UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN

TRABAJO FINAL DE GRADO PARA OBTENER EL TÍTULO DE LICENCIADO EN ESTADÍSTICA

INFERENCIA BAYESIANA EN MODELOS DE VOLATILIDAD ESTOCÁSTICA

JUAN LUCIANO GARRIDO VIDAL
DIEGO ALTAMIRANDA

Tutor:
MARCO SCAVINO

Tribunal:
Leonardo Moreno
Fernando Massa
Marco Scavino

Montevideo
URUGUAY
2025

Resumen

En los modelos de volatilidad estocástica, la volatilidad es un proceso latente no observable directamente, lo que complica tanto la estimación de sus parámetros como su inferencia en tiempo real.

El objetivo principal de este trabajo es presentar un esquema de inferencia eficiente para estos modelos, abordando tanto la estimación de los parámetros como la predicción del estado latente de la volatilidad. Para ello, se adopta un enfoque basado en un modelo bayesiano jerárquico, en el que la volatilidad sigue una estructura probabilística multinivel. En este marco, se implementa un método de muestreo de Gibbs dentro de un esquema cadena de Markov Monte Carlo (MCMC), lo que permite obtener estimaciones consistentes de los parámetros del modelo. Sin embargo, debido a la naturaleza *batch* del muestreo MCMC, este método no resulta óptimo para realizar predicciones en línea. Para superar esta limitación, se introduce un filtro de partículas, que permite estimar y actualizar la volatilidad en tiempo real a medida que se observan nuevos datos.

Para ilustrar la aplicación de las técnicas propuestas, se realiza un estudio empírico utilizando datos reales correspondientes al precio diario de Bitcoin (BTC) en dólares estadounidenses (USD) durante el periodo 2022-2024, un activo que se caracteriza por su marcada volatilidad.

Palabras Clave

Bayes Jerárquico, Volatilidad Estocástica, Cadenas de Markov Monte Carlo, Metropolis-Hastings, Muestreo de Gibbs, Filtro de Partículas, Hammersley-Clifford, Bitcoin.

Índice general

Índice de figuras

"Lo esencial es invisible a los ojos."

— Antoine de Saint-Exupéry, *El Principito*

En esta tesis, nos proponemos estudiar precisamente eso que es invisible.
Una fuerza subyacente, imperceptible a simple vista, pero fundamental.
Como en la obra de Saint-Exupéry, lo que no se ve es, muchas veces, lo que más importa.

Capítulo 1

Modelo de Volatilidad Estocástica

En el análisis de series financieras, es común observar que la varianza de los retornos cambia a lo largo del tiempo, un fenómeno conocido como *heteroscedasticidad*. Este comportamiento ha sido ampliamente documentado en estudios empíricos (Bollerslev, 1986; Engle, 1982), lo que ha motivado el desarrollo de modelos específicos para describir la evolución temporal de la volatilidad.

Uno de los enfoques más utilizados es el *Modelo de Volatilidad Estocástica* (SV), el cual introduce una estructura más flexible al permitir que la volatilidad evolucione como un proceso estocástico (Taylor, 1982). Este modelo es especialmente adecuado para describir la dinámica de los retornos logarítmicos y_t de un activo financiero, definidos como:

$$y_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \quad t = 1, 2, \ldots, T, \tag{1.1}$$

donde P_t representa el precio del activo en el tiempo t. Los retornos logarítmicos son ampliamente utilizados en econometría financiera debido a sus propiedades estadísticas favorables, como la aproximación a la normalidad en períodos cortos y la estabilidad de su escala (Cont, 2001).

Para capturar la variabilidad en los retornos, el modelo de volatilidad estocástica asume que y_t sigue un proceso con varianza condicional e^{h_t}, donde h_t es una variable que no se puede observar. Formalmente, se supone que:

$$y_t \mid h_t \sim N(0, e^{h_t}), \tag{1.2}$$

lo que implica que, condicionado a h_t, los retornos y_t son distribuidos normalmente con media cero y varianza e^{h_t}. La media cero refleja la hipótesis de eficiencia de los mercados en el corto plazo (Fama, 1970), mientras que la varianza condicional h_t modela

la incertidumbre presente en el mercado.

La densidad condicional de y_t bajo este modelo se obtiene a partir de la distribución normal:

$$p(y_t \mid h_t) \propto h_t^{-0.5} \exp\Big(-\frac{0.5\, y_t^2}{h_t}\Big) \tag{1.3}$$

A diferencia de otros enfoques paramétricos, como los modelos GARCH, donde la volatilidad se modela mediante una estructura determinista dependiente de valores pasados, el modelo SV asume que h_t sigue un proceso estocástico. Esto permite una mayor flexibilidad en la dinámica de la volatilidad, capturando fenómenos como efectos de memoria larga, presencia de saltos y comportamiento asimétrico en los retornos (Ghysels et al., 1996; Kim et al., 1998).

En términos de aplicabilidad, los modelos SV no solo han demostrado ser útiles en el ámbito financiero, sino que también han sido utilizados en otras disciplinas donde la variabilidad de un proceso cambia en el tiempo, como la biología, la climatología y la ingeniería de señales (Jacquier et al., 2002).

1.1. Modelo Autorregresivo de la Volatilidad

La volatilidad latente h_t se modela como un proceso Autorregresivo de primer orden con reversión a la media (Taylor, 1982). Este modelo sigue la ecuación:

$$h_t = \mu + \phi_h(h_{t-1} - \mu) + \sigma_h \eta_t, \quad \eta_t \sim \mathcal{N}(0,1) \tag{1.4}$$

$$h_t \in R,\ \mu \in R,\ \phi_h \in (-1,1),\ \sigma_h \in R^+$$

donde μ representa el nivel medio de volatilidad, ϕ_h controla la persistencia temporal y σ_h es la desviación estándar del término de innovación. La condición $|\phi_h| < 1$ garantiza estacionariedad en el proceso.

Bajo este enfoque, la volatilidad h_t depende linealmente de su rezago inmediato y un error aleatorio, lo que implica que, condicional en h_{t-1} y los parámetros del modelo $\omega = (\mu, \phi_h, \sigma_h^2)$, su distribución es:

$$h_t \mid h_{t-1}, \omega \sim N(\mu + \phi_h(h_{t-1} - \mu), \sigma_h^2) \tag{1.5}$$

Este modelo exhibe reversión a la media, es decir, h_t tiende a regresar a su nivel medio μ con una velocidad determinada por ϕ_h. La esperanza condicional lo refleja claramente:

$$E[h_t \mid h_{t-1}] = \mu + \phi_h(h_{t-1} - \mu) \tag{1.6}$$

Si h_{t-1} es mayor que μ, la volatilidad tiende a reducirse en la siguiente iteración, y viceversa. Así, el modelo captura de manera eficiente la dinámica de la volatilidad en series temporales.

1.2. Modelos Jerárquicos

En la formulación tradicional de un modelo sin variables latentes, la verosimilitud de un conjunto de datos $y = \{y_1, \ldots, y_T\}$, dado un conjunto de parámetros ω, se define como la probabilidad conjunta de los datos. Bajo el supuesto de independencia entre las observaciones, esta verosimilitud se expresa como:

$$\ell(\omega) = p(y \mid \omega) = \prod_{t=1}^{T} p(y_t \mid \omega) \tag{1.7}$$

donde $p(y_t \mid \omega)$ representa la función de densidad de cada dato individual bajo los parámetros del modelo. Este enfoque es directo y computacionalmente manejable cuando no existen dependencias complejas entre las observaciones.

No obstante, en muchos problemas prácticos, los datos dependen de variables latentes no observadas, lo que complica la construcción de la verosimilitud. Este es el caso de los *modelos jerárquicos*, una clase de modelos estadísticos diseñados para representar relaciones entre datos observados, variables latentes y parámetros a través de niveles organizados jerárquicamente (Gelman et al., 2013). Este enfoque resulta especialmente útil cuando las dependencias en los datos pueden descomponerse en niveles conceptuales.

En el modelo jerárquico que se presenta a continuación, los datos observados y_t de-

penden de variables latentes h_t, que evolucionan en el tiempo según parámetros globales ω. Desde una perspectiva bayesiana, estos parámetros ω son considerados aleatorios y se les asigna una distribución a priori. La probabilidad conjunta se puede expresar como:

$$p(y, h, \omega) = \prod_{t=1}^{T} p(y_t \mid h_t)\, p(h_t \mid \omega)\, p(\omega) \tag{1.8}$$

donde

$p(y_t \mid h_t)$ describe cómo los datos observados dependen de las volatilidades latentes h_t,

$p(h_t \mid \omega)$ captura la evolución temporal de esas volatilidades bajo los parámetros ω,

$p(\omega)$ es el prior sobre los parámetros globales.

Esta estructura jerárquica introduce una mayor flexibilidad al modelo (Robert and Casella, 2004), permitiendo capturar dependencias complejas, pero también hace más difícil el cálculo de la verosimilitud marginal $p(y \mid \omega)$, ya que en este caso es necesario integrar sobre todas las configuraciones posibles de las variables latentes h:

$$\ell(\omega) \;=\; p(y \mid \omega) \;=\; \int p(y \mid h)\, p(h \mid \omega)\, dh.^{1}$$

Este promedio ponderado incluye todas las posibles realizaciones de las variables latentes h, lo que con frecuencia conduce a una integral no resoluble de forma analítica debido a la alta dimensionalidad de h.

Para abordar esta complejidad, en lugar de trabajar directamente con la verosimilitud marginal $p(y \mid \omega)$, el enfoque bayesiano reestructura el problema en términos de la distribución posterior conjunta:

Teorema 1 (Teorema de Bayes)**.** *Sea $p(y, h, \omega)$ la probabilidad conjunta de los datos observados y, las variables latentes h y los parámetros ω. El Teorema de Bayes establece que la distribución posterior conjunta de h y ω dado y se expresa como*

$$\pi(h, \omega \mid y) = \frac{p(y, h, \omega)}{p(y)} \tag{1.9}$$

[1] Aquí, $p(y \mid h) = \prod_{t=1}^{T} p(y_t \mid h_t)$, y $p(h \mid \omega)$ captura la dependencia entre las volatilidades h_t y los parámetros globales ω.

donde $p(y)$ es la verosimilitud marginal, definida como $p(y) = \int p(y, h, \omega)\, d(h, \omega)$.

Observando que $p(y, h, \omega) = p(y \mid h)\, p(h \mid \omega)\, p(\omega)$ por la regla de la probabilidad condicionada, se obtiene que

$$\pi(h, \omega \mid y) \;=\; \frac{p(y \mid h)\, p(h \mid \omega)\, p(\omega)}{p(y)} \qquad \propto \qquad p(y \mid h)\, p(h \mid \omega)\, p(\omega),$$

donde el término $p(y)$ es constante con respecto a (h, ω) y, por tanto, puede omitirse al escribir la densidad en forma proporcional. Esta reformulación permite descomponer el problema en componentes más manejables, separando explícitamente la contribución de los datos y la estructura jerárquica impuesta por las variables latentes y los parámetros globales.

1.3. Análisis Bayesiano y Cadenas de Markov Monte Carlo (MCMC)

En lugar de calcular directamente $p(y \mid \omega)$, se recurre a métodos de simulación basados en cadenas de Markov Monte Carlo (MCMC), que permiten muestrear de la distribución posterior sin necesidad de resolver explícitamente la integral (Chib and Greenberg, 1995).

El enfoque MCMC construye una cadena de Markov cuya distribución invariante es la posterior $\pi(h, \omega \mid y)$. Esto permite obtener muestras representativas de h y ω, facilitando la inferencia sobre estos parámetros sin necesidad de evaluaciones directas de la verosimilitud marginal. Métodos como el muestreo de Gibbs y el algoritmo de Metropolis-Hastings son ampliamente utilizados para este propósito, y han demostrado ser herramientas fundamentales en la inferencia bayesiana para modelos de volatilidad estocástica (Kim et al., 1998).

1.3.1. Cadenas de Markov y su Aplicación en la Inferencia Bayesiana

La idea principal de MCMC es construir una cadena de Markov cuya distribución estacionaria sea precisamente la distribución posterior $\pi(h, \omega \mid y)$, de manera que, tras

un número suficiente de iteraciones, las muestras generadas por la cadena se comporten como si provinieran directamente de la posterior.

Definición 1 (Cadena de Markov)**.** *Una cadena de Markov es un proceso estocástico* $\{X_t\}_{t=0}^{\infty}$ *que satisface la propiedad de Markov, es decir, el estado futuro* X_{t+1} *depende únicamente del estado presente* X_t *y no de los estados anteriores:*

$$P(X_{t+1} = x \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = P(X_{t+1} = x \mid X_t = x_t).$$

En nuestro caso, el espacio de estados de la cadena de Markov está compuesto por las configuraciones posibles de las variables latentes h y los parámetros ω. El objetivo es hacer que esta cadena de Markov converja a una distribución estacionaria que coincida con la posterior conjunta $\pi(h, \omega \mid y)$.

1.3.2. Construcción y Análisis de la Cadena de Markov

Para aplicar el método MCMC, el primer paso es construir una cadena de Markov cuya distribución estacionaria coincida con la distribución posterior conjunta $\pi(h, \omega \mid y)$. Esto se logra definiendo una regla de transición que describe cómo la cadena evoluciona de un estado $(h^{(t)}, \omega^{(t)})$ al siguiente estado $(h^{(t+1)}, \omega^{(t+1)})$, asegurando que la secuencia de muestras generadas eventualmente se distribuye según la posterior (Robert and Casella, 2004).

Los algoritmos más comunes para implementar esta estrategia son el algoritmo de Metropolis-Hastings y el muestreo de Gibbs (Chib and Greenberg, 1995; Gelman et al., 2013). Cada uno de estos métodos emplea reglas específicas para proponer nuevos estados en la cadena y aceptar o rechazar los movimientos de manera que la distribución invariante de la cadena sea la deseada.

Propiedades necesarias para la convergencia

Para garantizar que las muestras generadas converjan a la distribución estacionaria, la cadena de Markov debe cumplir las siguientes propiedades (Tierney, 1994):

- **Irreducibilidad**: La cadena debe ser capaz de alcanzar cualquier estado en el espacio de parámetros con probabilidad positiva, partiendo de cualquier estado

inicial. Esto asegura que no existan regiones inaccesibles dentro del soporte de la distribución posterior.

- **Aperiodicidad**: La cadena no debe quedar atrapada en ciclos regulares, lo que evita oscilaciones deterministas entre estados. En la práctica, esta propiedad se garantiza permitiendo que la cadena permanezca en el mismo estado con probabilidad positiva.

El cumplimiento de estas condiciones permite justificar que las muestras generadas tras un período de calentamiento (*burn-in*) son representativas de la distribución posterior y pueden ser utilizadas para la inferencia de los parámetros (Robert and Casella, 2004; Gelman et al., 2013).

1.3.3. Teorema de Hammersley-Clifford

Este teorema establece un vínculo entre las distribuciones conjuntas y las estructuras de independencia condicional en grafos probabilísticos, proporcionando una herramienta clave para descomponer distribuciones conjuntas en factores más manejables (Hammersley and Clifford, 1971; Besag, 1974; Lauritzen, 1996).

Teorema 2 (Teorema de Hammersley-Clifford)**.** *Sea $G = (V, E)$ un grafo no dirigido donde los nodos V representan un conjunto de variables aleatorias $X = \{X_1, X_2, \ldots, X_n\}$. Una distribución de probabilidad conjunta $p(X)$ que es estrictamente positiva, es decir, $p(X) > 0$ para todo X, satisface la propiedad de Markov respecto a G si y solo si puede factorizarse como un producto de funciones de los clics máximos de G:*

$$p(X) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(X_C),$$

donde $\mathcal{C}$ es el conjunto de clics máximos de G, $\psi_C(X_C)$ son funciones de potencial no negativas, y Z es una constante de normalización (Lauritzen, 1996).

Este resultado implica que si la estructura de independencia condicional de un conjunto de variables puede representarse mediante un grafo de Markov, la distribución conjunta puede descomponerse en términos de distribuciones condicionales locales. Esta propiedad es la base de algoritmos como el *muestreo de Gibbs*, ya que permite muestrear

de la distribución conjunta a través de distribuciones condicionales completas (Geman and Geman, 1984; Besag, 1986).

Corolario 1. *Si la distribución conjunta $p(A, B, C, \dots)$ es estrictamente positiva y satisface la propiedad de Markov respecto a un grafo no dirigido, entonces puede descomponerse completamente en términos de sus distribuciones condicionales completas:*

$$p(A, B, C, \dots) \propto p(A \mid B, C, \dots) \cdot p(B \mid A, C, \dots) \cdots .$$

Dado que en modelos de volatilidad estocástica la estructura del modelo puede representarse mediante grafos no dirigidos, el teorema justifica el uso del muestreo de Gibbs como un método eficiente para la simulación de la posterior conjunta $\pi(h, \omega \mid y)$ (Jacquier et al., 2002). Esto se debe a que, en estos modelos, la volatilidad latente y los parámetros globales presentan relaciones de independencia condicional que permiten la factorización de la posterior en términos de distribuciones condicionales completas, reduciendo así la complejidad del muestreo.

Modelos bayesianos como grafos no dirigidos

En los modelos bayesianos jerárquicos, la estructura probabilística suele representarse mediante grafos probabilísticos, que pueden ser dirigidos (redes bayesianas) o no dirigidos (campos aleatorios de Markov). La representación como grafo no dirigido es especialmente útil en modelos donde las relaciones de independencia condicional no imponen un orden causal estricto, sino que reflejan simetrías entre variables (Bishop, 2006; Jordan, 2003).

En nuestro caso, el modelo de volatilidad estocástica se expresa como:

$$p(y, h, \omega) = p(y \mid h)\, p(h \mid \omega)\, p(\omega) \tag{1.10}$$

lo que implica una estructura de independencia condicional entre y y ω, dado h. Esta estructura encaja naturalmente en un grafo no dirigido, ya que no hay una relación causal entre los parámetros globales y las observaciones, sino una relación de dependencia estructural que puede representarse mediante un grafo de Markov (Lauritzen, 1996).

Aplicación en Modelos Jerárquicos

En el contexto de modelos jerárquicos, el teorema de Hammersley-Clifford asegura que la distribución posterior conjunta $\pi(h, \omega \mid y)$ puede determinarse completamente a partir de las distribuciones condicionales:

$$p(h \mid \omega, y) \quad \text{y} \quad p(\omega \mid h, y).$$

Esto significa que podemos aproximar la posterior conjunta muestreando alternadamente de estas distribuciones condicionales, lo cual es la base del muestreo de Gibbs.

Además si la distribución condicional de los parámetros $p(\omega \mid h, y)$ es demasiado compleja para muestrear directamente, el teorema nos permite descomponerla en componentes más manejables. Supongamos que el vector de parámetros ω pueden dividirse en k componentes $(\omega_1, \omega_2, \ldots, \omega_k)$. La distribución condicional completa de ω puede descomponerse como:

$$p(\omega \mid h, y) = p(\omega_1 \mid \omega_2, \ldots, \omega_k, h, y) \cdot p(\omega_2 \mid \omega_1, \omega_3, \ldots, \omega_k, h, y) \cdots p(\omega_k \mid \omega_1, \ldots, \omega_{k-1}, h, y).$$

Esta partición adicional simplifica el problema, reduciendo la dimensionalidad y facilitando el muestreo.

1.3.4. Muestreador de Gibbs

El muestreador de Gibbs es uno de los algoritmos más utilizados en los métodos de Monte Carlo por cadenas de Markov (MCMC) para aproximar distribuciones complejas (Geman and Geman, 1984; Casella and George, 1992; Robert and Casella, 2004). A diferencia de otros métodos como el algoritmo de Metropolis-Hastings, el muestreador de Gibbs se basa directamente en el teorema de Hammersley-Clifford, el cual asegura que una distribución conjunta puede determinarse completamente a partir de sus distribuciones condicionales completas. Esto simplifica considerablemente el proceso de muestreo, ya que permite descomponer la distribución conjunta en pasos más manejables.

El objetivo del muestreador de Gibbs es generar una secuencia de muestras $(h^{(t)}, \omega^{(t)})$ que se distribuyan según la posterior conjunta $\pi(h, \omega \mid y)$. En lugar de intentar muestrear directamente de esta distribución, lo hacemos alternando entre las distribuciones

condicionales:

$$p(h \mid \omega, y) \quad \text{y} \quad p(\omega \mid h, y).$$

Definición 2 (Muestreador de Gibbs)**.** *Sea $\pi(x_1, x_2, \ldots, x_d)$ la densidad conjunta objetivo definida sobre el espacio $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_d$ donde cada $\mathcal{X}_i$ denota el espacio de valores posibles de la componente x_i. Denotamos por*

$$\pi(x_i \mid x_{-i})$$

la densidad condicional completa de la i-ésima componente, donde $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$.

El muestreador de Gibbs construye una cadena de Markov $\{X^{(n)}\}_{n\geq 0}$ con $X^{(n)} = (x_1^{(n)}, x_2^{(n)}, \ldots, x_d^{(n)})$ cuya distribución estacionaria es π mediante el siguiente procedimiento iterativo (Casella and George, 1992):

1. ***Inicialización:*** *Se elige un valor inicial arbitrario $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_d^{(0)})$.*
2. ***Actualización:*** *Para cada iteración $n = 0, 1, 2, \ldots$, se actualizan secuencialmente las componentes según:*

$$x_1^{(n+1)} \sim \pi(x_1 \mid x_2^{(n)}, \ldots, x_d^{(n)}),$$

$$x_2^{(n+1)} \sim \pi\big(x_2 \mid x_1^{(n+1)}, x_3^{(n)}, \ldots, x_d^{(n)}\big),$$

$$\vdots$$

$$x_d^{(n+1)} \sim \pi\big(x_d \mid x_1^{(n+1)}, \ldots, x_{d-1}^{(n+1)}\big).$$

Bajo condiciones apropiadas (como irreducibilidad y aperiocidad), la cadena $\{X^{(n)}\}$ converge en distribución a la densidad $\pi(x_1, x_2, \ldots, x_d)$ (Robert and Casella, 2004).

El algoritmo sigue un procedimiento iterativo. Primero, se inicializan $h^{(0)}$ y $\omega^{(0)}$ con valores arbitrarios. Luego, en cada iteración t, se realizan los siguientes pasos: muestrear $h^{(t)}$ de la distribución condicional $p(h \mid \omega^{(t-1)}, y)$ y, a continuación, muestrear $\omega^{(t)}$ de la distribución condicional $p(\omega \mid h^{(t)}, y)$. Este proceso se repite hasta alcanzar el número deseado de iteraciones.

Una de las grandes ventajas del muestreador de Gibbs es su simplicidad, ya que no requiere calcular una función de aceptación como en el algoritmo de Metropolis-Hastings. Además, al muestrear directamente de las distribuciones condicionales completas, se

evita el problema de rechazar propuestas, lo que mejora la eficiencia del algoritmo (Bishop, 2006).

Sin embargo, para que el muestreador de Gibbs funcione correctamente, es crucial que las distribuciones condicionales sean fáciles de muestrear. Si estas son complicadas o no tienen una forma analítica conocida, puede ser necesario recurrir a métodos adicionales para aproximarlas (Robert and Casella, 2004).

1.3.5. Algoritmo de Metropolis-Hastings

El algoritmo de Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970; Robert and Casella, 2004) es un método MCMC que permite muestrear distribuciones de probabilidad complejas cuando no se dispone de una forma analítica para simular directamente desde ellas. Su objetivo es construir una cadena de Markov cuya distribución estacionaria coincida con la distribución objetivo $\pi(x)$, sin necesidad de conocer su constante de normalización.

El algoritmo de Metropolis-Hastings se describe a continuación de manera esquemática

Algorithm 1 Algoritmo de Metropolis-Hastings

Entrada: $\pi(x)$ (distribución objetivo), $q(x' \mid x)$ (distribución de propuesta), $x^{(0)}$ (estado inicial), N (número de iteraciones)
Salida: Conjunto de muestras $\{x^{(t)}\}_{t=1}^{N}$ aproximadamente distribuidas según $\pi(x)$

for $t = 1$ hasta N **do**
  Proponer un nuevo estado $x^* \sim q(x^* \mid x^{(t-1)})$
  Calcular la probabilidad de aceptación:

$$\alpha = \text{mín}\left(1,\ \frac{\pi(x^*)q(x^{(t-1)} \mid x^*)}{\pi(x^{(t-1)})q(x^* \mid x^{(t-1)})}\right)$$

  Generar $u \sim \mathcal{U}(0,1)$
  if $u \leq \alpha$ **then**
    Aceptar la propuesta: $x^{(t)} \leftarrow x^*$
  else
    Rechazar la propuesta y mantener $x^{(t)} \leftarrow x^{(t-1)}$
  end if
end for
Devolver: $\{x^{(t)}\}_{t=1}^{N}$

El atractivo principal de Metropolis-Hastings radica en que no exige una forma cerrada para muestrear $\pi(x)$. Basta con poder evaluar $\pi(x)$ hasta una constante de proporcionalidad y definir una distribución de propuesta $q(x^* \mid x)$ que cubra de manera efectiva el espacio de parámetros (**?**). El balance entre exploración y aceptación se controla ajustando la varianza de $q(\cdot \mid \cdot)$:

- Si q es demasiado *estrecha*, se proponen cambios muy pequeños y la cadena explora lentamente la distribución objetivo, aunque con una alta tasa de aceptación.
- Si q es demasiado *amplia*, la mayoría de propuestas serán rechazadas, lo que también se traduce en una exploración lenta.

En nuestro contexto, el algoritmo de Metropolis-Hastings puede aplicarse cuando las distribuciones condicionales $p(h \mid \omega, y)$ o $p(\omega \mid h, y)$ no son directamente simulables. En tal caso, se define una distribución de propuesta adecuada (por ejemplo, normal multivariante centrada en el valor actual de (h, ω)) y se itera siguiendo los pasos anteriores. Este procedimiento genera una cadena de Markov con distribución estacionaria $\pi(h, \omega \mid y)$, siempre que la cadena sea irreducible, aperiódica y ergódica (Tierney, 1994).

1.4. Estimación de los Parámetros

Como ya vimos, podemos descomponer la distribución posterior conjunta de h y ω como:

$$\pi(h, \omega \mid y) = p(h \mid \omega, y)p(\omega \mid h, y). \tag{1.11}$$

1.4.1. Muestrear ω desde su distribución condicional $p(\omega \mid h, y)$

El muestreo de los parámetros $\omega = (\mu, \phi_h, \sigma_h)$, que gobiernan la dinámica de la volatilidad en el modelo Autorregresivo con reversión a la media, se basa en la distribución condicional $p(\omega \mid h, y)$. Considerando el modelo de la ecuación 1.4 es necesario especificar distribuciones a priori que reflejen nuestras creencias iniciales sobre estos parámetros antes de observar los datos. En este caso, la prior conjunta de los parámetros del modelo es:

$$p\left(\mu, \phi_h, \sigma_h\right)$$

Para facilitar la inferencia, utilizamos una descomposición que separa la varianza del proceso σ_h^2 del resto de los parámetros:

$$p(\mu, \phi_h, \sigma_h) = p(\mu, \phi_h \mid \sigma_h)\, p(\sigma_h) \tag{1.12}$$

Esta formulación es conveniente porque permite modelar la varianza σ_h^2 de manera independiente, mientras que la estructura conjunta de (μ, ϕ_h) se captura condicionalmente en σ_h.

Vamos a suponer que la varianza σ_h^2 sigue una distribución gamma inversa,

$$\sigma_h^2 \sim IG(v_0, s_0^2) \tag{1.13}$$

donde los hiperparámetros v_0 y s_0^2 determinan la concentración y la escala de la distribución. La elección de esta distribución se debe a su conjugación con la verosimilitud normal del modelo, lo que permite una actualización posterior sencilla. Además, al ser una distribución que solo admite valores positivos, es una opción natural para modelar la varianza de un proceso estocástico (Gelman et al., 2013).

Por otro lado, dado σ_h, los parámetros μ y ϕ_h se modelan como una distribución normal multivariada:

$$(\mu, \phi_h) \mid \sigma_h \sim N(\bar{\beta}, \sigma_h^2 A^{-1}) \tag{1.14}$$

Esta distribución se selecciona porque mantiene la conjugación [2] con la verosimilitud y permite capturar posibles correlaciones entre los parámetros. El vector de medias $\bar{\beta} = (\bar{\mu}, \bar{\phi}_h)$ representa nuestras expectativas iniciales sobre los valores de μ y ϕ_h antes de observar los datos. Generalmente, $\bar{\mu}$ se puede fijar en torno a una creencia previa , mientras que $\bar{\phi}_h$ puede seleccionarse cerca de uno si asumimos una fuerte persistencia

[2] Cuando decimos que una distribución previa (prior) mantiene la conjugación con la verosimilitud, nos referimos a que la posterior resultante pertenece a la misma familia de distribuciones que la prior.

En términos generales, en inferencia bayesiana tenemos que:

$$\text{Posterior} \propto \text{Verosimilitud} \times \text{Prior}$$

Si la prior es seleccionada de tal manera que la posterior tiene la misma forma funcional que la prior, entonces decimos que la prior es conjugada con la verosimilitud.

en la volatilidad.

La matriz de precisión A juega un papel crucial en la estructura de la distribución previa.

Definición 3 (Matriz de precisión)**.** *Sea* $\mathbf{x} \in R^d$ *un vector aleatorio con distribución normal multivariante* $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, donde* $\boldsymbol{\mu} \in R^d$ *es el vector de medias y* $\boldsymbol{\Sigma} \in R^{d \times d}$ *es la matriz de covarianza definida positiva. La matriz de precisión* $\mathbf{A}$ *se define como el inverso de la matriz de covarianza:*

$$\mathbf{A} = \boldsymbol{\Sigma}^{-1}.$$

Desde el punto de vista probabilístico, la matriz de precisión $\mathbf{A}$ describe la estructura de dependencia entre las componentes de $\mathbf{x}$ y tiene las siguientes propiedades clave:

- **Elementos diagonales** a_{ii} indican la precisión (o la inversa de la varianza marginal) de cada componente x_i. Valores altos de a_{ii} corresponden a menor varianza, mientras que valores bajos indican mayor variabilidad en x_i.
- **Elementos fuera de la diagonal** a_{ij} (con $i \neq j$) representan la *correlación parcial* entre las componentes x_i y x_j, dado el resto de las variables. Si $a_{ij} = 0$, las variables x_i y x_j son condicionalmente independientes, dado el resto de las componentes.

En el contexto de distribuciones previas conjuntas para parámetros como μ y ϕ_h, la matriz de precisión regula tanto la varianza individual de cada parámetro como la relación entre ellos. Por ejemplo, si la matriz de precisión es:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix},$$

entonces el valor de a_{12} describe la dependencia lineal entre μ y ϕ_h. Un valor $a_{12} = 0$ implica que los dos parámetros son condicionalmente independientes, mientras que un valor diferente de cero indica correlación parcial entre ellos, condicionada al resto de las variables. Esta matriz debe ser simétrica y definida positiva. La correcta especificación de esta matriz es clave, ya que un valor demasiado alto en los elementos diagonales puede restringir excesivamente la varianza de los parámetros, limitando la capacidad del modelo para adaptarse a los datos.

Dado que hemos definido las distribuciones a priori para σ_h^2 y $(\mu, \phi_h) \mid \sigma_h$, es posible derivar una forma analítica para muestrear de la distribución condicional conjunta:

$$p(\omega \mid h, y) = p(\mu, \phi_h \mid h, y) \cdot p(\sigma_h \mid h, y) \tag{1.15}$$

pero necesitamos conocer la forma explícita de las distribuciones condicionales posteriores para cada uno de estos parámetros.

Muestreo de σ_h^2 de su distribución posterior Para calcular la distribución condicional posterior de $\sigma_h^2 \mid h$, aplicamos directamente el Teorema de Bayes, que nos dice que la posterior es proporcional al producto de la verosimilitud y la distribución a priori:

$$p(\sigma_h^2 \mid h) \propto p(h \mid \sigma_h^2) \cdot p(\sigma_h^2) \tag{1.16}$$

Primero, calculamos la verosimilitud de h dado σ_h^2. Recordemos que el modelo Autorregresivo para h_t está definido por:

$$h_t = \mu + \phi_h(h_{t-1} - \mu) + \sigma_h \eta_t, \quad \eta_t \sim N(0, 1).$$

Nuestro objetivo es demostrar que el vector de volatilidades latentes $\mathbf{h} = (h_2, \ldots, h_T)$ sigue una distribución normal multivariada de la forma:

$$\mathbf{h} \mid \mu, \phi_h, \sigma_h^2 \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma_h^2 \mathbf{I}_{T-1}\right) \tag{1.17}$$

donde:

$$\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{Hh} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \phi_h \end{bmatrix}.$$

Para lograr esto, reescribimos el modelo Autorregresivo en forma matricial. Definimos el vector de volatilidades latentes, el vector de errores y la matriz de desplazamiento

de la siguiente manera:

$$\mathbf{h} = \begin{bmatrix} h_2 \\ h_3 \\ \vdots \\ h_T \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_2 \\ \eta_3 \\ \vdots \\ \eta_T \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

El vector de errores $\boldsymbol{\eta}$ sigue una distribución normal multivariada $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{T-1})$. Reescribimos el modelo Autorregresivo para $t = 2, \ldots, T$ de forma matricial:

$$\mathbf{h} = \mu\mathbf{1} + \phi_h \mathbf{H}\mathbf{h} + \sigma_h \boldsymbol{\eta} \tag{1.18}$$

donde $\mathbf{1}$ es un vector columna de unos de longitud $T - 1$.

Reorganizando los términos, podemos expresar el modelo en la forma:

$$\mathbf{h} = \mathbf{X}\boldsymbol{\beta} + \sigma_h \boldsymbol{\eta} \tag{1.19}$$

Como $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{T-1})$, el término $\sigma_h \boldsymbol{\eta}$ sigue una distribución normal multivariada $\mathcal{N}(\mathbf{0}, \sigma_h^2 \mathbf{I}_{T-1})$. Por la propiedad de linealidad de la distribución normal, el vector $\mathbf{h} = \mathbf{X}\boldsymbol{\beta} + \sigma_h \boldsymbol{\eta}$ sigue una distribución normal multivariada:

$$\mathbf{h} \mid \mu, \phi_h, \sigma_h^2 \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma_h^2 \mathbf{I}_{T-1}\right) \tag{1.20}$$

La función de densidad de una distribución normal multivariada es:

$$p(\mathbf{h} \mid \sigma_h^2) = \frac{1}{(2\pi)^{\frac{T-1}{2}} (\sigma_h^2)^{\frac{T-1}{2}}} \exp\left(-\frac{1}{2\sigma_h^2}(\mathbf{h} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{h} - \mathbf{X}\boldsymbol{\beta})\right) \tag{1.21}$$

El término cuadrático $(\mathbf{h} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{h} - \mathbf{X}\boldsymbol{\beta})$ corresponde a la suma de residuos al cuadrado (RSS), dada por:

$$\text{RSS} = \sum_{t=2}^{T} \left(h_t - \mu - \phi_h (h_{t-1} - \mu)\right)^2 \tag{1.22}$$

Sustituyendo este término en la función de densidad, la expresión de la verosimilitud condicional se simplifica a:

$$p\left(h \mid \sigma_h^2\right) \propto (\sigma_h^2)^{-\frac{T-1}{2}} \exp\left(-\frac{\text{RSS}}{2\sigma_h^2}\right) \tag{1.23}$$

En esta expresión, el factor $(\sigma_h^2)^{-\frac{T-1}{2}}$ refleja el efecto del parámetro σ_h^2 en la escala de la distribución, mientras que el término exponencial penaliza las configuraciones de $\mathbf{h}$ que se desvían significativamente del valor esperado $\mathbf{X}\boldsymbol{\beta}$. La suma de residuos al cuadrado RSS mide precisamente esa discrepancia, lo que justifica su aparición en el término exponencial.

Por otro lado, la prior de σ_h^2 es una distribución gamma inversa con parámetros v_0 y s_0^2:

$$p(\sigma_h^2) \propto (\sigma_h^2)^{-(v_0+1)} \exp\left(-\frac{s_0^2}{\sigma_h^2}\right) \tag{1.24}$$

Sustituyendo las expresiones de la verosimilitud y la prior, tenemos:

$$p(\sigma_h^2 \mid h) \propto (\sigma_h^2)^{-\frac{T-1}{2}} \exp\left(-\frac{\text{RSS}}{2\sigma_h^2}\right) (\sigma_h^2)^{-(v_0+1)} \exp\left(-\frac{s_0^2}{\sigma_h^2}\right) \tag{1.25}$$

Reorganizando los términos:

$$p(\sigma_h^2 \mid h) \propto (\sigma_h^2)^{-\left(v_0+\frac{T-1}{2}+1\right)} \exp\left(-\frac{s_0^2 + \frac{\text{RSS}}{2}}{\sigma_h^2}\right) \tag{1.26}$$

Esta expresión corresponde a la densidad de una distribución gamma inversa:

$$\sigma_h^2 \mid h \sim IG\left(v_0 + \frac{T-1}{2},\ s_0^2 + \frac{\text{RSS}}{2}\right) \tag{1.27}$$

Esto permite muestrear σ_h^2 directamente en cada iteración del muestreo de Gibbs, disponiendo de los valores previamente muestreados del vector de variables latentes h. Es interesante notar que, en esta formulación, la distribución condicional de σ_h^2 no depende explícitamente de los datos observados $(Y_1, \ldots, Y_T)$, sino únicamente a través de los valores actuales de h.

Muestreo de $(\mu, \phi_h) \mid \sigma_h, h, y$ En el contexto de la regresión bayesiana, consideramos el modelo de volatilidad estocástica reparametrizado en términos de los coeficientes μ

y ϕ_h, de modo que la ecuación para la volatilidad latente como la ecuación 1.19

Asumimos que los errores son gaussianos independientes, lo que implica que la verosimilitud para $\mathbf{h}$ dado $\boldsymbol{\beta}$ sigue una distribución normal

$$p(\mathbf{h} \mid \boldsymbol{\beta}, \mathbf{X}, \sigma_h^2) = \frac{1}{(2\pi\sigma_h^2)^{(T-1)/2}} \exp\left(-\frac{1}{2\sigma_h^2}(\mathbf{h} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{h} - \mathbf{X}\boldsymbol{\beta})\right) \tag{1.28}$$

Adicionalmente, se asume una distribución a priori normal para $\boldsymbol{\beta}$,

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{b}_0, \sigma_h^2\boldsymbol{\Sigma}_0) \tag{1.29}$$

lo que implica que su densidad está dada por

$$p(\boldsymbol{\beta}) = \frac{1}{(2\pi)^{d/2}|\sigma_h^2\boldsymbol{\Sigma}_0|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{b}_0)^\top(\sigma_h^2\boldsymbol{\Sigma}_0)^{-1}(\boldsymbol{\beta} - \boldsymbol{b}_0)\right) \tag{1.30}$$

donde $\boldsymbol{b}_0$ representa la media previa de $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_0$ es la matriz de covarianza del prior, y $d = 2$, ya que $\boldsymbol{\beta}$ es un vector bidimensional. Aplicando la regla de Bayes, la distribución posterior de $\boldsymbol{\beta}$ se obtiene como

$$p(\boldsymbol{\beta} \mid \mathbf{h}) \propto p(\mathbf{h} \mid \boldsymbol{\beta})\, p(\boldsymbol{\beta}) \tag{1.31}$$

Para derivar la expresión explícita de esta posterior, expandemos el término cuadrático en la exponencial de la verosimilitud:

$$(\mathbf{h} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{h} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{h}^\top\mathbf{h} - 2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{h} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \tag{1.32}$$

De manera similar, el término cuadrático en la exponencial del prior se puede expandir como

$$(\boldsymbol{\beta} - \boldsymbol{b}_0)^\top\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{b}_0) = \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta} - 2\boldsymbol{b}_0^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta} + \boldsymbol{b}_0^\top\boldsymbol{\Sigma}_0^{-1}\boldsymbol{b}_0 \tag{1.33}$$

Multiplicando ambas distribuciones y combinando términos, se obtiene la siguiente expresión exponencial:

$$\exp\left(-\frac{1}{2\sigma_h^2}\left[\boldsymbol{\beta}^\top(\mathbf{X}^\top\mathbf{X} + \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\beta} - 2\boldsymbol{\beta}^\top(\mathbf{X}^\top\mathbf{h} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{b}_0)\right]\right) \tag{1.34}$$

Esta expresión corresponde al núcleo de una distribución normal para $\boldsymbol{\beta}$, con parámetros actualizados. Identificamos entonces la **varianza-covarianza posterior** como

$$\boldsymbol{\Sigma}_{\text{post}} = (\mathbf{X}^\top \mathbf{X} + \boldsymbol{\Sigma}_0^{-1})^{-1} \tag{1.35}$$

y la **media posterior** como

$$\boldsymbol{m}_\beta = \boldsymbol{\Sigma}_{\text{post}}(\mathbf{X}^\top \mathbf{h} + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{b}_0) \tag{1.36}$$

Por lo tanto, la distribución posterior de $\boldsymbol{\beta}$ queda completamente caracterizada como

$$\boldsymbol{\beta} \mid \mathbf{h}, \mathbf{X}, \sigma_h^2 \sim \mathcal{N}(\boldsymbol{m}_\beta, \sigma_h^2 \boldsymbol{\Sigma}_{\text{post}}) \tag{1.37}$$

Dado que $\boldsymbol{\Sigma}_{\text{post}} = A^{-1}$, la posterior se puede reescribir como:

$$(\mu, \phi_h) \sim \mathcal{N}(\bar{\beta}, \sigma_h^2 A^{-1}),$$

Por lo que el algoritmo para muestrear $\omega = (\mu, \phi_h, \sigma_h)$ nos queda de la siguiente manera:

Algorithm 2 Muestreo de $\omega = (\mu, \phi_h, \sigma_h)$

Entrada: Valores actuales de las variables latentes h, parámetros iniciales (μ, ϕ_h, σ_h), hiperparámetros v_0, s_0^2, número de iteraciones N.

Salida: Secuencia de muestras $\{(\mu^{(i)}, \phi_h^{(i)}, \sigma_h^{(i)})\}$ para $i = 1, \ldots, N$.

1: **for** $i = 1$ hasta N **do**

2: **Actualizar** σ_h^2:

3: Calcular la suma de residuos al cuadrado (RSS).

4: Obtener σ_h^2 de la distribución gamma inversa:

$$\sigma_h^2 \sim IG\left(v_0 + \frac{T-1}{2},\ s_0^2 + \frac{\text{RSS}}{2}\right).$$

5: **Actualizar** $(\mu, \phi_h) \mid \sigma_h$:

6: Calcular la matriz de covarianza A^{-1}.

7: Obtener (μ, ϕ_h) de la distribución normal multivariada:

$$(\mu, \phi_h) \sim \mathcal{N}(\bar{\beta}, \sigma_h^2 A^{-1}).$$

8: **end for**

1.4.2. Muestrear h desde $p(h \mid \omega, y)$

Para estimar la volatilidad latente $h = \{h_1, \ldots, h_T\}$, necesitamos trabajar con la distribución conjunta de h, condicionada en los parámetros del modelo ω y los datos observados $y = \{y_1, \ldots, y_T\}$:

$$p(h \mid \omega, y) = p(h_1, \ldots, h_T \mid \omega, y) \tag{1.38}$$

Esta distribución conjunta refleja la dependencia entre las volatilidades en distintos instantes de tiempo, así como la influencia de los datos observados. Dadas las complejas interacciones temporales del proceso Autorregresivo, la distribución conjunta es *intractable*, es decir, no puede ser muestreada de forma directa. Pero como ya vimos, podemos descomponer esta distribución conjunta en términos de las distribuciones condicionales de cada variable de manera secuencial, mediante el teorema de Hammersley-Clifford:

$$p(h \mid \omega, y) = \prod_{t=1}^{T} p\big(h_t \mid h_{-t}, \omega, y\big) \tag{1.39}$$

donde $h_{-t} = \{h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_T\}$ representa todos los valores de h excepto h_t. Si podemos muestrear de cada $p(h_t \mid h_{-t}, \omega, y)$, podemos reconstruir la distribución conjunta a través de un proceso iterativo que actualiza cada h_t sucesivamente, generando una *cadena de Markov* que converge a $p(h \mid \omega, y)$. Gracias a la estructura del proceso AR(1), esta dependencia condicional se reduce a

$$p(h_t \mid h_{-t}, \omega, y) = p(h_t \mid h_{t-1}, h_{t+1}, \omega, y) \tag{1.40}$$

En otras palabras, dado el modelo AR(1), h_t es condicionalmente independiente del resto de las volatilidades latentes h_{-t} si conocemos los valores de h_{t-1} y h_{t+1}.

Por lo tanto, la actualización de h_t en el proceso de muestreo se basa únicamente en los valores inmediatamente adyacentes h_{t-1} y h_{t+1}. Esta característica permite que para cada actualización solo necesitamos calcular la probabilidad condicional utilizando las dos observaciones vecinas:

$$p(h_t \mid h_{t-1}, h_{t+1}, \omega, y) \propto p\left(y_t \mid h_t\right) \times p\left(h_t \mid h_{t-1}, h_{t+1}, \omega\right) \tag{1.41}$$

Aunque el modelo dicta que h_t "herede" información solo de h_{t-1}, el muestreo de Gibbs requiere consistencia con la distribución conjunta completa $p(h \mid \omega, y)$. Por ello, cuando se actualiza h_t, se tienen en cuenta los valores actuales tanto de h_{t-1} (su antecesor directo) como de h_{t+1} (el siguiente instante), incluso si en la dinámica básica $AR(1)$ no existe una dependencia causal "hacia delante". El parámetro h_{t+1} no afecta a h_t en sentido físico o temporal; más bien, en la cadena de Gibbs, se exige que todas las variables latentes sean muestreadas de forma compatible con la posterior conjunta, lo que introduce este término al condicionar completamente en h_{-t}.

Componentes de la distribución condicional

El primer término, $p(y_t \mid h_t)$, proviene de la verosimilitud del modelo para los datos y_t. Suponemos que, dada la volatilidad $\exp(h_t)$, las observaciones y_t siguen una

distribución normal con media cero y varianza $\exp(h_t)$:

$$y_t \mid h_t \sim \mathcal{N}\big(0,\ e^{h_t}\big). \tag{1.42}$$

Ignorando constantes de normalización, se obtiene

$$p\big(y_t \mid h_t\big) \propto \big(e^{h_t}\big)^{-\frac{1}{2}}\ \exp\!\Big(-\tfrac{y_t^2}{2\,e^{h_t}}\Big) = e^{-\frac{h_t}{2}}\ \exp\!\Big(-\tfrac{y_t^2}{2\,e^{h_t}}\Big). \tag{1.43}$$

El segundo factor, $p(h_t \mid h_{t-1}, h_{t+1}, \omega)$, encapsula la **dependencia temporal** de la volatilidad latente. Suponiendo el siguiente modelo base:

$$h_t = \mu+\phi_h\big(h_{t-1}-\mu\big)+\sigma_h\,\eta_t,\ \ \eta_t \sim \mathcal{N}(0,1), \qquad h_{t+1} = \mu+\phi_h\big(h_t-\mu\big)+\sigma_h\,\eta_{t+1},\ \ \eta_{t+1} \sim \mathcal{N}(0,1),$$

buscamos

$$p\big(h_t \mid h_{t-1},\ h_{t+1},\ \omega\big).$$

Definamos $x = h_t - \mu,\ a = h_{t-1} - \mu,\ v = h_{t+1} - \mu$, de modo que

$$\begin{cases} x = \phi_h\, a\ +\ \sigma_h\,\eta_t, \\ v = \phi_h\, x\ +\ \sigma_h\,\eta_{t+1}. \end{cases} \tag{1.44}$$

El vector (x, v) condicionado en a es una transformación lineal de (η_t, η_{t+1}), cuyos componentes son i.i.d. $\mathcal{N}(0,1)$. En concreto,

$$\begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} \phi_h\, a \\ \phi_h^2\, a \end{pmatrix} + \sigma_h \begin{pmatrix} 1 & 0 \\ \phi_h & 1 \end{pmatrix} \begin{pmatrix} \eta_t \\ \eta_{t+1} \end{pmatrix}. \tag{1.45}$$

Recordemos que, si $\begin{pmatrix} x \\ v \end{pmatrix} \sim \mathcal{N}_2\big((M_x,\ M_v)^\top, \Sigma\big)$ con $\Sigma = \begin{pmatrix} \sigma_{xx}^2 & \sigma_{xv} \\ \sigma_{xv} & \sigma_{vv}^2 \end{pmatrix}$, entonces

$$x \mid v \sim \mathcal{N}\Big(M_x + \tfrac{\sigma_{xv}}{\sigma_{vv}^2}(v - M_v),\ \sigma_{xx}^2 - \tfrac{\sigma_{xv}^2}{\sigma_{vv}^2}\Big). \tag{1.46}$$

En nuestro caso,

$$M_x = \phi_h\, a, \quad M_v = \phi_h^2\, a, \quad \sigma_{xx}^2 = \sigma_h^2, \quad \sigma_{vv}^2 = \sigma_h^2(\phi_h^2 + 1), \quad \sigma_{xv} = \sigma_h^2\,\phi_h.$$

Así, $\frac{\sigma_{xv}}{\sigma_{vv}^2} = \frac{\phi_h}{\phi_h^2+1}$ y $\sigma_{xx}^2 - \frac{\sigma_{xv}^2}{\sigma_{vv}^2} = \frac{\sigma_h^2}{\phi_h^2+1}$. La media condicional es

$$\mu_{x|a,v} = \phi_h\, a + \frac{\phi_h}{\phi_h^2+1}\big(v - \phi_h^2\, a\big) = \frac{\phi_h}{\phi_h^2+1}\,(a+v), \tag{1.47}$$

y la varianza condicional $\sigma_{x|a,v}^2 = \frac{\sigma_h^2}{\phi_h^2+1}$. Por lo tanto,

$$x \mid (a,v) \;\sim\; \mathcal{N}\Big(\mu_{x|a,v},\, \sigma_{x|a,v}^2\Big) \quad\Longrightarrow\quad h_t \mid (h_{t-1}, h_{t+1}, \omega) \;\sim\; \mathcal{N}\Big(\tilde{\mu}_t,\, \tilde{\sigma}^2\Big),$$

donde, reexpresando en h,

$$\tilde{\mu}_t = \mu + \frac{\phi_h}{\phi_h^2+1}\,\big[(h_{t-1}-\mu) + (h_{t+1}-\mu)\big], \qquad \tilde{\sigma}^2 = \frac{\sigma_h^2}{\phi_h^2+1}.$$

La condicional completa resulta del producto entre la verosimilitud y el término que recoge la dependencia temporal (AR(1) con vecinos):

$$p\big(h_t \mid h_{t-1}, h_{t+1}, \omega, y_t\big) \;\propto\; \underbrace{e^{-\frac{h_t}{2}}\,\exp\!\Big(-\tfrac{y_t^2}{2\,e^{h_t}}\Big)}_{\text{verosimilitud}} \;\times\; \underbrace{\exp\!\Big(-\tfrac{\big(h_t-\tilde{\mu}_t\big)^2}{2\,\tilde{\sigma}^2}\Big)}_{\text{dependencia temporal (AR(1) con vecinos)}}.$$

La presencia conjunta de $e^{-h_t/2}$ y $\exp\!\big(-y_t^2/(2e^{h_t})\big)$ con el núcleo normal en h_t (media $\tilde{\mu}_t$, varianza $\tilde{\sigma}^2$) genera una forma posterior que no pertenece a familias estándar con conjugación directa, por lo que no es posible muestrear h_t de manera cerrada. En consecuencia, empleamos un paso de Metropolis–Hastings dentro de cada iteración de Gibbs. Usamos, por ejemplo, una propuesta simétrica

$$h_t^{(\text{prop})} \;\sim\; \mathcal{N}\Big(h_t^{(\text{curr})},\, \tau^2\Big),$$

donde τ^2 controla el tamaño de los saltos. La probabilidad de aceptación es

$$\alpha\big(h_t^{(\text{curr})}, h_t^{(\text{prop})}\big) = \text{mín}\left\{1,\; \frac{\exp\!\Big(-\frac{h_t^{(\text{prop})}}{2}\Big)\,\exp\!\Big(-\frac{y_t^2}{2\,e^{h_t^{(\text{prop})}}}\Big)\,\exp\!\Big(-\frac{\big(h_t^{(\text{prop})}-\tilde{\mu}_t\big)^2}{2\,\tilde{\sigma}^2}\Big)}{\exp\!\Big(-\frac{h_t^{(\text{curr})}}{2}\Big)\,\exp\!\Big(-\frac{y_t^2}{2\,e^{h_t^{(\text{curr})}}}\Big)\,\exp\!\Big(-\frac{\big(h_t^{(\text{curr})}-\tilde{\mu}_t\big)^2}{2\,\tilde{\sigma}^2}\Big)}\right\},$$

donde, al ser la propuesta normal centrada en el estado actual, los términos q se can-

celan.

El algoritmo de Metropolis-Hastings aplicado en este caso sigue los siguientes pasos:

Algorithm 3 Muestreo de h_t mediante MCMC

Entrada: Estado inicial $h_t^{(0)}$, varianza de propuesta τ^2, datos observados y_t, parámetros ω, número de iteraciones N.

Salida: Secuencia de muestras $\{h_t^{(i)}\}_{i=1}^{N}$ aproximando $p(h_t \mid h_{t-1}, h_{t+1}, \omega, y_t)$.

1: **for** $i = 1$ hasta N **do**

2: **Propuesta de un nuevo estado**:

3: Generar un candidato $h_t^{(\text{prop})} \sim \mathcal{N}(h_t^{(i-1)}, \tau^2)$.

4: **Cálculo de la razón de aceptación**:

5: Evaluar la probabilidad condicional en el nuevo estado:

$$p_{\text{prop}} = p(h_t^{(\text{prop})} \mid h_{t-1}, h_{t+1}, \omega, y_t).$$

6: Evaluar la probabilidad condicional en el estado actual:

$$p_{\text{curr}} = p(h_t^{(i-1)} \mid h_{t-1}, h_{t+1}, \omega, y_t).$$

7: Calcular la razón de aceptación:

$$r = \frac{p_{\text{prop}}}{p_{\text{curr}}}.$$

8: **Aceptación o rechazo**:

9: Generar $u \sim \mathcal{U}(0, 1)$.

10: **if** $u \leq \text{mín}(1, r)$ **then**

11: Aceptar la propuesta: $h_t^{(i)} \leftarrow h_t^{(\text{prop})}$.

12: **else**

13: Rechazar la propuesta: $h_t^{(i)} \leftarrow h_t^{(i-1)}$.

14: **end if**

15: **end for**

La deducción anterior corresponde al caso interior $2 \leq t \leq T-1$, donde h_t depende simultáneamente de h_{t-1} y h_{t+1}. En los extremos de la secuencia, la estructura se simplifica:

- Para h_1, la condicional completa involucra el término de verosimilitud $p(y_1 \mid h_1)$ y el prior inicial de h_1 (por ejemplo, la distribución estacionaria del proceso si se asume).

- Para h_T, la condicional depende únicamente de h_{T-1} y del dato observado y_T.

De esta forma, las expresiones anteriores deben adaptarse en los extremos de la cadena, aunque el esquema de muestreo mediante Metropolis–Hastings se mantiene sin cambios.

1.4.3. Estimación de la Distribución Conjunta de h y ω.

El objetivo es generar muestras de la distribución conjunta $p(h, \omega \mid y)$, donde h representa las volatilidades latentes y $\omega = (\mu, \phi_h, \sigma_h)$ son los parámetros globales. Para ello, se identifican las distribuciones condicionales:

$$p\big(\omega \mid h, y\big) \quad \text{y} \quad p\big(h \mid \omega, y\big),$$

y se implementa un *esquema iterativo* que alterna entre ambas. En cada iteración $(i+1)$, se realizan los siguientes pasos:

Algorithm 4 Muestreo de parámetros y volatilidades latentes

Entrada: Valores iniciales $\mu^{(0)}, \phi_h^{(0)}, \sigma_h^{(0)}$. Trayectoria inicial $h^{(0)}$. Número de iteraciones N. Desviación estándar de la propuesta σ_{prop}.

Inicialización:
    Asignar valores iniciales a (μ, ϕ_h, σ_h) y $h_t^{(0)}$.
    Definir contenedores para almacenar muestras.
for $i = 1$ **hasta** N **do**
    Actualización de parámetros $\omega = (\mu, \phi_h, \sigma_h)$:
    Muestrear $(\mu^{(i+1)}, \phi_h^{(i+1)}) \sim p\big(\mu, \phi_h \mid \sigma_h^{(i)}, h^{(i)}, y\big)$.
    Muestrear $\sigma_h^{(i+1)} \sim p\big(\sigma_h \mid \mu^{(i+1)}, \phi_h^{(i+1)}, h^{(i)}, y\big)$.
    Actualización de las volatilidades latentes h_t**:**
    for $t = 1$ **hasta** T **do**
        Proponer $h_t^* \sim \mathcal{N}(h_t^{(i)}, \sigma_{\text{prop}}^2)$.
        Calcular $r = \dfrac{p(h_t^* \mid h_{t-1}^{(i+1)}, h_{t+1}^{(i)}, \mu^{(i+1)}, \phi_h^{(i+1)}, \sigma_h^{(i+1)}, y_t)}{p(h_t^{(i)} \mid h_{t-1}^{(i+1)}, h_{t+1}^{(i)}, \mu^{(i+1)}, \phi_h^{(i+1)}, \sigma_h^{(i+1)}, y_t)}$.
        Generar $u \sim \mathcal{U}(0,1)$.
        if $u \leq \text{mín}(1, r)$ **then**
            $h_t^{(i+1)} \leftarrow h_t^*$.
        else
            $h_t^{(i+1)} \leftarrow h_t^{(i)}$.
        end if
    end for
end for

Salida: Secuencia de muestras $\{(\mu^{(i)}, \phi_h^{(i)}, \sigma_h^{(i)}, h^{(i)})\}_{i=1}^N$.

El resultado de este barrido sobre $t = 1, \ldots, T$ da la nueva trayectoria $h^{(i+1)}$.

Tras completar la actualización de ω y h, se obtiene la muestra $\big(\omega^{(i+1)}, h^{(i+1)}\big)$. Este proceso se repite el número de iteraciones deseado para construir una cadena de Markov cuyas realizaciones convergen a la distribución posterior:

$$p\big(\mu, \phi_h, \sigma_h, h \mid y\big) \tag{1.48}$$

Al descartar un periodo de *burn-in* y extraer submuestras (si es necesario reducir

autocorrelación), se obtienen muestras finales que permiten estimar tanto la volatilidad latente h como los parámetros globales ω.

1.4.4. Cómo verificamos la convergencia

Para garantizar que las muestras generadas mediante el muestreo de Gibbs o cualquier otro método MCMC sean representativas de la distribución posterior, es fundamental verificar la convergencia de la cadena de Markov. Existen varios criterios que utilizamos en la práctica para evaluar esta convergencia:

- **Inspección visual de trazas**: Se analizan las trazas de las cadenas para asegurarnos de que exploran adecuadamente el espacio paramétrico. Una cadena convergente debería mostrar un comportamiento estacionario sin tendencias ni patrones oscilatorios persistentes, lo que indicaría una exploración deficiente del espacio de estados.

- **Análisis de autocorrelación**: Evaluamos la correlación entre las muestras consecutivas. Una alta autocorrelación implica que las muestras están fuertemente correlacionadas y que la cadena explora lentamente el espacio de parámetros. En estos casos, puede ser necesario aplicar un *thinning* o aumentar el número de iteraciones para mejorar la independencia de las muestras.

Una vez que la cadena ha alcanzado su estado estacionario y se han descartado las iteraciones iniciales (*burn-in*), podemos utilizar las muestras generadas para calcular estadísticas resumen, como la media y la mediana posterior, la varianza y los intervalos de credibilidad. Estas métricas permiten realizar inferencias sobre los parámetros de interés con base en la distribución posterior obtenida.

Capítulo 2

Simulación y Estimación

Para estudiar el comportamiento del algoritmo para la estimación de los parámetros del modelo de volatilidad estocástica, realizamos una simulación de datos basada en el modelo propuesto, definido por las siguientes ecuaciones:

$$y_t = \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \exp(h_t)), \tag{2.1}$$

$$h_t = \mu + \phi_h(h_{t-1} - \mu) + \sigma_h \eta_t, \quad \eta_t \sim \mathcal{N}(0, 1). \tag{2.2}$$

Para la simulación, generamos una serie temporal de tamaño $T = 500$ bajo los parámetros reales $\mu = 2.0$, $\phi_h = 0.92$ y $\sigma_h = 0.8$. Posteriormente, aplicamos un algoritmo de muestreo de Gibbs, combinado con pasos de Metropolis-Hastings, para estimar estos parámetros a partir de las observaciones simuladas. Para su implementación, se utilizaron los scripts `TESIS_simulacion.R` y `TESIS_funciones.R`, disponibles en el repositorio de GitHub.

Estudios previos han demostrado que la volatilidad en series de tiempo financieras es altamente persistente, lo que sugiere que el parámetro ϕ_h debería estar cerca de 1 (Engle, 1982; Montenegro, 2010; Matos, 2008; Bariviera, 2011). Esta información previa nos permite definir una expectativa inicial sobre ϕ_h, mientras que para μ y σ_h adoptamos prioris menos informativas debido a la falta de conocimiento específico sobre su valor. En este contexto, el vector $\beta = (\mu(1-\phi_h), \phi_h)$ se modela con una distribución normal multivariada $\mathcal{N}(\mathbf{b_0}, \sigma_h^2 \mathbf{\Sigma_0})$, donde $\mathbf{b_0} = (0, 0.7)$ refleja nuestra expectativa de que ϕ_h esté cerca de 1, y $\mathbf{\Sigma_0} = \text{diag}(100, 1)$ asigna una varianza amplia a μ y una más restringida a ϕ_h para capturar esta persistencia.

Para σ_h^2, utilizamos una distribución Gamma-Inversa con parámetros de forma y escala $(1, 0.02)$, asegurando una priori que permita explorar adecuadamente el espacio

paramétrico sin restringir excesivamente la varianza del proceso de volatilidad. Realizamos pruebas preliminares de análisis de sensibilidad a la priori para ajustar estos parámetros y verificar que el algoritmo sea capaz de manejar correctamente la información de persistencia en la volatilidad y, al mismo tiempo, logre estimar bien σ_h y μ incluso en ausencia de información previa específica.

El algoritmo se ejecuta durante 50,000 iteraciones, descartando las primeras 8,000 como *burn-in* para favorecer la aproximación al régimen estacionario. Adicionalmente, se realizaron pruebas bajo diversas configuraciones de parámetros, variando los valores verdaderos de (μ, ϕ_h, σ_h) utilizados en la simulación y las condiciones iniciales y diferentes semillas aleatorias. La tasa de aceptación fue monitoreada y ajustada para garantizar una adecuada exploración del espacio paramétrico. En todas las configuraciones consideradas, las trazas de las cadenas MCMC indicaron convergencia clara y las estimaciones posteriores de μ, ϕ_h y σ_h se mantuvieron próximas a los valores de simulación, lo que sugiere que el algoritmo propuesto es efectivo para este tipo de modelos.

2.1. Resultados de la Simulación

A continuación, se presenta la serie temporal generada para los datos y_t simulados a partir del modelo de volatilidad estocástica. La figura muestra el comportamiento característico de una serie con heterocedasticidad, donde la varianza varía a lo largo del tiempo:

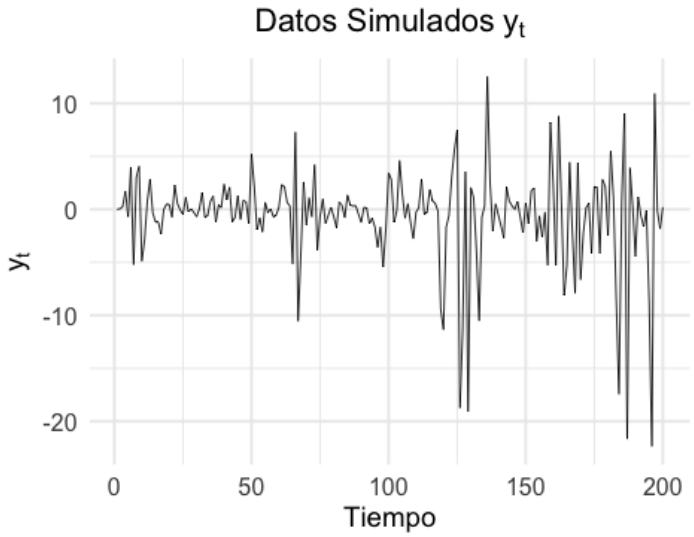

Figura 2.1: Serie temporal de y_t generada a partir del modelo SV.

Se puede observar a simple vista cómo en la simulación parece haber variabilidad en la varianza y evidencia de su persistencia. Esto es una característica típica de los retornos financieros, donde la volatilidad tiende a agruparse en períodos de alta o baja intensidad, reflejando la presencia de heterocedasticidad condicional (Mandelbrot, 1963; Cont, 2001; Bollerslev, 1986; Engle, 1982).

Una vez corrido el algoritmo de MCMC, procedemos a evaluar la convergencia de las cadenas. En un primer análisis se observa que la tasa de aceptación para la actualización de los estados h_t es de 0.518, lo cual es aceptable. Sin embargo, al examinar la función de autocorrelación (ACF) de los parámetros μ, ϕ_h y σ_h, se evidencia que las muestras sucesivas presentan una alta correlación, lo que implica que la información independiente en la cadena es menor de la esperada.

La elevada autocorrelación es un problema común en la inferencia MCMC, ya que puede reducir drásticamente el *tamaño efectivo de la muestra* (ESS, por sus siglas en inglés). El ESS se define como la cantidad de observaciones independientes que contienen la misma información que la cadena correlacionada, y se puede calcular a partir de la longitud de la cadena y de su función de autocorrelación. Para mitigar este efecto, se aplica la técnica de *thinning*, que consiste en retener, por ejemplo, cada n-ésima muestra de la cadena y descartar las intermedias. Esta estrategia tiene como objetivo reducir la dependencia entre las muestras, mejorando así la calidad de la estimación.

Después de aplicar thinning, se procede a calcular el tamaño efectivo de la muestra (ESS) mediante la función `effectiveSize()` del paquete `coda`. En nuestro caso, se obtuvieron los siguientes resultados:

- ESS para $\mu = 280$,
- ESS para $\phi_h = 230.26$,
- ESS para $\sigma_h = 230.97$.

Estos valores indican que, pese a la alta autocorrelación original, la cadena resultante tras el thinning cuenta con un número suficiente de muestras independientes para realizar inferencias robustas.

Con el fin de evaluar la convergencia y la estabilidad de las estimaciones, se corrieron varias cadenas MCMC independientes para cada parámetro del modelo. Se observa que las cadenas se mezclan adecuadamente y que todas arrojan valores consistentes entre sí, lo cual respalda la confiabilidad del muestreo. De todas maneras, para el análisis que sigue se reportan los resultados correspondientes a una de las cadenas seleccionadas a modo representativo.

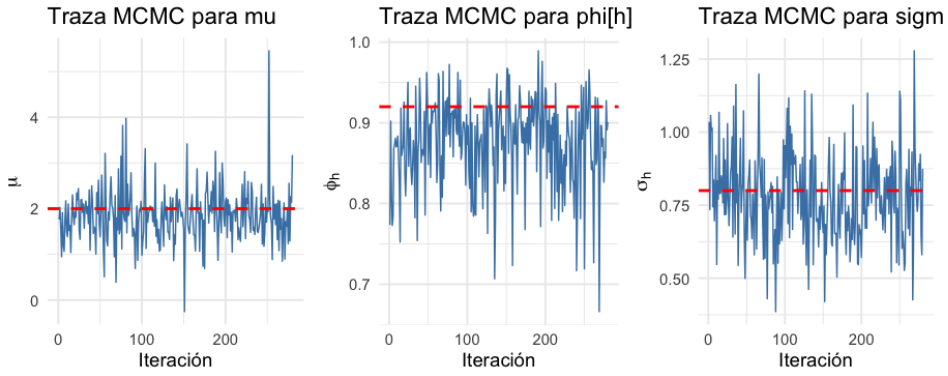

Figura 2.2: Trazas de los parámetros después del burn-in. En rojo se observa el valor verdadero del parámetro.

A continuación, se calculan las estimaciones puntuales de los parámetros a partir de las muestras posteriores (después del burn-in). En nuestro caso, se obtuvieron las siguientes estimaciones:

- $\hat{\mu} = 1.859 \quad \text{IC}_{95\,\%} = [0.861,\ 3.177] \quad$ (valor verdadero: $\mu_{\text{true}} = 2.0$),

- $\hat{\phi}_h = 0.876 \quad \text{IC}_{95\,\%} = [0.754,\ 0.963] \quad$ (valor verdadero: $\phi_{h,\text{true}} = 0.92$),

- $\hat{\sigma}_h = 0.779 \quad \text{IC}_{95\,\%} = [0.521,\ 1.118] \quad$ (valor verdadero: $\sigma_{h,\text{true}} = 0.8$).

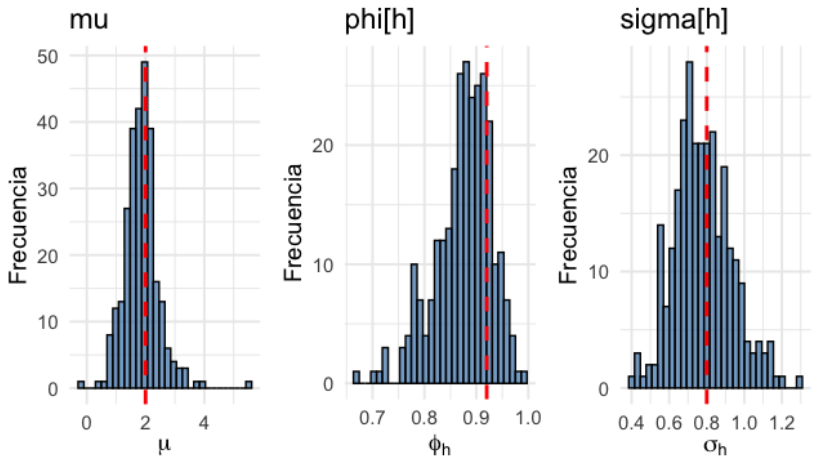

Figura 2.3: Histograma de las estimaciones posteriores de los parámetros. Las líneas punteadas rojas indican los valores reales, permitiendo evaluar la distribución posterior y la concordancia con los parámetros verdaderos.

A pesar de pequeñas discrepancias, las estimaciones muestran una buena concordancia con los parámetros verdaderos. En particular, la estimación de μ (1.859) se encuentra ligeramente por debajo del valor real (2.0), mientras que $\hat{\phi}_h$ (0.876) resulta algo menor que su valor verdadero (0.92) y $\hat{\sigma}_h$ (0.779) se aproxima de forma notable al valor fijado en la simulación (0.8). En todos los casos, los valores verdaderos se ubican dentro de los intervalos de credibilidad del 95 %, lo que indica que las diferencias observadas son consistentes con la variabilidad inherente al muestreo MCMC y la complejidad del proceso latente de volatilidad.

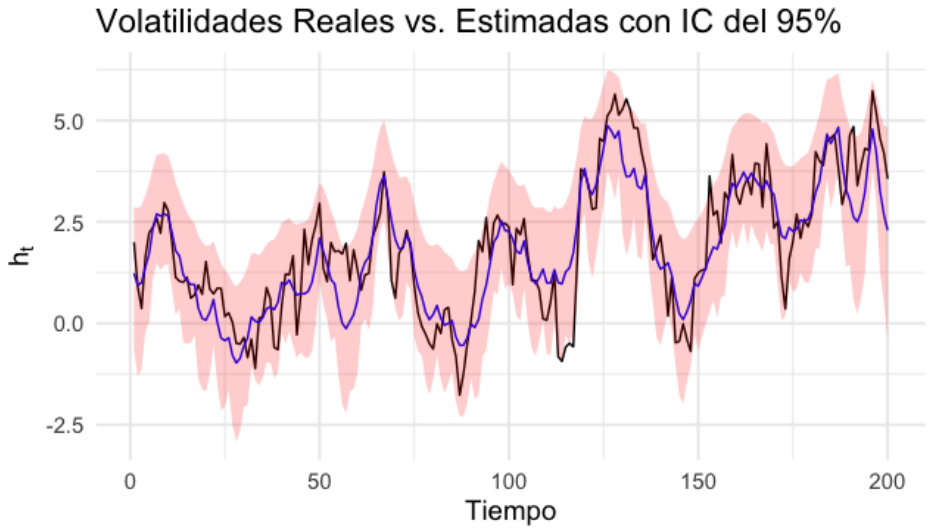

Figura 2.4: Comparación entre los valores reales de h_t y los intervalos de credibilidad del 95 % de los estados estimados. La linea negra corresponde a los valores reales y la azul a la predicción.

El hecho de que los h_t reales caigan sistemáticamente dentro del intervalo de credibilidad del 95 % respalda la validez del enfoque utilizado, pues sugiere que la distribución posterior obtenida es coherente con la información contenida en los datos simulados. En otras palabras, el modelo no solo proporciona estimaciones puntuales razonables, sino que también cuantifica de forma adecuada la variabilidad inherente al proceso de volatilidad.

Para finalizar, se realizaron múltiples simulaciones variando los parámetros reales del modelo y los datos con el objetivo de evaluar la robustez del algoritmo bajo diferentes condiciones.

Simulación	μ_{true}	μ_{estimado}	$\phi_{h,\text{true}}$	$\phi_{h,\text{estimado}}$	$\sigma_{h,\text{true}}$	$\sigma_{h,\text{estimado}}$
1	1	0.933	0.5	0.476	1	0.858
2	1	0.985	0.3	0.397	0.9	0.627
3	1	0.989	0.8	0.731	0.1	0.102
4	3	2.927	0.2	0.757	0.1	0.099

Tabla 2.1: Comparación entre parámetros verdaderos y estimados.

En las simulaciones se observa que el algoritmo estima el parámetro μ con un error

relativamente bajo, lo que evidencia una buena capacidad de aproximación en este caso. Por otro lado, la estimación de σ_h resulta bastante precisa en general, aunque en algunas simulaciones se percibe una leve tendencia a subestimar este parámetro. Sin embargo, el algoritmo presenta problemas para estimar correctamente ϕ_h cuando el valor verdadero de este parámetro es pequeño y, al mismo tiempo, σ_h también es bajo. En estas condiciones, a pesar de que μ y σ_h se estiman con precisión, la estimación de ϕ_h se ve comprometida, lo que sugiere la necesidad de ajustar el método o analizar con mayor profundidad las condiciones que generan estas discrepancias.

2.2. Limitaciones del Algoritmo MCMC

A pesar de la flexibilidad del enfoque bayesiano basado en MCMC, este método presenta diversas limitaciones que pueden afectar su eficiencia y precisión en aplicaciones prácticas.

Uno de los problemas fundamentales del método es su ineficiencia computacional. Un factor que agrava este problema es la alta autocorrelación entre muestras sucesivas. Como consecuencia, las cadenas MCMC tardan más en explorar el espacio paramétrico de manera eficiente.

Otro problema relevante es la suavización excesiva de la estimación de la volatilidad. Dado que el algoritmo utiliza la muestra completa para actualizar cada estado latente, la estimación de la volatilidad tiende a ser más suave de lo que realmente ocurre en los datos financieros. Esto se debe a que, en cada iteración del muestreo, se emplea información de toda la serie de tiempo, incluyendo datos futuros respecto a un instante t. Como consecuencia, la volatilidad inferida en un punto dado está influenciada por observaciones que aún no estarían disponibles en un contexto de predicción real.

Este problema es especialmente evidente en los bordes de la muestra, donde la falta de datos futuros genera una estimación menos precisa de la volatilidad latente. En el interior de la muestra, la estimación de h_t es estabilizada por la información disponible antes y después del instante t, lo que contribuye a una representación más homogénea del proceso de volatilidad. Sin embargo, en los extremos de la serie, la ausencia de observaciones posteriores provoca que las estimaciones sean más erráticas o menos confiables, afectando significativamente cualquier intento de extrapolación de la volatilidad más allá del último punto observado.

En consecuencia, al utilizar este modelo con fines predictivos, se introduce un sesgo debido a la forma en que la volatilidad ha sido inferida dentro de la muestra, lo que puede llevar a una subestimación o sobreestimación del riesgo en horizontes futuros. Este efecto compromete la utilidad del modelo SV con MCMC para aplicaciones en pronóstico.

Capítulo 3

Predicción en Modelos SV

En los capítulos anteriores, se presentó un modelo de *volatilidad estocástica (SV)* con una dinámica de *AR(1) con regresión a la media* para la log-volatilidad. La estimación de los parámetros y los estados ocultos del modelo se realizó utilizando un enfoque *batch*, en el cual se ajustan los datos utilizando toda la muestra de observaciones disponible.

Si bien este enfoque es útil para el análisis en retrospectiva, presenta serias limitaciones en los bordes de la muestra y en especial en la predicción,. En particular, debido a la naturaleza del ajuste batch:

- **La incertidumbre en los extremos de la muestra es mayor**, ya que no hay datos futuros que ayuden a refinar la estimación de los estados ocultos.
- **Se generan efectos de borde**, lo que puede llevar a predicciones inexactas de la volatilidad en las últimas observaciones.
- **La predicción no es adecuada para aplicaciones en tiempo real**, ya que cada nueva observación requeriría volver a ajustar el modelo con todos los datos disponibles.

Dado que en muchas situaciones es crucial predecir la volatilidad futura en tiempo real, es necesario un enfoque de estimación recursivo. En lugar de recalcular toda la estimación cada vez que se incorpora una nueva observación, un enfoque recursivo permite actualizar la distribución del estado latente de forma progresiva, integrando la información previa con la nueva evidencia de manera eficiente.

Esto nos lleva al concepto de **filtros bayesianos**, los cuales aprovechan la estructura secuencial del problema para descomponer la inferencia en una combinación de predicción y actualización. A medida que se reciben nuevas observaciones, la estimación del estado se ajusta dinámicamente sin necesidad de reoptimizar el modelo sobre toda la

serie de datos. De esta manera, se logra una propagación eficiente de la incertidumbre y una inferencia flexible que se adapta a la evolución del proceso subyacente (Särkkä, 2013).

3.1. Metodología

3.1.1. El Concepto de Filtrado

El filtrado es una técnica utilizada en modelos de espacio de estados[1] para estimar un estado latente x_t en función de las observaciones disponibles hasta el tiempo t. Es decir, en lugar de ajustar todo el conjunto de datos simultáneamente, como en el enfoque batch, el filtrado se basa en la actualización recursiva de la estimación.

Matemáticamente, el filtrado busca calcular la distribución posterior del estado latente x_t dado los datos observados hasta ese momento:

$$p(x_t \mid y_{1:t}). \tag{3.1}$$

Teorema 3 (Bayesian filtering equations)**.** *Las ecuaciones recursivas del Filtro Bayesiano para calcular la distribución predictiva* $p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1})$ *y la distribución filtrada* $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$ *en el instante de tiempo* k *están dadas por las siguientes ecuaciones:*

- ***Inicialización****: La recursión comienza con la distribución a priori* $p(\mathbf{x}_0)$.
- ***Paso de Predicción****: La distribución predictiva del estado* $\mathbf{x}_k$ *en el instante* k*, dada la dinámica del modelo, se obtiene mediante la ecuación de Chapman-Kolmogorov:*

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1})\, \mathrm{d}\mathbf{x}_{k-1}. \tag{3.2}$$

[1] Un **modelo de espacio de estados** es una representación matemática de un sistema dinámico en términos de variables de estado no observables ($\mathbf{x}_k$) y observaciones ruidosas ($\mathbf{y}_k$). Se define mediante dos ecuaciones: el *modelo de evolución del estado*, que describe la dinámica probabilística del sistema $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$, y el *modelo de observación*, que relaciona las mediciones con el estado $p(\mathbf{y}_k \mid \mathbf{x}_k)$. La estimación de los estados ocultos en estos modelos se realiza a través de filtrado bayesiano, utilizando métodos como el Filtro de Kalman (Kalman, 1960) o el Filtro de Partículas (Doucet et al., 2001), dependiendo de la linealidad y estructura probabilística del sistema.

- ***Paso de Actualización**: Dada la nueva observación $\mathbf{y}_k$ en el instante k, la distribución posterior del estado $\mathbf{x}_k$ se puede calcular mediante la regla de Bayes:*

$$p\left(\mathbf{x}_k \mid \mathbf{y}_{1:k}\right) = \frac{1}{Z_k} p\left(\mathbf{y}_k \mid \mathbf{x}_k\right) p\left(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}\right), \tag{3.3}$$

donde la constante de normalización Z_k está dada por:

$$Z_k = \int p\left(\mathbf{y}_k \mid \mathbf{x}_k\right) p\left(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}\right) \mathrm{d}\mathbf{x}_k. \tag{3.4}$$

En nuestro caso, aplicar un esquema de filtrado permitirá:

- Mejorar las estimaciones de volatilidad en los bordes de la muestra.
- Hacer predicciones en tiempo real sin necesidad de recalcular desde cero.[2]
- Incorporar nueva información de manera eficiente a medida que se van observando nuevos datos.

Dado que el modelo de volatilidad estocástica presenta una estructura no lineal, ya que la varianza de los retornos y_t depende exponencialmente del estado oculto h_t, es decir, $y_t \sim \mathcal{N}(0, e^{h_t})$, el filtrado no puede realizarse con el *Filtro de Kalman estándar*. En su lugar, se requieren métodos más avanzados como el Filtro de Partículas (Doucet et al., 2001; Gordon et al., 1993), que permiten manejar la no linealidad y la no gaussianidad del problema.

3.1.2. Muestreo por Importancia

El muestreo por importancia es un método Monte Carlo utilizado para aproximar integrales de esperanza cuando la distribución de probabilidad objetivo es difícil de muestrear directamente. En muchos problemas de inferencia, el objetivo es calcular la esperanza de una función $g(\mathbf{x}_k)$ respecto a una distribución posterior de estados latentes $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$:

$$E_p[g(\mathbf{x}_k)] = \int g(\mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) d\mathbf{x}_k. \tag{3.5}$$

[2]No obstante, conviene señalar que, a medida que se incorporan nuevas observaciones, puede ser prudente volver a estimar los parámetros del modelo (μ, ϕ_h, σ_h).

Sin embargo, si la distribución $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$ es difícil de muestrear o la integral no tiene solución analítica, esta esperanza es muy difícil de calcular analíticamente. Para resolver este problema, el muestreo por importancia reescribe la integral usando una distribución de importancia $\pi(\mathbf{x}_k)$, que sea más sencilla de muestrear:

$$E_p[g(\mathbf{x}_k)] = \int g(\mathbf{x}_k)\frac{p(\mathbf{x}_k \mid \mathbf{y}_{1:k})}{\pi(\mathbf{x}_k)}\pi(\mathbf{x}_k)d\mathbf{x}_k. \tag{3.6}$$

Esta reformulación permite aproximar la esperanza usando un conjunto de muestras $\mathbf{x}_k^{(i)} \sim \pi(\mathbf{x}_k)$, en lugar de $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$, y estimar la integral mediante un promedio ponderado:

$$E_p[g(\mathbf{x}_k)] \approx \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x}_k^{(i)})w_k^{(i)}, \tag{3.7}$$

donde los pesos de importancia están definidos como:

$$w_k^{(i)} = \frac{p(\mathbf{x}_k^{(i)} \mid \mathbf{y}_{1:k})}{\pi(\mathbf{x}_k^{(i)})}. \tag{3.8}$$

Dado que la distribución $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$ no suele conocerse de manera analítica, se puede escribir en términos de una función de verosimilitud y una distribución previa:

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) = \frac{1}{Z_k}p(\mathbf{y}_k \mid \mathbf{x}_k)p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}), \tag{3.9}$$

Por lo tanto los pesos quedan

$$w_k^{(i)} = \frac{p(\mathbf{x}_k^{(i)} \mid \mathbf{y}_{1:k})}{\pi(\mathbf{x}_k^{(i)})} = \frac{\frac{1}{Z_k}p(\mathbf{y}_k \mid \mathbf{x}_k^{(i)})\, p(\mathbf{x}_k^{(i)} \mid \mathbf{y}_{1:k-1})}{\pi(\mathbf{x}_k^{(i)})} \tag{3.10}$$

Al normalizar, se define

$$\tilde{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^{N} w_k^{(j)}} \tag{3.11}$$

reemplazando $w_k^{(i)}$,

$$\tilde{w}_k^{(i)} = \frac{\frac{1}{Z_k}p(\mathbf{y}_k \mid \mathbf{x}_k^{(i)})\, p(\mathbf{x}_k^{(i)} \mid \mathbf{y}_{1:k-1})/\pi(\mathbf{x}_k^{(i)})}{\sum_{j=1}^{N}\frac{1}{Z_k}p(\mathbf{y}_k \mid \mathbf{x}_k^{(j)})\, p(\mathbf{x}_k^{(j)} \mid \mathbf{y}_{1:k-1})/\pi(\mathbf{x}_k^{(j)})} \tag{3.12}$$

Dado que Z_k aparece en todos los términos, se cancela, dejando

$$\tilde{w}_k^{(i)} = \frac{p(\mathbf{y}_k \mid \mathbf{x}_k^{(i)})\, p(\mathbf{x}_k^{(i)} \mid \mathbf{y}_{1:k-1})/\pi(\mathbf{x}_k^{(i)})}{\sum_{j=1}^{N} p(\mathbf{y}_k \mid \mathbf{x}_k^{(j)})\, p(\mathbf{x}_k^{(j)} \mid \mathbf{y}_{1:k-1})/\pi(\mathbf{x}_k^{(j)})} \tag{3.13}$$

Esto permite calcular los pesos sin necesidad de conocer Z_k.

Para que el muestreo por importancia funcione correctamente, la distribución de importancia $\pi(\mathbf{x}_k)$ debe cumplir ciertas condiciones: (i) debe ser fácil de muestrear, (ii) debe cubrir la región de alta probabilidad de $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$, es decir, $\pi(\mathbf{x}_k) > 0$ donde $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$ es grande, y (iii) la relación $\frac{p(\mathbf{x}_k|\mathbf{y}_{1:k})}{\pi(\mathbf{x}_k)}$ no debe ser extremadamente variable, para evitar inestabilidad en los pesos (Douc et al., 2009).

3.1.3. Muestreo por Importancia Secuencial

El muestreo por importancia secuencial (*Sequential Importance Sampling*, SIS) es una extensión del muestreo por importancia que permite actualizar de manera recursiva la estimación a medida que se incorporan nuevas observaciones en problemas dinámicos. En muchos modelos de espacio de estados, el objetivo es estimar una secuencia de estados ocultos $\mathbf{x}_{1:T}$ dados datos observacionales $\mathbf{y}_{1:T}$. Sin embargo, calcular la distribución posterior $p(\mathbf{x}_{1:T} \mid \mathbf{y}_{1:T})$ no resoluble analíticamente en la mayoría de los casos.

Para abordar este problema, el enfoque SIS permite aproximar la distribución posterior mediante un conjunto de N muestras $\{\mathbf{x}_{0:k}^{(i)}\}$ y sus pesos asociados $w_k^{(i)}$, que se actualizan a medida que llegan nuevas observaciones.

Dado un modelo de espacio de estados de la forma:

$$\mathbf{x}_k \sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}), \tag{3.14}$$

$$\mathbf{y}_k \sim p(\mathbf{y}_k \mid \mathbf{x}_k), \tag{3.15}$$

el muestreo por importancia secuencial construye una aproximación recursiva de la distribución de estados usando una distribución de importancia $\pi(\mathbf{x}_{0:k} \mid \mathbf{y}_{1:k})$, que permite actualizar los pesos de manera eficiente en cada paso de tiempo.

Siguiendo un procedimiento similar al muestreo por importancia, la esperanza de una función $g(\mathbf{x}_k)$ se puede escribir como:

$$E[g(\mathbf{x}_k) \mid \mathbf{y}_{1:k}] = \int g(\mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) d\mathbf{x}_k. \tag{3.16}$$

Como en el caso estándar de muestreo por importancia, se introduce una distribución de importancia $\pi(\mathbf{x}_{0:k} \mid \mathbf{y}_{1:k})$ y se reescribe la integral como:

$$E[g(\mathbf{x}_k) \mid \mathbf{y}_{1:k}] = \int g(\mathbf{x}_k) \frac{p(\mathbf{x}_{0:k} \mid \mathbf{y}_{1:k})}{\pi(\mathbf{x}_{0:k} \mid \mathbf{y}_{1:k})} \pi(\mathbf{x}_{0:k} \mid \mathbf{y}_{1:k}) d\mathbf{x}_k. \tag{3.17}$$

Dado que el muestreo se realiza de manera secuencial, la distribución de importancia puede escribirse en términos de una relación recursiva. Siguiendo la regla de la cadena de la probabilidad:

$$\pi\left(\mathbf{x}_{0:k} \mid \mathbf{y}_{1:k}\right) = \pi\left(\mathbf{x}_k \mid \mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}\right) \pi\left(\mathbf{x}_{0:k-1} \mid \mathbf{y}_{1:k}\right) \tag{3.18}$$

Ahora, en el segundo término del lado derecho, aplicamos la propiedad de Markov del modelo de estados ocultos. En estos modelos, el estado $\mathbf{x}_k$ solo depende directamente del estado anterior $\mathbf{x}_{k-1}$ y de la observación $\mathbf{y}_k$, lo que implica que la información sobre $\mathbf{x}_{0:k-1}$ solo depende de las observaciones hasta el instante $k-1$, esto se expresa como:

$$\pi\left(\mathbf{x}_{0:k-1} \mid \mathbf{y}_{1:k}\right) = \pi\left(\mathbf{x}_{0:k-1} \mid \mathbf{y}_{1:k-1}\right) \tag{3.19}$$

Entonces,

$$\pi\left(\mathbf{x}_{0:k} \mid \mathbf{y}_{1:k}\right) = \pi\left(\mathbf{x}_k \mid \mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}\right) \pi\left(\mathbf{x}_{0:k-1} \mid \mathbf{y}_{1:k-1}\right) \tag{3.20}$$

Por otro lado, aplicando el mismo razonamiento, la distribución objetivo se descompone en:

$$p\left(\mathbf{x}_{0:k} \mid \mathbf{y}_{1:k}\right) \propto p\left(\mathbf{y}_k \mid \mathbf{x}_k\right) p\left(\mathbf{x}_k \mid \mathbf{x}_{k-1}\right) p\left(\mathbf{x}_{0:k-1} \mid \mathbf{y}_{1:k-1}\right) \tag{3.21}$$

Usamos estas dos descomposiciones en la ecuación original de los pesos y llegamos a que

$$w_k^{(i)} \propto \frac{p\left(\mathbf{y}_k \mid \mathbf{x}_k^{(i)}\right) p\left(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{k-1}^{(i)}\right) p\left(\mathbf{x}_{0:k-1}^{(i)} \mid \mathbf{y}_{1:k-1}\right)}{\pi\left(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{1:k}\right) \pi\left(\mathbf{x}_{0:k-1}^{(i)} \mid \mathbf{y}_{1:k-1}\right)} \tag{3.22}$$

Ahora, notamos que el peso en el paso anterior estaba definido como:

$$w_{k-1}^{(i)} \propto \frac{p\left(\mathbf{x}_{0:k-1}^{(i)} \mid \mathbf{y}_{1:k-1}\right)}{\pi\left(\mathbf{x}_{0:k-1}^{(i)} \mid \mathbf{y}_{1:k-1}\right)} \tag{3.23}$$

Sustituyendo esta expresión en la ecuación anterior podemos observar que el peso $w_t^{(i)}$ se actualiza en cada paso:

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p\left(\mathbf{y}_k \mid \mathbf{x}_k^{(i)}\right)\, p\left(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{k-1}^{(i)}\right)}{\pi\left(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{0:k-1}^{(i)}, \mathbf{y}_{1:k}\right)} \tag{3.24}$$

Cada término en esta ecuación cumple una función específica.

Cada partícula $\mathbf{x}_{0:t}^{(i)}$ representa una posible trayectoria del sistema desde el tiempo inicial 0 hasta el tiempo actual t. El peso asociado a cada partícula, $w_t^{(i)}$, indica la probabilidad de que dicha trayectoria sea consistente con los datos observados y la dinámica del sistema.

El peso previo $w_{t-1}^{(i)}$ mantiene la importancia relativa de la trayectoria de la partícula hasta el tiempo $t-1$. El término $p(\mathbf{y}_t \mid \mathbf{x}_t^{(i)})$ mide la verosimilitud de la observación actual $\mathbf{y}_t$, evaluando qué tan bien la partícula explica la medición. El factor $p(\mathbf{x}_t^{(i)} \mid \mathbf{x}_{t-1}^{(i)})$ representa la coherencia de la partícula con la dinámica del sistema, mientras que $\pi(\mathbf{x}_t^{(i)} \mid \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})$ corrige el peso según la distribución de importancia utilizada para generar las muestras.

Si el peso $w_t^{(i)}$ es alto, significa que la trayectoria de la partícula es consistente con las mediciones y con la evolución del sistema. Si el peso es bajo, indica que la partícula ha seguido un camino poco probable.

Este enfoque permite mantener una estimación eficiente de la distribución posterior de los estados ocultos sin necesidad de recomputar desde cero toda la historia de datos en cada paso de tiempo. Sin embargo, el muestreo por importancia secuencial presenta un problema conocido como degeneración de partículas, donde después de varias iteraciones, la mayoría de los pesos $w_k^{(i)}$ tienden a valores cercanos a cero, lo que reduce la efectividad del método ya que la mayoría de las partículas contribuyen muy poco o nada a la estimación del estado (Gordon et al., 1993).

3.1.4. Filtro de Partículas (SIR)

El muestreo por importancia secuencial (*Sequential Importance Sampling, SIS*) tiene la desventaja de que, tras varias iteraciones, la mayoría de los pesos de importancia $w_k^{(i)}$ tienden a valores muy pequeños. Esto provoca que unas pocas partículas dominen la estimación y muchas de ellas tengan una contribución despreciable. Este problema es conocido como degeneración de partículas, y reduce la eficiencia del método.

Para resolverlo, se introduce el remuestreo (*resampling*), dando lugar a una técnica más estable denominada muestreo por importancia secuencial con remuestreo (*Sequential Importance Resampling, SIR*), también conocido como filtro de partículas.

El filtro de partículas se basa en la misma estructura del SIS, pero incorpora una etapa de remuestreo en la que las partículas con mayor peso son replicadas y las de menor peso son eliminadas, asegurando que la distribución de partículas siga representando correctamente la distribución posterior.

Dado un modelo de espacio de estados:

$$x_k \sim p(x_k \mid x_{k-1}), \tag{3.25}$$

$$y_k \sim p(y_k \mid x_k), \tag{3.26}$$

y utilizando el modelo de transición como distribución de importancia:

$$\pi(x_k \mid x_{0:k-1}, y_{1:k}) = p(x_k \mid x_{k-1}), \tag{3.27}$$

los pesos se actualizan como:

$$w_k^{(i)} \propto w_{k-1}^{(i)} p(y_k \mid x_k^{(i)}). \tag{3.28}$$

Para evitar la degeneración de partículas, en cada paso de tiempo k se calcula el número efectivo de partículas:

$$N_{\text{ef}} = \frac{1}{\sum_{i=1}^{N} (w_k^{(i)})^2}. \tag{3.29}$$

donde N denota el número de partículas (o simulaciones independientes de la distribución de importancia) utilizadas en el algoritmo. Si N_{ef} cae por debajo de un umbral,

típicamente $N/2$, se activa el remuestreo, que consiste en generar un nuevo conjunto de partículas a partir de la distribución discreta definida por los pesos normalizados:

$$\tilde{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^{N} w_k^{(j)}}. \tag{3.30}$$

Las nuevas partículas $x_k^{(i)}$ se seleccionan mediante *resampling multinomial*, para redistribuir las partículas en función de sus pesos normalizados $\tilde{w}_k^{(j)}$. Este proceso se expresa matemáticamente como:

$$x_k^{(i)} \sim \sum_{j=1}^{N} \tilde{w}_k^{(j)} \delta \left(x_k - x_k^{(j)} \right),$$

donde $\delta(\cdot)$ representa la función delta de Dirac, lo que indica que las nuevas partículas $x_k^{(i)}$ son seleccionadas a partir del conjunto de partículas existentes $x_k^{(j)}$ con una probabilidad proporcional a sus pesos normalizados $\tilde{w}_k^{(j)}$.

En el *resampling multinomial*, se generan N nuevas partículas mediante un muestreo proporcional a los pesos $\tilde{w}_k^{(j)}$, es decir, las partículas con pesos mayores tienen más probabilidad de ser replicadas, mientras que las de peso bajo pueden desaparecer.

Tras el remuestreo, los pesos de las partículas se reinicializan uniformemente como:

$$w_k^{(i)} = \frac{1}{N} \tag{3.31}$$

Esto significa que después de redistribuir las partículas, todas reciben la misma ponderación inicial, eliminando acumulaciones de peso en unas pocas partículas y permitiendo que la nueva distribución capture mejor la dinámica del sistema en los siguientes pasos de la estimación.

El procedimiento general del filtro de partículas se resume en los siguientes pasos:

Algorithm 5 Filtro de Partículas (Bootstrap Filter)

1: **Inicialización:** Generar N partículas $x_0^{(i)} \sim p(x_0)$ con pesos $w_0^{(i)} = \frac{1}{N}$.
2: **for** $k = 1, \ldots, T$ **do**
3: **Predicción:** Generar $x_k^{(i)} \sim p(x_k \mid x_{k-1}^{(i)})$.
4: **Actualización de pesos (Bootstrap):**

$$w_k^{(i)} = w_{k-1}^{(i)}\, p(y_k \mid x_k^{(i)})$$

5: **Normalización:**

$$\tilde{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^{N} w_k^{(j)}}$$

6: **Remuestreo (opcional):** Si $N_{\text{ef}} = \frac{1}{\sum_{i=1}^{N} (\tilde{w}_k^{(i)})^2} < N_{\text{umbral}}$, muestrear $\{x_k^{(i)}\}$ con probabilidad $\tilde{w}_k^{(i)}$ y fijar $w_k^{(i)} = \frac{1}{N}$.
7: **end for**

Nota. En el caso del modelo de volatilidad estocástica, las partículas $x_t^{(i)}$ corresponden a las variables latentes de volatilidad h_t. Por lo tanto, en la etapa de predicción del filtro Bootstrap, muestrear $x_k^{(i)} \sim p(x_k \mid x_{k-1}^{(i)})$ equivale a generar realizaciones de h_k a partir de la dinámica del modelo.

Asimismo, en el filtro Bootstrap la distribución de propuesta se elige como

$$\pi(x_k \mid x_{0:k-1}, y_{1:k}) = p(x_k \mid x_{k-1}),$$

es decir, la dinámica del modelo de estado. Esta elección hace que el cociente en la expresión general de los pesos se simplifique, dando lugar a la actualización

$$w_k^{(i)} \propto w_{k-1}^{(i)}\, p(y_k \mid x_k^{(i)}),$$

que es la que se implementa en el algoritmo.

3.2. Aplicación del Filtro de Partículas al Modelo SV

El problema consiste en estimar la secuencia de estados ocultos $\{h_t\}$ dados los retornos observados $\{y_t\}$

Un modelo de espacio de estados se compone de:

Ecuación de estado: describe la evolución del estado oculto h_t.

Ecuación de observación: describe cómo las observaciones y_t dependen del estado h_t.

En el modelo SV, estas ecuaciones están dadas por:

$$y_t \mid h_t \sim \mathcal{N}(0, e^{h_t}), \tag{3.32}$$

$$h_t = \mu + \phi_h(h_{t-1} - \mu) + \sigma_h \xi_t, \quad \xi_t \sim \mathcal{N}(0,1). \tag{3.33}$$

Nuestro objetivo es predecir h_t en cada instante t basándonos en las observaciones previas $y_{1:t}$.

El filtro de partículas consiste en representar la distribución del estado $p(h_t \mid y_{1:t})$ mediante un conjunto de N muestras ponderadas, denominadas partículas $\{h_t^{(i)}\}_{i=1}^N$, con pesos $w_t^{(i)}$ que reflejan su importancia relativa. El procedimiento sigue los siguientes pasos:

1. **Estimación de parámetros:** Antes de la implementación del filtro de partículas, se estiman los parámetros del modelo autorregresivo utilizando el método de volatilidad estocástica (SV). Esto permite determinar los valores óptimos de los parámetros μ, ϕ_h y σ_h que describen la evolución del estado latente.

2. **Inicialización:** Se generan N partículas $h_0^{(i)} \sim \mathcal{N}(\mu, \sigma_h^2)$ con pesos iniciales iguales $w_0^{(i)} = \frac{1}{N}$.

3. **Predicción:** En cada paso temporal, las partículas evolucionan según la ecuación de estado:
$$h_t^{(i)} = \mu + \phi_h(h_{t-1}^{(i)} - \mu) + \sigma_h \xi_t^{(i)}, \quad \xi_t^{(i)} \sim \mathcal{N}(0,1). \tag{3.34}$$

4. **Actualización de pesos:** Se evalúa la verosimilitud de cada partícula bajo el

modelo de observación:

$$p(y_t \mid h_t^{(i)}) = \frac{1}{\sqrt{2\pi e^{h_t^{(i)}}}} \exp\left(-\frac{y_t^2}{2e^{h_t^{(i)}}}\right). \tag{3.35}$$

Para evitar problemas numéricos, se trabaja con la verosimilitud en escala logarítmica:

$$\log p(y_t \mid h_t^{(i)}) = -\frac{1}{2}\left(\frac{y_t^2}{e^{h_t^{(i)}}} + h_t^{(i)}\right). \tag{3.36}$$

Los pesos normalizados se calculan como:

$$w_t^{(i)} = \frac{\exp(\log p(y_t \mid h_t^{(i)}) - \max_i \log p(y_t \mid h_t^{(i)}))}{\sum_{j=1}^{N} \exp(\log p(y_t \mid h_t^{(j)}) - \max_i \log p(y_t \mid h_t^{(i)}))}. \tag{3.37}$$

5. **Remuestreo:** Debido a que los pesos tienden a concentrarse en unas pocas partículas con el tiempo, se realiza un remuestreo multinomial donde se generan nuevas partículas a partir de las partículas existentes con probabilidad proporcional a $w_t^{(i)}$:

$$h_t^{(i)} = h_t^{(I_i))}, \tag{3.38}$$

donde los índices I_i se seleccionan según:

$$I_i \sim \text{Multinomial}(\{w_t^{(1)}, w_t^{(2)}, ..., w_t^{(N)}\}). \tag{3.39}$$

Después del remuestreo, los pesos se reinician a valores uniformes:

$$w_t^{(i)} = \frac{1}{N}, \quad \forall i. \tag{3.40}$$

6. **Estimación del estado:** Finalmente, el valor estimado del estado oculto h_t se obtiene como la media de las partículas:

$$\hat{h}_t = \frac{1}{N}\sum_{i=1}^{N} h_t^{(i)}. \tag{3.41}$$

Este procedimiento permite realizar una inferencia secuencial sobre la volatilidad latente h_t sin necesidad de almacenar todas las observaciones previas, lo que lo hace

adecuado para su implementación en tiempo real. Además, el remuestreo mitiga el problema de degeneración de partículas, asegurando que la estimación siga siendo eficiente a lo largo del tiempo.

La estructura del filtro de partículas implementado responde precisamente a la estructura general de un filtro bayesiano.

3.3. Predicción de y_{T+1}

Para realizar la predicción, se emplea la dinámica del estado latente h_t junto con la estructura del modelo observacional de los retornos y_t.

Dado que el proceso latente de volatilidad sigue un modelo AR(1):

$$h_{t+1} = \mu + \phi_h(h_t - \mu) + \sigma_h \xi_t, \quad \xi_t \sim \mathcal{N}(0,1) \tag{3.42}$$

podemos generar muestras de h_{t+1} a partir de la distribución condicional:

$$h_{t+1}^{(i)} \sim \mathcal{N}\left(\mu + \phi_h(h_t - \mu), \sigma_h^2\right), \quad i = 1, \ldots, N \tag{3.43}$$

donde N es el número de partículas utilizadas en la estimación.

El modelo observacional establece que los retornos condicionales siguen una distribución normal con media cero y varianza dada por la volatilidad latente:

$$y_{t+1} \mid h_{t+1} \sim \mathcal{N}\left(0, e^{h_{t+1}}\right) \tag{3.44}$$

Dado un conjunto de N muestras generadas para h_{t+1}, podemos construir una distribución empírica de y_{t+1} como:

$$y_{t+1}^{(i)} \sim \mathcal{N}\left(0, e^{h_{t+1}^{(i)}}\right), \quad i = 1, \ldots, N \tag{3.45}$$

Esta distribución permite generar predicciones sobre el comportamiento futuro de los retornos del activo. En la implementación computacional, se simulan N valores de h_{t+1}, y para cada uno de ellos se extraen valores correspondientes de y_{t+1}, obteniendo así un conjunto de trayectorias posibles de retornos futuros.

La estimación puntual de y_{t+1} puede realizarse a través de la media de las simulaciones:

$$\hat{y}_{t+1} = \frac{1}{N} \sum_{i=1}^{N} y_{t+1}^{(i)} \tag{3.46}$$

Asimismo, es posible calcular intervalos de credibilidad para capturar la incertidumbre en la predicción. Un intervalo de credibilidad del 95 % para y_{t+1} se puede construir a partir de los cuantiles de la distribución empírica de $y_{t+1}^{(i)}$:

$$\left[Q_{2.5\,\%}(y_{t+1}^{(i)}), Q_{97.5\,\%}(y_{t+1}^{(i)}) \right].$$

Con esta metodología, se obtiene una predicción probabilística de los retornos futuros basada en la dinámica inferida de la volatilidad estocástica y su evolución en el tiempo.

3.4. Diagnóstico

Como h_t es un estado latente y no conocemos su verdadero valor, la validación del modelo no puede hacerse directamente comparando la estimación de h_t con un valor real. En su lugar, la estrategia adoptada es evaluar si los valores observados de y_t son consistentes con la estructura probabilística asumida en el modelo.

Dado que estamos interesados en evaluar la calidad de la predicción de la volatilidad, la manera más natural de hacerlo es a través de la distribución condicional de Y_t^2, no de Y_t. En este sentido, se evalúa la probabilidad acumulada asociada al valor observado y_{t+1}^2 bajo la distribución condicional predicha (Diebold et al., 1998; Diebold, 2006):

$$\Pr(Y_{t+1}^2 \leq y_{t+1}^{\text{obs}^2} \mid h_{t+1}). \tag{3.47}$$

Este cálculo se basa en la propiedad de que, si el modelo está correctamente especificado, los valores transformados por su función de distribución acumulada deben seguir una distribución uniforme $U(0,1)$ (Rosenberg, 1974). Es decir, para cada t:

Dado que $Y_t^2 \mid h_t \sim e^{h_t} \cdot \chi_1^2$, la distribución acumulada de u_t está dada por la función de distribución acumulada de una chi-cuadrado con un grado de libertad:

$$u_t = \Pr\left(Y_t^2 \leq y_t^2 \mid h_t\right) = F_{\chi_1^2}\left(\frac{y_t^2}{e^{h_t}}\right) \tag{3.48}$$

Si el modelo está correctamente especificado, entonces $u_t \sim U(0,1)$ (Hastie et al., 2009).

Esta propiedad se conoce como la Transformada de Probabilidad Integral (*Probability Integral Transform*) y es ampliamente utilizada para evaluar la calibración de modelos probabilísticos (Diebold et al., 1998). Si los valores u_t presentan sesgos sistemáticos, esto indicaría una mala especificación del modelo:

- Si la varianza está subestimada, los valores de u_t tenderán a concentrarse en valores cercanos a 1.
- Si la varianza está sobreestimada, los valores de u_t estarán más cercanos a 0.

Para verificar si los valores u_t siguen una distribución uniforme, se realiza la transformación a una normal estándar mediante la función inversa de la distribución normal:

$$n_t = \Phi^{-1}(u_t), \quad n_t \sim \mathcal{N}(0,1). \tag{3.49}$$

Si el modelo es correcto, la secuencia n_t debería seguir una distribución normal estándar $\mathcal{N}(0,1)$ sin autocorrelación. Para validar esta condición, se aplican los siguientes diagnósticos:

- **Test de Ljung-Box:** Verifica si la secuencia n_t es independiente, como se esperaría en una serie bien modelada (Box and Pierce, 1970; Ljung and Box, 1978).
- **Test de Shapiro-Wilk:** Evalúa si los valores n_t siguen una distribución normal estándar (Shapiro and Wilk, 1965).
- **Test de heterocedasticidad:** Se realiza una regresión de n_t^2 contra el tiempo para detectar posibles patrones en la varianza de los errores.

Este procedimiento permite evaluar si la varianza estimada e^{h_t} captura correctamente la incertidumbre de los datos y, en consecuencia, si el modelo SV ajusta adecuadamente la estructura de volatilidad de la serie temporal.

Capítulo 4

Datos Reales

En este capítulo se lleva a cabo un estudio empírico con el objetivo de evaluar el desempeño de los modelos de volatilidad estocástica y del filtro de partículas sobre datos reales del mercado financiero. Para ello, se analiza la serie temporal correspondiente al precio de Bitcoin (BTC) en dólares estadounidenses (USD), obtenida de la plataforma *Yahoo Finance* para el período comprendido entre el 1 de enero de 2022 y el 1 de enero de 2025.

La elección de este activo responde tanto a su relevancia en los mercados financieros contemporáneos como a su marcada volatilidad, lo cual lo convierte en un caso de estudio particularmente interesante para los modelos propuestos.

La metodología a seguir consiste, en primer lugar, en la estimación de los parámetros estructurales del modelo de volatilidad estocástica —concretamente, μ, ϕ y σ— a partir de los datos observados. Posteriormente, dichos parámetros serán utilizados como insumo para la aplicación del filtro de partículas, con el objetivo de estimar y predecir la evolución de la volatilidad latente en la serie analizada. Para la implementación, se utilizaron los scripts `TESIS_datos.R` y `TESIS_filtro_de_particulas.R`, disponibles en el repositorio de GitHub.

4.0.1. Introducción a Bitcoin (BTC)

Bitcoin es una criptomoneda descentralizada creada en 2009 por un individuo o grupo bajo el seudónimo de Satoshi Nakamoto. Se basa en la tecnología de cadena de bloques (blockchain) y utiliza un sistema de consenso llamado Prueba de Trabajo (Proof-of-Work) para validar transacciones y asegurar la red. A diferencia de las monedas fiduciarias tradicionales, Bitcoin no está respaldado por un banco central ni por ninguna entidad gubernamental, lo que lo convierte en un activo digital con caracterís-

ticas únicas.

Desde su creación, Bitcoin ha mostrado una alta volatilidad en su precio. A diferencia de los mercados tradicionales, que suelen presentar un comportamiento más estable, Bitcoin puede experimentar cambios abruptos en cortos períodos de tiempo.

Debido a esta naturaleza inusual, modelar la volatilidad de Bitcoin representa un desafío particular.

4.0.2. Descripción de los Datos y Cálculo de Retornos

Para analizar la dinámica de Bitcoin, primero calculamos los retornos logarítmicos a partir de la serie de precios. Dado que los precios financieros suelen seguir un proceso no estacionario, el uso de retornos nos permite trabajar con una serie más estable y apropiada para la modelización estadística.

El retorno logarítmico en el instante t se define como:

$$y_t = \log\left(\frac{P_t}{P_{t-1}}\right), \tag{4.1}$$

donde P_t representa el precio de Bitcoin en el tiempo t. Esta transformación es ampliamente utilizada en finanzas, ya que facilita la interpretación de los cambios porcentuales y hace que la serie sea más adecuada para modelización (Campbell et al., 1997; Shreve, 2004; Fama, 1965; Plerou et al., 1999).

4.0.3. Visualización de los Retornos

Una vez calculados los retornos, procedemos a graficarlos para visualizar su comportamiento a lo largo del tiempo. Como es característico en series financieras, observamos que los retornos presentan períodos de alta y baja volatilidad, reflejando la naturaleza heterocedástica de la serie. Además, se pueden identificar ciertos eventos extremos que corresponden a movimientos abruptos en el precio.

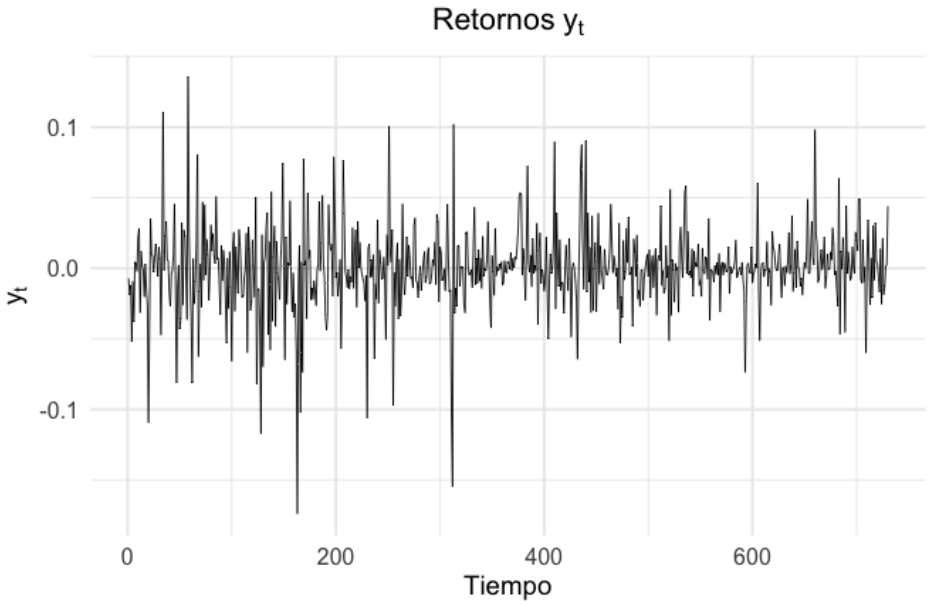

Figura 4.1: Evolución de los retornos de Bitcoin

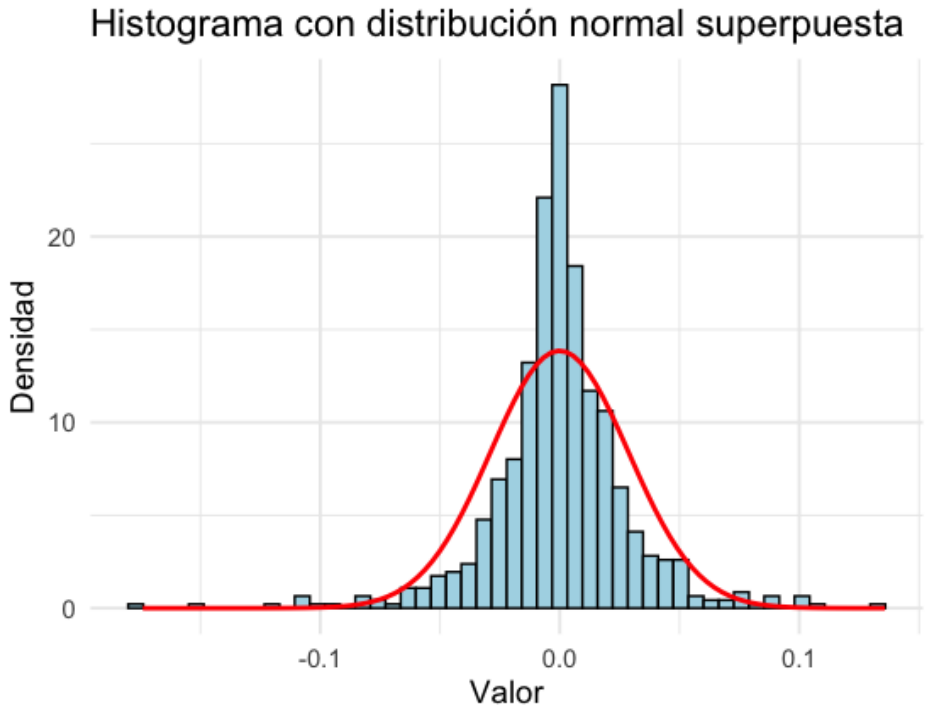

Figura 4.2: Histograma de los retornos con distribución normal superpuesta

Análisis de la Distribución de los Retornos

Para evaluar la distribución empírica de los retornos, construimos un histograma y lo superponemos con una distribución normal teórica con la misma media y desviación estándar que los datos observados.

En el gráfico resultante se observa que la distribución empírica de los retornos es más puntiaguda en el centro en comparación con la normal, lo que sugiere una mayor concentración de valores en torno a la media. Sin embargo, a diferencia de lo que se esperaría en una distribución normal, la presencia de colas más extendidas indica que los retornos extremos—tanto positivos como negativos—ocurren con mayor frecuencia de lo que la normal predice. Este comportamiento es una manifestación de la leptocurtosis, una propiedad bien documentada en los retornos financieros, la cual indica que los mercados exhiben cambios abruptos más a menudo de lo que un modelo gaussiano con varianza constante sugeriría (Cont, 2001; Mandelbrot, 1963; Bouchaud and Potters, 2003; Gopikrishnan et al., 1998).

Este fenómeno es especialmente relevante en el estudio de activos financieros, ya que la aparición frecuente de eventos extremos impacta directamente la evaluación del riesgo y la gestión de carteras.

Dicho de otra manera, los retornos de Bitcoin no solo tienden a agruparse en torno a su media con mayor intensidad que lo que predeciría una normal, sino que también exhiben una mayor incidencia de valores alejados de esta. Esto implica que el mercado de Bitcoin es propenso a experimentar episodios de alta volatilidad con movimientos bruscos en el precio.

4.1. Aplicación del Modelo de Volatilidad Estocástica

4.1.1. Especificaciones de los Parámetros y Priori

Definimos las distribuciones a priori para los parámetros μ, ϕ_h y σ_h. Se espera que μ sea cercano a 0, reflejando una volatilidad moderada en promedio, mientras que ϕ_h se asume alrededor de 0.8, indicando alta persistencia. Para facilitar la inferencia, reparametrizamos como $\beta_1 = \mu(1 - \phi_h)$ y $\beta_2 = \phi_h$, con priori:

$$b_0 = \begin{bmatrix} 0 \\ 0.8 \end{bmatrix}, \quad \Sigma_0 = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}.$$

Esta elección permite gran flexibilidad en μ y una varianza más acotada en ϕ_h.

Configuración del Muestreador MCMC

Se especifica un número total de $50,000$ iteraciones por cadena, lo que asegura la convergencia del muestreo y permite evaluar la mezcla entre cadenas. Además, se establecen los siguientes valores para los hiperparámetros de la distribución gamma a priori de σ_h:

- **Forma de la distribución a priori para** σ_h: Gamma$(1, 0.02)$.
- **Varianza de la propuesta en la actualización de** h_t: 1.5.

La ejecución del algoritmo se realizó con cuatro cadenas independientes, cada una con condiciones iniciales diferentes para favorecer la exploración del espacio paramétrico:

$$\begin{aligned}
&\text{Cadena 1:} \quad \mu_0 = 0.0, \quad \phi_{h,0} = 0.10, \quad \sigma_{h,0} = 0.10, \\
&\text{Cadena 2:} \quad \mu_0 = -1.0, \quad \phi_{h,0} = 0.70, \quad \sigma_{h,0} = 0.30, \\
&\text{Cadena 3:} \quad \mu_0 = 0.5, \quad \phi_{h,0} = -0.40, \quad \sigma_{h,0} = 0.20, \\
&\text{Cadena 4:} \quad \mu_0 = 1.0, \quad \phi_{h,0} = 0.95, \quad \sigma_{h,0} = 0.50.
\end{aligned}$$

4.2. Resultados de la Estimación

- **Burn-in:** Se descartan las primeras iteraciones del muestreo para evitar la influencia de los valores iniciales del algoritmo. En este caso, se han eliminado las primeras 3000 iteraciones, asegurando que las muestras utilizadas provienen de una fase estable de la cadena, reduciendo el impacto de los valores iniciales arbitrarios.

- **Thinning:** Se selecciona una muestra cada 100 iteraciones con el objetivo de reducir la autocorrelación dentro de la cadena MCMC. Dado que los algoritmos de muestreo MCMC generan valores secuenciales donde cada nueva muestra depende de la anterior, la autocorrelación puede sesgar la estimación de la distribución posterior. Aplicando un *thinning step* de 100, se asegura que las muestras sean más representativas de la verdadera distribución posterior, eliminando redundancias.

4.2.1. Tasa de Aceptación del Muestreo

Un aspecto clave en los algoritmos MCMC es la tasa de aceptación, que indica la proporción de iteraciones en las que se aceptó una nueva muestra en lugar de rechazarla. En este caso, la tasa de aceptación del muestreo de la volatilidad latente h_t fue la siguiente en cada una de las cuatro cadenas independientes:

Cadena	Tasa de aceptación en h_t
1	0.539
2	0.537
3	0.539
4	0.533

Estos valores se encuentran dentro del rango recomendado (entre 0.2 y 0.6) para un muestreo eficiente, lo que sugiere que la propuesta utilizada en el algoritmo logra un balance adecuado entre exploración y aceptación de nuevas muestras. No obstante, la eficiencia del algoritmo puede verse afectada por la presencia de autocorrelación en las muestras generadas, motivo por el cual se analizó la autocorrelación de las cadenas obtenidas y se evaluó el impacto del *thinning* como estrategia para mitigar este problema.

La Figura 4.3 muestra la función de autocorrelación (ACF) de las muestras MCMC obtenidas antes de aplicar cualquier corrección. Se observa que la autocorrelación es significativamente alta.

Para reducir la autocorrelación en las cadenas, se aplicó un thinning seleccionando una de cada k iteraciones. La Figura 4.4 muestra la función de autocorrelación tras la aplicación de esta técnica. Se observa que la autocorrelación ha disminuido notablemente en comparación con la Figura 4.3. Sin embargo, es importante destacar que el

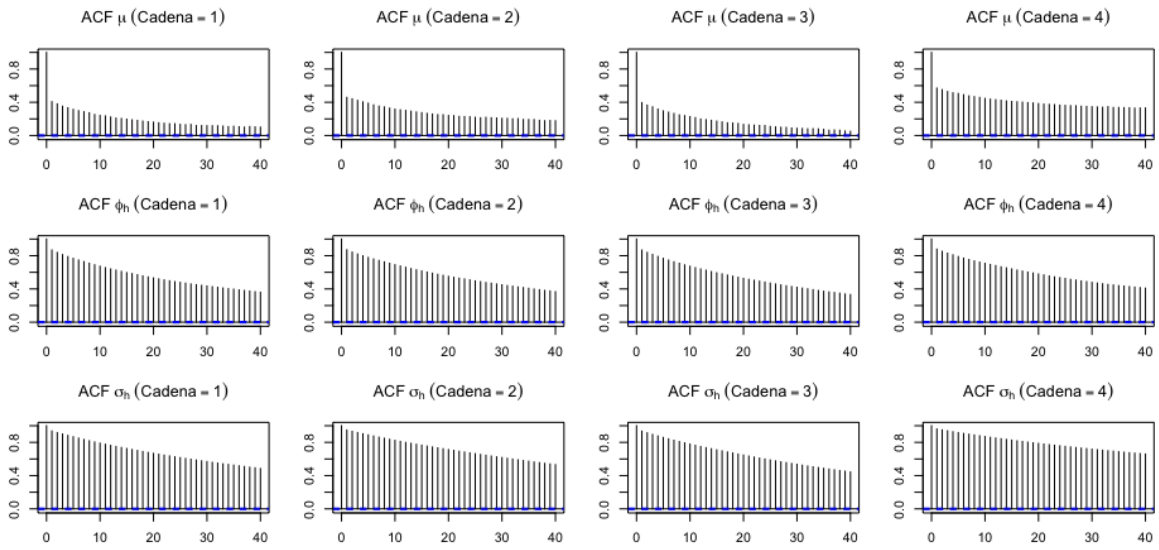

Figura 4.3: ACF

thinning implica una reducción del tamaño efectivo de la muestra, lo que puede afectar la precisión de las estimaciones si no se ha generado una cantidad suficiente de iteraciones en la cadena original.

Para cuantificar la eficiencia del muestreo, se calcula el **Tamaño de Muestra Efectivo (ESS)**, que mide cuántas muestras independientes equivalentes se han obtenido tras el muestreo.

Los valores calculados para cada parámetro son:

Cadena	ESS para μ	ESS para ϕ_h	ESS para σ_h
1	309.61	307.49	259.09
2	375.00	375.00	304.79
3	375.00	311.95	279.56
4	375.00	397.56	304.53

Tabla 4.1: Tamaños de muestra efectivos (ESS) por cadena

Estos valores indican que, aunque se han generado 50,000 iteraciones en el muestreo, el número de muestras efectivas es menor debido a la autocorrelación de las cadenas. Sin embargo, los valores de ESS obtenidos son suficientemente grandes para realizar inferencias confiables sobre los parámetros.

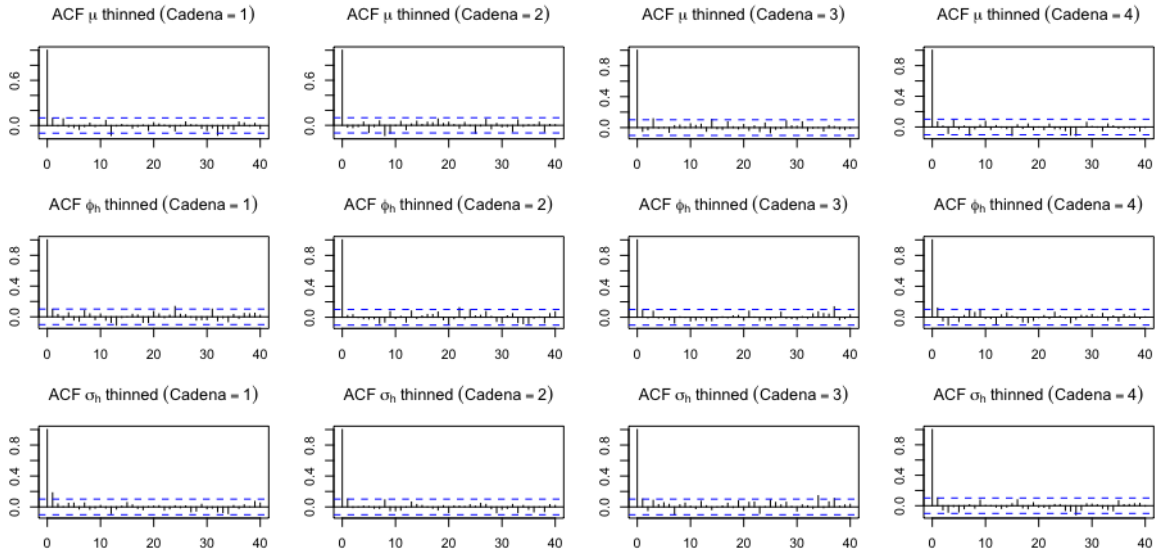

Figura 4.4: ACF thinned

Para evaluar la calidad de las cadenas generadas hacemos el análisis de las trazas, que permite visualizar la evolución de los parámetros muestreados a lo largo de las iteraciones. Para que el muestreo sea válido, es necesario que las cadenas converjan.

En el gráfico de las trazas obtenidas para los parámetros μ, ϕ_h y σ_h, se observa que todas las cadenas muestran un comportamiento estacionario después de un número inicial de iteraciones. No se identifican patrones de tendencia ni grandes oscilaciones fuera de lo esperado, lo que sugiere que el algoritmo ha alcanzado la convergencia y que las muestras pueden ser utilizadas para la inferencia posterior.

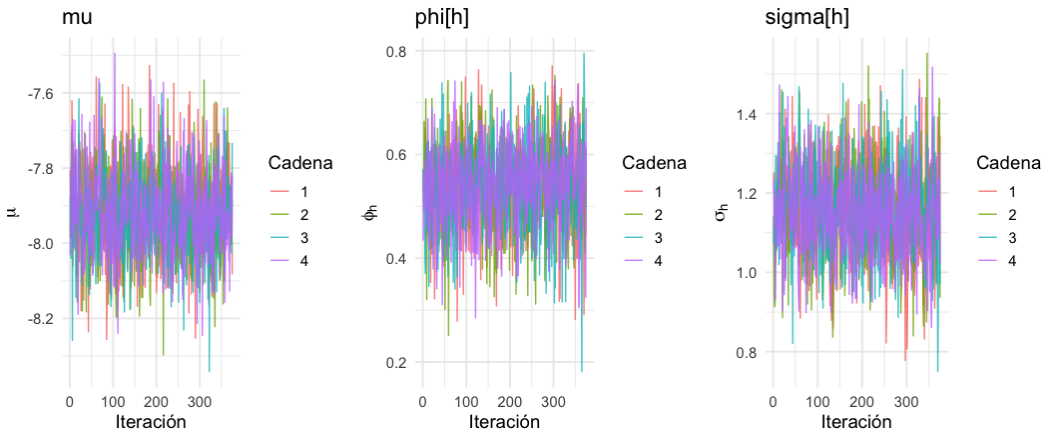

Figura 4.5: Trazas de las 4 cadenas de los parámetros μ, ϕ_h y σ_h.

La convergencia de las cadenas se evaluó mediante el diagnóstico de Gelman–Rubin, que compara la varianza dentro de cada cadena con la varianza entre cadenas. Un valor cercano a 1 indica que las cadenas han convergido hacia la misma distribución estacionaria. En este caso, los factores de reducción de escala potencial ($\hat{R}$) fueron:

$$\hat{R}_{\mu} = 1.00 \quad (UCI = 1.01), \qquad \hat{R}_{\phi_h} = 1.00, \qquad \hat{R}_{\sigma_h} = 1.00.$$

Estos resultados sugieren que las cuatro cadenas se mezclaron adecuadamente y alcanzaron la convergencia, ya que todos los valores de $\hat{R}$ se encuentran en el rango aceptado (≤ 1.1).

Dado que todas las trazas han alcanzado la estacionariedad y muestran un comportamiento adecuado, podemos concluir que el muestreo ha sido exitoso y que los valores obtenidos reflejan correctamente la incertidumbre sobre los parámetros.

4.2.2. Estimaciones Posteriores de los Parámetros

Los valores estimados de los parámetros del modelo se obtienen como el promedio de sus respectivas distribuciones posteriores, junto con sus intervalos de credibilidad al 95 %:

$$\hat{\mu} = -7.922 \quad \text{IC}_{95\,\%} = [-8.150,\ -7.668],$$

$$\hat{\phi}_h = 0.539 \quad \text{IC}_{95\,\%} = [0.358,\ 0.701],$$

$$\hat{\sigma}_h = 1.154 \quad \text{IC}_{95\,\%} = [0.922,\ 1.392].$$

Estos resultados sugieren que el parámetro μ tiene un valor negativo, lo que podría indicar un sesgo en la evolución logarítmica de la volatilidad. El valor estimado de ϕ_h, que mide la persistencia en la volatilidad, es menor a la expectativa previa (0.8), lo que sugiere que la volatilidad presenta menor dependencia en el tiempo de lo que se había supuesto a priori. Finalmente, el valor estimado de σ_h indica la magnitud de las fluctuaciones en la volatilidad, siendo relativamente alta, lo que es consistente con la naturaleza altamente volátil del mercado de criptomonedas.

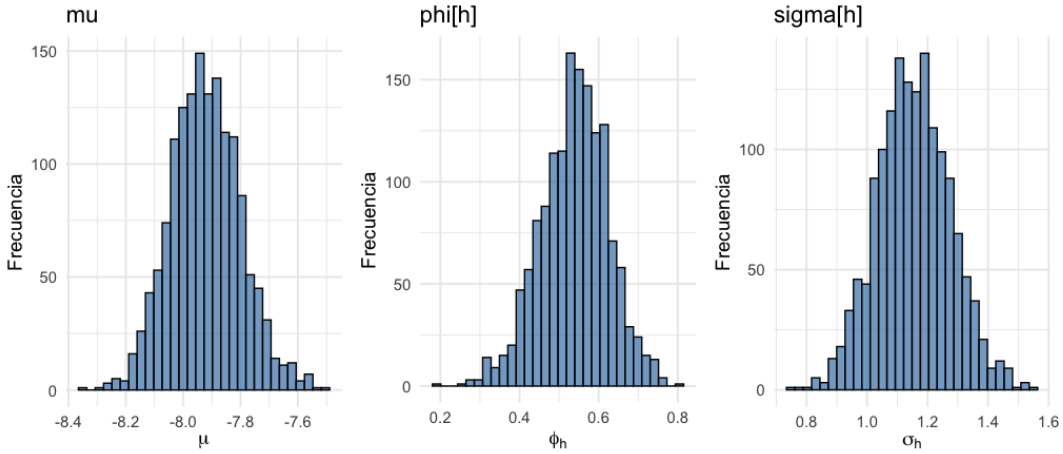

Figura 4.6: Histogramas de las cadenas MCMC correspondientes a cada parámetro estimado.

4.2.3. Resultados de la Estimación de la Volatilidad

Estimación de h_t con Intervalo de credibilidad

La figura 4.7 muestra la estimación posterior de h_t junto con su intervalo de credibilidad del 95 %. Esta representación es fundamental para visualizar cómo evoluciona la volatilidad a lo largo del tiempo y evaluar la incertidumbre en la predicción.

Lo que podemos observar es que, si bien hay cierta persistencia en la volatilidad, esta también es muy ruidosa. Cabe destacar que en mercados tradicionales, la volatilidad tiende a presentar una mayor persistencia y menor ruido, lo cual ha sido ampliamente documentado en la literatura. Esta diferencia en el comportamiento de la volatilidad

podría ser una de las características que dificultan la estimación del precio de este activo, ya que la alta variabilidad introduce mayor incertidumbre en los modelos de predicción.

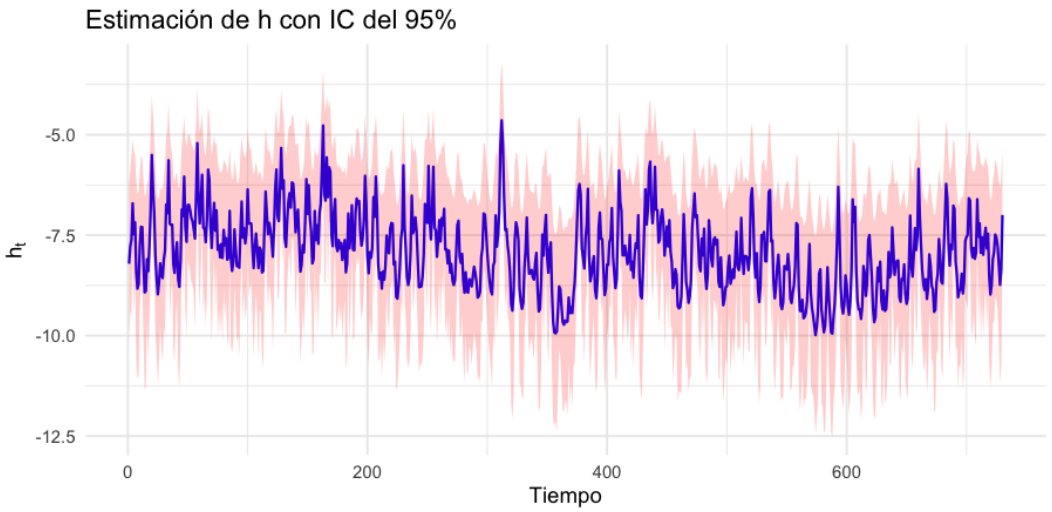

Figura 4.7: Estimación de la volatilidad h_t con intervalo de credibilidad del 95 %.

Se observa que la volatilidad estimada no es constante, sino que varía significativamente a lo largo del tiempo, con períodos de alta y baja volatilidad. En ciertos momentos, la volatilidad aumenta abruptamente

4.2.4. Relación entre la Volatilidad Estimada y los Retornos Logarítmicos

Para analizar la relación entre la volatilidad estimada y los retornos logarítmicos, se presentan dos gráficos superpuestos. En el segundo, se muestra la desviación estándar estimada de la volatilidad ($\exp(h_t/2)$), mientras que en el primero se grafican los retornos logarítmicos.

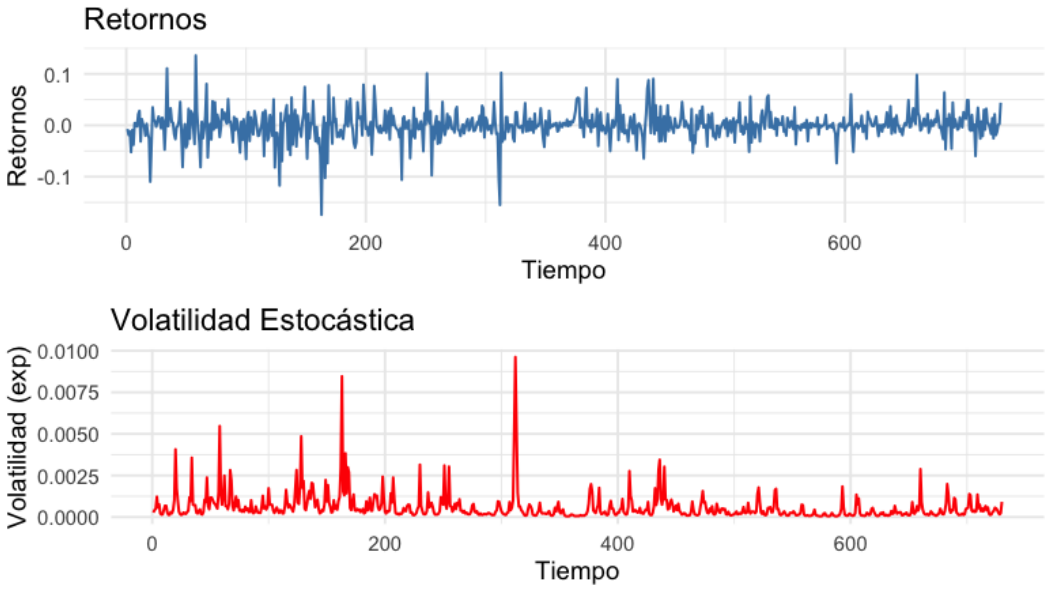

Figura 4.8: Comparación entre la desviación estándar de la volatilidad estimada y los retornos logarítmicos.

El análisis conjunto de estos gráficos permite observar cómo la volatilidad estimada varía en función de la magnitud de los retornos. En los períodos donde los retornos exhiben movimientos bruscos, la desviación estándar estimada de h_t tiende a incrementarse, reflejando un aumento en la incertidumbre y en la variabilidad del mercado.

4.3. Resultados de la Predicción

Los resultados del diagnóstico aplicado a la etapa de predicción muestran un desempeño razonable del modelo, aunque con ciertos matices a considerar. La prueba de independencia de Ljung-Box arroja un p-valor de 0.068, cercano al umbral típico de significancia (0.05), lo cual sugiere la posibilidad de una ligera autocorrelación remanente en los residuos transformados. Por otro lado, la prueba de normalidad (Shapiro-Wilk) no rechaza la hipótesis nula, lo cual es un buen indicio de que la distribución de los residuos predichos se aproxima a una normal estándar. Finalmente, el R^2 del modelo de heterocedasticidad es prácticamente nulo, indicando que no se observa tendencia en la varianza de los residuos transformados.

En conjunto, estos resultados respaldan parcialmente la validez del modelo para

la predicción de la volatilidad futura, aunque la cercanía del p-valor de la prueba de independencia a los valores críticos sugiere que podrían explorarse mejoras adicionales para capturar completamente la dinámica temporal del proceso.

Los resultados delas pruebas son los siguientes

- **Prueba de independencia (Ljung-Box)**: p-valor = 0.06798. Este resultado indica que no hay evidencia significativa de correlación en los residuos transformados n_t, lo que sugiere que el modelo captura correctamente la dinámica de la volatilidad.

- **Prueba de normalidad (Shapiro-Wilk)**: p-valor = 0.1389. Este valor sugiere que los residuos transformados están razonablemente cerca de seguir una distribución normal, aunque se observa una ligera desviación.

- **Prueba de heterocedasticidad ($\mathbf{R^2}$ de la regresión)**: $R^2 = 0.00034$. Un valor cercano a cero indica que no hay una tendencia sistemática en la varianza de los residuos transformados.

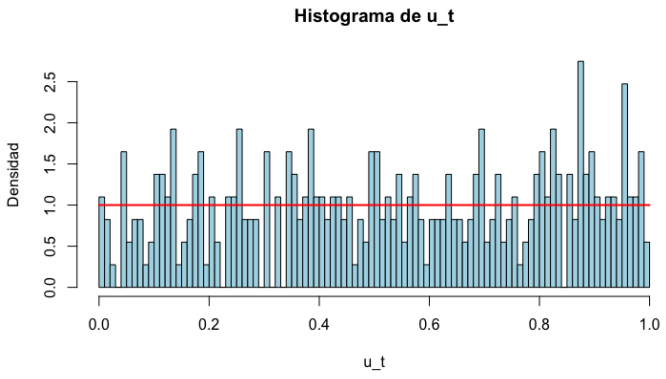

Figura 4.9: Histograma de la transformada de probabilidad integral u_t y su transformación normal n_t.

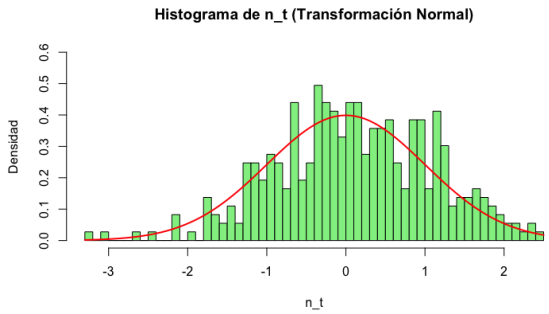

Figura 4.10: Histograma de la transformada de probabilidad integral u_t y su transformación normal n_t.

En la Figura 4.10, se presenta la visualización de la transformada de probabilidad integral (PIT) y su transformación a una distribución normal estándar. Esta representación permite evaluar visualmente si los valores transformados u_t siguen una distribución uniforme $U(0,1)$ y si la transformación normal $n_t = \Phi^{-1}(u_t)$ se ajusta a una distribución normal estándar.

Evolución de las distribuciones condicionales de y_t

El modelo de volatilidad estocástica plantea que, en cada instante de tiempo t, la variable observada y_t sigue una distribución normal con media cero y varianza condicional dada por $\exp(h_t)$. Como consecuencia, la distribución condicional de y_t cambia a lo largo del tiempo conforme evoluciona h_t.

Para visualizar esta dinámica, se construyó un conjunto de densidades condicionales de $y_t \mid h_t \sim \mathcal{N}(0, \exp(h_t))$ a distintos momentos t, utilizando las predicciones obtenidas mediante el filtro de partículas.

Este análisis permite observar de forma clara cómo se ajusta la forma de la distribución condicional a los distintos niveles de volatilidad implícitos en h_t. En momentos de alta volatilidad, las distribuciones se ensanchan reflejando mayor dispersión de y_t, mientras que en momentos de baja volatilidad se contraen, reflejando mayor certeza en torno al valor esperado cero.

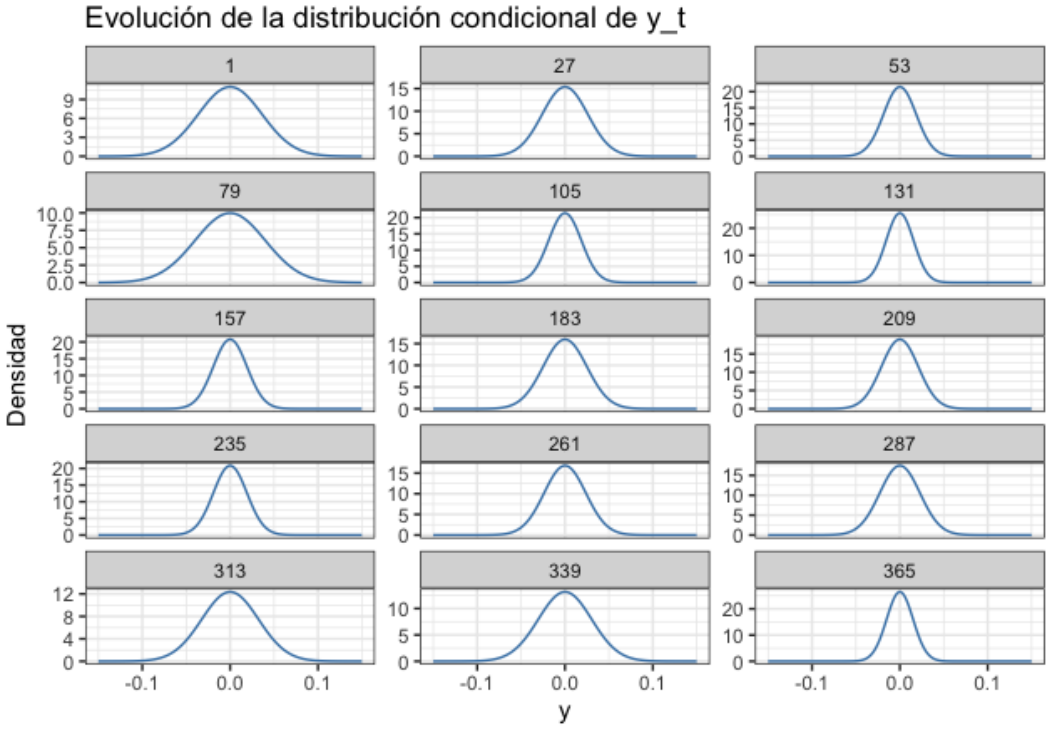

Figura 4.11: Evolución individual de Y_t.

Esta visualización resulta especialmente útil para entender el carácter dinámico y no estacionario del modelo, y refuerza la necesidad de utilizar enfoques de inferencia que incorporen la evolución temporal de la incertidumbre, como el filtro de partículas implementado en esta TFG.

Distribución marginal de y_t como mezcla de normales

La evolución temporal de la distribución condicional de los retornos y_t puede visualizarse mediante las densidades $p(y_t \mid h_t)$ para distintos valores de t, lo cual permite apreciar cómo varía la forma de la distribución a medida que cambia la volatilidad estimada h_t. Estas distribuciones, que en este modelo son normales centradas en cero con varianza $\exp(h_t)$, reflejan la naturaleza heterocedástica de la serie.

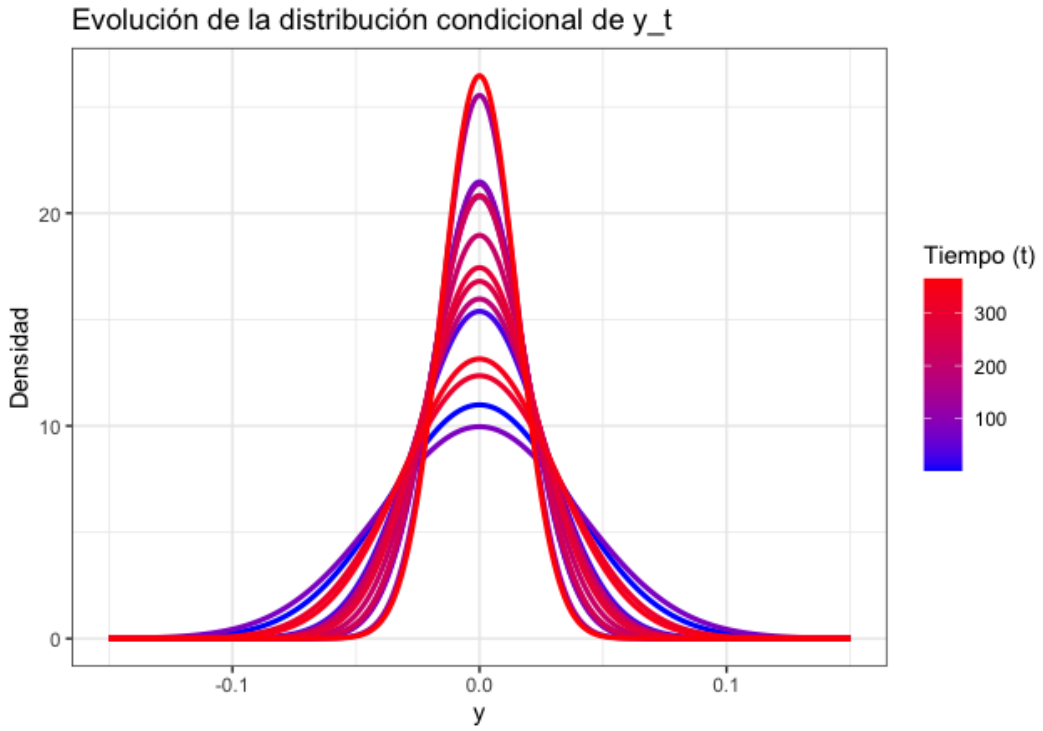

Figura 4.12: Evolucion conjunta de Y_t.

En lugar de analizar cada distribución condicional $p(y_t \mid h_t)$ por separado, es posible representar en un único gráfico la *evolución conjunta* de las densidades a lo largo del tiempo. Esta visualización permite apreciar cómo, en momentos de alta volatilidad, la distribución de y_t se vuelve más dispersa (con colas más pesadas), mientras que en períodos de menor volatilidad, se concentra más alrededor del cero.

Esta dinámica sugiere una interpretación probabilística más general: dado que h_t es una variable latente, su predicción a lo largo del tiempo induce una *distribución empírica* sobre sus posibles valores. En consecuencia, la variable y_t, condicionada a distintos valores de h_t, puede interpretarse como una *mezcla de distribuciones normales*, donde cada componente tiene varianza $\exp(h_t)$ y la mezcla está determinada por la distribución empírica de los h_t.

Desde esta perspectiva, la probabilidad de observar un valor específico de y_t *no es fija*, sino que depende de la incertidumbre asociada a h_t. La *densidad marginal* de y_t se construye como una superposición ponderada de distribuciones normales, y puede aproximarse mediante:

$$p(y_t) \approx \frac{1}{T} \sum_{t=1}^{T} \mathcal{N}(y_t \mid 0, \exp(h_t)) \tag{4.2}$$

donde T es la longitud de la serie de tiempo y h_t son las estimaciones puntuales del proceso latente.

Esta construcción representa una *mezcla empírica de normales* con varianzas heterogéneas. A diferencia de la suposición clásica de normalidad con varianza constante, aquí se tiene una mezcla que refleja los distintos niveles de volatilidad que el modelo identificó a lo largo del tiempo.

En términos prácticos, se comparó la densidad marginal empírica obtenida con esta mezcla contra el histograma de los datos observados y_t. El resultado mostró que la mezcla captura de forma adecuada la forma leptocúrtica (colas pesadas) de los retornos, algo que no puede ser replicado por una sola distribución normal con varianza constante. Adicionalmente, se realizó un test de Kolmogorov–Smirnov (KS) entre la CDF empírica de los datos y la función de distribución acumulada estimada de la mezcla, obteniendo un p-valor no significativo, lo que sugiere una buena concordancia entre ambas distribuciones.

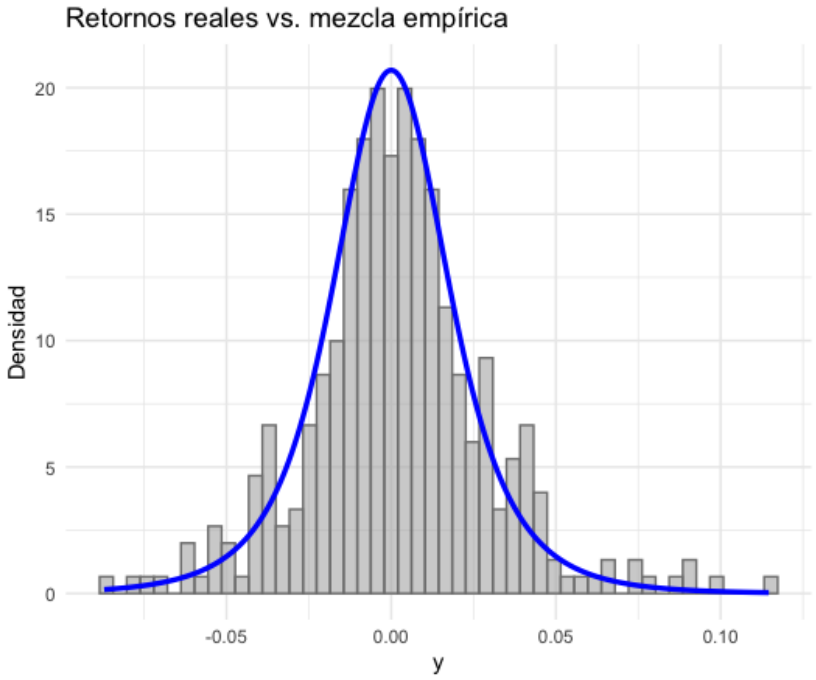

Figura 4.13: Densidad marginal de y_t estimada como mezcla empírica de normales, comparada con el histograma de los retornos reales.

Este enfoque proporciona no solo una representación más realista de la distribución de los retornos, sino también evidencia adicional de que el modelo es capaz de capturar adecuadamente la heterocedasticidad presente en los datos financieros.

4.4. Perspectivas y líneas futuras de investigación

Los resultados obtenidos en este trabajo proporcionan una base sólida para el estudio de modelos de volatilidad estocástica y su implementación mediante técnicas de inferencia bayesiana. Sin embargo, existen múltiples direcciones en las que se puede extender esta investigación para abordar estructuras más flexibles y representativas de la realidad financiera.

Una primera línea de trabajo consiste en la incorporación de modelos más complejos que capturen dinámicas adicionales en la volatilidad. En particular, los modelos con saltos, donde el proceso de volatilidad incluye movimientos abruptos que reflejan eventos extremos, representan una extensión natural del enfoque actual. Estos modelos pueden

expresarse como procesos de difusión con saltos de Poisson (**?**Bates, 1996) o mediante estructuras híbridas donde la volatilidad sigue un proceso de Markov oculto con cambios discretos de régimen.

Otro modelo ampliamente utilizado en finanzas es el modelo de Heston (Heston, 1993), que introduce una estructura de volatilidad estocástica continua basada en un proceso de difusión de Cox-Ingersoll-Ross (CIR). A diferencia del modelo SV tradicional considerado en este trabajo, en Heston la varianza sigue una ecuación diferencial estocástica de la forma:

$$dv_t = \kappa(\theta - v_t)dt + \sigma_v\sqrt{v_t}dW_t, \tag{4.3}$$

donde v_t representa la varianza instantánea, κ es la velocidad de reversión hacia la media θ, y σ_v controla la volatilidad de la varianza. Este modelo permite capturar de manera más realista las sonrisas de volatilidad observadas en los mercados financieros (**?**), y su inferencia mediante métodos Monte Carlo, incluyendo filtros de partículas y MCMC, es un área de estudio relevante (Jasra et al., 2011).

A pesar de la mayor complejidad de estos modelos, es importante destacar que las herramientas de inferencia utilizadas en este trabajo siguen siendo aplicables con modificaciones mínimas. El filtro de partículas puede adaptarse para manejar estructuras de volatilidad con procesos de difusión más sofisticados (Pitt and Shephard, 1999), mientras que el muestreo de Gibbs y otros métodos MCMC pueden extenderse para incluir parámetros adicionales relacionados con la dinámica de la volatilidad.

Otra dirección interesante es el estudio de modelos con memoria larga, donde la volatilidad no solo depende del estado inmediato anterior, sino que exhibe correlaciones persistentes en horizontes temporales más amplios. Modelos como el *Fractional Stochastic Volatility* (FSV) (Comte and Renault, 1998) permiten describir mejor la persistencia observada en datos financieros, y su implementación mediante métodos de Monte Carlo representa un desafío computacional interesante.

Finalmente, una posible extensión de esta investigación involucra la integración de técnicas de aprendizaje automático para la estimación de parámetros y la selección de modelos. Métodos basados en redes neuronales bayesianas (Gupta et al., 2011) o enfoques de inferencia variacional (Blei et al., 2017) pueden mejorar la eficiencia computacional y permitir estimaciones más robustas en entornos de alta dimensionalidad.

En conclusión, si bien este trabajo se centra en una familia específica de modelos, las metodologías presentadas ofrecen una base flexible para explorar estructuras más sofisticadas sin necesidad de modificar significativamente las herramientas de estimación. Las futuras investigaciones pueden enfocarse en adaptar y extender estas técnicas para capturar mejor las características estocásticas presentes en los mercados financieros y en otros procesos económicos dinámicos.

Capítulo 5

Conclusiones

En este trabajo se presentó un modelo básico de *volatilidad estocástica* (SV), el cual, a pesar de su simplicidad, planteó un desafío significativo en la estimación de sus parámetros y en la predicción del estado latente de la volatilidad. Para abordar la estimación de los parámetros del modelo, se implementó un esquema basado en el muestreo de Gibbs y métodos MCMC, lo que permitió obtener estimaciones consistentes de los parámetros del modelo. Sin embargo, debido a la naturaleza *batch* del muestreo en MCMC, este enfoque no resulta óptimo para realizar predicciones en línea del estado latente h_t.

Para solventar esta limitación, recurrimos a la utilización de un **filtro de partículas**, el cual, utilizando los parámetros estimados en el modelo SV, nos permite predecir la volatilidad h_t en tiempo real. Esta metodología combina la robustez de la estimación bayesiana con la flexibilidad de los métodos de inferencia secuenciales, permitiendo actualizar las creencias sobre el estado latente conforme ingresan nuevas observaciones.

En la aplicación realizada sobre los retornos de Bitcoins, el diagnóstico realizado sobre los residuos indica que el modelo es adecuado para los datos analizados, presentando independencia temporal, normalidad y ausencia de heterocedasticidad significativa. No obstante, aún existen oportunidades de mejora en la modelización de la dinámica de la volatilidad. En particular, se podrían considerar extensiones que incorporen saltos, memoria larga o estructuras más flexibles como el modelo de Heston, con el fin de capturar mejor las características empíricas observadas en series financieras.

Más allá de las mejoras en la especificación del modelo, es importante destacar que la lógica subyacente utilizada en este trabajo para la estimación de los parámetros y la inferencia del estado latente sigue siendo válida para modelos más complejos. La combinación de MCMC para la estimación de parámetros y el filtro de partículas para la inferencia en línea proporciona un marco metodológico generalizable que puede aplicarse

a una amplia variedad de modelos de volatilidad estocástica y otros procesos financieros dinámicos.

En conclusión, este trabajo contribuye al estudio de la volatilidad estocástica al proporcionar un esquema de inferencia eficiente que combina métodos bayesianos y técnicas de filtrado secuencial. Los resultados obtenidos validan la aplicabilidad de estas herramientas y sientan las bases para futuras extensiones hacia modelos más sofisticados sin perder la eficiencia computacional y la solidez en la estimación de parámetros.

Bibliografía

Bariviera, A. F. (2011). The Influence of Liquidity on Informational Efficiency: The Case of the Thai Stock Market. *Physica A: Statistical Mechanics and its Applications*, 390(23-24):4426–4432. doi:10.1016/j.physa.2011.07.008.

Bates, D. S. (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *The Review of Financial Studies*, 9(1):69–107. doi:10.1093/rfs/9.1.69.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society: Series B*, 36(2):192–236. doi:10.1111/j.2517-6161.1974.tb00999.x.

Besag, J. (1986). On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society: Series B*, 48(3):259–302. doi:10.1111/j.2517-6161.1986.tb01412.x.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. doi:10.1117/1.2819119.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877. doi:10.1080/01621459.2017.1285773.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307–327. doi:10.1016/0304-4076(86)90063-1.

Bouchaud, J.-P. and Potters, M. (2003). *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*. Cambridge University Press. doi:10.1017/CBO9780511753893.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press. doi:10.1515/9781400830213.

Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46(3):167–174. doi:10.1080/00031305.1992.10475878.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49(4):327–335. doi:10.1080/00031305.1995.10476177.

Comte, F. and Renault, E. (1998). Long Memory in Continuous-Time Stochastic Volatility Models. *Mathematical Finance*, 8(4):291–323. doi:10.1111/1467-9965.00057.

Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, 1(2):223–236. doi:10.1080/713665670.

Diebold, F. X. (2006). *Elements of Forecasting*. South-Western College Publishing.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–883. doi:10.2307/2527342.

Douc, R., Cappé, O., and Moulines, E. (2009). *Sequential Monte Carlo Methods for Filtering*. Springer Series in Statistics. Springer. doi:10.1007/978-0-387-75959-3.

Doucet, A., De Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer. doi:10.1007/978-1-4757-3437-9.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1008. doi:10.2307/1912773.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1):34–105. doi:10.1086/294743.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25(2):383–417. doi:10.2307/2325486.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC, 3rd edition. doi:10.1201/b16018.

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(6):721–741. doi:10.1109/TPAMI.1984.4767596.

Ghysels, E., Harvey, A. C., and Renault, E. (1996). Stochastic Volatility. *Handbook of Statistics*, 14:119–191. doi:10.1016/S0169-7161(96)14007-4.

Gopikrishnan, P., Meyer, M., Amaral, L. A. N., and Stanley, H. E. (1998). Inverse Cubic Law for the Distribution of Stock Price Variations. *The European Physical Journal B - Condensed Matter and Complex Systems*, 3(2):139–140. doi:10.1007/s100510050292.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. *IEE Proceedings F - Radar and Signal Processing*, 140(2):107–113. doi:10.1049/ip-f-2.1993.0015.

Gupta, A., Lam, M. S., and Ghosh, J. (2011). Bayesian Learning in Neural Networks. *Neural Computation*, 23(5):1342–1371. doi:10.1162/NECO_a_00113.

Hammersley, J. M. and Clifford, P. (1971). Markov Fields on Finite Graphs and Lattices. *Unpublished Manuscript*.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. doi:10.1007/978-0-387-84858-7.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109. doi:10.1093/biomet/57.1.97.

Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343. doi:10.1093/rfs/6.2.327.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2002). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 20(1):69–87. doi:10.1198/073500102753410408.

Jasra, A., Stephens, D. A., Doucet, A., and Tsagaris, T. (2011). Inference for Lévy-Driven Stochastic Volatility Models via Adaptive Sequential Monte Carlo. *Scandinavian Journal of Statistics*, 38(1):1–22. doi:10.1111/j.1467-9469.2010.00726.x.

Jordan, M. I. (2003). An Introduction to Probabilistic Graphical Models. *Unpublished Manuscript*.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering*, 82(1):35–45. doi:10.1115/1.3662552.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65(3):361–393. doi:10.1111/1467-937X.00050.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press.

Ljung, G. M. and Box, G. E. P. (1978). On a measure of a lack of fit in time series models. *Biometrika*, 65(2):297–303.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4):394–419. doi:10.1086/294632.

Matos, J. A. (2008). Series de Tiempo Financieras. Un Enfoque Alternativo. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 5:28–39.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092. doi:10.1063/1.1699114.

Montenegro, R. (2010). Medición de la Volatilidad en Series de Tiempo Financieras: Una Evaluación a la Tasa de Cambio Representativa del Mercado (TRM) en Colombia. *Revista de Economía Institucional*, 12(22):125–146.

Pitt, M. K. and Shephard, N. (1999). Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 94(446):590–599. doi:10.2307/2670179.

Plerou, V., Gopikrishnan, P., Amaral, L. A. N., Meyer, M., and Stanley, H. E. (1999). Scaling of the Distribution of Price Fluctuations of Individual Companies. *Physical Review E*, 60(6):6519–6529. doi:10.1103/PhysRevE.60.6519.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, 2nd edition. doi:10.1007/978-1-4757-4145-2.

Rosenberg, B. (1974). Extra-Market Components of Covariance in Security Returns. *The Journal of Financial and Quantitative Analysis*, 9(2):263–274. doi:10.2307/2330008.

Shapiro, S. S. and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3-4):591–611. doi:10.1093/biomet/52.3-4.591.

Shreve, S. E. (2004). *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*. Springer. doi:10.1007/978-0-387-22527-2.

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press. doi:10.1017/CBO9781139344203.

Taylor, S. J. (1982). Financial Returns Modeled by the Product of Two Stochastic Processes—a Study of the Daily Sugar Prices 1961–75. *Time Series Analysis: Theory and Practice*, 1:203–226. doi:10.1016/B978-0-12-437860-6.50018-9.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 22(4):1701–1762. doi:10.1214/aos/1176325750.