



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN
CIENCIAS MATEMÁTICAS**

**IMPLEMENTACIÓN DEL MUESTREO DE GIBBS EN
“R” PARA LA ESTIMACIÓN DE PARÁMETROS DEL
MODELO DE VOLATILIDAD ESTOCÁSTICA CON
APLICACIÓN A DATOS DE CONTAMINACIÓN.**

**TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS**

**PRESENTA:
VERÓNICA DE JESÚS ROMO**

**DIRECTORA DE TESIS:
DRA. ELIANE REGINA RODRIGUES
INSTITUTO DE MATEMÁTICAS**

MÉXICO, D.F. FEBRERO DEL 2013



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

*A mis padres y hermanas.
A David, quien estuvo en todo momento a mi lado.
A ustedes, infinitas gracias. Les amo.*

Índice general

1. Introducción	1
2. Análisis bayesiano	3
2.1. Introducción	3
2.2. Regla de Bayes	3
2.3. Familias conjugadas	7
3. Métodos Monte Carlo vía cadenas de Markov	11
3.1. Introducción	11
3.2. Fundamentos de la simulación Monte Carlo	12
3.3. Cadenas de Markov	14
3.4. Algoritmos Monte Carlo vía cadenas de Markov	21
3.5. El algoritmo Metropolis-Hastings	22
3.6. Muestreo de Gibbs	23
3.7. Metropolis-Hastings dentro del muestreo de Gibbs	24
3.8. Diagnósticos de convergencia	26
3.8.1. Geweke	26
3.8.2. Gelman-Rubin	27
3.8.3. Raftery-Lewis	28
3.8.4. Heidelberger-Welch	29
4. Un modelo de volatilidad estocástica	31
4.1. Introducción	31
4.2. Estructura de datos	32

4.3. Modelo de volatilidad estocástica univariado	33
4.4. Modelo de volatilidad estocástica multivariado.	36
4.5. Análisis bayesiano	39
4.5.1. Función de verosimilitud	40
4.5.2. Densidad conjunta de las variables latentes condicional a los parámetros del modelo	41
4.5.3. Densidades <i>a priori</i> de los parámetros	42
4.5.4. Densidad <i>a posteriori</i> de φ dada \mathbf{D}	44
4.6. Densidades marginales condicionales completas <i>a posteriori</i>	45
4.6.1. Densidades marginales condicionales completas <i>a posteriori</i> de las variables latentes	45
4.6.2. Densidades marginales condicionales completas <i>a posteriori</i> de μ_j, ϕ_j y σ_{η_j}	47
4.7. Pseudocódigo	52
5. Aplicación	55
5.1. Introducción	55
5.2. Estimación de parámetros del modelos VE	55
5.3. Evaluación de la convergencia	61
5.4. Ajuste y pronóstico.	63
6. Conclusiones	73
Apéndices	
A. Resultados de los diagnósticos de convergencia	79
B. Gráficas de las cadenas por región	83
C. Muestreo de Gibbs	87
D. Diagnóstico de convergencia y gráficas de resultados	93

Capítulo 1

Introducción

Dado que los habitantes de las grandes ciudades dañan su salud con altos niveles de contaminación atmosféricos, es importante estudiar el comportamiento de los contaminantes presentes en aire de las grandes ciudades. Se puede estudiar el problema de la contaminación desde distintos puntos de vista. Este trabajo se concentrará en el estudio de la variabilidad de los promedios semanales de un contaminante en particular, específicamente el ozono. El estudio del ozono fue motivado porque es uno de los contaminantes atmosféricos de mayor emisión en las grandes ciudades.

Varios países han implementado normas para regular su emisión. En México, la norma de calidad del aire para el ozono es que una persona no debe estar expuesta a una concentración superior a 0.11 partes por millón (0.11ppm) en promedio por una hora o más (NOM, 2002). En la década de los noventa, los niveles de ozono en la Ciudad de México alcanzaron niveles mayores a 0.30ppm convirtiéndose así en una de las ciudades más contaminadas del mundo.

Con el objetivo de mejorar la calidad del aire en la Ciudad de México, a partir de 1979 se han realizado acciones por parte de las autoridades ambientales para controlar la contaminación del aire. De las varias acciones podemos citar el *Programa Coordinado para el Mejoramiento de la Calidad del Aire en el Valle de México* 1979-1982, el programa *100 Acciones Necesarias para la Ecología* en 1986, el *PICCA* en 1990, el *Programa para Mejorar la Calidad del Aire de la Zona Metropolitana del Valle de México* conocido como *PROAIRE* en 1995 y *PROAIRE* 2002-2010. Las medidas que establece el *PROAIRE* 2002-2010 para disminuir la contaminación atmosférica se dividen en 38 medidas para vehículos y transporte, 7 para la industria, 9 para el servicio, 15 para la conservación de recursos naturales, 8 para la protección de la salud, 4 para la educación ambiental y 8 para el fortalecimiento institucional. También se formuló un programa *PROAIRE* para el periodo 2011-2020 retomando las experiencias de sus anteriores ediciones e introduciendo cambios en la concepción y tratamiento de los procesos de generación de contaminación atmosférica (CAM,

2011).

Ante la práctica y consecución de los programas, surge la necesidad de evaluar resultados. Evidentemente hay que verificar una tendencia a la baja en la concentración de los contaminantes, como, por ejemplo, el ozono. Sin embargo, es de igual relevancia verificar una disminución no sólo en la tendencia sino también en la variabilidad de las mediciones de la concentración de los contaminantes. Por tales motivos, el objetivo de este trabajo es modelar variabilidad de las series temporales del contaminante ozono en cinco regiones de la Zona Metropolitana de la Ciudad de México. Así al ser posible modelar la variabilidad de estas series también será posible predecir posibles contingencias.

Se utilizará una metodología que generalmente se usa para modelar la volatilidad de instrumentos financieros, denominada modelo de volatilidad estocástica (VE). La estimación de los parámetros del modelo se hará usando el muestreo de Gibbs.

Lo novedoso de este trabajo respecto a investigaciones previas sobre el tema (Achcar *et al.* (2008, 2010, 2011) y Zozolotto (2010)) es que la estimación no se hace por medio del software WinBUGS, sino que el algoritmo se programó en lenguaje “R”. Por lo que fue necesario calcular la distribución *a posteriori* y las densidades marginales condicionales completas de cada parámetro para estimarlos mediante el muestreo de Gibbs. La estimación de las variables latentes que el modelo supone, se realizó usando un paso del algoritmo Metropolis-Hastings dentro del algoritmo del muestreo de Gibbs.

El trabajo está organizado de la siguiente forma. En el capítulo 2 se hará una exposición básica sobre las herramientas matemáticas que se usarán, que en el presente trabajo es la estadística bayesiana. En el capítulo 3 se presentan las técnicas de estimación y evaluación utilizadas. En el capítulo 4 se desarrolla el modelo, se calculan los elementos necesarios para la estimación de los parámetros del mismo. También en el capítulo 4 se presenta el pseudocódigo del algoritmo implementado para las estimaciones y diagnósticos de convergencia. En el capítulo 5 se encuentran los resultados y su interpretación. El capítulo 6 está dedicado a las conclusiones del trabajo. Finalmente en un apéndice se presenta el código en “R” del algoritmo programado.

Capítulo 2

Análisis bayesiano

2.1. Introducción

Este trabajo utilizará el enfoque de la estadística bayesiana para realizar la estimación de los parámetros de un modelo. La característica principal de este enfoque es que los parámetros desconocidos se suponen variables aleatorias con alguna distribución *a priori* considerada conocida.

El objetivo de este capítulo es presentar un listado de los resultados que serán utilizados en la parte principal del trabajo. Las demostraciones de los resultados se pueden encontrar en los siguientes libros Carlin y Louis (2001) y a Mood *et al.*(1974).

2.2. Regla de Bayes

Un concepto muy importante para la estadística bayesiana es la probabilidad condicional:

Definición 2.2.1 *Sea un espacio de probabilidad (Ω, \mathcal{F}, P) . Si $A, B \in \mathcal{F}$ y $P(B) > 0$, entonces **la probabilidad condicional** del evento A dado B se define como:*

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

Existe la definición de la probabilidad condicional para variables aleatorias continuas. En este caso no se habla de probabilidades sino de densidades.

Definición 2.2.2 *Si X, Y son variables aleatorias con densidad conjunta $f(x, y)$ y X tiene como densidad a $f(x) > 0$, entonces la densidad condicional de Y dada X , se define y denota como*

$$f_{Y|X=x}(y | x) = \frac{f(x,y)}{f(x)}.$$

De esta definición podemos ver que si X, Y son variables aleatorias independientes, entonces $f_{Y|X=x}(y | x) = f_Y(y)$. También podemos calcular las densidades condicionales para varias variables:

Definición 2.2.3 Sean X_1, X_2, \dots, X_n, Y y Y_1, Y_2, \dots, Y_n variables aleatorias. Definimos sus densidades condicionales multivariadas como:

- $f_{X_1, X_2, \dots, X_n | Y=y}(x_1, x_2, \dots, x_n | y) = \frac{f_{X_1, X_2, \dots, X_n, Y}(x_1, x_2, \dots, x_n, y)}{f_Y(y)}$.
- $f_{X_1, X_2, \dots, X_n | Y_1=y_1, Y_2=y_2, \dots, Y_n=y_n}(x_1, x_2, \dots, x_n | y_1, y_2, \dots, y_n) =$
 $f_{X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n}(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n) / f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n)$.

Suponiendo que las respectivas densidades de la derecha de las igualdades existen.

A continuación se presentan dos resultados fundamentales en el análisis bayesiano.

Proposición 2.2.4 Ley de la probabilidad total.

1. **Para eventos.**

Para todo evento $A \in \mathcal{F}$ y una partición $\{B_1, \dots, B_k\}$ del espacio muestral Ω , (es decir, $\cup_{i=1}^k B_i = \Omega$ y $B_i \cap B_j = \emptyset$) se cumple:

$$P(A) = \sum_{i=1}^k P(A | B_i)P(B_i). \quad (2.1)$$

2. **Para variables aleatorias discretas.**

$$f(x) = \sum_y f_{X|Y=y}(x | Y = y)P(Y = y). \quad (2.2)$$

3. **Para variables aleatorias continuas.**

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y=y}(x | y)f_Y(y)dy. \quad (2.3)$$

Una implicación de la proposición 2.2.4 es el teorema de Bayes (o regla de Bayes) el cual es la base de las aplicaciones en la estadística bayesiana.

Teorema 2.2.5 Teorema de Bayes. Bajo los supuestos de la proposición 2.2.4,

1. *Para eventos.*

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{i=1}^k P(A | B_i)P(B_i)}. \quad (2.4)$$

2. *Para variables aleatorias continuas.*

$$f_{Y|X=x}(y | x) = \frac{f_{X|Y=y}(x | y)f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y=u}(x | u)f_Y(u)du}. \quad (2.5)$$

Observación: El caso para variables aleatorias discretas es similar al caso para variables aleatorias continuas, pero con una suma en lugar de la integral.

Uno de los objetivos de la estadística bayesiana es el siguiente. Teniendo un modelo donde conocemos una muestra de datos $\mathbf{x} = (x_1, x_2, \dots, x_N)$, que sabemos provienen de una distribución que depende de un vector de parámetros desconocidos θ , hacer inferencia sobre θ .

La base del análisis es suponer que θ es una variable aleatoria con distribución conocida llamada *a priori* y se determina una distribución de probabilidad condicionada $\pi(\theta | \mathbf{x})$, denominada densidad *a posteriori*, que permita conocer características del parámetro desconocido tomando en cuenta la información proporcionada por los datos \mathbf{x} .

La importancia del teorema de Bayes radica en que por medio de él podemos calcular la densidad *a posteriori*. A continuación se definen formalmente los elementos necesarios para hacer inferencia bayesiana.

- **Modelo paramétrico:** En general el modelo paramétrico es representado por

$$P(\mathbf{x} | \theta),$$

donde $P(\cdot | \theta)$ es una función de densidad de una variable (o vector) aleatoria(o).

- **Espacio paramétrico:** Es el conjunto de todos los valores que θ puede tomar, y es denotado por Θ .
- **Familia paramétrica de distribuciones.** Es el conjunto de distribuciones

$$\mathcal{P} = \{P(\mathbf{x} | \theta) : \theta \in \Theta\}.$$

Definición 2.2.6 Sea $\mathbf{x} = (x_1, x_2, \dots, x_N)$ una muestra aleatoria y $\theta \in \Theta$ el parámetro desconocido del modelo que describe el experimento que produjo la muestra \mathbf{x} .

La **distribución a priori** de θ es la distribución del parámetro θ que se le asigne sin tener en cuenta la muestra obtenida del experimento y su función de densidad se denotará por $\pi(\theta)$.

Definición 2.2.7 Se dice que una distribución **a priori** es **no informativa** para θ si no favorece ningún valor de θ sobre otros.

La distribución *a priori* es elegida por el investigador del experimento a partir del conocimiento previo que posee del parámetro. La distribución *a priori* que describe un estado de ignorancia o carencia de información del parámetro es la llamada distribución no informativa representada por la distribución uniforme.

Definición 2.2.8 La **función de verosimilitud** $L(\mathbf{x} | \theta)$ es la densidad conjunta de la muestra $\mathbf{x} = (x_1, x_2, \dots, x_N)$ condicionada a θ . Si las observaciones son independientes, esta función está dada por:

$$L(\mathbf{x} | \theta) = f_{X_1, X_2, \dots, X_N | \Theta = \theta}(x_1, x_2, \dots, x_N | \theta) = \prod_{i=1}^N f(x_i | \theta).$$

Como se ha expuesto aquí, la probabilidad condicional será la herramienta básica en todos los procedimientos bayesianos.

Definición 2.2.9 Distribución a posteriori. Sea $\mathbf{x} = (x_1, x_2, \dots, x_N)$ una muestra aleatoria, $\theta \in \Theta$ el parámetro desconocido del modelo que describe el experimento que produjo la muestra \mathbf{x} , $\pi(\theta)$ la distribución a priori de θ y $L(\mathbf{x} | \theta)$ la función de verosimilitud. La **distribución a posteriori** es la distribución del parámetro θ cuando se toma en cuenta la información de la muestra obtenida y su función de densidad se denotará por $\pi(\theta | \mathbf{x})$.

Observación: Notemos que:

$$\begin{aligned} \pi(\theta | \mathbf{x}) &= f_{\theta | X_1=x_1, X_2=x_2, \dots, X_N=x_N}(\theta | x_1, x_2, \dots, x_N) \\ &= \frac{f_{X_1, X_2, \dots, X_N | \theta}(x_1, x_2, \dots, x_N | \theta)\pi(\theta)}{f_{X_1, X_2, \dots, X_N}(x_1, x_2, \dots, x_N)} \\ &= \frac{\left(\prod_{i=1}^N f(x_i | \theta) \right) \pi(\theta)}{\int_{\Theta} \left(\prod_{i=1}^N f(x_i | \theta') \right) \pi(\theta') d\theta'} \\ &= \frac{L(\mathbf{x} | \theta)\pi(\theta)}{\int_{\Theta} L(\mathbf{x} | \theta')\pi(\theta') d\theta'}. \end{aligned}$$

Hay que notar que $f_{X_1, \dots, X_N}(x_1, x_2, \dots, x_N)$, también conocida como la constante normalizadora, es constante respecto a θ . Por tanto podemos expresar a la densidad *a posteriori* como proporcional a la función de verosimilitud por la densidad *a priori*, es decir:

$$\pi(\theta | \mathbf{x}) \propto L(\mathbf{x} | \theta)\pi(\theta). \quad (2.6)$$

2.3. Familias conjugadas

Observemos que tanto $\pi(\theta)$ como $\pi(\theta | \mathbf{x})$ son distribuciones de probabilidad de la variable aleatoria θ . Es conveniente que ambas distribuciones pertenezcan a la misma familia porque se facilitan los cálculos. La siguiente definición de familia conjugada se puede encontrar en el libro de James O. Berger (1985).

Definición 2.3.1 Sea \mathcal{F} la clase de funciones de densidad $f(x | \theta)$ parametrizada por θ . Una clase \mathcal{P} de distribuciones a priori se dice que es una familia conjugada para \mathcal{F} si $\pi(\theta | \mathbf{x}) = f(x | \theta)\pi(\theta)$ está en la clase \mathcal{P} para toda $f \in \mathcal{F}$ y $\pi \in \mathcal{P}$.

A continuación se presentan dos ejemplos que ilustran la definición.

Ejemplo 2.3.2 Supongamos que $X | \Theta = \theta$ se distribuye $N(\theta, \sigma^2)$, donde σ^2 es conocida y Θ se distribuye a priori $N(\mu, \tau^2)$ con parámetros conocidos. Entonces veremos que la distribución normal es una familia conjugada para si misma como lo muestra el siguiente cálculo.

$$\pi(\theta | x) = \frac{f(x|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)d\theta}.$$

Por una parte,

$$\begin{aligned} f(x | \theta)\pi(\theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right] \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{1}{2}\left(\frac{\theta-\mu}{\tau}\right)^2\right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left[-\frac{1}{2\sigma^2}(x^2 - 2x\theta + \theta^2)\right] \exp\left[-\frac{1}{2\tau^2}(\theta^2 - 2\theta\mu + \mu^2)\right] \\ &= \frac{1}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2}\right] \exp\left[-\frac{1}{2\sigma^2}(-2x\theta + \theta^2) - \frac{1}{2\tau^2}(\theta^2 - 2\theta\mu)\right] \\ &= \frac{1}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2}\right] \exp\left[-\frac{1}{2}\left(\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\theta^2 - 2(x + \mu)\theta\right)\right]. \end{aligned}$$

Sean $A = \frac{1}{\sigma^2} + \frac{1}{\tau^2}$ y $B = x + \mu$, entonces nos queda

$$\begin{aligned}
f(x | \theta)\pi(\theta) &= \frac{1}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2}\right] \exp\left[-\frac{1}{2}(A\theta^2 - 2B\theta)\right] \\
&= \frac{1}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2}\right] \exp\left[-\frac{A}{2}\left(\theta^2 - 2\frac{B}{A}\theta + \frac{B^2}{A^2} - \frac{B^2}{A^2}\right)\right] \\
&= \frac{1}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} + \frac{B^2}{2A}\right] \exp\left[-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2\right] \\
&= \frac{\sqrt{2\pi A^{-1}}}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} + \frac{B^2}{2A}\right] \frac{1}{\sqrt{2\pi A^{-1}}} \exp\left[-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2\right].
\end{aligned}$$

Notar que los primeros dos factores no dependen de θ y los dos últimos son la densidad de una $N\left(\frac{B}{A}, \frac{1}{A}\right)$. Por lo tanto,

$$\int_{-\infty}^{\infty} f(x | \theta)\pi(\theta)d\theta = \frac{\sqrt{2\pi A^{-1}}}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} + \frac{B^2}{2A}\right].$$

Finalmente sustituyendo en la distribución a posteriori:

$$\begin{aligned}
\pi(\theta | x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)d\theta} \\
&= \frac{\frac{\sqrt{2\pi A^{-1}}}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} + \frac{B^2}{2A}\right] \frac{1}{\sqrt{2\pi A^{-1}}} \exp\left[-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2\right]}{\frac{\sqrt{2\pi A^{-1}}}{\sqrt{4\pi^2\sigma^2\tau^2}} \exp\left[-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} + \frac{B^2}{2A}\right]} \\
&= \frac{1}{\sqrt{2\pi A^{-1}}} \exp\left[-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2\right].
\end{aligned}$$

Por lo tanto la distribución a posteriori pertenece a la familia normal.

Ejemplo 2.3.3 Supongamos que $X | \Theta = \theta$ se distribuye $Poisson(\theta)$ y Θ se distribuye a priori $Gamma(\alpha, \beta)$ con parámetros conocidos. Entonces veremos que la distribución gamma es una familia conjugada para la Poisson como lo muestra el siguiente cálculo.

$$\pi(\theta | x) = \frac{f(x|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)d\theta}.$$

Por una parte,

$$\begin{aligned}
f(x | \theta)\pi(\theta) &= e^{-\theta} \frac{\theta^x}{x!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\
&= \frac{1}{x!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^x e^{-\theta} \theta^{\alpha-1} e^{-\beta\theta} \\
&= \frac{\beta^\alpha}{x\Gamma(\alpha)} \theta^{x+\alpha-1} e^{-(1+\beta)\theta}.
\end{aligned}$$

Sean $C = x + \alpha$ y $D = 1 + \beta$, entonces nos queda

$$\begin{aligned} f(x | \theta)\pi(\theta) &= \frac{\beta^\alpha}{x!\Gamma(\alpha)}\theta^{C-1}e^{-D\theta} \\ &= \frac{\Gamma(C)\beta^\alpha}{D^C x!\Gamma(\alpha)}\frac{D^C}{\Gamma(C)}\theta^{C-1}e^{-D\theta}. \end{aligned}$$

Notar que el primer factor no dependen de θ y el resto es la densidad de una $\text{Gamma}(C, D)$. Por lo tanto,

$$\int_0^\infty f(x | \theta)\pi(\theta)d\theta = \frac{\Gamma(C)\beta^\alpha}{D^C x!\Gamma(\alpha)}.$$

Finalmente sustituyendo en la distribución a posteriori:

$$\begin{aligned} \pi(\theta | x) &= \frac{\frac{\Gamma(C)\beta^\alpha}{D^C x!\Gamma(\alpha)}\frac{D^C}{\Gamma(C)}\theta^{C-1}e^{-D\theta}}{\frac{\Gamma(C)\beta^\alpha}{D^C x!\Gamma(\alpha)}} \\ &= \frac{D^C}{\Gamma(C)}\theta^{C-1}e^{-D\theta}. \end{aligned}$$

Por lo tanto la distribución a posteriori pertenece a la familia Gamma .

En este capítulo se vio que la distribución *a posteriori* contiene toda la información de la muestra y de la densidad *a priori*, así que podemos proceder maximizándola para estimar a θ . Debido a que la distribución *a posteriori* es función de θ también podemos estimar a θ a través de la media o la mediana de esta distribución.

En este trabajo nos interesará sobre todo estimar cada parámetro del modelo usando la media de la distribución *a posteriori* y calculando sus intervalos de confianza. Como en el caso de una distribución *a posteriori* multivariada el cálculo de tales estimadores no es fácil, en el siguiente capítulo se exponen las técnica que se usarán para tales cálculos.

Capítulo 3

Métodos Monte Carlo vía cadenas de Markov

3.1. Introducción

Como hemos mencionado en el capítulo anterior, una forma de estimar los parámetros de un modelo es calcular la media a partir de la distribución *a posteriori* de los parámetros. Sin embargo las expresiones de las densidades *a posteriori* pueden llegar a ser muy complicadas. Por ejemplo, si intentamos encontrar una densidad marginal, necesitamos integrar con respecto a las otras variables y no siempre es posible por métodos analíticos encontrar esta integral.

En este capítulo se expondrán los métodos Monte Carlo vía cadenas de Markov, que proveen métodos para muestrear valores de distribuciones complicadas o multivariadas, como suele ser en el caso de las distribuciones *a posteriori* de la estadística bayesiana.

La idea es generar una cadena de Markov cuya distribución estacionaria es una distribución objetivo (por ejemplo la distribución *a posteriori*) y la simulación vía métodos Monte Carlo es utilizada para aproximar las integrales que comúnmente surgen. El uso de los Métodos Monte Carlo vía cadenas de Markov en la estadística bayesiana puede consultarse en Albert (2009), Congdon (2003, 2010), Geweke (2005), Marin y Robert (2007), Ntzoufras (2009) y Robert y Casella (2010), entre muchos otros.

A continuación se exponen los fundamentos teóricos de los métodos Monte Carlo vía cadenas de Markov y los criterios de convergencia que se aplicarán en este trabajo. Una exposición más amplia puede consultarse en Gilks. *et al.* (1996), Robert y Casella (2004, 2010), Roberts y Smith (1994), Smith y Roberts (1993) y Tierney (1994).

3.2. Fundamentos de la simulación Monte Carlo

1. Ley de los Grandes Números

La llamada *Ley de los Grandes Números* afirma que al obtener un promedio de variables aleatorias que cumplen con ciertas características, éste tiende a un número fijo conforme el número de variables aleatorias crezca. A continuación se enuncian estas características.

Definición 3.2.1 Diremos que X_1, X_2, \dots, X_N es una **muestra aleatoria** para referirnos a que X_1, X_2, \dots, X_N es una sucesión de variables aleatorias independientes e idénticamente distribuidas. El número de variables "N" se llama el tamaño de la muestra.

Las muestras aleatorias en la vida real, se construyen a partir de datos capturados periódicamente. Con métodos estadísticos se determina con cierta confianza si la muestra aleatoria proviene de alguna distribución. La Ley de los Grandes Números establece que con probabilidad uno, el promedio de una muestra aleatoria de gran tamaño tiende a ser la esperanza común de las variables aleatorias que generaron la muestra.

Definición 3.2.2 Sea X_1, X_2, \dots, X_N una muestra aleatoria. Se define la **media muestral** como:

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i.$$

Suponiendo que $E(X_i) = \mu$ y $\text{Var}(X_i) = \sigma^2$ para cualquier i , la nueva variable aleatoria \bar{X}_N tiene las siguientes propiedades:

- $E(\bar{X}_N) = \mu$,
- $\text{Var}(\bar{X}_N) = \frac{\sigma^2}{N}$.

Teorema 3.2.3 Ley Fuerte de los grandes números. Sea X_1, X_2, \dots, X_N una muestra aleatoria generada por un sucesión de variables aleatorias, cada una con esperanza $\mu < \infty$, entonces:

$$P\left(\lim_{N \rightarrow \infty} \bar{X}_N = \mu\right) = 1.$$

Demostración: Ver, por ejemplo, Feller (1968).

La versión de la Ley de los Grandes Números que presentamos es la más usada. Una variante con las mismas hipótesis, pero un resultado más amplio es el siguiente:

Proposición 3.2.4 *Sea X_1, X_2, \dots, X_N una muestra aleatoria, cada una con esperanza $\mu < \infty$, g una función continua, entonces:*

$$P\left(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N g(X_i) = E(g(X_1))\right) = 1.$$

El resultado anterior es el fundamento de la simulación Monte Carlo.

2. Integración Monte Carlo

a) Para aproximar a través del método de Monte Carlo el valor de $\int_a^b g(x)dx$, se puede usar la siguiente técnica:

- Si debemos encontrar $\int_a^b g(x)dx$, hacemos el cambio de variable

$$x = (b - a)u + a \text{ y } dx = (b - a)du,$$

para $0 < u < 1$ y llegamos a la integral equivalente:

$$\int_a^b g(x)dx = (b - a) \int_0^1 g((b - a)u + a)du.$$

Así, tenemos que generar una muestra aleatoria, u_1, \dots, u_N a partir de la distribución $U(0, 1)$ y usar la Ley de los Grandes Números para aproximar:

$$(b - a) \int_0^1 g((b - a)u + a)du \approx (b - a)E[g((b - a)u + a)],$$

dando como resultado que:

$$\int_a^b g(x)dx \approx (b - a) \frac{1}{N} \sum_{i=1}^N g((b - a)u_i + a).$$

b) Para aproximar a través del método de Monte Carlo el valor de $\int_0^\infty g(x)f(x)dx$, se puede usar la siguiente técnica:

- Si debemos encontrar $\int_0^\infty g(x)f(x)dx$ y f es una densidad conocida con soporte $(0, \infty)$, entonces:

$$\int_0^\infty g(x)f(x)dx = E(g(x)).$$

Así, tenemos que generar una muestra a partir de la función de densidad f y usar la Ley de los Grandes Números para aproximar esa esperanza.

- Si debemos encontrar $\int_0^\infty g(x)f(x)dx$ y f no es una función de densidad conocida, podemos usar una función de densidad conocida, por ejemplo $h > 0$ en $(0, \infty)$ y utilizar el hecho que

$$\begin{aligned} \int_0^\infty g(x)f(x)dx &= \int_0^\infty g(x)f(x)\frac{h(x)}{h(x)}dx \\ &= \int_0^\infty \frac{g(x)f(x)}{h(x)}h(x)dx \\ &= E\left(\frac{f(x)g(x)}{h(x)}\right). \end{aligned}$$

Así, tenemos que generar una muestra de valores a través de la función de densidad común h y usar la Ley de los Grandes Números para aproximar esa esperanza.

A continuación se presentan las herramientas básicas para el estudio de las cadenas de Markov.

3.3. Cadenas de Markov

La aplicación de cadenas de Markov es muy amplia. Este tipo de procesos pueden ser discretos o continuos y modelan situaciones donde el comportamiento de un fenómeno futuro depende solamente del pasado más próximo.¹Una cadena de Markov es una sucesión de variables aleatorias definidas como sigue:

Definición 3.3.1 *Sea $\{X_n\}_{n=0}^\infty$ una sucesión de variables aleatorias con soporte común $S \subset \mathbb{R}$. Diremos que $\{X_n\}_{n=0}^\infty$ es una cadena de Markov si para cualquier t , la distribución condicional de X_{t+1} dado X_n con $n = 0, 1, \dots, t$, es la misma que la distribución condicional de X_{t+1} dado X_t . Al conjunto S se llama espacio de estados y a la densidad de X_{t+1} dado X_t se le llama kernel de transición de la cadena y se denota como $K(X_{t+1} | X_t)$, donde $K(\cdot | x)$ es una medida de probabilidad.*

Caso con espacio de estados discreto. Si $S \subset \mathbb{N}$, el kernel, también llamado probabilidad de transición, es representado por $P_{i,j}$, $i, j \in S$ y es tal que la entrada i, j está dada por:

$$P_{i,j} = P(X_n = j | X_{n-1} = i), i, j \in S, n \geq 1. \quad (3.1)$$

Una definición equivalente a la 3.3.1 es:

¹La teoría que se expone en esta sección está basada en Robert y Casella (2004), excepto en los casos que se indique otra referencia.

Definición 3.3.2 Una sucesión $\{X_n\}_{n=0}^{\infty}$ de variables aleatorias con kernel de transición $K(\cdot | \cdot)$ es una **cadena de Markov** si para cualquier $n > 0$ y $A \subseteq S$,

$$\begin{aligned} P(X_n \in A | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) &= P(X_n \in A | X_{n-1} = x_{n-1}) \\ &= \int_A K(x | x_{n-1}) dx. \end{aligned}$$

Definición 3.3.3 Una cadena de Markov con espacio de estados discreto es **irreducible** si todos los estados son comunicantes, o sea, si

$$P(\tau_y < \infty | X_0 = x) > 0, \text{ para todo } x, y \in S, \quad (3.2)$$

donde $\tau_y = \inf\{n \geq 1 : X_n = y\}$, es el primer momento que la cadena visita y .

Definición 3.3.4 Dada una medida ψ , una cadena de Markov $\{X_n\}_n^{\infty}$ con espacio de estados no numerable es ψ -**irreducible**, si para todo $A \subset S$ con $\psi(A) > 0$, existe algún n tal que $K^n(A | x) > 0$ para todo $x \in S$.

Definición 3.3.5 La cadena de Markov $\{X_n\}_n^{\infty}$ tiene un átomo $\alpha \subset S$ si existe una medida no cero ν tal que

$$K^n(A | x) = \nu(A) \text{ para todo } x \in \alpha, \text{ para todo } A \subset S.$$

Si $\{X_n\}_n^{\infty}$ es ψ -irreducible, el átomo es accesible cuando $\psi(\alpha) > 0$.

La definición de cadena irreducible es muy importante, ya que ayudará a garantizar la convergencia de la cadena de Markov a un estado de equilibrio.

Definición 3.3.6 El **periodo** de un estado $w \in S$ con S discreto, se define como:

$$d(w) = m.c.d.\{m \geq 1 : K^m(w | w) > 0\}, \quad (3.3)$$

donde *m.c.d.* indica el máximo común divisor y donde para $\{X_n\}_{n=0}^{\infty}$ la cadena de Markov con kernel $K(\cdot | \cdot)$, se tiene que $K^m(y | x) = P(X_m = y | X_0 = x)$. Se puede demostrar que si la cadena es irreducible, todos los estados tienen el mismo periodo. Si éste es uno, diremos que la cadena es **aperiódica**.

Definición 3.3.7 En un espacio de estados finito (y de forma similar para el caso continuo), denotamos como el número de visitas al estado w como:

$$\eta_w = \sum_{i=1}^{\infty} 1_{\{w\}}(X_i),$$

donde $1_{\{x\}}(X) = 1$ si $X = x$ y es cero en otro caso.

Si $E(\eta_w \mid X_0 = w) = \infty$, entonces el estado w es llamado **recurrente**, si $E(\eta_w \mid X_0 = w) < \infty$ el estado w es llamado transitorio. Además, un estado recurrente se dice **recurrente positivo** si el tiempo medio de retorno es finito, en otro caso se dice recurrente nulo.

Teorema 3.3.8 Una cadena ψ -irreducible o bien es recurrente o transitoria.

Demostración. Ver Robert, y Casella (2004).

Definición 3.3.9 Un conjunto A es **recurrente de Harris** in $P(\eta_A = \infty) = 1$ para todo $x \in A$. La cadena $\{X_n\}_n^\infty$ es **recurrente de Harris** si existe una medida ψ tal que $\{X_n\}_n^\infty$ es ψ -irreducible y para todo conjunto A con $\psi(A) > 0$, A es recurrente de Harris.

La propiedad de la recurrencia de Harris sólo en el caso cuando el espacio de estados S es no numerable.

Medidas Estacionarias. La estabilidad de una cadena de Markov se incrementa conforme la distribución marginal de X_n se vuelva independiente de n y del valor inicial. Formalmente, se tiene la siguiente definición.

Definición 3.3.10 Una medida p es **estacionaria o invariante** para el kernel de transición $K(\cdot \mid \cdot)$ si para todo $B \subseteq S$

$$p(B) = \int_S K(B \mid x)p(dx). \quad (3.4)$$

En particular en el caso discreto $p(i)$ con $i \in S$, es una función de probabilidad que cumple:

$$p(j) = \sum_{i \in S} p(i)P_{i,j}, \text{ para todo } j \in S.$$

Por el teorema 3.3.8 cuando existe una medida de probabilidad invariante para una cadena ψ -irreducible, la cadena es positiva.

Teorema 3.3.11 Si una cadena de Markov $\{X_n\}_{n=0}^\infty$ es irreducible, recurrente positiva y aperiódica entonces tiene una distribución estacionaria única y coincide con la distribución límite de la cadena.

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

En otras palabras, si una cadena de Markov tiene una distribución estacionaria $p(\cdot)$, entonces a partir de alguna “ n_0 ” suficientemente grande, X_n es aproximadamente distribuida como $p(\cdot)$ para toda $n > n_0$.

Definición 3.3.12 Una cadena de Markov $\{X_n\}_{n=0}^\infty$ irreducible con distribución estacionaria $p(\cdot)$ es **recurrente de Harris** si para todo A subconjunto del espacio de estados y todo x en el espacio de estados,

$$P(\tau_A < \infty \mid X_0 = x) = 1, \quad (3.5)$$

donde τ_A es el tiempo del primer arribo al conjunto A .²

La siguiente exposición está basada en el libro de Robert y Casella (2004) y trata sobre el comportamiento de las cadenas de Markov a largo plazo.

Teorema 3.3.13 Sea $\{X_n\}_n^\infty$ una cadena de Markov ψ -irreducible con un átomo α . La cadena es positiva si y sólo si $E[\tau_\alpha] < \infty$. En este caso, la distribución invariante $p(\cdot)$ para $\{X_n\}_n^\infty$ satisface

$$p(\alpha) = (E[\tau_\alpha])^{-1}.$$

Demostración. Ver Robert y Casella (2004).

Los siguientes teoremas son la base fundamental de la convergencia de los algoritmos Monte Carlo vía cadenas de Markov. La demostración de estos teoremas requiere aspectos teóricos que están fuera del alcance de este trabajo. La demostración del caso general puede consultarse en Meyn y Tweedie (2009) y en Athreya *et al.* (1996).

Definición 3.3.14 Para una cadena $\{X_n\}_n^\infty$ recurrente positiva de Harris, con distribución invariante p , un átomo es **ergódico** si:

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

Para establecer la existencia de un átomo ergódico, en el caso de espacio de estados numerable es suficiente la convergencia conforme a la norma de variación total (TV) entre dos medidas, es decir,

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

²Roberts y Rosenthal (2006).

Proposición 3.3.15 Si $\{X_n\}_n^\infty$ es recurrente positiva sobre el espacio de estados numerable e infinito S y si existe un átomo ergódico $\alpha \subset S$, entonces, para todo $x \in S$,

$$\lim_{n \rightarrow \infty} \|K^n(\cdot | x) - p\|_{TV} = 0.$$

Demostración. Ver Robert y Casella (2004).

Teorema 3.3.16 Para una cadena $\{X_n\}_n^\infty$ recurrente positiva y aperiódica sobre un espacio de estados numerable S , para todo estado inicial x se cumple,

$$\lim_{n \rightarrow \infty} \|K^n(\cdot | x) - p\|_{TV} = 0. \quad (3.6)$$

La demostración de Robert y Casella (2004), utiliza la técnica de acople para el caso general de espacio de estados. En este caso el acoplamiento entre dos cadenas, $\{X_n\}_n^\infty$ y $\{X'_n\}_n^\infty$, asociadas con el mismo kernel se da cuando se encuentran en α .

Esbozo de la demostración. Como $\{X_n\}_n^\infty$ es recurrente positiva, $E[\tau_\alpha]$ es finita por el Teorema 3.3.13. El resultado se sigue de la Proposición 3.3.15. Por lo tanto m_p , tiempo medio de excursión (ver lema 6.49 de Robert y Casella (2004)) es finito. Además se cumple que todo átomo es ergódico y por la proposición 3.3.15 se da (3.6).

A continuación se presenta la demostración para el caso de espacio de estados numerable.

Demostración: El método de esta demostración se conoce como técnica de acople. Sea Y_n una cadena de Markov independiente de la original X_n pero con la misma matriz de probabilidades de transición. Entonces el proceso $Z_n = (X_n, Y_n)$ es una cadena de Markov con probabilidades de transición

$$P(Z_{n+1} = (x_{n+1}, y_{n+1}) | Z_n = (x_n, y_n)) = p_{x_n, x_{n+1}} p_{y_n, y_{n+1}},$$

y tiene distribución estacionaria $\pi_{x_n, y_n} = \pi_{x_n} \pi_{y_n}$. Puede comprobarse que la cadena Z es recurrente positiva. Además es irreducible pues como X y Y son aperiódicas, existe un número natural N tal que $p_{i,j}(n) p_{k,l}(n) > 0$, para toda $n > N$.

Sea j cualquier estado. Definamos el primer momento en el que la cadena Z visita el estado (j, j) como $\tau_j = \min\{n \geq 1 : Z_n = (j, j)\}$. Sea además $\tau = \min\{n \geq 1 : X_n = Y_n\}$. Este es el momento de acople de las dos cadenas. Como Z es recurrente, $P(\tau < \infty) = 1$. Además $\tau \leq \tau_j$. Por la propiedad de Markov

$$\begin{aligned}
P(X_n = x, \tau \leq n) &= \sum_j \sum_{r=1}^n P(X_n = x, X_r = j, \tau = r) \\
&= \sum_j \sum_{r=1}^n P(X_n = x \mid X_r = j, \tau = r) P(X_r = j, \tau = r) \\
&= \sum_j \sum_{r=1}^n P(Y_n = x \mid Y_r = j, \tau = r) P(Y_r = j, \tau = r) \\
&= \sum_j \sum_{r=1}^n P(Y_n = x \mid Y_r = j) P(Y_r = j, \tau = r) \\
&= P(Y_n = x, \tau \leq n),
\end{aligned}$$

es decir, sobre el evento $(\tau \leq n)$, las variables X_n y Y_n tienen la misma distribución de probabilidad. Por otro lado

$$\begin{aligned}
P(X_n = j) &= P(X_n = j, \tau \leq n) + P(X_n = j, \tau > n) \\
&= P(Y_n = j, \tau \leq n) + P(X_n = j, \tau > n) \\
&\leq P(Y_n = j) + P(\tau > n).
\end{aligned}$$

De manera análoga, $P(Y_n = j) \leq P(X_n = j) + P(\tau > n)$. Por lo tanto,

$$|P(X_n = j) - P(Y_n = j)| \leq P(\tau > n) \rightarrow 0, \quad (3.7)$$

cuando $n \rightarrow \infty$. Si ahora se toma $X_0 = i$ y Y_0 con la distribución estacionaria π , entonces $P(X_n = j) = P(X_0 = i)p_{ij}(n) = p_{ij}(n)$, y por lo tanto (3.7) establece que $|p_{ij}(n) - \pi_j| \rightarrow 0$. ■

El caso de **espacio de estados general** requiere que la cadena sea **recurrente de Harris** para derivar la convergencia de K^n a p .

Ahora estudiaremos la versión de la **Ley de los grandes números para cadenas de Markov**. Sean X_1, X_2, \dots, X_n observaciones de una cadena de Markov, nos interesa estudiar el comportamiento de las sumas parciales

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

cuando n tiende a infinito.

Definición 3.3.17 Una función medible h es armónica para la cadena $\{X_n\}_n^\infty$ si

$$E(h(X_{n+1}) \mid x_n) = h(x_n).$$

Estas funciones son invariantes para el kernel de transición y caracterizan la recurrencia de Harris como sigue.

Proposición 3.3.18 *Si las funciones constantes son las únicas funciones armónicas acotadas para una cadena de Markov positiva, entonces la cadena es recurrente de Harris.*

Demostración. Ver Robert y Casella (2004).

La proposición anterior es más práctica para establecer la recurrencia de Harris de algunos algoritmos Monte Carlo vía cadenas de Markov. El regreso de la proposición también se cumple y se puede encontrar la demostración en el libro de Meyn y Tweedie (1993).

Lema 3.3.19 *Para las cadenas de Markov con recurrencia de Harris, las funciones constantes son las únicas funciones armónicas acotadas.*

Una consecuencia del lema es que si la cadena $\{X_n\}_n^\infty$ es Harris positiva con distribución estacionaria p y si $S_n(h)$ converge casi seguramente a

$$\int_S h(x)p(dx),$$

para una distribución inicial μ_0 , esta convergencia ocurre para toda distribución inicial μ .

El resultado más importante para justificar los algoritmos Monte Carlo vía cadenas de Markov es la llamada ley de los grandes números para cadenas de Markov (también conocido como teorema ergódico), que garantiza la convergencia de $S_n(h)$.

Teorema 3.3.20 *Teorema Ergódico. Si $\{X_n\}_n^\infty$ tiene una medida invariante σ -finita p , los siguientes enunciados son equivalentes:*

1. Si f y $g \in L^1(\pi)$ con $\int g(x)dp(x) \neq 0$, entonces

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

2. La cadena de Markov $\{X_n\}_n^\infty$ es recurrente de Harris.

Demostración: Si 1. se cumple, sea $f(x) = 1_A(x)$ la función indicadora para un conjunto A con medida finita y g una función arbitraria con integral finita y positiva. Si $p(A) > 0$,

$$P_x(X \in A \text{ una infinidad de veces}) = 1$$

para todo $x \in S$, lo cual establece la recurrencia de Harris. El regreso es una aplicación de la ley de los grandes números para variables aleatorias independientes e idénticamente distribuidas y se puede consultar en Robert y Casella (2004).

La validez de los algoritmos Monte Carlo vía cadenas de Markov que se presentan en la siguiente sección, están fundamentados en el teorema 3.3.20, puesto que depende de que la cadena construida converja a su distribución estacionaria.

3.4. Algoritmos Monte Carlo vía cadenas de Markov

La idea de los algoritmos Monte Carlo vía cadenas de Markov es crear una cadena de Markov que tenga distribución estacionaria aproximada a una función de densidad de interés. Para ser más precisos, supongamos que deseamos generar una muestra a partir de la función de densidad $\pi(\theta | y)$ para $\theta \in \Theta$ pero que esto no se puede hacer directamente. Supongamos que podemos construir una cadena de Markov con espacio de estados Θ , la cual sea fácil de simular y tenga como distribución estacionaria a $\pi(\theta | y)$. Si simulamos una trayectoria de la cadena lo suficientemente grande, los valores simulados pueden usarse para resumir algunas propiedades de $\pi(\theta | y)$.

Lo que se quiere es utilizar algoritmos Monte Carlo vía cadenas de Markov que produzcan cadenas de Markov ergódicas con la distribución estacionaria de interés. Es necesario hacer las siguientes observaciones:

- Estamos interesados en el caso cuando la distribución estacionaria de la cadena construida corresponde a la densidad *a posteriori* $\pi(\theta | y)$.
- La propiedad ergódica significa que la distribución de θ_t converge a la distribución límite $\pi(\theta | y)$ para cualquier valor inicial θ_0 . Por lo tanto para t grande, θ_t proviene de la distribución $\pi(\theta | y)$ sin importar el punto inicial. Un problema es decidir qué tan grande tomar t , pero esto puede ser establecido haciendo pruebas de hipótesis sobre la estacionariedad de la cadena (véase sección 3.8).
- Las cadenas de Markov obtenidas por los métodos Monte Carlo vía cadenas de Markov son casi siempre homogéneas en el tiempo, o sea, la distribución de $(\theta_{t_0+1}, \theta_{t_0+2}, \dots, \theta_{t_0+k} | \theta_{t_0})$ es igual que la distribución de $(\theta_1, \theta_2, \dots, \theta_k | \theta_0)$ para cualquier $t_0 \geq 0$ y $k > 0$.
- Otra importante propiedad de estabilidad es la llamada reversibilidad, que significa que la dirección del tiempo de la cadena no importa. La cadena es reversible con respecto a $\pi(\theta | y)$ si el kernel de transición satisface la condición de balance:

$$K(\theta | \theta')\pi(\theta | y) = K(\theta' | \theta)\pi(\theta' | y).$$

La condición de balance es una condición suficiente (no necesaria) para que $\pi(\theta | y)$ sea una densidad estacionaria asociada al kernel de transición.

De acuerdo a lo anterior, los valores generados $\theta_1, \theta_2, \dots$ no son independientes. Sin embargo, la independencia requerida para calcular algunas características de $\pi(\theta | y)$, es obtenida al tomar valores generados en iteraciones suficientemente distantes, para componer la muestra usada en el Teorema Ergódico.

3.5. El algoritmo Metropolis-Hastings

Desde los noventa, los métodos Monte Carlo vía cadenas de Markov se han usado cada vez más frecuentemente para obtener información respecto a distribuciones complejas multivariadas. El algoritmo Metropolis-Hastings es un método Monte Carlo vía cadenas de Markov. Fue desarrollado por Metropolis *et al.* (1953) en el contexto de la física estadística. En 1970 fue generalizado por Hastings (1970) en el contexto puramente estadístico. El algoritmo se describe a continuación.

Dada una densidad $\pi(\theta | y)$, normalizada o no y una densidad condicional $q(\cdot | \theta)$, el algoritmo consiste en generar una cadena $\{\theta_1, \theta_2, \dots\}$ usando los siguientes pasos:

1. Inicializar la cadena en θ_0 , cualquier valor dentro del soporte de $\pi(\theta | y)$.
2. Para $t = 1, 2, \dots, T$
 - a) Generar un valor θ^* que provenga de $q(\cdot | \theta_t) > 0$.
 - b) Aceptar el valor θ^* generado con probabilidad

$$\rho(\theta_t, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^* | y)q(\theta_t | \theta^*)}{\pi(\theta_t | y)q(\theta^* | \theta_t)} \right\} \quad (3.8)$$

o sea, hay que simular un valor u a partir de una $U(0,1)$ y establecer $\theta_{t+1} = \theta^*$ si $u < \rho(\theta_t, \theta^*)$ y en caso contrario hacer $\theta_{t+1} = \theta_t$ y regresar a a).

Observaciones:

- A la densidad $\pi(\theta | y)$ se le llama la densidad objetivo.
- A la función $q(\cdot | \theta)$ se le llama densidad propuesta.

Una exposición más amplia del algoritmo Metropolis- Hastings puede consultarse en Chib y Greenberg (1995).

3.6. Muestreo de Gibbs

El muestreo de Gibbs es un algoritmo utilizado para generar una muestra aleatoria a partir de una distribución, sin tener que utilizar directamente la función de densidad de ésta. El método comenzó por el trabajo de Geman y Geman (1984). Sus raíces se remontan a los 50's desde el trabajo de Metropolis *et al.* (1953), actualmente se puede encontrar abundante literatura acerca del tema, tal es el caso de Casella y George (1992), Smith *et al.* (1993), Roberts y Smith (1994), Carter *et al.* (1994), Robert y Casella (2004).

Para ilustrar la idea del muestreo de Gibbs, consideremos el caso de dos variables aleatorias X y Y con densidad conjunta $f_{X,Y}(x, y)$. Supongamos que la simulación de y de $f_Y(\cdot)$ no puede hacerse directamente pero podemos fácilmente generar una muestra de $f_{Y|x}(\cdot | X)$. De forma similar podemos generar x de $f_{X|Y}(\cdot | y)$. Entonces, el muestreo de Gibbs consiste en generar la muestra como sigue:

1. Iniciar con un valor arbitrario x_0 en el soporte de X .
2. Para $t = 1, 2, \dots, T$:
 - a) simular y_t de $f_{Y|x}(\cdot | x_{t-1})$,
 - b) simular x_t de $f_{X|Y}(\cdot | y_t)$.

Incrementar t y regresar al paso a).

En Roberts y Smith (1994) se establecen condiciones necesarias para asegurar que las cadenas generadas mediante el muestreo de Gibbs cumplan el teorema 3.3.20, es decir sobre las cuales $f_{X,Y}(x, y)$ es una distribución estacionaria de la cadena $\{(x_t, y_t), t = 1, 2, \dots\}$; que es ergódica con una distribución límite $f(x, y)$, y también que la distribución de y_t converge a $f_Y(y)$ para t grande.

La generalización del muestreo de Gibbs a un caso multidimensional es como sigue. Consideremos un vector aleatorio $\bar{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ con distribución conjunta $\pi(\bar{\theta})$. Sea $\bar{\theta}_{-i}$ el vector $\bar{\theta}$ sin la i -ésima coordenada, es decir, $\bar{\theta}_{-1} = (\theta_2, \theta_3, \dots, \theta_d)$, $\bar{\theta}_{-d} = (\theta_1, \theta_2, \dots, \theta_{d-1})$ y para $i = 2, \dots, d-1$, $\bar{\theta}_{-i} = (\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \theta_{i+2}, \dots, \theta_d)$. Denotemos a las **densidades marginales condicionales completas** por $\pi_i(\theta_i | \bar{\theta}_{-i})$, para $i = 1, \dots, d$. Así, teniendo bien definidas las densidades marginales condicionales completas y la manera en cómo se pueden muestrear valores a partir de ellas, hay que seguir los siguientes pasos:

1. Inicializar $\theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_d^{(0)}$ con valores arbitrarios dentro del soporte de $\pi_i(\cdot)$.

2. Para $t = 0, 1, \dots, T$, sea $\bar{\theta}_t = (\theta_1^{(t)}, \theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_d^{(t)})$
 - a) simular $\theta_1^{(t+1)}$ de $\pi_1(\theta_1 | \theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_d^{(t)})$;
 - b) simular $\theta_i^{(t+1)}$ de $\pi_i(\theta_i | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \theta_{i+2}^{(t)}, \dots, \theta_d^{(t)})$, $i = 2, 3, \dots, d-1$, y
 - c) simular $\theta_d^{(t+1)}$ de $\pi_d(\theta_d | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{d-1}^{(t+1)})$.
3. Hacer $\bar{\theta}^{(t+1)} = (\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_d^{(t+1)})$, incrementar t y regresar al paso 2.a).

Este algoritmo es fácil de implementar siempre y cuando podamos obtener de una manera práctica las muestras aleatorias de las densidades condicionales completas, lo cual no ocurre en la mayoría de los casos en la estadística aplicada.

Recientemente el software WinBUGS³ ha facilitado la implementación de los métodos Monte Carlo vía cadenas de Markov (ver Lunn *et al.* (2000) para una breve descripción del programa), aunque puede ser lento y en ocasiones las advertencias (avisos de errores) son poco útiles para solucionar problemas. Congdon (2003, 2010) y Ntzoufras (2009), por ejemplo, utilizan el software WinBUGS para la implementación de los métodos Monte Carlo vía cadenas de Markov en estadística bayesiana.

En este trabajo la implementación del algoritmo del muestreo de Gibbs se realizó por medio de un programa usando el lenguaje “R”, mismo que se presenta en la sección 4.7 del siguiente capítulo por medio del pseudocódigo y de manera íntegra en el Apéndice C.

3.7. Metropolis-Hastings dentro del muestreo de Gibbs

Para realizar un muestreo de Gibbs es necesario saber cómo muestrear valores de las densidades marginales condicionales completas $\pi_i(\theta_i | \bar{\theta}_{-i})$. Si éstas tienen una forma estándar no hay problema puesto que es fácil muestrear usando el método de la función inversa o bien el método de aceptación-rechazo. Sin embargo hay ocasiones, como en el modelo que vamos a estudiar en este trabajo, donde algunas de las densidades marginales condicionales completas no son de forma estándar y esto representa un problema para implementar el algoritmo.

³WinBUGS forma parte del proyecto BUGS (Bayesian inference Using Gibbs Sampling), que tiene como objetivo hacer prácticos los métodos Monte Carlo vía cadenas de Markov disponibles para la estadística aplicada.

El proyecto BUGS se refiere a un software flexible para el análisis bayesiano de modelos estadísticos complejos utilizando métodos Monte Carlo vía cadenas de Markov. El proyecto se inició en 1989 en la Unidad de Bioestadística del Medical Research Council en Cambridge, posteriormente en colaboración con el St. Mary’s Hospital del Imperial College, en Londres se desarrolló el software WinBUGS.

Parece atractivo solucionar este inconveniente, si se conoce una función aproximada $q_i(\theta_i | \bar{\theta}_{-i})$ a la densidad marginal condicional completa $\pi_i(\theta_i | \bar{\theta}_{-i})$, de la cual es fácil obtener muestras. Ya que la usaríamos en vez de π_i pero esta forma no hará que la distribución estacionaria corresponda exactamente con π_i ; no importa lo larga que sea la trayectoria de la cadena simulada. (Gilks et al, 1996).

El estadístico Gelman en 1992 (Gelman, 1992) sugirió una manera de muestrear a partir de una densidad marginal condicional completa aproximada cuya distribución estacionaria fuera exactamente la requerida para la cadena. Lo cual implica usar la función aproximada q_i como distribución propuesta en el algoritmo Metropolis-Hastings. Es decir, supongamos que se realiza un muestreo de Gibbs para el vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{i-1}, \theta_i, \theta_{i+1}, \dots, \theta_d)$ y la densidad condicional marginal completa de θ_i no es de forma estándar. Entonces se aplica un paso del algoritmo Metropolis-Hastings dentro del muestreo de Gibbs para generar valores del parámetro θ_i . En Geweke (2005) y Robert y Casella (2004) dicho algoritmo se expone como sigue:

Para $i = 1, \dots, d$, dados $(\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)})$:

1. Generar θ^* de $q_i(\cdot | \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)})$;
2. generar un valor u de una $U(0, 1)$;
3. establecer $\theta_i^{(t+1)} = \theta^*$ si u es menor o igual a

$$\min \left\{ \frac{\pi_i(\theta^* | \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)}) q_i(\theta_i^{(t)} | \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta^*, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)})}{\pi_i(\theta_i^{(t)} | \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)}) q_i(\theta^* | \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \theta_{i+1}^{(t)}, \dots, \theta_d^{(t)})}, 1 \right\},$$

4. en otro caso establecer $\theta_i^{(t+1)} = \theta_i^{(t)}$.

En este caso sólo una iteración del Metropolis-Hastings es necesaria (Robert y Casella, 2004). La validez del algoritmo se da debido a que la cadena generada cumple las propiedades necesarias para la convergencia. En el artículo de Roberts y Rosenthal (2006) se estudia la recurrencia de Harris y la convergencia de las cadenas generadas mediante este algoritmo.

En las secciones siguientes se usará este algoritmo para muestrear valores de las densidades marginales condicionales completas de las variables latentes.

3.8. Diagnósticos de convergencia

Al hacer una simulación utilizando los algoritmos Monte Carlo vía Cadenas de Markov, deseamos obtener una muestra que provenga de la distribución objetivo. ¿Cómo saber si nuestra simulación fue adecuada en el sentido de que la distribución de la cadena ya es su distribución estacionaria? Existen dos maneras, la gráfica y los diagnósticos de convergencia. Dentro de los métodos gráficos encontramos los siguientes:

1. Valores muestreados contra iteración (Trace plot): Si el resultado es una trayectoria que no recorre el soporte de la distribución objetivo y que en periodos grandes no cambia de valor, es indicio de un mal muestreo.
2. Histograma: Un histograma de los valores muestreados nos sirve para observar la distribución que toman éstos, de nuevo un histograma muy deforme (en el sentido de que no parezca una densidad definida en el soporte de la distribución de interés), nos da información de un mal muestreo.
3. Gráfico de autocorrelaciones: Nos da información acerca de la dependencia entre los valores muestreados. El número de pasos que tarde en que la autocorrelación sea aproximadamente cero debe ser muy chico para que la muestra sea independiente.
4. Gráfico de promedios ergódicos: Nos sirve para observar la velocidad con la que el promedio de la muestra converge a un valor.

Los diagnósticos de convergencia, son métodos estadísticos que nos permiten conocer si una cadena simulada con los algoritmos Monte Carlo vía cadenas de Markov ha logrado converger al equilibrio. En algunas de estas pruebas los resultados nos sirven para mejorar el análisis modificando el periodo de calentamiento; es decir el tiempo que tarda en llegar a un estado estacionario, o el tamaño de la simulación.

Los diagnósticos de Gelman y Rubin (1992) y de Raftery y Lewis (1992) son los más populares debido a su fácil implementación. En este trabajo también usaremos el diagnóstico de Geweke (1992) y el de Heidelberger y Welch (1983). Para realizar tal tarea, se utilizaron las funciones del paquete CODA de “R”, Plummer *et al.* (2006). A continuación se exponen los detalles de las pruebas tomando como referencia a Cowles y Carlin (1996) y Robert y Casella (2010).

3.8.1. Geweke

El diagnóstico de Geweke (1992) evalúa la convergencia de la media de cada parámetro de una cadena después que n valores fueron muestreados. Así, para una

cadena de Markov $\{X_n\}_{n=0}^{\infty}$ y para cualquier función h tal que $E(|h|) < \infty$, la varianza asintótica del promedio $\mu_n = \frac{1}{n} \sum_{j=1}^n h(x^{(j)})$, se calcula utilizando la densidad espectral de $h(x^{(t)})$ como:

$$S_h(w) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \text{cov}(h(x^{(0)}), h(x^{(t)})) e^{tw},$$

donde ι es el número complejo $\sqrt{-1}$.

La metodología consiste en calcular el promedio a partir de las $n_A = 0.1n$ primeras iteraciones de la cadena como $\delta_A = \frac{1}{n_A} \sum_{j=1}^{n_A} h(x^{(j)})$ y estimar σ_A^2 a partir de $S_h(0)$. Después calcular otro promedio con las últimas $n_B = 0.5n$ iteraciones de la cadena como $\delta_B = \frac{1}{n_B} \sum_{j=n-n_B+1}^n h(x^{(j)})$ y estimar σ_B^2 a partir de $S_h(0)$.

La hipótesis a contrastar es que δ_A es igual a δ_B , si ésta se rechaza hay evidencia de que el promedio μ_n no ha convergido. El estadístico de prueba está dado por:

$$Z = \frac{\delta_A - \delta_B}{\sqrt{\hat{\sigma}_B^2/n_B + \hat{\sigma}_A^2/n_A}}.$$

Si las relaciones n_A/n y n_B/n son fijas y $n_A + n_B < n$, entonces por el teorema de límite central, el estadístico de prueba se distribuye como una variable aleatoria normal estándar para una n suficientemente grande.

Cuando el estadístico de prueba en valor absoluto es mayor a 2, con un 95 % de confianza se dice que no hay una estacionariedad adecuada y será necesario tomar un periodo de calentamiento mayor.

3.8.2. Gelman-Rubin

Una manera de identificar la ausencia o la presencia de convergencia es simular M cadenas $\{x_m^{(j)}\}_{j=1}^n$ ($1 \leq m \leq M$) con distintos valores iniciales. El criterio de Gelman y Rubin (1992) está construido sobre la comparación de la varianza dentro de las cadenas y la varianza entre las cadenas, definidas como:

$$B = \frac{1}{M-1} \sum_{m=1}^M (\delta_m - \delta)^2 \text{ para la varianza entre cadenas y}$$

$$W = \frac{1}{M-1} \sum_{m=1}^M \frac{1}{n-1} \sum_{j=1}^n (h(x_m^{(j)}) - \delta_m)^2 \text{ para la varianza dentro la cadena, donde } h \text{ es}$$

cualquier función tal que $E(|h|) < \infty$, $\delta_m = \frac{1}{n} \sum_{j=1}^n h(x_m^{(j)})$ y $\delta = \frac{1}{M} \sum_{j=1}^M \delta_m$.

Un estimador de la varianza de $h(x^{(j)})$ está dado por

$$\hat{\sigma}^2 = \frac{n-1}{n}W + B.$$

En este método, se calcula un estadístico R que contienen la información de estas varianzas,

$$R = \frac{\hat{\sigma}^2 + \frac{B}{M}}{W} \frac{\nu+1}{\nu+3},$$

donde $\nu = 2(\hat{\sigma}^2 + \frac{B}{M})^2$ son los grados de libertad estimados. La convergencia ha sido alcanzada si R es aproximadamente 1. Un valor que se aleje de 1, nos indicará que no hay convergencia y es necesario hacer más iteraciones.

3.8.3. Raftery-Lewis

La metodología de Raftery y Lewis (1992) verifica si se ha alcanzado con un grado de exactitud previamente establecido, cuantiles de interés de la distribución de la cadena.

En esta metodología, bajo el supuesto de independencia, se puede determinar el número mínimo de iteraciones (n_{min}) necesarias para estimar el cuantil de interés con la precisión previamente especificada, así como el tamaño del periodo de calentamiento. Finalmente también se calcula un estadístico $I = \frac{n}{n_{min}}$ que mide el incremento en el número de iteraciones debido a la dependencia en la muestra. Valores de I más grandes que 5 indican alto nivel de dependencia y problemas que pueden ser resueltos haciendo cambios en la implementación. Tal dependencia puede provenir de valores iniciales muy malos o alta correlación que se corrige tomando cada k valores de la muestra original para formar una nueva muestra.

El paquete CODA se encuentra instrumentado este diagnóstico, el usuario debe indicar los siguientes valores:

- q = el cuantil a ser estimado,
- r = el margen de error para la estimación y
- s = la probabilidad de obtener un estimador en el intervalo $(q - r, q + r)$.

Los valores predeterminados en CODA son $q = 0.025$, $r = 0.005$ y $s = 0.95$.

3.8.4. Heidelberg-Welch

El diagnóstico de Heidelberg y Welch (1983) consta de dos partes. En la primera se contrasta la hipótesis nula de que la cadena o alguna parte de ella proviene de una distribución estacionaria. En la segunda parte se calcula un intervalo de confianza para la media usando únicamente la parte de la cadena que fue estacionaria, esto con el fin de analizar su longitud. A continuación se describe el desarrollo de la metodología de las pruebas.

Sea $S_h(0)$ la densidad espectral de la muestra evaluada en 0, obtenida a partir de la segunda mitad de la cadena (para evitar un error de estimación que provenga de una posible fase inicial transitoria). Considere la siguiente función:

$$B_n(t) = \frac{Y_{[nt]} - [nt]\bar{X}}{(n\hat{S}_n(0))^{1/2}}, \quad 0 \leq t \leq 1,$$

donde $Y_0 = 0$, $Y_n = \sum_{j=1}^n X^j$, $\bar{X} = \frac{1}{n} \sum_{j=1}^n X^j$ y $[u]$ denota el mayor entero menor o igual a u . Entonces, bajo la hipótesis nula de estacionariedad, se tiene que $B_n = B_n(t)$ se distribuye aproximadamente como un puente Browniano⁴ para una n suficientemente grande. El estadístico de Cramer-von-Mises se define como

$$T = \int_0^1 B_n(t)^2 dt,$$

y se puede utilizar para contrastar la hipótesis nula.

La prueba se aplica sucesivamente, comenzando con la cadena entera, si se rechaza la hipótesis nula se descarta el primer 10 % de la muestra y se vuelve a realizar el test. Se continua realizando la prueba después de descartar el primer 20 %, 30 %, ... de la cadena hasta que la hipótesis nula no se rechaza o el 50 % de la cadena ha sido descartado. En el caso de que la hipótesis nula se continúe rechazando después de eliminar el 50 % de la muestra, es indicio de que se deben tomar más iteraciones.

En caso de que en algún momento no se rechace la hipótesis de estacionariedad, se realiza la segunda parte de la metodología de Heidelberg y Welch, usando sólo la parte de la cadena que pasó la prueba para calcular el error estándar de la media como $\sqrt{\frac{\hat{S}_n(0)}{n_p}}$, donde n_p es el tamaño de la nueva cadena, y posteriormente se construye un intervalo de confianza para la media.

Si la mitad del ancho del intervalo de confianza generado es menor que ϵ veces la media muestral de las iteraciones ó el cociente

⁴Ver Ross (1996).

$$c = \frac{\text{mitad de ancho del intervalo}}{\text{media}} < \epsilon,$$

se considera que el tamaño de la nueva cadena es suficiente como para estimar la media con precisión. El valor de ϵ generalmente utilizado en este procedimiento es 0.1 y el nivel de confianza es de 95 %.

Capítulo 4

Un modelo de volatilidad estocástica

4.1. Introducción

Los modelos de volatilidad han sido muy utilizados en el análisis de la varianza de series de tiempo de retornos financieros, ver por ejemplo Taylor (1982, 1986), Harvey *et al.* (1994), Jacquier *et al.* (1994), Shephard (2004), Broto y Ruíz (2004), Shephard y Anderson (2009), Asai *et al.* (2006). Las metodologías existentes para la modelación de la volatilidad se pueden clasificar en aquellas que se caracterizan por modelar la varianza condicional en función de las observaciones pasadas de la serie de los retornos (modelos ARCH propuestos por Engle, 1982) y de sus propios valores rezagados (modelos GARCH propuestos por Bollerslev, 1986), y en las que se caracterizan por modelar la volatilidad de una serie de tiempo como función de un proceso estocástico no observable, específicamente los modelos de **Volatilidad Estocástica** (VE) (propuesto por Taylor, 1982)

La estimación de los parámetros de los modelos VE es de mayor dificultad debido a la complejidad de la expresión analítica de la función de verosimilitud. Esto se debe a que las perturbaciones aleatorias que aparecen en las ecuaciones del modelo, lo hacen de forma no lineal. Como consecuencia, no es posible utilizar directamente los métodos de estimación habituales basados en Máxima Verosimilitud. Por lo tanto los métodos de Monte Carlo vía cadenas de Markov son recomendados.

A continuación se exponen la estructura de los datos, el modelo de volatilidad estocástica univariado y el multivariado (VEM). Posteriormente se presenta el análisis de la estimación bayesiana; es decir, se mostrará el cálculo de la densidad *a posteriori*, las densidades condicionales marginales completas de cada parámetro

del modelo y al final del capítulo se presenta el pseudocódigo del algoritmo para la simulación de valores a partir de estas densidades.

4.2. Estructura de datos

Denotemos por $X_j(t)$ la observación del máximo de ozono registrado en la región j -ésima en el día t -ésimo, con $j = 1, 2, \dots, K$ y $t = 1, 2, \dots, T$. Donde K es el número total de regiones a considerar y T es el total de observaciones diarias.

Sea $Z_j(t)$ el promedio semanal de los máximos diarios de ozono en la región j -ésima en la semana t -ésima, con $j = 1, 2, \dots, K$ y $t = 1, 2, \dots, N$, donde N representa el número de semanas a considerar, es decir:

$$Z_j(1) = \frac{1}{7} \sum_{t=1}^7 X_j(t), \quad Z_j(2) = \frac{1}{7} \sum_{t=8}^{14} X_j(t),$$

y así sucesivamente.

Para los promedios semanales de ozono en la región j -ésima definimos el **log-retorno** al tiempo t como el logaritmo natural de la medición relativa del promedio de la semana t respecto al de la semana $t - 1$ y es denotado por $Y_j(t)$, es decir:

$$Y_j(t) = \ln \left[\frac{Z_j(t)}{Z_j(t-1)} \right] \quad \text{con } j = 2, \dots, K \text{ y } t = 1, 2, \dots, N. \quad (4.1)$$

Un valor positivo de $Y_j(t)$ indica que la medición promedio en cuestión es mayor al promedio inmediato anterior, por lo que los log-retornos nos indican incrementos respecto al periodo anterior. De manera análoga un valor negativo de $Y_j(t)$ indica un decremento respecto al periodo anterior.

Sea $\mathbf{Y}(t)$ el vector que contiene los log-retornos al tiempo t de cada una de las K -regiones:

$$\mathbf{Y}(t) = (Y_1(t), Y_2(t), \dots, Y_k(t))',$$

donde v' indica el vector v transpuesto.

Así estamos en posibilidad de definir la matriz $\mathbf{D}_{K \times N}$ como el conjunto de datos observados a modelar, es decir,

$$\mathbf{D} = \{\mathbf{Y}(1), \mathbf{Y}(2), \dots, \mathbf{Y}(N)\},$$

o bien,

$$\mathbf{D} = \begin{pmatrix} Y_1(1) & Y_1(2) & Y_1(3) & \dots & Y_1(N) \\ Y_2(1) & Y_2(2) & Y_2(3) & \dots & Y_2(N) \\ \vdots & \vdots & \vdots & & \vdots \\ Y_K(1) & Y_K(2) & Y_K(3) & \dots & Y_K(N) \end{pmatrix}_{K \times N}.$$

Notar que los renglones de la matriz $\mathbf{D}_{K \times N}$ son las series de tiempo de los log-retornos de cada región.

4.3. Modelo de volatilidad estocástica univariado

Como hemos mencionado, el modelo de volatilidad estocástica fue inspirado en el contexto de la finanzas (Taylor, 1982, Shephard, 2004), específicamente fue pensado para modelar la volatilidad de la capitalización continua de los log-retornos diarios de acciones, calculada como la diferencia del logaritmo del precio de la acción entre dos días consecutivos. La volatilidad de cualquier serie $Y(t)$, $t = 1, 2, \dots, N$, por definición es el segundo momento condicional de la serie $Y(t)$, hasta la información conocida a tiempo $t - 1$, es decir

$$\sigma^2(t) = E[Y^2(t) \mid Y(1), Y(2), \dots, Y(t-1)]. \quad (4.2)$$

Siguiendo el planteamiento de Taylor (1982), se supone que la serie de tiempo $Y(t)$, $t = 1, 2, \dots, N$ puede ser modelada como:

$$Y(t) = \sigma(t)\epsilon(t). \quad (4.3)$$

Esto significa que el modelo supone que $Y(t)$ se comporta según la raíz cuadrada de su segundo momento condicional perturbado por una cantidad aleatoria $\epsilon(t)$. El supuesto sobre $\epsilon(t)$ es que sigue una distribución normal $\mathcal{N}(0, 1)$ y es independiente de $\sigma(t)$.

La especificación acerca de la volatilidad de la serie $Y(t)$, $t = 1, 2, \dots, N$ es que está dada por:

$$\sigma^2(t) = e^{h(t)}, \quad (4.4)$$

y por tanto

$$\sigma(t) = e^{h(t)/2}, \quad (4.5)$$

donde $h(t)$ es un proceso gaussiano lineal con media diferente a cero. Específicamente será supuesto como un proceso autorregresivo de primer orden definido por:

$$\begin{cases} h(t) = \mu + \eta(t) & , \quad t = 1 \\ h(t) = \mu + \phi(h(t-1) - \mu) + \eta(t) & , \quad t = 2, \dots, N, \end{cases} \quad (4.6)$$

donde μ es un número real, ϕ es el parámetro que está relacionado con la persistencia de los shocks a la volatilidad. Para asegurar que $\{h(t)\}_{t=1}^N$ sea estacionario suponemos que $|\phi| < 1$ y $\eta(t)$ se distribuye $\mathcal{N}(0, \sigma_\eta^2)$, $t = 2, \dots, N$ y son variables aleatorias independientes e idénticamente distribuidas. Las variables $h(t)$, $t = 1, 2, \dots, N$ son conocidas como variables latentes o de estado.

Así el modelo univariado queda conformado de la siguiente manera:

$$Y(t) = e^{h(t)/2} \epsilon(t), \quad t = 1, \dots, N, \quad (4.7)$$

donde, $Y(t)$ es la series de tiempo de los log-retornos, $\epsilon(t)$ $t = 1, \dots, N$ son variables aleatorias independientes y se distribuyen $\mathcal{N}(0, 1)$, y $h(t)$ son las variables latentes definidas en (4.6).

Lema 4.3.1 *Si Z se distribuye $\mathcal{N}(0, \sigma^2)$ y si $Y = aZ + \mu$, entonces Y se distribuye $\mathcal{N}(\mu, a^2\sigma^2)$ donde a y μ son números reales.*

Demostración: Sabemos que la función generadora de momentos de una variable aleatoria X que se distribuye $\mathcal{N}(\mu, \sigma^2)$ está dada por $\mathcal{M}_X(t) = e^{t\mu + \sigma^2 t^2/2}$. Ahora encontramos la función generadora de momentos de la variable Y .

$$\begin{aligned} \mathcal{M}_Y(t) &= E[e^{(aZ + \mu)t}] \\ &= e^{t\mu} E[e^{(at)Z}] \\ &= e^{t\mu} \mathcal{M}_Z(at) \\ &= e^{t\mu} e^{(\sigma^2 a^2 t^2)/2} \\ &= e^{t\mu + (\sigma^2 a^2) t^2/2}, \end{aligned}$$

por lo tanto, Y se distribuye $\mathcal{N}(\mu, a^2\sigma^2)$. ■

Proposición 4.3.2 *Bajo los supuestos del modelo de VE, planteados en esta sección:*

a) *La variable aleatoria $h(1)$ tiene una distribución $\mathcal{N}(\mu, \sigma_\eta^2)$.*

- b) La variable aleatoria $h(t)$ dado $h(t-1)$, tiene una distribución $\mathcal{N}(\mu + \phi[h(t-1) - \mu], \sigma_\eta^2)$, $t = 2, \dots, N$.
- c) $E(Y(t) | h(t)) = 0$.
- d) $\text{Var}(Y(t) | h(t)) = e^{h(t)}$.
- e) La variable aleatoria $Y(t)$ dado $h(t)$, tiene una distribución $\mathcal{N}(0, e^{h(t)})$, $t = 1, \dots, N$.

Demostración:

- a) Como $\eta(t)$ se distribuye $\mathcal{N}(0, \sigma_\eta^2)$ y $h(1) = \mu + \eta(1)$, por el lema 4.3.1 es inmediato que $h(1)$ tiene una distribución $\mathcal{N}(\mu, \sigma_\eta^2)$.
- b) Como $\eta(t)$ se distribuye $\mathcal{N}(0, \sigma_\eta^2)$ y $h(t) = \mu + \phi(h(t-1) - \mu) + \eta(t)$, $t = 2, \dots, N$, y por el lema 4.3.1, $h(t)$ dado $h(t-1)$ se distribuye $\mathcal{N}(\mu + \phi(h(t-1) - \mu), \sigma_\eta^2)$.
- c) Como en el modelo $Y(t) = e^{h(t)/2}\epsilon(t)$ y $\epsilon(t)$ es una variable aleatoria con distribución normal $\mathcal{N}(0, 1)$, entonces

$$\begin{aligned} E[Y(t) | h(t)] &= E[e^{h(t)/2}\epsilon(t) | h(t)] \\ &= e^{h(t)/2}E[\epsilon(t) | h(t)] \\ &= 0. \end{aligned}$$

- d) Por c), la varianza de $Y(t)$ dado $h(t)$ será igual al segundo momento de $Y(t)$ dado $h(t)$:

$$\begin{aligned} \text{Var}[Y(t) | h(t)] &= E[Y^2(t) | h(t)] \\ &= E[e^{h(t)}\epsilon^2(t) | h(t)] \\ &= e^{h(t)}E[\epsilon^2(t) | h(t)] \\ &= e^{h(t)}. \end{aligned}$$

Esto implica que $h(t)$ puede ser entendida como el logaritmo de la varianza condicional de los log-retornos.

- e) De c) y d) sabemos que $E[Y(t) | h(t)] = 0$ y $\text{Var}[Y(t) | h(t)] = e^{h(t)}$ y como $\epsilon(t)$ tiene una distribución $\mathcal{N}(0, 1)$, entonces usando el lema 4.3.1 se tiene que dado $h(t)$ la variable aleatoria $Y(t) = e^{h(t)/2}\epsilon(t)$ tiene una distribución $\mathcal{N}(0, e^{h(t)})$. ■

Finalmente por e) de la proposición 4.3.2 la función de verosimilitud del modelo está dada por:

$$L(\mathbf{Y} \mid h(t), \theta) = \prod_{t=1}^N \frac{1}{\sqrt{2\pi e^{h(t)}}} \exp \left\{ -\frac{Y^2(t)}{2e^{h(t)}} \right\}. \quad (4.8)$$

Así el Modelo de Volatilidad Estocástica es un modelo de series de tiempo no lineal que consta de dos ecuaciones. La primera para $Y(t)$ y la segunda representando la evolución de las variables latentes $h(t)$ dado $h(t-1)$. Es decir, en el modelo están involucrados dos procesos estocásticos uno para las observaciones y otro para la volatilidad latente, donde los parámetros desconocidos a estimar son: μ , ϕ y σ_η^2 , o bien el vector de parámetros a estimar es $\theta = (\mu, \phi, \sigma_\eta^2)$.

4.4. Modelo de volatilidad estocástica multivariado.

Actualmente existe abundante literatura sobre los modelos de volatilidad estocástica multivariada (VEM), un excelente revisión de las distintas versiones del modelo fue realizada por Chib *et al.* (2008) en el artículo *Multivariate Stochastic Volatility*. A continuación se presenta la extensión del modelo presentado en la sección 4.3 para el análisis de varias series temporales.

Sea $\mathbf{Y}(t) = (Y_1(t), Y_2(t), \dots, Y_K(t))'$, $t = 1, 2, \dots, N$ el vector de los log-retornos en cada una de las K regiones. Se supone que la serie de tiempo $Y_j(t)$, $t = 1, 2, \dots, N$ puede ser escrita como:

$$Y_j(t) = \sigma_j(t)\epsilon_j(t). \quad (4.9)$$

Además suponemos que la volatilidad de la serie $Y_j(t)$ está dada por:

$$\sigma_j^2(t) = e^{h_j(t)}. \quad (4.10)$$

Así el modelo queda determinado por:

$$Y_j(t) = e^{h_j(t)/2}\epsilon_j(t) \quad , \quad t = 1, \dots, N \quad , \quad j = 1, \dots, K, \quad (4.11)$$

donde:

- $Y_j(t)$ es la serie de tiempo de los log-retornos, en la región j -ésima,
- $h_j(t)$ son las variables latentes definidas en (4.6),

Para asegurar que $\{h_j(t)\}_{t=1}^N$ sea estacionario, suponemos que $|\phi_j| < 1$.

En forma matricial, la ecuación que describe el comportamiento de los log-retornos está dada por:

$$\mathbf{Y}(t) = \mathbf{H}(t)\boldsymbol{\epsilon}(t), \quad (4.12)$$

donde $\mathbf{H}(t)$ es la matriz diagonal de orden K definida por

$$\mathbf{H}(t) = \text{diag}(e^{h_1(t)/2}, e^{h_2(t)/2}, \dots, e^{h_K(t)/2}),$$

y el vector $\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \epsilon_2(t), \dots, \epsilon_K(t))'$, es el vector de perturbaciones aleatorias asociado a los log-retornos. Así

$$\begin{pmatrix} Y_1(t) \\ Y_2(t) \\ \vdots \\ Y_K(t) \end{pmatrix} = \begin{pmatrix} e^{h_1(t)/2} & 0 & \dots & 0 \\ 0 & e^{h_2(t)/2} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & e^{h_K(t)/2} \end{pmatrix} \begin{pmatrix} \epsilon_1(t) \\ \epsilon_2(t) \\ \vdots \\ \epsilon_K(t) \end{pmatrix}$$

Denotemos por $\mathbf{h}(t) = (h_1(t), h_2(t), \dots, h_K(t))'$, el vector integrado por las variables latentes al tiempo t en cada una de las K -regiones.

Se suponen las siguientes hipótesis que definen al modelo VEM:

- $\boldsymbol{\epsilon}(t) = (\epsilon_1(t), \epsilon_2(t), \dots, \epsilon_K(t))'$ es el vector de las perturbaciones asociados a los observaciones (log-retornos), y siguen una distribución normal multivariada $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$ con vector de medias $\mathbf{0} = (0, 0, \dots, 0)'$ y una matriz de varianza-covarianza $\boldsymbol{\Sigma}_\epsilon$ dada por

$$\boldsymbol{\Sigma}_\epsilon = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1K} \\ \rho_{21} & 1 & \rho_{23} & \dots & \rho_{2K} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{K1} & \rho_{K2} & \rho_{K3} & \dots & 1 \end{pmatrix}, \quad (4.13)$$

donde $\rho_{i,j} = \text{Cov}(\epsilon_i(t), \epsilon_j(t))$, $i, j = 1, 2, \dots, K$, $i \neq j$ y $\text{Var}(\epsilon_i(t)) = 1$.

- $\boldsymbol{\eta}(t) = (\eta_1(t), \eta_2(t), \dots, \eta_K(t))'$ es el vector de las perturbaciones aleatorias asociados a las variables latentes y tiene distribución normal multivariada $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$ con vector de medias $\mathbf{0} = (0, 0, \dots, 0)'$ y una matriz de varianza-covarianza $\boldsymbol{\Sigma}_\eta$, dada por $\boldsymbol{\Sigma}_\eta = \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2, \dots, \sigma_{\eta_K}^2)$.

A continuación se presenta el resultado equivalente a la proposición 4.3.2, pero ahora para el caso multivariado.

Proposición 4.4.1 *Con las hipótesis del modelo VEM, planteadas en esta sección:*

- a) $h_j(1)$ se distribuye $\mathcal{N}(\mu_j, \sigma_{\eta_j}^2)$, $j = 1, 2, \dots, K$.
- b) Dado $h_j(t-1)$, $h_j(t)$ tiene distribución $\mathcal{N}(\mu_j + \phi_j[h_j(t-1) - \mu_j], \sigma_{\eta_j}^2)$, $j = 1, 2, \dots, K$ y $t = 2, \dots, N$.
- c) Dado el vector $\mathbf{h}(t)$, $\mathbf{Y}(t)$ tiene una distribución normal multivariada $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Y}(t)})$ con vector de medias $\mathbf{0} = (0, 0, \dots, 0)'$ y una matriz de varianza-covarianza $\boldsymbol{\Sigma}_{\mathbf{Y}(t)}$ dada por

$$\boldsymbol{\Sigma}_{\mathbf{Y}(t)} = \begin{pmatrix} e^{h_1(t)} & \rho_{12}e^{(h_1(t)+h_2(t))/2} & \dots & \rho_{1K}e^{(h_1(t)+h_K(t))/2} \\ \rho_{21}e^{(h_2(t)+h_1(t))/2} & e^{h_2(t)} & \dots & \rho_{2K}e^{(h_2(t)+h_K(t))/2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{K1}e^{(h_K(t)+h_1(t))/2} & \rho_{K2}e^{(h_K(t)+h_2(t))/2} & \dots & e^{h_K(t)} \end{pmatrix}.$$

Omitimos la prueba de este resultado puesto que en este trabajo se estudiará el modelo VEM bajo el supuesto de que $\rho_{ij} = 0$, $i, j = 1, 2, \dots, K$, $i \neq j$. Así, se ha alterado la matriz de varianza-covarianza del vector de las perturbaciones aleatorias $\boldsymbol{\epsilon}(t)$ pero no su distribución, es decir, $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}})$, donde $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \text{diag}(1, 1, \dots, 1)$, por lo que los ϵ_j son mutuamente independientes. De esta forma se pueden tomar los casos de cada región y analizarlos separadamente. A continuación se presenta el resultado bajo el supuesto antes mencionado.

Proposición 4.4.2 *La función de densidad condicional conjunta del vector $\mathbf{Y}(t)$ dado $\mathbf{h}(t)$ está dada por:*

$$f(\mathbf{Y}(t) | \mathbf{h}(t)) = (2\pi)^{-K/2} \exp \left\{ -\frac{1}{2} \sum_{j=1}^K (h_j(t) + e^{-h_j(t)} [Y_j(t)]^2) \right\}. \quad (4.14)$$

Demostración: Notemos que

$$\begin{aligned}
f(\mathbf{Y}(t) \mid \mathbf{h}(t)) &= f((Y_1(t), Y_2(t), \dots, Y_K(t))' \mid (h_1(t), h_2(t), \dots, h_K(t))') \\
&= \prod_{j=1}^K \frac{1}{\sqrt{2\pi}e^{h_j(t)/2}} \exp \left\{ \frac{-[Y_j(t)]^2}{2e^{h_j(t)}} \right\} \\
&= (2\pi)^{-K/2} \exp \left\{ -\frac{1}{2} \sum_{j=1}^K h_j(t) \right\} \exp \left\{ -\frac{1}{2} \sum_{j=1}^K e^{-h_j(t)} [Y_j(t)]^2 \right\} \\
&= (2\pi)^{-K/2} \exp \left\{ -\frac{1}{2} \sum_{j=1}^K (h_j(t) + e^{-h_j(t)} [Y_j(t)]^2) \right\}. \blacksquare
\end{aligned}$$

Esta proposición nos dice que el vector de los log-retornos, dadas las variables latentes sigue una distribución normal multivariada $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Y}(t)})$ con vector de medias $\mathbf{0} = (0, 0, \dots, 0)'$ y una matriz de varianza-covarianza $\boldsymbol{\Sigma}_{\mathbf{Y}(t)}$ dada por

$$\boldsymbol{\Sigma}_{\mathbf{Y}(t)} = \text{diag}(e^{h_1(t)}, e^{h_2(t)}, \dots, e^{h_K(t)}). \quad (4.15)$$

Notemos que en cada una de las K -regiones debemos estimar μ_j , ϕ_j y $\sigma_{\eta_j}^2$ $j = 1, \dots, K$. Denotemos por $\boldsymbol{\Theta}$ el vector de parámetros a estimar en el modelo, es decir,

$$\begin{aligned}
\boldsymbol{\Theta} &= (\boldsymbol{\mu}; \boldsymbol{\phi}; \boldsymbol{\sigma}_{\eta}^2) \\
&= ((\mu_1, \mu_2, \dots, \mu_K)'; (\phi_1, \phi_2, \dots, \phi_K)'; (\sigma_{\eta_1}^2, \sigma_{\eta_2}^2, \dots, \sigma_{\eta_K}^2)').
\end{aligned}$$

La estimación de estos parámetros se realizará mediante análisis bayesiano, específicamente por métodos de Monte Carlo vía cadenas de Markov. En la siguiente sección se exponen los detalles teóricos de la estimación.

4.5. Análisis bayesiano

Iniciamos definiendo el vector de parámetros y variables latentes del modelo que denotaremos por $\boldsymbol{\varphi}$. De esta forma se tiene que

$$\begin{aligned}
\boldsymbol{\varphi} &= (\boldsymbol{\Theta}, \mathbf{h}) \\
&= ((\mu_1, \mu_2, \dots, \mu_K)'; (\phi_1, \phi_2, \dots, \phi_K)'; (\sigma_{\eta_1}^2, \sigma_{\eta_2}^2, \dots, \sigma_{\eta_K}^2)'; \mathbf{h}),
\end{aligned}$$

donde \mathbf{h} es la matriz de variables latentes en cada una de las K -regiones, y en cada uno de los N -tiempos y está dada por

$$\mathbf{h} = \begin{pmatrix} h_1(1) & h_1(2) & h_1(3) & \dots & h_1(N) \\ h_2(1) & h_2(2) & h_2(3) & \dots & h_2(N) \\ \vdots & \vdots & \vdots & & \vdots \\ h_K(1) & h_K(2) & h_K(3) & \dots & h_K(N) \end{pmatrix}.$$

El objetivo de esta sección es determinar la **densidad *a posteriori*** de φ dado \mathbf{D} , es decir,

$$\begin{aligned} P(\varphi \mid \mathbf{D}) &\propto L(\mathbf{D} \mid \varphi)P(\varphi) \\ &\propto L(\mathbf{D} \mid \varphi)P(\Theta, \mathbf{h}) \\ &\propto L(\mathbf{D} \mid \varphi)P(\mathbf{h} \mid \Theta)P(\Theta), \end{aligned} \tag{4.16}$$

donde:

- $P(\varphi \mid \mathbf{D})$ es la densidad *a posteriori* de los parámetros dada la información de la muestra.
- $L(\mathbf{D} \mid \varphi)$ es la función de verosimilitud del modelo.
- $P(\mathbf{h} \mid \Theta)$ es la densidad condicional de las variables latentes dados los parámetros del modelo.
- $P(\Theta)$ es la densidad *a priori* conjunta de los parámetros.

A continuación se calculan cada una de las funciones que integran la densidad *a posteriori* del modelo.

4.5.1. Función de verosimilitud

Para calcular la densidad *a posteriori* de los parámetros a estimar dados los datos es necesario conocer la función de verosimilitud de los datos dados los parámetros.

Proposición 4.5.1 *La función de verosimilitud del modelo está dada por:*

$$L(\mathbf{D} \mid \varphi) \propto \exp \left(-\frac{1}{2} \sum_{t=1}^N \sum_{j=1}^K [h_j(t) + e^{-h_j(t)} [Y_j(t)]^2] \right). \tag{4.17}$$

Demostración:

Usando la densidad condicional conjunta del vector $\mathbf{Y}(t)$ dado $\mathbf{h}(t)$ (ecuación (4.14)) se tiene:

$$\begin{aligned}
L(\mathbf{D} \mid \boldsymbol{\varphi}) &= f(\mathbf{Y}(1), \mathbf{Y}(2), \dots, \mathbf{Y}(N) \mid \boldsymbol{\Theta}, \mathbf{h}) \\
&= \prod_{t=1}^N P(\mathbf{Y}(t) \mid \boldsymbol{\Theta}, \mathbf{h}) \\
&= \prod_{t=1}^N (2\pi)^{-K/2} \exp\left(-\frac{1}{2} \sum_{j=1}^K (h_j(t) + e^{-h_j(t)} [Y_j(t)]^2)\right) \\
&\propto \exp\left(-\frac{1}{2} \sum_{t=1}^N \sum_{j=1}^K [h_j(t) + e^{-h_j(t)} [Y_j(t)]^2]\right). \blacksquare
\end{aligned}$$

4.5.2. Densidad conjunta de las variables latentes condicional a los parámetros del modelo

Como se denotamos antes \mathbf{h} es la matriz de variables latentes, dada por $\mathbf{h} = (\mathbf{h}(1), \mathbf{h}(2), \dots, \mathbf{h}(N))$, donde $\mathbf{h}(t) = (h_1(t), h_2(t), \dots, h_K(t))'$. Así la densidad conjunta de las variables latentes condicional a los parámetros del modelo, está dada por:

$$\begin{aligned}
f(\mathbf{h} \mid \boldsymbol{\Theta}) &= f(\mathbf{h}(1), \mathbf{h}(2), \dots, \mathbf{h}(N) \mid \boldsymbol{\Theta}) \\
&= f(\mathbf{h}(1) \mid \boldsymbol{\Theta}) \prod_{t=2}^N f(\mathbf{h}(t) \mid \mathbf{h}(t-1), \boldsymbol{\Theta}). \tag{4.18}
\end{aligned}$$

Usando el resultado **a)** de la proposición (4.4.1) se tiene:

$$\begin{aligned}
f(\mathbf{h}(1) \mid \boldsymbol{\Theta}) &= f(h_1(1), h_2(1), \dots, h_K(1) \mid \boldsymbol{\Theta}) \\
&= \prod_{j=1}^K f(h_j(1) \mid \boldsymbol{\Theta}) \\
&= \prod_{j=1}^K \frac{1}{\sqrt{2\pi}(\sigma_{\eta_j}^2)^{1/2}} \exp\left(-\frac{1}{2\sigma_{\eta_j}^2} [h_j(1) - \mu_j]^2\right) \\
&\propto \prod_{j=1}^K (\sigma_{\eta_j}^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_{\eta_j}^2} [h_j(1) - \mu_j]^2\right).
\end{aligned}$$

Por lo tanto:

$$f(\mathbf{h}(1) \mid \Theta) \propto \prod_{j=1}^K (\sigma_{\eta_j}^2)^{-1/2} \exp \left(-\frac{1}{2\sigma_{\eta_j}^2} [h_j(1) - \mu_j]^2 \right). \quad (4.19)$$

Para $t = 2, \dots, N$, usando el resultado **b)** de la proposición (4.4.1) se tiene que:

$$\begin{aligned} f(\mathbf{h}(t) \mid \mathbf{h}(t-1), \Theta) &= f(h_1(t), h_2(t), \dots, h_K(t) \mid h_1(t-1), h_2(t-1), \dots, h_K(t-1), \Theta) \\ &= \prod_{j=1}^K f(h_j(t) \mid h_j(t-1), \Theta) \\ &= \prod_{j=1}^K \frac{1}{\sqrt{2\pi(\sigma_{\eta_j}^2)^{1/2}}} \exp \left(\frac{1}{2\sigma_{\eta_j}^2} [h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2 \right) \\ &\propto \prod_{j=1}^K (\sigma_{\eta_j}^2)^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma_{\eta_j}^2} [h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2 \right). \end{aligned}$$

Por lo tanto:

$$f(\mathbf{h}(t) \mid \mathbf{h}(t-1), \Theta) \propto \prod_{j=1}^K (\sigma_{\eta_j}^2)^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma_{\eta_j}^2} [h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2 \right). \quad (4.20)$$

Sustituyendo las expresiones (4.19) y (4.20) en (4.18), tenemos la densidad de las variables latentes dados los parámetros.

$$\begin{aligned} f(\mathbf{h} \mid \Theta) &\propto \left[\prod_{j=1}^K (\sigma_{\eta_j}^2)^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma_{\eta_j}^2} [h_j(1) - \mu_j]^2 \right) \right] \\ &\quad \prod_{t=2}^N \left[\prod_{j=1}^K (\sigma_{\eta_j}^2)^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma_{\eta_j}^2} [h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2 \right) \right]. \end{aligned} \quad (4.21)$$

4.5.3. Densidades *a priori* de los parámetros

En el análisis bayesiano es necesario suponer *a priori* que cada uno de los parámetros a estimar sigue una determinada distribución con sus respectivos parámetros (hiperparámetros). La elección de las distribuciones *a priori* se ha hecho considerando anteriores investigaciones, específicamente Achcar *et al.* (2008),

Achcar *et al.*(2010), Achcar *et al.*(2011) y Zozolotto (2010). Sin embargo en este trabajo no se utiliza la distribución $Beta(\alpha, \beta)$ como distribución *a priori* para el parámetros ϕ_j . Así suponemos que para $j = 1, \dots, K$:

μ_j sigue una distribución *a priori* Normal con hiperparámetros e_j y f_j , es decir:

$$\begin{aligned} P(\mu_j) &= \mathcal{N}(e_j, f_j) \\ &\propto \exp\left(\frac{-1}{2f_j}(\mu_j - e_j)^2\right). \end{aligned} \quad (4.22)$$

ϕ_j sigue una distribución *a priori* Normal con hiperparámetros a_j y b_j , es decir:

$$\begin{aligned} P(\phi_j) &= \mathcal{N}(a_j, b_j) \\ &\propto \exp\left(\frac{-1}{2b_j}(\phi_j - a_j)^2\right). \end{aligned} \quad (4.23)$$

$\sigma_{\eta_j}^2$ sigue una distribución *a priori* Gamma Inversa con hiperparámetros c_j y d_j , es decir:

$$\begin{aligned} P(\sigma_{\eta_j}^2) &= \mathcal{GI}(c_j, d_j) \\ &\propto \frac{\exp(-d_j/\sigma_{\eta_j}^2)}{(\sigma_{\eta_j}^2)^{c_j+1}}, \end{aligned} \quad (4.24)$$

o equivalentemente su recíproco $\frac{1}{\sigma_{\eta_j}^2}$, se distribuye Gamma con hiperparámetros c_j y d_j conocidos, es decir:

$$\begin{aligned} P(1/\sigma_{\eta_j}^2) &= \mathcal{G}(c_j, d_j) \\ &\propto \frac{\exp(-d_j/\sigma_{\eta_j}^2)}{(\sigma_{\eta_j}^2)^{c_j-1}}. \end{aligned} \quad (4.25)$$

Los hiperparámetros $a_j, b_j, c_j, d_j, e_j, f_j, j = 1, \dots, K$ se consideran constantes conocidas y serán especificadas en el capítulo siguiente.

Así bajo la hipótesis de independencia *a priori* de los parámetros, la densidad *a priori* conjunta está dada por:

$$P(\Theta) = \prod_{j=1}^K P(\mu_j)P(\phi_j)P(\sigma_{\eta_j}^2). \quad (4.26)$$

Finamente las variables $h_j(t)$ serán vistas como parámetros que hay que estimar, por lo que también se requiere asignarles densidades *a priori*. Así el supuesto acerca de las densidades *a priori* son:

- Para $h_j(1)$ tomamos

$$\begin{aligned} P(h_j(1)) &= \mathcal{N}(\mu_j, \sigma_{\eta_j}^2) \\ &\propto \exp\left(-\frac{1}{2\sigma_{\eta_j}^2}(h_j(1) - \mu_j)^2\right). \end{aligned} \quad (4.27)$$

- Para $h_j(t)$, $t = 2, 3, \dots, N$ tomamos

$$\begin{aligned} P(h_j(t) | h_j(t-1)) &= \mathcal{N}(\mu_j + \phi_j(h_j(t-1) - \mu_j), \sigma_{\eta_j}^2) \\ &\propto \exp\left(-\frac{1}{2\sigma_{\eta_j}^2}(h_j(t) - (\mu_j + \phi_j(h_j(t-1) - \mu_j)))^2\right). \end{aligned} \quad (4.28)$$

4.5.4. Densidad *a posteriori* de φ dada \mathbf{D}

Una vez que se han calculado las funciones en las que se ha descompuesto la densidad *a posteriori*, es decir, la función de verosimilitud del modelo (4.17), la condicional de las variables latentes (4.21) y la *a priori* conjunta de los parámetros (4.26), sólo falta sustituirlas en (4.16). Así llegamos a la expresión:

$$\begin{aligned} P(\varphi | \mathbf{D}) &\propto L(\mathbf{D} | \varphi)P(\mathbf{h} | \Theta)P(\Theta) \\ &\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^N \sum_{j=1}^K (h_j(t) + e^{-h_j(t)}[Y_j(t)]^2)\right\} \\ &\quad \times \prod_{j=1}^K (\sigma_{\eta_j}^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2}(h_j(1) - \mu_j)^2\right\} \\ &\quad \times \prod_{t=2}^N \prod_{j=1}^K (\sigma_{\eta_j}^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2}[h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2\right\} \\ &\quad \times \prod_{j=1}^K e^{-\frac{1}{2f_j}(\mu_j - e_j)^2} e^{-\frac{1}{2b_j}(\phi_j - a_j)^2} (\sigma_{\eta_j}^2)^{1-c_j} e^{-d_j/\sigma_{\eta_j}^2}. \end{aligned} \quad (4.29)$$

A partir de (4.29) debemos estimar cada uno de los parámetros del modelo, como la esperanza condicional del parámetro dados los datos. Considerando la forma y la dimensionalidad de la densidad *a posteriori*, la estimación los parámetros se hará haciendo uso de los métodos Monte Carlo vía cadenas de Markov. En la siguiente sección se calculan las densidades marginales condicionales completas que se utilizaran en la simulación por métodos Monte Carlo vía cadenas de Markov.

4.6. Densidades marginales condicionales completas *a posteriori*

Para muestrear valores de la densidad *a posteriori* de φ se usaremos el método del **Muestreo de Gibbs**, para lo cual es necesario establecer las **densidades *a posteriori* marginales condicionales completas** de cada parámetro, también conocidas como **densidades de Gibbs**.

4.6.1. Densidades marginales condicionales completas *a posteriori* de las variables latentes

En el modelo VE las variables latentes tienen distintas expresiones dependiendo de si $t = 1$ ó $t = 2, \dots, N$. Sin embargo, como a continuación se verá, en ambos casos **no tienen densidad condicional completa con forma conocida (estándar)**. A partir de la expresión de la densidad *a posteriori* (4.29) se obtiene la densidad completa de $h_j(1)$ para $j = 1, \dots, K$:

$$\begin{aligned} \pi(h_j(1) \mid \varphi - \{h_j(1)\}, Y_j(1)) &\propto L(\mathbf{D} \mid \varphi)P(\mathbf{h} \mid \Theta) \\ &\propto \exp \left\{ -\frac{1}{2}(h_j(1) + e^{-h_j(1)}[Y_j(1)]^2) \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2}(h_j(1) - \mu_j)^2 \right\} \\ &\propto \Psi_{h_j(1)}(h_j(1), Y_j(1)) \cdot \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2}(h_j(1) - \mu_j)^2 \right\}, \end{aligned}$$

donde:

$$\Psi_{h_j(1)}(h_j(1), Y_j(1)) = \exp \left\{ -\frac{1}{2}(h_j(1) + e^{-h_j(1)}[Y_j(1)]^2) \right\}. \quad (4.30)$$

La notación $\boldsymbol{\varphi} - \{h_j(1)\}$ significa todo el vector de parámetros $\boldsymbol{\varphi}$ excepto la coordenada $h_j(1)$.

De manera análoga para $t = 2, \dots, N$ la densidad marginal condicional completa de $h_j(t)$ dado $h_j(t-1)$ para $j = 1, \dots, K$ es:

$$\begin{aligned} \pi(h_j(t) \mid \boldsymbol{\varphi} - \{h_j(t)\}, Y_j(t)) &\propto L(\mathbf{D} \mid \boldsymbol{\varphi})P(\mathbf{h} \mid \boldsymbol{\Theta}) \\ &\propto \exp \left\{ -\frac{1}{2}(h_j(t) + e^{-h_j(t)}[Y_j(t)]^2) \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2} [h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2 \right\} \\ &\propto \Psi_{h_j(t)}(h_j(t), Y_j(t)) \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2} [h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2 \right\}, \end{aligned}$$

donde:

$$\Psi_{h_j(t)}(h_j(t), Y_j(t)) = \exp \left\{ -\frac{1}{2}(h_j(t) + e^{-h_j(t)}[Y_j(t)]^2) \right\}. \quad (4.31)$$

La notación $\boldsymbol{\varphi} - \{h_j(t)\}$, $t = 2, \dots, N$ significa todo el vector de parámetros $\boldsymbol{\varphi}$ excepto la coordenada $h_j(t)$.

Debido a que **las densidades condicionales completas de $h_j(1)$ y $h_j(t)$ dado $h_j(t-1)$ no poseen forma estándar se usará un paso Metropolis-Hastings para cada simulación de Gibbs.** En el paso Metropolis-Hastings, se usa como densidad propuesta la densidad *a priori* de cada variable latente, dadas por las expresiones (4.27) para $h_j(1)$, y por (4.28) para $h_j(t)$ dado $h_j(t-1)$.

De esta forma, para la simulación de $h_j(1)$, si se denotará por x_{t-1} al valor obtenido de la simulación anterior y la densidad propuesta es una $\mathcal{N}(\mu_j, \sigma_{\eta_j}^2)$ y y es el valor simulado del valor propuesto. La probabilidad de aceptación ρ del paso Metropolis-Hastings; dada por la expresión (3.8), está dada por:

$$\begin{aligned}
 \rho(x_{t-1}, y) &= \min \left\{ \frac{\pi(y \mid \varphi - \{h_j(1)\}, Y_j(t)) \cdot e^{-\frac{1}{2\sigma_{\eta_j}^2}(x_{t-1}-\mu_j)^2}}{\pi(x_{t-1} \mid \varphi - \{h_j(1)\}, Y_j(t)) \cdot e^{-\frac{1}{2\sigma_{\eta_j}^2}(y-\mu_j)^2}}, 1 \right\} \\
 &= \min \left\{ \frac{\Psi_{h_j(1)}(y, Y_j(1))}{\Psi_{h_j(1)}(x_{t-1}, Y_j(1))}, 1 \right\}. \tag{4.32}
 \end{aligned}$$

De manera análoga, para la simulación del valor $h_j(t)$ dado $h_j(t-1)$, $t = 2, \dots, N$, si x_{t-1} es el valor de la simulación anterior y usamos como densidad propuesta una $\mathcal{N}(\mu_j + \phi_{jj}(h_j(t-1) - \mu_j), \sigma_{\eta_j}^2)$ y y es el valor simulado de la densidad propuesta entonces ρ el cociente de prueba del paso Metropolis-Hastings; dado por la expresión (3.8), toma la forma:

$$\begin{aligned}
 \rho(x_{t-1}, y) &= \min \left\{ \frac{\pi(y \mid \varphi - \{h_j(t)\}, Y_j(t)) \cdot e^{-\frac{1}{2\sigma_{\eta_j}^2}(x_{t-1}-\mu_j-\phi_j(h_j(t-1)-\mu_j))^2}}{\pi(x_{t-1} \mid \varphi - \{h_j(t)\}, Y_j(t)) \cdot e^{-\frac{1}{2\sigma_{\eta_j}^2}(y-\mu_j-\phi_j(h_j(t-1)-\mu_j))^2}}, 1 \right\} \\
 &= \min \left\{ \frac{\Psi_{h_j(t)}(y, Y_j(t))}{\Psi_{h_j(t)}(x_{t-1}, Y_j(t))}, 1 \right\}. \tag{4.33}
 \end{aligned}$$

Como se observa para la simulación de $h_j(t)$, $t = 1, 2, \dots, N$ el cociente de prueba del paso Metropolis-Hastings se forma con las funciones $\Psi_{h_j(1)}$ y $\Psi_{h_j(t)}$.

A continuación se obtendrán **las densidades condicionales completas a posteriori de los parámetros μ_j, ϕ_j y $\sigma_{\eta_j}^2$ que a diferencia de las de las variables latentes dados los datos, son de forma estándar.**

4.6.2. Densidades marginales condicionales completas a posteriori de μ_j, ϕ_j y σ_{η_j}

Partiendo de (4.29) calculamos las densidades marginales condicionales completas descartando los parámetros que no dependan de la variable de interés.

Proposición 4.6.1 *Densidad marginal condicional completa a posteriori de μ_j . Supongamos que se cumplen las características de la expresión (4.29), entonces la densidad condicional completa a posteriori de μ_j para $j = 1, \dots, K$ es proporcional a una $\mathcal{N}(\frac{B}{A}, \frac{1}{A})$, donde:*

$$\blacksquare A = \frac{1}{\sigma_{\eta_j}^2} \left[1 + (N-1)(1 - \phi_j)^2 \right] + \frac{1}{f_j}$$

$$\blacksquare B = \frac{1}{\sigma_{\eta_j}^2} \left[h_j(1) + (1 - \phi_j) \sum_{t=2}^N (h_j(t) - \phi_j h_j(t-1)) \right] + \frac{e_j}{f_j}.$$

Demostración:

$$\begin{aligned} \pi(\mu_j \mid \varphi - \{\mu_j\}, Y_j(t)) &\propto \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2} (h_j(1) - \mu_j)^2 \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N [h_j(t) - \mu_j - \phi_j (h_j(t-1) - \mu_j)]^2 \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2f_j} (\mu_j - e_j)^2 \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2} (h_j(1)^2 + \mu_j^2 - 2h_j(1)\mu_j) \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N [(h_j(t) - \phi_j h_j(t-1)) - \mu_j(1 - \phi_j)]^2 \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2f_j} (\mu_j^2 + e_j^2 - 2\mu_j e_j) \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2} (h_j(1)^2 + \mu_j^2 - 2h_j(1)\mu_j) \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N [(1 - \phi_j)^2 \mu_j^2 + (h_j(t) - \phi_j h_j(t-1))^2 \right. \\ &\quad \left. - 2(1 - \phi_j)(h_j(t) - \phi_j h_j(t-1))\mu_j] \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2f_j} (\mu_j^2 + e_j^2 - 2e_j \mu_j) \right\}. \end{aligned}$$

Removiendo los términos constantes con respecto a μ_j , se tiene

$$\begin{aligned}
\pi(\mu_j \mid \varphi - \{\mu_j\}, Y_j(t)) &\propto \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2}(\mu_j^2 - 2h_j(1)\mu_j)\right\} \\
&\times \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N [(1 - \phi_j)^2 \mu_j^2 - 2(1 - \phi_j)(h_j(t) - \phi_j h_j(t-1))\mu_j]\right\} \\
&\times \exp\left\{-\frac{1}{2f_j}(\mu_j^2 - 2e_j\mu_j)\right\} \\
&\propto \exp\left\{\left(-\frac{1}{2\sigma_{\eta_j}^2} - \frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N (1 - \phi_j)^2 - \frac{1}{2f_j}\right) \mu_j^2\right. \\
&\quad \left.- 2\left(-\frac{h_j(1)}{2\sigma_{\eta_j}^2} - \frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N (1 - \phi_j)(h_j(t) - \phi_j h_j(t-1)) - \frac{e_j}{2f_j}\right) \mu_j\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma_{\eta_j}^2} + \frac{1}{\sigma_{\eta_j}^2} \sum_{t=2}^N (1 - \phi_j)^2 + \frac{1}{f_j}\right) \mu_j^2\right.\right. \\
&\quad \left.\left.- 2\left(\frac{h_j(1)}{\sigma_{\eta_j}^2} + \frac{1}{\sigma_{\eta_j}^2} \sum_{t=2}^N (1 - \phi_j)(h_j(t) - \phi_j h_j(t-1)) + \frac{e_j}{f_j}\right) \mu_j\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma_{\eta_j}^2} \left[1 + \sum_{t=2}^N (1 - \phi_j)^2\right] + \frac{1}{f_j}\right) \mu_j^2\right.\right. \\
&\quad \left.\left.- 2\left(\frac{1}{\sigma_{\eta_j}^2} \left[h_j(1) + \sum_{t=2}^N (1 - \phi_j)(h_j(t) - \phi_j h_j(t-1))\right] + \frac{e_j}{f_j}\right) \mu_j\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[A\mu_j^2 - 2B\mu_j\right]\right\}.
\end{aligned}$$

Completando cuadrados llegamos a la expresión

$$\begin{aligned}\pi(\mu_j \mid \varphi - \{\mu_j\}, Y_j(t)) &\propto \exp\left\{-\frac{1}{2}\left[A\mu_j^2 - 2B\mu_j\right]\right\} \\ &= \exp\left\{-\frac{A}{2}\left[\mu_j^2 - 2\frac{B}{A}\mu_j\right]\right\} \\ &\propto \exp\left\{-\frac{A}{2}\left[\mu_j - \frac{B}{A}\right]^2\right\},\end{aligned}$$

por lo tanto, μ_j dado $(\varphi - \{\mu_j\}, Y_j(t))$ tiene una distribución

$$\mathcal{N}\left(\frac{B}{A}, \frac{1}{A}\right). \blacksquare \quad (4.34)$$

Proposición 4.6.2 Densidad marginal condicional completa a posteriori de ϕ_j . Supongamos que se cumplen las características de la expresión (4.29), entonces la densidad condicional completa a posteriori de ϕ_j para $j = 1, \dots, K$ es proporcional a una $\mathcal{N}\left(\frac{D}{C}, \frac{1}{C}\right)$, donde:

- $C = \left(\frac{1}{\sigma_{\eta_j}^2} \sum_{t=2}^N (h_j(t-1) - \mu_j)^2 + \frac{1}{b_j}\right)$
- $D = \left(\frac{1}{\sigma_{\eta_j}^2} \sum_{t=2}^N (h_j(t) - \mu_j)(h_j(t-1) - \mu_j) + \frac{a_j}{b_j}\right).$

Demostración:

$$\begin{aligned}\pi(\phi_j \mid \varphi - \{\phi_j\}, Y_j(t)) &\propto \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N [h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2\right\} \\ &\quad \times \exp\left\{-\frac{1}{2b_j}(\phi_j - a_j)^2\right\} \\ &= \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N [(h_j(t) - \mu_j)^2 - 2(h_j(t) - \mu_j)\phi_j(h_j(t-1) - \mu_j) \right. \\ &\quad \left. + \phi_j^2(h_j(t-1) - \mu_j)^2]\right\} \exp\left\{-\frac{1}{2b_j}(\phi_j^2 + a_j^2 - 2\phi_j a_j)\right\}.\end{aligned}$$

Removiendo los términos constantes con respecto a ϕ_j , se tiene

$$\begin{aligned}
 \pi(\phi_j \mid \varphi - \{\phi_j\}, Y_j(t)) &\propto \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2} \left[\phi_j^2 \sum_{t=2}^N (h_j(t-1) - \mu_j)^2 \right. \right. \\
 &\quad \left. \left. - 2\phi_j \sum_{t=2}^N (h_j(t) - \mu_j)(h_j(t-1) - \mu_j) \right] \right\} \exp\left\{-\frac{1}{2b_j} (\phi_j^2 - 2\phi_j a_j)\right\} \\
 &= \exp\left\{\left(-\frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N (h_j(t-1) - \mu_j)^2 - \frac{1}{2b_j}\right) \phi_j^2 \right. \\
 &\quad \left. - 2\left(-\frac{1}{2\sigma_{\eta_j}^2} \sum_{t=2}^N (h_j(t) - \mu_j)(h_j(t-1) - \mu_j) - \frac{1}{2b_j} a_j\right) \phi_j\right\} \\
 &= \exp\left\{-\frac{1}{2} \left[\left(\frac{1}{\sigma_{\eta_j}^2} \sum_{t=2}^N (h_j(t-1) - \mu_j)^2 + \frac{1}{b_j} \right) \phi_j^2 \right. \right. \\
 &\quad \left. \left. - 2 \left(\frac{1}{\sigma_{\eta_j}^2} \sum_{t=2}^N (h_j(t) - \mu_j)(h_j(t-1) - \mu_j) + \frac{a_j}{b_j} \right) \phi_j \right] \right\} \\
 &= \exp\left\{-\frac{1}{2} [C\phi_j^2 - 2D\phi_j]\right\}.
 \end{aligned}$$

Completando cuadrados llegamos a la expresión:

$$\begin{aligned}
 \pi(\phi_j \mid \varphi - \{\phi_j\}, Y_j(t)) &\propto \exp\left\{-\frac{1}{2} [C\phi_j^2 - 2D\phi_j]\right\} \\
 &= \exp\left\{-\frac{C}{2} \left[\phi_j^2 - 2\frac{D}{C}\phi_j\right]\right\} \\
 &\propto \exp\left\{-\frac{C}{2} \left[\phi_j - \frac{D}{C}\right]^2\right\},
 \end{aligned}$$

por lo tanto, ϕ_j dado $(\varphi - \{\phi_j\}, Y_j(t))$ tiene función de densidad

$$\mathcal{N}\left(\frac{D}{C}, \frac{1}{C}\right). \blacksquare \tag{4.35}$$

Proposición 4.6.3 *Densidad marginal condicional completa a posteriori de $\sigma_{\eta_j}^2$.* Supongamos que se cumplen las características de la expresión (4.29), entonces la densidad marginal condicional completa a posteriori de $\sigma_{\eta_j}^2$ para $j = 1, \dots, K$ es proporcional a una $\mathcal{GI}\left(\frac{N}{2} + c_j, \frac{1}{2}(E + F + 2d_j)\right)$, donde:

- $E = (h_j(1) - \mu_j)^2$
- $F = \sum_{t=2}^N [h_j(t) - \mu_j - \phi_j(h_j(t-1) - \mu_j)]^2$.

Demostración:

$$\begin{aligned} \pi(\sigma_{\eta_j}^2 \mid \varphi - \{\sigma_{\eta_j}^2\}, Y_j(t)) &\propto (\sigma_{\eta_j}^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2} E\right\} [(\sigma_{\eta_j}^2)^{-1/2}]^{N-1} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2} F\right\} e^{-d_j/\sigma_{\eta_j}^2} / (\sigma_{\eta_j}^2)^{c_j-1} \\ &\propto (\sigma_{\eta_j}^2)^{1-(N/2+c_j)} \exp\left\{-\frac{1}{2\sigma_{\eta_j}^2} (E + F + 2d_j)\right\}, \end{aligned}$$

por lo tanto, $\sigma_{\eta_j}^2$ dado $(\varphi - \{\phi_j\}, Y_j(t))$ tiene función de densidad

$$\mathcal{GI}\left(\frac{N}{2} + c_j, \frac{1}{2}(E + F + 2d_j)\right). \blacksquare \quad (4.36)$$

En todos los casos, a diferencia de las variables latentes, las distribuciones marginales condicionales completas poseen formas conocidas de las cuales fácilmente se pueden obtener muestras aleatorias en cada paso de Gibbs.

4.7. Pseudocódigo

Para cada una de las K regiones $j = 1, 2, \dots, K$, se obtienen el número de iteraciones deseadas (ID) como sigue:

1. Calcular a

$$\Psi_{h_j(1)}, \Psi_{h_j(2)}, \dots, \Psi_{h_j(N)},$$

usando las ecuaciones (4.30) y (4.31).

2. Proporcionar los valores de los hiperparámetros, a_j y b_j , c_j y d_j e_j y f_j .
3. Tomar las densidades *a priori* de los parámetros como:
 - ◇ $\mathcal{N}(e_j, f_j)$, para μ_j .
 - ◇ $\mathcal{N}(a_j, b_j)$, para ϕ_j .
 - ◇ $\mathcal{GI}(c_j, d_j)$, para μ_j .
4. Ingresar la base de datos observados.
5. Primera iteración del Muestreo de Gibbs, es decir para $n = 1$.

a) Simular $h_j(1)$ con un paso de Metropolis-Hastings¹.

- 1) Definir un valor inicial $x^{(0)}$, generándolo de una $\mathcal{N}(\mu_j, \sigma_{\eta_j}^2)$.
- 2) Generar un número aleatorio $u \sim U(0, 1)$, que se usará en el criterio de decisión del paso Metropolis-Hastings.
- 3) Generar un valor ‘‘ y ’’ que provenga de la densidad propuesta, que en este caso es la densidad de $h_j(1)$ según (4.30).
- 4) Calcular la probabilidad de aceptación, según la expresión (3.8) y que en términos del modelo considerado aquí está dada por la expresión (4.32).

$$\rho = \min\left\{\frac{\Psi_{h_j(1)}(y, Y_j(1))}{\Psi_{h_j(1)}(x^{(0)}, Y_j(1))}, 1\right\}.$$

- 5) Si $u \leq \rho$, entonces $x^{(1)} = y$, en caso contrario $x^{(1)} = x^{(0)}$.

b) Simular $h_j(t)$ para $t = 2, 3, \dots, N$.

- 1) Definir un valor inicial $x^{(0)}$, generándolo de una $\mathcal{N}(h_j(t-1), \sigma_{\eta_j}^2)$.
- 2) Generar un número aleatorio $u \sim U(0, 1)$, que se usará en el criterio de decisión del Metropolis-Hastings.
- 3) Generar un valor ‘‘ y ’’ que provenga de la densidad propuesta, que en este caso es la densidad de $h_j(t)$ según (4.31).
- 4) Calcular la probabilidad de aceptación según la expresión (3.8) y que en términos del modelo considerado aquí está dada por la expresión (4.33).

$$\rho = \min\left\{\frac{\Psi_{h_j(t)}(y, Y_j(t))}{\Psi_{h_j(t)}(x^{(0)}, Y_j(t))}, 1\right\}.$$

- 5) Si $u \leq \rho$, entonces $x^{(1)} = y$, en caso contrario $x^{(1)} = x^{(0)}$.

c) Simular μ_j de su densidad condicional completa dada por la expresión (4.34), utilizando los valores actualizados de $h(t)$, $t = 1, 2, \dots, N$ obtenidos en el paso previo del muestreo de Gibbs.

¹Note que en el modelo de volatilidad estocástica las variables latentes tienen una formulación para $t = 1$ y otra distinta para $t = 2, 3, \dots, N$, por lo que la programación es diferente en cada caso.

- d*) Simular ϕ_j de su densidad condicional completa dada por la expresión (4.35), utilizando el valor actualizado de μ_j .
 - e*) Simular $\sigma_{\eta_j}^2$ de su densidad condicional completa dada por la expresión (4.36), utilizando el valor actualizado de ϕ_j .
6. Para las siguientes $n = 2, 3, \dots, ID$ iteraciones del Muestreo de Gibbs.
- a*) Actualizar las simulaciones de las variables latentes y de los parámetros de la iteración anterior.
 - b*) Simular $h_j(t)$ para $t = 1, 2, \dots, N$ con un paso de Metropolis-Hastings como en $n = 1$, pero definiendo el valor inicial $x^{(0)}$ como el valor de $h_j(t)$ que se aceptó en la iteración anterior, es decir, en la iteración $(n - 1)$.
 - c*) Simular μ_j , utilizando el valor actualizado de $h(t)$ con $t = 1, 2, \dots, N$.
 - d*) Simular ϕ_j , utilizando el valor actualizado de μ_j .
 - e*) Simular $\sigma_{\eta_j}^2$, utilizando el valor actualizado de ϕ_j .
 - f*) Repetir el procedimiento de *a*) - *e*) para cada iteración.

Capítulo 5

Aplicación

5.1. Introducción

En este capítulo se utilizan los métodos Monte Carlo vía cadenas de Markov específicamente muestreo de Gibbs y Metropolis-Hastings dentro del muestreo de Gibbs para estimar los parámetros del modelo de volatilidad estocástica para los datos de concentraciones de ozono en las cinco regiones de la Zona Metropolitana de la Ciudad de México (ZMCM).

Tanto el análisis bayesiano del modelo, que se presentó en el capítulo anterior, como los resultados de la aplicación a datos de ozono en la ZMCM, que a continuación se exponen y el código del programa dieron origen al artículo *A Gibbs Sampling Algorithm to Estimate the Parameters of a Volatility Model: An Application to Ozone Data*. De Jesús *et al.* (2012).

5.2. Estimación de parámetros del modelos VE

El objetivo práctico de este trabajo es estudiar el comportamiento del contaminante ozono en la ZMCM, en el periodo de 01 de Enero de 2008 al 31 de Diciembre de 2010, haciendo uso del modelo de volatilidad estocástica presentado en el Capítulo 4. Se tomará en cuenta la división espacial considerada por las autoridades ambientales de la Ciudad de México, es decir, las regiones Noreste (NE), Noroeste(NW), Centro (CE), Sureste (SE) y Suroeste (SW).

Se hizo uso de la base de datos del Sistema de Monitoreo Atmosférico (disponible en página web <http://www.calidadaire.df.gob.mx/calidadaire/productos/basesdedatos/bases.php>), de los máximos diarios de ozono, medidos en partes por millón (ppm), registrados en cinco regiones en las cuales está dividida la

Ciudad de México.

El dato de la concentración máxima de ozono de un día se calcula de la siguiente manera. Las estaciones miden minuto a minuto la concentración de ozono y se obtiene un promedio por hora, el dato del máximo diario para cada región es el máximo de los promedios de todas las estaciones en la región considerada registrados durante las 24 horas del día.

La Figura 5.1 muestra los máximos diarios de ozono en cada región a partir del 01 de Enero de 2000 hasta el 31 de diciembre de 2010. Se puede observar una tendencia decreciente en un primer periodo, aproximadamente entre el año 2000 y 2006, y un cambio en la variabilidad observable en las concentraciones posteriores al año 2007.

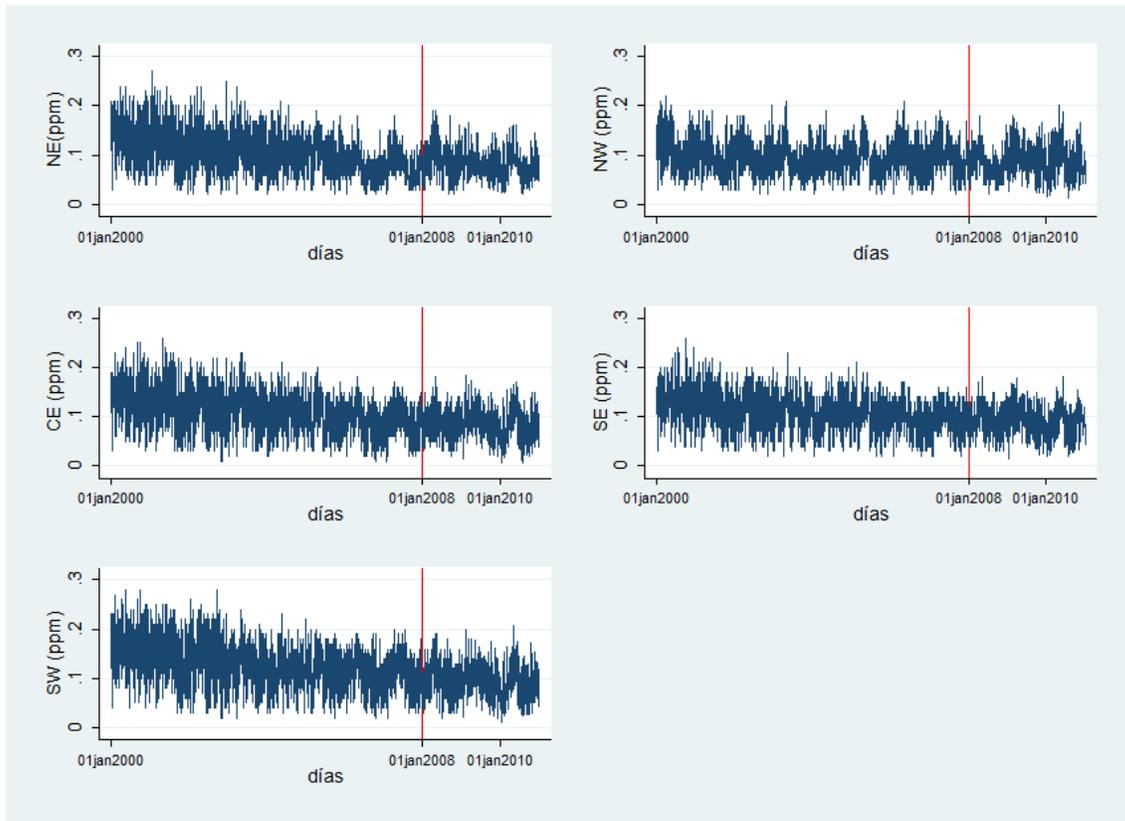


Figura 5.1: Máximos diarios de ozono(ppm) en Regiones de la ZMCM en el periodo 01 de Enero de 2000 a 31 de diciembre 2010.

En el Cuadro 5.1 se presentan la media y la desviación estándar de las series de tiempo de los máximos diarios de ozono por región. Se puede observar que el promedio anual del 2010 es mucho menor al registrado en el año 2000 en todas las

Cuadro 5.1: Estadísticas de los máximos de ozono por región.

periodo	Estadísticas	NE	NW	CE	SE	SW
2000	Media	0.1327	0.1054	0.1391	0.1344	0.1589
	DE	0.0456	0.0370	0.0426	0.0392	0.0477
2010	Media	0.0778	0.0829	0.0825	0.0859	0.0926
	DE	0.0274	0.0338	0.0325	0.0293	0.0350
2000-2006	Media	0.1078	0.0939	0.1140	0.1125	0.1302
	DE	0.0432	0.0324	0.0425	0.0380	0.0467
2008-2010	Media	0.0829	0.0851	0.0876	0.0911	0.0977
	DE	0.0286	0.0303	0.0306	0.0294	0.0341

regiones. De hecho el promedio anual en el año 2000 aún era superior a la norma de protección a la salud que es de 0.11 ppm, mientras que para el año 2010 en todos los casos estuvo por debajo de la misma.

La Figura 5.2 presenta los máximos diarios de ozono en cada región durante el periodo de tiempo considerada en el trabajo, es decir, de 01 de Enero de 2008 al 31 de Diciembre de 2010.

Se calcularon las medias semanales (ver Figura 5.3) y posteriormente se calcularon los log-retornos semanales de cada una de las cinco regiones, según la expresión

$$y_j(t) = \ln \left[\frac{Z_j(t)}{Z_j(t-1)} \right],$$

los cuales, constituyen el conjunto de datos a modelar y que podemos observar en la Figura 5.4

Así, en la notación del Modelo, el número de regiones es $K = 5$ donde $j = 1, 2, \dots, K$ corresponde a la región NE, NW, CE, SE y SW, respectivamente. El total de observaciones diarias en los 3 años a estudiar es $T = 1092$ y el total de log retornos es $N = 155$. Así para $j = 1, 2, \dots, K$

- $X_j(t)$ con $t = 1, 2, \dots, T$ corresponde a la serie de tiempo de los máximos diarios en cada región en el periodo del 01 de Enero de 2008 al 31 de Diciembre de 2010.
- $Z_j(t)$ con $t = 1, 2, \dots, (N + 1)$ y corresponde a la serie de tiempo de los promedios semanales de ozono en cada región.
- $Y_j(t)$ con $t = 1, 2, \dots, N$ y corresponde a la serie de tiempo de los log-retornos de ozono en cada región.

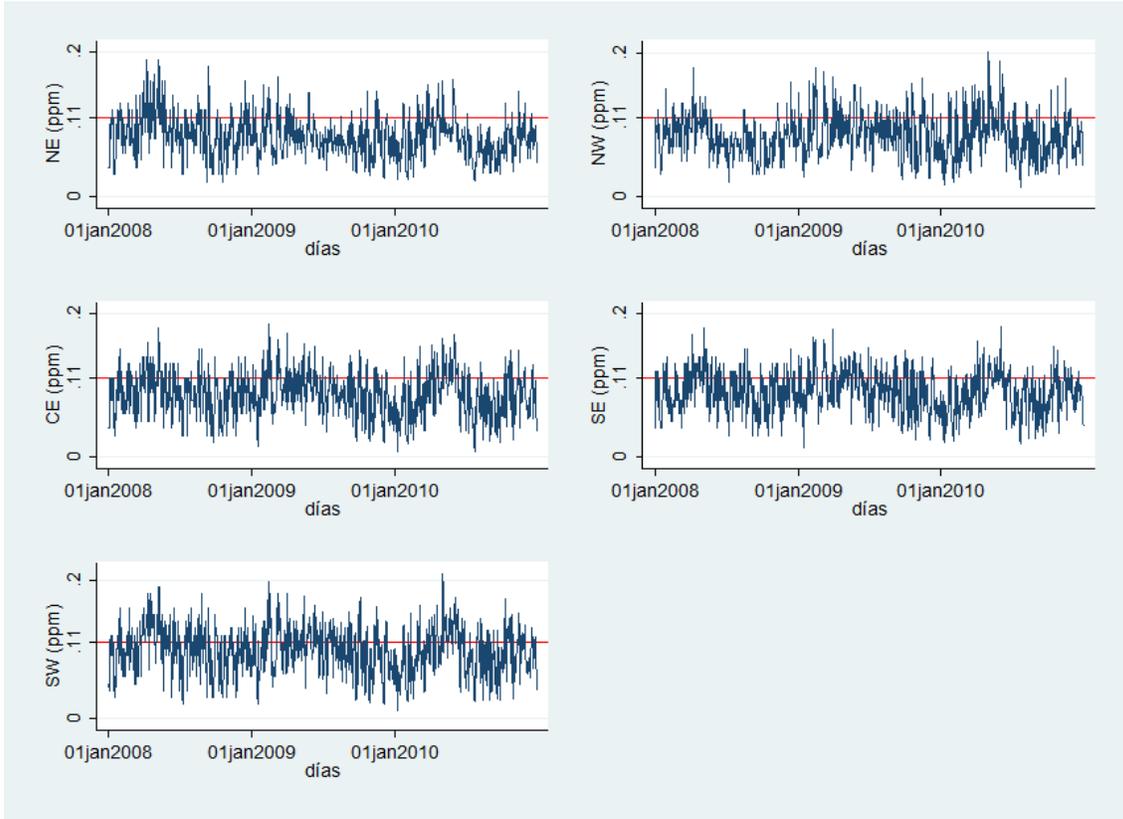


Figura 5.2: Máximos diarios de ozono(ppm) en Regiones de la ZMCM en el periodo 01 de Enero de 2008 a 31 de diciembre 2010.

Los detalles computacionales para la estimación bayesiana se describieron en la sección 4.7, del pseudocódigo, mismos que fue implementado en “R” y se puede consultar en el Apéndice C. Se usó Metropolis-Hastings dentro del muestreo de Gibbs para generar las simulaciones de las variables latentes, dado que sus densidades marginales completas no tienen una forma estándar. Para la simulación de los parámetros μ , ϕ y σ_{η}^2 se utilizó únicamente muestreo de Gibbs.

Recordar que hemos supuesto las siguientes distribuciones como distribuciones *a priori* de los parámetros:

$$P(\mu_j) = \mathcal{N}(e_j, f_j), P(\phi_j) = \mathcal{N}(a_j, b_j) \text{ y } P(\sigma_{\eta_j}^2) = \mathcal{GI}(c_j, d_j).$$

Los hiperparámetros de las densidades *a priori* asociadas a cada parámetro que se utilizaron en $j = 1, 2, \dots, 5$ fueron los siguientes.

- Para μ_j , $e_j = 0$, $f_j = 10$.

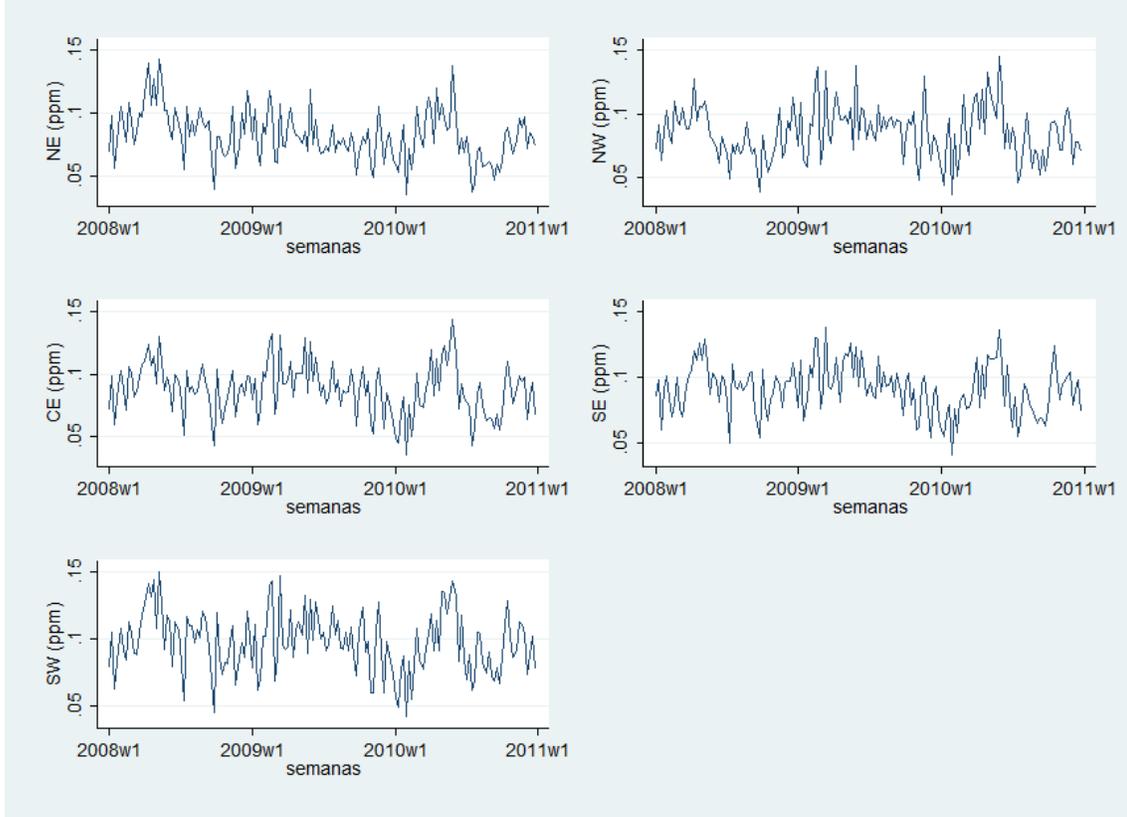


Figura 5.3: Promedios semanales de ozono por región de la ZMCM en el periodo 2008-2010. (La notación 2008w1 significa semana 1 del año 2008.)

- Para ϕ_j , $a_j = 0, b_j = 1$.
- Para $\sigma_{\eta_j}^2$, $c_j = 3, d_j = 3$.

Observación: Como se puede notar en el caso de ϕ_j se ha supuesto media cero y desviación estándar de 1, debido a que en el modelo de volatilidad estocástica, $|\phi_j| < 1$ para asegurar la estacionaridad de $\{h_j(t)\}_{t=1}^N$. En los casos de μ_j y $\sigma_{\eta_j}^2$ los valores de los hiperparámetros se han elegido de modo que las densidades *a priori* resulten no informativos y se observe una mezcla adecuada de los valores generados. Las estimaciones para todas las regiones se realizaron suponiendo estos mismos hiperparámetros.

Para el cálculo de cada uno de los parámetros *a posteriori*, $\mu_j, \phi_j, \sigma_{\eta_j}^2$, $j = 1, 2, \dots, 5$ se generaron tres cadenas de 21,000 iteraciones, cada una con valores iniciales diferentes. Los valores iniciales de la cadena 1 fueron $\mu_0 = 0.270, \phi_0 = 0.550, \sigma_{\eta_0}^2 = 1.293$, los de la cadena 2 fueron $\mu_0 = -0.57, \phi_0 = -0.095, \sigma_{\eta_0}^2 = 3.284$ y los de la cadena 3 fueron $\mu_0 = 0.874, \phi_0 = -0.447, \sigma_{\eta_0}^2 = 0.641$. Estos valores

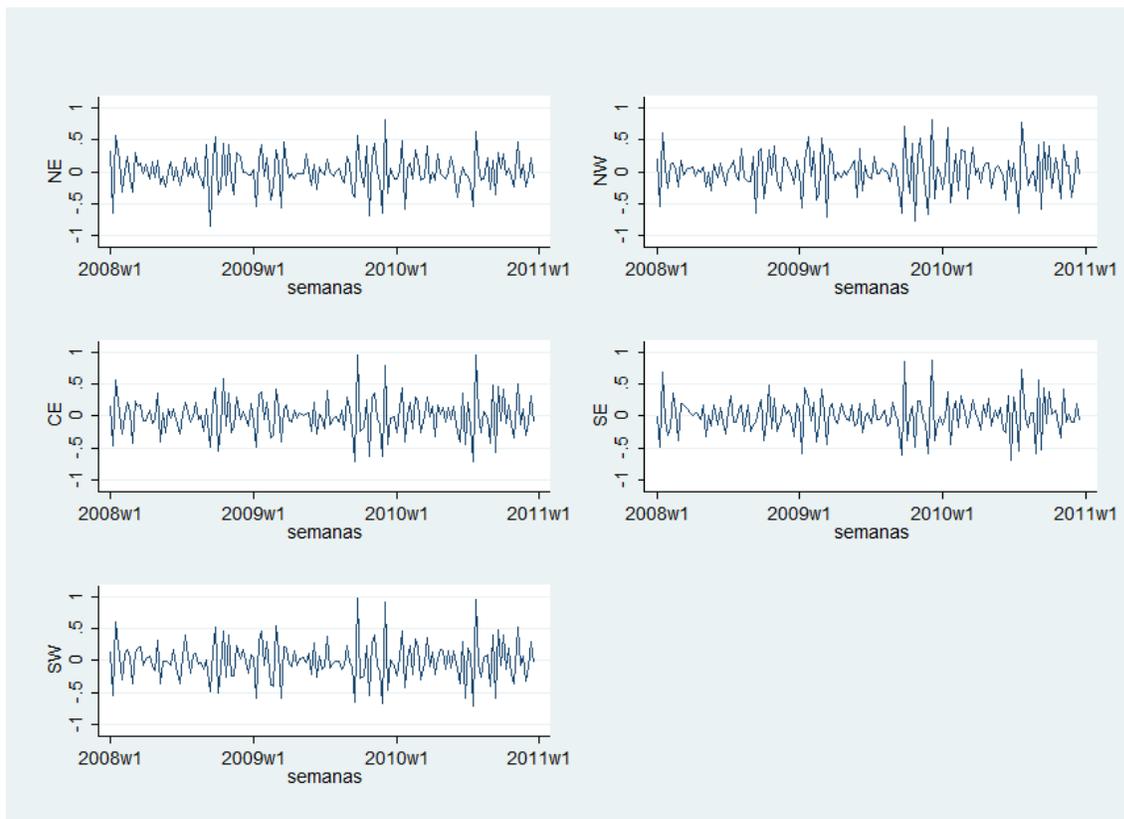


Figura 5.4: Log-retornos semanales de ozono por región de la ZMCM en el periodo 2008-2010. (La notación 2008w1 significa semana 1 del año 2008.)

iniciales serán los mismos para los datos de cada región.

Finalmente las estimaciones *a posteriori* se calcularon con las muestras de las tres cadenas, después de eliminar los valores generados durante un periodo de calentamiento de 2000 iteraciones en cada cadena y seleccionar valores muestreados cada quinto valor. Así los parámetros *a posteriori* fueron estimados a partir de una muestra de 11,400 iteraciones; 3,800 por cada una de las tres cadenas.

El Cuadro 5.2 muestra las estimaciones *a posteriori* de los parámetros del modelo para cada región junto con su correspondiente desviación estándar (DE) e intervalo de credibilidad al 95 %.

Las Figuras B.1 a B.5 contienen las gráficas de las muestras contra el número de iteración (Trace plot), los histogramas y las funciones de autocorrelación.

Cuadro 5.2: Estimaciones *a posteriori*

Región	Parámetro	Media	DE	Intervalo de credibilidad al 95 %
NE	μ	-2.864	0.174	-3.205 , -2.524
	ϕ	0.154	0.150	-0.140 , 0.448
	σ_η^2	0.837	0.281	0.287 , 1.387
NW	μ	-2.689	0.182	-3.046 , -2.332
	ϕ	0.163	0.152	-0.135 , 0.460
	σ_η^2	0.912	0.328	0.269 , 1.555
CE	μ	-2.689	0.167	-3.016 , -2.362
	ϕ	0.141	0.151	-0.155 , 0.437
	σ_η^2	0.757	0.255	0.257 , 1.256
SE	μ	-2.921	0.173	-3.260 , -2.582
	ϕ	0.135	0.143	-0.145 , 0.416
	σ_η^2	0.888	0.292	0.315 , 1.460
SW	μ	-2.668	0.171	-3.003 , -2.333
	ϕ	0.131	0.150	-0.163 , 0.425
	σ_η^2	0.839	0.287	0.277 , 1.401

A continuación se presenta los resultados de los diagnósticos de convergencia de las cadenas. Esto con el fin de verificar que la validez de las estimaciones.

5.3. Evaluación de la convergencia

Inspección gráfica. En las figuras B.1 a B.5 del apéndice B, se observan los valores generados contra el número de iteración (Trace plot), los histogramas de las cadenas y las gráficas de la función de autocorrelación (FAC) de las cadenas de los valores generados por región.

A partir de las gráficas de los valores generados y de los histogramas se puede concluir una mezcla adecuada en la cadena de cada parámetro en todas las regiones. Al observar las gráficas de las autocorrelaciones de los valores generados de cada región se evidencia independencia al interior de las cadenas que simulan las distribuciones de los parámetros μ y ϕ . En el caso de las gráficas de las autocorrelaciones de la cadena de σ_η^2 , la independencia no es tan clara. Sin embargo la prueba de Raftery–Lewis concluye que se satisface la hipótesis de independencia.

Resultados de los diagnósticos de convergencia. La evaluación de la convergencia se realizó utilizando el paquete CODA de “R”, específicamente se utilizaron las metodologías de Geweke, Raftery–Lewis, Heidelberger–Welch, y finalmente la de Gelman–Rubin. Las tres primeras metodologías evalúan la convergencia in-

dividual de cada cadena y la de Gelman–Rubin evalúa la convergencia de las tres cadenas generadas.

Los resultados de estos diagnósticos se presentan en los Cuadros A.1 al A.5, del apéndice A, donde se reporta:

- El z -score, en el diagnóstico de Geweke.
- El factor de dependencia I asociado a $q = 0.025$, $r = 0.005$, $s = 0.95$, $\epsilon_{converg} = 0.001$, en el diagnóstico de Raftery-Lewis.
- El p -value asociado a la hipótesis nula de que los valores muestreados provienen de una distribución estacionaria asociado a un nivel de significancia de 0.05, en la prueba de Heidelberger–Welch de estacionariedad.
- El *cociente* de la mitad del ancho del intervalo de confianza para la media al 95 % entre la media, la prueba se realizó con $\epsilon = .1$, en Heidelberger–Welch de "media ancho", y
- El estadístico R en el diagnóstico de Gelman-Rubin.

De la información contenida en los los Cuadros A.1 al A.5, se puede concluir que no hay evidencia estadística para suponer que las cadenas no hayan convergido, dado que:

- El diagnóstico de Geweke, el z - score en valor absoluto es menor a 2. Lo que implica que a un nivel de confianza de 95 %, se ha logrado la convergencia.
- En el diagnóstico de Raftery-Lewis, el factor de dependencia ($I = \frac{n_{min}}{n}$) siempre es cercano a 1 en las simulaciones de μ y ϕ . En el caso de las simulaciones de σ_η^2 el factor de dependencia (I) está alrededor de 2, lo cual aún es indicador de una mezcla adecuada y la autocorrelación que se aprecia en las gráficas se puede tolerar. El valor $I = 5$, es el valor a partir del cual falla la convergencia.
- En la prueba de Heidelberger–Welch donde la hipótesis nula de que los valores muestreados provienen de un proceso estacionario, son satisfactorios en todos los casos, ya que el p-value asociado a la prueba es mayor al nivel de significancia (0.05), con lo que en ningún caso se rechaza la hipótesis nula con una confianza del 95 %.
- En la prueba de Heidelberger–Welch de "mitad de ancho" los resultados indican que las simulaciones deben de generar estimaciones *a posteriori* de los parámetros que cumplen con el criterio de precisión de 0.1, ya que en todos los casos el *cociente* de la mitad del ancho del intervalo de confianza para la media al 95 % entre la media es menor a 0.1.

- Finalmente en la metodología de Gelman–Rubin, el estadístico R siempre es muy cercano a 1, con lo cual se evidencia la convergencia.

De esta forma, se puede concluir que las estimaciones del Cuadro 5.2 provienen de simulaciones sobre las cuales hay evidencia estadística suficiente para afirmar que han convergido al estado estacionario. Por lo tanto las podemos utilizar para estimar los parámetros del modelo de volatilidad estocástica para modelar la contaminación promedio semanal en las regiones de la ZMCM.

5.4. Ajuste y pronóstico.

Así, para obtener la serie de los log-retornos **estimados** en cada región $j = 1, \dots, 5$, se procedió de la siguiente manera:

Se generan las variables latentes $h_j(t)$ dado $h_j(t-1)$ de su distribución; es decir, al tiempo $t = 1$, $h_j(1)$ de la distribución $\mathcal{N}(\hat{\mu}_j, \hat{\sigma}_{\eta j}^2)$ y al tiempo $t = 2, \dots, N$, $h_j(t)$ dado $h(t-1)$ de la distribución $\mathcal{N}(\hat{\mu}_j + \hat{\phi}_j[h_j(t-1) - \hat{\mu}_j], \hat{\sigma}_{\eta j}^2)$.

Posteriormente se estimó las serie $\hat{\sigma}_j(t)$, la volatilidad de la serie $Y(t)$ según la ecuación (4.4). Las estimaciones de los promedios semanales $\hat{Z}_j(t)$ dado el pasado inmediato $Z_j(t-1)$ se obtuvieron usando las ecuaciones (4.1) y (4.3):

$$\hat{Z}_j(t) = Z_j(t-1) \exp\{\hat{\sigma}_j(t)\epsilon(t)\}, t = 2, \dots, N,$$

donde $\epsilon_j(t)$ se distribuye $\mathcal{N}(0, 1)$.

Finalmente los log-retornos estimados $\hat{Y}_j(t)$ se obtienen como:

$$\hat{Y}_j(t) = \ln \left[\frac{\hat{Z}_j(t)}{\hat{Z}_j(t-1)} \right] \text{ con } j = 1, 2, \dots, 5 \text{ y } t = 2, 3, \dots, N.$$

Las series de tiempo de los log-retornos del promedio semanal de ozono estimados y observados en cada región se muestran en las figuras 5.5 a 5.9. A partir de estas gráficas se puede observar el ajuste que se obtuvo a partir de la metodología implementada.

La utilidad principal del modelo es realizar un pronóstico. El cuadro 5.3 contiene el pronóstico y los datos observados de los **log-retornos** de las semanas de

enero del año 2011 en las cinco regiones de la ZMCM. En el Cuadro 5.4 se presenta el pronóstico y los datos observados de los **promedios semanales** del mes de enero del año 2011 en las cinco regiones de la ZMCM.

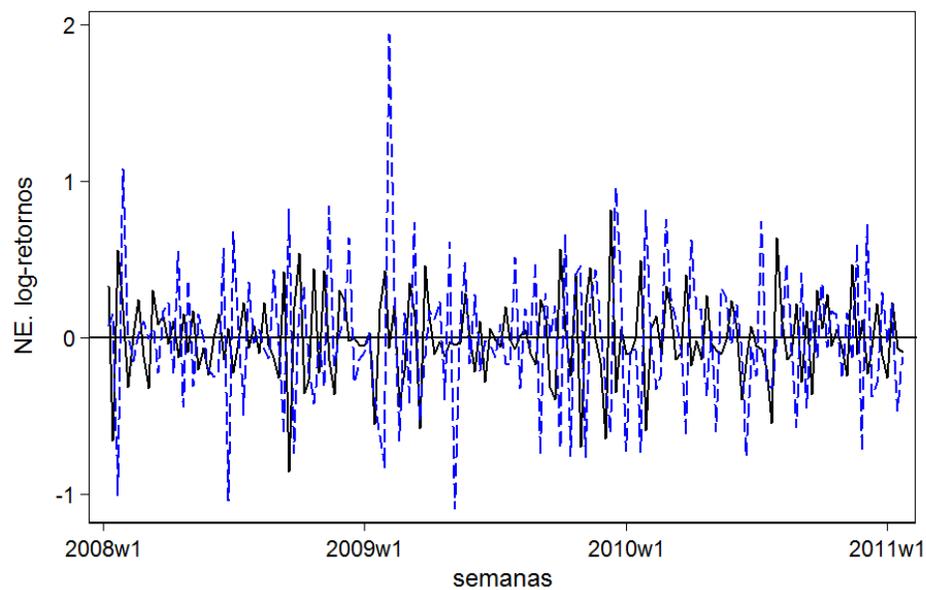
Cuadro 5.3: Log-retornos observados y estimados semanales de Ozono.

Enero 2011					
Observados					
Semana	NE	NW	CE	SE	SW
1	-0.251	-0.264	-0.415	-0.372	-0.392
2	0.224	0.287	0.380	0.360	0.302
3	-0.061	-0.163	-0.186	-0.233	-0.121
4	-0.089	-0.052	-0.063	-0.051	-0.099
Estimados					
Semana	NE	NW	CE	SE	SW
1	-0.042	-0.045	-0.045	-0.045	-0.040
2	0.224	0.254	0.234	0.248	0.224
3	-0.463	-0.529	-0.480	-0.501	-0.454
4	-0.123	-0.137	-0.130	-0.132	-0.118

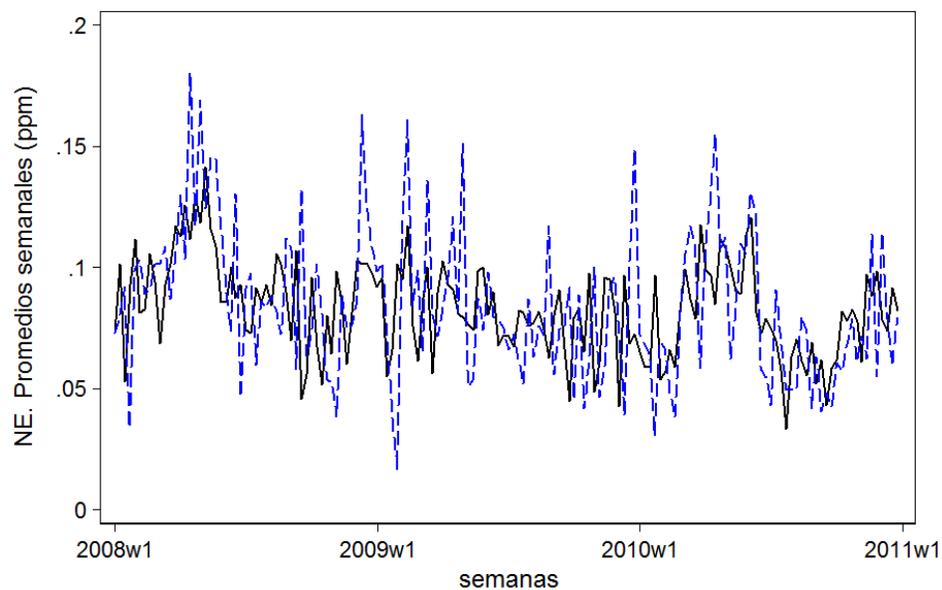
Cuadro 5.4: Promedios semanales observados y estimados semanales de Ozono.

Enero 2011					
Observados					
Semana	NE	NW	CE	SE	SW
1	0.064	0.063	0.057	0.063	0.069
2	0.080	0.084	0.084	0.085	0.098
3	0.076	0.071	0.070	0.076	0.078
4	0.069	0.068	0.065	0.069	0.074
Estimados					
Semana	NE	NW	CE	SE	SW
1	0.076	0.070	0.079	0.083	0.084
2	0.095	0.091	0.099	0.103	0.108
3	0.060	0.053	0.061	0.066	0.065
4	0.053	0.047	0.054	0.058	0.057

Recordemos que los log-retornos se aproximan a una tasa de crecimiento, por lo que se puede decir que el modelo pronostica un decremento aproximado en la primera semana de 2011 respecto a la semana 52 de 2010 de -4.2 %, -4.5 %, -4.5 %, -4.5 % y de -4.0 % en la región NE, NW, CE, SE y SW respectivamente. Los promedios semanales de ozono de la primera semana de Enero de 20011 pronosticados fueron 0.076, 0.070, 0.079, 0.083 y 0.084 en la región NE, NW, CE, SE y SW respectivamente. De los cuadros 5.3 y 5.4 vemos que el pronóstico del modelo fue mejor para los promedios semanales que para los log-retornos. A continuación se presentan las gráficas del ajuste de los log-retornos y promedios semanales por región.

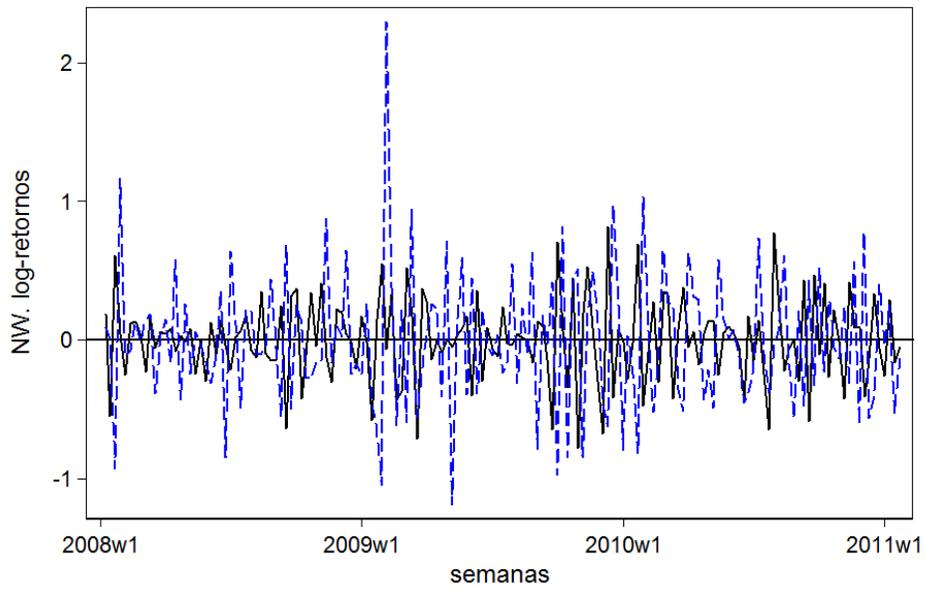


(a)

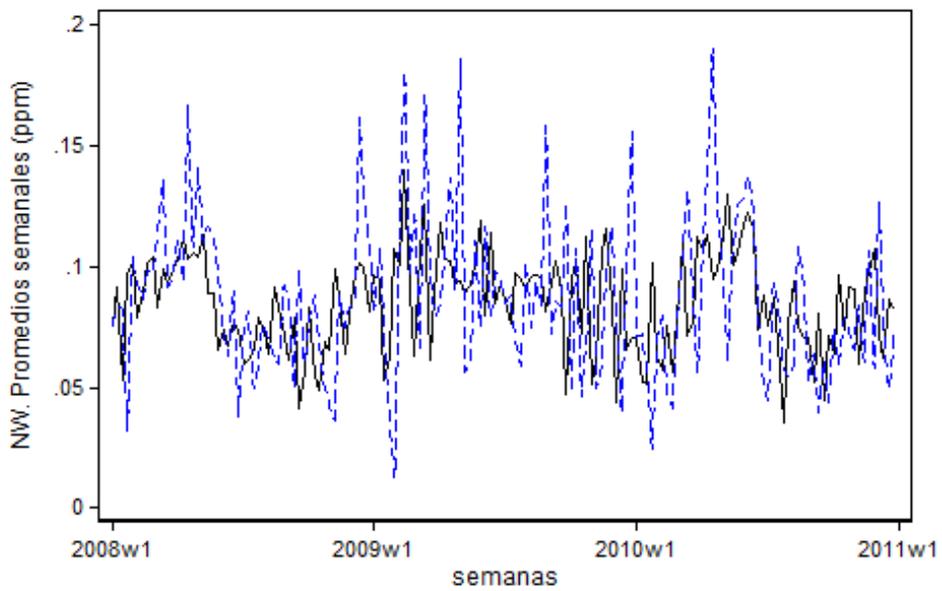


(b)

Figura 5.5: Región NE. En (a) log-retornos del promedio semanal de ozono observados (línea sólida) y estimados (línea punteada). En (b) los promedios semanales de ozono observados (línea sólida) y estimados (línea punteada).

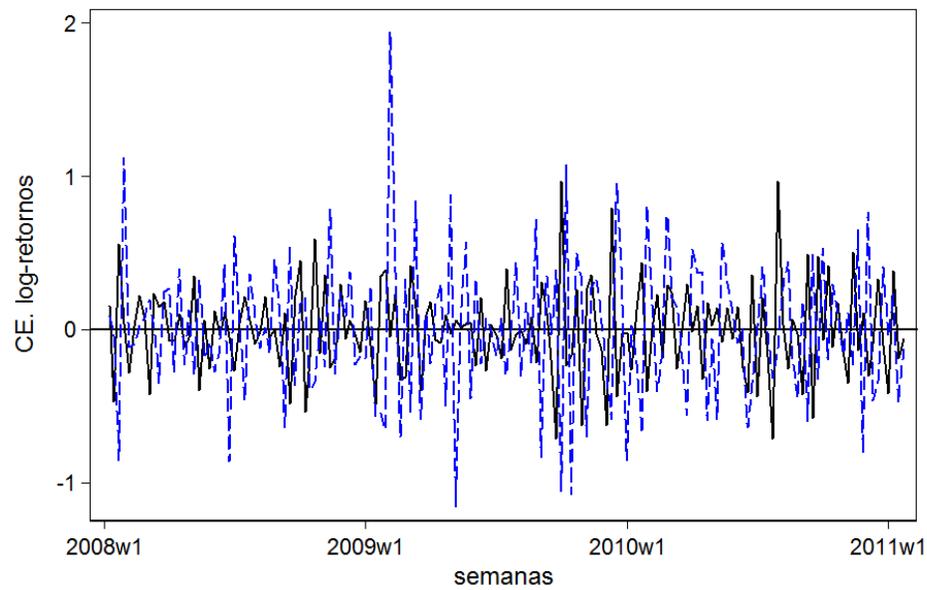


(a)

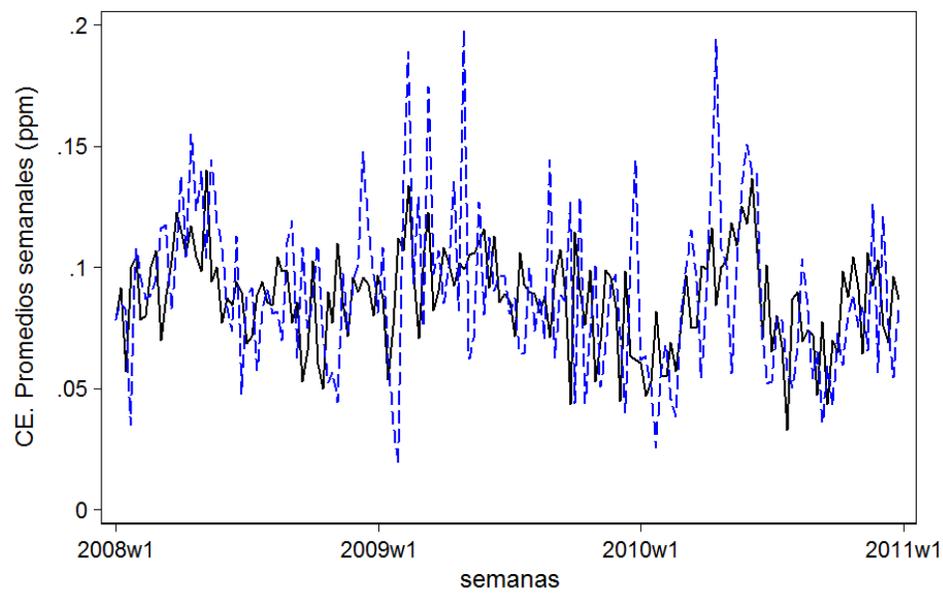


(b)

Figura 5.6: Región NW. En (a) log-retornos del promedio semanal de ozono observados (línea sólida) y estimados (línea punteada). En (b) los promedios semanales de ozono observados (línea sólida) y estimados (línea punteada).

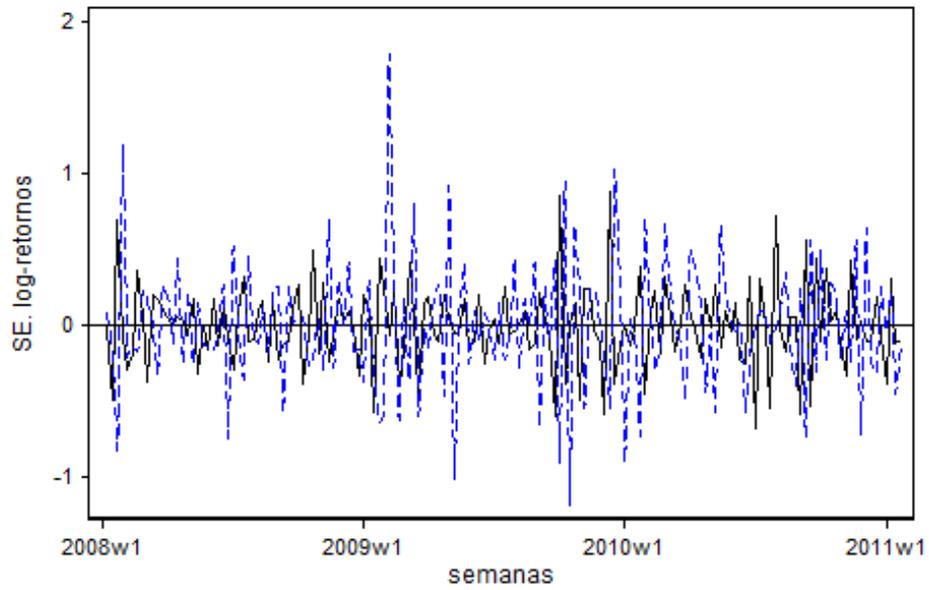


(a)

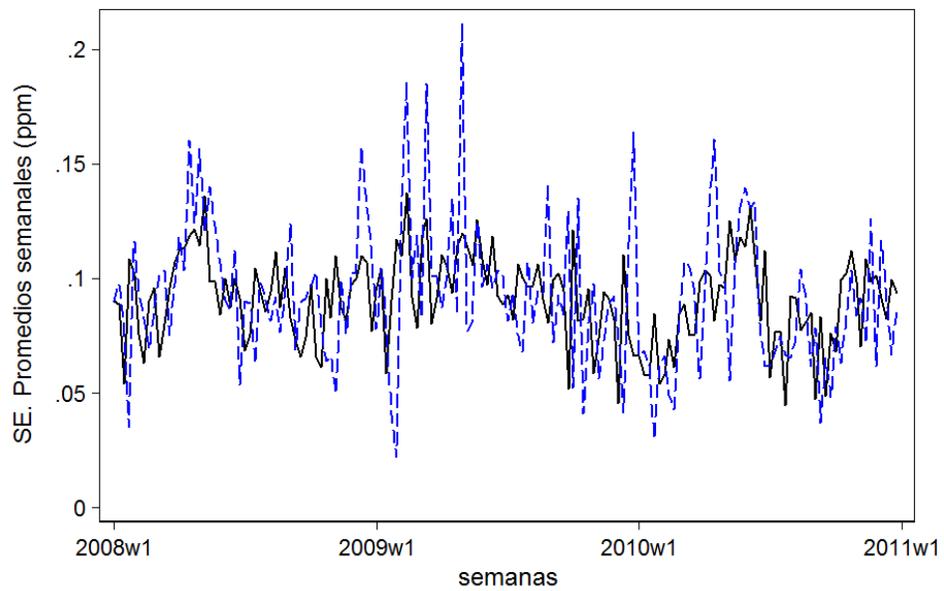


(b)

Figura 5.7: Región CE. En (a) log-retornos del promedio semanal de ozono observados (línea sólida) y estimados (línea punteada). En (b) los promedios semanales de ozono observados (línea sólida) y estimados (línea punteada).

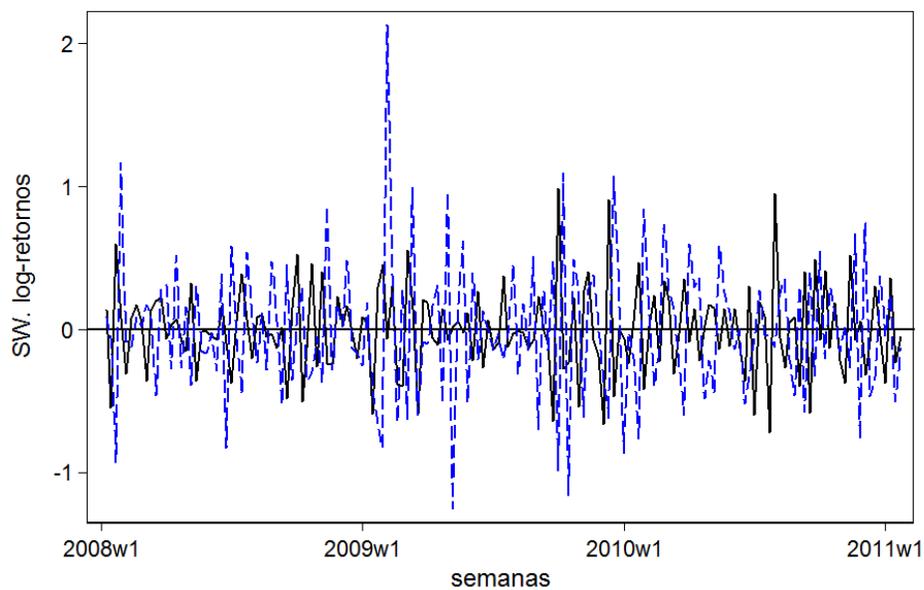


(a)

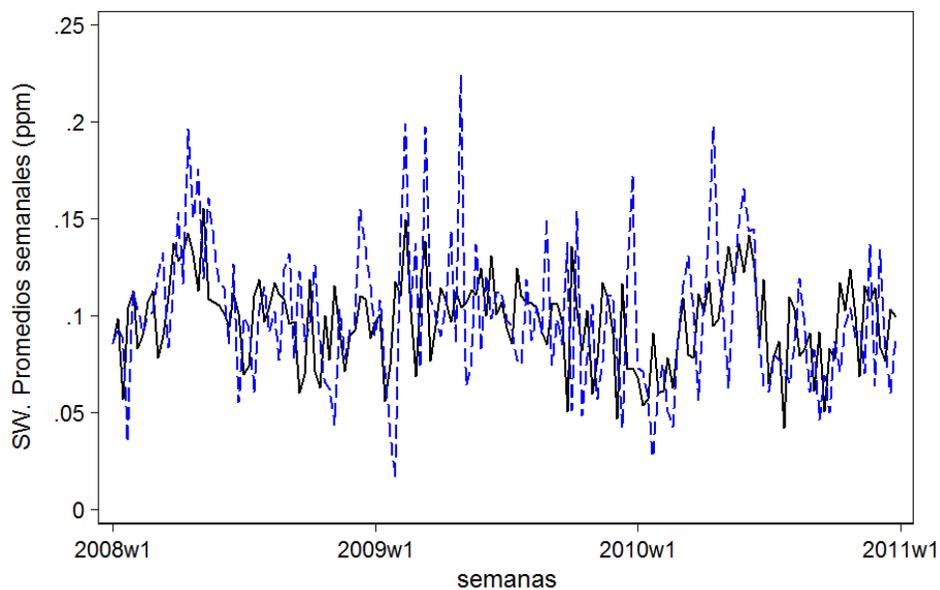


(b)

Figura 5.8: Región SE. En (a) log-retornos del promedio semanal de ozono observados (línea sólida) y estimados (línea punteada). En (b) los promedios semanales de ozono observados (línea sólida) y estimados (línea punteada).



(a)



(b)

Figura 5.9: Región SW. En (a) log-retornos del promedio semanal de ozono observados (línea sólida) y estimados (línea punteada). En (b) los promedios semanales de ozono observados (línea sólida) y estimados (línea punteada).

Finalmente se presentan gráficas del pronóstico y los datos reales que se observaron en las cinco regiones.

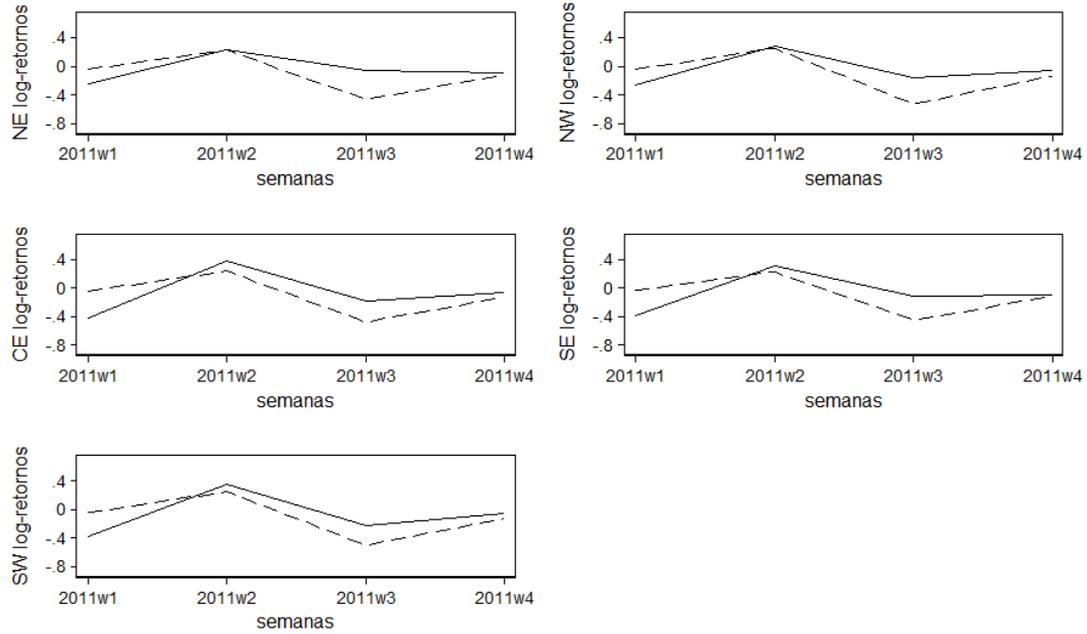


Figura 5.10: Pronóstico para Enero 2011. Log-retornos semanales de ozono por región de la ZMCM. Datos estimados (línea sólida) y observados (línea punteada).

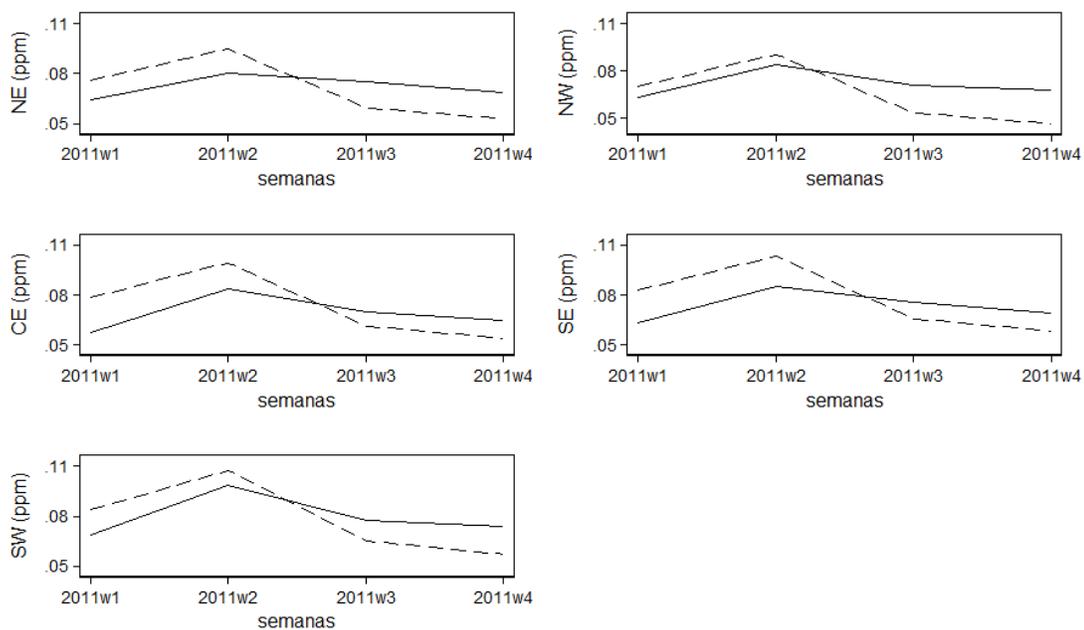


Figura 5.11: Pronóstico para Enero 2011. Promedios semanales de ozono por región de la ZMCM. Datos estimados (línea sólida) y observados (línea punteada).

Capítulo 6

Conclusiones

En este trabajo se utilizó estadística bayesiana y métodos Monte Carlo vía cadenas de Markov para realizar la estimación de los parámetros de un modelo de volatilidad estocástica. Específicamente, se hizo un muestreo de Gibbs. En uno de los pasos de este muestreo se utilizó un paso del algoritmo Metropolis-Hastings para la estimación de las variables latentes del modelo.

Se realizó una aplicación de esta metodología a datos de concentraciones de ozono de la Ciudad de México en el periodo de 2008 a 2010. Para el análisis se tomó la división espacial considerada por las autoridades ambientales de la Ciudad de México, es decir las regiones Noreste (NE), Noroeste (NW), Centro (CE), Sureste (SE) y Suroeste (SW).

Los datos modelados fueron los logaritmos de las mediciones relativas (log-retornos) de los promedios semanales de las concentraciones diarias máximas por región.

Los log-retornos indican incrementos o decrementos de ozono respecto a periodos consecutivos. En el modelo éstos sólo dependen de perturbaciones aleatorias y del valor que tomaron en un instante de tiempo previo. Además se supuso independencia entre las cinco regiones.

Para realizar la estimación de los parámetros del modelo, se escribió un programa usando el software “R” donde se implementó el algoritmo del muestreo de Gibbs. Las muestras de las funciones de densidad marginales completas con distribución estándar fueron obtenidas con funciones contenidas en “R”. Para las que no tuvieron distribución conocida se programó un paso del algoritmo Metropolis-Hastings dentro del muestreo de Gibbs.

La evaluación de la convergencia fue satisfactoria y se realizó por medio de cuatro de los criterios de convergencia contenidos en la paquetería CODA de “R”; tales criterios fueron los de Geweke, Raftery-Lewis, Heidelberger-Welch y el de Gelman-Rubin.

Se observa que las series de los log-retornos presentan un comportamiento muy homogéneo en de las 5 regiones durante el periodo de estudio.(ver Figura 5.4). Lo anterior justifica que los parámetros estimados no difieran mucho entre ellos.

Los resultados muestran que el efecto del parámetro μ sobre las variables latentes es negativo. Este efecto también fue observado en estudios similares que analizan periodos de tiempo más grandes, ver Achcar *et al.* (2008, 2010 y 2011).

El efecto de ϕ sobre las variables latentes en todas las regiones es muy pequeño (alrededor de 0.15), lo que significa que el efecto del periodo pasado sobre el valor actual es muy pequeño. Este comportamiento se observa en todas las regiones. De hecho es menor respecto a los resultados de Achcar *et al.* (2011).

Finalmente la varianza estimada de la perturbación aleatoria asociada a las variables latentes, en todas las regiones fue aceptable y se ve reflejada en que el ajuste de las gráficas es bueno, ver figuras 5.5 a 5.9.

Empleando las estimaciones obtenidas de la tabla 5.2 se aplicó el modelo en cada región para calcular lo siguiente:

- Los log-retornos y los promedios semanales estimados, mismos que se presentan en las figuras 5.5 a 5.9 junto con los log-retornos observados para observar el ajuste de la estimación. En estas gráficas podemos observar que en general se obtuvo un buen ajuste.
- Un pronóstico de los log-retornos y los promedio semanales para las primeras 4 semanas de 2011 que se muestra en las tablas 5.3 y 5.4.

Los pronósticos de los promedios semanales fueron muy cercano a los observados y la predicción de los log-retornos fue acertada en cuanto al crecimiento o decrecimiento de los log-retornos observados.

Finalmente es importante señalar que existen muchas extensiones del modelo de volatilidad estocástica que podrían ayudar a un mejor ajuste de los datos que se

han modelado (ver Asai *et al.* 2006, Chib *et al.* 2008, Philipov y Glickman, 2006, Pitt y Shephard, 1999 y Yu y Meyer, 2006).

Hay evidencia de que se debe incorporar un término de interacción entre las regiones que explique la relación entre sus log-retornos (versión multivariada), ya que si bien se ha obtenido un buen ajuste haciendo uso de una versión univariada del modelo, el efecto de periodos anteriores es muy pequeño y la varianza del componente aleatorio de las variables latentes podría ser menor.

El uso de una versión multivariada del modelo muy probablemente explique la dirección del viento en la Ciudad de México y la forma en que reaccionan los precursores del ozono .

Los parámetros de dichas versiones del modelo de volatilidad estocástica se pueden estimar haciendo algunas modificaciones al programa que en este trabajo se desarrolló. Así el trabajo aquí presentado puede ser retomado en investigaciones futuras para modelar la volatilidad de estos y otros datos. De igual manera se pueden trabajar con otras versiones del modelo con supuestos menos restrictivos que permitan un mejor ajuste.

Cabe resaltar que aunque los hiperparámetros de las distribuciones *a priori* fueron obtenidos a partir de estudios previos, así como de información proporcionada por los propios datos, se podría por ejemplo, utilizar el criterio de entropía. Es decir, elegir las distribuciones *a priori* que maximicen la entropía¹ . Intuitivamente, lo que el enfoque plantea es elegir las distribuciones *a priori* que incorporan el conocimiento previo que resulten lo menos informativas posible, ver Bernardo (1979), Berger (1985) y O'Hagan(1994). Este proceso de elección de las distribuciones *a priori*, también puede considerarse en futuros trabajos.

¹Se define la entropía de la densidad f como $E(-\log(f(\Theta)))$, Bernardo (1979).

Apéndice

Apéndice A

Resultados de los diagnósticos de convergencia

En las siguientes tablas se reporta:

- El *z-score*, en el diagnóstico de Geweke.
- El factor de dependencia I asociado a $q = 0.025$, $r = 0.005$, $s = 0.95$, $\epsilon_{converg} = 0.001$, en el diagnóstico de Raftery-Lewis.
- El *p-value* asociado a la hipótesis nula de que los valores muestreados provienen de una distribución estacionaria asociado a un nivel de significancia de 0.05, en la prueba de Heidelbergger–Welch de estacionariedad.
- El *cociente* de la mitad del ancho del intervalo de confianza para la media al 95 % entre la media, la prueba se realizó con $\epsilon = .1$, en Heidelbergger–Welch de "media ancho", y
- El estadístico R en el diagnóstico de Gelman-Rubin.

Cuadro A.1: NE. Diagnósticos de Convergencia

Diagnóstico	Cadena	μ	ϕ	σ_η^2
Geweke	c1	-0.3837	-0.0593	0.0833
	c2	-0.5159	-1.3970	-0.2380
	c3	-1.5110	0.0493	-1.8320
Raftery–Lewis	c1	1.180	1.110	2.610
	c2	1.160	1.060	3.460
	c3	1.160	1.170	1.730
Heidelberger–Welch Estacionalidad	c1	0.990	0.733	0.902
	c2	0.876	0.550	0.903
	c3	0.245	0.577	0.253
Heidelberger–Welch mitad de ancho	c1	0.0090	0.0052	0.0242
	c2	0.0091	0.0055	0.0229
	c3	0.0097	0.0060	0.0194
Gelman–Rubin	3 cadenas	1.000	1.000	1.000

Cuadro A.2: NW. Diagnósticos de Convergencia

Diagnóstico	Cadena	μ	ϕ	σ_η^2
Geweke	c1	-0.0711	-0.1434	-0.3287
	c2	-1.4280	-0.3156	1.4860
	c3	-0.2325	-0.4031	-2.3820
Raftery–Lewis	c1	1.160	1.080	2.310
	c2	1.210	0.951	1.850
	c3	1.230	1.260	2.480
Heidelberger–Welch Estacionalidad	c1	0.974	0.841	0.975
	c2	0.258	0.656	0.430
	c3	0.923	0.479	0.175
Heidelberger–Welch mitad de ancho	c1	0.0127	0.0065	0.0328
	c2	0.0112	0.0067	0.0289
	c3	0.0104	0.0062	0.0291
Gelman–Rubin	3 cadenas	1.000	1.000	1.000

Cuadro A.3: CE. Diagnósticos de Convergencia

Diagnóstico	Cadena	μ	ϕ	σ_η^2
Geweke	c1	-0.1973	0.0481	-0.3204
	c2	-0.8114	-1.1010	0.6419
	c3	-0.5439	0.0013	-2.0920
Raftery–Lewis	c1	1.230	1.060	1.730
	c2	1.260	1.080	1.690
	c3	1.290	1.130	1.810
Heidelberger–Welch Estacionalidad	c1	0.974	0.841	0.975
	c2	0.258	0.656	0.430
	c3	0.923	0.479	0.175
Heidelberger–Welch mitad de ancho	c1	0.0127	0.0065	0.0328
	c2	0.0112	0.0067	0.0289
	c3	0.0104	0.0062	0.0291
Gelman–Rubin	3 cadenas	1.000	1.000	1.000

Cuadro A.4: SE. Diagnósticos de Convergencia

Diagnóstico	Cadena	μ	ϕ	σ_{η}^2
Geweke	c1	-0.3944	-0.6037	0.5887
	c2	-1.0190	-0.9321	1.0530
	c3	-0.7077	1.1580	-1.7670
Raftery–Lewis	c1	1.260	1.040	2.140
	c2	1.130	1.040	1.990
	c3	1.080	1.080	1.770
Heidelberger–Welch Estacionalidad	c1	0.324	0.233	0.693
	c2	0.511	0.367	0.505
	c3	0.949	0.453	0.671
Heidelberger–Welch mitad de ancho	c1	0.0088	0.0045	0.0262
	c2	0.0092	0.0047	0.0211
	c3	0.0090	0.0061	0.0205
Gelman–Rubin	3 cadenas	1.000	1.000	1.000

Cuadro A.5: SW. Diagnósticos de Convergencia

Diagnóstico	Cadena	μ	ϕ	σ_{η}^2
Geweke	c1	0.2446	-0.2240	-0.2362
	c2	-0.4887	-1.2030	0.6248
	c3	-0.4086	0.1522	-1.6150
Raftery–Lewis	c1	1.320	1.130	1.810
	c2	1.130	1.010	2.040
	c3	1.210	1.080	1.850
Heidelberger–Welch Estacionalidad	c1	0.918	0.479	0.870
	c2	0.184	0.392	0.623
	c3	0.924	0.611	0.422
Heidelberger–Welch mitad de ancho	c1	0.0095	0.0053	0.0265
	c2	0.0090	0.0056	0.0201
	c3	0.0088	0.0061	0.0203
Gelman–Rubin	3 cadenas	1.000	1.000	1.000

Apéndice B

Gráficas de las cadenas por región

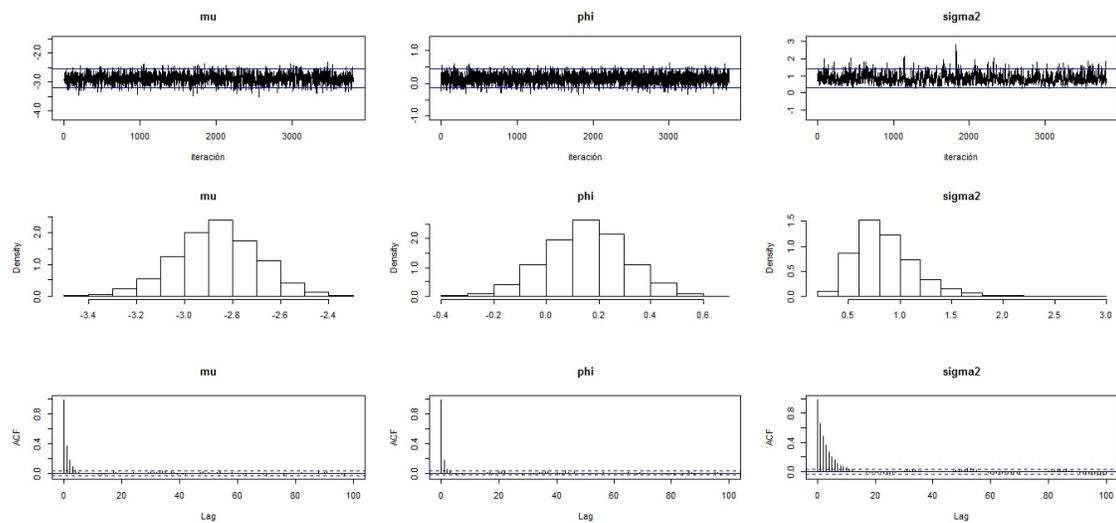


Figura B.1: Región NE. Trace plot, histograma y FAC de la simulación de μ , ϕ y σ^2_{η} .

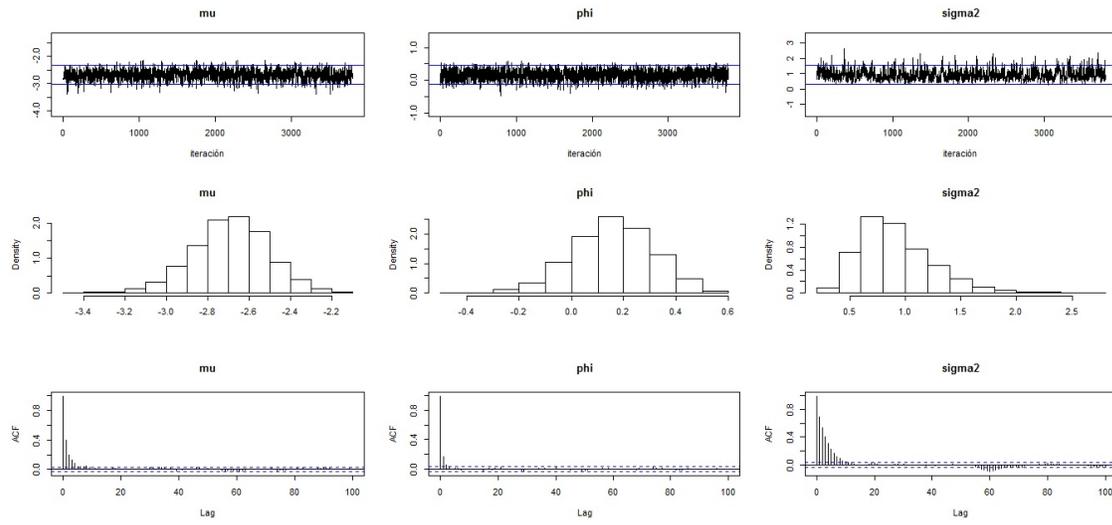


Figura B.2: Región NW. Trace plot, histograma y FAC de la simulación de μ , ϕ y σ_η^2 .

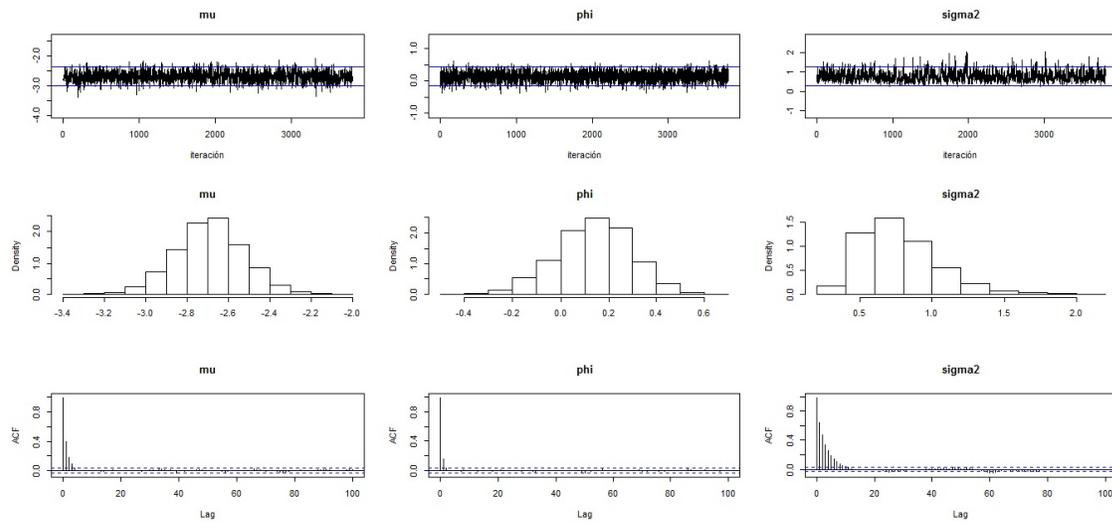


Figura B.3: Región CE. Trace plot, histograma y FAC de la simulación de μ , ϕ y σ_η^2 .

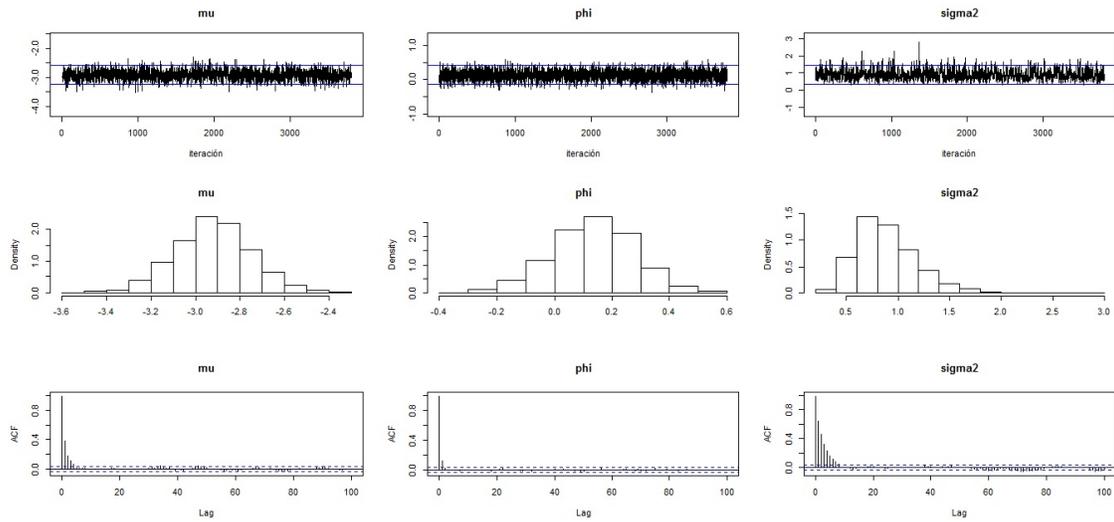


Figura B.4: Región SE. Trace plot, histograma y FAC de la simulación de μ , ϕ y σ_η^2 .

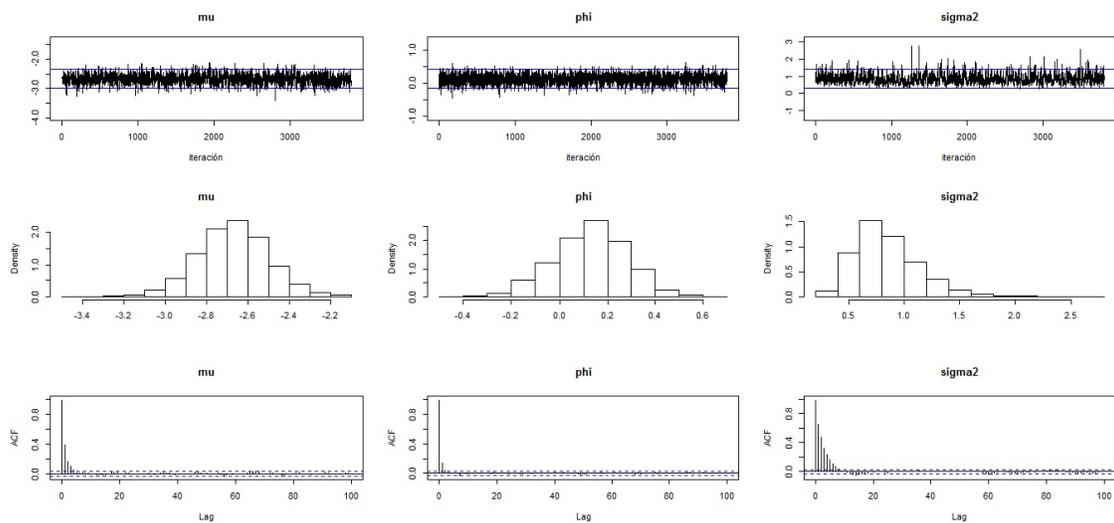


Figura B.5: Región SW. Trace plot, histograma y FAC de la simulación de μ , ϕ y σ_η^2 .

Apéndice C

Muestreo de Gibbs

```
1 #####
2 #   MUESTREO DE GIBBS:           #####
3 #####
4 #Funciones auxiliares para el paso M-H:
5 psi_hj1=function(hj1,yj1){
6 b_2=exp(-1/2*(hj1+exp(-hj1)*yj1^2))
7 b_2
8 }
9 #####
10 psi_hjt=function(hjt,yjt){
11 b_1=exp(-1/2*(hjt+exp(-hjt)*yj1^2))
12 b_1
13 }
14 #####
15 #Datos iniciales
16 n=21000                               #número de iteraciones
17 burn=2000                             #tamaño del periodo de calentamiento
18 n.cadenas=3                           #número de cadenas a simular
19 salto=5                               #tomar valor cada "salto" iteraciones
20 region=1                              #simulación para la región número
21 seed=987                               #semilla para recuperar simulación
22 if(seed>0){set.seed(seed)}
23 a1=0 ;b1=1                             #hiperparámetros phi
24 c1=3 ;d1=3                             #hiperparámetros sigma2
25 e1=0 ;f1=10                           #hiperparámetros mu
26 #Generación de valores iniciales
27 in_mu=rnorm(n.cadenas,0,1)
28 in_phi=rnorm(n.cadenas,0,0.35)
29 in_sigma2=1/rgamma(n.cadenas,3,2)
30 #####
31 mide.tiempo=proc.time()
32 if(region==1){y1=scan("c:/r/rsne3.txt")}
33 if(region==2){y1=scan("c:/r/rsnw3.txt")}
```

```

34 if(region==3){y1=scan("c:/r/rsce3.txt")}
35 if(region==4){y1=scan("c:/r/rsse3.txt")}
36 if(region==5){y1=scan("c:/r/rssw3.txt")}
37 N=155                                     #num de semanas (3 años)
38 espmu=e1; espphi=a1; espsigma2=d1/(c1-1)
39 #####
40 #Almacenamiento de las cadenas
41 SIMmu=matrix(0,n,n.cadenas)
42 SIMphi=matrix(0,n,n.cadenas)
43 SIMsigma2=matrix(0,n,n.cadenas)
44 Hjtn=matrix(0,n,N*n.cadenas)           #guarda simul. de h.
45 Contador=matrix(0,N,n.cadenas)        #guarda valores aceptados en MH
46 #####
47 #####
48 for(m in 1:n.cadenas){
49 #####
50 # Simulación Gibbs n=1 latentes iniciales #
51 # h_j para t=1,2,...,155; N=155 #
52 #####
53 # Paso M-H dentro de GIBBS #
54 # Primero para t=1 #
55 #####
56 anterior=in_mu[m]                       #valor inicial para h_j(1)
57 u=runif(1)
58 y=rnorm(1,in_mu[m],sqrt(in_sigma2[m]))   #densidad propuesta
59 num=psi_hj1(y,y1[1])
60 den=psi_hj1(anterior,y1[1])
61 rho=min(1,num/den)                      #proba de aceptación M-H
62 Hjtn[1,(N*(m-1)+1)]=anterior+(y-anterior)*(u<=rho)#1er sim. Gibbs.
63 Contador[1,m]=(u<=rho)
64 #####
65 # simulamos para t=2,3,...,N #
66 #####
67 for(t in 2:N){
68 hj_ant=Hjtn[1,(N*(m-1)+t-1)]           #actualización GIBBS
69 anterior=rnorm(1,hj_ant,sqrt(in_sigma2[m])) #valor inicial
70 u=runif(1)
71 #####                                     #densidad propuesta
72 y=rnorm(1,in_mu[m]+in_phi[m]*(hj_ant-in_mu[m]),sqrt(in_sigma2[m]))
73 num=psi_hjt(y,y1[t])
74 den=psi_hjt(anterior,y1[t])
75 rho=min(1,num/den)                      #proba de aceptación M-H
76 Hjtn[1,(N*(m-1)+t)]=anterior+(y-anterior)*(u<=rho)
77 Contador[t,m]=(u<=rho)
78 }
79 #####
80 # mu_1--->n=1 #
81 #####
82 h_1=Hjtn[1,(N*(m-1)+1):(N*(m-1)+N)]   #actualización gibbs

```

```

83 #   Parámetros de la condicional completa de mu           #####
84 A=1/f1+(1+(1-in_phi[m])^2*(N-1))/in_sigma2[m]
85 diferencias=numeric(N)
86 for(t in 2:N){
87 diferencias[t]=h_1[t]-in_phi[m]*h_1[t-1]}
88 B=e1/f1+(h_1[1]+(1-in_phi[m])*sum(diferencias))/in_sigma2[m]
89 C=B/A; D=1/A
90 #####
91 SIMmu[1,m]=rnorm(1,C,sqrt(D))           #simulación mu
92 #####
93 #               phi--->n=1                #
94 #####
95 mu_1=SIMmu[1,m]           #actualización gibbs
96 #   Parámetros de la condicional completa de phi       ##
97 diferencias1=numeric(N)
98 for(t in 2:N){
99 diferencias1[t]=(h_1[t-1]-mu_1)^2}
100 A=1/b1+sum(diferencias1)/in_sigma2[m]
101 diferencias2=numeric(N)
102 for(t in 2:N){
103 diferencias2[t]=(h_1[t]-mu_1)*(h_1[t-1]-mu_1)}
104 B=a1/b1+sum(diferencias2)/in_sigma2[m]
105 C=B/A; D=1/A
106 #####
107 SIMphi[1,m]=rnorm(1,C,sqrt(D))           #simulación phi
108 #####
109 #               sigma2--->n=1              #
110 #####
111 phi_1=SIMphi[1,m]           #actualización gibbs
112 #   Parámetros de la condicional completa de sigma2    ###
113 A=c1+N/2
114 diferencias=numeric(N)
115 for(t in 2:N){
116 diferencias[t]=(h_1[t]-mu_1-phi_1*(h_1[t-1]-mu_1))^2}
117 B=1/2*((h_1[1]-mu_1)^2+sum(diferencias))+d1
118 #####
119 SIMsigma2[1,m]=1/rgamma(1,A,B)           #simulación sigma2
120 #####
121 #####
122 #   Simulación de Gibbs                                #
123 #   Para n=2,3,...,#simul.                            #
124 #####
125 #####
126 for(k in 2:n){
127                                     #actualización de parámetros
128 mu_1=SIMmu[k-1,m]
129 phi_1=SIMphi[k-1,m]
130 sigma2_1=SIMsigma2[k-1,m]
131 #   Simulación con M-H para h(1)                       #####

```

```

132 anterior=Hjtn[k-1,(N*(m-1)+1)]
133 u=runif(1)
134 y=rnorm(1,mu_1,sqrt(sigma2_1))
135   num=psi_hj1(y,y1[1])
136   den=psi_hj1(anterior,y1[1])
137   rho=min(1,num/den)                                #proba de aceptación M-H
138 Hjtn[k,(N*(m-1)+1)]=anterior+(y-anterior)*(u<=rho)
139 Contador[1,m]=Contador[1,m]+(u<=rho)
140 #####
141 #       Simulación con M-H h_j con t=2,3,...,N
142 #####
143 for(t in 2:N){
144   hj_ant=Hjtn[k,(N*(m-1)+t-1)]
145   anterior=Hjtn[k-1,(N*(m-1)+t)]
146   u=runif(1)
147   y=rnorm(1,mu_1+phi_1*(hj_ant-mu_1),sqrt(sigma2_1))
148   num=psi_hjt(y,y1[t])
149   den=psi_hjt(anterior,y1[t])
150   rho=min(1,num/den)                                #proba de aceptación M-H
151   Hjtn[k,(N*(m-1)+t)]=anterior+(y-anterior)*(u<=rho)
152   Contador[t,m]=Contador[t,m]+(u<=rho)
153 }
154 #####
155 #       mu-->>n=k
156 #####
157 h_1=Hjtn[k,(N*(m-1)+1):(N*(m-1)+N)]                #actualización gibbs
158 #   Parámetros de la condicional completa mu        #####
159 A=1/f1+(1+(1-phi_1)^2*(N-1))/sigma2_1
160 diferencias=numeric(N)
161 for(t in 2:N){
162   diferencias[t]=h_1[t]-phi_1*h_1[t-1]}
163 B=e1/f1+(h_1[1]+(1-phi_1)*sum(diferencias))/sigma2_1
164 C=B/A; D=1/A
165 #####
166 SIMmu[k,m]=rnorm(1,C,sqrt(D))                        #simulación mu
167 #####
168 #       phi-->>n=k
169 #####
170 mu_1=SIMmu[k,m]                                     #actualización gibbs
171 #   Parámetros de la condicional completa phi        #####
172 diferencias1=numeric(N)
173 for(t in 2:N){
174   diferencias1[t]=(h_1[t-1]-mu_1)^2}
175 A=1/b1+sum(diferencias1)/sigma2_1
176 diferencias2=numeric(N)
177 for(t in 2:N){
178   diferencias2[t]=(h_1[t]-mu_1)*(h_1[t-1]-mu_1)}
179 B=a1/b1+sum(diferencias2)/sigma2_1
180 C=B/A; D=1/A

```

```

181 #####
182 SIMphi[k,m]=rnorm(1,C,sqrt(D))           #simulación phi
183 #####
184 #           sigma2-->n=k
185 #####
186 phi_1=SIMphi[k,m]                       #actualización gibbs
187 #   Parámetros de la condicional completa sigma2   #####
188 A=c1+N/2
189 diferencias=numeric(N)
190 for(t in 2:N){
191 diferencias[t]=(h_1[t]-mu_1-phi_1*(h_1[t-1]-mu_1))^2}
192 B=1/2*((h_1[1]-mu_1)^2+sum(diferencias))+d1
193 #####
194 SIMsigma2[k,m]=1/rgamma(1,A,B)          #simulación sigma2
195 #####
196 }}

```

GibbsFinal1.R

Apéndice D

Diagnóstico de convergencia y gráficas de resultados

```
1 #####
2 #           RESULTADOS:           #
3 #####
4 library(lattice) #Paquetes que sirven
5 library(coda)    #para hacer las pruebas.
6 #####
7 burning=burn+1
8 #####
9 las_h1=matrix(0,N,m)
10 las_h2=matrix(0,N,m)
11 SIMmu1=matrix(0,n-burn,m)
12 SIMphi1=matrix(0,n-burn,m)
13 SIMsigma21=matrix(0,n-burn,m)
14 #####
15 nuevo_tno=(n-burn)/salto
16 SIMmu2=matrix(0,nuevo_tno,m)
17 SIMphi2=matrix(0,nuevo_tno,m)
18 SIMsigma22=matrix(0,nuevo_tno,m)
19 #####
20 mmu1=1; mphi1=1; msigma21=1
21 vmu1=1; vphi1=1; vsigma21=1
22 mmu2=1; mphi2=1; msigma22=1
23 vmu2=1; vphi2=1; vsigma22=1
24
25 for(m in 1:n.cadenas){
26 #####
27 ## Estimaciones a posteriori de las h's ##
28   for(t in 1:N){las_h1[t,m]=mean(Hjtn[burning:n,(N*(m-1)+t)])}
29   help.h=matrix(0,nuevo_tno,N);
30   for(t in 1:N){for(i in 1:nuevo_tno){
31     help.h[i,t]=Hjtn[(burn+salto*i),(N*(m-1)+t)]}}
```

```

32     for(t in 1:N){las_h2[t,m]=mean(help.h[,t])}
33 #####
34 ## Estimaciones para muestra sin minar ##
35 SIMmu1[,m]=SIMmu[burning:n,m]
36 SIMphi1[,m]=SIMphi[burning:n,m]
37 SIMsigma21[,m]=SIMsigma2[burning:n,m]
38 mmu1[m]=mean(SIMmu1[,m])
39 mphi1[m]=mean(SIMphi1[,m])
40 msigma21[m]=mean(SIMsigma21[,m])
41 vmu1[m]=var(SIMmu1[,m])
42 vphi1[m]=var(SIMphi1[,m])
43 vsigma21[m]=var(SIMsigma21[,m])
44 ## Estimaciones para muestra minada ##
45 for(i in 1:nuevo_tno){
46 SIMmu2[i,m]=SIMmu1[salto*i,m]
47 SIMphi2[i,m]=SIMphi1[salto*i,m]
48 SIMsigma22[i,m]=SIMsigma21[salto*i,m]}
49 mmu2[m]=mean(SIMmu2[,m])
50 mphi2[m]=mean(SIMphi2[,m])
51 msigma22[m]=mean(SIMsigma22[,m])
52 vmu2[m]=var(SIMmu2[,m])
53 vphi2[m]=var(SIMphi2[,m])
54 vsigma22[m]=var(SIMsigma22[,m])}
55 #####
56 #####
57 m=1
58 par(mfrow=c(3,3))
59 plot(SIMmu2[,m],type="l",ylim=c(mmu2[m]-4*1.96*sqrt(vmu2[m]),mmu2[m]
60     +4*1.96*sqrt(vmu2[m])))
61 abline(h=c(mmu2[m]-1.96*sqrt(vmu2[m]),mmu2[m]+1.96*sqrt(vmu2[m])),
62     col="blue")
63 plot(SIMphi2[,m],type="l",ylim=c(mphi2[m]-4*1.96*sqrt(vphi2[m]),
64     mphi2[m]+4*1.96*sqrt(vphi2[m])))
65 abline(h=c(mphi2[m]-1.96*sqrt(vphi2[m]),mphi2[m]+1.96*sqrt(vphi2[m]
66     )),col="blue")
67 plot(SIMsigma22[,m],type="l",ylim=c(msigma22[m]-4*1.96*sqrt(
68     vsigma22[m]),msigma22[m]+4*1.96*sqrt(vsigma22[m])))
69 abline(h=c(msigma22[m]-1.96*sqrt(vsigma22[m]),msigma22[m]+1.96*sqrt
70     (vsigma22[m])),col="blue")
71 hist(SIMmu2[,m])
72 hist(SIMphi2[,m])
73 hist(SIMsigma22[,m])
74 acf(SIMmu2[,m],100)
75 acf(SIMphi2[,m],100)
76 acf(SIMsigma22[,m],100)
77 simtime=proc.time()-mide.tiempo
78 rbind(paste("Resultados para muestra usando las",n.cadenas,"cadenas
79     ."),

```

```

73 paste("Parámetros estimados cada",salto,"datos:",mean(mmu2),mean(
    mphi2),mean(msigma22)),
74 paste("Parámetros estimados muestra completa:",mean(mmu1),mean(
    mphi1),mean(msigma21)),
75 paste("En todos los casos el calentamiento fue de",burn),
76 paste("Tamaño de la muestra cada cadena (n-burn):",n-burn),
77 paste("Tamaño de la muestra cada cadena usando cada",salto,"datos
    ((n-burn)/salto):",nuevo_tno),
78 paste("Resultados para cadena ",m),
79 paste("Valores iniciales:","mu0=",in_mu[m],"phi0=",in_phi[m],"
    sigma20=",in_sigma2[m]),
80 paste("Parámetros estimados cada",salto,"datos:",mmu2[m],mphi2[m],
    msigma22[m]),
81 paste("Varianzas muestrales cada",salto,"datos:",vmu2[m],vphi2[m],
    vsigma22[m]),
82 paste("Parámetros estimados muestra completa:",mmu1[m],mphi1[m],
    msigma21[m]),
83 paste("Varianzas muestrales con muestra completa:",vmu1[m],vphi1[m]
    ],vsigma21[m]),
84 paste("Hiperp(mu,phi,sigma2)=",e1,f1,a1,b1,c1,d1),
85 paste("Esperanzas a priori (mu,phi,sigma2)=",espmu,espphi,espsigma2
    ),
86 paste("Varianzas a priori(mu,phi,sigma2)=",f1,b1,(espsigma2^2/(c1
    -2))),
87 paste("Tiempo simulación:",simtime[3]))
88 ##### Pruebas con coda de R
89 #####
90 ##1: Geweke
91 Gibbs_mu1=mcmc(SIMmu1[,m])
92 Gibbs_phi1=mcmc(SIMphi1[,m])
93 Gibbs_sigma21=mcmc(SIMsigma21[,m])
94 geweke.diag(Gibbs_mu1)
95 geweke.diag(Gibbs_phi1)
96 geweke.diag(Gibbs_sigma21)
97 ##
98 Gibbs_mu2=mcmc(SIMmu2[,m])
99 Gibbs_phi2=mcmc(SIMphi2[,m])
100 Gibbs_sigma22=mcmc(SIMsigma22[,m])
101 geweke.diag(Gibbs_mu2)
102 geweke.diag(Gibbs_phi2)
103 geweke.diag(Gibbs_sigma22)
104 ##2: Gelman and Rubin
105 GibbsGR_mu11=mcmc(SIMmu1[,1])
106 GibbsGR_phi11=mcmc(SIMphi1[,1])
107 GibbsGR_sigma211=mcmc(SIMsigma21[,1])
108 GibbsGR_mu12=mcmc(SIMmu1[,2])
109 GibbsGR_phi12=mcmc(SIMphi1[,2])
110 GibbsGR_sigma212=mcmc(SIMsigma21[,2])

```

```

111 GibbsGR_mu13=mcmc(SIMmu1[,3])
112 GibbsGR_phi13=mcmc(SIMphi1[,3])
113 GibbsGR_sigma213=mcmc(SIMsigma21[,3])
114 GibbsGR_mu1=mcmc.list(list(GibbsGR_mu11,GibbsGR_mu12,GibbsGR_mu13))
115 GibbsGR_phi1=mcmc.list(list(GibbsGR_phi11,GibbsGR_phi12,GibbsGR_
    phi13))
116 GibbsGR_sigma21=mcmc.list(list(GibbsGR_sigma211,GibbsGR_sigma212,
    GibbsGR_sigma213))
117 gelman.diag(GibbsGR_mu1)
118 gelman.diag(GibbsGR_phi1)
119 gelman.diag(GibbsGR_sigma21)
120 ##
121 GibbsGR_mu21=mcmc(SIMmu2[,1])
122 GibbsGR_phi21=mcmc(SIMphi2[,1])
123 GibbsGR_sigma221=mcmc(SIMsigma22[,1])
124 GibbsGR_mu22=mcmc(SIMmu2[,2])
125 GibbsGR_phi22=mcmc(SIMphi2[,2])
126 GibbsGR_sigma222=mcmc(SIMsigma22[,2])
127 GibbsGR_mu23=mcmc(SIMmu2[,3])
128 GibbsGR_phi23=mcmc(SIMphi2[,3])
129 GibbsGR_sigma223=mcmc(SIMsigma22[,3])
130 GibbsGR_mu2=mcmc.list(list(GibbsGR_mu21,GibbsGR_mu22,GibbsGR_mu23))
131 GibbsGR_phi2=mcmc.list(list(GibbsGR_phi21,GibbsGR_phi22,GibbsGR_
    phi23))
132 GibbsGR_sigma22=mcmc.list(list(GibbsGR_sigma221,GibbsGR_sigma222,
    GibbsGR_sigma223))
133 gelman.diag(GibbsGR_mu2)
134 gelman.diag(GibbsGR_phi2)
135 gelman.diag(GibbsGR_sigma22)
136 #gelman.plot(GibbsGR_mu1)
137
138 ##3: Raftery and Lewis
139 raftery.diag(Gibbs_mu1)
140 raftery.diag(Gibbs_phi1)
141 raftery.diag(Gibbs_sigma21)
142 ##
143 raftery.diag(Gibbs_mu2)
144 raftery.diag(Gibbs_phi2)
145 raftery.diag(Gibbs_sigma22)
146
147 ##4: Heidelberger and Welch
148 heidel.diag(Gibbs_mu1)
149 heidel.diag(Gibbs_phi1)
150 heidel.diag(Gibbs_sigma21)
151 ##
152 heidel.diag(Gibbs_mu2)
153 heidel.diag(Gibbs_phi2)
154 heidel.diag(Gibbs_sigma22)

```

Gráfica de series observada y estimada.

```

1 grafo=function(seed,m,p,s2,z,y_real,region,linf){
2 #####
3 set.seed(seed)
4 epsilon=rnorm(160)
5 h=1; z_est=1; y_est=numeric(160); volatilidad=1
6 #####
7 h[1]=rnorm(1,m,sqrt(s2))
8 volatilidad[1]=exp(h[1])
9 #y_est[1]=exp(h[1]/2)*epsilon[1]
10 z_est[1]=z[1]
11 for(i in 2:156){
12 h[i]=rnorm(1,(m+p*(h[i-1]-m)),sqrt(s2))
13 volatilidad[i]=exp(h[i])
14 z_est[i]=z[i-1]*exp(sqrt(volatilidad[i])*epsilon[i])
15 y_est[i]=log(z_est[i]/z_est[i-1])}
16 #####
17 for(i in 157:160){
18 h[i]=rnorm(1,(m+p*(h[i-1]-m)),sqrt(s2))
19 volatilidad[i]=exp(h[i])
20 z_est[i]=z_est[i-1]*exp(sqrt(volatilidad[i])*epsilon[i])
21 y_est[i]=log(z_est[i]/z_est[i-1])}
22 #####
23 graf=matrix(0,160,2)
24 graf[,1]= z
25 graf[,2]= z_est
26 graf_y=matrix(0,160,2)
27 graf_y[,1]= c(0,y_real)
28 graf_y[,2]= y_est
29 errores=abs(graf[,1]-graf[,2])
30 error=sum(errores)
31 volat=sum(volatilidad)
32 #matplot(1:160,graf,type='l',lwd=2.5,ylim=c(-1,1),ylab=" log-
    retorno ",xlab=" Número de semana", main=region)
33 #par(mfrow=c(3,2))
34 par(mfrow=c(1,1))
35 #plot(1:160,volatilidad,type='l',col=2)
36 #plot(linf:lsup,volatilidad[linf:lsup],type='l',col=2)
37 #plot(1:160,errores,type='l',col=2)
38 plot(1:160, graf[,1],type='n',lwd=2.5,ylim=c(0,.3),ylab="
    promedio semanal",xlab=" Número de semana", main=region)
39 ver=seq(0,160,1)
40 hor=seq(-1,1,.1)
41 imp=c(-1,-.5,0,.5,1)
42 abline(h=hor,v=ver,lwd=.5,col='gray')
43 abline(h=0,lwd=1.5)
44 legend(70,-.8, c('observada','estimada'),lty=1:2, col=1:2, lwd=2)
45 lines(1:156, graf[1:156,1],type='l',lwd=2.5)
46 lines(156:160, graf[156:160,1],type='o',col="blue",lwd=2.5)

```

```

47 lines(1:160, graf[1:160,2],type='o',lwd=2.5, col=2, lty=2)
48 abline(h=0,lwd=1.5)
49 #plot(linf:160, graf[linf:160,1],type='n',lwd=2.5, ylim=c(0, .3),
      ylab="promedio semanal",xlab=" Número de semana", main=region)
50 #hor=seq(-1,1,.1)
51 #imp=c(-1,-.5,0,.5,1)
52 #abline(h=hor,lwd=.5,col='gray')
53 #abline(h=0,lwd=1.5)
54 #legend(70,-.8, c('observada','estimada'),lty=1:2, col=1:2, lwd=2)
55 #lines(linf:156, graf[linf:156,1],type='l',lwd=2.5)
56 #lines(156:160, c(graf[156,1],.06414,.080286,.07557,.069143),type='
      o',col="blue",lwd=2.5)
57 #lines(156:160, graf[156:160,1],type='o',col="blue",lwd=2.5)
58 #lines(linf:160, graf[linf:160,2],type='l',lwd=2.5, col=2, lty=2)
59 #abline(h=0,lwd=1.5)
60
61 #plot(1:160, graf_y[,1],type='n',lwd=2.5, ylim=c(-1,1), ylab=" log-
      retorno ",xlab=" Número de semana", main=region)
62 #hor=seq(-1,1,.1)
63 #imp=c(-1,-.5,0,.5,1)
64 #abline(h=hor,lwd=.5,col='gray')
65 #abline(h=0,lwd=1.5)
66 #legend(70,-.8, c('observada','estimada'),lty=1:2, col=1:2, lwd=2)
67 #lines(2:156, graf_y[2:156,1],type='l',lwd=2.5)
68 #lines(2:160, graf_y[2:160,2],type='l',lwd=2.5, col=2, lty=2)
69 #abline(h=0,lwd=1.5)
70 #plot(linf:160, graf_y[linf:160,1],type='n',lwd=2.5, ylim=c(-1,1),
      ylab=" log-retorno ",xlab=" Número de semana", main=region)
71 #hor=seq(-1,1,.1)
72 #imp=c(-1,-.5,0,.5,1)
73 #abline(h=hor,lwd=.5,col='gray')
74 #abline(h=0,lwd=1.5)
75 #legend(70,-.8, c('observada','estimada'),lty=1:2, col=1:2, lwd=2)
76 #lines(linf:156, graf_y[(linf-1):155,1],type='l',lwd=2.5)
77 #lines(156:160,graf_y[155:159,1],col="blue",type='o',lwd=2.5)
78 #lines(linf:160, graf_y[(linf):160,2],type='l',lwd=2.5, col=2, lty
      =2)
79 #abline(h=0,lwd=1.5)
80 #c(y_est,error,volat)
81 paste("promsem",z_est[157:160],"log-retornos",y_est[157:160],"error
      y volatilidad totales",error,volat)
82 write.table(graf_y,file="c:/r/grafy_nw_140.xls")
83 }
84
85 #z=scan("c:/r/psne3.txt")
86 #y1=scan("c:/r/rsne3.txt")
87 #grafo(240,-2.864,0.154,0.837,z,y1, 'NE', 140)
88 z=scan("c:/r/psnw3.txt")
89 y1=scan("c:/r/rsnw3.txt")

```

```
90 grafo(240,-2.689,0.1625, 0.912,z,y1, 'NW',140)
91 #z=scan("c:/r/psce3.txt")
92 #y1=scan("c:/r/rsce3.txt")
93 #grafo(240,-2.689,0.141,0.757,z,y1, 'CE',140)
94 #z=scan("c:/r/psse3.txt")
95 #y1=scan("c:/r/rsse3.txt")
96 #grafo(240,-2.921,0.135,0.888,z,y1, 'SE',140)
97 #z=scan("c:/r/pssw3.txt")
98 #y1=scan("c:/r/rssw3.txt")
99 #grafo(240,-2.668,0.131,0.839,z,y1, 'SW',140)
```

grafo_modzs.R

Bibliografía

- [1] Achcar, J.A., Zozolotto, H.C. y Rodrigues, E.R. (2008). Bivariate volatility models applied to air pollution data, *Rev. Bras. Biom.* Vol. 26: pp.67-81.
- [2] Achcar, J.A., Zozolotto, H.C. y Rodrigues, E.R. (2010). Bivariate stochastic volatility models applied to Mexico City ozone pollution data, in G.C. Romano & A.G. Conti (eds.), *Air Quality in the 21st Century*, Nova Publishers, New York, pp. 241-267
- [3] Achcar, J.A., Rodrigues, E.R. y Tzintzun, G. (2011). Using stochastic volatility models to analyse weekly ozone averages in Mexico City, *Environmental and Ecological Statistics* 18, pp. 271 – 290.
- [4] Albert, J. (2009). *Bayesian Computation with R*. Use R, 2nd ed. Springer.
- [5] Anderson *et al.*, Editors(2009). *Handbook of Financial Time Series*. Springer-Verlag, Berlin.
- [6] Asai, M., McAleer, M. y Yu, J. (2006). Multivariate stochastic volatility: A review. *Econometric Reviews* 25, pp. 145 – 175.
- [7] Athreya, K. B., Doss, H. y Sethuraman, J. (1996). On the convergence of the Markov chain simulation method. *Annals of Statist.* Volume 24, Number 1, pp.69-100
- [8] Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, second edition, Springer-Verlag, NewYork.
- [9] Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Royal Statistical Society B* 41, pp. 113-147 (with discussion)
- [10] Bollerslev T., Generalized Autoregressive Conditional Heterocedasticity, *Journal of Econometrics*, Vol. 31, No. 3, 1986, pp. 307-327.
- [11] Broto, C. y Ruiz, E. (2004). Estimation methods for stochastic volatility models: A survey. *Journal of Economic Survey* 18, pp. 613 – 649.
- [12] Carlin, B. y Louis, T. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*, second edition, Chapman and Hall, New York.

- [13] Carter, C., y Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika* 81, pp. 541 – 553.
- [14] Casella, G., y George, E (1992). Explaining the Gibbs Sampler, *The American Statistician*, 46 pp. 167 – 174.
- [15] Congdon, P.(2003). *Applied Bayesian Modelling*, Wiley Series in Probability and Statistics, John Wiley and Sons Ltd.
- [16] Congdon, P. (2010). *Applied Bayesian Hierarchical Methods*, Chapman and Hall/CRC.
- [17] Comisión Ambiental Metropolitana (2011), Programa para Mejorar la Calidad del Aire de la Zona Metropolitana del Valle de México 2011-2020, D.F., México.
- [18] Cowles M. K., y Carlin B. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *Journal of the American Statistical Association*, Vol. 91, No. 434, pp. 883 – 904.
- [19] Chib, S., y Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician* 49, pp. 327 – 335.
- [20] Chib, S., Omori, Y. y Asai, M. (2008). Multivariate stochastic volatility. En: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.): *Handbook of Financial Time Series*, pp. 365 – 400. Springer, New York.
- [21] De Jesús V., Rodrigues E. and Tzintzun G., A Gibbs Sampling Algorithm to Estimate the Parameters of a Volatility Model: An Application to Ozone Data, *Applied Mathematics*, Vol. 3 No. 12A, 2012, pp. 2178-2190.
- [22] Engle R. F., Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, Vol. 50, No. 4, 1982, pp. 987-1007.
- [23] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. 1, 3rd. ed., John Wiley, New York.
- [24] Gelman, A.(1992). Iterative y Non-Iterative Simulation Algorithms, in *Computing Science and Statistic (Interface Proceedings)*, 24, pp. 433 – 438
- [25] Gelman, A. y Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 7, pp. 457 – 511.
- [26] Geman S. y Geman D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, pp. 721 – 741.
- [27] Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*, John Wiley & Sons, Inc.

- [28] Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. En: Bayesian Statistics 4, eds. J. M. Bernardo *et al.*, Oxford, U.K.: Oxford University Press, pp. 169 – 193.
- [29] Gilks W. R., Richardson S. y Spiegelhalter D. J. (Eds.) (1996). Markov chain Monte Carlo in Practice. Chapman and Hall, London, UK.
- [30] Harvey A., Ruiz E, Shephard N.(1994). Multivariate Stochastic Variance Models, The Review of Economic Studies, Vol. 61, No. 2, pp. 247-264.
- [31] Hastings W.K. (1970). Monte Carlo sampling-based methods using Markov chains and their applications. Biometrika 57 pp. 97 – 109.
- [32] Heidelberger, P. y Welch, P. (1983). A spectral method for confidence interval generation and run length control in simulations. Comm. Assoc. Comput. Machinery, 24: pp. 233-245.
- [33] Jacquier, E., Polson, N. G. y Rossi, P. E. (1994): Bayesian analysis of stochastic volatility models (with discussion). Journal of Business and Economic Statistics 12, pp. 371 – 389.
- [34] Lunn, D. J., A. Thomas, N. G. Best, y D. J. Spiegelhalter (2000). WinBUGSA Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10, pp. 325 – 337.
- [35] Marin, J., y Robert, C. P. (2007). Bayesian Core: A Practical Approach to Computational Bayesian Statistics. New York: Springer.
- [36] Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, y E. Teller (1953). Equation of State Calculations by Fast ComputingMachines, The Journal of Chemical Physics 21: pp. 1087-1092.
- [37] Meyn, S. P. y Tweedie, R. L. (2009). Markov Chains and Stochastic Stability, second edition, Springer, London.
- [38] Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). Introduction to the Theory of Statistics, third edition. McGraw-Hill.
- [39] NOM (2002). Modificación a la norma oficial mexicana NOM-020-SSA1-1997, Diario Oficial de la Federación, 30 de Octubre de 2002.
- [40] Ntzoufras I. (2009). Bayesian Modeling Using WinBUGS, Wiley Series in Computational Statistics, John Wiley & Sons, Inc.
- [41] O’Hagan, A. (1994). Kendall’s Advanced Theory of Statistics, Vol.2b: Bayesian Inference, Cambridge: Edward Arnold.
- [42] Philipov, A. y Glickman, M. E. (2006). Multivariate stochastic volatility via Wishart processes. Journal of Business and Economic Statistics 24, pp. 313 – 328.

- [43] Pitt, M. K. y Shephard, N. (1999). Time varying covariances: a factor stochastic volatility approach. En: Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F.M. (Eds.): Bayesian Statistics 6, pp. 547 – 570. Oxford University Press, Oxford.
- [44] Plummer, M., Best, N., Cowles, K. y Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC, *R News* 6(1), pp. 7-1.
- [45] Raftery, A. y Lewis, S. (1992). How many iterations in the Gibbs sampler?, En J. Bemardo, J. Berger, A. Dawid, and A. Smith, eds., Bayesian Statistics, Vol. 4, Claredon Press, Oxford pp. 763-774.
- [46] Robert, C.P. y Casella, G. (2004). Monte Carlo Statistical Methods, second edition, Springer-Verlang, New York.
- [47] Robert, C.P. y Casella, G. (2010). Introducing Monte Carlo Methods with R. Use R. Springer.
- [48] Roberts, G. O. y Smith A. F. M.(1994), Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms, *Stochastic Processes and Their Applications* 49: pp 207-216.
- [49] Roberts G. O. y Rosenthal J. S. (2006). Harris recurrence of Metropolis with Gibbs and trans-dimensional Markov chains. *The Annals of Applied Probability*, Vol. 16, No. 4, pp. 2123 – 2139.
- [50] Ross, S. M. (1996) Stochastic processes, second edition, John Wiley&Sons, Inc.
- [51] Shephard, N. (2004). Stochastic Volatility: Selected Readings. Oxford University Press, Oxford.
- [52] Shephard, N. y Andersen T.G.(2009). Stochastic Volatility: Origins and Overview. En : Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.): Handbook of Financial Time Series, pp. 233 – 254. Springer, New York.
- [53] Smith, A.F.M. y Roberts G. O. (1993). Bayesian computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society (Series B)* 55, pp. 3-23.
- [54] Taylor, S. J. (1982). Financial Returns Modelled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices, 1961-1979. En Shephard, N. Stochastic Volatility: Selected Readings. Oxford University Press, Oxford, pp.60-82.
- [55] Taylor, S. J. (1986). Modelling Financial Time Series. New York: Wiley.
- [56] Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions (with discussion and rejoinder), *Annals of Statistics* 22, pp. 1701-1762.

- [57] Yu, J. y Meyer, R. (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews* 25, pp. 361 – 384.
- [58] Zozolotto, H. C. (2010). Aplicação de modelos de volatilidade estocástica em dados de poluição do ar de duas grandes cidades: cidade do México e São Paulo. 2010. Dissertação. Tesis de Maestría.