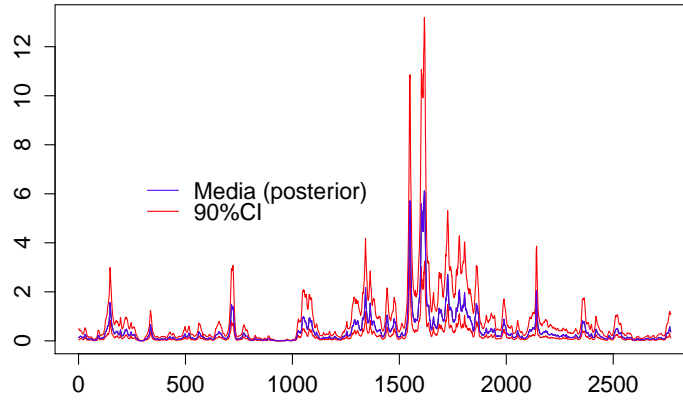


Figura 4.8: Volatilidad estimada de la TRM (Modelo SVMHI)

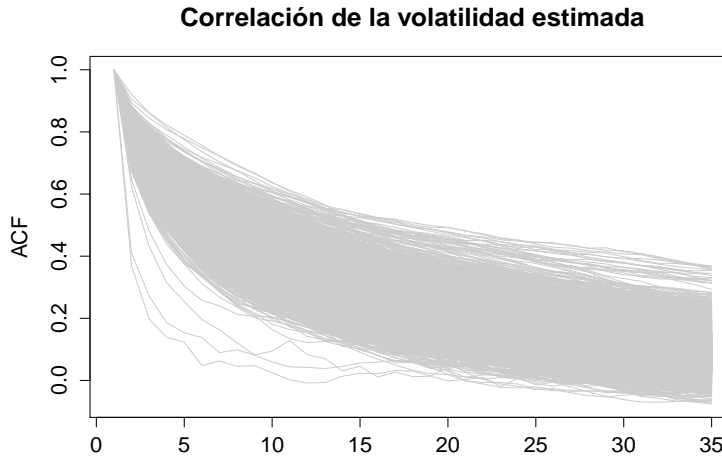


Figura 4.9: Correlación simple h_t (Modelo SVMHI)

En los procedimientos anteriores, se resuelve el problema mediante el cálculo de las *condicionales*, de los parámetros y de los estados para cada momento del tiempo. Sin embargo, West and Harrison [2007] muestran que el sistema (4.2) puede ser expresado como una clase de modelo dinámico lineal normal condicionado, entonces es factible muestrear la *condicionada* conjunta de los estados. En otras palabras, se hace el muestreo para todo el conjunto h^T proveniente de $p(h^T|y^T, \theta)$. La condicionada está dada por

$$p(h^T|y^T, \theta) = p(h_T|y^T, \theta) \prod_{t=1}^{T-1} p(h_t|h_{t+1}, \theta, y_t) \quad (4.11)$$

Utilizando el algoritmo FFBS (*Forward Filtering Backward Sampling*, Ver (3.20) del ca-

pítulo 3) se puede resolver el cálculo de 4.11. Como resultado se obtuvo la siguiente estimación de los parámetros del modelo SV con el algoritmo FFBS (SVFFBS)

Parámetro	μ	ϕ	τ^2
Estimador	-0.025492	0.983282	0.234235
Des. est.	(0.009164)	(0.004240)	(0.027640)

Cuadro 4.4: Parámetros estimados para el SV(1) (SVFFBS)

Como resultado de la estimación por (SVFFBS), también se aprecia que la volatilidad tiene una persistencia alta, dado que el parámetro ϕ está cercano a uno. Para cada una de las simulaciones se pueden apreciar la dinámica de los parámetros y algunas de sus características como la función de correlación simple y el histograma (Figuras 4.10 y 4.11). Asimismo la volatilidad estimada (h_t) ajustada por este algoritmo se presenta en las siguientes figuras 4.12 y 4.13.

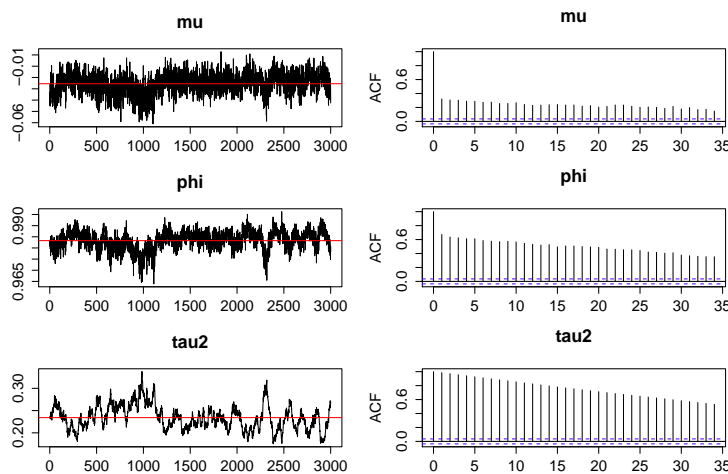


Figura 4.10: Evolución de los parámetros en cada iteración y función de correlación simple (Modelo SVFFBS)

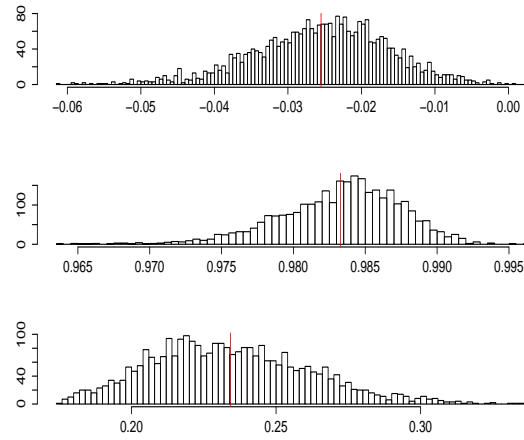


Figura 4.11: Histograma de los parámetros (Modelo SVFFBS)

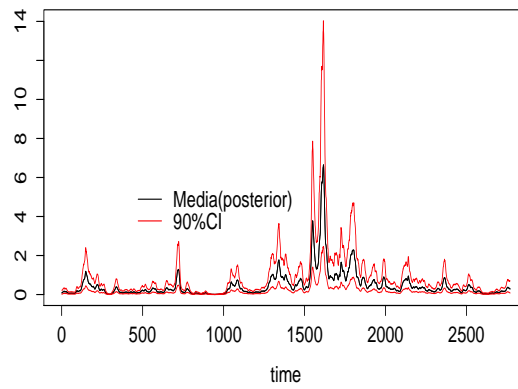


Figura 4.12: Volatilidad estimada de la TRM (Modelo SVFFBS)

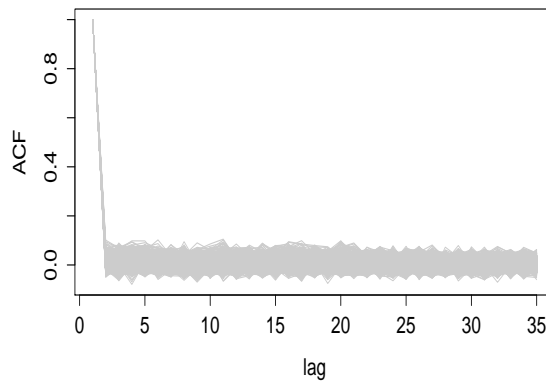


Figura 4.13: Correlación simple h_t (Modelo SVFFBS)

Para superar la dificultad asociada con no normalidad que se presenta en el modelo SV, Kim et al. [1998] utilizan una mezcla de siete distribuciones normales para aproximar la distribución de $\log y_t^2$. Para apreciar esto, partiendo de un sistema como 4.2 y reexpresándolo como $\log y_t^2 = h_t + \epsilon_t$ donde $\epsilon_t = \log \varepsilon_t^2$, se tiene que ϵ_t sigue una distribución $\log \chi_1^2$ con media -1.27 y varianza 4.94. La aproximación está dada por (Ver Tsay [2010])

$$p(\epsilon_t) = \sum_{i=1}^7 \pi_i p_N(\epsilon_t; \bar{m}_i, \bar{v}_i) \quad (4.12)$$

Componente	Probabilidad (π_i)	Media (\bar{m}_i)	Varianza (\bar{v}_i)
1	0.00730	-11.4004	5.7960
2	0.10556	-5.2432	2.6137
3	0.00002	-9.8373	5.1795
4	0.04395	1.5075	0.1674
5	0.34001	-0.6510	0.6401
6	0.24566	0.5248	0.3402
7	0.25750	-2.3586	1.2626

Cuadro 4.5: Mixtura de Normales

Con base en esta mixtura de Normales y utilizando el algoritmo (FFBS con Mixtura de Normales (FFBSM)) descrito en (3.24, 3.25, 3.26 y 3.27) se puede resolver el problema de encontrar la distribución conjunta para los estados del modelo SV (4.11). Como resultado se obtuvo la siguiente estimación de los parámetros del modelo SV con el algoritmo FFBS con Mixturas de Normales (SVFFBSM)

Parámetro	μ	ϕ	τ^2
Estimador	-0.041676	0.973202	0.301462
Des. est.	(0.010854)	(0.004356)	(0.021841)

Cuadro 4.6: Parámetros estimados para el SV(1) (SVFFBM)

Para cada una de las simulaciones se pueden apreciar la dinámica de los parámetros y algunas de sus características como la función de correlación simple y el histograma (Figuras 4.14 y 4.15). Asimismo la volatilidad estimada (h_t) ajustada por este algoritmo se presenta en las siguientes figuras (Figuras 4.16 y 4.17).

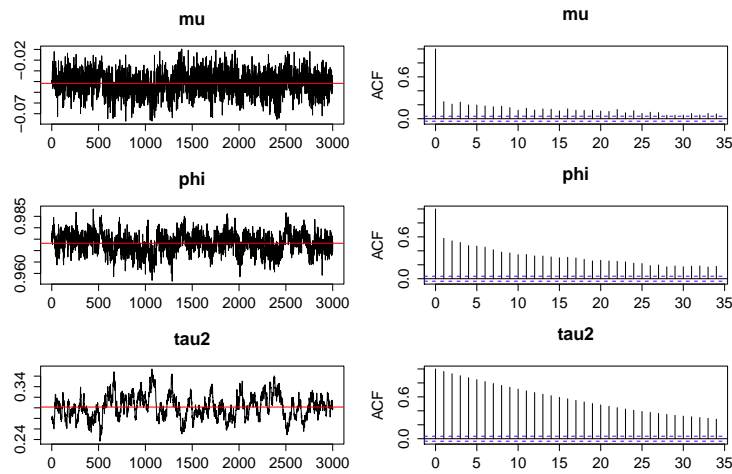


Figura 4.14: Evolución de los parámetros en cada iteración y función de correlación simple (Modelo SVFFBSM)

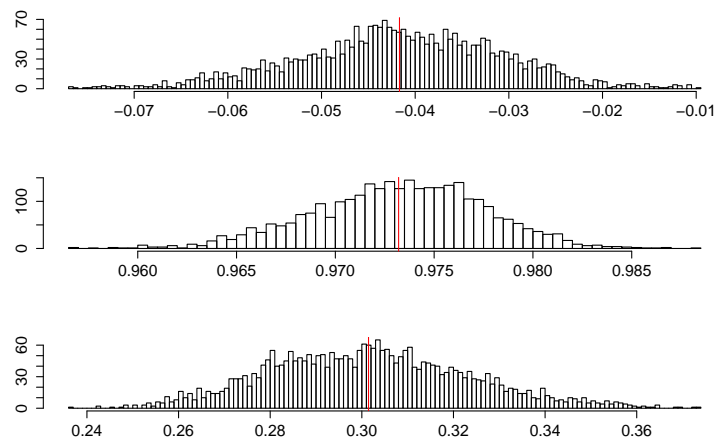


Figura 4.15: Histograma de los parámetros (Modelo SVFFBSM)

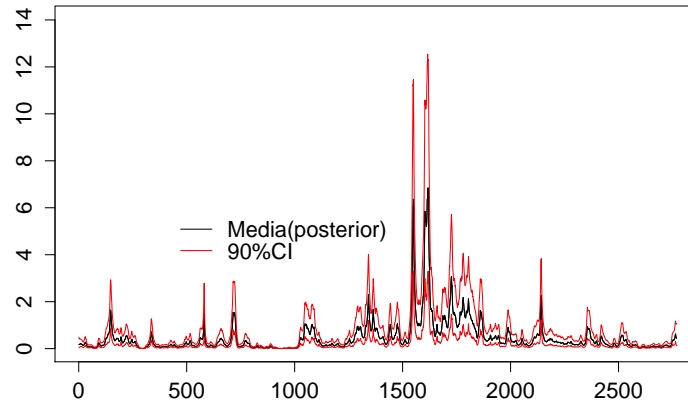


Figura 4.16: Volatilidad estimada de la TRM (Modelo SVFFBSM)

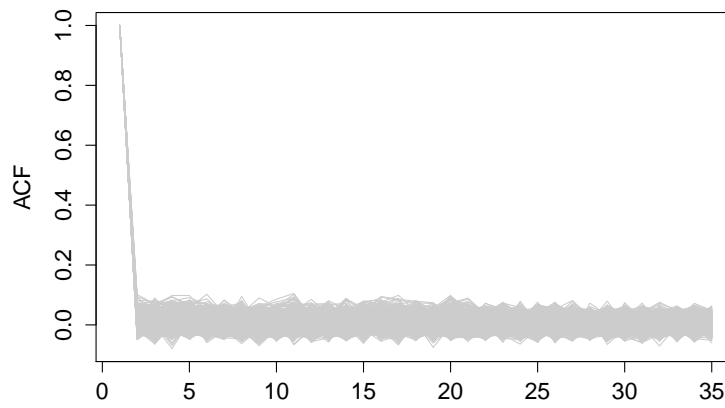


Figura 4.17: Correlación simple h_t (Modelo SVFFBSM)

4.2.2. Inferencia de la Posterior vía Monte Carlo Secuencial

En esta sección se muestra la aplicación del método de Monte Carlo Secuencial sobre la estimación del modelo SV (4.2), más conocido como filtro de partícula. Este usa los métodos de Monte Carlo, que incorporan dentro del algoritmo procesos de remuestreo, tal como el algoritmo SISR (Ver Sección 3.6.3). Con este último se busca remuestrear y reponderar secuencialmente realizaciones de una densidad de propagación. En general, el procedimiento se descompone en dos partes: *Predicción* y *Suavizamiento*, que en el caso del modelo SV (4.2), están dados por

$$p(h_t|y^{t-1}) = \int f_N(h_t; \mu + \phi h_{t-1}, \tau^2) p(h_{t-1}|y^{t-1}) dh_{t-1} \quad (4.13)$$

$$p(h_t|y^t) \propto p_N(y_t; 0, \exp\{h_t\}) p(h_t|y^{t-1}) \quad (4.14)$$

Teniendo en cuenta lo anterior, el objetivo del filtro de partículas es generar un conjunto de partículas, independientes e idénticamente distribuidas (*i.i.d*) $\{h_t^{(i)}\}_{i=1}^T$, que se aproximen a $p(h_t|y^t)$ mediante el uso de un conjunto de partículas *i.i.d* $\{h_{t-1}^{(i)}\}_{i=1}^T$ que aproximen $p(h_{t-1}|y^{t-1})$. Los filtros más populares que desarrollan estas aproximaciones son los filtros *Bootstrap* (BF) y el de *partículas auxiliar* (APF), los cuales se describieron en la sección 3.6.3 (Ver Gordon et al. [1993] y Pitt and Shephard [1999]). Como se discutió antes, estos algoritmos operan bajo el supuesto de parámetros conocidos, es decir que ya previamente se tienen los valores de (μ, ϕ, τ^2) .

Para modelar (4.2) y realizar el método Secuencial de Monte Carlo, con el problema asociado al desconocimiento de los parámetros (μ, ϕ, τ^2) , se desarrolló una técnica denominada Filtro con *Aprendizaje de Parámetros* (*Parameter learning*) (Ver Lopes and Tsay [2011]), en el cual se involucra un procedimiento secuencial de aprendizaje conjunto de h_t y $\theta = (\mu, \phi, \tau^2)$. La esencia de este tipo de filtros se basa superar el problema asociado con no remuestrear θ a través del tiempo, ya que, en general, en pocos pasos del tiempo el conjunto de partículas tiende a contenerse en una sola partícula. Gordon et al. [1993] sugirieron incorporar un ruido artificial en la evolución de θ para implementar la idea del aprendizaje. A continuación se implementa el Filtro de Liu y West (*LW filter*).

El filtro propuesto por Liu and West [2001] combina tres aspectos: i) el filtro APF (Sección 3.6.3), ii) una aproximación del núcleo suavizado para $p(\theta, y^{t-1})$ via una mixtura de normales multivariadas, y iii) un factor de contracción que incorpora la idea de una evolución artificial para θ sin pérdidas asociadas a la información. En particular, para el problema del modelo SV (ecuación 4.2), el propósito es encontrar un conjunto de partículas (*i.i.d*) $\{h_{t-1}^{(i)}, \theta_{t-1}^{(i)}\}_{i=1}^T$ que aproximen $p(h_{t-1}, \theta|y^{t-1})$ tal que

$$p(\theta|y^{t-1}) \approx \frac{1}{T} \sum_{j=1}^T p_N(\theta; m^{(j)}, V) \quad (4.15)$$

donde $m^{(j)} = a\theta_{t-1}^{(j)} + (1-a)\bar{\theta}$, $\bar{\theta} = \sum_{j=1}^T \theta_{t-1}^{(j)}/T$, $V = r^2 \sum_{j=1}^T (\theta_{t-1}^{(j)} - \bar{\theta})(\theta_{t-1}^{(j)} - \bar{\theta})'/T$, a es un factor de contracción y $r^2 = 1 - a^2$ es un factor de suavizamiento⁴. Siguiendo el filtro APF de Pitt and Shephard [1999], el problema de generar los estados y los parámetros (h_t, θ_t) como $p(h_t, h_{t-1}, \theta_t, \theta_{t-1}|y_t, y^{t-1})$ es descompuesto en dos pasos:

$$\begin{aligned} \text{Remuestreo: } & p(y_t|h_{t-1}, \theta_{t-1})p(h_{t-1}|\theta_{t-1}, y^{t-1})p(\theta_{t-1}|y^{t-1}) \\ \text{Propagación: } & p(h_t|h_{t-1}, \theta_t, y^t)p(\theta_t|\theta_{t-1}, y^t) \end{aligned} \quad (4.16)$$

Liu and West [2001], siguiendo los pasos de Pitt and Shephard [1999], proponen remuestrear a partir de la *propuesta (proposal)* $p(y_t|g(h_{t-1}, m(\theta_{t-1})))$, donde $g(\cdot)$ es usualmente el valor esperado, la mediana o la moda de $p(h_t|h_{t-1})$ y m evoluciona de acuerdo al factor de contracción y θ como se describió en el párrafo anterior. De esta forma, θ_t es muestreado a través de una transición artificial $p(\theta_t|\theta_{t-1})$ y h_t es muestreado de la densidad de evolución $p(h_t|h_{t-1}, \theta_t)$. Por último, las partículas propagadas (h_t, θ_t) están relacionadas con las ponderaciones ω_t proporcionales a $p(y_t|h_t, \theta_t)/p(y_t|g(h_{t-1}, m(\theta_{t-1})))$. Para comprender lo anterior, el algoritmo propuesto en el filtro LW está definido como sigue

⁴El desempeño del Filtro LW depende de la elección del parámetro de ajuste a , ya que éste se encuentra en el factor de contracción y de suavizamiento de la aproximación por medio de la Normal. En la práctica, muchos investigadores fijan este parámetro a alrededor de 0.98

Algoritmo Filtro de Liu y West (Con Aprendizaje de Parámetros)

1. Remuestrear $\{(\tilde{h}_{t-1}, \tilde{\theta}_{t-1})^{(i)}\}_{i=1}^T$ a partir de $\{(h_{t-1}, \theta_{t-1})^{(i)}\}_{i=1}^T$ con pesos $\omega_t^{(i)} \propto p_N(y_t; 0, \exp\{\tilde{m}_0^{(i)} + \tilde{m}_1^{(i)} h_{t-1}^{(i)}\})$ donde $\tilde{m}_0^{(i)}$ y $\tilde{m}_1^{(i)}$ se definen de acuerdo a la ecuación 4.15.
2. Propagar
 - a) $\{\tilde{\theta}_{t-1}^{(i)}\}_{i=1}^T$ para $\{\hat{\theta}_t^{(i)}\}_{i=1}^T$ a partir de $N(\tilde{m}^{(i)}, V)$, y luego
 - b) $\{\tilde{h}_{t-1}^{(i)}\}_{i=1}^T$ para $\{\hat{h}_t^{(i)}\}_{i=1}^T$ a partir de $p_N(h_t; \hat{\mu}^{(i)} + \hat{\phi}^{(i)} \tilde{h}_{t-1}^{(i)}, \hat{\tau}^{2(i)})$
3. Remuestrear $\{(h_t, \theta_t)^{(i)}\}_{i=1}^T$ a partir de $\{(\hat{h}_t, \hat{\theta}_t)^{(i)}\}_{i=1}^T$ con pesos $\omega_t^{(i)} \propto \frac{p_N(y_t; 0, \exp\{\hat{h}_t^{(i)}\})}{p_N(y_t; 0, \exp\{\tilde{m}_0^{(i)} + \tilde{m}_1^{(i)} \tilde{h}_{t-1}^{(i)}\})}$

Como resultado se obtuvo la siguiente estimación de los parámetros del modelo SV con el algoritmo del filtro LW con aprendizaje de parámetros (SVSMC). Se aprecia que la volatilidad tiene una persistencia alta, dado que el parámetro ϕ está cercano a uno.

Parámetro	μ	ϕ	τ^2
Estimador	-0.024818	0.9893292	0.061014
Des. est.	(0.020068)	(0.022877)	(0.014165)

Cuadro 4.7: Parámetros estimados para el SV(1) (SVSMC)

Para cada una de las simulaciones se pueden apreciar la dinámica de los parámetros (Figura 4.18). Asimismo la volatilidad estimada (h_t) ajustada por este algoritmo se presenta en la figura 4.19.

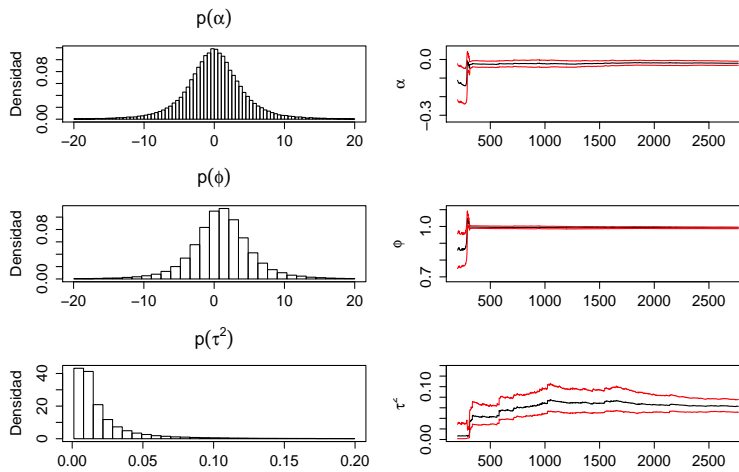


Figura 4.18: Distribución Prior y evolución de los parámetros en cada iteración (Modelo SVSMC)

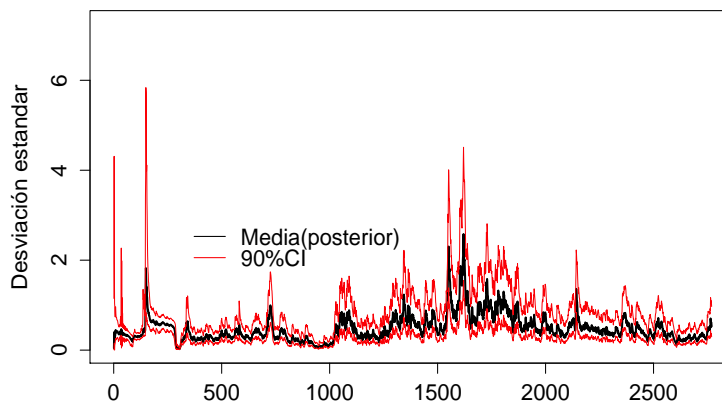


Figura 4.19: Volatilidad estimada de la TRM (Modelo SVSMC)

4.3. Diagnóstico de los modelos

La mayoría de modelos logra replicar los principales hechos estilizados de la tasa de cambio peso-dólar: i) existen conglomerados de volatilidad, ii) los retornos al cuadrado están correlacionados, lo cual se refleja en la estructura de correlación de h_t y iii) los retornos tienen colas pesadas. De acuerdo a los sistemas que rigen la dinámica de los retornos y la volatilidad (ecuaciones 4.1 y 4.2) se establecieron unos supuestos relacionados con las distribuciones asumidas por las perturbaciones. Allí, se planteó el uso de la distribución normal para caracterizar tanto el modelo de heteroscedasticidad como

el modelo de volatilidad estocástica.

Por ello, una forma de verificar algún tipo de diagnóstico sobre el ajuste de los modelos, siguiendo a Choi [2011], es evaluando si al estandarizar los retornos con la medida de volatilidad obtenida⁵, dicha medida se comporta como una distribución Normal. Es decir, el retorno estandarizado es igual a $\frac{y_t}{\hat{\sigma}_t}$ donde $\hat{\sigma}_t = \exp\{\frac{h_t}{2}\}$.

Bajo esta consideración, se verificó el supuesto de Normalidad sobre la medida estandarizada asociada al retorno $(\frac{y_t}{\hat{\sigma}_t})$.

Modelo	Jarque-Bera P-Valor	Shapiro-Wilk P-valor
SVMRW	2.2e-16	1.037e-10
SVMHI	2.2e-16	1.087e-08
SVFBBS	2.2e-16	2.2e-16
SVFBBSM	2.2e-16	1.456e-08
SVSMC	0.0004622	1.409e-05
GARCH	2.638e-12	3.559e-06

H_0 : Hay Normalidad vs. H_a : No hay Normalidad

Cuadro 4.8: Pruebas de Normalidad para las medidas estandarizadas

Aunque ningún modelo cumple con el supuesto de normalidad, se puede apreciar que a partir de las volatilidades estimadas se logra modelar en parte los excesos de kurtosis, que en la serie original era de 9.09. En la siguiente tabla se observa como estos excesos se reducen mediante la estandarización de los retornos ó, en su defecto, del residual de la ecuación de la media que se obtiene al dividirlo por las estimaciones realizadas de la volatilidad en cada uno de los modelos. Allí, se obtienen valores para los excesos de kurtosis mucho menores. Frente a esta debilidad, un paso a seguir en futuras investigaciones, sería incorporar el uso de otras distribuciones que tengan colas más pesadas como la *t-student* o la *GED*, con lo que se podría mejorar el ajuste frente a la presencia de eventos extremos que afectan el peso de las colas de la distribución empírica de los retornos estandarizados.

Pese a todo esto, el resultado de este trabajo de tesis es satisfactorio en la medida que se logra corregir en alguna medida ciertos problemas que la serie original presenta en el modelamiento tradicional de series de tiempo, como son los excesos de kurtosis, conglomerados de volatilidad y la varianza condicional no constante a lo largo del tiempo.

⁵Esto si los retorno siguen un mercado eficiente. En caso contrario, una posibilidad es utilizar una estructura ARIMA para la media y calcular el residual de dicho modelo. Teniendo esto, se estandariza el residual al dividirlo por la volatilidad estimada. Así se hizo para el caso de la tasa de cambio peso-dólar.

Modelo	media	asimetría	Exceso kurtosis*
serie original	-0.01	0.19	9.09
SVMRW	-0.04	0.01	1.59
SVMHI	-0.04	0.02	1.25
SVFBBS	-0.04	-0.09	3.25
SVFBBSM	-0.04	0.03	1.23
SVSMC	-0.01	0.05	-0.48
GARCH	0	0.14	0.85

* El programa escala la medida comparando con la normal, es decir al valor registrado originalmente le resta 3, que es la la kurtosis de la distribución normal.

Cuadro 4.9: Estadísticas descriptivas de los retornos de la TRM y las medidas estandarizadas.

La mejoría en términos de modelar los excesos de kurtosis también se puede observar al comparar la distribución empírica de los retornos estandarizados frente a la serie original de los retornos de la TRM, los cuales fueron presentados en el capítulo 1.

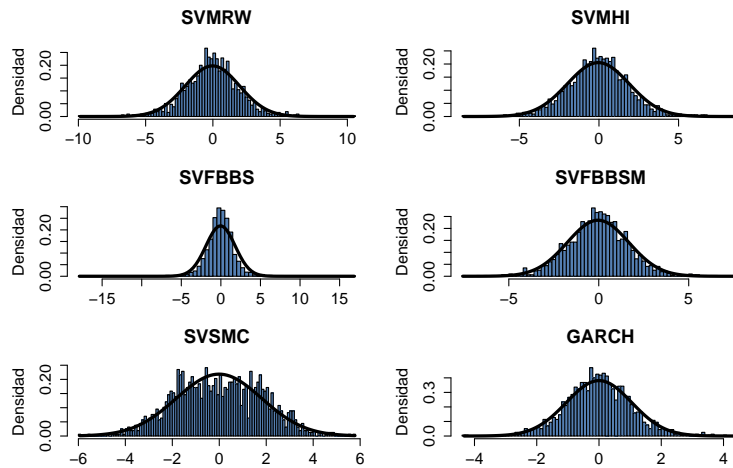


Figura 4.20: Distribución de los retornos estandarizados (Todos los modelos)

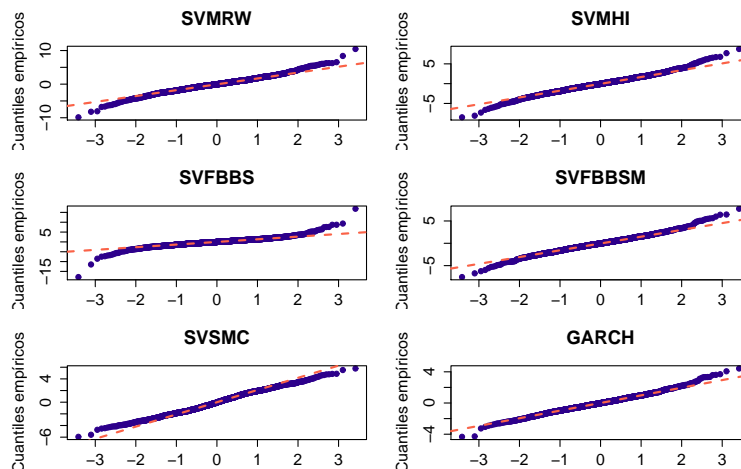


Figura 4.21: QQ-plot (Todos los modelos)

4.4. Aplicaciones en Economía

El objetivo principal de la tesis era estimar la volatilidad de la tasa de cambio peso-dólar en el período comprendido entre enero de 2002 y junio de 2013, mediante el uso de modelos de volatilidad estocástica. Luego, uno de los propósitos era comparar los resultados con respecto a otras posibles mediciones de la varianza como lo son las estimaciones de modelos tradicionales de series de tiempo con heteroscedasticidad condicional (familia GARCH), éstas últimas son las más utilizadas por la facilidad en su implementación computacional. A continuación, se busca contextualizar al lector sobre la utilización de dichas estimaciones en áreas como la economía y las finanzas. En especial, en la segunda parte se presenta una aplicación ampliamente utilizada en el campo de la medición de riesgo de mercado como lo es el Valor en Riesgo (*Value at Risk- VaR*).

4.4.1. Volatilidad del mercado: Comparación entre las diferentes estimaciones.

Como se aprecia en la figura (4.22), el modelo de heteroscedasticidad condicional (GARCH(1,1)) tiende a arrojar una estimación de volatilidad de la tasa de cambio peso-dólar mayor que la reportada por los modelos de volatilidad estocástica. Sin embargo, no se puede afirmar cuál de las estimaciones presenta el mejor ajuste puesto que la volatilidad es una variable *latente*, es decir no es observable, y por lo tanto, no se puede realizar evaluación de pronóstico. Usualmente, se utilizan medidas de evaluación de pronóstico para ver el ajuste del modelo fuera de muestra, las cuales comparan la proyección frente a la observación en cada momento del tiempo. Con ello se busca concluir si el modelo estimado replica el comportamiento de la variable objetivo del estudio.

Pese a las limitaciones para determinar el mejor ajuste, es relevante contextualizar al lector sobre la principal utilidad de las estimaciones presentadas. Al comparar diversas mediciones realizadas para la volatilidad de la tasa de cambio, se observa como tienen un comportamiento similar. En especial, todas ellas aumentan considerablemente en períodos de alta incertidumbre y turbulencia financiera. Asimismo, se aprecia que las estimaciones ratifican que la volatilidad cambia en el tiempo, presentando períodos tanto de alta como de baja variabilidad; lo que está asociado con el hecho estilizado de los conglomerados de volatilidad (Ver capítulo 1).

El período de mayor volatilidad de la muestra se presenta en el último trimestre de 2008, cuando la variabilidad diaria de la tasa de cambio peso-dólar alcanzó valores entre 5,5 % y 7 %. Esto se explica por la crisis financiera internacional que se presentó en dicho período, en donde de acuerdo con Toro et al. [2013] la turbulencia financiera originada en Estados Unidos (EEUU), se expandió mediante diversos canales a nivel mundial. Todo ello causó una disminución generalizada en los precios de las acciones, bonos y monedas, de tal forma que el pánico entre los inversionistas produjo gran volatilidad en los mercados financieros. Los modelos presentados en esta tesis permiten cuantificar la volatilidad, por lo que podrían servir de ayuda para responder el tipo de interrogantes que se acaba de plantear. Asimismo, estos modelos se convierten en herramientas útiles para el seguimiento del mercado cambiario.

Otros eventos de alta volatilidad, los cuales se aprecian en la gráfica son: i) entre finales de 2010 y comienzos de 2011, se observa de nuevo un período de alta volatilidad con respecto a períodos de calma, esto se debe a las presiones que ejercía la incertidumbre sobre la deuda soberana de países europeos y la posible recesión que ellos presentarían, ii) En el segundo trimestre de 2013 cuando existía gran incertidumbre sobre la continuación de la flexibilidad monetaria por parte de la Reserva Federal y los debates sobre los límites de la deuda pública en EEUU. En entornos como éstos, el contar con

estimaciones de dicha volatilidad permite a las autoridades económicas (como el Banco de la República) intervenir y regular el mercado cambiario, con el fin de reducir la variabilidad que afecta las decisiones de inversión y gasto de empresas y hogares. De la misma manera, para inversionistas privados estas mediciones cobran importancia al permitir cuantificaciones del riesgo de mercado al que están expuestos, y por lo tanto, generar cubrimientos a través de instrumentos como los contratos a futuro y opciones de compra o venta de divisas.

Volatilidades estimadas (TRM)

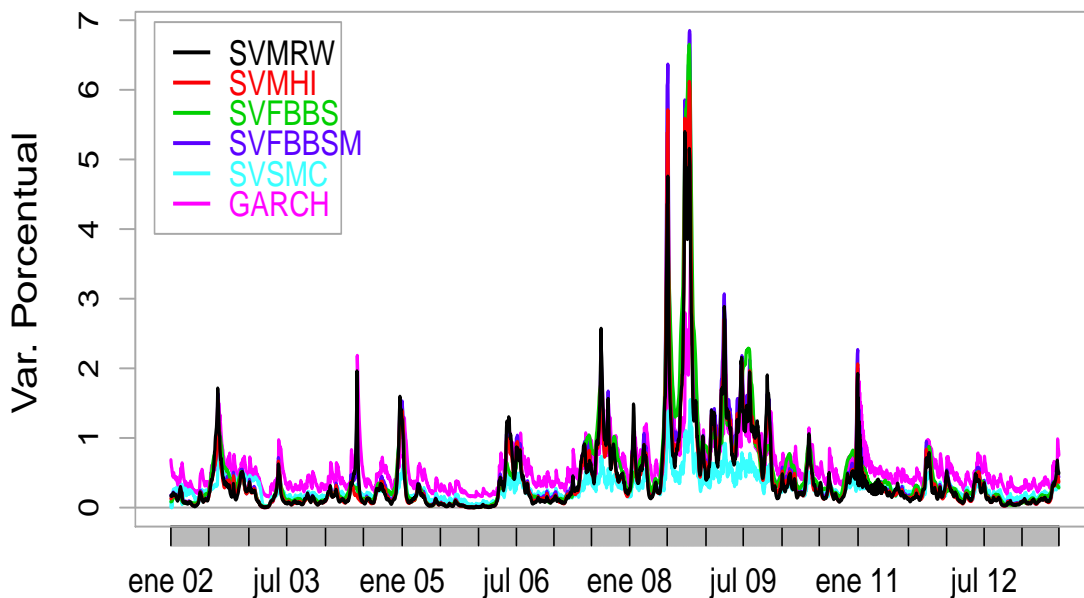


Figura 4.22: Volatilidad estimada de la TRM (Todos los modelos)

Pronósticos fuera de muestra

Como se mencionó en el capítulo 1, para un exportador es conveniente tener una estimación de valores futuros de la tasa de cambio peso-dólar, puesto que con esa previsión podría cubrir sus operaciones y transacciones con la divisa en el futuro. Cabe señalar que dado que la tasa de cambio sigue un proceso estocástico, la incertidumbre a futuro siempre existirá y no necesariamente el ajuste de un modelo permite conocer

con exactitud los valores que tomará la variable días o semanas después.

A continuación, dadas las estimaciones de los parámetros realizadas mediante el uso de los modelos presentados en los capítulos anteriores, se pronostican fuera de muestra el valor de los retornos de la tasa de cambio peso-dólar y su volatilidad un paso adelante. Para ello se realizó el procedimiento de pronóstico para 41 observaciones fuera de muestra entre el 2 de julio de 2013 y el 30 de agosto de 2013 con cada uno de los modelos propuestos. También se incorporó la estimación de la varianza muestral, a partir de los datos intra-día con intervalos de tiempo de 10 minutos (*DES_EST*).

Como resultado, fuera de muestra, el modelo GARCH tiende a sobreestimar la volatilidad con respecto a las estimaciones realizadas de los modelos de volatilidad estocástica. Asimismo, se observa que utilizar la estimación muestral de la varianza intra-día puede conducir a subestimaciones de la volatilidad de la tasa de cambio peso-dólar (Figura 4.23).

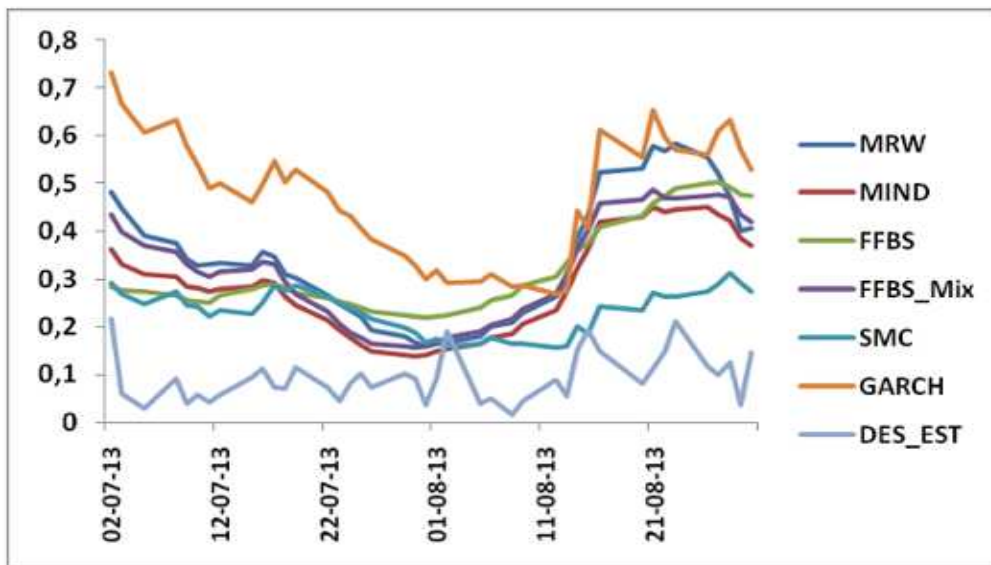


Figura 4.23: Pronósticos de la volatilidad diaria un paso adelante (Variaciones porcentual)

4.4.2. Medición del Valor en Riesgo (Value at Risk).

Una aplicación ampliamente utilizada en finanzas para la medición de riesgo de mercado es el Valor en Riesgo, el cual es una medida de la incertidumbre en el valor futuro de un portafolio. El Valor en riesgo se utiliza para medir y cuantificar el nivel de riesgo financiero de un activo o portafolio de inversión durante un periodo específico de tiempo, sus siglas responden a las palabras en inglés, *Value at Risk*.

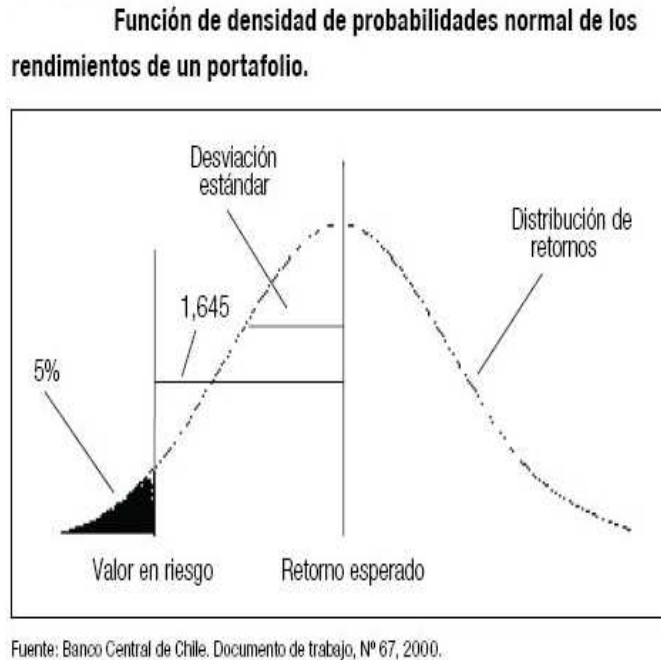


Figura 4.24: Valor en Riesgo

Siguiendo a Jorion [2003] y Alexander [2008b], el *VaR* describe las posibles pérdidas mediante el cuantil de la distribución proyectada de los retornos de un activo durante un horizonte de tiempo específico (k) y un nivel de confianza (α). De esta forma, el *VaR* puede expresarse como $VaR_{t+k|t}^\alpha$ y corresponde a la mínima pérdida en el $(1-\alpha) \times 100\%$ de los peores casos. Generalmente, $\alpha = 95\%$, por lo que el 5% restante representa el conjunto de las máximas pérdidas posibles que está dispuesto a asumir el inversionista (área sombreada en la figura 4.24). Así, el *VaR* nace con el propósito de guiar al inversionista en el manejo de la volatilidad de sus inversiones y para cuantificar su nivel de exposición al riesgo de mercado, es decir a las posibles pérdidas que puede asumir por las variaciones en los precios de los activos. En particular, si se evalúa el *VaR* sobre la tasa de cambio, se está midiendo el riesgo cambiario al que se expone un inversionista extranjero que invierte en pesos o un colombiano que invierte en activos en dólares.

Teniendo en cuenta esto, es común utilizar el supuesto de normalidad de los retornos

para calcular el Valor en Riesgo. De acuerdo con Melo and Becerra [2005], su cálculo puede ser derivado de la siguiente manera:

$$\begin{aligned}
 P(y_t \geq VaR_t) &= P\left(\frac{y_t - \mu}{\sigma} \geq \frac{VaR_t - \mu}{\sigma}\right) = 1 - \alpha \\
 \frac{VaR_t - \mu}{\sigma} &= \Phi^{-1}(\alpha) = z_\alpha \\
 VaR_t &= \mu + z_\alpha \sigma
 \end{aligned} \tag{4.17}$$

donde $\Phi^{-1}(\alpha)$ es la función inversa de la distribución normal acumulada y z_α es el α -ésimo cuantil de la distribución normal. Con todo ello, los autores muestran que al utilizar modelos de volatilidad variante en el tiempo (familia GARCH y volatilidad estocástica vistos en capítulos anteriores), el cálculo del Valor en Riesgo para un período de un día ($k = 1$) está definido como:

$$VaR_{t+1} = \mu_{t+1} + z_\alpha \sigma_{t+1} \tag{4.18}$$

donde μ_{t+1} es el pronóstico un paso adelante fuera de muestra que proviene del modelo ARIMA seleccionado en el capítulo 1 ($ARIMA(1, 0, 2)$) y σ_{t+1} es la proyección que se obtiene un paso adelante fuera de muestra en cada uno de los modelos estimados en esta tesis.

En la figura (4.25) se compara el VaR estimado para cada una de las estimaciones propuestas y la realización real de la tasa de cambio peso-dólar observada. Los puntos en cada gráfica implican que, en ese día la pérdida real superó la estimación del VaR , en otras palabras el inversionista (exportador, en este caso) sufrió una pérdida mayor a la que esperaba.

Cabe resaltar que si se utiliza simplemente la media y varianza muestral del día hoy para pronosticar el VaR del día de mañana, dicha estimación tiende a subestimar significativamente el riesgo (Ver *DES_EST*), por lo que el inversionista puede incurrir en mayores riesgos. Esto puede estar explicado por la ausencia de una estructura de correlación que permita modelar la persistencia de la volatilidad. Lo anterior, muestra la relevancia del presente estudio ya que se puede contar con diversas medidas que permiten modelar correctamente el riesgo a partir de la estimación de la media (μ_t) y varianzas (σ_t) que cambian en el tiempo. Estas dependen de una estructura de correlación temporal.

Asimismo, a continuación se muestra el porcentaje de error de cada una de las metodologías expuestas, es decir, se cuantifica el número de veces que el retorno real superó la estimación realizada para el Valor en Riesgo de la tasa de cambio para un horizonte de un día. Esto se realiza con los 41 pronósticos realizados fuera de muestra. Aunque la mayoría de modelos no cumple con uno de los criterios estadísticos usuales, que los errores o fallos sean menores al 5 %, se observa que al utilizar modelos econométricos mejora el porcentaje de fallo respecto a realizar las proyecciones con estimadores muestrales de media y varianza (*DES_EST*, 34,15 %).

Modelo	Número de Fallos	Porcentaje de fallo
SVMRW	4	9.76 %
SVMHI	4	9.76 %
SVFBBS	3	7.32 %
SVFBBSM	3	7.32 %
SVSMC	6	14.63 %
GARCH	2	4.88 %
DES_EST	14	34.15 %

Cuadro 4.10: Porcentajes de fallo por metodología.

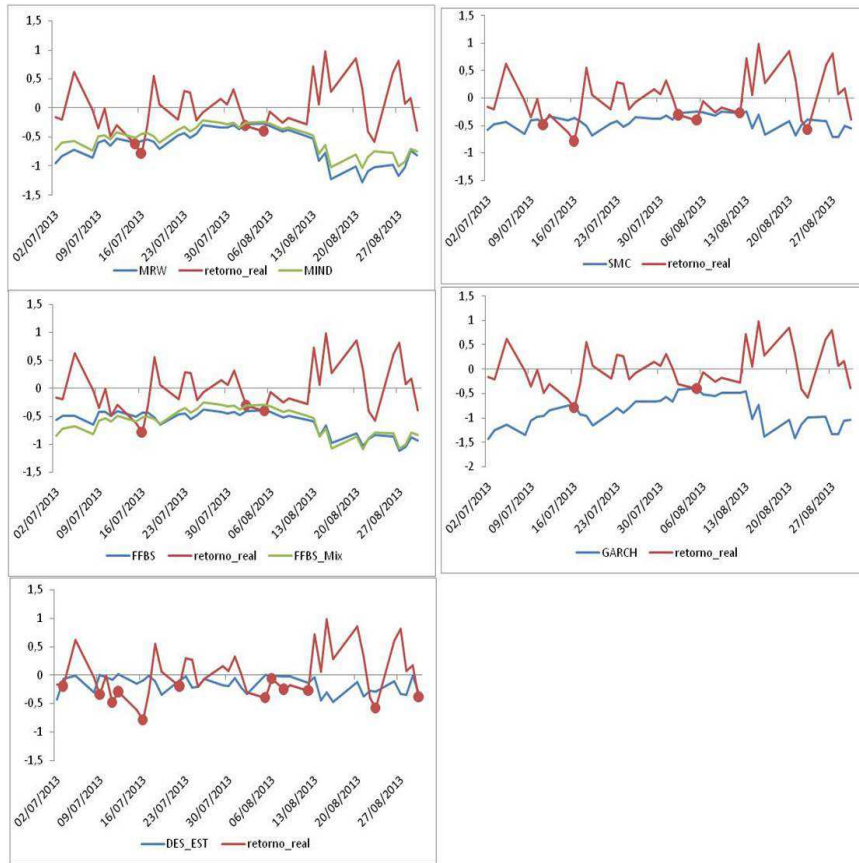


Figura 4.25: Evaluación de la medición de riesgo- Fuera de muestra

Conclusiones

En este trabajo de tesis se describieron teóricamente dos tipos de modelos para estimar volatilidad, el modelo de heteroscedasticidad condicional (GARCH) de Bollerslev [1986] y el de volatilidad estocástica (SV) de Taylor [1982], los cuales son usualmente utilizados en el campo de las finanzas. Se enunciaron las principales propiedades de cada uno de ellos y se realizó una aplicación para la estimación de la volatilidad de la tasa de cambio peso-dólar en el periodo comprendido entre enero de 2002 y junio de 2013, con información diaria.

Desde un punto de vista empírico, el mayor atractivo de los modelos de volatilidad estocástica (SV) es la flexibilidad para presentar los momentos muestrales que se presentan habitualmente en los retornos de activos financieros, en otras palabras replican de manera adecuada la mayoría de hechos estabilizados mencionados en el capítulo 1.

Como se discutió a lo largo del documento, el modelo SV tiene las siguientes ventajas sobre el tradicional GARCH: i) son aproximaciones discretas de los modelos que usualmente se utilizan en finanzas, por lo que luego de realizar las estimaciones los valores de los parámetros pueden ser incorporados en modelos de precios activos, ii) incorpora las noticias y choques nuevos directamente sobre la volatilidad (Clark [1973]), es decir, agrega un componente estocástico a la ecuación de comportamiento de la volatilidad y iii) permite generar distribuciones sobre los parámetros del modelo, lo cual podría ser utilizado en la inferencia estadística.

Los principales resultados encontrados en la aplicación realizada en el capítulo 4 son: i) la volatilidad de la tasa de cambio peso-dólar presenta alta persistencia; esto lo sugiere el hecho de que en cada uno de los modelos el coeficiente asociado al rezago de ella misma en la ecuación que dirige su dinámica es cercano a uno, ii) el estimador de la persistencia de la volatilidad es mayor en el modelo SV que en el GARCH, ($\beta_1=0.82$ en el caso del GARCH frente a $\phi =0.97$ en promedio para las medidas del SV), iii) los modelos replican los hechos estilizados de las series financieras, en particular, conglomerados de volatilidad, correlaciones de los retornos al cuadrado y excesos de kurtosis, iv) pese a que ningún modelo cumple con el supuesto de normalidad, al utilizar las estimaciones se logra reducir en alguna medida los excesos de kurtosis (el modelo que mejor se ajustó fue el (SVSMC), v) el modelo GARCH tiende a sobreestimar la volatilidad respecto a las estimaciones realizadas por SV, y vi) al pronosticar

fuera de muestra, el modelo GARCH continúa siendo el modelo que reporta mayores valores para la volatilidad, lo cual, en principio, podría implicar una sobrestimación del riesgo cambiario en Colombia mediante el uso de modelos GARCH. Sin embargo, al utilizar la metodología de Valor en Riesgo (*VaR*), se aprecia que mediante el uso del GARCH, se obtienen estimaciones que no son desfavorables para el riesgo del mercado derivado por las fluctuaciones de la tasa de cambio peso-dólar, como era de esperarse dado que sobreestima la varianza.

En este trabajo de tesis se estudió el caso simple de los modelos de volatilidad estocástica, un proceso autoregresivo de orden uno, en el cual existe simetría en la volatilidad. Sin embargo, en la práctica se aprecia que algunas divisas y activos financieros presentan una correlación temporal mayor, por lo que podría extenderse en modelo para un orden de rezagos más alto. Asimismo, en la realidad existen asimetrías en el comportamiento de los retornos financieros, el denominado efecto apalancamiento. Este sucede porque la volatilidad responde en la mayoría de casos de forma asimétrica, en especial cuando hay una noticia desfavorable los inversionistas sobreaccionan generando un impacto mayor en términos de la variabilidad del activo. Por otro lado, cuando la noticia es favorable, esto afecta la volatilidad pero en menor medida.

Por último, el uso de información diaria y la alta volatilidad observada en los retornos financieros hacen que los excesos de kurtosis sean elevados, por lo que el supuesto de normalidad podría relajarse mediante la utilización de otras funciones de distribución de colas más pesadas. Esto permitiría un mejor ajuste en las estimaciones de este tipo de modelos. El tratamiento de estos temas planteados podría convertirse en una extensión de esta tesis.

Apéndice A

Conceptos Básicos de probabilidad y matemáticas

En esta sección se presentan varios conceptos con el fin de brindar al lector algunas nociones básicas y diferentes notaciones estadísticas, matemáticas, econométricas y financieras que se utilizaron a lo largo del documento. Para ello se tomó como guía los siguientes libros de texto: Blanco [2004], Guerrero [2003], Peña [2005], Venegas [2007] y Alexander [2008a].

A.1. Introducción

Las matemáticas y la estadística son una herramienta de trabajo indispensable para el diseño de modelos en el ámbito de las ciencias económicas, las finanzas, los negocios y la dirección de empresas. Tanto su lenguaje como sus técnicas permiten afrontar problemas que, de otra manera, serían difíciles de resolver. Por ejemplo, en el caso de la tasa de cambio peso-dólar, existen diversas teorías que buscan responder cuáles son sus determinantes, qué explica su comportamiento, cómo es su relación con el conjunto de la economía de un país, entre otras. Los primeros modelos desarrollados para contestar dichas preguntas modelaban el problema mediante representaciones determinísticas, centrando su atención en la aplicación de las matemáticas al razonamiento deductivo, y no del estudio inductivo. Estos avances se dieron en el campo de lo que se conoce como la economía matemática, la cual se refiere a la aplicación de las matemáticas a los aspectos puramente teóricos del análisis económico.

Una referencia obligada en este campo, ya que es uno de los escritos más famosos en economía sobre el comportamiento de la tasa de cambio, se encuentra en Dornbusch [1976]. Allí, el autor ofreció una explicación a la dinámica errática que siguió el dólar después de 1973 con respecto a las divisas más importantes de Europa. El modelo que propuso fue un sistema de ecuaciones diferenciales para los precios internos y la tasa

de cambio. Como resultado de la modelación obtuvo que en el corto plazo una expansión monetaria eleva los precios internos y disminuye las tasas de interés reales, lo cual presiona una salida de capital que induce a una depreciación inmediata de la moneda nacional (sobre-reacción), y una posterior fluctuación de ésta y de los términos de intercambio.

Sin embargo, en la economía y otras ciencias sociales, se presentan fenómenos repentinos por lo que se requiere utilizar variables aleatorias en la modelación, lo que a su vez exige que el modelo a desarrollar sea estocástico. En el caso de la tasa de cambio, se tiene que factores políticos y económicos, tanto internos como externos, afectan su dinámica en el tiempo. Frente a este aspecto en la literatura se pueden encontrar aplicaciones del cálculo estocástico para la implementación de opciones sobre la tasa de cambio (Hull [2008]), así como el desarrollo reciente de modelos de equilibrio general dinámico estocásticos (DSGE), que pueden reconocer la dinámica de esta variable junto con el conjunto de la economía. También se encuentran otros trabajos en los cuáles se calcula la volatilidad por medio de herramientas estadísticas. En este último caso, se trata del estudio de observaciones empíricas por métodos estadísticos de estimación y prueba de hipótesis, temas que se podrían ubicar dentro del marco de referencia de la econometría. Esta tesis utilizó en el estudio teórico del comportamiento de la tasa de cambio peso-dólar herramientas de modelación matemática y econométrica que permitieron formular un modelo que replicó las características de los datos empíricos. Los conceptos básicos detrás de los modelos presentados se enuncian a continuación.

A.2. Conceptos básicos de matemáticas y probabilidad

A.2.1. Variable aleatoria

Antes de introducir el concepto de variable aleatoria se deben considerar las siguientes definiciones:

Definición A.1. El conjunto de Ω de todos los posibles resultados de un experimento aleatorio se llama **espacio muestral**. Los elementos $\omega \in \Omega$ son llamados puntos muestrales o eventos.

Definición A.2. Sea $\Omega \neq \Phi$. Una colección \mathfrak{F} de subconjuntos de Ω es una **σ -álgebra** sobre Ω , si:

- I) $\Omega \in \mathfrak{F}$.
- II) Si $A \in \mathfrak{F}$ entonces $A^c \in \mathfrak{F}$.
- III) Si $A_1, A_2, \dots \in \mathfrak{F}$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}$.

Los elementos de \mathfrak{F} se llaman eventos.

Definición A.3. Sean $\Omega \neq \Phi$ y \mathfrak{F} una σ -álgebra sobre Ω . La pareja (Ω, \mathfrak{F}) se llama **espacio medible**.

Definición A.4. Sea (Ω, \mathfrak{F}) un espacio medible. Una función P definida sobre \mathfrak{F} y de valor real que satisface las siguientes condiciones se llama medida de probabilidad sobre (Ω, P) :

- I) $P(A) \geq 0$ para todo $A \in \mathfrak{F}$.
- II) $P(\Omega) = 1$.
- III) Si A_1, A_2, \dots son elementos de \mathfrak{F} mutuamente excluyentes, esto es que $A_i \cap A_j = \Phi$ para todo $i \neq j$, entonces $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

La tripla $(\Omega, \mathfrak{F}, P)$ se llama **espacio de probabilidad**.

Definición A.5. La menor σ -álgebra sobre \mathbb{R} que contiene todos los intervalos de la forma $(-\infty, a]$ con $a \in \mathbb{R}$ se llama **σ -álgebra de Borel** y se denota como \mathfrak{B} .

Puesto que \mathfrak{B} es una σ -álgebra entonces los siguientes conjuntos también pertenecen a \mathfrak{B}

$$\begin{aligned} (a, \infty) &= \mathbb{R} - (-\infty, a], \\ (a, b] &= (-\infty, b] \cap (a, \infty), \\ (-\infty, a) &= \bigcup_{n=1}^{\infty} (-\infty, a - \frac{1}{n}], \\ [a, \infty) &= \mathbb{R} - (-\infty, a), \\ (a, b) &= (-\infty, b) \cap (a, \infty), \\ [a, b) &= \mathbb{R} - ((-\infty, a) \cup (b, \infty)), \\ \{a\} &= [a, a], \\ \mathbb{N} &= \bigcup_{n=0}^{\infty} \{n\}, \\ \mathbb{Q} &= \bigcup_{r \in \mathbb{Q}} \{r\}, \\ \mathbb{Q}^c &= \mathbb{R} - \mathbb{Q} \end{aligned}$$

donde $a, b \in \mathbb{R}$ y $a < b$.

En un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ los elementos del espacio muestral Ω no tienen que ser números, pero la asignación de valores numéricos a los eventos ω dentro

del espacio de probabilidad es la clave para definir el concepto de variable aleatoria. De esta forma, una aplicación X definida sobre el espacio muestral Ω que fije a cada evento ω un valor numérico $X(\omega)$ es una variable aleatoria. Dado esto, y utilizando el concepto de σ -álgebra de Borel, se puede generalizar la definición de variable aleatoria y en particularizar cuando ésta es real.

Definición A.6. Sean $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad y (Ω', \mathfrak{F}') un espacio medible. Una \mathfrak{F} - \mathfrak{F}' -variable aleatoria es una aplicación $X: \Omega \rightarrow \Omega'$ tal que, para todo $A \in \mathfrak{F}'$ se tiene que $X^{-1}(A) \in \mathfrak{F}$.

Si $(\Omega', \mathfrak{F}') = (\mathbb{R}, \mathcal{B})$, entonces, se dice que X es una variable aleatoria real.

Teniendo en cuenta lo anterior, podemos decir que una variable aleatoria es una variable cuyos valores son estocásticos. Esto significa que existe incertidumbre sobre los valores que ésta puede tomar, y que cada uno de los posibles eventos tiene asociado una probabilidad de ocurrencia. Todo ello podrá verse a través de la distribución de la variable aleatoria.

Teorema A.1. Suponga que X es una variable aleatoria definida sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ y con valores en el espacio medible (Ω', \mathfrak{F}') . La función P_X definida sobre la σ -álgebra \mathfrak{F}' por medio de :

$$P_X(B) = P(\{X \in B\})$$

para todo $B \in \mathfrak{F}'$, es una medida de probabilidad sobre (Ω', \mathfrak{F}') llamada distribución de la variable aleatoria X .

Una vez establecido lo anterior, se define la probabilidad condicional y se enuncia la regla de Bayes, en el cual están basados los algoritmos de inferencia bayesiana que se mostraron en el capítulo 3.

Definición A.7. Sean $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad. si $A, B \in \mathfrak{F}$ con $P(A) > 0$, entonces se define la probabilidad del evento B bajo la condición A como sigue:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

A esto se denomina la *probabilidad condicionada*, en este caso la probabilidad condicionada de B dado A .

Teorema A.2. teorema de probabilidad total Sea A_1, A_2, \dots , una partición finita o numerable de Ω , es decir $A_i \cap A_j = \emptyset$ para todo $i \neq j$ y $\bigcup_{i=1}^{\infty} A_i = \Omega$, tal que $P(A_i) > 0$ para todo i . Entonces para cualquier $B \in \mathfrak{F}$ se satisface:

$$P(B) = \sum_n P(B|A_n)P(A_n)$$

Como **corolario** se obtiene el resultado conocido como **Regla de Bayes**. Sea A_1, A_2, \dots una partición finita o numerable de Ω con $P(A_i) > 0$ para todo i , entonces se satisface para todo $B \in \mathfrak{F}$ con $P(B) > 0$ lo siguiente:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_j P(B|A_j)P(A_j)} \quad \text{para todo } i$$

A.2.2. Funciones de distribución de probabilidad y de densidad de probabilidad

Una realización (también llamada a menudo observación) de una variable aleatoria X puede ser pensada como un número que está asociado con un cambio de resultado, donde cada uno de ellos es determinado por un cambio de evento y tiene una medida de probabilidad asociada. Así, el conjunto de todos los posibles resultados y sus probabilidades asociadas son llamadas **medidas de probabilidad**.

Una vía para representar una medida de probabilidad de una variable aleatoria es a través de la función de distribución de probabilidad. Esta función asigna la probabilidad que X tome un valor menor o igual a x . Algunas veces también se le denomina función de distribución de probabilidad acumulada. Formalmente, su definición general es:

Definición A.8. Sea X una variable aleatoria real. La función F_X definida sobre \mathbb{R} por medio de:

$$\begin{aligned} F_X(x) &= P_X((-\infty, x]) \\ &= P(X \leq x) \end{aligned}$$

se llama función de distribución de la variable aleatoria X .

Otra forma de representar una medida de probabilidad es a través de la función de densidad de la variable aleatoria, la cual se utiliza en estadística con el propósito de conocer cómo se distribuyen las probabilidades de un suceso o evento, en relación al resultado del suceso. Este último concepto es más complicado de definir ya que depende de la característica de la variable, si es discreta o continua. Por ejemplo, en el caso

de las discretas¹ la función de densidad puede ser definida como la probabilidad de observar un valor particular.

Por el contrario, en el caso de las variables aleatorias continuas la probabilidad de observar un valor específico es 0, por lo que las probabilidades se obtendrán a partir de la integración de la función de densidad sobre un intervalo de interés. En particular para alguna variable aleatoria X que tome valores x dentro de los \mathbb{R} , su función de densidad, si existe, está dada por una función integrable f de x que es no negativa y tiene la propiedad que:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Así, generalmente la función de densidad de probabilidad que se representa por $f(x)$ puede también pensarse matemáticamente como la derivada de la función de distribución de probabilidad (para el caso continuo). En el Apéndice B.1 se muestran algunas de las distribuciones que se utilizaron en esta tesis, junto con sus principales características.

A.2.3. Valor esperado, varianza, simetría y curtosis

Como se mencionó en los apartados anteriores, la variable X es estocástica porque puede tomar cualquier valor x aleatoriamente y encontrarse en cierto intervalo con una probabilidad. Las propiedades probabilísticas de la variable X están dadas por su función de distribución $F_X(\cdot)$. Pero también es de interés conocer algunas medidas que generen información útil sobre la variable aleatoria X , tal es el caso de los momentos de la distribución de probabilidad de la variable aleatoria. A continuación se describen algunos de estos, especialmente, los primeros cuatro que son útiles para el desarrollo de esta tesis:

Valor esperado y media muestral

El primer momento de la distribución de probabilidad de la variable aleatoria X es denominado valor esperado o esperanza matemática de X , el cual se denota como $E(X)$. Este indicador representa una medición de localización del centro de la distribución² y se asocia al valor promedio que toma la variable X .

Definición A.9. Sea X una variable aleatoria real definida sobre el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$. El valor esperado se define como:

¹Variabes discretas que tomen valores finitos.

²Esto si la distribución es simétrica.

1. En el caso que X sea una variable aleatoria discreta, con valores x_1, x_2, \dots , su valor esperado o media poblacional es:

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j)$$

2. En el caso que X sea una variable aleatoria continua con función de densidad f_X , su valor esperado o media poblacional es:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Por otra parte, muchas veces en la práctica se cuenta con un número finito de realizaciones de la variable aleatoria X , o simplemente no se conoce la función de densidad. Así que se procede a calcular los momentos muestrales como una aproximación de los poblacionales. Un estimador insesgado del valor esperado de la distribución de una variable aleatoria X es la media o promedio muestral (Ver demostración en apéndice C de Greene [2011]). Esta última no es más que el promedio aritmético simple de las observaciones en una muestra, por ejemplo, si el conjunto de realizaciones está compuesto por $\{1,2,3,4,5\}$, la media muestral sería 3.

Formalmente, para una muestra compuesta por n observaciones³ de la variable aleatoria X , la media muestral denotada por \bar{x} estará dada por la siguiente formula:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Como se discutió antes, un estimador insesgado de la media poblacional (μ) es la media muestral (\bar{x}). Así que también se tiene la siguiente propiedad $E(X) = E(\bar{X}) = \mu$.

Varianza

La varianza de la distribución de probabilidad de una variable aleatoria X se denota como $V(X)$, o en notación de parámetro como σ^2 , y mide o representa la dispersión de las realizaciones de la variable X de su centro de densidad o media. Cuando X es discreta

$$\sigma^2 = V(X) = \sum_{-\infty}^{\infty} (x - \mu)^2 P(X = x)$$

o cuando X es continua con función de densidad $f(x)$

³El conjunto $\{x_1, x_2, \dots, x_n\}$.

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Una definición más general, parte de la utilización del concepto de valor esperado, como sigue:

$$V(X) = E([X - E(X)]^2) = E(X^2) - [E(X)]^2$$

De esta definición se deriva que la varianza sea el segundo momento sobre la esperanza matemática, o el segundo momento central de la distribución de la variable aleatoria X .

Simetría y curtosis

Teniendo en cuenta lo anterior, la varianza podría también denotarse como μ_2 , y así extender la definición a momentos de mayor orden. De esta forma, el k -ésimo momento central de una variable aleatoria X puede ser definido como:

$$\mu_k = E([X - E(X)]^k) = E([X - \mu]^k)$$

La asimetría y la curtosis son el tercer y cuarto momentos centrales estandarizados, esto último significa que para cualquier momento de orden k se divide por σ^k y se obtiene el momento estandarizado. Así, la asimetría denotada por ι y la curtosis por κ estarían dadas por:

$$\iota = \frac{\mu_3}{\sigma_3}, \kappa = \frac{\mu_4}{\sigma_4}$$

A.2.4. Procesos estocásticos

El concepto de variable aleatoria sirve para introducir incertidumbre en el modelamiento matemático y económico. Esto permite a los investigadores acercarse aún más a la realidad ya que existen factores o choques que afectan el comportamiento de las variables de interés. Por ejemplo, los precios de un activo en bolsa (acción) responden a un componente determinístico pero también a uno estocástico. Así, en cada momento del tiempo el precio de la acción responde a un valor fundamental (situación del activo, fortalezas, diversos indicadores que lo describen) y a un choque inesperado que puede ser ocasionado por noticias o circunstancias que afectan el entorno del activo (elecciones presidenciales, condiciones climáticas, etc). De esta manera, cuando se estudia un fenómeno que ocurre a lo largo del tiempo, éste puede ser representado por medio de una colección de variables aleatorias que tienen como rasgo característico el estar

indexadas al tiempo. Esto se denomina como “*proceso estocástico*”. Por ejemplo, si X_t representa el precio de un activo financiero en el tiempo t , $\{X_t\}_t$ modela la evolución del precio en el tiempo.

Definición A.10. Un **proceso estocástico** es una colección de variables aleatorias $X = \{X_t\}_{t \in T}$, con $T \neq \emptyset$, definidas todas sobre el mismo espacio de probabilidad, con valores en un espacio medible llamado espacio de estados. El conjunto de parámetros T se llama el dominio de definición del proceso. Se dice que X es real si las variables aleatorias X_t son de valor real para todo $t \in T$.

Si $X = \{X_t\}_{t \in T}$ es un proceso real, para cada $\omega \in \Omega$, la aplicación

$$t \mapsto X_t(\omega)$$

se denomina una trayectoria o ω -trayectoria o realización del proceso estocástico real.

Para un proceso estocástico $X = \{X_t\}_{t \in T}$, al fijar un conjunto de instantes finitos $M = \{0 \leq t_1 < t_2 < \dots < t_n\} \subseteq \mathbb{R}$ se obtiene un vector aleatorio

$$\begin{aligned} (X_{t_1}, \dots, X_{t_n}) : \Omega &\rightarrow \mathbb{R}^n \\ \omega &\mapsto (X_{t_1}(\omega), \dots, X_{t_n}(\omega)) \end{aligned}$$

La distribución de $(X_{t_1}, \dots, X_{t_n})$ está dada por:

$$P_M(A) = P_{t_1, \dots, t_n}(A) = (P \circ (X_{t_1}, \dots, X_{t_n})^{-1})(A) = P((X_{t_1}, \dots, X_{t_n}) \in A)$$

para A un booreliano en \mathbb{R}^n .

La familia de distribuciones de probabilidades $\{P_M\}_{M \subseteq T}$, $\emptyset \neq M \subseteq T$ finito se llama familia de distribuciones finito dimensionales del proceso estocástico X . Las propiedades pueden ser consultadas en Billingsley [1986] o en Blanco and Muñoz [2003].

Definición A.11. Sea $\{X\}$ un proceso estocástico, donde cada X_t asume un valor en el espacio Θ . El proceso $\{X_t\}$ es un *proceso de Markov*, si este tiene dentro de sus propiedades lo siguiente: dado el valor de X_t , los valores de X_j , $j > t$, no dependen de los valores de X_s , $s < t$. En otras palabras, $\{X_t\}$ es un *proceso de Markov* si su función de distribución condicional satisface:

$$P(X_j | X_s, s \leq t) = P(X_j | X_t), \quad h > t \tag{A.1}$$

Si $\{X\}$ es un proceso estocástico discreto, la anterior propiedad se vuelve:

$$P(X_j | X_t, X_{t-1}, \dots) = P(X_j | X_t), \quad h > t \tag{A.2}$$

Sea \mathbf{A} un subconjunto de Θ . La función:

$$P_t(\theta, j, \mathbf{A}) = P(X_j \in \mathbf{A} | X_t = \theta), \quad h > t \tag{A.3}$$

es llamada la función de transición del *proceso de Markov*. Si la probabilidad de transición depende de $h - t$, pero no de t , entonces el proceso tiene una distribución de transición estacionaria.

A.2.5. Martingalas

En teoría de probabilidad, la martingala (galicismo de martingale) es un determinado proceso estocástico. Por ejemplo, un método de apuesta en juegos de azar consistente en multiplicar sucesivamente en caso de pérdida una apuesta inicial determinada. En el momento de ganar la apuesta, el proceso se iniciaría de nuevo. En la ruleta, por ejemplo, la martingala consistiría en comenzar apostando una determinada cantidad, por ejemplo 1 dólar, al rojo. En caso de pérdida, se apostaría de nuevo al rojo, pero esta vez, duplicando la cantidad y así, sucesivamente, hasta ganar la apuesta. Llegado ese momento se compensarían las pérdidas y obtendríamos como beneficio la primera cantidad apostada. Originalmente la martingala se refería a un tipo de estrategia de apuesta muy popular en Francia en el siglo XVIII. La más simple de estas estrategias fue diseñada para un juego en que el apostador gana la apuesta en caso de que al lanzar una moneda caiga de cara y pierde en caso de que salga cruz.

El concepto de la martingala en la teoría de probabilidades fue introducido por Paul Pierre Lévy, y una gran parte del desarrollo original de la teoría lo realizó Joseph Leo Doob. Parte de la motivación era demostrar la inexistencia de estrategias de juego infalibles. Recientemente, en el mercado bursátil el método estocástico de martingalas, adaptado en Forex porque trabaja en dos posiciones (alza-baja).⁴

Definición A.12. Sea $X = \{X_t\}_{t \geq 0}$ un proceso estocástico, se dice que X es una martingala con respecto a la filtración $\{\mathfrak{F}_t\}_{t \geq 0}$, si:

1. $E(|X_t|) < \infty$, para todo $t \geq 0$.
2. X es adaptada con respecto a la filtración $\{\mathfrak{F}_t\}_{t \geq 0}$.
3. $E(X_t | \mathfrak{F}_s) = X_s$, para todo $t \geq s$.

Se dice que X es una supermartingala si la condición 3 se reemplaza por:

$$E(X_t | \mathfrak{F}_s) \leq X_s, \text{ para todo } t \geq s.$$

Se dice que X es una submartingala si la condición 3 se reemplaza por:

$$E(X_t | \mathfrak{F}_s) \geq X_s, \text{ para todo } t \geq s.$$

Observación A.1. La *martingala* en el caso discreto es un modelo apropiado para describir un juego justo. Supóngase que la variable aleatoria X_n denota la fortuna del jugador luego de la n -ésima ronda de juego y que la σ -álgebra \mathfrak{F}_m contiene toda la información acerca del desarrollo del juego hasta la m -ésima ronda de juego, entonces el juego es justo si el valor esperado de los valores futuros de X_n con $n \geq m$, dada la información contenida en \mathfrak{F}_m , es igual a la fortuna del jugador en el tiempo m , es decir, la predicción es justamente el valor actual. En términos probabilísticos se espera que:

$$E(X_n | \mathfrak{F}_m) = X_m; \quad P - c.s \quad \text{para todo } m \leq n$$

⁴Este apartado fue tomado de <http://es.wikipedia.org/wiki/Martingala>.

Una consecuencia de la tercera condición de martingala es la siguiente:

Lema A.1. Sea $X = \{X_t\}_{t \geq 0}$ una martingala con respecto a la filtración $\{\mathfrak{F}_t\}_{t \geq 0}$, entonces para $s \leq t$

$$E(X_t - X_s | \mathfrak{F}_s) = 0 \quad (\text{A.4})$$

Demostración.

$$\begin{aligned} E(X_t - X_s | \mathfrak{F}_s) &= E(X_t | \mathfrak{F}_s) - E(X_s | \mathfrak{F}_s) \\ &= X_s - E(X_s | \mathfrak{F}_s) \\ &= X_s - X_s \\ &= 0 \end{aligned} \quad \blacksquare$$

Del anterior lema se concluye que si $X = \{X_t\}_{t \geq 0}$ es una martingala, la función de media es constante. Como se verá más adelante esto tiene una implicación económica en términos de eficiencia de los mercados, lo cual sugiere que el mejor pronóstico que se puede realizar para un precio que sigue una martingala es el último precio observado. Por lo tanto, no hay ganancias extraordinarias ni arbitraje.

A.2.6. Movimiento browniano

En 1827, el médico y botánico escocés Robert Brown (1773-1858) mientras examinaba partículas de polen en el microscopio, observó que cuando éstas se encontraban suspendidas en agua y en otros líquidos mostraban comportamientos erráticos. En principio, pensó que las partículas tenían vida propia, pero luego descubrió que este fenómeno obedecía a choques aleatorios que sufrían los átomos de polen junto con las moléculas del líquido. Sin embargo, su formulación matemática sólo apareció hasta comienzos del siglo XX cuando el físico Albert Einstein escribió un artículo sobre mecánica estadística, que proporciona la formulación matemática del movimiento Browniano, de la cual se deriva que la dispersión promedio del desplazamiento de la partícula en un líquido, en un tiempo dado, es proporcional a dicho tiempo.

Otro destacado matemático asociado con la axiomática del movimiento Browniano es Norbert Wiener, quién alrededor de los años 20 presentó un modelo matemático para este movimiento basado en la teoría de procesos estocásticos y teoría de la medida. Allí expresó su formulación a través de filtraciones, denominando su formulación como proceso de Wiener. En la literatura es muy común asociar los movimientos brownianos con los procesos de Wiener, la diferencia fundamental es que en el segundo el concepto de filtración está presente, mientras que en el primero no ⁵.

En el caso unidimensional, la definición del movimiento browniano como proceso estocástico es la siguiente:

⁵Este apartado se realizó con base en la reseña histórica del Capítulo 3 de Venegas [2007].

Definición A.13. Un proceso estocástico $W = \{W_t\}_{t \geq 0}$ es llamado **movimiento browniano** estándar o proceso **Wiener**, si se satisfacen las siguientes condiciones:

1. $W_0 = 0$.
2. $W_{t+h} - W_t \stackrel{d}{\sim} N(0, h)$ para $h > 0$.
3. W tiene incrementos independientes.
4. Las trayectorias del proceso son funciones continuas.

Las propiedades y características del movimiento browniano, pueden ser consultadas con mayor detalle en Blanco and Muñoz [2003] o en Tudor [1997].

Algunos procesos estocásticos relacionados con el movimiento browniano son:

1. **El puente browniano:**

$$X_t = W_t + tW_1$$

para $t \in [0,1]$.

2. **El movimiento browniano con derive:**

$$X_t = \sigma W_t + \mu t$$

donde $t \geq 0$, $\sigma > 0$ y $\mu \in \mathbb{R}$ son constantes.

3. **Movimiento browniano geométrico:**

$$X_t = e^{\sigma W_t + \mu t}$$

donde $t \geq 0$, $\sigma > 0$ y $\mu \in \mathbb{R}$ son constantes. Es decir, la exponencial del movimiento browniano con derive.

Movimiento browniano geométrico

El movimiento browniano estándar no describe apropiadamente la dinámica de los precios de un activo ya que, por ejemplo, el supuesto que los precios parten de cero es poco realista. Asimismo, empíricamente se ha observado que los incrementos que rigen la ecuación de comportamiento de los precios pueden tener medias diferentes de cero, ó también podrían tener varianzas que no son necesariamente proporcionales al tiempo.

Para solucionar algunos de estos limitantes se introdujo el movimiento browniano geométrico, el cual es una transformación exponencial del movimiento browniano estándar, que tiene algunas de las características empíricas de los precios tales como: los precios de los activos comienzan en valores diferentes de cero, tienen incrementos con

medias diferentes de cero, varianzas que no son proporcionales al tiempo, y covarianzas diferentes de cero.

Así, la representación de un movimiento browniano geométrico del precio de un activo $S(t)$ es de la siguiente forma:

$$S_t = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right\}$$

donde W_t es un movimiento browniano estándar, μ es una constante asociada a la tendencia, σ es una constante positiva asociada a la volatilidad y S_0 es un precio inicial predeterminado y conocido. Este proceso es bastante utilizado en los modelos financieros para describir el cambio porcentual (retorno o rendimiento) del precio de un activo⁶. Nótese que una expresión similar que recoge lo anterior es:

$$\ln(S_t) = \ln(S_0) + \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t$$

De esta representación, se tiene que la distribución de $\ln(S_t)$ es normal con media o valor esperado de $\ln(S_0) + \left(\mu - \frac{1}{2}\sigma^2\right)t$ y varianza dada por $\sigma^2 t$.

A.2.7. Integral estocástica y diferencial estocástica

Después de la introducción de los movimientos brownianos, que permitían incorporar en el análisis la aleatoriedad propia de los activos financieros, surgió la necesidad de encontrar herramientas útiles que permitieran modelar adecuadamente las variables financieras. Así, una de las ramas de la teoría económica y financiera que ha cobrado fuerza es el cálculo estocástico en la medida que brinda la posibilidad de modelar el comportamiento del precio de activos en tiempo continuo y en entornos de incertidumbre o aleatorios. Conceptos clave dentro de estas teorías son la integral estocástica y la diferencial estocástica, siendo la segunda el término más utilizado y convencional dentro del estudio de modelos estocásticos en tiempo continuo en la actualidad. En términos estrictos, el objeto del cálculo estocástico es la integral y no la diferencial, sin embargo, en la práctica cuando se escribe una ecuación diferencial estocástica realmente se está pensando en una integral estocástica. Por lo tanto, una ecuación diferencial estocástica es una notación simplificada de una integral estocástica. A continuación se tratará de describir intuitivamente la relación entre estos dos conceptos.

Cabe resaltar que el comportamiento del precio de un activo financiero a través del tiempo es un ejemplo clásico de un sistema dinámico con perturbaciones aleatorias que

⁶Más adelante se hará una breve descripción del concepto de retorno y su relación con las propiedades del logaritmo.

pueden derivar su origen de eventos naturales o situaciones políticas o económicas, tales como inundaciones, elecciones, guerras, etc. En general, es muy común utilizar la siguiente ecuación estocástica como ley de movimiento del precio de un activo X_t , a ésta se le denomina ecuación diferencial estocástica:

$$dX_t = \mu X_t dt + \sigma X_t dW_t \quad (\text{A.5})$$

donde $X_0 = x_0 > 0$ y μ y $\sigma > 0$ son constantes.

Esta ecuación es de fácil interpretación y puede describirse de la siguiente forma. Para explicar el primer término, debe suponerse la ausencia del componente estocástico dado por el movimiento browniano. De esta manera, la ecuación se reduciría a $dX_t = \mu X_t dt$ cuya solución es $X_t = x_0 e^{\mu t}$, la cual revisa el comportamiento de un capital $x_0 > 0$ que crece de manera continua y determinística a una tasa efectiva de μ por ciento. Por otro lado, el segundo término ($\sigma X_t dW_t$) corresponde a la volatilidad de una inversión con riesgo sujeta a las fluctuaciones de los mercados financieros. El modelo asume que dicha variabilidad es proporcional al valor de la inversión.

Es curioso que la ecuación anterior se conozca como ecuación diferencial estocástica cuando se sabe que las trayectorias del movimiento browniano no son diferenciables. Teniendo en cuenta esto, en 1942 el matemático japonés Kiyosi Itô (1915-2008) dió una rigurosa representación de la parte estocástica de la ecuación (A.5) escribiéndola en forma integral:

$$\int_0^t X_1(s) dW_s \quad (\text{A.6})$$

y determinando para cuáles procesos $\{X_1(s)\}$ la anterior expresión tiene sentido. En primer lugar él definió la integral para ciertos procesos llamados simples, luego por una metodología de paso al límite lo definió para procesos más generales. El conjunto de procesos para los cuales se define la integral estocástica, es el siguiente:

Definición A.14. Sea $\mathcal{N} = \mathcal{N}[S, T]$ la clase de funciones de

$$f(t, w) : [S, T] \times \Omega \longrightarrow \mathbb{R}$$

tales que:

1. $(t, w) \longrightarrow f(t, w)$ es $\mathcal{B} \otimes \mathfrak{F}$ -medible.
2. $f(t, w)$ es \mathfrak{F}_t -adaptado.
3. $E \left(\int_S^T f(t, w)^2 dt \right) < \infty$ (trayectorias cuadrado integrables).

Observación A.2. El conjunto \mathcal{N} es un subconjunto de L^2 , siendo L^2 el conjunto de todas las funciones cuadrado integrable.

La construcción y propiedades de la integral estocástica, así como su extensión n -dimensional se pueden ver con mayor detalle en Oksendal [2005], en Blanco and Muñoz [2003] o en Karatzas and Shreve [1988].

Por medio de la integral estocástica se puede definir un proceso estocástico especial, el cual se conoce como proceso de Itô.

Definición A.15. Sea $W = \{W_t\}_{t \geq 0}$ un movimiento browniano unidimensional sobre $(\Omega, \mathfrak{F}, P)$. Un **proceso de Itô** (o **integral estocástica**) unidimensional, es un proceso estocástico $X = \{X_t\}_{t \geq 0}$ sobre $(\Omega, \mathfrak{F}, P)$ tal que

$$X_t(w) = X_0(w) + \int_0^t U(s, w)ds + \int_0^t V(s, w)dW_s(w) \quad (\text{A.7})$$

donde U es un proceso medible y $\{\mathfrak{F}_t\}_{t \geq 0}$ -adaptado con trayectorias integrables y V es un proceso en \mathcal{N} tal que

$$P \left[\int_0^t V(s, \omega)^2 ds < \infty, \text{ para todo } t \geq 0 \right] = 1.$$

Al igual que ocurre con el cálculo diferencial e integral clásico, la evaluación de la integral a partir de la definición es un trabajo difícil. Sin embargo, el siguiente resultado conocido como Fórmula de Itô permite calcular, en algunos casos, la integral estocástica.

Teorema A.3 (Fórmula de Itô). Sea $X = \{X_t\}_{t \geq 0}$ un proceso de Itô, esto es, supóngase que

$$dX_t = U_t dt + V_t dW_t$$

y sea $g(t, x) \in C^{2,1}([0, \infty) \times \mathbb{R})$,

donde $Y_t = g(t, X_t)$ es de nuevo una integral estocástica con $dt \cdot dt = dt \cdot dW_t = dW_t \cdot dW_t = 0$ y $dW_t \cdot dW_t = dt$.

Entonces:

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2$$

Demostración. Oksendal [2005]. ■

Retomando el problema inicial de la ecuación (A.5) y aplicando el concepto anterior de la fórmula de Itô, y suponiendo que $Y_t = g(t, x) = \ln X_t$ entonces se tiene que:

$$d \ln X_t = \left(\mu - \frac{1}{2} \sigma^2 \right) dt + \sigma dW_t \quad (\text{A.8})$$

Expresión cuya correspondiente integral estocástica y solución está dada por:

$$\begin{aligned} \ln X_t &= \ln X_0 + \mu \int_0^t d\mu - \frac{1}{2} \sigma^2 \int_0^t (dW_\mu)^2 + \sigma \int_0^t dW_s \\ &= \ln X_0 + (\mu - \frac{1}{2} \sigma^2) t + \sigma W_t \end{aligned}$$

Esta última es exactamente la misma ecuación del movimiento browniano geométrico visto en la sección anterior.

Por último, discretizando la ecuación (A.5) con $\Delta t = T - t$ entonces se tiene:

$$\ln(X_T) - \ln(X_t) = (\mu - \frac{1}{2} \sigma^2)(T - t) + \sigma(\sqrt{T - t})\varepsilon$$

con $\varepsilon \sim \mathbb{N}(0, 1)$. Por lo tanto, el supuesto del movimiento browniano geométrico para los precios en tiempo continuo es equivalente a suponer una caminata aleatoria para el logaritmo de los precios en tiempo discreto, esto es útil cuando se quiere evaluar si un mercado es eficiente.

Apéndice B

Funciones de distribución

B.1. Funciones de distribución

B.1.1. Distribución Normal

Sea una variable aleatoria X , si ésta se distribuye normal con media μ y varianza σ^2 , su función de distribución está definida por:

$$f(X|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(X - \mu)^2}{2\sigma^2}\right\} \quad (\text{B.1})$$

B.1.2. Distribución log-normal

Sea una variable aleatoria x , si ésta se distribuye log-normal, su función de distribución está definida por (para $x \geq 0$):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{[(\ln x - \mu)]^2}{2\sigma^2}\right\} \quad (\text{B.2})$$

Una variable log-normal tiene una media igual a $(\exp\{\mu + \sigma^2/2\})$ y una varianza igual a $(\exp\{2\mu + \sigma^2\} (\exp\{\sigma^2\} - 1))$. Si y es log-normal (μ, σ^2) con media μ y varianza σ^2 , el logaritmo de y se distribuye Normal. $\ln y$ se distribuye $N(\mu, \sigma^2)$.

B.2. Descripción probabilidades condicionadas del Modelo de volatilidad estocástica

Partiendo del siguiente modelo de volatilidad estocástica se plantea el problema de estimación a resolver y se calculan las funciones de densidad condicionada¹

$$\begin{aligned} y_t &= \exp\left\{\frac{h_t}{2}\right\} \sigma_\varepsilon \varepsilon_t, \text{ con } \varepsilon_t \stackrel{i.i.d.}{\sim} \mathbf{N}(0, 1) \\ h_t &= \mu + \phi h_{t-1} + \tau \eta_t, \text{ con } \eta_t \stackrel{i.i.d.}{\sim} \mathbf{N}(0, 1) \end{aligned} \quad (\text{B.3})$$

Como se apreció en el capítulo 3, los métodos de estimación bayesianos parten de una condición inicial para la variable de estado, que en este caso es la log-volatilidad (h_t) en el momento cero. Esta es definida como

$$h_0 \sim \mathbf{N}\left(\frac{\mu}{1-\phi}, \frac{\tau^2}{1-\phi^2}\right) \quad (\text{B.4})$$

De esta forma la función de verosimilitud es

$$\begin{aligned} \mathbf{L}(\mu, \phi, \tau | y_1, y_2 \dots y_T) = \\ \int_{\mathbb{R}^T} \prod_{t=1}^T \frac{1}{\sqrt{2\pi \exp\{h_t\}}} \exp\left\{-\frac{1}{2} \exp\{-h_t\} y_t^2\right\} \frac{1}{\sqrt{2\pi \tau^2}} \exp\left\{-\frac{1}{2\tau^2} (h_t - \mu - \phi h_{t-1})^2\right\} dh_1 dh_2 \dots dh_T \end{aligned} \quad (\text{B.5})$$

Este es el problema de integración que surge sobre un *gran espacio dimensional* (en este caso T sobre el modelo B.3) puede ser resuelto usualmente al incluir las variables *latentes* dentro del vector de parámetros, para indagar más al respecto el lector puede ver el capítulo 5 y 9 de Robert and Casella [2004]. De acuerdo con Robert and Casella [2004], el problema se reduce a²

¹Se puede fijar $\sigma_\varepsilon^2 = \exp\{\mu\}$.

²Se omite la dependencia sobre las condiciones iniciales (h_0, y_1), las cuales se asumen son conocidas.

$$\mathbf{L}(\mu, \phi, \tau | h_1, h_2 \dots h_T, y_1, y_2 \dots y_T) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi \exp\{h_t\}}} \exp\left\{-\frac{1}{2} \exp\{-h_t\} y_t^2\right\} \frac{1}{\sqrt{2\pi \tau^2}} \exp\left\{-\frac{1}{2\tau^2} (h_t - \mu - \phi h_{t-1})^2\right\} \quad (\text{B.6})$$

En principio, se asume que los parámetros μ, ϕ, τ tienen las siguientes distribuciones *a priori*

$$\begin{aligned} \mu &\sim N(\bar{\mu}, \sigma_\mu^2) \\ \phi &\sim N(\bar{\phi}, \sigma_\phi^2) \\ \tau^2 &\sim \frac{1}{\tau^2} \end{aligned} \quad (\text{B.7})$$

Los parámetros μ y ϕ tendrán distribuciones *a priori* y las *condicionadas* de acuerdo a la distribución Normal-Gaussiana. Mientras que τ^2 se comporta como una función de distribución Inversa-Gamma (Ver Apéndice B.1).

De esta forma, definiendo los siguientes conjuntos de información: $Y = (y_1, \dots, y_T)'$, $H = (h_1, \dots, h_T)'$ y $H_{-t} = (h_1, h_2, \dots, h_{t-1}, h_{t+1}, \dots, h_T)'$, las funciones de densidad condicionales de los parámetros (similar a 3.13) pueden ser aproximadas como sigue

Condional para μ

$$\begin{aligned} \Pi(\mu | \phi, \tau, Y, H) &\propto \prod_{t=1}^T \exp\left\{-\frac{1}{2\tau^2} (h_t - \mu - \phi h_{t-1})^2\right\} \exp\left\{-\frac{1}{2\sigma_\mu^2} (\mu - \bar{\mu})^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\tau^2} \sum_{t=1}^T (\mu^2 - 2\mu(h_t - \phi h_{t-1})) - \frac{1}{2\sigma_\mu^2} (\mu^2 - 2\mu\bar{\mu})\right\} \\ &\propto \exp\left\{-\frac{1}{2} \left(\mu^2 \left(\frac{T}{\tau^2} + \frac{1}{\sigma_\mu^2}\right) - 2\mu \left(\frac{1}{\tau^2} \sum_{t=1}^T (h_t - \phi h_{t-1}) + \frac{\bar{\mu}}{\sigma_\mu^2}\right)\right)\right\} \\ &\propto \mathbf{N}(\bar{\mu}, \bar{\sigma}_\mu^2) \end{aligned} \quad (\text{B.8})$$

donde $\bar{\sigma}_\mu^2 = \left(\frac{T}{\tau^2} + \frac{1}{\sigma_\mu^2}\right)^{-1}$

y $\bar{\mu} = \bar{\sigma}_\mu^2 \left(\frac{1}{\tau^2} \sum_{t=1}^T (h_t - \phi h_{t-1}) + \frac{\bar{\mu}}{\sigma_\mu^2}\right)$

Condicional para ϕ

$$\begin{aligned}
 \Pi(\phi|\mu, \tau, Y, H) &\propto \prod_{t=1}^T \exp\left\{-\frac{1}{2\tau^2}(h_t - \mu - \phi h_{t-1})^2\right\} \exp\left\{-\frac{1}{2\sigma_\phi^2}(\phi - \bar{\phi})^2\right\} \\
 &\propto \prod_{t=1}^T \exp\left\{-\frac{1}{2}\left(\frac{1}{\tau^2}\phi^2 h_{t-1}^2 - 2\phi h_{t-1}(h_t - \mu)\right)\right\} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_\phi^2}(\phi^2 - 2\bar{\phi}\phi)\right)\right\} \\
 &\propto \exp\left\{-\frac{1}{2}\left(\phi^2\left(\frac{1}{\tau^2}\sum_{t=1}^T h_{t-1}^2 + \frac{1}{\sigma_\phi^2}\right)\right)\right\} \exp\left\{-\frac{1}{2}\left(-2\phi\left(\frac{1}{\tau^2}\sum_{t=1}^T (h_t - \mu)h_{t-1} + \frac{\bar{\phi}}{\sigma_\phi^2}\right)\right)\right\} \\
 &\propto \mathbf{N}(\bar{\phi}, \bar{\sigma}_\phi^2)
 \end{aligned} \tag{B.9}$$

donde $\bar{\sigma}_\phi^2 = \left(\frac{1}{\tau^2}\sum_{t=1}^T h_{t-1}^2 + \frac{1}{\sigma_\phi^2}\right)^{-1}$

y $\bar{\phi} = \bar{\sigma}_\phi^2\left(\frac{1}{\tau^2}\sum_{t=1}^T (h_t - \mu)h_{t-1} + \frac{\bar{\phi}}{\sigma_\phi^2}\right)$

Condicional para τ^2

$$\begin{aligned}
 \Pi(\tau^2|\mu, \phi, Y, H) &\propto \prod_{t=1}^T \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(h_t - \mu - \phi h_{t-1})^2\right\} \\
 &\propto (\tau^2)^{\frac{T}{2}-\frac{1}{2}} \exp\left\{-\frac{1}{\tau^2}\left(\frac{1}{2}\sum_{t=1}^T (h_t - \mu - \phi h_{t-1})^2\right)\right\} \\
 &\propto \mathbf{IG}(\bar{m}_t, \bar{V}_t)
 \end{aligned} \tag{B.10}$$

donde $\bar{m}_t = \frac{T-1}{2}$ y $\bar{V}_t = \left(\frac{1}{2}\sum_{t=1}^T (h_t - \mu - \phi h_{t-1})^2\right)$.

Por último, Casarin [2012] demuestra que las *condicionales* para las variables *latentes* están dadas por

Condicional para h_t , $t = 2, \dots, T - 1$

$$\begin{aligned}
 \Pi(h_t|\mu, \phi, \tau, Y, H_{-t}) &\propto \\
 &\propto \exp\left\{-\frac{h_t}{2}\right\} \exp\left\{-\frac{y_t^2}{2} \exp\{-h_t\}\right\} \exp\left\{-\frac{1}{2\tau^2}(h_t - \mu - \phi h_{t-1})^2\right\} \exp\left\{-\frac{1}{2\tau^2}(h_{t+1} - \mu - \phi h_t)^2\right\} \\
 &\propto \exp\left\{-\frac{y_t^2}{2} \exp\{-h_t\}\right\} \exp\left\{-\frac{h_t}{2} - \frac{1}{2\tau^2}(h_t^2 - 2h_t(\mu + \phi h_{t-1}) + h_t^2\phi^2 - 2h_t\phi(h_{t+1} - \mu))\right\} \\
 &\propto \exp\left\{-\frac{y_t^2}{2} \exp\{-h_t\}\right\} \exp\left\{-\frac{1}{2}\left(h_t^2\left(\frac{1+\phi^2}{\tau^2}\right) - 2h_t\frac{1}{\tau^2}(\mu(1-\phi) + \phi(h_{t-1} + h_{t+1}) - \frac{\tau^2}{2})\right)\right\}
 \end{aligned} \tag{B.11}$$

Condicional para h_1

$$\begin{aligned}
 \Pi(h_1|\mu, \phi, \tau, Y, H_{-1}) &\propto \exp\left\{-\frac{h_1}{2}\right\} \exp\left\{-\frac{y_1^2}{2} \exp\{-h_1\}\right\} \exp\left\{-\frac{1}{2\tau^2}(h_2 - \mu - \phi h_1)^2\right\} \\
 &\propto \exp\left\{-\frac{y_1^2}{2} \exp\{-h_1\}\right\} \exp\left\{-\frac{h_1}{2}\right\} \mathbf{N}\left(\frac{h_2 - \mu}{\phi}, \frac{\tau^2}{\phi^2}\right)
 \end{aligned} \tag{B.12}$$

Condicional para h_T

$$\begin{aligned}
 \Pi(h_T, \mu, \phi, \tau, Y, H_{-T}) &\propto \exp\left\{-\frac{h_T}{2}\right\} \exp\left\{-\frac{y_T^2}{2} \exp\{-h_T\}\right\} \exp\left\{-\frac{1}{2\tau^2}(h_T - \mu - \phi h_{T-1})^2\right\} \\
 &\propto \exp\left\{-\frac{h_T}{2}\right\} \exp\left\{-\frac{y_T^2}{2}\right\} \mathbf{N}(\mu + \phi h_{T-1}, \tau^2)
 \end{aligned} \tag{B.13}$$

De esta forma, queda demostrado que los parámetros $(\mu, \phi|\tau^2)$ se distribuyen normal y que τ^2 sigue una distribución inversa-gamma. Esto es utilizado para realizar los métodos de simulación de Monte Carlo.

Apéndice C

Conceptos Básicos de Series de Tiempo

C.1. Series de tiempo

Una serie de tiempo es una sucesión de realizaciones de una generadas por un proceso estocástico, con la particularidad que su índice toma una relación con al tiempo. De esta manera, toda inferencia que se realice, se hará sobre las características del proceso estocástico generador de la serie observada. Cabe mencionar de paso, que así como existen procesos estocástico discretos y continuos, de igual manera existen series de tiempo discretas y continuas. Por ejemplo, en el caso de una serie de tiempo discreta, que son mucho más comunes en el ámbito económico, las observaciones de la variable aleatoria (Y) se toman de los momentos t_1, t_2, \dots, t_T , el proceso estocástico respectivo se denotará por $\{Y(t_1), Y(t_2), \dots, Y(t_T)\}$. Una característica que tendrán las series de tiempo es que sus realizaciones están equi-espaciadas, es decir, los períodos t_1, t_2, \dots, t_T son del mismo tamaño (ej: días, meses, años).

C.1.1. Operador rezago y operador diferencia

Un operador muy conocido en el trabajo de serie de tiempo es el denominado Operador de rezago o retraso, que se denotará por la letra L. Dicho operador se define mediante la siguiente relación:

$$LY_t = Y_{t-1}, L^m Y_t = Y_{t-m}$$

para todo t . Debe tenerse en cuenta que el operador sólo se aplica sobre una serie de tiempo o proceso estocástico que tenga asociado un índice de tiempo. Por ejemplo, al aplicar este operador a una constante, el resultado es ella misma (ej: $L\pi = \pi$).

A partir de esto se deriva el operador diferencia que es muy utilizado para realizar transformaciones a series que no son estacionarias como vamos a ver más adelante. El operador diferencia se define mediante la siguiente relación:

$$(1 - L)Y_t = Y_t - Y_{t-1} = \Delta Y_t$$

C.1.2. Procesos estacionarios y procesos ruido blanco

Las series de tiempo son procesos estocásticos cuyos valores sucesivos pueden ser altamente dependientes, puede ser considerada generada a partir de una serie de choques aleatorios independientes ($\{\varepsilon_t\}$). Estos últimos se suponen que son realizaciones independientes de una variable aleatoria cuya media y varianza es constante. Este tipo de sucesión de variables aleatorias es conocida como proceso ruido blanco. El nombre de ruido blanco tiene su origen en la ingeniería electrónica y de comunicaciones, en donde es necesario distinguir las señales con mensaje de las del ruido.

Definición C.1. Una serie de tiempo $\{\varepsilon_t\}$ es un proceso ruido blanco si cumple las siguientes propiedades:

- I) $E(\varepsilon_t) = 0 \quad \forall t$
- II) $V(\varepsilon_t) = \sigma_\varepsilon^2 \quad \forall t$
- III) $\text{cov}(\varepsilon_t, \varepsilon_{t-s}) = 0 \quad \forall t, s \neq 0$

Otro concepto de suma importancia es el de estacionariedad. En general, para caracterizar completamente un proceso estocástico es necesario conocer la función de densidad conjunta de las variables aleatorias que lo componen, para encontrar condiciones o propiedades que éste tenga a lo largo del tiempo. Esto conduce al concepto de estacionariedad:

Definición C.2. Un proceso estocástico (serie de tiempo) $\{Y_t\}$ es estacionario en sentido estricto si:

- I) Las distribuciones marginales de todas las variables son idénticas
- II) Las distribuciones finito-dimensionales de cualquier conjunto de variables sólo dependen de los retardos entre ellas

Estas dos condiciones se pueden resumir al establecer que la función de distribución conjunta de un conjunto de variables aleatorias $Y_t, Y_{t+1}, \dots, Y_{t+m}$ es invariante respecto a desplazamientos en el tiempo, es decir, un proceso es estacionario en sentido estricto si:

$$F(Y_t, Y_{t+1}, \dots, Y_{t+m}) = F(Y_{t+j}, Y_{t+j+1}, \dots, Y_{t+j+m}) \text{ para toda } t, m \text{ y } j.$$

Sin embargo, en la práctica es poco realista el pretender verificar la anterior condición ya que muchas veces lo único que se tiene es una realización en cada momento del tiempo. No obstante, si podemos estudiar el comportamiento de las variables aleatorias a través del estudio de los primeros dos momentos de sus distribuciones. Esto dado que como se había mencionado con anterioridad la media, la varianza y las covarianzas resumen en buena medida el comportamiento de la distribución de dichas variables. Así, surge el concepto de estacionariedad en sentido débil, el cual es el más utilizado en el campo de las series de tiempo y será utilizado en el presente trabajo.

Definición C.3. Un proceso estocástico (serie de tiempo) $\{Y_t\}$ es estacionario en sentido débil si se cumplen las siguientes condiciones:

$$\text{I) } E(Y_t) = \mu_y \quad \forall t$$

$$\text{II) } V(Y_t) = \sigma_Y^2 \quad \forall t$$

$$\text{III) } \text{cov}(Y_t, Y_{t-s}) = \gamma_s \quad \forall t$$

Esto último puede interpretarse como que para cada variable aleatoria que compone el proceso estocástico o serie de tiempo, la media y la varianza son la misma, y las covarianzas entre las variables es constante y sólo depende del espacio de tiempo al cual están separadas. Una serie de tiempo que es estacionaria es una serie que es integrada de orden cero o $I(0)$.

C.1.3. Transformaciones aplicadas a series no estacionarias

Ahora, en la práctica es poco probable encontrar series de tiempo que sean estacionarias ó $I(0)$ y se utilizan algunas transformaciones para obtener esta propiedad. La diferenciación es una de la más utilizada a menudo para estacionarizar series, en este se permite transformar la serie original de tal manera que se va eliminando la presencia de raíces unitarias. Así por ejemplo, una serie $\{Y_t\}$ es integrada de orden d , si $\{Y_t\}$ es estacionaria después de ser diferenciada d veces:

$$\text{Si } Y_t \sim I(d) \Rightarrow (1 - L)^d Y_t \sim I(0)$$

Es decir, que la serie original contaba con mínimo d raíces que hacían que el proceso no fuera estacionario, o análogamente, en el contexto de ecuaciones en diferencias que no se contaba con estabilidad. Por ejemplo, si $Y_t \sim I(2)$ entonces se requiere diferenciar dos veces la serie para obtener estacionariedad:

$$\begin{aligned} Z_t &= Y_t - Y_{t-1} = (1 - L)Y_t \\ Z_t - Z_{t-1} &= (1 - L)Z_t = (1 - L)^2 Y_t \Rightarrow (1 - L)^2 Y_t \sim I(0) \end{aligned}$$

En general, al momento de modelar series de tiempo se utilizan algunas de las siguientes herramientas que permiten estabilizar varianza y media de la serie de tiempo $\{Y_t\}$.

1. Estabilizar varianza

a) Logaritmos: $Z_t = \ln(Y_t)$

b) Transformación Box-Cox: $Z_t = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln(Y_t) & \text{si } \lambda = 0 \end{cases}$

2. Estabilizar media

a) Extracción de la tendencia determinística: $Y_t = \alpha + \beta t + e_t \Rightarrow y_t = x_t - \hat{\alpha} - \hat{\beta}t$

b) Diferenciación: $Z_t = Y_t - Y_{t-1}$

Apéndice D

Discretización

D.1. Discretización

Para tratar de hacer alguna inferencia de un modelo en tiempo continuo se debe tener confianza en un número N de realizaciones discretas. De esta manera, se divide el período de la muestra $[0, T]$ en N intervalos correspondientes a los datos en tiempo discreto (generalmente las observaciones son equidistantes). Luego el proceso en tiempo continuo se sustituye por un proceso constante a trozos y en el cual para cada intervalo $[t_i, t_{i+1})$, $i = 1, 2, \dots, N - 1$ se asume que el proceso es constante, aunque si cambio de un intervalo a otro. A continuación se mostrarán dos métodos de discretización para un proceso de Itô: el esquema de Euler y el esquema de Milstein (Ver una referencia detallada de estos métodos en Kloeden and Platen [1992]).

D.1.1. Método de Euler

Considere un proceso estocástico $X = \{X_t, t_0 \leq t \leq T\}$ que satisface la siguiente ecuación diferencial:

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t$$

sobre $t_0 \leq t \leq T$ y para un valor inicial $X_{t_0} = X_0$. Entonces, si se divide el intervalo de tiempo $[t_0, T]$ en N intervalos donde $t_0 < t_1 < \dots < t_n = T$, el esquema de Euler es:

$$Y_{n+1} = Y_n + a(t_n, Y_n)(t_{n+1} - t_n) + b(t_n, Y_n)(W_{t_{n+1}} - W_{t_n}) \quad (\text{D.1})$$

para $n = 0, 1, 2, \dots, N - 1$ con valor inicial $Y_0 = X_0$. Ahora, si los intervalos son igualmente espaciados entonces $\Delta = t_n - t_{n-1} = \frac{T - t_0}{N}$ y se puede reescribir la ecuación (D.1) de la siguiente forma:

$$Y_{n+1} = Y_n + a\Delta + b\Delta W_n \quad (\text{D.2})$$

donde $\Delta W_n = W_{t_{n+1}} - W_{t_n}$.

D.1.2. Método de Milstein

Considere el mismo proceso que en el apartado anterior, tal que $X = \{X_t, t_0 \leq t \leq T\}$ satisface la siguiente ecuación diferencial:

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t$$

Entonces en el esquema de Milstein sería de la forma:

$$Y_{n+1} = Y_n + a(t_n, Y_n)(t_{n+1} - t_n) + b(t_n, Y_n)(W_{t_{n+1}} - W_{t_n}) + \frac{1}{2}b(t_n, Y_n)b'(t_n, Y_n)\{(W_{t_{n+1}} - W_{t_n})^2 - (t_{n+1} - t_n)\}$$

lo cual en el caso de intervalos igualmente espaciados resulta ser:

$$Y_{n+1} = Y_n + a\Delta + b\Delta W_n + \frac{1}{2}bb'\{(\Delta W_n)^2 - \Delta\}$$

El esquema de Milstein es igual al esquema de Euler con el término adicional $\frac{1}{2}bb'\{(\Delta W_n)^2 - \Delta\}$.

Apéndice E

Salidas programas econométricos

E.1. Pruebas de raíz unitaria

```
>retorno.adf = ur.df(retorno,type = "none", selectlags = .AIC")
>summary(retorno.adf)
```

Augmented Dickey-Fuller Test Unit Root Test

Test regression none

Call:

```
lm(formula = z.diff z.lag.1 - 1 + z.diff.lag)
```

Residuals:

Min 1Q Median 3Q Max

-5.9534 -0.3078 -0.0152 0.2816 5.2028

Coefficients:

Estimate Std. Error t value Pr(> |t|)

*z.lag.1 -0.852500 0.024754 -34.439 <2e-16 ****

z.diff.lag 0.006253 0.019018 0.329 0.742

—

*Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Residual standard error: 0.6811 on 2765 degrees of freedom

Multiple R-squared: 0.4236, Adjusted R-squared: 0.4232

F-statistic: 1016 on 2 and 2765 DF, p-value: <2.2e-16

Value of test-statistic is: -34.4387

Critical values for test statistics:

1pct 5pct 10pct

tau1 -2.58 -1.95 -1.62

```
>retorno.pp = ur.pp(retorno, type = "Z-alpha", model = "constant", lags = "short")
```

```
>summary(retorno.pp)
```

Phillips-Perron Unit Root Test

Test regression with intercept

Call:

```
lm(formula = y ~ y.l1)
```

Residuals:

Min 1Q Median 3Q Max

-5.9546 -0.3011 -0.0104 0.2859 5.2090

Coefficients:

Estimate Std. Error t value Pr(> |t|)

(Intercept) -0.00535 0.01294 -0.413 0.679

y.l1 0.15273 0.01879 8.128 6.52e-16 ***

—

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.681 on 2766 degrees of freedom

Multiple R-squared: 0.02333, Adjusted R-squared: 0.02297

F-statistic: 66.06 on 1 and 2766 DF, p-value: 6.518e-16

Value of test-statistic, type: Z-alpha is: -2249.725

aux. Z statistics

Z-tau-mu -0.4113

```
>retorno.kpss = ur.kpss(retorno, type = "mu")
```

```
>summary(retorno.kpss)
```

*KPSS Unit Root Test**Test is of type: mu with 9 lags.**Value of test-statistic is: 0.0944**Critical value for a significance level of:**10pct 5pct 2.5pct 1pct**critical values 0.347 0.463 0.574 0.739**>retorno.ers= ur.ers(retorno, type = c("DF-GLS"), model = c(çonstant"), lag.max = 4)**>summary(retorno.ers)**Elliot, Rothenberg and Stock Unit Root Test**Test of type DF-GLS**detrending of series with intercept**Call:**lm(formula = dfgls.form, data = data.dfgls)**Residuals:**Min 1Q Median 3Q Max**-6.0085 -0.3320 -0.0374 0.2591 5.1821**Coefficients:**Estimate Std. Error t value Pr(> |t|)**yd.lag -0.89554 0.03769 -23.762 <2e-16 *****yd.diff.lag1 0.05303 0.03409 1.555 0.1200**yd.diff.lag2 0.05212 0.02960 1.761 0.0783 .**yd.diff.lag3 0.00518 0.02488 0.208 0.8351**yd.diff.lag4 0.04525 0.01902 2.379 0.0174 ***—**Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**Residual standard error: 0.6805 on 2759 degrees of freedom**Multiple R-squared: 0.4258, Adjusted R-squared: 0.4248**F-statistic: 409.2 on 5 and 2759 DF, p-value: <2.2e-16*

Value of test-statistic is: -23.762

Critical values of DF-GLS are:

1pct 5pct 10pct

critical values -2.57 -1.94 -1.62

E.2. Modelo de heteroscedasticidad condicional

* GARCH Model Fit *

Conditional Variance Dynamics

GARCH Model : sGARCH(1,1)

Mean Model : ARFIMA(1,0,2)

Distribution : norm

Optimal Parameters

Estimate Std. Error t value Pr(> |t|)

mu -0.018577 0.008743 -2.1248 0.033602

ar1 -0.949454 0.033496 -28.3452 0.000000

ma1 1.171334 0.038860 30.1428 0.000000

ma2 0.217473 0.021660 10.0403 0.000000

omega 0.003908 0.001106 3.5328 0.000411

alpha1 0.169782 0.015164 11.1960 0.000000

beta1 0.829218 0.015327 54.1006 0.000000

Robust Standard Errors:

Estimate Std. Error t value Pr(> |t|)

mu -0.018577 0.010124 -1.8350 0.066506

ar1 -0.949454 0.023558 -40.3028 0.000000

ma1 1.171334 0.037586 31.1638 0.000000

ma2 0.217473 0.028415 7.6535 0.000000

omega 0.003908 0.003373 1.1585 0.246652

alpha1 0.169782 0.036665 4.6306 0.000004

beta1 0.829218 0.049690 16.6879 0.000000

LogLikelihood : -2113.696

Information Criteria

Akaike 1.5317

Bayes 1.5467

Shibata 1.5317

Hannan-Quinn 1.5372

Q-Statistics on Standardized Residuals

statistic p-value

Lag[1] 1.275 0.25874

Lag[p+q+1][4] 4.733 0.02959

Lag[p+q+5][8] 6.735 0.24109

d.o.f=3

H0 : No serial correlation

Q-Statistics on Standardized Squared Residuals

statistic p-value

Lag[1] 0.000815 0.9772

Lag[p+q+1][3] 0.357249 0.5500

Lag[p+q+5][7] 0.838546 0.9745

d.o.f=2

ARCH LM Tests

Statistic DoF P-Value ARCH Lag[2] 0.01378 2 0.9931

ARCH Lag[5] 0.42311 5 0.9947

ARCH Lag[10] 2.93725 10 0.9829

Adjusted Pearson Goodness-of-Fit Test:

group statistic p-value(g-1)

1 20 55.78 1.763e-05

2 30 60.24 5.763e-04

3 40 79.93 1.222e-04

4 50 93.17 1.447e-04

E.3. Pruebas de Normalidad

```
nom= c("SVMRW", "SVMHI", "SVFBBS", "SVFBBSM", "SVSMC", "GARCH")
```

```
>shapiro.test(SVMRW)
```

Shapiro-Wilk normality test

data: SVMRW

W = 0.987, p-value = 1.037e-10

```
>shapiro.test(SVMHI)
```

Shapiro-Wilk normality test

data: SVMHI

W = 0.9903, p-value = 1.087e-08

```
>shapiro.test(SVFBBS)
```

Shapiro-Wilk normality test

data: SVFBBS

W = 0.9022, p-value <2.2e-16

```
>shapiro.test(SVFBBSM)
```

Shapiro-Wilk normality test

data: SVFBBSM

W = 0.9905, p-value = 1.456e-08

```
>shapiro.test(SVSMC)
```

Shapiro-Wilk normality test

data: SVSMC

W = 0.9945, p-value = 1.409e-05

```
>shapiro.test(GARCH)
```

Shapiro-Wilk normality test

data: GARCH

W = 0.9938, p-value = 3.559e-06

```
>jarque.bera.test(SVMRW)
```

JarqueBera Test

data: SVMRW

X-squared = 166.2885, df = 2, p-value <2.2e-16

```
>jarque.bera.test(SVMHI)
```

JarqueBera Test

data: SVMHI

X-squared = 103.0723, df = 2, p-value <2.2e-16

```
>jarque.bera.test(SVFBBS)
```

JarqueBera Test

data: SVFBBS X-squared = 11565.3, df = 2, p-value <2.2e-16

```
>jarque.bera.test(SVFBBSM)
```

JarqueBera Test

data: SVFBBSM

X-squared = 99.8743, df = 2, p-value <2.2e-16

```
>jarque.bera.test(SVSMC)
```

JarqueBera Test

data: SVSMC

X-squared = 15.3589, df = 2, p-value = 0.0004622

>jarque.bera.test(GARCH)

JarqueBera Test

data: GARCH

X-squared = 53.3222, df = 2, p-value = 2.638e-12

Bibliografía

- Alberg, D., Shalit, H., and Yosef, R. (2002). Estimating stock market volatility using asymmetric GARCH models. *Journal of Banking and Finance*, 26:1487–1503.
- Alexander, C. (2008a). *Market Risk Analysis: Quantitative Methods in Finance*, volume I. John Wiley and Sons, first edition.
- Alexander, C. (2008b). *Market Risk Analysis: Value at Risk Models*, volume IV. John Wiley and Sons, first edition.
- Andersen, T., Bollerslev, T., Diebold, F., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76.
- Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55.
- Anderson, B. D. and Moore, J. B. (1979). *Optimal Filtering*. Dover Publications.
- Arize, A., Osang, T., and Slottje, D. (2000). Exchange rate volatility and foreign trade: Evidence from thirteen LDCs. *Journal of Business and Economic Statistics*, 18.
- Awartani, B. and Corradi, V. (2005). Predicting the volatility of the SP-500 stock index via GARCH models: the role of asymmetriess. *International Journal of Forecasting*, 21:167–183.
- Bailey, M. and Tavlas, G. (1991). Exchange rate variability and direct investment. *The ANNALS of the American Academy of Political and Social Science*, 516:106–116.
- Banrep (1998). *Introducción al análisis económico: El caso colombiano*. Banco de la República y Siglo del Hombre Editores.
- Baum, L. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37:1554–1563.
- Baum, L. E., Petrie, T. P., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 41:164–171.

- Berkes, I., Horváth, L., and Kokoszka, P. (2003). Garch process: structure and estimation. *Bernoulli*, 9:201–227.
- Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons, second edition.
- Blanco, L. (2004). *Probabilidad*. Universidad Nacional de Colombia, first edition.
- Blanco, L. and Muñoz, M. (2003). *Notas de Clase: Análisis Estocástico*. Universidad Nacional de Colombia, second edition.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.
- Bollerslev, T., Chou, R., and Kroner, K. (1992). Arch modelling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, 32:5–59.
- Bozdogan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370.
- Brandt, M. and Jones, C. (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business and Economic Statistics*, 24(4):470–486.
- Carnero, M., Pena, D., and Ruiz, E. (2004). Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics*, 2:319–342.
- Casarin, R. (2012). Lectures Notes: Bayesian Methods in Economics and Finance. Technical Report 1, Summer School on Bayesian Methods in Economics and Finance, Trieste University.
- Casas, M. and Cepeda, E. (2008). Modelos ARCH, GARCH y EGARCH: Aplicaciones a series financieras. *Cuadernos de Economía*, 28(48):287–319.
- Choi, E. (2011). *Stochastic Volatility Models and Simulated Maximum Likelihood Estimation*. PhD thesis, University of Waterloo, Canada.
- Christoffersen, P., Jacobs, K., and Mimouni, K. (2010). Volatility dynamics for the SP500: Evidence from realized volatility, daily returns, and option prices. *Review of Financial Studies*, 23(8):3141–3189.
- Chui, C. K. and Chen, G. (1991). *Kalman Filtering with Real-Time Applications*. New York: Springer-Verlag, second edition.
- Clark, P. (1973). A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica*, 41:135–155.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

- Côte, A. (1994). Exchange rate volatility and trade: A survey. Technical Report 94, Working Papers-Bank of Canada.
- Creal, D. (2012). A survey of sequential monte carlo methods for economics and finance. *Econometric Reviews*, 31(3):245–296.
- Danielsson, J. (1994). Stochastic volatility in asset prices estimation with simulated maximum likelihood. *Journal of Econometrics*, 64:375–400.
- Danielsson, J. and Richard, J. F. (1993). Accelerated gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics*, 8:153–77.
- de Jong, D., Dharmarajan, H., Liesenfeld, R., Moura, G. V., and Richard, J. F. (2009). Efficient likelihood evaluation of state space representations. Technical report, Unpublished manuscript, Department of Economics, University of Pittsburgh.
- Dickey, D. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4):1057–72.
- Dornbusch, R. (1976). Expectations and exchange rate dynamics. *Journal of Political Economy*, 84:1161–1176.
- Dornbusch, R. (1987). Exchanges rates and prices. *American Economic Review*, 77:93–106.
- Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. New York, Springer-Verlag, first edition.
- Drost, F. and Nijman, T. (1993). Temporal aggregation of GARCH processes. *Econometrica*, 61(4):909–927.
- Drost, F. and Werker, B. (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Econometrica*, 61(4):909–927.
- Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, first edition.
- Durbin, J. and Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika*, 84:669–684.
- Echavarría, J. J., Melo, L., , Téllez, S., and Villamizar, M. (2013). The impact of pre-announced day-to-day interventions on the Colombian Exchange Rate. Technical Report 767, Borrador de Economía, Banco de la República.
- Echavarría, J. J., Vásquez, D. M., and Villamizar, M. (2009). Impacto de las intervenciones cambiarias sobre el nivel y la volatilidad de la tasa de cambio en Colombia. Technical Report 561, Borrador de Economía, Banco de la República.

- Elliott, G., Rothenberg, T., and Stock, J. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64:813–836.
- Enders, W. (2004). *Applied Econometric Time Series*. Wiley, second edition.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of uk inflation. *Econometrica*, 50(4):987–1008.
- Engle, R. (1983). Estimates of the variance of us inflation based upon the arch model. *Journal of Money, Credit and Banking*, 15:286–301.
- Engle, R. and Bollerslev, T. (1986). Modelling the persistence of conditional variance. *Econometric Reviews*, 5(1):1–50.
- Engle, R. and Ng, K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(1):1749–78.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, pages 383–417.
- Fleming, J. M. (1962). Domestic financial policies under fixed and under floating exchange rates. *International Monetary Fund Staff papers*, 9:369–79.
- Franses, P. H. and van Dijk, D. (2003). *Non-Linear Time Series Models in Empirical Finance*. Cambridge University Press, first edition.
- Gallant, A. R., Hsieh, D., and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, 8:153–173.
- Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC, second edition.
- Garber, P. and Spencer, M. (1996). Dynamic hedging and the interest rate defense. In Frankel, J., Galli, G., and Giovannini, A., editors, *The microstructure of foreign Exchange Markets*. The University of Chicago Press.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398 – 409.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC, second edition.
- Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

- Ghysels, E., Harvey, A. C., and Renault, E. (1996). Stochastic volatility. In Rao, C. R. and Maddala, G. S., editors, *Handbook of Statistics: Statistical Methods in Finance*, pages 119–91. Amsterdam: North-Holland.
- Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801.
- González, A., Rincón, H., and Rodríguez, N. (2010). La transmisión de los choques a la tasa de cambio sobre la inflación de los bienes importados en presencia de asimetrías. In Jalil, M. and Mahadeva, L., editors, *Mecanismos de Transmisión de la Política Monetaria en Colombia*. Banco Central de Colombia and Universidad Externado de Colombia.
- Gordon, N., Salmond, D., and Smith, A. (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proceedings*, 140(2):107–113.
- Greene, W. (2011). *Econometric Analysis*. Prentice Hall, seventh edition.
- Guerrero, G. (2003). *Análisis estadístico de series de tiempo económicas*. Thomson Learning, second edition.
- Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press, first edition.
- Hammersley, J. M. and Morton, K. W. (1954). Poor man monte carlo. *Journal of the Royal Statistical Society*, 16:23–38.
- Handschin, J. (1970). Monte carlo techniques for prediction and filtering of non-linear stochastic processes. *Automatica*, 6:555–563.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 8:85–118.
- Hansen, R. and Lunde, A. (2001). A comparison of volatility models: Does anything beat a garch(1,1)? Technical Report 84, Centre for Analytical Finance- University of Aarhus.
- Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264.
- Harvey, A., Ruiz, E., and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, 14:429–434.
- Hasting, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.

- Hoque, H. and Kim, J. Soo, C. (2007). A comparison of variance ratio test of random walk: A case of Asian emerging stock markets. *International Review of Economics and Finance*, 16:488 – 502.
- Hull, J. (2008). *Options, Futures, and Other Derivatives*. Prentice Hall, seventh edition.
- Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300.
- Isard, P. (1977). How far can we push the law of one price? *American Economic Review*, 67:942–48.
- Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:371–389.
- Johannes, M. and Polson, N. (2010). MCMC methods for continuous-time financial econometrics. In Yaít-Sahalia, Y. and Hansen, L. P., editors, *Handbook of Financial Econometrics*, volume 2, pages 1–72. Princeton University Press.
- Jorion, P. (2003). *Valor en Riesgo*. Limusa Noriega Editores, second edition.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.
- Karatzas, I. and Shreve, S. (1988). *Brownian motion and stochastic calculus*. Springer-Verlag Berlin Heidelberg.
- Kim, J. and Kim, S. (2004). Empirical comparison of alternative stochastic volatility option pricing models: Evidence from korean KOSPI 200 index options market. *Pacific Basin Finance Journal*, 12:117–142.
- Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies*, 65(3):361–93.
- Kloeden, P. and Platen, E. (1992). *Numerical solution of stochastic differential equations*. Springer Verlag Berlin Heidelberg.
- Kloek, T. and van Dijk, H. K. (1978). Bayesian estimations of equation system parameters: an application of Monte Carlo Integration. *Econometrica*, 46(1):1–20.
- Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Political Economy*, 54:159–178.
- Lam, K. and Ng, H. (2009). Intra-daily information of range-based volatility for MEM-GARCH. *Mathematics and Computers in Simulation*, 79:2625–2632.

- Lega, P. F., Murcia, A., Vásquez, D., and Venegas, T. (2007). Volatilidad de la tasa de cambio nominal en Colombia y su relación con algunas variables. Technical Report 407, Borrador de Economía, Banco de la República.
- Liesenfeld, R. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, 10:505–531.
- Lildholdt, P. (2002). Estimation of GARCH models based on open, close, high, and low prices. Technical Report 128,, Centre for Analytical Finance- University of Aarhus.
- Liu, J. and West, M. (2001). Combined parameters and state estimation in simulation-based filtering. In Doucet, A., De Freitas, N., and Gordon, N., editors, *Sequential Monte Carlo Methods in Practice*, pages 197–223. New York, Springer-Verlag.
- Liu, Z. (2007). *Direct simulation method for multiple changepoint problems*. PhD thesis, Lancaster University.
- Lo, A. and Mackinlay, A. (1988). Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies*, 1:41–66.
- Lopes, H. F. and Tsay, R. S. (2011). Particle filters and bayesian inference in financial econometrics. *Journal of Forecasting*, 30(1):168–209.
- Mapa, D. (2004). A forecast comparison of financial volatility models: GARCH (1,1) is not enough. Technical Report 21028, Munich Personal RePEc Archive - MPRA.
- Marshall, A. W. (1956). The use of multi-stage sampling schemes in Monte Carlo computations. In Meyer, M., editor, *Symposium on Monte Carlo Methods*, pages 123–140. New York-Wiley.
- Matei, M. (2009). Assessing volatility forecasting models: Why GARCH models take the lead. *Journal of Political Economy*, 4:42–65.
- Meddahi, N. and Renault, E. (2004). Temporal aggregation of volatility models. *Journal of Econometrics*, 119(2):355–377.
- Meddahi, N. and Renault, E. (2006). GARCH and irregularly spaced data. *Economics Letters*, 90:200–204.
- Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45:239–265.
- Melo, L. and Becerra, O. (2005). *Medidas de riesgo, características y técnicas de una medición: una aplicación del VaR y el ES a la tasa interbancaria de Colombia*. Universidad del Rosario, first edition.
- Mesa, D. C. (2010). Spillovers de volatilidad entre tasa de interés y tasa de cambio en Colombia. Technical report, Universidad del Rosario.

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.
- Micchiché, S., Bonanno, G., Lillo, F., and Mantegna, R. N. (2002). Volatility in financial markets: stochastic models and empirical results. *Physica*, 314:756–761.
- Mills, T. C. and Markellos, R. N. (2008). *The Econometric Modelling of Financial Time Series*. Cambridge University Press, third edition.
- Mundell, R. A. (1963). Capital mobility and stabilization policy under fixed and flexible exchange rates. *Canadian Journal of Economics and Political Science*, 29:475–85.
- Nelson, D. (1990). Stationarity and persistence in the garch(1,1) model. *Econometric Theory*, 6:318–334.
- Nelson, D. (1991). Conditional heteroscedasticity in asset returns: a new approach. *Econometrica*, 59:347–370.
- Nelson, D. and Cao, C. (1992). Inequality constraints in the univariate garch model. *Journal of Business and Economics Statistics*, 10:229–235.
- Oksendal, P. (2005). *Stochastic Differential Equations An Introduction with Applications*. Springer-Verlag Berlin Heidelberg.
- Parra, J. C. (2010). La sensibilidad de los precios del consumidor a la tasa de cambio en colombia: Una medición de largo plazo. In Jalil, M. and Mahadeva, L., editors, *Mecanismos de Transmisión de la Política Monetaria en Colombia*. Banco Central de Colombia and Universidad Externado de Colombia.
- Peña, D. (2005). *Análisis de series temporales*. Alianza editorial, first edition.
- Phillips, P. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75:335–346.
- Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.
- Posedel, P. (2005). Properties and Estimation of GARCH(1,1) Model. *Metodoloski zvezki*, 2(2):243–257.
- Praetz, P. (1972). The distribution of share price changes. *Journal of Business*, 45:49–55.
- Pulgarin, L. (2001). *Applied comparison between GARCH and stochastic volatility models*. PhD thesis, University of Pittsburgh.
- Qu, Z. and Perron, P. (2013). A stochastic volatility model with random level shifts and its applications to SP500 and NASDAQ return indices. *The Econometrics Journal*, 16(3):309–339.

- Raftery, A. and Lewis, S. (2010). How many iterations in the Gibbs Sampler? In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics*, volume 4, pages 763–773. Oxford University Press.
- Richard, J. F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411.
- Rincón, H., Caicedo, E., and Rodríguez, N. (2005). Exchange rate pass-through effects: A disaggregate analysis of colombian imports of manufactured goods. Technical Report 330, Banco de la República.
- Rincón, H. and Toro, J. (2010). Are capital controls and central bank intervention effective? Technical Report 625, Borradores de Economía, Banco de la República.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. New York: Springer Press, second edition.
- Rosenberg, B. (1972). The behaviour of random variables with nonstationary variance and the distribution of security price. Technical report.
- Rubin, D. (1988). Using the SIR algorithm to simulate posterior distributions. In Bernardo, J. M., DeGroot, M. H., and Smith, A., editors, *Bayesian Statistics*, New-York. Oxford University Press.
- Ruiz, E. and Veiga, H. (2008). Modelos de volatilidad estocástica: una alternativa atractiva y factible para modelizar la evolución de la volatilidad. *Anales de Estudios Económicos y Empresariales*, 18:9–68.
- Sandmann, G. and Koopman, S. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, 87:271–301.
- Sarno, L. and Taylor, M. (2002). *The Economics of Exchange Rates*. Cambridge university press.
- Shephard, N. and Pitt, M. K. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84:653–667.
- Sánchez, F., Fernández, A., and Armenta, A. (2005). Historia monetaria de colombia en el siglo xx: Grandes tendencias y episodios relevantes. Technical Report 30, Universidad de los Andes-Documento CEDE.
- So, M. K. and Yu, P. L. (2006). Empirical analysis of GARCH models in value at risk estimation. *International Financial Markets, Institutions and Money*, 16:180–197.
- Taylor, J. B. (2000). Low inflation, pass-through, and the pricing power of firms. *European Economic Review*, 44:1389–1408.

- Taylor, S. (1982). Financial returns modelled by the product of two stochastic process, a study of daily sugar prices. In *Time Series Analysis: Theory and Practice*. Amsterdam, North Holland.
- Taylor, S. (1986). *Modelling Financial Time Series*. Wiley.
- Taylor, S. J. (2005). *Asset Price Dynamics, Volatility, and Prediction*. Princeton University Press, Princeton and Oxford.
- Terasvirta, T. and Zhao, Z. (2007). Stylized facts of return series, robust estimates and three popular models of volatility. Technical Report 662, SSE/EFI Working Paper Series in Economics and Finance.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1762.
- Toro, J., Mora, R., and Parra, D. (2013). Flujos de capitales, la crisis financiera internacional y los desbalances macroeconómicos. In Rincon, H. and Velasco, A., editors, *Flujos de capitales, choques externos y respuestas de política en países emergentes*. Banco de la República - Colombia.
- Tsay, R. (2010). *Analysis of Financial Time Series*. Wiley, third edition.
- Tudor, C. (1997). *Procesos estocásticos*. Sociedad Matemática Mexicana.
- Venegas, F. (2007). *Riesgos financieros y económicos. Productos derivados y decisiones económicas bajo incertidumbre*. Thomson Learning, first edition.
- West, M. and Harrison, J. (2007). *Bayesian Forecasting and Dynamic Models*. New York: Springer, second edition.
- Wright, J. (2000). Alternative variance-ratio tests using ranks and signs. *Journal of Business and Economic Statistics*, 18(1):1–9.