

UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS

---



FACULTAD DE CIENCIAS  
ESCUELA DE MATEMÁTICA

MODELOS AUTORREGRESIVOS VECTORIALES  
INTEGRADOS CON VOLATILIDAD ESTOCÁSTICA  
MULTIVARIADA

T E S I S

PARA OBTENER EL TÍTULO DE:

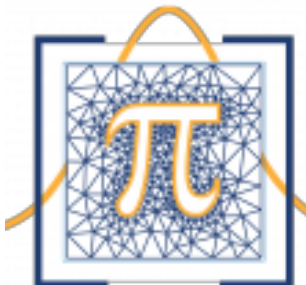
MÁSTER EN MATEMÁTICA CON ORIENTACIÓN EN  
ESTADÍSTICA MATEMÁTICA

PRESENTA:

MARVIN LEVI VILLAFRANCA RIVERA

CUENTA:

MMM100216



ASESOR:

DR. CRISTIAN ANDRES CRUZ TORRES

Ciudad Universitaria, Tegucigalpa, Mayo, 2022

# Resumen

Los modelos autorregresivos vectoriales (VAR) han demostrado ser eficientes para capturar las relaciones dinámicas de las series de tiempo multivariadas. Los modelos de volatilidad estocástica multivariada (MSV) muestran ser útiles para modelar la varianza cuando cambia en el tiempo. Por lo anterior, en esta tesis se propone la integración de un modelo VAR con un modelo MSV con efecto de apalancamiento cruzado (VAR-MSV). La escogencia del VAR más adecuado se lleva a cabo por medio del Criterio de Información de Desviación (DIC). El modelo es validado mediante un ejemplo simulado y con datos reales bivariados, se hizo una aplicación para los precios de Apple y Microsoft y se interpretaron los resultados. Para estimar los parámetros se usan métodos de Monte Carlo vía Cadenas de Markov (MCMC). Los resultados indican que el modelo VAR-MSV captura las relaciones dinámicas así como la varianza cambiando en el tiempo de manera eficaz.

**Palabras clave:** Volatilidad Estocástica, Apalancamiento Cruzado, VAR-MSV, Muestreador Multimovimiento.

# Abstract

Vector autoregressive (VAR) models have proven to be efficient in capturing the dynamic relationships of multivariate time series. Multivariate Stochastic Volatility (MSV) models are shown to be useful for modeling variance as it changes over time. Therefore, in this thesis the integration of a VAR model with a MSV model with cross-leverage effect (VAR-MSV) is proposed. The choice of the most appropriate VAR is carried out through the Deviation Information Criterion (DIC). The model is validated by means of a simulated example and with real bivariate data, an application was made for the prices of Apple and Microsoft and the results were interpreted. To estimate the parameters, Monte Carlo methods via Markov Chains (MCMC) are used. The results indicate that the VAR-MSV model captures the dynamic relationships as well as the variance changing over time effectively.

**Keywords:** Stochastic Volatility, Cross Leverage, VAR-MSV, Multi-Move Sampler.

# Agradecimientos

En primer lugar quisiera agradecer a Dios por proveerme las personas, recursos y sabiduría para poder realizar este trabajo. Gracias Dr. Cristian Andrés Cruz Torres por su dirección y consejos en este trabajo de tesis. Gracias a mis Padres por el apoyo económico brindado. Agradezco a mi amigo Mauricio Alfredo Zelaya Aguilar por su apoyo en este trabajo.

*En la multitud de consejeros hay seguridad...*

Proverbios 11:14.

# Índice general

Índice de figuras	VI
Índice de tablas	IX
<b>1 Introducción</b>	<b>1</b>
<b>2 Procesos Autorregresivos Vectoriales Estables</b>	<b>6</b>
§2.1 Propiedades y suposiciones de los modelos VAR . . . . .	7
§2.2 Métodos de Estimación . . . . .	12
§2.2.1 Método de mínimos cuadrados multivariado . . . . .	13
§2.2.2 Método de máxima verosimilitud . . . . .	15
§2.2.3 Método de estimación bayesiano . . . . .	18
§2.3 Métodos de Selección del Orden VAR . . . . .	21
§2.4 Verificación del Modelo Adecuado . . . . .	24
§2.4.1 Verificación de los Residuales . . . . .	24
§2.4.2 Prueba de Portmanteau . . . . .	26
§2.4.3 Prueba de Normalidad . . . . .	27
<b>3 Modelos de Volatilidad Estocástica Multivariado</b>	<b>29</b>
§3.1 Modelos GARCH . . . . .	32
§3.2 Modelos SV . . . . .	37
§3.3 Modelos MSV . . . . .	39

§3.3.1 Efecto de apalancamiento . . . . .	40
§3.3.2 Efecto de umbral . . . . .	42
§3.3.3 Efecto asimétrico general . . . . .	44
§3.3.4 Modelo Factor . . . . .	45
§3.4 Métodos de Estimación . . . . .	48
§3.5 Comparación de Modelos . . . . .	57
<b>4 Modelo Autorregresivo Vectorial con Volatilidad Estocástica Multivariada</b>	<b>61</b>
§4.1 Revisión de la Literatura de Modelos VAR-MSV . . . . .	62
§4.2 Modelo autorregresivo vectorial con volatilidad estocástica y efecto de apalancamiento	72
§4.3 Método de Estimación . . . . .	75
§4.4 Selección del orden VAR-MSV . . . . .	76
§4.5 Ejemplo . . . . .	77
§4.6 Resultados Empíricos . . . . .	92
<b>5 Conclusiones</b>	<b>109</b>
§5.1 Trabajos futuros . . . . .	110
<b>Bibliografía</b>	<b>112</b>

# Índice de figuras

4.1	Gráfico de convergencia de las muestras de $\phi_1$ y $\phi_2$ respectivamente. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.	
	<b>Fuente: Elaboración propia</b> . . . . .	79
4.2	Gráfico de convergencia de las muestras de los elementos de $\Sigma$ . La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.	
	<b>Fuente: Elaboración propia</b> . . . . .	81
4.3	Gráfico de convergencia de las muestras de los elementos del vector termino de intercepción y las matrices de coeficientes VAR. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.	
	<b>Fuente: Elaboración propia</b> . . . . .	82
4.4	Gráfico de convergencia de las muestras de los elementos del vector termino de intercepción y las matrices de coeficientes VAR. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.	
	<b>Fuente: Elaboración propia</b> . . . . .	83
4.5	Función de autocorrelación de $\phi_1$ y $\phi_2$ respectivamente.	
	<b>Fuente: Elaboración propia</b> . . . . .	84
4.6	Función de autocorrelación de los elementos de $\Sigma$ .	
	<b>Fuente: Elaboración propia</b> . . . . .	85
4.7	Función de autocorrelación del vector termino de interceptación y las matrices de coeficientes VAR.	
	<b>Fuente: Elaboración propia</b> . . . . .	86
4.8	Función de autocorrelación de las matrices de coeficientes VAR.	
	<b>Fuente: Elaboración propia</b> . . . . .	87

4.9	Histograma de $\phi_1$ y $\phi_2$ respectivamente.	
	<b>Fuente: Elaboración propia</b> . . . . .	88
4.10	Histograma de los elementos de $\Sigma$ .	
	<b>Fuente: Elaboración propia</b> . . . . .	89
4.11	Histograma del vector termino de interceptación y las matrices de coeficientes VAR.	
	<b>Fuente: Elaboración propia</b> . . . . .	90
4.12	Histograma de las matrices de coeficientes VAR.	
	<b>Fuente: Elaboración propia</b> . . . . .	91
4.13	Gráficos temporales diarios de los precios de cierre desde el 12 de enero del 2004 hasta el 13 de abril del 2022: (a) Precios de cierre de Apple y (b) Precios de cierre de Microsoft.	
	<b>Fuente: Elaboración propia</b> . . . . .	92
4.14	Gráficos temporales de los retornos: (a) Retornos de Apple y (b) Retornos de Microsoft.	
	<b>Fuente: Elaboración propia</b> . . . . .	93
4.15	Histograma de los retornos: (a) Retornos de Apple y (b) Retornos de Microsoft.	
	<b>Fuente: Elaboración propia</b> . . . . .	94
4.16	Gráfico $Q - Q$ normal de los retornos: (a) Retornos de Apple y (b) Retornos de Microsoft.	
	<b>Fuente: Elaboración propia</b> . . . . .	94
4.17	Función de autocorrelación de los retornos y los retornos al cuadrado: (a) Retornos de Apple, (b) Retornos de Microsoft, (c) Retornos al cuadrado de Apple y (d) Retornos al cuadrado de Microsoft.	
	<b>Fuente: Elaboración propia</b> . . . . .	95
4.18	Gráfico de convergencia de las muestras de $\phi_1$ y $\phi_2$ respectivamente. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.	
	<b>Fuente: Elaboración propia</b> . . . . .	99



4.19	Gráfico de convergencia de las muestras de los elementos de la matriz de covarianza $\Sigma$ . La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.	
	<b>Fuente: Elaboración propia</b> . . . . .	100
4.20	Gráfico de convergencia de las muestras de los elementos del vector término de intercepción y las matrices de coeficientes VAR. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.	
	<b>Fuente: Elaboración propia</b> . . . . .	101
4.21	Gráfico de convergencia de las muestras de los elementos del vector término de intercepción y las matrices de coeficientes VAR. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.	
	<b>Fuente: Elaboración propia</b> . . . . .	102
4.22	Función de autocorrelación de $\phi_1$ y $\phi_2$ respectivamente.	
	<b>Fuente: Elaboración propia</b> . . . . .	102
4.23	Función de autocorrelación de los elementos de la matriz de covarianza $\Sigma$ .	
	<b>Fuente: Elaboración propia</b> . . . . .	103
4.24	Función de autocorrelación de los elementos del vector término de intercepción y las matrices de coeficientes VAR.	
	<b>Fuente: Elaboración propia</b> . . . . .	104
4.25	Función de autocorrelación de los elementos de las matrices de coeficientes VAR.	
	<b>Fuente: Elaboración propia</b> . . . . .	105
4.26	Histograma de $\phi_1$ y $\phi_2$ respectivamente.	
	<b>Fuente: Elaboración propia</b> . . . . .	105
4.27	Histograma de los elementos de la matriz de covarianza $\Sigma$ .	
	<b>Fuente: Elaboración propia</b> . . . . .	106
4.28	Histograma de los elementos del vector término de intercepción y las matrices de coeficientes VAR.	
	<b>Fuente: Elaboración propia</b> . . . . .	107
4.29	Histograma de los elementos de las matrices de coeficientes VAR.	
	<b>Fuente: Elaboración propia</b> . . . . .	108

# Índice de tablas

3.1	Métodos de estimación de modelos SV. . . . .	31
4.1	Media posterior, Desviación estándar e Intervalos de credibilidad del 95 % de los elementos de la matriz $\Phi$ y la matriz de covarianza $\Sigma$ . . . . .	79
4.2	Media posterior, Desviación estándar e Intervalos de credibilidad del 95 % de los elementos del vector término intersección $v$ y las matrices de coeficientes $A_1, A_2, \dots, A_6$ . . . . .	80
4.3	Estadísticos descriptivos para los retornos de Apple y Microsoft. . . . .	96
4.4	Medias de las muestras, errores estándar, valores más grandes y valores más pequeños DIC de los modelos VAR(4)-MSV, VAR(5)-MSV y VAR(6)-MSV. . . . .	96
4.5	Media posterior, Desviación estándar e Intervalos de credibilidad del 95 % de los elementos de la matriz $\Phi$ y la matriz de covarianza $\Sigma$ . . . . .	97
4.6	Media posterior, Desviación estándar e Intervalos de credibilidad del 95 % de los elementos del vector término intersección $v$ y las matrices de coeficientes $A_1, A_2, \dots, A_5$ . . . . .	98

# Capítulo 1

## Introducción

En el área de las finanzas es de interés los precios de los activos, cómo varían en el tiempo y el riesgo asociado con esa variación, J.-C. Duan et al. [35] comentan: *"hay muchos factores que motivan el estudio de datos financieros. Los inversores, especuladores y operadores buscan una ventaja sobre otros en la negociación de los activos financieros. Los académicos a menudo encuentran un motivo diferente para estudiar datos financieros solo por los desafíos de desarrollar modelos para los movimientos de precios. Finalmente, los reguladores gubernamentales y otros están motivados por el interés de mantener un mercado justo y ordenado. Con más y más jubilados que dependen de las inversiones del capital para su sustento, se vuelve muy importante comprender y controlar el riesgo en las carteras de acciones corporativas. El estudio de las características de una corporación y su relación con el estado financiero actual y futuro de la cooperación es un tema importante en el campo de las finanzas"*.

En el subcampo de las finanzas conocido como "matemáticas financieras", un objetivo principal es desarrollar y estudiar modelos del movimiento de precios de mercado de los activos financieros básicos. El propósito de estos modelos no es predecir precios futuros; más bien, el propósito es proporcionar una descripción del comportamiento estocástico de los precios. Para lograr describir este comportamiento, usualmente, en lugar del precio en sí, se considera el cambio relativo en el precio, es decir, la tasa de los retornos, durante un intervalo de tiempo. La tasa de los retornos tiene un fuerte componente estocástico, en su forma más simple es solo la diferencia de precio entre

dos puntos de tiempo dividida por el precio en el primer punto de tiempo, pero más a menudo es la diferencia en el logaritmo del precio. Los retornos dependen de la duración del intervalo de tiempo, por lo que se puede hablar de retornos "semanales", retornos "diarios", etc.

Para entender mejor los retornos de activos. Sea  $P_t$  el precio de un activo en el índice de tiempo  $t$ , y supongamos que el activo no paga dividendos (al pagar dividendos los retornos se modelan diferente). Existen muchas definiciones de retornos de activos de las cuales dos se mencionarán en este texto. Más definiciones de retornos de activos se pueden encontrar en Campbell et al. [16].

La primera es la de retornos simples o retornos simples netos en un período de tiempo que se define de la siguiente manera

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1.1)$$

donde  $1 + R_t = \frac{P_t}{P_{t-1}}$  es el retorno bruto simple, en el que el activo se mantiene durante un período desde el tiempo  $t - 1$  al tiempo  $t$ .

De esta definición se desprende la idea de retorno bruto simple de  $h$  períodos, que es cuando el activo se mantiene entre los tiempos  $t - h$  y  $t$ . EL retorno bruto simple de  $h$  períodos es dado por

$$1 + R_t(h) = (1 + R_t)(1 + R_{t-1}) \dots (1 + R_{t-h+1}) = \frac{P_t}{P_{t-h}}. \quad (1.2)$$

De esta manera, el retorno bruto simple de  $h$  períodos es solo el producto de los  $h$  retornos brutos simples involucrados de un período. La ecuación (1.2) es llamada retornos de períodos múltiples o retornos compuestos.

El retorno neto simple del período  $h$  es

$$R_t(h) = \frac{P_t - P_{t-h}}{P_{t-h}}. \quad (1.3)$$

Una segunda definición es la de retornos continuamente compuestos, en el que se calcula el logaritmo natural del retorno bruto simple de un activo y se escribe de la siguiente manera

$$r_t = \ln(1 + R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) = p_t - p_{t-1}, \quad p_t = \ln(P_t). \quad (1.4)$$

La ventaja de los retornos continuamente compuestos es cuando se considera los retornos de períodos múltiples

$$r_t(h) = \ln(1 + R_t(h)) = \sum_{i=1}^h \ln(1 + R_{t-i+1}) = \sum_{i=1}^h r_{t-i+1} \quad (1.5)$$

y por tanto, los retornos de períodos múltiples es simplemente la suma de los retornos de un sólo período compuesto continuamente.

Otra ventaja es que el modelado del comportamiento estadístico de los retornos compuestos continuamente de los activos a lo largo del tiempo, es mucho más fácil para derivar las propiedades de las series de tiempo.

Los retornos compuestos continuamente se conocen también como retornos logarítmicos. Cuando existe una relación lineal entre un retorno logarítmico en el tiempo  $t$  y la información disponible antes del tiempo  $t$  el análisis de series de tiempo lineales proporciona un marco natural para estudiar la estructura dinámica de dichas series.

La teoría de series de tiempo lineal incluye estacionaridad, dependencia dinámica, función de autocorrelación, modelado y pronóstico. Los modelos más sencillos de series de tiempo lineal son los autorregresivos (AR), media móvil (MA), media móvil autorregresivo (ARMA), que combina las ideas de los modelos AR y MA en una forma compacta para que el número de parámetros se mantenga pequeño, y el autorregresivo integrado media móvil (ARIMA), en el que la tendencia y la estacionalidad son eliminadas mediante la diferenciación al comienzo del análisis.

En el caso de múltiples retornos logarítmicos el análisis de series de tiempo multivariado es apropiado para entender la relación dinámica entre los retornos. El modelo autorregresivo vectorial (VAR) es útil para modelar los múltiples retornos. Para una comprensión más detallada de los modelos AR, MA, ARMA, ARIMA y VAR ver Campbell et al. [16], Tsay [100], Lütkepohl [65], Brandt y Williams [14].

La variabilidad de los retornos logarítmicos es una de las cantidades más importantes en los estudios financieros. La desviación estándar de los retornos logarítmicos es una medida de esta variabilidad y se conoce como "*volatilidad*". La volatilidad no es directamente observable y no es constante en el tiempo. Otras características de la volatilidad es que puede ser alta para ciertos períodos de tiempo

y baja para otros períodos, estos períodos de alta y baja volatilidad se conocen como cluster de volatilidad. Como se menciona anteriormente la volatilidad evoluciona con el tiempo de manera continua por lo que los saltos de volatilidad son raros. Los saltos son retornos extremadamente grandes, tanto positivos como negativos.

La volatilidad no diverge hasta el infinito, es decir, la volatilidad varía dentro de un rango fijo. Hablando estadísticamente, esto significa que la volatilidad suele ser estacionaria. También, la volatilidad parece reaccionar de manera diferente a un gran aumento de precios o una gran caída de precios, lo que se conoce como efecto de apalancamiento. Este efecto para una empresa en particular dice que un shock negativo (un retorno por debajo de su valor esperado) implica que la empresa está más apalancada, es decir, tiene una relación más alta entre la deuda y el valor de las acciones y, por lo tanto, es más riesgosa, por lo que la volatilidad debería aumentar.

Para medir la volatilidad muchos modelos han sido propuestos. Los modelos de retornos de un solo activo se conocen como modelos de volatilidad univariados. Algunos modelos de volatilidad univariados más conocidos son los modelos heterocedástico condicional autoregresivo (ARCH) introducido por Engle [38] y los modelos heterocedástico condicional autoregresivo generalizado (GARCH) introducidos por Bollerslev [12]. Variaciones de los modelos tipo ARCH y GARCH así como propiedades y estructuras subyacentes se pueden encontrar en Gouriéoux [48], Francq y Zakoian [42] respectivamente.

Un enfoque diferente es propuesto por French et al. [43] en el que se usan datos de alta frecuencia (por ejemplo, retornos intradiarios de 10 minutos ) para calcular la volatilidad de retornos de baja frecuencia (por ejemplo, retornos diarios). Gracias a la amplia disponibilidad de bases de datos este enfoque ha sido de gran interés en los últimos años y es conocido como volatilidad realizada (VR). Una explicación más detallada sobre la volatilidad realizada se puede encontrar en Anderson et al. [4] y Anderson et al. [3]. Otra clase de modelos univariados son los modelos de volatilidad estocástica (SV) en los que se postula que la volatilidad es impulsada por su propio proceso estocástico. Para un mejor entendimiento sobre los modelos de volatilidad estocástica ver Taylor [95].

Los modelos de múltiples retornos de activos se conocen como modelos de volatilidad multivariados y deben tomar en cuenta la correlación entre los retornos. Las volatilidades multivariadas tienen

muchas aplicaciones financieras importantes, por ejemplo, desempeñan un papel importante en la selección de carteras y la asignación de activos y se pueden utilizar para calcular el valor en riesgo de una posición financiera que consta de múltiples activos. Muchos autores han generalizado los modelos de volatilidad univariado al caso multivariado, entre ellos están los modelos GARCH multivariados, ver por ejemplo Bauwens et al. [9].

Los modelos de covarianza realizada en el que la información adicional en los datos de alta frecuencia, también se pueden explotar cuando se observa la covarianza, la correlación y las regresiones simples. Algunas referencias sobre los modelos de covarianza realizada son Anderson et al. [4], Anderson et al. [3], Barndorff-Nielsen y Shepard [7]. Los modelos SV también han sido generalizados al caso multivariado (MSV). Se pueden encontrar más especificaciones de los modelos MSV en Asai et al. [6], Yu y Meyer [104].

Al integrar los modelos VAR con los modelos MSV se obtiene un tipo de modelo diferente llamado modelos autorregresivos vectoriales con volatilidad estocástica multivariados (VAR-MSV). La volatilidad estocástica en los modelos VAR-MSV está destinada a capturar la posible heteroscedasticidad de los choques y no linealidades en las relaciones simultáneas entre las variables del modelo. Para una comprensión más detallada de los modelos VAR-MSV ver Primiceri [78] y Triantafyllopoulos [99].

En este trabajo se propone integrar un modelo VAR con un modelo MSV con efecto de apalancamiento cruzado. Para estimar los parámetros se usan métodos de Monte Carlo vía Cadenas de Markov (MCMC). El modelo propuesto se aplica a series bivariadas financieras compuesto de precios de Apple y de Microsoft.

En el capítulo 2 se hablará sobre los procesos autorregresivos vectoriales (VAR) , algunas de sus propiedades, métodos de estimación, selección del orden y comprobación de modelo adecuado. En el capítulo 3 se exponen algunos modelos de volatilidad estocástica multivariados métodos de estimación y comprobación de modelo adecuado . En el capítulo 4 se presenta la integración de un modelo VAR con un modelo de volatilidad estocástica multivariado (MSV), se explicaran algunas de sus propiedades y se expondrán algunos ejemplos para su mayor comprensión. En el capítulo 5 se exponen las conclusiones.

## Capítulo 2

# Procesos Autorregresivos Vectoriales Estables

En una serie de artículos el macroeconomista Christopher A. Sims [86–89] propuso el uso de los modelos autorregresivos vectoriales (VAR), para modelar la dinámica y las relaciones causales entre un conjunto de variables macroeconómicas. Desde entonces los modelos VAR han demostrado ser especialmente útiles para describir el comportamiento dinámico de series temporales económicas y financieras y para realizar pronósticos.

Los modelos VAR se utilizan para la inferencia estructural y análisis de políticas. Su configuración es tal que los valores actuales de un conjunto de variables se explican en parte por los valores pasados de las variables involucradas, esto se debe a que a menudo el valor de una variable no solo está relacionada con sus predecesores en el tiempo, sino que, además depende de valores pasados de otras variables. En los modelos VAR para cada variable endógena en el sistema de ecuaciones, se construye una ecuación que hace que estas variables dependan de sus propios valores pasados y los valores pasados de todas las demás variables endógenas. Normalmente, esto se hará con el mismo número de retrasos o valores pasados de cada variable en cada ecuación.

Los modelos VAR se pueden configurar de tres maneras: Forma reducida, recursiva y estructural.

- i) Un VAR en su forma reducida expresa cada variable como función lineal de sus propios valores



pasados, los valores pasados de todas las demás variables y un término de error serialmente no correlacionado.

- ii) Un VAR recursivo construye los términos de error en cada ecuación para que no estén correlacionados con el error en las ecuaciones anteriores. Esto se hace incluyendo juiciosamente algunos valores contemporáneos como regresores.
- ii) Un VAR estructural utiliza la teoría económica para clasificar los vínculos contemporáneos entre las variables.

El resto de este capítulo está organizado de la siguiente manera. En la sección 2.1 se presentan algunas suposiciones y propiedades de los modelos VAR. En la sección 2.2 se presentan algunos métodos de estimación. En la sección 2.3 se presentan métodos para seleccionar el orden VAR. 2.4 se exponen algunos métodos para verificación del modelo adecuado.

## 2.1. Propiedades y suposiciones de los modelos VAR

El modelo VAR ( $k$ ) puede ser escrito de la siguiente manera

$$y_t = v + A_1 y_{t-1} + \dots + A_k y_{t-k} + \varepsilon_t \quad , t = 0, \pm 1, \pm 2, \dots \quad (2.1)$$

donde  $y_t = (y_{1t}, \dots, y_{pt})^T$  es un vector aleatorio de dimensión  $p \times 1$ , los  $A_i$  son matrices de coeficientes fijas de dimensión  $p \times p$ ,  $v = (v_1, \dots, v_p)^T$  es un vector fijo de los términos de intercepción de dimensión  $p \times 1$ , y  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{pt})^T$  es un proceso de ruido blanco de dimensión  $p \times 1$ , donde  $E(\varepsilon_t) = 0$ ,  $E(\varepsilon_t \varepsilon_t^T) = \Sigma_\varepsilon$  y  $E(\varepsilon_t \varepsilon_s^T) = 0$  para  $s \neq t$ . En este capítulo supondremos que la matriz de covarianza  $\Sigma_\varepsilon$  es no singular.

A manera de ejemplo, supongamos el caso donde  $p = 2$  y  $k = 1$ , en este caso tenemos un modelo VAR (1) el cual se puede escribir de la siguiente manera.

$$\begin{cases} y_{1t} = v_1 + A_{11}^1 y_{1,t-1} + A_{12}^1 y_{2,t-1} + \varepsilon_{1t} \\ y_{2t} = v_2 + A_{21}^1 y_{1,t-1} + A_{22}^1 y_{2,t-1} + \varepsilon_{2t} \end{cases} \quad (2.2)$$

donde  $v_i$  es el  $i$ -ésimo elemento de  $v$  y  $A^1_{ij}$  es el  $(i, j)$ -ésimo elemento de  $A_1$ . En la ecuación (1.2)  $A^1_{12}$  denota la dependencia lineal de  $y_{1t}$  en  $y_{2,t-1}$  en la presencia de  $y_{1,t-1}$ . Por lo que  $A^1_{12}$  es el efecto condicional de  $y_{2,t-1}$  en  $y_{1t}$  dado  $y_{1,t-1}$ . Si  $A^1_{12} = 0$ , entonces  $y_{1t}$  no depende de  $y_{2,t-1}$ , y el modelo muestra que  $y_{1t}$  únicamente depende de su propio pasado.

Consideremos las dos ecuaciones conjuntamente. Si  $A^1_{12} = 0$  y  $A^1_{21} \neq 0$  entonces hay una relación unidireccional de  $y_{1t}$  a  $y_{2t}$ . Si  $A^1_{12} = A^1_{21} = 0$  entonces  $y_{1t}$  y  $y_{2t}$  están desacoplados. Si  $A^1_{12} \neq 0$  y  $A^1_{21} \neq 0$ , entonces hay una relación de retroalimentación entre las dos series. Por lo tanto, los coeficientes de la matriz  $A_1$  miden la dependencia dinámica de  $y_t$ . La relación concurrente entre  $y_{1t}$  y  $y_{2t}$  es mostrada por el elemento  $\sigma_{12}$  de  $\Sigma_\varepsilon$ . Si  $\sigma_{12} = 0$ , entonces no hay relación lineal concurrente entre los dos componentes de la serie.

Suponiendo  $p > 0$  y  $k = 1$  en la ecuación (2.1) entonces el modelo resultante es de la siguiente manera

$$y_t = v + A_1 y_{t-1} + \varepsilon_t. \quad (2.3)$$

Haciendo algunos cálculos recursivos comenzando en  $t = 1$  en la ecuación (2.3) obtenemos la siguiente expresión

$$y_t = (I_p + A_1 + \dots + A_1^{t-1})v + A_1^t y_0 + \sum_{i=0}^{t-1} A_1^i \varepsilon_{t-i}. \quad (2.4)$$

La ecuación (2.4) muestra que cualquiera de los vectores  $y_1, \dots, y_t$  pueden ser determinados únicamente por  $y_0, \varepsilon_1, \dots, \varepsilon_t$ . También la distribución conjunta de  $y_1, \dots, y_t$  es determinada por la distribución conjunta de  $y_0, \varepsilon_1, \dots, \varepsilon_t$ .

El modelo VAR ( $k$ ) de la ecuación (2.1) puede ser escrito de diferentes formas (ver por ejemplo Canova [17]; Kadiyala y Karlsson [58]). En este capítulo lo escribiremos por conveniencia en su forma VAR (1) de la siguiente manera

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}\mathbf{y}_{t-1} + \varepsilon_t \quad (2.5)$$

donde

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{k-1} & A_k \\ I_p & 0 & \dots & 0 & 0 \\ 0 & I_p & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_p & 0 \end{bmatrix}, \quad \text{de dimensión } pk \times pk$$

$$\mathbf{y}_t = [y_t^T, y_{t-1}^T, \dots, y_{t-k+1}^T]^T, \quad \text{de dimensión } pk \times 1$$

$$\mathbf{v} = [v^T, 0, \dots, 0]^T, \quad \text{de dimensión } pk \times 1$$

$$\boldsymbol{\varepsilon}_t = [\varepsilon_t^T, 0, \dots, 0]^T, \quad \text{de dimensión } pk \times 1$$

La ecuación (2.5) la podemos escribir de la siguiente manera

$$\mathbf{y}_t = (I_{pk} + \mathbf{A} + \dots + \mathbf{A}^{t-1})\mathbf{v} + \mathbf{A}^t \mathbf{y}_0 + \sum_{i=0}^{t-1} \mathbf{A}^i \boldsymbol{\varepsilon}_{t-i}. \quad (2.6)$$

Si además se supone que el proceso ha comenzado en el pasado infinito y todos los autovalores de  $\mathbf{A}$  son menores que 1, se puede demostrar (ver, Lutkepohl [65]) que la ecuación (2.6) es un proceso bien definido y se puede escribir de la siguiente manera

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}^i \boldsymbol{\varepsilon}_{t-i}, \quad t = 0, \pm 1, \pm 2, \dots \quad (2.7)$$

donde

$$\mathbf{E}(\mathbf{y}_t) = \boldsymbol{\mu} = (I_{pk} - \mathbf{A})^{-1} \mathbf{v} \quad (2.8)$$

es el vector media y las autocovarianzas son

$$\Gamma_{\mathbf{y}}(h) = \sum_{i=0}^{\infty} \mathbf{A}^{h+i} \Sigma_{\boldsymbol{\varepsilon}} (\mathbf{A}^i)^T, \quad (2.9)$$

$$\Sigma_{\boldsymbol{\varepsilon}} = \mathbf{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}_t^T).$$

Cuando todos los autovalores de  $\mathbf{A}$  son menores que 1 se dice que  $\mathbf{y}_t$  es estable. La condición de estabilidad es equivalente a

$$\det(I_{pk} - \mathbf{A}z) \neq 0, \quad \text{para } |z| \leq 1. \quad (2.10)$$

Para obtener el proceso  $y_t$  se define la matriz  $J = [I_p : 0 : \dots : 0]$  de dimensión  $p \times pk$  tal que  $y_t = J \mathbf{y}_t$ . Además, como el proceso estocástico  $\mathbf{y}_t$  es bien definido, entonces el proceso  $y_t$  es bien definido. El vector medio y las autocovarianzas son de la siguiente manera

$$E(y_t) = J \boldsymbol{\mu}, \quad \Gamma_y(h) = J \Gamma_{\mathbf{y}}(h) J^T.$$

Un método más formal pero equivalente de comprobar que  $y_t$  es estable es de la siguiente manera

$$\det(I_p - A_1 z - \dots - A_k z^k) \neq 0, \quad \text{para } |z| \leq 1. \quad (2.11)$$

Por lo cual, se puede decir que  $y_t$  es un proceso VAR ( $k$ ) estable si las condiciones de la ecuación (2.11) se mantienen y

$$y_t = J \mathbf{y}_t = J \boldsymbol{\mu} + J \sum_{i=0}^{\infty} \mathbf{A}^i \boldsymbol{\varepsilon}_{t-i}. \quad (2.12)$$

Por la ecuación (2.12) se puede notar que el proceso  $y_t$  puede ser determinado por el proceso de ruido blanco  $\boldsymbol{\varepsilon}_t = (\varepsilon_t^T, 0, \dots, 0)^T$  por lo que a menudo, se hacen suposiciones específicas con respecto a  $\varepsilon_t$ . Por ejemplo, si se supone que  $\varepsilon_t$  es un ruido blanco gaussiano, esto es,  $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$  para todo  $t$ ,  $\varepsilon_t$  y  $\varepsilon_s$  son independientes para  $s \neq t$ , entonces  $y_t$  es un proceso gaussiano.

El proceso de la ecuación (2.7) es conocida como media móvil (MA) donde  $\mathbf{y}_t$  es expresado en términos de vectores de errores pasados y presentes  $\boldsymbol{\varepsilon}_t$  y el término medio  $\boldsymbol{\mu}$ . La representación MA de  $y_t$  es de la siguiente forma

$$\begin{aligned} y_t = J \mathbf{y}_t &= J \boldsymbol{\mu} + \sum_{i=0}^{\infty} J \mathbf{A}^i J^T J \boldsymbol{\varepsilon}_{t-i} \\ &= \boldsymbol{\mu} + \sum_{i=0}^{\infty} \Phi_i \boldsymbol{\varepsilon}_{t-i}. \end{aligned} \quad (2.13)$$

donde  $\mu = J\boldsymbol{\mu}$ ,  $\Phi_i = J\mathbf{A}^i J^T$  y  $\Phi_0 = I_p$ . Esto es posible debido a que el proceso de ruido blanco  $\boldsymbol{\varepsilon}_t$  tiene las siguientes propiedades

$$\boldsymbol{\varepsilon}_t = J^T J \boldsymbol{\varepsilon}_t \quad \text{y} \quad J \boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}_t.$$

La media y autocovarianza del proceso  $y_t$  de la ecuación (2.13) es de la siguiente manera

$$E(y_t) = \mu, \tag{2.14}$$

$$\Gamma_y(h) = \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_{\varepsilon} \Phi_i^T. \tag{2.15}$$

Un método para encontrar las matrices  $\Phi_i$  está en Lutkepohl [65].

Una propiedad importante de un proceso estocástico es la de estacionariedad en el que el primer y segundo momento son invariantes en el tiempo. En otras palabras un proceso estocástico  $y_t$  es estacionario si

$$E(y_t) = \mu, \quad \forall t. \tag{2.16}$$

y

$$E[(y_t - \mu)(y_{t-h} - \mu)^T] = \Gamma_y(h) = \Gamma_y(-h)^T, \quad \forall t \text{ y } h = 0, 1, \dots \tag{2.17}$$

De las ecuaciones (2.14)-(2.15) se puede ver que un proceso VAR ( $k$ ) estable es estacionario.

Ajustando el proceso (2.1) a la forma ajustada a la media y multiplicando posteriormente ambos lados de la ecuación por el término  $(y_{t-h} - \mu)$  y tomando la esperanza se obtienen las ecuaciones *Yule-Walker* que pueden ser utilizadas para calcular las autocovarianzas de manera recursiva y así evitar las sumas infinitas de las ecuaciones (2.9) y (2.15). Siempre que  $A_1, \dots, A_k$  y  $\Gamma_y(k-1), \dots, \Gamma_y(0)$  sean conocidas

$$\Gamma_y(h) = \begin{cases} A_1 \Gamma_y(h-1) + \dots + A_k \Gamma_y(h-k), & h > 0 \\ A_1 \Gamma_y(1)^T + \dots + A_k \Gamma_y(k)^T + \Sigma_{\varepsilon}, & h = 0. \end{cases} \tag{2.18}$$

Las ecuaciones de Yule-Walker del proceso (2.5) son de la siguiente manera

$$\Gamma_{\mathbf{y}}(h) = \begin{cases} \mathbf{A} \Gamma_{\mathbf{y}}(h-1), & h > 0 \\ \mathbf{A} \Gamma_{\mathbf{y}}(0) \mathbf{A}^T + \Sigma_{\boldsymbol{\varepsilon}}, & h = 0. \end{cases} \quad (2.19)$$

De la ecuación (2.19) se pueden obtener las matrices de autocovarianza iniciales

$$\begin{aligned} \text{vec}(\Gamma_{\mathbf{y}}(0)) &= (I_{(pk)^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec}(\Sigma_{\boldsymbol{\varepsilon}}) \Gamma_{\mathbf{y}}(0) \\ &= \begin{bmatrix} \gamma_{11}(0) & \dots & \gamma_{1p}(0) \\ \vdots & \ddots & \vdots \\ \gamma_{p1}(0) & \dots & \gamma_{pp}(0) \end{bmatrix} \end{aligned}$$

donde el operador  $\text{vec}$  apila las columnas de una matriz en un solo vector. Mientras que  $\otimes$  denota el producto de Kronecker.

Debido a que las autocovarianzas suelen ser medidas variantes de escala de las dependencias lineales entre las variables del sistema, suele ser más conveniente usar las autocorrelaciones que se pueden escribir de la siguiente manera

$$R_{\mathbf{y}}(h) = D^{-1} \Gamma_{\mathbf{y}}(h) D^{-1} \quad (2.20)$$

donde  $D = \text{diag}(\sqrt{\gamma_{11}(0)}, \dots, \sqrt{\gamma_{pp}(0)})$ .

## 2.2. Métodos de Estimación

Existen muchos métodos para estimar los parámetros del modelo  $\text{VAR}(k)$  de la ecuación (2.1) de los cuales se presentan algunos en esta sección.

Suponiendo que los vectores de series de tiempo  $y_1, \dots, y_n$  son conocidos, esto es, se tiene una

muestra de tamaño  $n$ . Además,  $k$  valores pre-muestrales  $y_{-k+1}, \dots, y_0$  están disponibles. Se define

$$\begin{aligned}
 Y &= (y_1, \dots, y_n) \quad \text{de dimensión } p \times n \\
 B &= (v, A_1, \dots, A_k) \quad \text{de dimensión } p \times (pk + 1) \\
 Z_t &= [1 \quad y_t^T \quad \dots \quad y_{t-p+1}^T]^T \quad \text{de dimensión } (pk + 1) \times 1 \\
 Z &= (Z_0, \dots, Z_{n-1})^T \quad \text{de dimensión } (pk + 1) \times n \\
 \varepsilon &= (\varepsilon_1, \dots, \varepsilon_n) \quad \text{de dimensión } p \times n \\
 \mathbf{y} &= \text{vec}(Y) \quad \text{de dimensión } pn \times 1 \\
 \boldsymbol{\beta} &= \text{vec}(B) \quad \text{de dimensión } (p^2k + p) \times 1 \\
 \mathbf{b} &= \text{vec}(B^T) \quad \text{de dimensión } (p^2k + p) \times 1 \\
 \boldsymbol{\varepsilon} &= \text{vec}(\varepsilon) \quad \text{de dimensión } pn \times 1
 \end{aligned}$$

Usando esta notación, el modelo VAR ( $k$ ) de la ecuación (2.1) puede ser escrito de las siguientes maneras

$$Y = BZ + \varepsilon, \tag{2.21}$$

$$\mathbf{y} = (Z^T \otimes I_p) \boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{2.22}$$

### 2.2.1. Método de mínimos cuadrados multivariado

El estimador de mínimos cuadrados multivariados (GLS) se puede escribir de las siguientes maneras

$$\hat{\boldsymbol{\beta}} = [(ZZ^T) Z \otimes I_p] \mathbf{y}, \tag{2.23}$$

$$\hat{B} = YZ^T (ZZ^T)^{-1}, \tag{2.24}$$

$$\hat{\mathbf{b}} = [I_p \otimes (ZZ^T)^{-1} Z] \text{vec}(Y^T). \tag{2.25}$$

Mann y Wald [66] demostraron que si  $y_t$  es un proceso VAR ( $k$ ) estable y estacionario, y bajo condiciones deseables para  $\varepsilon_t$  entonces

$$\Gamma = p \lim \frac{1}{n} Z Z^T \quad \text{existe y es no singular} \quad (2.26)$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \text{vec} (\varepsilon_t Z_{t-1}^T) = \frac{1}{\sqrt{n}} \text{vec} (\varepsilon Z^T) = \frac{1}{\sqrt{n}} (Z \otimes I_k) \boldsymbol{\varepsilon} \quad \xrightarrow[n \rightarrow \infty]{d} N(0, \Gamma \otimes \Sigma_\varepsilon). \quad (2.27)$$

La ecuación (2.26) significa que la matriz  $\frac{1}{n} Z Z^T$  cuando  $n \rightarrow \infty$  converge en probabilidad a la matriz no singular  $\Gamma$ . La ecuación (2.27) significa que cuando  $n \rightarrow \infty$  la secuencia de variables aleatorias converge en distribución a una distribución normal con media cero y matriz de covarianza  $\Gamma \otimes \Sigma_\varepsilon$ .

Una condición suficiente para que el proceso  $\varepsilon_t$  tenga las condiciones deseables para que (2.26)-(2.27) se mantengan es que  $\varepsilon_t$  es un vector aleatorio continuo satisfaciendo que  $E(\varepsilon_t) = 0$ ,  $\Sigma_\varepsilon = E(\varepsilon_t \varepsilon_t^T)$  es no singular,  $\varepsilon_s$  y  $\varepsilon_t$  son independientes para  $s \neq t$  y para alguna constante finita  $c$

$$E|\varepsilon_{it} \varepsilon_{jt} \varepsilon_{kt} \varepsilon_{mt}| \leq c \quad \text{para } i, j, k, m = 1, \dots, p \text{ y para todo } t .$$

Esta condición es llamada ruido blanco estándar.

Además, se puede demostrar (ver, Fuller [44] ) que si  $y_t$  es un proceso VAR ( $k$ ) estable con  $\varepsilon_t$  siendo un ruido blanco estándar, entonces (2.26)-(2.27) se mantienen.

Una propiedad importante del estimador GLS cuando  $y_t$  es un proceso VAR ( $k$ ) estable como en (2.1) con los vectores  $\varepsilon_t$  siendon ruidos blancos estándar es

$$p \lim \hat{B} = B, \quad (2.28)$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{n} \text{vec}(\hat{B} - B) \xrightarrow{d} N(0, \Gamma^{-1} \otimes \Sigma_\varepsilon), \quad (2.29)$$

$$\sqrt{n}(\hat{\mathbf{b}} - \mathbf{b}) = \sqrt{n} \text{vec}(\hat{B}^T - B^T) \xrightarrow{d} N(0, \Sigma_\varepsilon \otimes \Gamma^{-1}). \quad (2.30)$$



Estimadores consistentes para  $\Sigma_\varepsilon$  y  $\Gamma$  se pueden escribir de la siguiente manera

$$\hat{\Sigma}_\varepsilon = \frac{1}{n}(Y - \hat{B}Z)(Y - \hat{B}Z)^T, \quad (2.31)$$

$$\hat{\Gamma} = \frac{1}{n}ZZ^T. \quad (2.32)$$

Una demostración de (2.28)-(2.30) se puede encontrar en Lütkepohl [65] y en Tsay [100].

### 2.2.2. Método de máxima verosimilitud

Para este método se supone que la distribución del proceso es conocida. Suponiendo que  $\varepsilon_t$  es Gaussiano entonces

$$\boldsymbol{\varepsilon} \sim N_{pn}(0, I_n \otimes \Sigma_\varepsilon), \quad (2.33)$$

$$f_\varepsilon(\boldsymbol{\varepsilon}) = \frac{1}{(2\pi)^{pn/2}} \frac{1}{|I_n \otimes \Sigma_\varepsilon|^{-1/2}} \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}^T (I_n \otimes \Sigma_\varepsilon^{-1}) \boldsymbol{\varepsilon}\right). \quad (2.34)$$

Ajustando el proceso (2.1) a la forma ajustada a la media y asumiendo que  $v = 0$  se puede escribir  $\boldsymbol{\varepsilon}$  de la siguiente manera

$$\boldsymbol{\varepsilon} = \mathbf{y} - \boldsymbol{\mu}^* - (X^T \otimes I_p)\boldsymbol{\alpha}, \quad (2.35)$$

donde

$$\boldsymbol{\mu}^* = (\mu^T, \dots, \mu^T)^T \quad \text{de dimensión } pn \times 1,$$

$$\boldsymbol{\alpha} = \text{vec}(A_1, \dots, A_k) \quad \text{de dimensión } p^2k \times 1,$$

$$X = (y_0^0, \dots, y_{n-1}^0) \quad \text{de dimensión } pk \times n,$$

$$y_t^0 = [(y_t - \mu)^T, \dots, (y_{t-k+1} - \mu)^T]^T \quad \text{de dimensión } pk \times 1.$$

La densidad de probabilidad de  $\mathbf{y}$  viene dada por

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\pi)^{pn/2} |I_n \otimes \Sigma_\varepsilon|} \exp \left[ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}^* - (X^T \otimes I_p) \boldsymbol{\alpha})^T (I_n \otimes \Sigma_\varepsilon^{-1}) (\mathbf{y} - \boldsymbol{\mu}^* - (X^T \otimes I_p) \boldsymbol{\alpha}) \right] \quad (2.36)$$

por lo que los estimadores de máximo verosimilitud (ML) de  $\mu$ ,  $\boldsymbol{\alpha}$  y  $\Sigma_\varepsilon$  son

$$\hat{\mu} = \frac{1}{n} \left( I_p - \sum_{i=1}^k \hat{A}_i \right)^{-1} \sum_{t=1}^n \left( y_t - \sum_{i=1}^k \hat{A}_i y_{t-i} \right), \quad (2.37)$$

$$\hat{\boldsymbol{\alpha}} = [(\tilde{X} \tilde{X}^T)^{-1} \tilde{X} \otimes I_p] (\mathbf{y} - \tilde{\boldsymbol{\mu}}^*), \quad (2.38)$$

$$\hat{\Sigma}_\varepsilon = \frac{1}{n} (\tilde{Y}^0 - \tilde{A} \tilde{X}) (\tilde{Y}^0 - \tilde{A} \tilde{X})^T, \quad (2.39)$$

donde

$$A = (A_1, \dots, A_k) \quad \text{de dimensión } p \times pk,$$

$$Y^0 = (y_1 - \mu, \dots, y_n - \mu)^T \quad \text{de dimensión } p \times n.$$

$\tilde{X}$  y  $\tilde{Y}^0$  son obtenidos reemplazando  $\mu$  por la media muestral de  $y_1, \dots, y_n$  en  $X$  y  $Y^0$  respectivamente.

Una propiedad de los estimadores ML es la siguiente:

**Teorema.** Si  $\hat{\delta}$  es un estimador ML del parámetro  $\delta_0$  entonces

$$\hat{\delta} \quad \text{es consistente} \quad (2.40)$$

y

$$\sqrt{n} (\hat{\delta} - \delta_0) \xrightarrow{d} N(0, I_a(\delta_0)^{-1}), \quad (2.41)$$

donde

$$I_a(\delta_0) = \lim_{n \rightarrow \infty} \frac{1}{n} I(\delta_0),$$

$I(\delta_0)$  es la matriz de información de  $\delta_0$ .

Una demostración de (2.40)-(2.41) se puede encontrar en Newey y McFadden [72]. Por (2.40)-(2.41) se puede demostrar (ver, Lütkepohl [65]) que los estimadores ML de (2.37)-(2.39) son consistentes y tienen distribuciones asintóticas

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N_p \left( 0, \left( I_p - \sum_{i=1}^k A_i \right)^{-1} \Sigma_\varepsilon \left( I_p - \sum_{i=1}^k A_i \right)^{-1} \right), \quad (2.42)$$

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} N_{p^2k} (0, \Gamma_{\mathbf{y}}(0)^{-1} \otimes \Sigma_\varepsilon), \quad (2.43)$$

$$\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{d} N(0, 2D_p^+(\Sigma_\varepsilon \otimes \Sigma_\varepsilon)(D_p^+)^T), \quad (2.44)$$

donde

$$\sigma = \text{vech}(\Sigma_\varepsilon),$$

$$D_p^+ = (D_p^T D_p)^{-1} D_p^T \quad \text{es la inversa generalizada Moore-Penrose de la matriz duplicación.}$$

La matriz duplicación  $D_p$  es de dimensión  $(p^2 \times \frac{1}{2}p(p+1))$  y es definida tal que, para cualquier matriz simétrica  $A$  de dimensión  $(p \times p)$

$$\begin{aligned} \text{vec}(A) &= D_p \text{vech}(A), \\ \text{vech}(A) &= \text{vech} \begin{bmatrix} a_{11} & a_{21} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & & \ddots & \vdots \\ a_{p1} & \dots & \dots & a_{pp} \end{bmatrix} \\ &= [a_{11}, a_{12}, \dots, a_{p1}, a_{22}, a_{32}, \dots, a_{p2}, a_{33}, a_{43}, \dots, a_{pp}]^T, \end{aligned}$$

donde  $\Gamma_{\mathbf{y}}(0)$  puede ser estimado por  $\frac{1}{n} \tilde{X} \tilde{X}^T$ .

### 2.2.3. Método de estimación bayesiano

Para entender el método de estimación bayesiano se reescribe la ecuación (2.1) de la siguiente manera

$$y_t = [v, A_1, A_2, \dots, A_k] \begin{bmatrix} 1 \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-k} \end{bmatrix} + \varepsilon_t = (Y_t^T \otimes I_p) \beta + \varepsilon_t, \quad (2.45)$$

donde  $(Y_t^T \otimes I_p)$  es de dimensión  $p \times (kp + 1)$ ,  $Y_t = [1, y_{t-1}^T, y_{t-2}^T, \dots, y_{t-k}^T]^T$  es de dimensión  $(kp + 1) \times 1$ ,  $\beta = \text{vec}(v, A_1, A_2, \dots, A_k)$  es de dimensión  $(kp^2 + p) \times 1$  y,  $y_0, \dots, y_{-k+1}$  son valores premuestrales.

Si se supone que  $\varepsilon_t$  es Gaussiano con  $E(\varepsilon_t) = 0, \forall t$ ;  $E(\varepsilon_t \varepsilon_s^T) = 0$  para  $t \neq s$  y,  $E(\varepsilon_t \varepsilon_t^T) = \Sigma_\varepsilon$  los parámetros desconocidos de (2.45) son  $\beta$  y  $\Sigma_\varepsilon$ .

Uno de los métodos usados para hacer estimaciones de los parámetros  $\beta$  y  $\Sigma_\varepsilon$  es el muestreador de Gibbs. Este método requiere del conocimiento de la distribución posterior condicional completa de los parámetros de interés, a saber,  $\Pr(\beta | \Sigma_\varepsilon, Y)$  y  $\Pr(\Sigma_\varepsilon | \beta, Y)$ , establecer valores iniciales para  $\beta^{(0)}$  y  $\Sigma_\varepsilon^{(0)}$  y, la elección de una distribución a priori para los parámetros del modelo.

El muestreador Gibbs muestrea recursivamente de la siguiente manera:

1. Inicializar los valores iniciales  $\beta^{(0)}$  y  $\Sigma_\varepsilon^{(0)}$ .
2. Generar  $\beta^{(i)} | \Sigma_\varepsilon^{(i-1)}, Y$ .
3. Generar  $\Sigma_\varepsilon^{(i)} | \beta^{(i)}, Y$ .
4. Ir al paso 2 hasta que la convergencia es alcanzada.

Note que  $\left( (\beta^{(1)}, \Sigma_\varepsilon^{(1)}), (\beta^{(2)}, \Sigma_\varepsilon^{(2)}), \dots \right)$  definen una cadena de Markov ya que el cambio probabilístico en la iteración  $i$  depende solo de los valores de la cadena en el paso  $i - 1$ . Además,

las  $i$ -ésimas muestras  $\beta^{(i)}$  y  $\Sigma_\epsilon^{(i)}$  para un número suficientemente grande, a saber,  $i \geq j$ , pueden considerarse como muestras de las verdaderas distribuciones posteriores marginales.

Establecer los valores iniciales  $\beta^{(0)}$  y  $\Sigma_\epsilon^{(0)}$  se puede hacer de manera arbitraria. Para la elección de las distribuciones a priori, existe una gran variedad que pueden ser usadas para los modelos VAR, algunas de ellas son la priori difusa, la priori conjugada, la priori normal-Wishart, la priori normal difusa. En esta sección solo hablaremos de la priori de Litterman (ver, Doan et al. [32], Litterman [64]).

Suponiendo que la distribución priori de  $\beta$  es normal multivariante con vector media priori  $\mu_\beta$  y matriz de covarianza priori  $\Sigma_\beta$ . Entonces la priori de Litterman es de la siguiente manera

$$f(\beta) = \left(\frac{1}{2\pi}\right)^{\frac{kp^2+p}{2}} |\Sigma_\beta|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(\beta - \mu_\beta)^T \Sigma_\beta^{-1} (\beta - \mu_\beta) \right\}. \quad (2.46)$$

La priori Litterman en esta sección configura  $\mu_\beta = 0$  y  $\Sigma_\beta \neq 0$ . Esto significa que se cree que la dependencia interporal de las variables es débil. La matriz de covarianza priori  $\Sigma_\beta$  es configurada como una matriz diagonal de la siguiente manera

$$\Sigma_{\beta,ij,r} = \begin{cases} \left(\frac{\lambda}{r}\right)^2 & \text{si } i = j, \\ \left(\frac{\lambda\theta\sigma_i}{r\sigma_j}\right)^2 & \text{si } i \neq j, \end{cases} \quad (2.47)$$

donde  $\Sigma_{\beta,ij,r}$  es la varianza priori de  $A_{ij,r}$ ,  $\lambda$  es la desviación estándar priori de los coeficientes  $A_{\ell\ell,1}$ ,  $\ell = 1, 2, \dots, p$ ,  $0 < \theta < 1$  y,  $\sigma_i^2$  es el elemento de la  $i$ -ésima diagonal de  $\Sigma_\epsilon$ .

Algunas consideraciones a tomar en cuenta (Para mas detalles ver Lütkepohl [65]) son:

- i) Para cada ecuación,  $\lambda$  controla qué tan estrechamente se cree que el coeficiente del primer rezago de la variable dependiente está centrada alrededor de cero.
- ii) Se pueden probar diferentes valores de  $\lambda$ .
- iii) Se puede considerar el uso de diferentes  $\lambda$  en diferentes ecuaciones.
- iv) Al aumentar el rezago  $r$ , la matriz de covarianza disminuye. Esto se debe a que se cree que

los coeficientes de lag de alto orden probablemente estén cerca de cero.

- v) A los coeficientes de las variables distintas de la variable dependiente se les asigna una varianza más pequeña en términos relativos, eligiendo  $\theta$  entre 0 y 1 debido a que se cree que la mayor parte de la variación en cada una de las variables se explica por los propios lags.
- vi) La razón  $\frac{\sigma_i^2}{\sigma_j^2}$  se incluye para cuidar las diferencias en la variabilidad de las diferentes variables.
- vii) Se prefieren las varianzas residuales sobre las varianzas  $y_k$  debido a que la respuesta de una variable a otra está determinada en gran medida por los movimientos inesperados reflejados en las varianzas residuales.
- viii) El supuesto de la matriz diagonal  $\Sigma_\beta$  significa que se especifican distribuciones priori independientes de los diferentes coeficientes.
- ix) Las varianzas priori de los términos interceptados son infinitas, reflejando que no se tiene ninguna conjetura a priori para estos coeficientes.

Para  $\Sigma_\varepsilon$  una distribución priori es la Wishart inversa ( $IW$ ), por lo que

$$\Sigma_\varepsilon \sim IW(S^*, \eta^*). \quad (2.48)$$

La función de densidad de probabilidad de los datos, condicional a los parámetros del modelo es dada por

$$f(y_1, \dots, y_n | \beta, \Sigma_\varepsilon) = \prod_{t=1}^n f(y_t | \beta, \Sigma_\varepsilon) \propto |\Sigma_\varepsilon|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^n (y_t - (Y_t^T \otimes I_p) \beta)^T \Sigma_\varepsilon^{-1} (y_t - (Y_t^T \otimes I_p) \beta) \right\}. \quad (2.49)$$

Usando (2.49) se obtienen las distribuciones posterior condicionales completas de los parámetros de interés y están dadas por

$$\begin{aligned}
 f(\beta|\Sigma_\varepsilon, y_1, \dots, y_n) &\propto f(y_1, \dots, y_n|\beta, \Sigma_\varepsilon) f(\beta) \\
 &\propto \exp \left\{ -\frac{1}{2} \left[ \beta - \left( \Sigma_\beta^{-1} + \sum_{t=1}^n (Y_t \otimes I_p) \Sigma_\varepsilon^{-1} (Y_t^T \otimes I_p) \right)^{-1} \left( \sum_{t=1}^n (Y_t \otimes I_p) \Sigma_\varepsilon^{-1} y_t \right) \right]^T \right. \\
 &\quad \left[ \Sigma_\beta^{-1} + \sum_{t=1}^n (Y_t \otimes I_p) \Sigma_\varepsilon^{-1} (Y_t^T \otimes I_p) \right] \\
 &\quad \left. \left[ \beta - \left( \Sigma_\beta^{-1} + \sum_{t=1}^n (Y_t \otimes I_p) \Sigma_\varepsilon^{-1} (Y_t^T \otimes I_p) \right)^{-1} \left( \sum_{t=1}^n (Y_t \otimes I_p) \Sigma_\varepsilon^{-1} y_t \right) \right] \right\} \quad (2.50)
 \end{aligned}$$

y

$$\begin{aligned}
 f(\Sigma_\varepsilon|\beta, y_1, \dots, y_n) &\propto f(y_1, \dots, y_n|\beta, \Sigma_\varepsilon) f(\Sigma_\varepsilon) \\
 &\propto |\Sigma_\varepsilon|^{-\frac{n_1+p+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left( \left[ \sum_{t=1}^n (y_t - (Y_t^T \otimes I_p) \beta) (y_t - (Y_t^T \otimes I_p) \beta)^T + S^{*-1} \right] \Sigma_\varepsilon^{-1} \right) \right\}, \quad (2.51)
 \end{aligned}$$

donde  $n_1 = n + n^*$ .

## 2.3. Métodos de Selección del Orden VAR

Un elemento crítico en la especificación de los modelos VAR es la determinación del orden. Hamilton [49] indica que las matrices de coeficientes respuesta impulso son construidas a partir de las matrices de coeficientes autorregresivas por lo que la precisión de la estimación de las matrices de coeficientes respuesta impulso dependen de la precisión de la estimación de las matrices de los coeficientes autorregresivos.

Lutkepohl [65] señala que bajo una medida de error cuadrático medio, escoger un orden VAR más alto que el verdadero resulta en pronósticos inferiores que si se escogiera el orden verdadero. En práctica generalmente el orden VAR es desconocido. En esta sección se mencionan algunos métodos para la selección del orden de un modelo VAR.

1. El primer método es conocido como *Prueba de razón verosimilitud secuencial* (para más detalles ver Lutkepohl [65]). La idea de ese método es encontrar el mínimo entero  $k$  tal que  $A_i = 0$  para  $i > k$  y  $A_k \neq 0$ . Este número  $k$  será llamado el orden del modelo VAR. Para lograr esto se supone que se sabe que el orden VAR no puede ser mayor que un entero  $k^*$ , luego se procede a hacer una prueba de hipótesis secuencial de la siguiente manera:

Secuencia	Hipótesis nula ( $H_0$ )	Hipótesis alternativa ( $H_1$ )
1	$A_{k^*} = 0$	$A_{k^*} \neq 0$
2	$A_{k^*-1} = 0$	$A_{k^*-1} \neq 0 \mid A_{k^*} = 0$
$\vdots$	$\vdots$	$\vdots$
$i$	$A_{k^*-i+1} = 0$	$A_{k^*-i+1} \neq 0 \mid A_{k^*} = \dots = A_{k^*-i+1} = 0$
$\vdots$	$\vdots$	$\vdots$
$k^*$	$A_1 = 0$	$A_1 \neq 0 \mid A_{k^*} = \dots = A_2 = 0.$

La secuencia de prueba de hipótesis termina cuando la primera hipótesis nula ( $i$ ) es rechazada,  $\hat{k} = k^* - i + 1$  será elegido como el estimador del orden del modelo VAR.

Para probar la  $i$ -ésima hipótesis se usa el estadístico razón verosimilitud y es dado por

$$M_{LR}(i) = n \left( \ln \left| \hat{\Sigma}_\varepsilon(k^* - i) \right| - \ln \left| \hat{\Sigma}_\varepsilon(k^* - i + 1) \right| \right), \quad (2.52)$$

donde  $\hat{\Sigma}_\varepsilon(j)$  denota el estimador ML de  $\Sigma_\varepsilon$  cuando un modelo VAR ( $j$ ) es ajustado a una serie de tiempo de tamaño  $n$ .  $M_{LR}$  sigue una distribución chi cuadrado con  $p^2$  grados de libertad.

2. Otro método para estimar el orden de un modelo VAR es el error de predicción final (FPE) que es definido de la siguiente manera

$$\text{FPE}(i) = \left[ \frac{n + pi + 1}{n - pi - 1} \right]^p \det(\hat{\Sigma}_\varepsilon(i)), \quad (2.53)$$

donde  $\hat{\Sigma}_\varepsilon(i)$  es el estimador ML de la matriz de covarianza. El estimador del orden VAR es obtenido de la siguiente manera

$$\hat{k}_{\text{FPG}} = \min \{ \text{FPE}(i) \mid i = 0, 1, \dots, k^* \}. \quad (2.54)$$



3. El siguiente método es el criterio de información de Akaike (AIC) definido de la siguiente manera

$$.AIC(i) = \ln |\hat{\Sigma}_\varepsilon(i)| + \frac{2ip^2}{n}. \quad (2.55)$$

El criterio AIC que mejor estima el orden VAR es de la siguiente manera

$$\hat{k}_{AIC} = \text{mín} \{AIC(i) \mid i = 0, 1, \dots, k^*\}. \quad (2.56)$$

4. Un cuarto método para estimar el orden VAR es el criterio Hannan-Quinn (HQ) el cual es obtenido de la siguiente manera

$$HQ(i) = \ln |\hat{\Sigma}_\varepsilon(i)| + \frac{2 \ln(\ln(n))}{n} ip^2. \quad (2.57)$$

El criterio HQ que mejor estima el orden VAR es de la siguiente manera

$$\hat{k}_{HQ} = \text{mín} \{HQ(i) \mid i = 0, 1, \dots, k^*\}. \quad (2.58)$$

5. El siguiente método que se presenta en esta sección para estimar el orden VAR es el criterio bayesiano o Schwarz (BIC) que es de la siguiente manera

$$BIC(i) = \ln |\hat{\Sigma}_\varepsilon(i)| + \frac{\ln(n)}{n} ip^2. \quad (2.59)$$

El criterio BIC que mejor estima el orden VAR es de la siguiente manera

$$\hat{k}_{BIC} = \text{mín} \{BIC(i) \mid i = 0, 1, \dots, k^*\}. \quad (2.60)$$

Para una revisión más detallada de estos métodos de estimación del orden VAR ver Lutkepohl [65], Tsay [100], Brandt y Williams [14].

## 2.4. Verificación del Modelo Adecuado

En la construcción de modelos VAR un elemento de suma importancia es garantizar que el modelo ajustado a los datos sea el adecuado. Existen muchos métodos para comprobar si un modelo VAR ajustado a los datos es el adecuado. En esta sección se presentan algunos de estos métodos.

### 2.4.1. Verificación de los Residuales

Un primer método para verificar que el modelo VAR ajustado a los datos es el adecuado es verificar que los residuos de la serie de un modelo VAR no están correlacionados en serie. Siempre que los residuos no estén correlacionados en serie la estimación de los modelos VAR es bastante robusta.

Para verificar que los residuos no están correlacionados en serie defina  $\hat{\varepsilon}_t = y_t - (\hat{V} + \hat{A}_1 y_{t-1} + \dots + \hat{A}_k y_{t-k})$  el estimador de los residuos donde  $\hat{V}, \hat{A}_1, \dots, \hat{A}_k$  son estimadores ML. Entonces el estimador de la matriz de covarianza de los residuos es de la siguiente manera

$$\hat{C}_i = \frac{1}{n} \sum_{t=i+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-i}^T = \frac{1}{n} \hat{\varepsilon} F_i \hat{\varepsilon}^T \quad i = 0, 1, \dots, h < n, \quad (2.61)$$

donde  $F_i$  es una matriz de dimensión  $n \times n$  tal que los elementos  $(i+l, j)$  son igual a uno para  $l, j = 1, \dots, n-i$ , y cero en el resto de los elementos,  $\hat{\varepsilon} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)$  es de dimensión  $p \times n$ . Sea  $F = (F_1, \dots, F_h)$  de dimensión  $n \times nh$  y defina

$$\hat{C}_h = (C_1, \dots, C_h) = \frac{1}{n} \hat{\varepsilon} F (I_h \otimes \hat{\varepsilon}^T), \quad (2.62)$$

$$\hat{c}_h = \text{vec}(\hat{C}_h). \quad (2.63)$$

Las matrices de autocorrelación estimadas son de la siguiente manera

$$\hat{R}_i = \hat{D}^{-1} \hat{C}_i \hat{D}^{-1} \quad i = 0, 1, \dots, h, \quad (2.64)$$

donde  $\hat{D}$  es una matriz diagonal de dimensión  $p \times p$  cuyos elementos son las raíces cuadradas de

los elementos diagonales de  $\hat{C}_0$ . De igual manera que (2.62)-(2.63) defina

$$\hat{\mathbf{R}}_h = (\hat{R}_1, \dots, \hat{R}_h), \quad (2.65)$$

$$\hat{\mathbf{r}}_h = \text{vec}(\hat{\mathbf{R}}_h). \quad (2.66)$$

Dado que (2.1) se puede escribir como (2.21) entonces

$$\hat{\varepsilon} = Y - \hat{B}Z. \quad (2.67)$$

Se puede demostrar (ver Lutkepohl [65]) que si  $y_t$  es un proceso VAR ( $k$ ) estable de dimensión  $p \times 1$ , y estacionario con  $\varepsilon_t$  siendo un proceso de ruido blanco estándar idénticamente distribuidos y  $B$  es un estimador ML, entonces la distribución asintótica de las autocovarianzas residuales y las autocorrelaciones residuales son de la siguiente manera

$$\sqrt{n}\hat{\mathbf{c}}_h \xrightarrow{d} N(0, \Sigma_{\mathbf{c}}(h)), \quad (2.68)$$

donde  $\Gamma$  es definida en (2.26) y

$$\Sigma_{\mathbf{c}}(h) = ((I_h \otimes \Sigma_{\varepsilon}) - \tilde{G}^T \Gamma \tilde{G}) \otimes \Sigma_{\varepsilon},$$

$$\tilde{G} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \Sigma_{\varepsilon} & \Phi_1 \Sigma_{\varepsilon} & \dots & \Phi_{h-1} \Sigma_{\varepsilon} \\ 0 & \Sigma_{\varepsilon} & \dots & \Phi_{h-2} \Sigma_{\varepsilon} \\ \vdots & & & \vdots \\ 0 & 0 & \dots & \Phi_{h-p} \Sigma_{\varepsilon} \end{bmatrix} \quad \text{es de dimensión } ((pk + 1) \times ph,$$

$$\Phi_i = J A^i J \quad i = 1, 2, \dots, \quad \Phi_0 = I_p,$$

$$J = [I_p : 0 : \dots : 0],$$

$$A = \begin{bmatrix} A_1 & A_2 & \dots & A_{k-1} & A_k \\ I_p & 0 & \dots & 0 & 0 \\ 0 & I_p & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_p & 0 \end{bmatrix} \quad \text{es de dimensión } pk \times pk,$$

$$\sqrt{n}\hat{\boldsymbol{r}}_h \xrightarrow{d} N(0, \Sigma_{\boldsymbol{r}}(h)), \quad (2.69)$$

donde  $\Sigma_{\boldsymbol{r}}(h) = [(I_h \otimes R_\varepsilon) - G_0^T \Gamma^{-1} G_0] \otimes R_\varepsilon$ ,  $G_0 = \tilde{G}(I_h \otimes D^{-1})$  y  $D$  es una matriz diagonal de dimensión  $p \times p$  con las raíces cuadradas de  $\Sigma_\varepsilon$  en la diagonal, además

$$\sqrt{n} \text{vec}(\hat{R}_j) \xrightarrow{d} N(0, \Sigma_R(j)) \quad j = 1, 2, \dots, \quad (2.70)$$

donde

$$\Sigma_R(j) = \left( R_\varepsilon - D^{-1} \Sigma_\varepsilon [0 : \Phi_{j-1}^T : \dots : \Phi_{j-p}^T]^T \Gamma^{-1} \begin{bmatrix} 0 \\ \Phi_{j-1} \\ \vdots \\ \Phi_{j-p} \end{bmatrix} \Sigma_\varepsilon D^{-1} \right) \otimes R_\varepsilon, \quad (2.71)$$

con  $\Phi_i = 0$  para  $i < 0$  y  $R_\varepsilon = D^{-1} \Sigma_\varepsilon D^{-1}$ .

Los parámetros  $\Gamma$ ,  $\Sigma_\varepsilon$  y  $B$  generalmente son desconocidos por lo que se sustituyen por estimadores para obtener errores estándar de las autocorrelaciones residuales y pruebas de hipótesis específicas con respecto a las autocorrelaciones.

Denotando por  $\rho_{ij}(\ell)$  como los verdaderos coeficientes de correlación correspondientes a los  $r_{ij,\ell}(\ell)$ .

Una prueba con un nivel aproximado de 5%, de la hipótesis nula es dada por

$$H_0 : \rho_{ij}(\ell) = 0 \quad \text{contra} \quad H_1 : \rho_{ij}(\ell) \neq 0, \quad (2.72)$$

donde  $H_0$  es rechazado si  $\frac{-2}{\sqrt{n}} < r_{ij,\ell} < \frac{2}{\sqrt{n}}$ , en otras palabras, la hipótesis de ruido blanco es rechazada si alguno de los coeficientes de correlación estimados sale del intervalo.

### 2.4.2. Prueba de Portmanteau

Un segundo método para verificar que el modelo VAR ajustado es el adecuado consiste en considerar los coeficientes de correlación como grupo.

Este método es conocido como prueba de Pormonteau

$$H_0 : \boldsymbol{R}_h = 0 \quad \text{contra} \quad H_1 : \boldsymbol{R}_h \neq 0. \quad (2.73)$$

El estadístico de prueba es de la siguiente manera

$$Q_h = n \sum_{i=1}^h \text{tr} \left( \hat{C}_i^T \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1} \right), \quad (2.74)$$

donde  $\hat{C}_i$  se define en (2.61). Se puede demostrar (ver Ahn [2], Li y McLeod [62], Hosking [52] y Lutkepohl [65]) que bajo ciertas condiciones de regularidad como las de (2.68) la distribución aproximada del estadístico de Pormonteau es chi-cuadrado con  $p^2(h - k)$  grados de libertad.

### 2.4.3. Prueba de Normalidad

Un tercer método para verificar que el modelo VAR ajustado a los datos es el adecuado consiste en notar que  $\varepsilon_t = y_t - (A_1 y_{t-1} + \dots + A_k y_{t-k})$  se distribuye normal con vector media 0 y matriz de covarianza  $\Sigma_\varepsilon$ . Debido a que en la práctica no se cuenta con  $\varepsilon_t$  entonces es reemplazado por su estimador  $\hat{\varepsilon}_k$ . Por lo que la prueba de normalidad se aplica a estos últimos. La prueba es basada en el tercer y cuarto momento. El estadístico de prueba para el tercer y cuarto momento y sus distribuciones asintóticas son de la siguiente manera

$$\begin{aligned} \hat{\lambda}_3 &= \frac{n}{6} \hat{a}_1^T \hat{a}_1 \xrightarrow{d} \chi^2(p), \\ \hat{\lambda}_4 &= \frac{n}{24} (\hat{a}_2 - 3p)^T (\hat{a}_2 - 3p) \xrightarrow{d} \chi^2(p), \\ \hat{\lambda}_{34} &= \hat{\lambda}_3 + \hat{\lambda}_4 \xrightarrow{d} \chi^2(2p), \end{aligned}$$

donde

$$\begin{aligned} \hat{a}_1 &= (\hat{a}_{11}, \dots, \hat{a}_{p1})^T, & \hat{a}_{i1} &= \frac{1}{n} \sum_{t=1}^n \hat{b}_{it}^3, \quad i = 1, \dots, p, \\ \hat{a}_2 &= (\hat{a}_{12}, \dots, \hat{a}_{p2})^T, & \hat{a}_{i2} &= \frac{1}{n} \sum_{t=1}^n \hat{b}_{it}^4, \quad i = 1, \dots, p, \\ b_t &= (\hat{b}_{1t}, \dots, \hat{b}_{pt})^T = \hat{\Sigma}_\varepsilon^{-1/2} \hat{\varepsilon}_t, \\ \hat{\Sigma}_\varepsilon^{-1/2} \left( \hat{\Sigma}_\varepsilon^{1/2} \right)^T &= \hat{\Sigma}_\varepsilon, \\ \hat{\Sigma}_\varepsilon &= \frac{1}{n - pk - 1} \sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_t^T \end{aligned}$$

$$\hat{\varepsilon}_t = (y_t - \bar{y}) - \left( \hat{A}_1(y_{t-1} - \bar{y}) + \dots + \hat{A}_k(y_{t-k} - \bar{y}) \right), \quad t = 1, \dots, n,$$

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t.$$

$\hat{\lambda}_3$  puede ser usado para probar

$$H_0 : E \begin{bmatrix} b_{1t}^3 \\ \vdots \\ b_{pt}^3 \end{bmatrix} = 0 \quad \text{contra} \quad H_1 : E \begin{bmatrix} b_{1t}^3 \\ \vdots \\ b_{pt}^3 \end{bmatrix} \neq 0 \quad (2.75)$$

y  $\hat{\lambda}_4$  puede ser usado para probar

$$H_0 : E \begin{bmatrix} b_{1t}^4 \\ \vdots \\ b_{pt}^4 \end{bmatrix} = 3p \quad \text{contra} \quad H_1 : E \begin{bmatrix} b_{1t}^4 \\ \vdots \\ b_{pt}^4 \end{bmatrix} \neq 3p. \quad (2.76)$$

$\hat{\lambda}_{34}$  puede ser usado para una prueba conjunta de las hipótesis nulas en (2.75)-(2.76). Para más detalles de la prueba de normalidad ver Lutkepohl [65].

# Capítulo 3

## Modelos de Volatilidad Estocástica Multivariado

Los modelos de volatilidad estocástica (SV) han sido ampliamente discutidos en la literatura y muestran ser útiles para modelar la varianza, analizando su cambio en el tiempo de retornos de activos en econometría financiera. Los modelos SV postulan que la volatilidad es conducida por su propio proceso estocástico a diferencia de los modelos autorregresivo generalizado condicional heterocedástico (GARCH) por lo que son más flexibles para representar las propiedades empíricas características de los retornos financieros.

Melino y Turnbull [67] demuestran que las suposiciones de distribución de probabilidad lognormal para los tipos de cambio y volatilidad constante en el modelo estándar para fijar los precios en moneda extranjera, son inapropiadas y que una hipótesis más plausible es que la volatilidad cambie estocásticamente, por lo que el modelo de volatilidad estocástica hace un trabajo superior al igualar la distribución del tipo de cambio empírico y corrige muchos de los sesgos observados en los precios de opciones pronosticados.

La motivación económica de los modelos SV se basa en la así llamada hipótesis de mezcla de distribuciones, que establece que los retornos financieros son impulsados por una convolución de dos variables aleatorias, siendo una un término de ruido independiente, y el otro un proceso estocástico que representa un proceso de llegada de información. Una característica común de los modelos

motivados por la hipótesis de mezcla de distribuciones es que condicionado a la variable latente  $\sigma_t$ , los retornos siguen una distribución normal:  $z_t|\sigma_t \sim N(0, \sigma_t^2)$ .

Sin embargo, como se supone que  $\sigma_t$  es una variable aleatoria, la distribución incondicional de  $z_t$  ya no es gaussiana. En particular, los modelos SV tienen colas más pesadas que la distribución normal, lo que corresponde a la evidencia empírica de los retornos financieros. El modelo SV autoregresivo gaussiano estándar en tiempo discreto fue introducido por primera vez en esta forma por Taylor [94], y es dado por

$$y_t - \mu_t = z_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1), \quad (3.1)$$

$$\log \sigma_{t+1}^2 = w + \phi \log \sigma_t^2 + \sigma_\eta \eta_t, \quad \eta_t \sim N(0, 1). \quad (3.2)$$

Note que la ecuación (3.2) es un proceso autorregresivo gaussiano de primer orden de la log volatilidad de los retornos. En el modelo de Taylor [94] las innovaciones  $\varepsilon_t$  y  $\eta_t$  son independientes.

La principal diferencia con los modelos GARCH es que, condicional al conjunto de información la verosimilitud no se conoce, sino más bien una variable aleatoria no observada, por lo que implementar estimadores máximo verosimilitud es muy difícil. Para ver esto, la función verosimilitud del modelo SV para una muestra de  $n$  observaciones puede ser escrita como

$$L(\theta; Y_n) \propto \int f(Y_n|H_n; \theta) f(H_n|\theta) dH_n, \quad (3.3)$$

donde  $Y_n$  es un vector conteniendo todos los retornos observables,  $H_n = (\sigma_1^2, \dots, \sigma_n^2)^T$  es el vector conteniendo todas las volatilidades latentes, y  $\theta$  es el vector de parámetros, el cual en el modelo clásico es  $\theta = (w, \phi, \sigma_\eta)^T$ . El problema de la función verosimilitud es la integral que aparece en la ecuación (3.3), la cual es una integral múltiple de dimensión  $n$ . No se puede resolver analíticamente, y los métodos numéricos discretos son inviables incluso para muestras moderadamente grandes. La dificultad para estimar el modelo SV radica en que la ecuación de observación (3.1) depende de forma no lineal de los estados  $\sigma_t$ .

Existen muchas formas diferentes de realizar la estimación de los parámetros del modelo SV. La tabla 3.1 es tomada de Bauwens et al. [8] donde resumen algunos métodos para la estimación de modelos SV.



<b>Método (acrónimo)</b>	<b>Referencia central</b>
Verosimilitud Cuasi-máxima (QML)	Harvey et al. [50]
Muestreo de mezcla gaussiana (GMS)	Kim et al. [59]
Método simulado de momentos (SMM)	Gallant y Taucher [46]
Muestreo de importancia (IS)	Durbin y Koopman [36]
Muestreo de importancia Eficiente (EIS)	Richard y Zhang [80]
Muestreo de importancia mejorado (IIS)	Nguyen [73]
Muestreo de sitio único (SSS)	Carter y Kohn [18]
Muestreo de multimovimiento (MMS)	Shepard y Pitt [85]

Tabla 3.1: Métodos de estimación de modelos SV.

Estos métodos para estimar modelos SV surgen como respuesta para superar el problema de no poder emplear soluciones analíticas. Una descripción detallada de estos métodos se puede encontrar en Bauwens et al. [8].

La variedad de modelos de volatilidad estocástica multivariante (MSV) es notable, desde un modelo rígido con volatilidades independientes y correlaciones constantes hasta modelos altamente complejos que incorporan correlaciones dinámicas y efectos de apalancamiento. Una revisión de modelos MSV propuestos se pueden encontrar en Asai et al. [6], Yu y Meyer [104].

El modelo básico de Harvey [50] se puede escribir como

$$\varepsilon_t = H_t^{1/2} \varepsilon_t, \quad (3.4)$$

$$H_t = \text{diag}(\exp(h_{1t}), \dots, \exp(h_{pt})), \quad (3.5)$$

$$h_{t+1} = \mu + \phi \odot h_t + \eta_t, \quad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\eta \end{pmatrix} \right], \quad (3.6)$$

donde  $h_t = (h_{1t}, \dots, h_{pt})^T$  es el vector de volatilidad  $\mu$  y  $\phi$  son vectores de parámetros  $p \times 1$ ,  $\Sigma_\varepsilon$  es una matriz de correlación,  $\Sigma_\eta$  es una matriz de covarianza definida positiva, y el operador  $\odot$  denota el producto Hadamard.

El resto de este capítulo está organizado de la siguiente manera. En la Sección 3.1 se presentan algunos modelos GARCH y sus propiedades. En la Sección 3.2 se describen propiedades de los modelos SV de tiempo discreto. En la sección 3.3 se presentan algunos modelos MSV de tiempo discreto. En la sección 3.4 se discuten algunos métodos de estimación como el muestreo de múltiples movimientos (multimove sampler). En la sección 3.5 se examina el criterio de información de

desviación (DIC) para comprobación de modelo.

### 3.1. Modelos GARCH

Una de las principales características de los retornos de activos son los grupos de volatilidad (clusters), a saber, los grandes choques tienden a ser seguidos por grandes choques, de cualquier signo, de igual manera los pequeños choques tienden a ser seguidos por pequeños choques. En otras palabras, la volatilidad medida por los retornos al cuadrado es persistente y, por lo tanto hasta cierto punto predecible. El modelo ARCH propuesto por Engle [38] es el primer modelo que provee un marco sistemático para modelar la correlación en serie de la volatilidad. El modelo ARCH(k) es dado por

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_k a_{t-k}^2, \\ a_t &= \sigma_t \varepsilon_t,\end{aligned}\tag{3.7}$$

donde  $\varepsilon_t$  son idénticamente distribuidos e independientes con media cero y varianza 1,  $\alpha_0 > 0$ , y  $\alpha_i \geq 0$  para  $i > 0$ .

Los coeficientes  $\alpha_i$  deben seguir alguna condición de regularidad para garantizar que la varianza incondicional de  $a_t$  sea finita. Por la estructura del modelo se puede ver que grandes choques cuadrados pasados  $a_{t-1}, \dots, a_{t-k}$  implican una gran varianza condicional  $\sigma_t^2$  para la innovación  $a_t$  por lo que tiende a asumir valores grandes. Esto significa que bajo el marco ARCH, grandes choques tienden a ser seguidos por grandes choques. Esta propiedad de los modelos ARCH cumple la característica de los clusters de volatilidad de los retornos de activos. A menudo el modelo ARCH requiere de muchos coeficientes para describir adecuadamente el proceso de volatilidad de un retorno de activo. Bollerslev [12] como una forma de modelar los movimientos persistentes en la volatilidad sin estimar un gran número de coeficientes en un polinomio de alto orden sugirió los modelos GARCH. El modelo GARCH( $k_1, k_2$ ) es dado por

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{k_1} \alpha_i a_{t-i}^2 + \sum_{j=1}^{k_2} \beta_j \sigma_{t-j}^2, \quad (3.8)$$

$$a_t = y_t - \mu_t = \sigma_t \varepsilon_t,$$

donde  $y_t$  es una serie de retornos logarítmicos y  $\mu_t$  su respectiva media condicional. De nuevo, los  $\varepsilon_t$  son idénticamente distribuidos e independientes con media cero y varianza 1,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0, \beta_0 \geq 0$ , y  $\sum_{i=1}^{\max(k_1, k_2)} (\alpha_i + \beta_i) < 1$ . Además,  $\alpha_i = 0$  para  $i > k_1$  y  $\beta_j = 0$  para  $j > k_2$ . El modelo más utilizado en la clase GARCH es el simple GARCH(1,1) dado por

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \quad (3.9)$$

El coeficiente  $\alpha_1$  mide la medida en que un choque de volatilidad actual se refleja en la volatilidad del próximo período, mientras que  $(\alpha_1 + \beta_1)$  mide la velocidad a la que este efecto desaparece con el tiempo. Para series de retornos financieros, estimaciones de  $\alpha_1$  y  $\beta_1$  están muy a menudo en los rangos  $[0.02, 0.25]$  y  $[0.75, 0.98]$  respectivamente. En los modelos GARCH(1,1) es fácil construir pronósticos de múltiples períodos adelante de la volatilidad. Entonces, para el modelo (3.9) el pronóstico un paso adelante es dado por

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 a_t^2 + \beta_1 \sigma_t^2,$$

donde  $a_t^2$  y  $\sigma_t^2$  son conocidas hasta el tiempo  $t$ . Para el pronóstico de múltiples pasos adelante se configura  $a_t^2 = \sigma_t^2 \varepsilon_t^2$  y se reescribe la ecuación (3.9) de la siguiente manera

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2 + \alpha_1 \sigma_t^2 (\varepsilon_t^2 - 1). \quad (3.10)$$

Calculando la esperanza condicional al conjunto de información en ambos lados de la ecuación se obtiene

$$\sigma_t^2(h) = \alpha_0 + (\alpha_1 + \beta_1) \sigma_t^2 (h - 1), h > 1. \quad (3.11)$$

Si en el modelo GACRH(1,1) de la ecuación (3.9) se establece la restricción ( $\alpha_1 + \beta_1 = 1$ ) el modelo tiene una raíz autorregresiva unitaria por lo que la volatilidad de hoy afecta los pronósticos de volatilidad en el futuro indefinido. Los modelos con esta restricción se conocen como GARCH(1,1) integrados o IGARCH(1,1). Cuando  $\sum_{i=1}^{k_1} \alpha_i + \sum_{j=1}^{k_2} \beta_j = 1$  el modelo es llamado IGARCH( $k_1, k_2$ ) (ver Engle y Bollerslev [39]). El modelo IGARCH(1,1) puede ser escrito como

$$\sigma_t^2 = \alpha_0 + \beta_1 a_{t-1}^2 + (1 - \beta_1) \sigma_{t-1}^2, 0 < \beta_1 < 1. \quad (3.12)$$

Una característica clave de los modelos IGARCH es que el impacto de los choques al cuadrado pasados es persistente. Para ver esto primero se define la ecuación de pronóstico múltiples pasos adelante de la siguiente manera

$$\sigma_t^2(h) = \sigma_t^2(1) + (h - 1)\alpha_0, h \geq 1. \quad (3.13)$$

De la ecuación (3.13) se puede ver que el efecto de  $\sigma_t^2(1)$  en volatilidades futuras es persistente, y los pronósticos de volatilidad forman una línea recta con pendiente  $\alpha_0$ .

Glosten et al. [47] propuso modelar las respuestas asimétricas en volatilidad condicional utilizando umbrales y encontraron apoyo para una relación negativa entre el retorno mensual esperado condicional y la varianza condicional del retorno mensual, utilizando un modelo GARCH-M modificado al permitir

1. Patrones estacionales en la volatilidad
2. Innovaciones positivas y negativas en los retornos que tienen diferentes impactos en la volatilidad condicional, y
3. Tasas de interés nominales para predecir la varianza condicional.

El modelo GARCH-M modificado general que propusieron es:

**Ecuación para la media condicional:**

$$E[X_{t+1}|G_t] = \mu(G_t),$$

donde  $\mu(\cdot)$  es una función que describe la naturaleza de la dependencia de la media condicional en los elementos del conjunto información  $G_t$ .

$$X_{t+1} = \mu(G_t) + \varepsilon_{t+1} \quad \text{con} \quad E[\varepsilon_{t+1}|G_t] = 0.$$

**Ecuación para la varianza condicional:**

$$\text{Var}(X_{t+1}|G - t) = \text{Var}(\varepsilon_{t+1}|G_t) = V(G_t),$$

donde  $V(\cdot)$  es una función que describe la naturaleza de la dependencia de la varianza condicional en los elementos del conjunto información  $G_t$ .

La literatura de los modelos GARCH es enorme, algunas variaciones son los GARCH exponencial (EGARCH) propuesto por Nelson [71], GARCH de umbral (TGARCH) (ver, Glosten et al. [47] y Zakoian [105]), GARCH extendido (GJR-GARCH) (ver Glosten [47]), poder asimétrico GARCH (APGARCH) propuesto por Ding et al. [31], Red neuronal artificial GARCH (ANN-GARCH) (ver Donaldson y Kamstra [33]), GARCH que cambia de volatilidad (VS-GARCH) de Fornari y Mele [41].

Muchas de las ideas que se consideraron en el contexto de los modelos GARCH univariados fueron generalizados al caso multivariado. El modelo GARCH(1,1) multivariado es dado por

$$\text{vech}(\Sigma_t) = w + \alpha \text{vech}(\Sigma_{t-1}) + \beta \text{vech}(\eta_t \eta_t^T), \quad (3.14)$$

donde  $w$  es un vector con dimensión  $\frac{p(p+1)}{2}$ ,  $\alpha$  y  $\beta$  son matrices de dimensión  $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ ,  $\Sigma_t$  es una matriz de covarianza definida positiva. Bollerslev, Engle y Wooldridge [13] generalizaron el enfoque de promedio móvil ponderado exponencialmente de la siguiente manera

$$\Sigma_t = A_0 + \sum_{i=1}^{k_1} A_i \odot (a_{t-i} a_{t-i}^T) + \sum_{j=1}^{k_2} B_j \odot \Sigma_{t-j}, \quad (3.15)$$

donde  $A_i$  y  $B_j$  son matrices simétricas y,  $\odot$  denota el producto Hadamard. El modelo (3.15) es conocido como vectorización diagonal (VEC( $k_1, k_2$ ) o DVEC( $k_1, k_2$ )).

El modelo Baba-Engle-Kraft-Kroner (BEKK) propuesto por Engle y Kroner [40] garantiza que  $\Sigma_t$  sea definida positiva. El modelo BEKK es dado por

$$\Sigma_t = CC^T + \sum_{i=1}^{k_1} A_i(a_{t-i}a_{t-i}^T)A_i^T + \sum_{j=1}^{k_2} B_j\Sigma_{t-j}B_j^T, \quad (3.16)$$

donde  $C$  es una matriz triangular inferior con  $\frac{p(p+1)}{2}$  parámetros,  $A_i$  y  $B_i$  son matrices de dimensión  $p \times p$  que garantizan que  $\Sigma_t$  sea definida positiva.

El modelo BEKK es un caso especial del modelo VEC. Algunos modelos GARCH multivariados son los de correlación condicional dinámica (DCC) en el que la matriz de covarianza se descompone de la siguiente manera

$$\Sigma_t = \Delta_t R_t \Delta_t, \quad (3.17)$$

donde  $\Delta_t = \text{diag}(\sqrt{\sigma_{11}}, \dots, \sqrt{\sigma_{pp}})$  y  $R_t$  es la matriz de correlación variando en el tiempo. Engle et al. [81] sugirieron una parametrización de la matriz de covarianza condicional usando la idea de que los comovimientos de los retornos de activos están impulsados por un pequeño número de variables subyacentes comunes llamados factores.

Engle et al. [81] suponen que la matriz de varianza condicional  $\Sigma_t$  es generada por  $K$ ,  $K < p$  volatilidades de los factores  $f_{kk,t}$  correspondientes a  $K$  factores subyacentes que no necesariamente son no correlacionados. El modelo es dado por

$$\Sigma_t = C + \sum_{i=1}^K w_i w_i^T f_{ii,t} = C + W F_t W^T, \quad (3.18)$$

donde  $C$  es una matriz semidefinida positiva de dimensión  $p \times p$ ,  $W$  es una matriz de pesos de dimensión  $p \times K$  con columnas  $w_1, \dots, w_K$  de dimensión  $k \times 1$  linealmente independientes, y  $F_t = \text{diag}(f_{11,t}, \dots, f_{KK,t})$  es una matriz diagonal de dimensión  $K \times K$ . Vrontos et al. [102] introdujo un modelo de factores completos dado por

$$y_t = \mu + \varepsilon_t, \quad (3.19)$$

$$\varepsilon_t = W X_t, X_t | \Phi_{t-1} \sim N_p(0, \Sigma_t), \quad (3.20)$$

donde  $\mu$  es un vector de constantes de dimensión  $p \times 1$ ,  $W$  es una matriz parámetro de dimensión  $p \times p$ ,  $\Phi_{t-1}$  es el conjunto información hasta el tiempo  $t-1$ ,  $X_t$  es un vector de factores de dimensión  $p \times 1$ .  $\Sigma_t = \text{diag}(\sigma_{1,t}^2, \dots, \sigma_{p,t}^2)$ ,  $\varepsilon_t$  es un vector término error de dimensión  $p \times 1$  y  $\varepsilon_t | \Phi_{t-1} \sim N(0, H_t)$  donde  $H_t = W \Sigma_t W^t$ .

Al igual que los modelos GARCH univariados en el caso multivariado la literatura es grande. Una revisión más detallada de los modelos GARCH multivariados se puede encontrar en Bauwens et al. [9].

### 3.2. Modelos SV

Los modelos de SV describen la evolución de la volatilidad de los retornos financieros introduciendo un término error en la varianza condicional. Al agregar el término error incrementa la flexibilidad del modelo en describir la evolución de la varianza condicional pero también incrementa la dificultad en estimar los parámetros debido a que la función verosimilitud exacta es difícil de evaluar y el estimador de ML de los parámetros no es directa como en los modelos GARCH. Broto y Ruiz [15] comentan: *La popularidad de los modelos SV también se debe porque se acercan a los modelos que se utilizan a menudo en la teoría financiera para representar el comportamiento de los precios financieros. Sus propiedades estadísticas son fáciles de derivar utilizando resultados bien conocidos en distribuciones logarítmicas normales. Además, en comparación con los modelos GARCH más populares, capturan de manera adecuada las principales propiedades empíricas que se observan a menudo en las series diarias de retornos financieros.* Un modelo SV es dado por

$$a_t = \sigma_t \varepsilon_t, \tag{3.21}$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \sigma_{t-1}^2 + \dots + \alpha_k \ln \sigma_{t-k}^2 + \eta_t, \tag{3.22}$$

donde los  $\varepsilon_t$  son distribuidos normal idénticamente e independientes con media igual a 0 y varianza igual a 1, los  $\eta_t$  son distribuidos normal idénticamente e independientes con media igual a 0 y varianza igual a  $\sigma_\eta^2$ . Además,  $\varepsilon_t$  y  $\eta_t$  son independientes,  $\alpha_0$  es constante y  $\sigma_\eta^2$  mide la incertidumbre de la volatilidad futura.

En el modelo autorregresivo gaussiano en tiempo discreto de Taylor [94] de la ecuación (3.1)-(3.2) el parámetro  $\phi$  a menudo se considera como una medida de la persistencia de los choques en la volatilidad. Cuando  $\phi$  es cercano a 1 y  $\sigma_\eta$  es cercano a cero, la evolución de la volatilidad sobre el tiempo es muy fluida. Si  $\phi = 1$  y  $\sigma_\eta = 0$  la volatilidad es constante sobre el tiempo y consecuentemente los retornos son homocedásticos. La varianza del retorno observado  $z_t$  condicional al conjunto de información hasta el tiempo  $t - 1$  no es en general igual a  $\sigma_t$  ya que este último no depende de las observaciones pasadas sino de algunos componentes no observados. Siguiendo el trabajo de Clarck [22] y, Tauchen y Pitts [93] una interpretación de la variable latente  $\sigma_t$  es representar el flujo aleatorio y desigual de nueva información, que es muy difícil de modelar directamente, en los mercados financieros. Como  $\varepsilon_t$  son idénticamente distribuidos e independientes  $z_t$  es una diferencia de martingalas satisfaciendo la hipótesis de mercado eficiente.

Una definición de la hipótesis de mercado eficiente tomada de Campbell et al. [16] dice: *El mercado es eficiente con respecto a algún conjunto de información si los precios no se verían afectados por revelar esa información a todos los participantes. Además, la eficiencia con respecto a un conjunto de información implica que es imposible obtener beneficios económicos negociando sobre la base de ese conjunto de información.* Note que incluso cuando  $\varepsilon_t$  es asumido a ser gaussiano la distribución de  $z_t$  condicional a las observaciones pasadas hasta el tiempo  $t - 1$  no es gaussiano. Liesenfeld y Richard [63] muestran que el modelo SV con errores t-student puede interpretarse como un modelo SV con dos procesos de volatilidad independientes. Los modelos SV se pueden linealizar tomando los logaritmos de los retornos al cuadrado, por ejemplo el modelo de Taylor [94] de la ecuación (3.1)-(3.2) es dado por

$$\log z_t^2 = \log \sigma_t^2 + \log \varepsilon_t^2, \quad (3.23)$$

$$\log \sigma_{t+1}^2 = w + \phi \log \sigma_t + \eta_t. \quad (3.24)$$

El modelo de la ecuación (3.23)-(3.24) es un modelo de espacio de estados lineal no gaussiano donde (3.23) es la ecuación de medida y (3.24) es la ecuación de transición. Como  $\varepsilon_t$  es asumido a tener distribución gaussiana,  $\log \varepsilon_t^2$  tiene distribución  $\log \chi_{(1)}^2$  con varianza igual a  $\frac{\pi^2}{2}$ . Otra propiedad del modelo de Taylor [94] es que se puede calcular la función de autocorrelación de los retornos al cuadrado  $z_t^2$  y la curtosis de  $\varepsilon_t$  de la siguiente manera



$$\rho(\tau) = \frac{\exp\left(\frac{\sigma_\eta^2}{(1-\phi^2)\phi^\tau}\right) - 1}{\kappa - 1}, \quad (3.25)$$

$$\kappa = 3 \exp\left(\frac{\sigma_\eta}{(1-\phi^2)}\right). \quad (3.26)$$

Note que la curtosis  $\kappa$  es estrictamente mas grande que 3 por lo que los retornos tienen una distribución leptocúrtica o de cola pesada que es una característica de los retornos financieros.

Para más detalles de los modelos SV y sus propiedades ver Broto y Ruiz [15], Cox et al. [27], Taylor [95] y Shephard [83].

### 3.3. Modelos MSV

El primer modelo MSV propuesto en la literatura se debe a Harvey et al. [50] descrito en la ecuación (3.4)-(3.6). Harvey también consideró la distribución  $t$  multivariante para  $\varepsilon_t$ , ya que está especificación permite una mayor curtosis que con el supuesto de Gaussiana. Se puede observar que la definición de definida positiva y la parsimonia se logran restringiendo  $H_t$  para que sea una matriz diagonal, con el elemento diagonal siguiendo la función exponencial de un proceso de vector gaussiano AR(1).

El efecto asimétrico en la volatilidad es que los efectos de los retornos positivos sobre la volatilidad son diferentes de los retornos negativos de magnitud similar. Por otro lado, el apalancamiento se refiere a la correlación negativa entre el retorno actual y la volatilidad futura. Por lo tanto, el apalancamiento denota asimetría. Pero no todos los efectos asimétricos muestran apalancamiento. Una explicación popular de asimetría es el efecto de apalancamiento propuesto por Black [11] y Christie [21].

La mayoría de los modelos MSV asimétricos se basan en las especificaciones básicas de los modelos SV y, por lo tanto, se garantiza que  $H_t$  sea definida positiva. Dependiendo de cómo se introduzca la asimetría, se tienen tres tipos de modelos MSV asimétricos, como se muestra a continuación.

### 3.3.1. Efecto de apalancamiento

Harvey y Shephard [51] propusieron por primera vez un modelo SV de tiempo discreto univariado con efectos de apalancamiento. El modelo de Harvey y Shephard [51] es de la forma

$$\begin{aligned}
 y_t &= \sigma_t \varepsilon_t, \\
 \ln \sigma_{t+1}^2 &= \mu + \phi \ln \sigma_t^2 + \sigma_\eta \eta_t, \\
 \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right],
 \end{aligned} \tag{3.27}$$

donde todas las variables son escalares.

En el contexto de SV, Yu [103] definió el efecto de apalancamiento como una relación negativa entre  $E(\ln \sigma_{t+1}^2 | y_t)$  y los retornos en el periodo  $t$  (definido por  $y_t$ ), manteniendo todo lo demás constante. Yu [103] mostró que

$$E(\ln \sigma_{t+1}^2 | y_t) = \mu + \frac{\mu\phi}{1-\phi^2} + \rho\sigma_\eta \exp\left(-\frac{\sigma_\eta^4}{4(1-\phi^2)^2} + \frac{\sigma_\eta^2\mu}{(1-\phi^2)(1-\phi)}\right) y_t$$

y por lo tanto, el modelo (3.27) asegura el efecto de apalancamiento siempre que  $\rho < 0$ .

Jacquier [55] propone dejar  $(\varepsilon_t, \eta_{t-1})$  siguiendo una distribución normal bivariada con correlación  $\rho$ . Ellos proponen métodos de estimación e inferencia para este modelo en un marco bayesiano. Una crítica de este modelo es el de martingalas en el sentido de que  $E(y_{t-1} | y_t, \sigma_{t-1}) \neq 0$ , violando la hipótesis del mercado eficiente, ver (Harvey and Shephard [51]).

Asai y McAleer [5] consideraron una extensión multivariada del modelo de Harvey y Shephard [51].

El modelo es dado por

$$Y_t = H_t^{1/2} \varepsilon_t, \quad (3.28)$$

$$H_t^{1/2} = \text{diag} (\exp(h_{1t}/2), \dots, \exp(h_{pt}/2)) = \text{diag} (h_t/2), \quad (3.29)$$

$$h_{t+1} = \mu + \phi \odot h_t + \eta_t, \quad (3.30)$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\varepsilon & L \\ L & \Sigma_\eta \end{pmatrix} \right], \quad (3.31)$$

$$L = \text{diag} (\lambda_1 \sigma_{\eta,11}, \dots, \lambda_m \sigma_{\eta,pp}),$$

donde el parámetro  $\lambda_i$ ,  $i = 1, \dots, p$ , se espera que sea negativa.

Un modelo MSV más completo propuesto por Ishihura y Omori [53] que es el que integraremos a un modelo VAR ( $k$ ) en el capítulo 4 es presentado de la siguiente manera.

Sea  $y_t = (y_{1t}, \dots, y_{pt})^T$  denotando un conjunto de observaciones en el tiempo  $t$  sobre  $p$  retornos activos, y sea  $\alpha_t = (\alpha_{1t}, \dots, \alpha_{pt})^T = h_t - \mu$ , donde  $h_t = (h_{1t}, \dots, h_{pt})^T$  es el correspondiente vector de log verosimilitud y  $\mu = (\mu_1, \dots, \mu_p)^T$  es su vector media. Entonces el modelo MSV es dado por

$$Y_t = V_t^{1/2} \varepsilon_t, \quad t = 1, \dots, n, \quad (3.32)$$

$$\alpha_{t-1} = \Phi \alpha_t + \eta_t, \quad t = 1, \dots, n-1, \quad (3.33)$$

$$\alpha_1 \sim N_p(0, \Sigma_0), \quad (3.34)$$

donde

$$V_t^{1/2} = \text{diag} (\exp(\alpha_{1t}/2), \dots, \exp(\alpha_{mt}/2)),$$

$$\Phi = \text{diag} (\phi_1, \dots, \phi_m),$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\eta} \\ \Sigma_{\eta\varepsilon} & \Sigma_{\eta\eta} \end{pmatrix}, \quad (3.35)$$

además

$\rho_{ij,\varepsilon\varepsilon} \sigma_{i,\varepsilon\varepsilon} \sigma_{j,\varepsilon\varepsilon}$ , es el elemento  $ij$ -ésimo de  $\Sigma_{\varepsilon\varepsilon}$ ,

$\rho_{ij,\varepsilon\eta} \sigma_{i,\varepsilon\varepsilon} \sigma_{j,\eta\eta}$ , es el elemento  $ij$ -ésimo de  $\Sigma_{\varepsilon\eta}$ ,

$\rho_{ij,\eta\eta} \sigma_{i,\eta\eta} \sigma_{j,\eta\eta}$ , es el elemento  $ij$ -ésimo de  $\Sigma_{\eta\eta}$ ,

con coeficiente de correlación

$$\rho_{ij,\varepsilon\varepsilon} = \text{CORR}(\varepsilon_{it}, \varepsilon_{jt}),$$

$$\rho_{ij,\varepsilon\eta} = \text{CORR}(\varepsilon_{it}, \eta_{jt}),$$

$$\rho_{ij,\eta\eta} = \text{CORR}(\eta_{it}, \eta_{jt}),$$

y desviación estándar

$$\sigma_{i,\varepsilon\varepsilon} = \sqrt{\text{Var}(\varepsilon_{it})},$$

$$\sigma_{i,\eta\eta} = \sqrt{\text{Var}(\eta_{it})}.$$

Para hacer cumplir la estacionariedad de  $\alpha_t$  suponemos que  $|\phi_i| < 1$  y el  $(i, j)$ -ésimo elemento de  $\Sigma_0$  es el  $(i, j)$ -ésimo elemento de  $\Sigma_{\eta\eta}$  dividido por  $1 - \phi_i \phi_j$  ( $i, j = 1, \dots, p$ ) o equivalentemente

$$\text{vec}(\Sigma_0) = (I_{p^2} - \Phi \otimes \Phi)^{-1} \text{vec}(\Sigma_{\eta\eta}) \quad (3.36)$$

tal que  $\Sigma_0 = \Phi \Sigma_0 \Phi + \Sigma_{\eta\eta}$  donde  $\otimes$  y  $\text{vec}(A)$  denota el producto Kronecker y una vectorización de la matriz  $A$ .

### 3.3.2. Efecto de umbral

So et al. [91] propusieron un modelo umbral SV. La no linealidad del umbral se incorpora en las especificaciones de la media y varianza del modelo. Ellos argumentan que dado que se cree que la varianza tiende a ser mayor bajo la influencia de malas noticias que bajo la influencia de buenas noticias, es probable que la influencia autorregresiva en (3.2) este determinado por el signo de los

rendimientos anteriores. Y por tanto, el tipo de umbral de modelos (Tong y Lim [98]; Tong, [96,97]) parece ser muy apropiado para describir la asimetría en el mecanismo generador de varianza. Además, Li y Lam [61] argumentaron que la magnitud de los retornos depende del signo de los retornos. El modelo es como sigue:

Defina un conjunto de variables aleatorias Bernoulli  $s_t$  por

$$s_t = \begin{cases} 0 & \text{si } r_{t-1} < 0, \\ 1 & \text{si } r_{t-1} \geq 0, \end{cases}$$

entonces el modelo SV de umbral (THSV) es dado de la siguiente manera

$$r_t = \Psi_{0s_t} + \Psi_{1s_t} r_{t-1} + y_t, \quad (3.37)$$

$$y_t = \sqrt{h_t} u_t, \quad u_t \sim N(0, 1), \quad (3.38)$$

$$\log h_{t+1} = \alpha_{s_{t+1}} + \phi_{s_{t+1}} \log h_t + \eta_t, \quad \eta_t \sim N(0, \sigma^2). \quad (3.39)$$

So et al. [90] introdujeron un modelo de volatilidad de umbral multivariado para series de tiempo de rendimiento financiero múltiple. El modelo multivariado THSV es como sigue:

Sea  $r_t = (r_{1t}, \dots, r_{pt})^T$  la serie de tiempo de dimensión  $m$  de interés. Defina un conjunto de variables aleatorias Bernoulli  $s_t$  por

$$s_t = \begin{cases} 0 & \text{si } z_{t-1} < 0, \\ 1 & \text{si } z_{t-1} \geq 0, \end{cases}$$

donde  $z_t$  es la variable umbral. En general,  $z_t$  puede ser una función de la variable de retorno  $r_t$  o cualquier variable exógena.

El modelo MTHSV es dado por

$$r_t = \Psi_{0_{S_t}} + \Psi_{1_{S_t}} r_{t-1} + y_t, \quad (3.40)$$

$$y_{it} = \sqrt{h_{it}} u_{it}, \quad i = 1, \dots, p, \quad \text{donde } u_t \sim N_p(0, I_p), \quad (3.41)$$

$$\alpha_{i+1} = \Phi_{0_{S_{t-1}}} + \Phi_{1_{S_{t-1}}} \alpha_t + \eta_t, \quad \eta_t \sim N_p(0, \Sigma_\eta), \quad (3.42)$$

donde  $u_t$  y  $\eta_t$  son procesos de ruido blanco multivariados independientes,  $y_t = (y_{1t}, \dots, y_{pt})^T$  y  $\alpha_t = (\log h_{1t}, \dots, \log h_{pt})^T$ . Note que las matrices de coeficientes  $\Psi_{0_{S_t}}, \Psi_{1_{S_t}}, \Phi_{0_{S_{t+1}}}$  y  $\Phi_{1_{S_{t+1}}}$  dependen de  $z_t$ .

### 3.3.3. Efecto asimétrico general

Danielsson [29] sugirió un tipo alternativo de modelo SV asimétrico en el que utiliza la función de valor absoluto para capturar el signo y la magnitud del valor anterior de los retornos normalizados para incorporar el comportamiento asimétrico.

Asai y McAleer [5] llaman este modelo SV con efecto de apalancamiento y tamaño, (SV-LSE), además comentan que el modelo de Danielsson [29] incorporó el retorno observado en la especificación SV, ya que no es computacionalmente sencillo en el marco SV las perturbaciones normalizadas. El modelo SV-LSE es de la siguiente manera

$$y_t = \sigma_t \varepsilon_t, \quad t = 1, \dots, n, \quad (3.43)$$

$$\varepsilon_t = \rho \varepsilon_{t-1}, \quad \varepsilon_t \sim N(0, 1), \quad (3.44)$$

$$\sigma_t^2 = \exp[\omega + \gamma_1 y_{t-1} + \gamma_2 |y_{t-1}| + \delta \ln \sigma_{t-1}^2 + v \eta_t], \quad \eta_t \sim N(0, 1), \quad (3.45)$$

donde  $n$  es el número de observaciones. El vector de parámetros a estimar es  $\theta = (\omega, \delta, v, \gamma_1, \gamma_2, \rho)^T$ .

Asai y McAleer [5] extendieron el modelo SV-LSE a el caso multivariado de la siguiente manera

$$y_t = D_t \varepsilon_t, \quad (3.46)$$

$$D_t = \text{diag} \{ \exp(h_{1t}/2), \dots, \exp(h_{pt}/2) \} = \text{diag} \{ \exp(0.5h_t) \}, \quad (3.47)$$

$$h_{t+1} = \mu + \gamma_1 \odot y_t + \gamma_2 \odot |y_t| + \phi \odot h_t + \eta_t, \quad (3.48)$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\varepsilon & 0 \\ 0 & \Sigma_\eta \end{pmatrix} \right). \quad (3.49)$$

### 3.3.4. Modelo Factor

Cuando en un modelo MSV la dimensión  $p$  es muy grande podría ser computacionalmente difícil estimarlo por la cantidad de parámetros y variables latentes. Además, en muchos casos, es más apropiado usar factores comunes para describir los comovimientos entre muchos retornos de activos. Para permitir una parametrización más parsimoniosa, se ha propuesto en la literatura una clase de modelos factoriales MSV. También comentan que los modelos MSV factor además de la capacidad de garantizar la parsimonia, pueden capturar las características comunes en los retornos y volatilidades de los activos y, por lo tanto, tienen un vínculo natural con la teoría de precios de arbitraje (APT) de Ross [82].

Harvey et al. [50] introdujeron por primera vez el modelo MSV factor aditivo modificando el siguiente modelo de espacio de estados

$$W_t = -1.27 \mathbf{1}_p + h_t + \xi_t, \quad \xi_t \sim N_p(0, \Sigma_\xi), \quad (3.50)$$

$$h_t = h_{t-1} + \eta_t, \quad \eta_t \sim N_p(0, \Sigma_\eta), \quad (3.51)$$

donde  $W_t$  y  $\xi_t$  son vectores  $p \times 1$  con elementos  $w_{it} = \log y_{it}^2$  y  $\xi_{it} = \log \varepsilon_{it}^2 + 1.27$ ,  $i = 1, \dots, p$ , respectivamente y  $\mathbf{1}_p$  es un vector  $p \times 1$  de unos.  $\varepsilon_t$  y  $\eta_t$  son no correlacionadas.

Harvey et al. [50] modificaron el modelo de la ecuación (3.50)-(3.51) concentrándose en el caso donde hay movimientos persistentes en la volatilidad, modelados por una caminata aleatoria

multivariante para obtener el modelo factor. El modelo es como sigue

$$W_t = -1.27 \mathbf{1}_p + \theta h_t + \bar{h} + \xi_t, \quad (3.52)$$

$$h_t = h_{t-1} + \eta_t, \quad (3.53)$$

donde  $\theta$  es una matriz de coeficientes  $p \times k$  con  $k \leq p$ ,  $h_t$  y  $\eta_t$  son vectores de  $k \times 1$ ,  $\Sigma_\eta$  es una matriz definida positiva  $k \times k$  y  $\bar{h}$  es un vector  $p \times 1$  en el que los primeros  $k$  elementos son iguales a cero, mientras que los  $p - k$  elementos restantes no están restringidos.

Asai et al. [6] comentan que la idea básica se toma prestada de los modelos ARCH multivariados de factores y, más generalmente de las descomposiciones factoriales de las estructuras de covarianza en el análisis multivariado, donde los retornos se descomponen en dos componentes aditivos. El primer componente tiene un número menor de factores, que captura la información relevante para la fijación de precios de todos los activos, mientras que el otro componente es el ruido idiosincrásico, que captura la información específica del activo.

El modelo de Harvey [50] fue extendido por Jacquier [54], Kim et al. [59], Pitt y Shephard [77], Aguilar y West [1]. Chib et al. [19] propusieron un modelo MSV factor general con errores de colas pesadas y saltos. El modelo es dado por

$$y_t = B f_t + K_t q_t + u_t, \quad (3.54)$$

donde  $y_t = (y_{1t}, \dots, y_{pt})^T$  denotan las  $p$  observaciones en el tiempo  $t$ ,  $f_t = (f_{1t}, \dots, f_{kt})^T$  denotan los  $k$  factores no observados,  $q_t$  denota  $p$  variables aleatorias independientes Bernoulli "saltos".  $B$  es una matriz de parámetros desconocidos, y  $K_t = \text{diag}(k_{1t}, \dots, k_{pt})$  son los tamaños de salto. Chib et al. [19] asumen que cada elemento  $q_{jt}$  de  $q_t$  toma el valor uno con probabilidad  $\kappa_j$  y el valor cero con probabilidad  $1 - \kappa_j$ . Cada elemento  $u_{jt}$  de  $u_t$  sigue una distribución  $t$ -student con grados de libertad  $v_j > 2$ , que expresan en forma jerárquica como

$$u_{jt} = \lambda_{jt}^{-1/2} \varepsilon_{jt}, \quad \lambda_{jt} \sim \text{Gamma} \left( \frac{v_j}{2}, \frac{v_j}{2} \right), \quad t = 1, \dots, n,$$



donde

$$\begin{pmatrix} \varepsilon_t \\ f_t \end{pmatrix} \Big| V_t, D_t, K_t, q_t \sim N_{p+k} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_t & 0 \\ 0 & D_t \end{pmatrix} \right),$$

$\varepsilon_t$  y  $f_t$  son vectores aleatorios gaussianos condicionalmente independientes y

$$V_t = V_t(h_t) = \text{diag} (\exp(h_{1t}), \dots, \exp(h_{pt})), \text{ es de dimensión } p \times p,$$

$$D_t = D_t(h_t) = \text{diag} (\exp(h_{p+1,t}), \dots, \exp(h_{p+k,t})), \text{ es de dimensión } k \times k,$$

donde  $h_t$  son las variables aleatorias no observables. Cada  $h_{jt}$  sigue un proceso de volatilidad estocástica independiente de tres parámetros  $(\mu_j, \phi_j, \sigma_j)$

$$\begin{aligned} h_{jt} - \mu_j &= \phi_j (h_{j,t-1} - \mu_j) + \sigma_j \eta_{jt}, & \eta_{jt} &\sim N(0, 1), \\ \ln(1 + \kappa_{jt}) &\sim N \left( -\frac{1}{2} \delta_j^2, \delta_j^2 \right), & j &\leq p, \\ \delta &= (\delta_1, \dots, \delta_p) \text{ son parámetros desconocidos.} \end{aligned}$$

Ishihara y Omori [74] extendieron el modelo MSV factor discutido por Chib et al. [19] al modelo con efectos de apalancamiento pero sin saltos. El modelo es de la siguiente manera

$$Y_t = B f_t + \lambda_t^{-1/2} V_{1t}^{1/2} \varepsilon_{1t}, \quad t = 1, \dots, n, \quad (3.55)$$

$$f_t = V_{2t}^{1/2} \varepsilon_{2t}, \quad t = 1, \dots, n, \quad (3.56)$$

$$\alpha_{t+1} = \Phi \alpha_t + \eta_t, \quad t = 1, \dots, n-1, \quad (3.57)$$

$$\alpha_{j1} \sim N \left( 0, \frac{\sigma_{j\varepsilon}^2}{(1 - \phi_j^2)} \right), \quad j = 1, \dots, p+k, \quad (3.58)$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N_{2(p+k)}(0, \Sigma), \quad \Sigma = \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\eta} \\ \Sigma_{\eta\varepsilon} & \Sigma_{\eta\eta} \end{pmatrix}, \quad (3.59)$$

$$\lambda_t \sim \text{Gamma} \left( \frac{\nu}{2}, \frac{\nu}{2} \right),$$

$$V_t^{1/2} = \text{diag} \left( V_{1t}^{1/2}, V_{2t}^{1/2} \right),$$

$$\varepsilon_t = (\varepsilon_{1t}^T, \varepsilon_{2t}^T)^T,$$

$$\Phi = \text{diag} (\phi_1, \dots, \phi_{p+k}),$$

$$\begin{aligned}
 V_{1t}^{1/2} &= \text{diag} (\exp(\alpha_{1t}/2), \dots, \exp(\alpha_{pt}/2)), \\
 V_{2t}^{1/2} &= \text{diag} (\exp(\alpha_{m+1,t}/2), \dots, \exp(\alpha_{p+k,t}/2)), \\
 \Sigma_{\varepsilon\varepsilon} &= \text{diag} (\sigma_{1\varepsilon}^2, \dots, \sigma_{p+k,\varepsilon}^2), \\
 \Sigma_{\eta\eta} &= \text{diag} (\sigma_{1\eta}^2, \dots, \sigma_{p+k,\eta}^2), \\
 \Sigma_{\varepsilon\eta} &= \text{diag} (\rho_1 \sigma_{1\varepsilon} \sigma_{1\eta}, \dots, \rho_{p+k} \sigma_{p+k,\varepsilon} \sigma_{p+k,\eta}), \\
 \alpha_t &= (\alpha_{1t}, \dots, \alpha_{p+k,t})^T = h_t - \mu.
 \end{aligned}$$

Quintana y West [79] consideraron por primera vez el modelo de factores multiplicativos también conocido como modelos de factores discontinuo. Un modelo factor multiplicativo con  $k$  factores es dado por

$$r_t = \mu_t + \exp\left(\frac{Wh_t}{2}\right) \xi_t, \quad (3.60)$$

$$h_{it+1} = \delta_i + \gamma_i h_{it} + \sigma_{\eta_i} \eta_{it}, \quad i = 1, \dots, k, \quad (3.61)$$

donde  $W$  es una matriz de  $N \times k$  de cargas factoriales que es de rango  $k$  y  $h_t = (h_{1t}, \dots, h_{kt})^T$ .

### 3.4. Métodos de Estimación

En los modelos GARCH la densidad predictiva de los retornos depende de la volatilidad, que se puede medir con respecto al conjunto de información implicando que la estimación por máxima verosimilitud sea sencilla. En los modelos MSV no es sencillo debido a que la volatilidad está latente y debe integrarse a partir de la función verosimilitud conjunta de los retornos y la volatilidad.

Por lo que los modelos MSV no tienen una expresión de forma cerrada para la función verosimilitud. Diferentes métodos se han implementado para la estimación tanto de los modelos SV como los MSV, una revisión completa de estos métodos de estimación los podemos encontrar en Broto y Ruiz [15], Asai [6] y Handbook [8].

Asai et al. [6] comentan cinco criterios que se deben tener en cuenta en la elección de un método de estimación:

1. Eficiencia estadística.
2. La estimación de la volatilidad latente.
3. Determinación de los métodos óptimos de filtrado, suavizado y pronóstico.
4. Eficiencia computacional.
5. Aplicabilidad para modelado flexible.

En éste apartado, se muestran algunos de los métodos de estimación aplicados en los modelos de la sección 3.2

Para el modelo de la ecuación (3.4)-(3.6) propuesto por Harvey [50] el método que propusieron para la estimación de los parámetros  $(\mu, \phi, \Sigma_\varepsilon, \Sigma_\eta, H)$  donde  $H = (h_1^T, \dots, h_n^T)^T$  es el método cuasi máxima verosimilitud (QML). Para lograr esto hicieron los siguientes cálculos

$$\begin{aligned}
 y_{it} &= \exp\left(\frac{h_{it}}{2}\right) \varepsilon_{it}, \\
 y_{it}^2 &= \exp(h_{it}) \varepsilon_{it}^2, \\
 \log(y_{it}^2) &= h_{it} + \xi_{it}, \quad \xi_{it} = \log(\varepsilon_{it}^2).
 \end{aligned}$$

Obteniendo el modelo de espacio de estados

$$y_t^* = h_t + \xi_t, \quad E(\xi_t) = 1.2704 \mathbf{1}_p, \quad (3.62)$$

$$h_{t+1} = \mu + \phi \odot h_t + \eta_t, \quad (3.63)$$

$\xi_t$  es no normal aunque en el caso univariado es aproximado a una densidad normal. Además, demostraron que el  $ij$ -ésimo elemento de la matriz de covarianza de  $\xi_t$ ,  $\text{Cov}(\xi_t) = \Sigma_\xi$  es dado por  $(\pi^2/2) \rho_{ij}^*$ , donde

$$\rho_{ij}^* = \begin{cases} 1 & i = j, \\ \frac{2}{\pi^2} \sum_{k=1}^{\infty} \frac{(k-1)!}{(1/2)_k k} \rho_{ij}^{2k} & i \neq j, \end{cases}$$

$(x)_k = x(x+1)\dots(x+k-1)$ . El valor absoluto de  $\rho_{ij}$  y la correlación cruzada entre las diferentes  $\xi_{it}$  puede ser estimada a partir de las observaciones. Los signos se pueden obtener retornando a las observaciones no transformadas y señalando que el signo de cada una de los dos pares  $\xi_i\xi_j$  sea el mismo que el correspondiente del valor observado  $y_i y_j$ . Así el signo de  $\rho_{ij}$  es estimado como positivo si más de la mitad de los pares  $y_i y_j$  son positivos. Tratando  $\xi_t$  como gaussiana, el modelo de espacio de estados (3.62)-(3.63) se puede estimar usando los estimadores QML obtenidos a partir del filtro de Kalman.

Para el modelo de la ecuación (3.28)-(3.31) propuesto por Asai y McAleer [5] los parámetros a estimar son  $(\mu, \phi, \Sigma_\varepsilon, \Sigma_\eta, L, H)$ . El método que propusieron para estimar los parámetros fue derivando la forma espacio de estados para el modelo con efecto de apalancamiento de la ecuación (3.28)-(3.31). El espacio de estados resultante es

$$y_t^2 = \ln(y_t^2) = h_t + \xi_t, \quad (3.64)$$

$$\xi_t = \ln(\varepsilon_t^2),$$

$$h_{t+1} = \mu_t + \mu_t^* + \phi \odot h_t + \eta_t^*, \quad (3.65)$$

$$\eta_t^* \sim N_p(0, \Sigma_{\eta_t^*}^*), \quad (3.66)$$

$$E(\eta_t^* \xi_t^T) = L_t^*,$$

$$\Sigma_{\eta_t^*}^* = \Sigma_\eta - LP_\varepsilon^{-1}L + LP_\varepsilon^{-1} \left[ \left\{ P_{|\varepsilon|} - \frac{2}{\pi} tt^T \right\} \odot (s_t s_t^T) \right] P_\varepsilon^{-1}L,$$

$$\mu_t^* = \sqrt{\frac{2}{\pi}} LP_\varepsilon^{-1} s_t,$$

$$L_t^* = LP_\varepsilon^{-1} \left[ \left\{ P_{|\varepsilon|} - c\sqrt{\frac{2}{\pi}} \right\} \odot (s_t t^T) \right].$$

El espacio de estado (3.64)-(3.66) se puede estimar al igual que Harvey [50] aplicando el método QML basado en el filtro de Kalman. Asai y McAleer [5] utilizan el método de verosimilitud de Monte-Carlo el cual explican en detalle.

Para el modelo con efecto de apalancamiento cruzado (3.32)-(3.36) propuesto por Ishihara y Omori [53] los parámetros a estimar son  $(\Phi, \Sigma, \alpha)$ . Ishihara y Omori [53] tomaron un enfoque bayesiano y emplearon el método de simulación MCMC para generar muestras de la distribución posterior para realizar inferencias estadísticas con respecto a los parámetros del modelo.

Tomando  $\alpha = (\alpha_1^T, \dots, \alpha_n^T)^T$ ,  $Y_n = (y_1^T, \dots, y_n^T)^T$ , y  $\phi = (\phi_1, \dots, \phi_p)^T$  implementaron el algoritmo MCMC en tres etapas:

1. Generar  $\alpha \mid \phi, \Sigma, Y_n$ .
2. Generar  $\Sigma \mid \phi, \alpha, Y_n$ .
3. Generar  $\phi \mid \Sigma, \alpha, Y_n$ .

Para muestrear  $\alpha$  discuten dos métodos. El primer método es el muestreo de un solo movimiento (single-move) en el que muestrean un estado  $\alpha_t$  dado los otros estados  $\alpha_j$ . La función de densidad de probabilidad condicional posterior de  $\alpha_t$  es

$$\pi(\alpha_t \mid \{\alpha_s\}_{s \neq t}, \phi, \Sigma, Y_n) \propto \exp \left\{ -\frac{1}{2}(\alpha_t - m_{\alpha_t})^T \Sigma_{\alpha_t}^{-1}(\alpha_t - m_{\alpha_t}) + g(\alpha_t) \right\}, \quad (3.67)$$

donde

$$m_{\alpha_t} = \begin{cases} \Sigma_{\alpha_1} \left( -\frac{1}{2}1_p + \Phi M_1 \alpha_2 \right), & t = 1, \\ \Sigma_{\alpha_t} \left( -\frac{1}{2}1_p + \Phi M_t \alpha_{t+1} + M_{t-1} \Phi \alpha_{t-1} + N_{t-1} \right), & 1 < t < n, \\ \Sigma_{\alpha_n} \left( -\frac{1}{2}1_p + M_{n-1} \Phi \alpha_{n-1} + N_{n-1} \right), & t = n, \end{cases}$$

$$\Sigma_{\alpha_t} = \begin{cases} (\Sigma_0^{-1} + \Phi M_1 \Phi)^{-1}, & t = 1, \\ (M_{t-1} + \Phi M_t \Phi)^{-1}, & 1 < t < n, \\ M_{n-1}^{-1}, & t = n, \end{cases}$$

$$M_t = \Sigma_{\eta\eta}^{-1} + \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} S_t^{-1} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1}, \quad N_t = \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} S_t^{-1} V_t^{-1/2} y_t,$$

$$S_t = \begin{cases} \Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon}, & t < n, \\ \Sigma_{\varepsilon\varepsilon}, & t = n. \end{cases}$$

$$g(\alpha_t) = -\frac{1}{2} y_t^T \Sigma_t^{-1} y_t + y_t^T \Sigma_t^{-1} \mu_t, \quad \mu_t = V_t^{1/2} m_t, \quad \Sigma_t = V_t^{1/2} S_t V_t^{1/2},$$

$$m_t = \begin{cases} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} (\alpha_{t+1} - \Phi \alpha_t), & t < n, \\ 0, & t = n. \end{cases}$$

Ishihara y Omori [53] muestrean de (3.47) usando el algoritmo de Metropolis-Hastings (MH) generando un candidato  $\alpha_t^* \sim N(m_{\alpha_t}, \Sigma_{\alpha_t})$  y aceptando con probabilidad

$$\min \{ \exp (g(\alpha_t^*) - g(\alpha_t)), 1 \}$$

para  $t = 1, \dots, n$ , donde  $\alpha_t$  es el valor actual.

El segundo método es el muestreo de múltiples movimientos (multi-move) propuesto por primera vez por Shephard y Pitt [85] para estimar modelos de series de tiempo no gaussianos. Los muestreadores de múltiples movimientos funcionan actualizando varias variables al mismo tiempo.

Omori y Watanabe [76] desarrollaron un muestreo de bloques para modelos de volatilidad estocásticos asimétricos. Primero derivaron un algoritmo recursivo para encontrar una moda posterior del vector de estado para un modelo de mediciones no gaussianas con una ecuación de estado lineal utilizando la expansión de Taylor del logaritmo de la densidad posterior condicional para las perturbaciones. Luego, definen una ecuación de medición lineal y gaussiana aproximada basada en la moda posterior obtenida.

Ishihara y Omori [53] extendieron el modelo de Omori y Watanabe [76] al caso multivariado. Primero, dividen  $\alpha = (\alpha_1^T, \dots, \alpha_n^T)^T$  en  $K + 1$  bloques  $(\alpha_{k_{i-1}+1}^T, \dots, \alpha_{k_i}^T)^T$ ,  $i = 1, \dots, K + 1$ , con  $k_0 = 0$ ,  $k_{K+1} = n$  y  $k_i - k_{i-1} \geq 2$ .  $K$  nodos  $(k_1, \dots, k_K)$  son generados aleatoriamente usando

$$k_i = \text{int} \left[ \frac{n \times (i + U_i)}{(K + 2)} \right], \quad i = 1, \dots, K,$$

donde los  $U_i$  son variables aleatorias uniformes independientes en  $(0, 1)$ .

Para encontrar la distribución de densidad conjunta condicional completa del  $i$ -ésimo bloque suponen que  $k_{i-1} = s$  y  $k_i = s + m$ . Denotan  $x_t = R_t^{-1} \eta_t$ , donde la matriz  $R_t$  denota una descomposición de Choleski de  $\Sigma_{\eta\eta} = R_t R_t^T$  para  $t > 0$ ,  $\Sigma_o = R_0 R_0^T$  para  $t = 0$ . Para construir una distribución propuesta para el algoritmo MH, se enfocan en la distribución de las perturbaciones  $x = (x_s^T, \dots, x_{s+m-1}^T)^T$ , que es fundamental en el sentido que deriva la distribución de  $a = (\alpha_{s+1}^T, \dots, \alpha_{s+m}^T)^T$ . Entonces, el logaritmo de la distribución de densidad conjunta condicional completa de  $x$ , excluyendo los términos constantes, es dada por

$$\log f(x|\alpha_s, \alpha_{s+m+1}, y_s, y_{s+m}) = -\frac{1}{2} \sum_{t=s}^{s+m-1} x_t^T x_t + L, \quad (3.68)$$

donde

$$L = \sum_{t=s}^{s+m} \ell_t - \frac{1}{2} (\alpha_{s+m+1} - \Phi \alpha_{s+m})^T \Sigma_{\eta\eta}^{-1} (\alpha_{s+m+1} - \Phi \alpha_{s+m}) I(s+m < n),$$

$$\ell_t = \text{const} - \frac{1}{2} 1_p^T \alpha_t - \frac{1}{2} (z_t - m_t)^T S_t^{-1} (z_t - m_t),$$

$$z_t = V_t^{-1/2} y_t,$$

const = constante.

Usando la expansión de Taylor de segundo orden en (3.68) alrededor de la moda  $\hat{x}$ , obtienen la densidad normal aproximada de  $f^*$

$$\log f(x|\alpha_s, \alpha_{s+m+1}, y_s, y_{s+m}) \approx \text{const} - \frac{1}{2} \sum_{t=s}^{s+m-1} x_t^T x_t + \hat{L} + \hat{d}^T (a - \hat{a}) - \frac{1}{2} (a - \hat{a})^T \hat{Q} (a - \hat{a}), \quad (3.69)$$

$$\approx \text{const} + \log f^*(x|\alpha_s, \alpha_{s+m+1}, y_s, y_{s+m}), \quad (3.70)$$

donde  $Q = -E \left( \frac{\partial^2 L}{\partial a \partial a^T} \right)$  y  $d = \frac{\partial L}{\partial a}$  son evaluadas en  $a = \hat{a}$  ( es decir,  $x = \hat{x}$  ).

Para encontrar una moda  $\hat{x}$  Ishihara y Omori [53] interpretaron la ecuación (3.70) como la densidad de probabilidad posterior derivada de un modelo de espacio de estados auxiliar el cual es dado por el siguiente algoritmo:

### Algoritmo 1

1. Inicializar  $\hat{x}$  y calcular  $\hat{a}$  en  $x = \hat{x}$  usando (3.33).
2. Evaluar  $\hat{d}_t$ ,  $\hat{A}_t$  y  $\hat{B}_t$  para  $t = s + 1, \dots, s + m$  en la moda del paso (1).  $d_t$ ,  $A_t$  y  $B_t$  son definidas de la siguiente manera:

$$\begin{aligned}
 d_t &= -\frac{1}{2}1_p + \frac{1}{2} \left\{ \text{diag}(z_t) - 2\Phi \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} I(t < \eta) \right\} S_t^{-1}(z_t - m_t) \\
 &\quad + \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} S_{t-1}^{-1}(z_{t-1} - m_{t-1}) I(t > 1) \\
 &\quad + \Phi \Sigma_{\eta\eta}^{-1}(\alpha_{t+1} - \Phi\alpha_t) I(t = s + m < n), \\
 A_t &= \frac{1}{4} \left\{ I_p + S_t^{-1} \odot (S_t + m_t m_t^T) \right\} + \Phi \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} S_t^{-1} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Phi I(t < n) \\
 &\quad - \frac{1}{2} \left\{ \Phi \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} S_t^{-1} \text{diag}(m_t) + \text{diag}(m_t) S_t^{-1} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Phi \right\} I(t < n) \\
 &\quad + \Sigma_{\eta\eta}^{-1} \Sigma_{\varepsilon\eta} S_{t-1}^{-1} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} I(t > 1) + \Phi \Sigma_{\eta\eta} \Phi I(t = s + m < n), \\
 B_t &= \frac{1}{2} \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} S_{t-1}^{-1} \left\{ \text{diag}(m_{t-1}) - 2\Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Phi \right\},
 \end{aligned}$$

donde  $B_{s+1} = 0$ .

3. Calcular recursivamente  $D_t, b_t$  y  $\hat{\gamma}_t$  usando  $\hat{d}_t, \hat{A}_t$  y  $\hat{B}_t$  calculadas en el paso (2) para  $t = s + 1, \dots, s + m$ . Notar que  $\hat{B}_s = \hat{B}_{s+m+1} = 0$ ,  $D_s = 0$  y  $b_s = 0$ .

$D_t, b_t$  y  $\hat{\gamma}_t$  son definidos de la siguiente manera:

$$D_t = \hat{A}_t - \hat{B}_t D_{t-1}^{-1} \hat{B}_t^T, \quad b_t = \hat{d}_t - \hat{B}_t D_{t-1}^{-1} b_{t-1}, \quad \hat{\gamma}_t = \hat{\alpha}_t + D_t^{-1} \hat{B}_{t+1}^T \hat{\alpha}_{t+1}.$$

4. Aplicar el filtro de Kalman y el suavizador de perturbaciones de Koopman [60] al siguiente modelo de espacio de estados gaussianos lineal aproximado para obtener la moda posterior  $\hat{x}$  y  $\hat{a}$

$$\hat{y}_t = Z_t \alpha_t + G_t u_t, \quad t = s + 1, \dots, s + m, \quad (3.71)$$

$$\alpha_{t+1} = \Phi \alpha_t + H_t u_t, \quad t = s + 1, \dots, s + m - 1, \quad (3.72)$$

donde

$$u_t = (u_{1t}^T, x_t^T)^T \sim N_{2p}(0, I_{2p}),$$

$$H_t = [0, R_t],$$

$$\hat{y}_t = \hat{\gamma}_t + D_t^{-1} b_t, \quad Z_t = I_p + D_t^{-1} \hat{B}_{t+1}^T \Phi, \quad G_t = [(K_t^T)^{-1}, D_t^{-1} \hat{B}_{t+1}^T R_t],$$

$$D_t = K_t K_t^T, \quad K_t \text{ es la descomposición de Choleski de } D_t.$$



5. Aplicar nuevamente el paso (2) con la moda posterior  $\hat{x}$  y  $\hat{\alpha}$  calculadas en el paso (4).

El procedimiento se repite hasta que las estimaciones suavizadas converjan, y así obtener la moda posterior de  $x_t$  y  $\alpha_t$ . Shephard y Pitt [85] sugieren usar entre 2 a 5 iteraciones.

Para muestrear  $x$  de la densidad posterior condicional Ishihara y Omori [53] usan la moda  $\hat{x}$  obtenida del modelo de espacio de estado gaussiano lineal aproximado, y muestrean  $x$  de la distribución condicional posterior mediante un algoritmo AR-MH de la siguiente manera:

### Algoritmo 2

1. Encontrar una moda  $\hat{x}$  usando el algoritmo 1.
2. Usar la moda  $\hat{x}$  del paso (1) para obtener el modelo gaussiano lineal aproximado (3.71)-(3.72).
3. Proponer un candidato  $x^*$  muestreando de  $f(x^*) \propto \min(f(x^*), c f^*(x^*))$  usando el algoritmo AR, donde  $c$  puede ser construido de un término constante y  $\hat{L}$  de (3.69).
  - a) Generar  $x^* \sim f^*$  usando un simulador más suave (simulation smoother) (por ejemplo, Jong y Shephard [57]; Durbin y Koopman [37] usando el modelo gaussiano lineal aproximado del paso (2)).
  - b) Aceptar  $x^*$  con probabilidad  $\min\left\{\frac{f(x^*)}{c f^*(x^*)}, 1\right\}$ . Si es rechazado, volver a (a).
4. Dado el valor actual  $x$ , se acepta  $x^*$  con probabilidad

$$\min\left\{1, \frac{f(x^*) \min(f(x), c f^*(x))}{f(x) \min(f(x^*), c f^*(x))}\right\}.$$

Si es rechazado, se acepta el actual  $x$  como muestra.

Para muestras  $\phi$  y  $\Sigma$ , Ishihara y Omori [53] suponen las siguientes distribuciones a priori

$$\begin{aligned} \frac{\phi_j + 1}{2} &\sim \beta(a_j, b_j), \quad j = 1, \dots, p, \\ \Sigma &\sim IW(n_0, R_0), \end{aligned}$$

donde  $\beta(a_j, b_j)$  y  $IW(n_0, R_0)$  denotan la distribución Beta y Inversa Wishart respectivamente con funciones de densidad de probabilidad

$$\begin{aligned}\pi(\phi_j) &\sim (1 + \phi_j)^{a_j-1} (1 - \phi_j)^{b_j-1}, \quad j = 1, \dots, p, \\ \pi(\Sigma) &\sim |\Sigma|^{(-n_0+p+1)/2} \exp \left\{ -\frac{1}{2} \text{tr} (R_0^{-1} \Sigma^{-1}) \right\}.\end{aligned}$$

La función de densidad de probabilidad posterior condicional de  $\phi$  es

$$\pi(\phi|\Sigma, \alpha, Y_n) \propto h(\phi) \times \exp \left\{ -\frac{1}{2} (\phi - \mu_\phi)^T \Sigma_\phi^{-1} (\phi - \mu_\phi) \right\}, \quad (3.73)$$

$$h(\phi) = |\Sigma_0|^{-1/2} \prod_{j=1}^P (1 + \phi_j)^{a_j-1} (1 - \phi_j)^{b_j-1} \exp \left\{ -\frac{1}{2} \alpha_1^T \Sigma_0 \alpha_1 \right\},$$

$$\mu_\phi = \Sigma_\phi b,$$

$$\Sigma_\phi^{-1} = \Sigma^{22} \odot A,$$

$$A = \sum_{t=1}^{n-1} \alpha_t \alpha_t^T,$$

$$b = \text{diag}(B_{11}, \dots, B_{pp}), \quad B_{ii} \text{ es el } (i,i)\text{-ésimo elemento de } B,$$

$$B = \sum_{t=1}^{n-1} \left\{ \alpha_t y_t^T V_t^{-1/2} \Sigma^{12} + \alpha_t \alpha_{t+1}^T \Sigma^{12} \right\},$$

donde

$$\Sigma^{ii} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$$

Ishihara y Omori [53] muestrean de (3.73) usando un algoritmo MH generando un candidato de una distribución normal truncada sobre la región  $R$ ,  $\phi^* \sim TN_R(\mu_\phi, \Sigma_\phi)$  y  $R = \{\phi : |\phi_j| < 1, j = 1, \dots, p\}$  y aceptando con probabilidad

$$\min \left\{ \frac{h(\phi^*)}{h(\phi)}, 1 \right\},$$

donde  $\phi$  es la muestra actual.

La función de densidad de probabilidad posterior condicional de  $\Sigma$  es

$$\begin{aligned} \Phi(\Sigma|\phi, \alpha, Y_n) \propto |\Sigma|^{(n_1+2p+1)/2} \exp \left\{ -\frac{1}{2} \text{tr}(R_1^{-1}\Sigma^{-1}) \right\} \times g(\Sigma), \\ g(\Sigma) = |\Sigma_0|^{-1/2} |\Sigma_{\varepsilon\varepsilon}|^{-1/2} \exp \left\{ -\frac{1}{2} (\alpha_1^T \Sigma_0 \alpha_1 + y_n^T V_n^{-1/2} \Sigma_{\varepsilon\varepsilon}^{-1} V_n^{-1/2} y_n) \right\}, \end{aligned} \quad (3.74)$$

donde  $n_1 = n_0 + n - 1$ ,  $R_1^{-1} = R_0^{-1} + \sum_{t=1}^{n-1} v_t v_t^T$  y

$$V_t = \begin{pmatrix} V_t^{-1/2} y_t \\ \alpha_{t-1} - \Phi \alpha_t \end{pmatrix}.$$

Ishihara y Omori [53] muestrean de (3.74) usando un algoritmo MH generando un candidato  $\Sigma^* \sim IW(n_1, R_1)$  y aceptando con probabilidad

$$\min \left\{ \frac{g(\Sigma^*)}{g(\Sigma)}, 1 \right\},$$

donde  $\Sigma$  es la muestra actual.

Para el modelo de volatilidad de umbral multivariado de la ecuación (3.40)-(3.42) propuesto por So [90], el método de estimación de los parámetros desconocidos es llevado a cabo utilizando técnicas de MCMC.

### 3.5. Comparación de Modelos

Para escoger el mejor modelo Ishihara y Omori [53] para cada modelo estimado calculan el Criterio de Información de Desviación (DIC) Spiegelhalter et al. [92]. Utilizan el filtro de partículas auxiliar propuesto por Shephard y Pitt [84] para calcular la función verosimilitud ordinaria dado los parámetros. Por último escogen el modelo que tiene la medida DIC más pequeña.

## Filtro de partículas

Para el modelo (3.32)-(3.36) propuesto por Ishihara y Omori [53] el filtro de partículas es como sigue: Sea  $Y_t = (y_1, \dots, y_t)$  el conjunto de información hasta el tiempo  $t$  y  $\theta$  los parámetros,  $f(\alpha_t|Y_t, \theta)$  la función de densidad de probabilidad de  $\alpha_t$  dado  $(Y_t, \theta)$  y, es aproximada por  $\hat{f}(\alpha_t|Y_t, \theta)$ . Considere el muestreo de la distribución conjunta condicional de  $(\alpha_{t+1}, \alpha_t)$  dado  $(Y_{t+1}, \theta)$  con función de densidad de probabilidad dado por

$$f(\alpha_{t+1}|Y_{t+1}, \theta) \propto f(y_{t+1}|\alpha_{t+1})f(\alpha_{t+1}|y_t, \alpha_t, \theta)f(\alpha_t|Y_t, \theta), \quad (3.75)$$

donde

$$f(y_t|\alpha_t) = (2\pi)^{-\frac{p}{2}} \left| V_t^{\frac{1}{2}} \Sigma_{\varepsilon\varepsilon} V_t^{\frac{1}{2}} \right| \exp \left\{ -\frac{1}{2} y_t^T V_t^{-\frac{1}{2}} \Sigma_{\varepsilon\varepsilon}^{-1} V_t^{-\frac{1}{2}} y_t \right\}, \quad (3.76)$$

$$f(\alpha_{t+1}|y_t, \alpha_t, \theta) = (2\pi)^{-\frac{p}{2}} |\Sigma_\alpha|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\alpha_{t+1} - \mu_{\alpha, t+1})^T \Sigma_\alpha^{-1} (\alpha_{t+1} - \mu_{\alpha, t+1}) \right\}, \quad (3.77)$$

$$\mu_{\alpha, t+1} = \Phi \alpha_t + \Sigma_{\eta\varepsilon} \Sigma_{\varepsilon\varepsilon}^{-1} V_t^{-\frac{1}{2}} y_t, \quad \Sigma_\alpha = \Sigma_{\eta\eta} - \Sigma_{\eta\varepsilon} \Sigma_{\varepsilon\varepsilon}^{-1} \Sigma_{\varepsilon\eta}.$$

La función de densidad de probabilidad de importancia es dada por:

$$g(\alpha_{t+1}^i|Y_{t+1}, \theta) \propto f(\alpha_{t+1}|y_t, \alpha_t^i, \theta)g(\alpha_t^i|Y_{t+1}, \theta),$$

donde

$$g(\alpha_t^i|Y_{t+1}, \theta) = \frac{f(y_{t+1}|\mu_{t+1}^i)\hat{f}(\alpha_t^i|Y_t, \theta)}{\sum_{i=1}^I f(y_{t+1}|\mu_{t+1}^i)\hat{f}(\alpha_t^i|Y_t, \theta)},$$

$$\mu_{\alpha, t+1}^i = \Phi \alpha_t^i + \Sigma_{\eta\varepsilon} \Sigma_{\varepsilon\varepsilon}^{-1} V_t^{i-\frac{1}{2}} y_t, \quad V_t^i|_{\alpha_t=\alpha_t^i}.$$

La función de densidad de probabilidad de importancia  $g(\alpha_{t+1}^i|Y_{t+1}, \theta)$  ayuda a los pesos  $w_i$  a ser más uniformes.

El filtro de partículas auxiliar se deriva de la siguiente manera:

1. Inicializar  $t = 1$ , y generar  $\alpha_1^i \sim N(0, \Sigma_0)$ , ( $i = 1, \dots, I$ ).

a) Calcular  $w_i = f(y_1 | \alpha_1^i, \theta)$  y almacenar  $\bar{w}_1 = \frac{1}{I} \sum_{i=1}^I w_i$ .

b) Sea  $\hat{f}(\alpha_1^i | Y_1, \theta) = \pi_1^i = \frac{w_i}{\sum_{j=1}^I w_j}$ , ( $i = 1, \dots, I$ ).

2. a) Para cada  $i$ , generar  $(\alpha_{t+1}^i, \alpha_t^i)$  de  $g(\alpha_{t+1}, \alpha_t^i | Y_{t+1}, \theta)$ , ( $i, j = 1, \dots, I$ ) de la siguiente manera: Remuestrear  $\alpha_t^i, i = 1, \dots, I$ , con probabilidad  $g(\alpha_t^j | Y_{t+1}, \theta), j = 1, \dots, I$ . Entonces generar  $\alpha_{t+1}^i, i = 1, \dots, I$ , de la distribución normal con función de densidad  $f(\alpha_{t+1} | y_t, \alpha_t^i, \theta)$ .

b) Calcular

$$w_i = \frac{f(y_{t+1} | \alpha_{t+1}^i) \hat{f}(\alpha_t^i | Y_t, \theta)}{g(\alpha_{t+1}^i | Y_{t+1}, \theta)}$$

para  $i = 1, \dots, I$ , y almacenar  $\bar{w}_t = \sum_{i=1}^I \frac{w_i}{I}$ .

c) Sea  $\hat{f}(\alpha_{t+1}^i | Y_{t+1}, \theta) = \pi_{t+1}^i = \frac{w_i}{\sum_{j=1}^I w_j}$ , ( $i = 1, \dots, I$ ).

3. Incrementar  $t$  una unidad y volver al paso 2.

Luego,

$$\sum_{t=1}^n \log \bar{w}_t \xrightarrow{p} \sum_{t=1}^n \log f(y_t | Y_{t-1}, \theta) \quad \text{cuando } I \rightarrow \infty. \quad (3.78)$$

es un estimador consistente de la verosimilitud logarítmica condicional.

## Criterio de Información de Desviación

La medida DIC Spiegelhalter et al. [92] es definido por

$$\text{DIC} = E_{\theta | Y_n} [D(\theta)] + P_D, \quad (3.79)$$

donde

$$P_D = E_{\theta|Y_n}[D(\theta)] - D(E_{\theta|Y_n}[\theta]), \quad D(\theta) = -2 \log f(Y_n|\theta). \quad (3.80)$$

Para calcular  $E_{\theta|Y_n}[D(\theta)]$  se puede aproximar mediante  $\frac{1}{M} \sum_{m=1}^M D(\theta^{(m)})$ ,  $\theta^{(m)}$  se vuelven a muestrear a partir de la distribución posterior. El error estándar del estimador es obtenido estimando repetidamente  $E_{\theta|Y_n}[D(\theta)]$ .  $D(E_{\theta|Y_n}\theta)$  es igual a  $D(\theta)$  evaluado en la media posterior. Ishihara y Omori [53] configuraron  $M = 100$ ,  $I = 10000$  y repitieron 10 veces  $E_{\theta|Y_n}[D(\theta)]$  para obtener el error estándar.

# Capítulo 4

## Modelo Autorregresivo Vectorial con Volatilidad Estocástica Multivariada

Los mercados financieros son dependientes entre sí, por esa razón es necesario considerarlos en conjunto para comprender mejor su estructura dinámica. Conocer cómo se interrelacionan los mercados es de gran importancia en las finanzas. Los modelos VAR son excelentes herramientas para conocer el comportamiento dinámico de las series financieras.

Una característica de las series financieras es la volatilidad. La volatilidad no es constante en el tiempo por lo que un modelo como en (2.1) en el que la matriz de covarianza es constante en el tiempo no es adecuado para modelar las series de tiempo financieras. Los modelos MSV son útiles para medir la volatilidad. Muchos estudios (ver por ejemplo Primiceri [78]) en los que se integran los modelos VAR con los MSV muestran evidencia empírica de una mejoría significativa en la descripción de los datos y el pronóstico de variables macroeconómicas. En este capítulo integramos un modelo VAR con un modelo MSV de apalancamiento cruzado.

El resto del capítulo está organizado de la siguiente manera. En la sección 4.1 se presentan diferentes modelos MSV-VAR propuestos y sus aplicaciones. En la sección 4.2 se presenta el modelo VAR-MSV con efecto de apalancamiento cruzado. En la sección 4.3 se presenta el método de estimación. En la sección 4.4 se expone el método para escoger el orden VAR-MSV. En la sección 4.5 se presenta un ejemplo con datos simulados. Por último en la sección 4.6 se presentan resultados empíricos

aplicando el modelo MSV-VAR a datos reales.

## 4.1. Revisión de la Literatura de Modelos VAR-MSV

Uhlig [101] introduce la volatilidad estocástica multivariada sin restricciones en el contexto de los modelos autorregresivos vectoriales. El modelo que propuso es de la siguiente manera

$$Y_t = A_0 V_t + B_1 y_{t-1} + \dots + A_k y_{t-k} + R_t^{-1} \varepsilon_t, \quad t = 1, \dots, n, \quad (4.1)$$

$$H_{t+1} = \frac{1}{\lambda} R_t^T \Sigma_t R_t, \quad t = 0, \dots, n-1, \quad (4.2)$$

donde

$$\begin{aligned} \varepsilon_t &\sim N(0, I_p), \\ \varepsilon_t &\sim \beta_p \left( \frac{v + c + kp}{2}, \frac{1}{2} \right), \end{aligned}$$

$Y_t$ ,  $t = 1 - k, \dots, n$  de dimensión  $p \times 1$  son datos observables.  $V_t$  de dimensión  $c \times 1$  denota regresores deterministas como una constante y una tendencia de tiempo. La matriz de coeficientes  $B_0$  es de dimensión  $p \times c$ . Las matrices de coeficientes  $B_i$ ,  $i = 1, \dots, k$  son de dimensión  $p \times p$ .  $v > p - 1$  y  $\lambda > 0$  son parámetros.  $\varepsilon_t$ ,  $t = 1, \dots, n$  son de dimensión  $p \times 1$ .  $\Sigma_t$ ,  $t = 1, \dots, n$  son de dimensión  $p \times p$  distribuidos independientemente.  $R_t$  denota la descomposición de Cholesky superior de  $H_t$  y  $\beta_m(a, b)$  denota la distribución beta multivariada. Uhlig [101] escogió la distribución beta multivariante para explotar una conjugación entre esa distribución y la distribución Wishart para que la integración sobre el choque no observado en la matriz de precisión se puede realizar de forma cerrada, lo que lleva a una generalización de las fórmulas estándar de filtro de Kalman, el problema de filtrado no lineal.

Cogley [25] para estimar la tendencia de crecimiento en la “nueva economía” desarrolló una estrategia de filtrado bayesiano. La especulación sobre el crecimiento es importante para una variedad de decisiones de política pública. Además, las estimaciones de crecimiento tendencial también son importantes para la política monetaria. En un modelo homocedástico como el de la ecuación



(2.1) del capítulo 2 la tendencia de crecimiento es constante en el tiempo. Cogley [25] supone que la tendencia de crecimiento varía con el tiempo. Define el crecimiento tendencial de la siguiente manera

$$\tau_t = S_{\Delta q}(I - A_t)^{-1}\mu_t \quad (4.3)$$

y el espectro de poder instantáneo se define como

$$f_{yy}(w, t) = S (I - A_t e^{-iw})^{-1} \left( \frac{\Sigma_t}{2\pi} \right) \left[ (I - A_t e^{-iw})^{-1} \right]^T S^T. \quad (4.4)$$

Cogley [25] calcula  $A_t, \mu_t, \Sigma_t, t = 1, \dots, n$  y  $S_{\Delta q}, S$  usando un modelo autorregresivo vectorial bayesiano con parámetros variando en el tiempo que es dado por

$$Y_t = X_t^T \beta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t), \quad (4.5)$$

donde  $X_t$  incluye constantes más rezagos de  $Y_t$ ,  $\beta_t$  es un vector de parámetros VAR variando en el tiempo.

Al aplicar las variables de manera correcta expresa (4.5) en forma complementaria como

$$z_t = \mu_t + A_t z_{t-1} + u_t, \quad (4.6)$$

donde  $z_t$  consiste en valores actuales y lags de  $Y_t$ ,  $\mu_t$  contiene las intersecciones en  $\beta_t$  y  $A_t$  contiene los parámetros autorregresivos en el tiempo  $t$ .  $S_{\Delta q}$  es un vector fila que selecciona el crecimiento del producto interno bruto (PIB) real de  $z_t$ .  $\beta_t$  evoluciona como una caminata aleatoria con barrera reflectante de la siguiente manera

$$\beta_t = \beta_{t-1} + W_t, \quad W_t \sim N(0, Q). \quad (4.7)$$

Para modelar  $\Sigma_t$  adopta una versión multivariada del modelo de volatilidad estocástica de Jacquier et al. [56] donde

$$\Sigma_t = B^{-1} H_t (B^{-1})^T, \quad (4.8)$$

$$H_t = \begin{pmatrix} h_{1t} & 0 \\ 0 & h_{2t} \end{pmatrix}, \quad (4.9)$$

$$B = \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}, \quad (4.10)$$

$h_{it}$ ,  $i = 1, 2, \dots$  son volatilidades estocásticas univariadas e independientes que evolucionan como caminata aleatoria geométricas sin deriva.

$$\log(h_{it}) = \log(h_{it-1}) + \sigma_i \eta_{it}, \quad \eta_{it} \sim N(0, 1), \quad (4.11)$$

donde  $\eta_{it}$  son independientes entre sí y de los otros choques del modelo,  $E(\xi_t W_t^T) = 0 \quad \forall s, t \quad \xi_t = \Sigma_t^{-1/2} \varepsilon_t$ . También se asume que ambos  $W_t$  y  $\xi_t$  son independientes entre  $\eta_t$ .

Cogley y Sargent [26] presentan densidades posteriores para varios objetos que son pertinentes para diseñar y evaluar política monetaria. Para este fin usan un modelo autorregresivo vectorial con parámetros variando en el tiempo y volatilidad estocástica como en (4.5), (4.7)-(4.11) con sus respectivas restricciones. Las variables que usan en este modelo son inflación, desempleo y tasa de interés.

Primiceri [78] estimó un modelo autorregresivo vectorial estructural con parámetros variando en el tiempo con el fin de investigar las causas potenciales del pobre desempeño económico de los Estados Unidos en los años 70 y principios de los 80 y, en qué medida la política monetaria desempeño un papel importante en esos episodios de alto desempleo e inflación. Primiceri [78] usa los parámetros variando en el tiempo para medir los cambios de política y los cambios implícitos en el comportamiento del sector privado. Las ecuaciones múltiples del modelo las usa para comprender cómo los cambios en la política han afectado al resto de la economía. A diferencia del modelo de Cogley [25] y, Cogley y Sargent [26] el modelo propuesto por Primiceri [78] permite que las correlaciones simultaneas entre las variables varíen en el tiempo. El modelo es de la siguiente manera

$$Y_t = V_t + A_{1,t}Y_{t-1} + \dots + A_{k,t}Y_{t-k} + B_t^{-1}\Sigma_t\varepsilon_t, \quad (4.12)$$

donde  $\text{VAR}(\varepsilon_t) = I_p$ .

$Y_t$  es un vector de dimensión  $p \times 1$  de variables endógenas observadas.  $V_t$  es un vector de dimensión  $p \times 1$  de coeficientes variando en el tiempo que multiplican términos constantes.  $A_{it}$ ,  $i = 1, \dots, k$  son matrices de dimensión  $p \times p$  de coeficientes variando en el tiempo

$$\Sigma_t = \text{diag}(\sigma_{1,t}, \dots, \sigma_{p,t}),$$

$$B_t = \begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{21,t} & 1 & & \\ \vdots & \ddots & \ddots & \vdots \\ a_{p1,t} & \dots & a_{pp-1,t} & 1 \end{bmatrix}.$$

Primiceri [78] reescribe (3.12) de la siguiente manera

$$Y_t = X_t^T \beta_t + B_t^{-1}\Sigma_t\varepsilon_t. \quad (4.13)$$

La dinámica de los parámetros variando en el tiempo del modelo los especifica de la siguiente manera

$$\beta_t = \beta_{t-1} + w_t, \quad (4.14)$$

$$\lambda_t = \lambda_{t-1} + \varrho_t, \quad (4.15)$$

$$\log(\sigma_t) = \log(\sigma_{t-1}) + \eta_t, \quad (4.16)$$

donde  $\lambda_t$  es el vector de elementos distintos de cero y distintos de uno de la matriz  $B_t$ ,  $\sigma_t$  es el vector de los elementos de la diagonal de  $\Sigma_t$ .

Primiceri [78] supone que todas las innovaciones en el modelo tienen una distribución normal conjunta con los siguientes supuestos sobre la matriz de covarianza

$$\text{Var} \begin{bmatrix} \left( \begin{array}{c} \varepsilon_t \\ w_t \\ \varrho_t \\ \eta_t \end{array} \right) \end{bmatrix} = \begin{bmatrix} I_p & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix}, \quad (4.17)$$

donde  $Q, S$  y  $W$  son matrices definidas positivas.

Benati [10] usa un modelo autorregresivo vectorial estructural bayesiano de parámetros variables en el tiempo con volatilidad estocástica como en (4.12)-(4.17), para investigar la contribución de la política monetaria a los cambios en el crecimiento de la producción y la dinámica de la inflación en los Estados Unidos.

Galí y Gambetti [45] usa un modelo autorregresivo vectorial estructural con parámetros variando en el tiempo y volatilidad estocástica como en (4.12)-(4.17) con el fin de proporcionar evidencia sobre algunos de los cambios experimentados por la economía de Estados Unidos durante el período posterior a la Segunda Guerra Mundial y alrededor del momento del quiebre de la volatilidad asociado con la Gran moderación.

Basándose en la evidencia de Cogley y Sargent [26] y Primiceri [78] de que los modelos autorregresivos vectoriales con parámetros variables en el tiempo y volatilidad estocástica son muy relevantes para caracterizar correctamente los cambios estructurales en la economía de los Estados Unidos, Gambetti et al. [28] propusieron un modelo como en (4.12)-(4.17) para pronosticar en tiempo real tres variables macroeconómicas de los Estados Unidos, a saber, tasa de desempleo, inflación y una tasa de interés a corto plazo con el objetivo de evaluar si el modelo de los cambios estructurales puede mejorar la precisión de los pronósticos de las series temporales macroeconómicas. Pronostican hasta 3 años hacia adelante usando errores de pronóstico cuadráticos medios y puntajes predictivos logarítmicos. Gambetti et al. [28] obtienen pronósticos utilizando la densidad predictiva posterior.

Clark [23] usa un modelo autorregresivo vectorial bayesiano con volatilidad estocástica para examinar la precisión de los pronósticos de densidad en tiempo real de las variables macroeconómicas de Estados Unidos. Evalúa los pronósticos para cuatro variables: Crecimiento de la producción,

tasa de desempleo, inflación y tasa de fondos federales. El modelo es de la siguiente manera

$$\Phi(L)(Y_t - A_0 d_t) = B^{-1} V_t^{1/2} \varepsilon_t, \quad \varepsilon_t \sim N(0, I_p), \quad (4.18)$$

$$V_t^{1/2} = \text{diag} \left( \frac{h_{1t}}{2}, \frac{h_{2t}}{2}, \dots, \frac{h_{pt}}{2} \right), \quad (4.19)$$

$$\log(h_{it}) = \log(h_{it-1}) + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma_i), \quad (4.20)$$

donde  $Y_t$  es de dimensión  $p \times 1$  son vectores observados,  $d_t$  de dimensión  $q \times 1$  son vectores de variables determinadas.  $\Phi(L) = I_p - A_1 L - A_2 L^2 - \dots - A_k L^k$ .  $L$  es un operador tal que  $L^i Y_t = Y_{t-i}$ .  $A_0$  es una matriz de dimensión  $p \times q$  de coeficientes en la variable determinista.  $B$  es una matriz triangular inferior con elementos 1 en la diagonal y coeficientes  $b_{ij}$  en la  $i$ -ésima fila y  $j$ -ésima columna,  $i = 2, \dots, p$ ,  $j = 1, \dots, i - 1$ .  $\eta_{it}$  son independientes e idénticamente distribuidas.

Clark y Ravazzolo [24] compara la precisión para el pronóstico puntual y de densidad en tiempo real de varios modelos. Los modelos que usa son los autorregresivos y los autorregresivos vectoriales. Ambas clases de modelos son probados con distintas configuraciones en la volatilidad. Las distintas configuraciones de la volatilidad son de la siguiente manera

- (i) Volatilidad constante.
- (ii) Volatilidad estocástica.
- (iii) Volatilidad estocástica estacionaria.
- (iv) Volatilidad estocástica con colas pesadas.
- (v) Volatilidad estocástica y parámetros variando en el tiempo.
- (vi) Autorregresivo generalizado condicional heterocedástico.

Las variables que pronostican son: Crecimiento del producto interno bruto, tasa de desempleo, inflación en el deflactor del producto interno bruto y la tasa de las letras del Tesoro de los Estados Unidos a 3 meses.

Chiu et al. [20] propone un modelo autorregresivo vectorial con volatilidad estocástica y errores  $t$  de Student para investigar el ajuste dentro de la muestra y el rendimiento de pronóstico fuera

de la muestra. Calcula la verosimilitud marginal con un filtro de partículas y compara el ajuste dentro de la muestra y el rendimiento de pronóstico fuera de la muestra con otros tres modelos, a saber, un modelo autorregresivo vectorial gaussiano, un modelo autorregresivo vectorial con error  $t$  de Student y un modelo autorregresivo vectorial con volatilidad estocástica. Las variables que usan son: Crecimiento de producción industrial, tasa de inflación, tasa de interés a corto plazo y el índice SP500 de Estados Unidos. El modelo que proponen es de la siguiente manera

$$Y_t = V + A_1 Y_{t-1} + \dots + A_k Y_{t-k} + \Sigma_t^{1/2} \varepsilon_t, \quad \varepsilon_t \sim N(0, I_p), \quad (4.21)$$

$$\Sigma_t = B^{-1} H_t (B^{-1})^T, \quad (4.22)$$

$$H_t = \text{diag} \left( \frac{h_{1t}^2}{\lambda_{1,t}}, \dots, \frac{h_{pt}^2}{\lambda_{p,t}} \right), \quad (4.23)$$

$$\ln(h_{it}) = \ln(h_{i,t-1}) + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma_i), \quad (4.24)$$

donde  $Y_t$  son vectores observados de dimensión  $p \times 1$ ,  $V$  es un vector intersección de dimensión  $p \times 1$ ,  $A_i$ ,  $i = 1, \dots, k$  son matrices de coeficientes de dimensión  $p \times p$ .  $B$  es una matriz triangular inferior. Chiu et al. [20] comentan que en su modelo propuesto la volatilidad estocástica pretende capturar la posible heteroscedasticidad de los shocks y las posibles no linealidades en las relaciones no dinámicas de las variables del modelo, que están relacionadas con los cambios de baja frecuencia en la volatilidad. La introducción de la  $t$ -student en la estructura del shock pretende capturar cambios de alta frecuencia en la volatilidad que a menudo son de magnitudes extremas, por lo que potencialmente brindan un tratamiento de valores atípicos y eventos extremos.

Mumtaz [68] introduce un modelo autorregresivo vectorial con volatilidad estocástica en media en el que correlaciona los choques de las ecuaciones de volatilidad y las ecuaciones de observación. El modelo es de la siguiente manera

$$Y_t = V_1 + \sum_{j=1}^k A_j Y_{t-j} + \sum_{i=1}^{k_1} D_i h_{t-i} + H_t^{1/2} \varepsilon_t, \quad (4.25)$$

$$h_t = V_2 + \Phi h_{t-1} + \sum_{i=1}^{k_1} G_i Y_{t-i} + \Sigma_{\eta\eta}^{1/2} \eta_t, \quad (4.26)$$

donde  $Y_t$  son variables endógenas de dimensión  $p \times 1$ ,  $h_t = (h_{1t}, \dots, h_{pt})^T$ .  $V_2$  es un término intercepto de dimensión  $p \times 1$ ,  $\Phi$  y  $G_i$  son matrices de coeficientes de dimensión  $p \times p$ ,  $\Sigma_{\eta\eta} =$

$\text{diag}(\sigma_{11}, \dots, \sigma_{pp})$ ,  $A_i$  y  $D_i$  son matrices de coeficientes de dimensión  $p \times p$  y  $V_2$  es un término intercepto de dimensión  $p \times 1$ ,

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim N_{2p}(0, \Sigma), \quad \Sigma = \begin{pmatrix} \Sigma_\eta & \Sigma_{\eta_t \varepsilon_t}^T \\ \Sigma_{\varepsilon_t \eta_t} & \Sigma_\varepsilon \end{pmatrix}. \quad (4.27)$$

Los elementos de la diagonal de  $\Sigma$  son igual a la matriz de covarianza variando en el tiempo de los residuos de forma reducida es de la siguiente manera

$$\Sigma_t = \begin{bmatrix} \Sigma_{\eta\eta}^{1/2} & 0 \\ 0 & H_t^{1/2} \end{bmatrix} \begin{bmatrix} \Sigma_\eta & \Sigma_{\eta_t \varepsilon_t}^T \\ \Sigma_{\varepsilon_t \eta_t} & \Sigma_\varepsilon \end{bmatrix} \begin{bmatrix} \Sigma_{\eta\eta}^{1/2} & 0 \\ 0 & H_t^{1/2} \end{bmatrix}. \quad (4.28)$$

Note que si  $k_1 = k_2 = 0$  entonces el modelo (4.25)-(4.26) se convierte en un modelo autorregresivo vectorial con volatilidad estocástica y choques correlacionados de la siguiente manera

$$Y_t = V_1 + \sum_{i=1}^k A_i Y_{t-i} + H_t \varepsilon_t, \quad (4.29)$$

$$h_t = V_2 + \Phi h_{t-1} + \Sigma_{\eta\eta}^{1/2} \eta_t. \quad (4.30)$$

A diferencia del efecto de apalancamiento en el que existe una correlación negativa entre los retornos actuales y la volatilidad futura, el modelo (4.25)-(4.26) tiene una correlación entre los retornos actuales y la volatilidad actual.

Mumtaz [69] generaliza el modelo (4.25)-(4.26) de Mumtaz [68] modificando la matriz covarianza  $\Sigma$  de (4.27) de la siguiente manera

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & 1 & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & 1 \end{bmatrix}. \quad (4.31)$$

Ding et al. [30] extienden el modelo autorregresivo vectorial con parámetros variando en el tiempo y volatilidad estocástica de Primiceri [78] introduciendo volatilidades estocásticas variables en el tiempo basadas en los shocks de Omori et al [75] y Nakajima [70]. Los shocks son modelados de la

siguiente manera

$$\beta_{t+1} = \beta_t + W_t, \quad (4.32)$$

$$\lambda_{t+1} = \lambda_t + \varrho_t, \quad (4.33)$$

$$h_{t+1} = h_t + \eta_t. \quad (4.34)$$

Ding et al [30] comentan que su modelo tiene la ventaja de que refleja tanto la respuesta impulso en cada punto en el tiempo como las relaciones variables en el tiempo no lineales simultaneas entre las variables durante todo el período de la muestra. Ding et al [30] usan su modelo para estudiar los impactos de las incertidumbres financieras y geopolíticas en los mercados de productos básicos en un marco variable en el tiempo.

Triantafyllopoulos [99] propuso un modelo autorregresivo vectorial con parámetros variando en el tiempo y volatilidad estocástica con el propósito de pronosticar series de tiempo multivariadas. El modelo que propone es de la siguiente manera

$$y_t^T = [1, y_{t-1}^T, \dots, y_{t-1t}^T] \begin{bmatrix} v_t^T \\ A_{1t}^T \\ \vdots \\ A_{kt}^T \end{bmatrix} + \varepsilon_t^T = Y_t^T A_t + \varepsilon_t^T, \quad (4.35)$$

donde  $Y_t$  son variables observadas de dimensión  $p \times 1$ ,  $v_t$  son términos de intersección de dimensión  $p \times 1$ ,  $A_{it}$  son matrices de coeficientes de dimensión  $p \times p$ .  $A_t$  es modelado de la siguiente manera

$$A_t = A_{t-1} + \Omega_t, \quad \Omega_t \sim N_{(kp+1) \times p}(0, W_t, \Sigma_t), \quad (4.36)$$

donde  $\Omega_t$  es independiente de  $\Omega_s$ ,  $\forall t \neq s$  y, además  $\Omega_t$  es independiente de  $\varepsilon_s$ ,  $\forall t, s$ . Para modelar la evolución de  $\Sigma_t$ , adopta una versión ligeramente modificada del modelo de Uhlig [101], basado en la convolución de Wishart y la distribución beta multivariada singular para permitir la inferencia bayesiana totalmente conjugada. La evolución de  $\Sigma_t$  es de la siguiente manera

$$(\Sigma_t)^{-1} = c \left( \Sigma_{t-1}^{-1/2} \right)^T B_t \Sigma_{t-1}^{-1/2}, \quad (4.37)$$



donde  $\Sigma_{t-1}^{-1/2}$  denota la matriz triangular superior de la descomposición de Choleski de  $\Sigma_{t-1}^{-1}$ ,  $B_t$  es independiente de  $\Sigma_{t-1}^{-1}$  y sigue una distribución beta multivariada y singular con grados de libertad  $n_1 = \frac{(n+p-1)}{2}$  y  $n_2 = \frac{1}{2}$ . La matriz de covarianza  $W_t$  es de dimensión  $(kp+1) \times (kp+1)$  y es especificada usando factores de descuento.

Para encontrar el orden  $K$  primero define la distribución de pronóstico un paso hacia adelante de la siguiente manera

$$P_r(Y_{t+1}|Y^t, \text{modelo}(i)) = \int_{\Sigma_t} \Pr(y_{t+1}|\Sigma_t, Y^t, \text{modelo}(i)) \Pr(\Sigma_t|Y^t, \text{modelo}(i)) d\Sigma_t, \quad (4.38)$$

donde  $\text{modelo}(i)$  es un modelo estimado como en (4.35) con orden  $K_i$  y  $Y^t = (y_1, \dots, y_t)$  es el conjunto de información hasta el tiempo  $t$ . Luego  $y_{t+1}|Y^t, \text{modelo}(i) \sim t_{p \times 1}(n, y_{it}(1), Q_{it}(1))$  es una distribución  $t$  de Student multivariada.  $y_{it}(1)$  es el pronóstico un paso adelante medio de  $y_{t+1}$  bajo el modelo  $i$ , y  $Q_{it}(t)$  es la matriz escala de la distribución  $t$ . Después de encontrar la distribución de pronóstico un paso hacia adelante procede a calcular el factor de Bayes de los diferentes modelos de la siguiente manera

$$\begin{aligned} H_t(1) &= \frac{\Pr(y_{t+1}|Y^t \text{ modelo}(i))}{\Pr(y_{t+1}|Y^t \text{ modelo}(j))} \\ &= \left( \frac{|Q_{jt}(1)|}{|Q_{it}(1)|} \right)^{1/2} \left( \frac{\theta n + e_{i,t+1}^T (Q_{it}(1))^{-1} e_{i,t+1}}{\theta n + e_{j,t+1}^T (Q_j(1))^{-1} e_{j,t+1}} \right)^{-\frac{(\theta n + p)}{2}}, \end{aligned} \quad (4.39)$$

donde  $0 < \theta < 1$  y  $e_{i,t+1}$  es el error de pronóstico un paso hacia adelante calculado con el filtro de Kalman.

Si  $H_t(1) > 1$  el factor de Bayes puede indicar preferencia por el modelo  $i$  mientras que si  $H_t(1) < 1$  la preferencia es para el modelo  $j$ . En el caso que  $H_t(1) = 1$  los dos modelos son equivalentes. Triantafyllopoulos [99] comenta que la ventaja de (4.39) es que se puede aplicar secuencialmente y así indicar la preferencia de un modelo en cada punto del tiempo.

Para pronosticar  $h$  pasos hacia adelante, donde  $h$  es un entero encuentra que la distribución de pronóstico  $h$  pasos hacia adelante es aproximadamente de la siguiente manera

$$y_{t+h}|Y^t \sim t_{p \times 1} \left( \theta n, m_t^T \hat{Y}_{t+h}, \left( \hat{Y}_{t+h}^T R_t(h) \hat{Y}_{t+h+1} \right) c^{-1} S_t \right), \quad (4.40)$$

donde

$$\begin{aligned}
 Y_{t+h}^T &= \begin{cases} [1, y_t^T(h-1), \dots, y_t^T(h-k)], & \text{si } h > k, \\ [1, y_t^T(h-1), \dots, y_t^T(1), y_t, \dots, y_{t+h-k}^T], & \text{si } h \leq k, \end{cases} \\
 R_t(h) &= P_t + \sum_{i=1}^h W_{t+i}, \quad P_t = R_t - K_t K_t^T / Q_t, \quad R_t = P_{t-1} + W_t, \\
 Q_t &= A_t^T R_t A_t + 1, \quad m_t = m_{t-1} + K_t e_t^T, \\
 K_t &= R_t A_t / Q_t \quad \text{es la ganancia de Kalman,} \\
 e_t &= y_t - m_{t-1}^T A_t, \\
 S_t &= c^{-1} S_{t-1} + e_t e_t^T / Q_t.
 \end{aligned}$$

El pronóstico medio  $h$  pasos adelante de  $y_{t+h}$  es dado por

$$y_t(h) = m_t^T \hat{F}_{t+h}. \quad (4.41)$$

La matriz de covarianza de pronóstico  $h$  pasos adelante es dada por

$$Q_t(h) = \frac{(\hat{Y}_{t+h}^T R_t(h) \hat{Y}_{t+h} + 1) (1 - \theta) c^{-1}}{3\theta - 2} S_t, \quad (4.42)$$

donde  $\theta > 2/3$ . El error de pronóstico  $h$  pasos adelante es aproximadamente dado por

$$e_t(h) | Y^t = y_t - y_t(h) | Y^t \sim t_{p \times 1} \left( \theta n, 0, \frac{(3\theta - 2)}{(1 - \theta)} Q_t(h) \right). \quad (4.43)$$

## 4.2. Modelo autorregresivo vectorial con volatilidad estocástica y efecto de apalancamiento

En esta sección se propone un modelo autorregresivo vectorial con volatilidad estocástica multivariada con efecto de apalancamiento cruzado. La volatilidad modelada es la propuesta por Ishihara y Omori [53] en el que los retornos de hoy están correlacionados negativamente con la volatilidad de mañana. A este efecto se le conoce como efecto de apalancamiento cruzado. El modelo es de la

siguiente manera

$$y_t = v + A_1 y_{t-1} + \dots + A_k y_{t-k} + V_t^{1/2} \varepsilon_t, \quad t = 1, \dots, n, \quad (4.44)$$

$$\alpha_{t+1} = \Phi \alpha_t + \eta_t, \quad t = 1, \dots, n-1, \quad (4.45)$$

$$\alpha_1 \sim N_p(0, \Sigma_0), \quad (4.46)$$

$$V_t = \text{diag}(\exp(\alpha_{1t}), \dots, \exp(\alpha_{pt})), \quad (4.47)$$

$$\Phi = \text{diag}(\phi_1, \dots, \phi_p), \quad (4.48)$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{pmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\eta} \\ \Sigma_{\eta\varepsilon} & \Sigma_{\eta\eta} \end{pmatrix}, \quad (4.49)$$

$$\text{vec}(\Sigma_0) = (I_p^2 - \Phi \otimes \Phi)^{-1} \text{vec}(\Sigma_{\eta\eta}), \quad (4.50)$$

$\alpha_t = h_t - \mu$  es de dimensión  $p \times 1$ .  $h_t$  es el correspondiente vector de log volatilidad y  $\mu$  es el vector media. La ecuación (4.50) asegura la estacionariedad de (4.44). El vector  $v$  es un término intercepto de dimensión  $p \times 1$  mientras que  $A_i, i = 1, \dots, k$  son matrices de coeficientes de dimensión  $p \times p$  que miden la dependencia dinámica de  $y_t$ . La volatilidad estocástica pretende capturar la posible heteroscedasticidad de los shocks y las posibles no linealidades en las relaciones simultaneas entre las variables del modelo (ver Primiceri [78]). La sub-matriz de covarianza  $\Sigma_{\varepsilon\eta}$  se usa para calcular la posible correlación entre los retornos de hoy y la volatilidad de mañana.

Para encontrar la función verosimilitud del modelo de las ecuaciones (4.44)-(4.50) se hace una leve modificación a la verosimilitud de Ishihara y Omori [53] reescribiendo el modelo (4.44) de la siguiente manera

$$y_t = (Y_t^T \otimes I_p) \beta + V_t^{1/2} \varepsilon_t, \quad (4.51)$$

donde  $Y_t = [1, y_{t-1}^T, \dots, y_{t-k}^T]^T$  es de dimensión  $(kp+1) \times 1$  y  $\beta = \text{vec}(v, A_1, \dots, A_k)$  es de dimensión  $(kp^2 + p) \times 1$ .

Defina  $\theta = (\phi, \Sigma, \beta)$ ,  $\phi = (\phi_1, \dots, \phi_p)^T$ ,  $\alpha = (\alpha_1^T, \dots, \alpha_n^T)^T$ ,  $Y^n = (y_1, \dots, y_n)$ ,  $Y^k = (y_{-k+1}, \dots, y_0)$  y  $1_p = [1, \dots, 1]^T$ , entonces por (4.49) y (4.51) se tiene que

$$y_n | \alpha_n, y_{n-1}, \dots, y_{n-k}, \theta = (Y_n^T \otimes I_p) \beta + V_n^{1/2} \varepsilon_t | \alpha_n, y_{n-1}, \dots, y_{n-k}, \theta \quad (4.52)$$

$$\sim N \left( (Y_n^T \otimes I_p) \beta, V_n^{1/2} \Sigma_{\varepsilon\varepsilon} V_n^{1/2} \right),$$

$$y_t, \alpha_{t+1} | \alpha_t, y_{t-1}, \dots, y_{t-k}, \theta = (Y_t^T \otimes I_p) \beta + V_t^{1/2} \varepsilon_t, \Phi \alpha_t + \eta_t | \alpha_t, y_{t-1}, \dots, y_{t-k}, \theta \quad (4.53)$$

$$\sim N \left( \begin{bmatrix} (Y_t^T \otimes I_p) \beta \\ \Phi \alpha_t \end{bmatrix}, \begin{bmatrix} V_t^{1/2} \Sigma_{\varepsilon\varepsilon} V_t^{1/2} & V_t^{1/2} \Sigma_{\varepsilon\eta} \\ \Sigma_{\eta\varepsilon} V_t^{1/2} & \Sigma_{\eta\eta} \end{bmatrix} \right), \quad t = 1, \dots, n-1$$

Por (4.46), (4.52) - (4.53) la función verosimilitud del modelo (4.51) es dada por

$$f(\alpha, Y^n | \theta, Y^k) = f(y_n | \alpha_n, y_{n-1}, \dots, y_{n-k}, \theta) \prod_{t=1}^{n-1} f(y_t, \alpha_{t+1} | \alpha_t, y_{t-1}, \dots, y_{t-k}, \theta) f(\alpha_1 | \theta) \quad (4.54)$$

$$\propto |V_n^{1/2} \Sigma_{\varepsilon\varepsilon} V_n^{1/2}|^{-1/2} \prod_{t=1}^{n-1} \left| V_t^{1/2} \Sigma_{\varepsilon\varepsilon} V_t^{1/2} - V_t^{1/2} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} V_t^{1/2} \right|^{-1/2} |\Sigma_0|^{-1/2}$$

$$\times \exp \left\{ -\frac{1}{2} (y_n - (Y_n^T \otimes I_p) \beta)^T (V_n^{1/2} \Sigma_{\varepsilon\varepsilon} V_n^{1/2})^{-1} (y_n - (Y_n^T \otimes I_p) \beta) \right\}$$

$$\times \exp \left\{ -\frac{1}{2} \sum_{t=1}^{n-1} \begin{pmatrix} y_t - (Y_t^T \otimes I_p) \beta \\ \alpha_{t+1} - \Phi \alpha_t \end{pmatrix}^T \begin{bmatrix} V_t^{1/2} \Sigma_{\varepsilon\varepsilon} V_t^{1/2} & V_t^{1/2} \Sigma_{\varepsilon\eta} \\ \Sigma_{\eta\varepsilon} V_t^{1/2} & \Sigma_{\eta\eta} \end{bmatrix}^{-1} \begin{pmatrix} y_t - (Y_t^T \otimes I_p) \beta \\ \alpha_{t+1} - \Phi \alpha_{t+1} \end{pmatrix} \right\}$$

$$\times \exp \left\{ -\frac{1}{2} \alpha_1^T \Sigma_0^{-1} \alpha_1 \right\}$$

$$\propto |\Sigma_0|^{-1/2} |\Sigma|^{-(\frac{n-1}{2})} |\Sigma_{\varepsilon\varepsilon}|^{-1/2} \exp \left\{ \sum_{t=1}^n \ell_t - \frac{1}{2} \alpha_1^T \Sigma_0^{-1} \alpha_1 - \frac{1}{2} \sum_{t=1}^{n-1} (\alpha_{t+1} - \Phi \alpha_t)^T \Sigma_{\eta\eta}^{-1} (\alpha_{t+1} - \Phi \alpha_t) \right\},$$

donde

$$\ell_t = \text{const} - \frac{1}{2} 1_p^T \alpha_T - \frac{1}{2} (y_t - ((Y_t^T \otimes I_p) + \mu_t))^T \Sigma_t^{-1} (y_t - ((Y_t^T \otimes I_p) + \mu_t)), \quad (4.55)$$

$$\mu_t = V_t^{1/2} m_t, \quad (4.56)$$

$$\Sigma_t = V_t^{1/2} S_t V_t^{1/2}, \quad (4.57)$$

$$m_t = \begin{cases} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} (\alpha_{t+1} - \Phi \alpha_t), & t < n, \\ 0, & t = n, \end{cases} \quad (4.58)$$

$$S_t = \begin{cases} \Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon}, & t < n, \\ \Sigma_{\varepsilon\varepsilon}, & t = n. \end{cases} \quad (4.59)$$

Esto es posible debido a a siguiente igualdad

$$\begin{aligned}
 & \begin{bmatrix} V_t^{1/2} \Sigma_{\varepsilon\varepsilon} V_t^{1/2} & V_t^{1/2} \Sigma_{\varepsilon\eta} \\ \Sigma_{\eta\varepsilon} V_t^{1/2} & \Sigma_{\eta\eta} \end{bmatrix}^{-1} \\
 &= \begin{bmatrix} V_t^{-1/2} (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon})^{-1} V_t^{-1/2} & -V_t^{-1/2} (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon})^{-1} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \\ -\Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon})^{-1} V_t^{-1/2} & \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon} (\Sigma_{\varepsilon\varepsilon} - \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} \Sigma_{\eta\varepsilon})^{-1} \Sigma_{\varepsilon\eta} \Sigma_{\eta\eta}^{-1} + \Sigma_{\eta\eta}^{-1} \end{bmatrix}^{-1}. \quad (4.60)
 \end{aligned}$$

### 4.3. Método de Estimación

Para estimar los parámetros se usa un algoritmo MCMC de seis bloques que es dado por

1. Inicializar  $\alpha, \phi, \Sigma$  y  $\beta$ .
2. Generar  $\alpha | \phi, \Sigma, \beta, Y^n, Y^k$ .
3. Generar  $\phi | \Sigma, \beta, \alpha, Y^n, Y^k$ .
4. Generar  $\Sigma | \beta, \alpha, \phi, Y^n, Y^k$ .
5. Generar  $\beta | \alpha, \phi, \Sigma, Y^n, Y^k$ .
6. Ir a 2.

#### Generación de $\alpha, \phi$ y $\Sigma$

Para generar  $\alpha, \phi$  y  $\Sigma$  se aplica el método de Ishihara y Omori [53] expuesto en la sección 3.4 sustituyendo  $y_t$  por  $y_t^* = y_t - (Y_t^T \otimes I_p) \hat{\beta}$ .

### Generación de $\beta$

La distribución priori  $f(\beta)$  de  $\beta$  es la de Litterman de las ecuaciones (2.46)-(2.47) del capítulo 2.

La función de densidad de probabilidad posterior de  $\beta$  es dada por

$$\begin{aligned}
 f(\beta|\alpha, \phi, \Sigma, Y^n, Y^k) &\propto f(\alpha, Y^n|\phi, \Sigma, \beta, Y^k) f(\beta) & (4.61) \\
 &\propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^n (y_t - ((Y_t^T \otimes I_p)\beta + \mu_t))^T \Sigma_t^{-1} (y_t - ((Y_t^T \otimes I_p)\beta + \mu_t)) \right\} f(\beta) \\
 &\propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^n \left( -(y_t - \mu_t)^T \Sigma_t^{-1} (Y_t^T \otimes I_p)^T \beta - \beta^T (Y_t \otimes I_p) \Sigma_t^{-1} (y_t - \mu_t) \right. \right. \\
 &\quad \left. \left. + \beta^T (Y_t \otimes I_p) \Sigma_t^{-1} (Y_t^T \otimes I_p) \beta \right) \right\} f(\beta) \\
 &\propto \exp \left\{ -\frac{1}{2} \left( - \left[ \sum_{t=1}^n (y_t - \mu_t)^T \Sigma_t^{-1} (Y_t^T \otimes I_p) \right] \beta \right. \right. \\
 &\quad \left. \left. - \left[ \sum_{t=1}^n (Y_t \otimes I_p) \Sigma_t^{-1} (y_t - \mu_t) \right] \beta + \beta \left[ \sum_{t=1}^n (Y_t \otimes I_p) \Sigma_t^{-1} (Y_t^T \otimes I_p) \right] \right) \right\} f(\beta) \\
 &\propto \exp \left\{ -\frac{1}{2} (\beta - \mu_\beta^*)^T \Sigma_\beta^{*-1} (\beta - \mu_\beta^*) \right\},
 \end{aligned}$$

donde

$$\begin{aligned}
 \Sigma_\beta^* &= \left[ \sum_{t=1}^n (Y_t \otimes I_p)^T \Sigma_t^{-1} (Y_t^T \otimes I_p) + \Sigma_\beta^{-1} \right]^{-1}, \\
 \mu_\beta^* &= \Sigma_\beta^* \left[ \sum_{t=1}^n (Y_t \otimes I_p) \Sigma_t^{-1} (y_t - \mu_t) \right].
 \end{aligned}$$

Se usa el algoritmo de muestreador de Gibbs para generar una muestra de  $\beta^* \sim N(\mu_\beta^*, \Sigma_\beta^*)$ .

$\mu_\beta$  es configurada igual a cero.

## 4.4. Selección del orden VAR-MSV

Para escoger el mejor orden VAR-MSV se aplican los siguientes pasos:

1. Suponiendo que se sabe que el orden VAR-MSV no puede exceder un entero  $K_1$  se procede a estimar los modelos VAR-MSV comenzando desde 0 hasta  $K_1$  y se almacenan sus parámetros estimados  $\theta_0, \theta_1, \dots, \theta_{K_1}$ , donde  $\theta_i$  son los parámetros estimados del modelo  $i$ .
2. Para cada modelo se calcula el DIC de la sección 3.5 sustituyendo  $y_t$  por  $y_t^* = y_t - (Y_t^T \otimes I_p) \hat{\beta}^i$ , donde  $\hat{\beta}^i$  son las matrices de coeficientes estimadas del modelo  $i$ .
3. Se escoge el modelo que tenga la menor medida DIC.

## 4.5. Ejemplo

En esta sección se aplica el método propuesto usando datos simulados para observar que tan bien se estiman los parámetros. Ishihara y Omori [53] generan sus resultados computacionales usando Ox [34]. Para este trabajo, nuestros resultados computacionales se generaron usando el software libre R. Se usaron las librerías Rcpp y RcppArmadillo para implementar los algoritmos que involucran álgebra lineal con el fin de encontrar un equilibrio entre velocidad y facilidad de uso.

Los datos son simulados usando el modelo (4.44)-(4.50). El orden del modelo MSV-VAR que se usa para simular datos es  $k = 6$  y la dimensión es  $p = 2$ . La matriz de covarianza  $\Sigma$  es tomada de Ishihara y Omori [53]. Los parámetros  $\Sigma$ ,  $\Phi$  y el término de intersección  $v$  así como las matrices de coeficientes son dados por

$$\Sigma = \begin{bmatrix} 1.44 & 0.864 & -0.096 & -0.072 \\ 0.864 & 1.44 & -0.072 & -0.096 \\ -0.096 & -0.072 & 0.04 & 0.028 \\ -0.072 & -0.096 & 0.028 & 0.04 \end{bmatrix},$$

$$\begin{aligned} \Phi &= \begin{bmatrix} 0.97 & 0 \\ 0 & 0.97 \end{bmatrix}, & v &= \begin{bmatrix} 1 & 0.5 \end{bmatrix}^T \\ A_1 &= \begin{bmatrix} 0.1 & -0.3 \\ 0.2 & -0.4 \end{bmatrix}, & A_2 &= \begin{bmatrix} -0.2 & -0.3 \\ 0.3 & 0.4 \end{bmatrix}, \\ A_3 &= \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}, & A_4 &= \begin{bmatrix} -0.2 & -0.3 \\ -0.2 & -0.4 \end{bmatrix}, \\ A_5 &= \begin{bmatrix} 0.09 & 0 \\ 0 & 0.09 \end{bmatrix}, & A_6 &= \begin{bmatrix} 0.07 & 0 \\ 0 & 0.07 \end{bmatrix} \end{aligned}$$

Ishihara y Omori [53] usan la matriz  $\Sigma$  para ilustrar el efecto de apalancamiento cruzado ya que

$$\begin{aligned} \rho_{i,\varepsilon\eta} &= \frac{\text{Cov}(\varepsilon_{it}, \eta_{it})}{\sqrt{\text{Var}(\varepsilon_{it}) \text{Var}(\eta_{it})}} = \frac{-0.096}{\sqrt{(1.44)(0.04)}} = -0.4, \\ \rho_{ij,\varepsilon\eta} &= \frac{\text{Cov}(\varepsilon_{it}, \eta_{jt})}{\sqrt{\text{Var}(\varepsilon_{it}) \text{Var}(\eta_{jt})}} = \frac{-0.072}{\sqrt{(1.44)(0.04)}} = -0.3. \end{aligned}$$

Los hiperparámetros de las distribuciones priori  $\phi$  y  $\Sigma$  son tomados de Ishihara y Omori [53] donde  $a_i = 20, i = 1, \dots, p; b_i = 1.5, i = 1, \dots, p. n_0 = 10$ , mientras que para  $R_0$  se usó una matriz con elementos iguales a 1000 en las diagonales y 0 en los elementos fuera de la diagonal. Ishihara y Omori [53] usan  $R_0 = (10\Sigma^*)^{-1}$  donde  $\Sigma^*$  es la matriz de covarianza verdadera. Para los hiperparámetros de la priori de Litterman se configura  $\mu_\beta = 0$  mientras que  $\theta=0.99$  y  $\lambda = 10$ .

Se generan 4,000 observaciones donde las primeras 6 observaciones son tomadas como valores premuestrales mientras que los últimos 3,994 serán tomados como la muestra. Usando el método de la sección 3.3 se generan 120,000 muestras y se descartan las primeras 20,000. Para el parámetro de ajuste se usa  $K = 0.05n$  y para evitar que el método se atasque firmemente debido al rechazo excesivo se sigue la sugerencia de Sheppard y Pitt [85] incrementando  $K$  unas pocas iteraciones en intervalos regulares. La tablas 4.1-4.2 muestran la media posterior, desviación estándar e intervalos de credibilidad del 95 % de los elementos de la matriz  $\Phi$ , la matriz de covarianza  $\Sigma$ , el termino de intercesión  $v$  y las matrices de coeficientes  $A_1, A_2, \dots, A_6$ . Todos los intervalos de credibilidad con un 95 % de confianza contienen los parámetros verdaderos a excepción de  $A_{11}^1$  y  $A_{22}^1$ . Aumentando los intervalos de credibilidad a 96.2 % todos los parámetros están contenidos.



	Parámetro	Media	Desviación Estándar	Intervalo de Credibilidad del 95 %
$\phi_1$	0.97	0.9689	0.0052	[0.9578, 0.9785]
$\phi_2$	0.97	0.9721	0.0046	[0.9622, 0.9805]
$\sigma_{11}$	1.44	1.2443	0.1232	[1.0289, 1.5231]
$\sigma_{12}$	0.864	0.7446	0.0741	[0.6124, 0.9066]
$\sigma_{13}$	-0.096	-0.0779	0.0136	[-0.1062, -0.0525]
$\sigma_{14}$	-0.072	-0.0527	0.0129	[-0.0789, -0.0280]
$\sigma_{21}$	0.864	0.7446	0.0741	[0.6124, 0.9066]
$\sigma_{22}$	1.44	1.1668	0.1295	[0.9362, 1.4464]
$\sigma_{23}$	-0.072	-0.0647	0.0131	[-0.0919, -0.0407]
$\sigma_{24}$	-0.096	-0.0771	0.0135	[-0.1047, -0.0517]
$\sigma_{31}$	-0.096	-0.0779	0.0136	[-0.1062, -0.0525]
$\sigma_{32}$	-0.072	-0.0647	0.0131	[-0.0919, -0.0407]
$\sigma_{33}$	0.04	0.0334	0.0054	[0.0237, 0.0450]
$\sigma_{34}$	0.028	0.0244	0.0038	[0.0174, 0.0326]
$\sigma_{41}$	-0.072	-0.0527	0.0129	[-0.0789, -0.0281]
$\sigma_{42}$	-0.096	-0.0771	0.0135	[-0.1047, -0.0517]
$\sigma_{43}$	0.028	0.0244	0.0038	[0.0174, 0.0326]
$\sigma_{44}$	0.04	0.0350	0.0053	[0.0257, 0.0462]

Tabla 4.1: Media posterior, Desviación estándar e Intervalos de credibilidad del 95 % de los elementos de la matriz  $\Phi$  y la matriz de covarianza  $\Sigma$ .

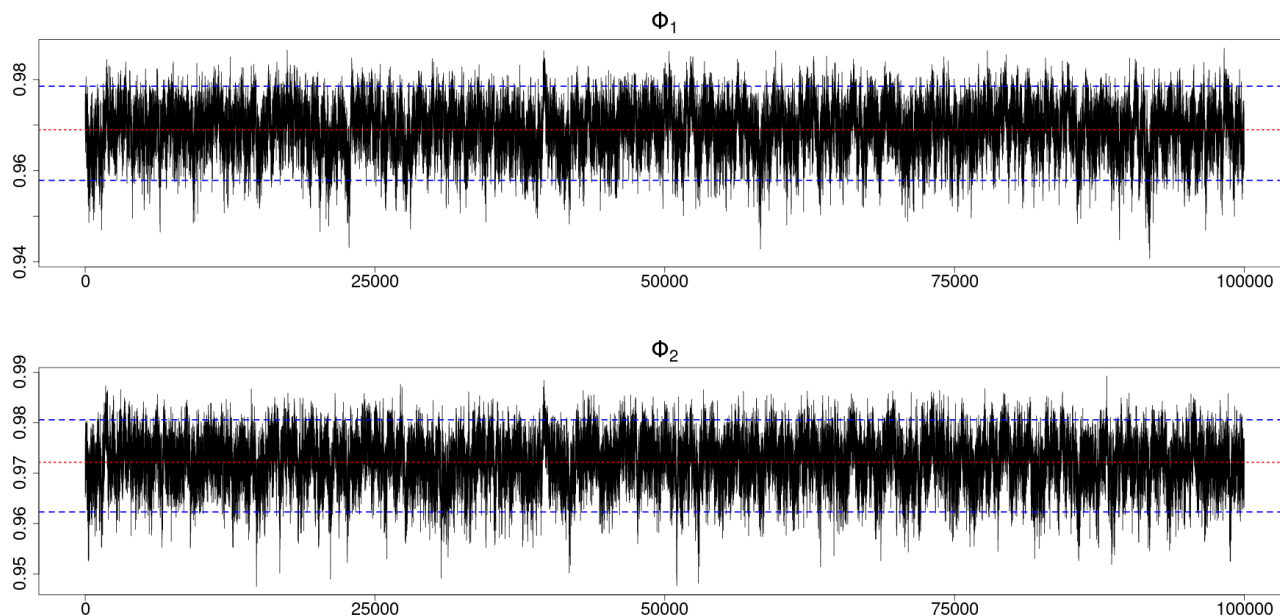


Figura 4.1: Gráfico de convergencia de las muestras de  $\phi_1$  y  $\phi_2$  respectivamente. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.

Fuente: Elaboración propia

	Parámetro	Media	Desviación Estándar	Intervalo de Credibilidad del 95 %
$v_1$	1	0.9852	0.0389	[0.9091, 1.0618]
$A_{11}^1$	0.1	0.1408	0.0198	[0.1022, 0.1795]
$A_{12}^1$	-0.3	-0.3029	0.0188	[-0.3401, -0.2660]
$A_{11}^2$	-0.2	-0.1889	0.0198	[-0.2281, -0.1501]
$A_{12}^2$	-0.3	-0.2801	0.0196	[-0.3187, -0.2415]
$A_{11}^3$	0.1	0.0982	0.0215	[0.0560, 0.1404]
$A_{12}^3$	0	0.0079	0.0218	[-0.0351, 0.0506]
$A_{11}^4$	-0.2	-0.1896	0.0209	[-0.2307, -0.1485]
$A_{12}^4$	-0.2	-0.2082	0.0216	[-0.2509, -0.1658]
$A_{11}^5$	0.09	0.0878	0.0190	[-0.0501, 0.1249]
$A_{12}^5$	0	-0.0044	0.0165	[-0.0370, 0.0281]
$A_{11}^6$	0.07	0.0546	0.0187	[0.0181, 0.0913]
$A_{12}^6$	0	0.0194	0.0151	[-0.0102, 0.0489]
$v_2$	0.5	0.4736	0.0343	[0.4064, 0.5410]
$A_{21}^1$	0.2	0.2442	0.0174	[0.2103, 0.2785]
$A_{22}^1$	-0.4	-0.4396	0.0197	[-0.4781, -0.4007]
$A_{21}^2$	0.3	0.3284	0.0174	[0.2941, 0.3627]
$A_{22}^2$	0.4	0.4033	0.0204	[0.3631, 0.4434]
$A_{21}^3$	0	0.0142	0.0196	[-0.0241, 0.0529]
$A_{22}^3$	0.1	0.1358	0.0217	[0.0931, 0.1784]
$A_{21}^4$	-0.3	-0.2800	0.0190	[-0.3172, -0.2424]
$A_{22}^4$	-0.4	-0.3938	0.0216	[-0.4362, -0.3514]
$A_{21}^5$	0	-0.0151	0.0181	[-0.0505, 0.0205]
$A_{22}^5$	0.09	0.0708	0.0159	[0.0397, 0.1021]
$A_{21}^6$	0	-0.0105	0.0178	[-0.0453, 0.0244]
$A_{22}^6$	0.07	0.0946	0.0146	[0.0658, 0.123]

Tabla 4.2: Media posterior, Desviación estándar e Intervalos de credibilidad del 95 % de los elementos del vector término intersección  $v$  y las matrices de coeficientes  $A_1, A_2, \dots, A_6$ .

Las figuras 4.1-4.4 muestran los gráficos de convergencia de los parámetros estimados. Las figuras 4.5-4.8 muestran los gráficos de las funciones de correlación de las muestras de los parámetros estimados. Los gráficos de autocorrelación de  $\Sigma$  y  $\Phi$  muestran la correlación usual que se ve en la literatura. Las funciones de autocorrelación de los elementos de las matrices de coeficientes VAR muestran poca correlación. Por otro lado, los gráficos de las funciones de autocorrelación de los términos de intersección alcanzan sus máximos en 0.24 para  $v_1$  y 0.249 para  $v_2$ , además decrecen lentamente mostrando la alta dependencia de valores pasados. Esto puede ser debido a la priori de Litterman ya que las varianzas priori son infinitas para  $v_1$  y  $v_2$  reflejando que no se tiene ninguna conjetura a priori para estos coeficientes. Los histogramas de las figuras 4.9-4.12 revelan que las cadenas de los parámetros estimados parecen tener distribución Gaussiana. Los diagnósticos de convergencia muestran que el enfoque bayesiano para estimar los parámetros es eficiente.

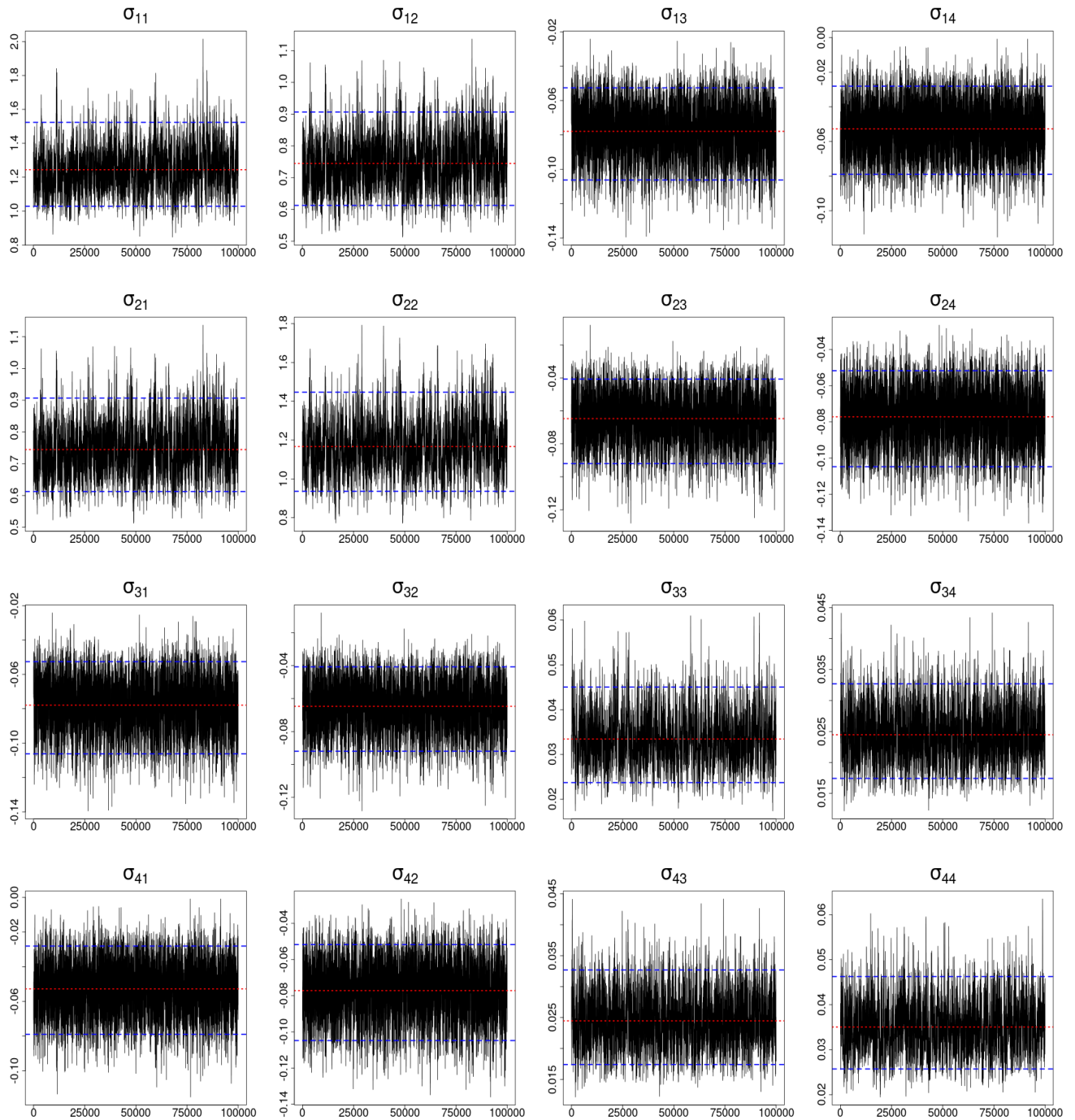


Figura 4.2: Gráfico de convergencia de las muestras de los elementos de  $\Sigma$ . La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.

Fuente: Elaboración propia

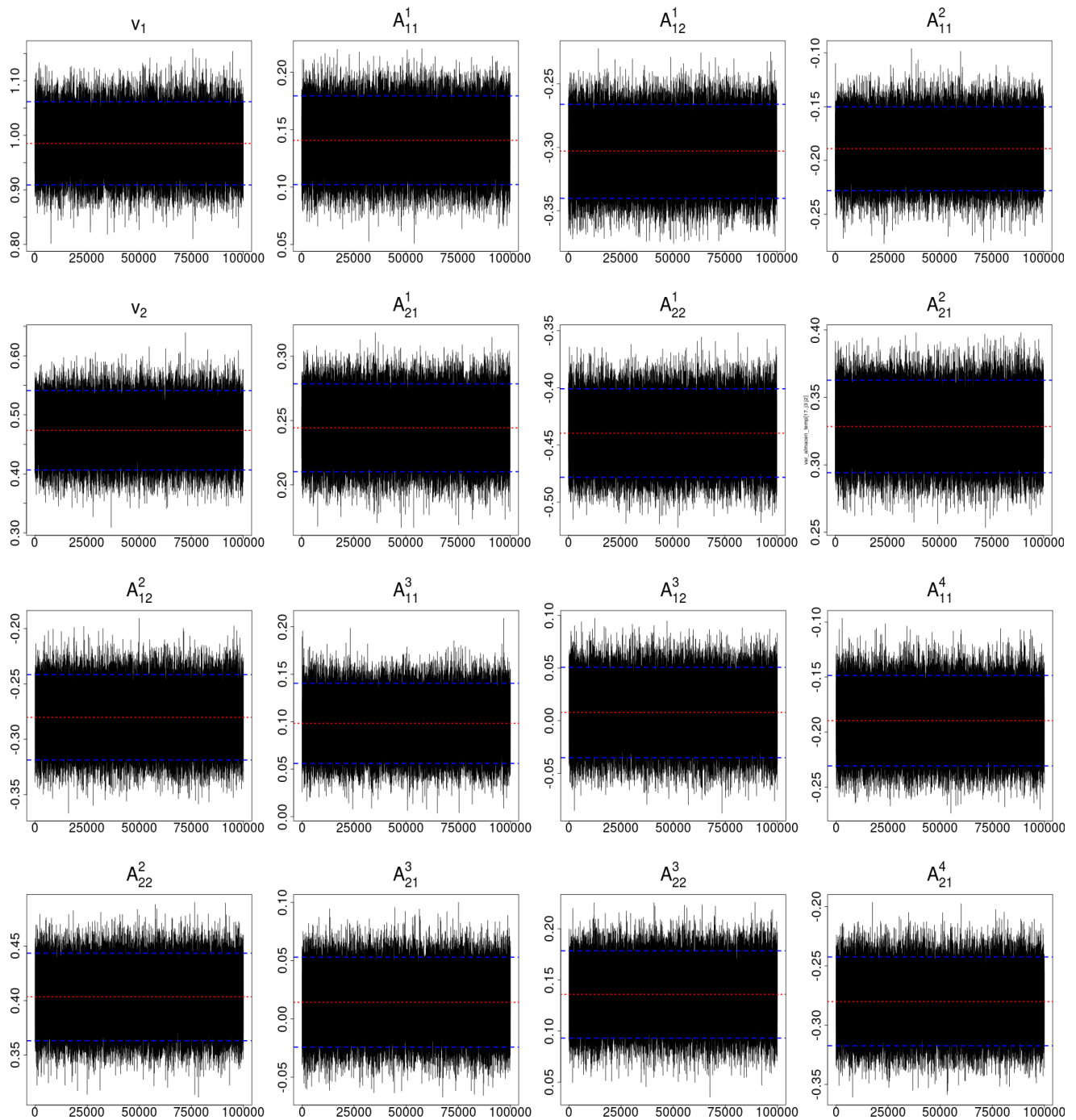


Figura 4.3: Gráfico de convergencia de las muestras de los elementos del vector termino de intercepción y las matrices de coeficientes VAR. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.

Fuente: Elaboración propia

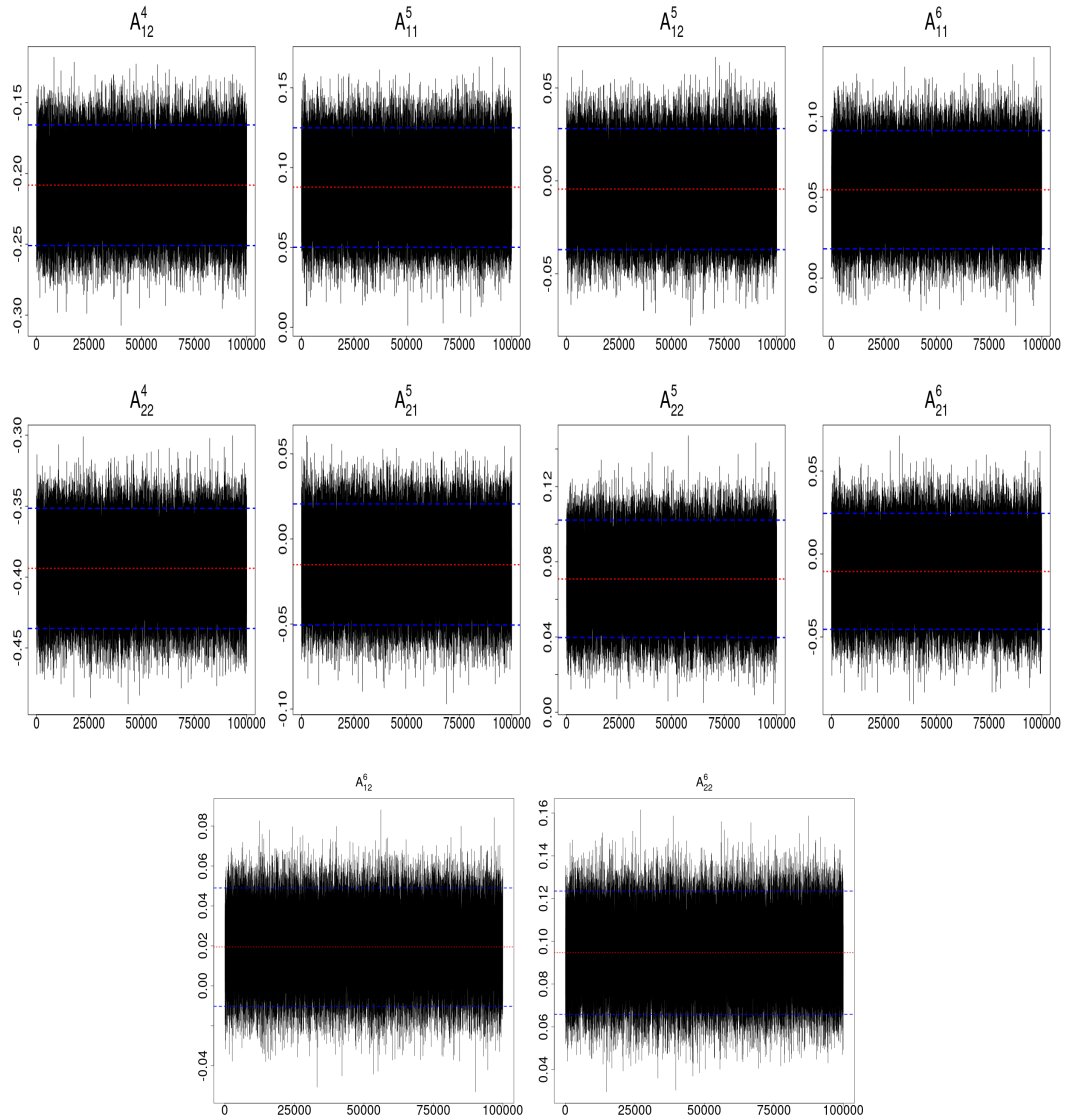


Figura 4.4: Gráfico de convergencia de las muestras de los elementos del vector termino de intercepción y las matrices de coeficientes VAR. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.

**Fuente:** Elaboración propia

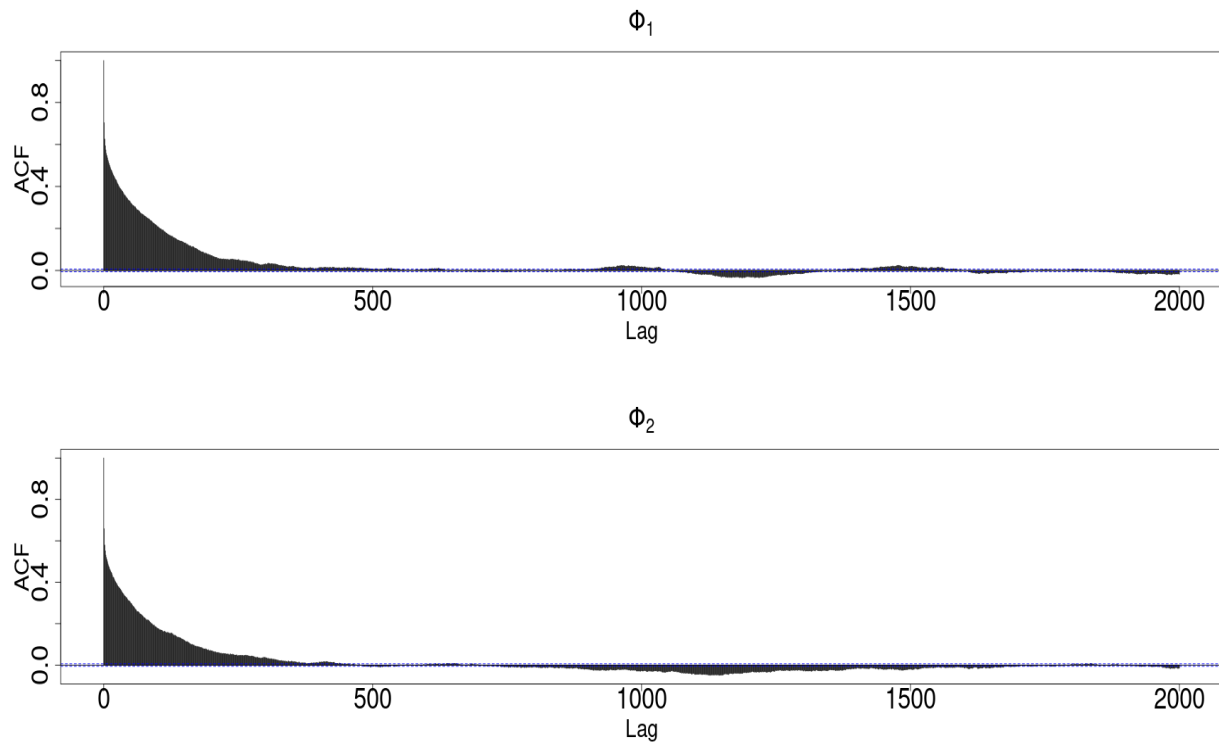


Figura 4.5: Función de autocorrelación de  $\phi_1$  y  $\phi_2$  respectivamente.  
Fuente: Elaboración propia

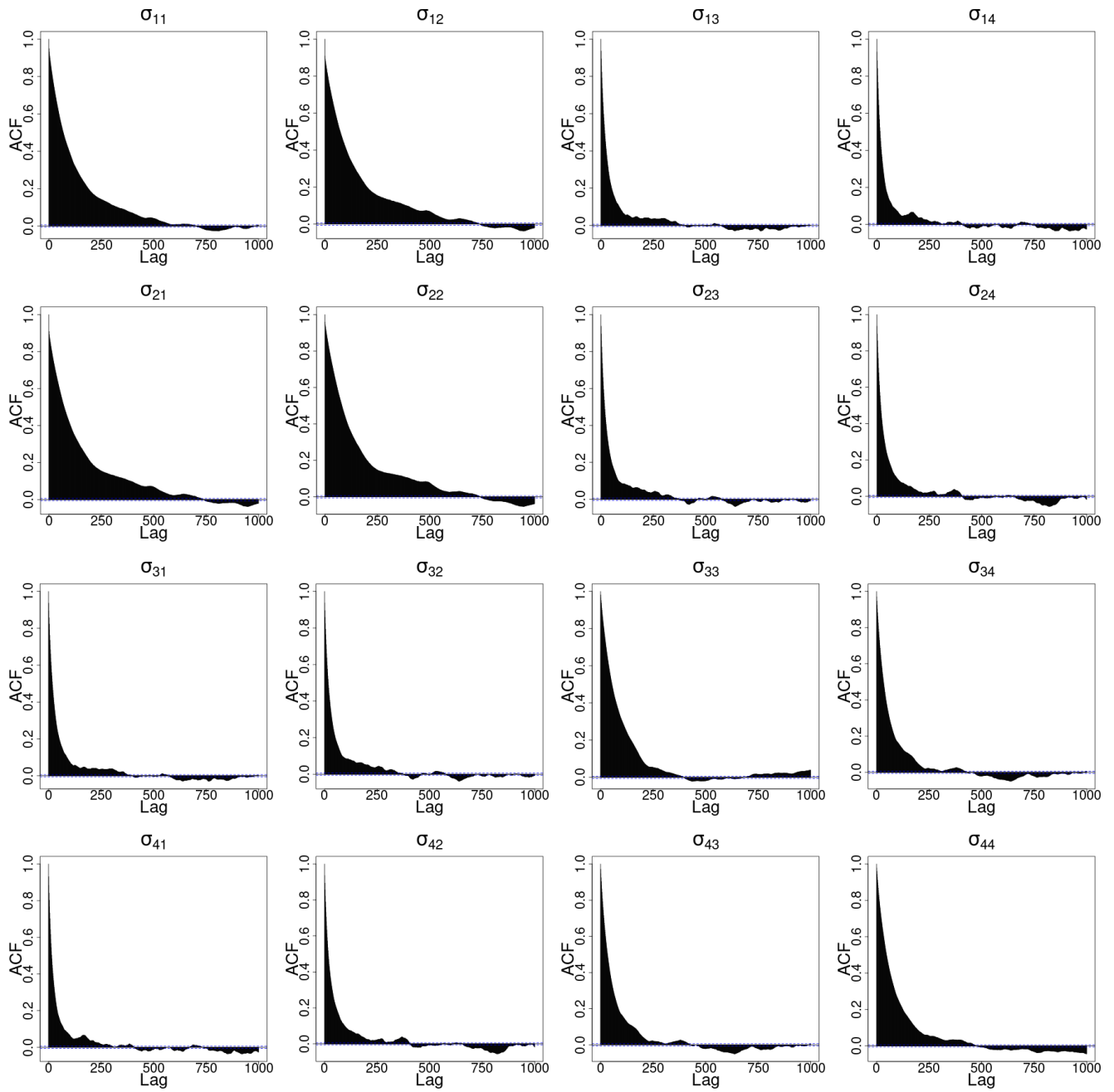


Figura 4.6: Función de autocorrelación de los elementos de  $\Sigma$ .

Fuente: Elaboración propia

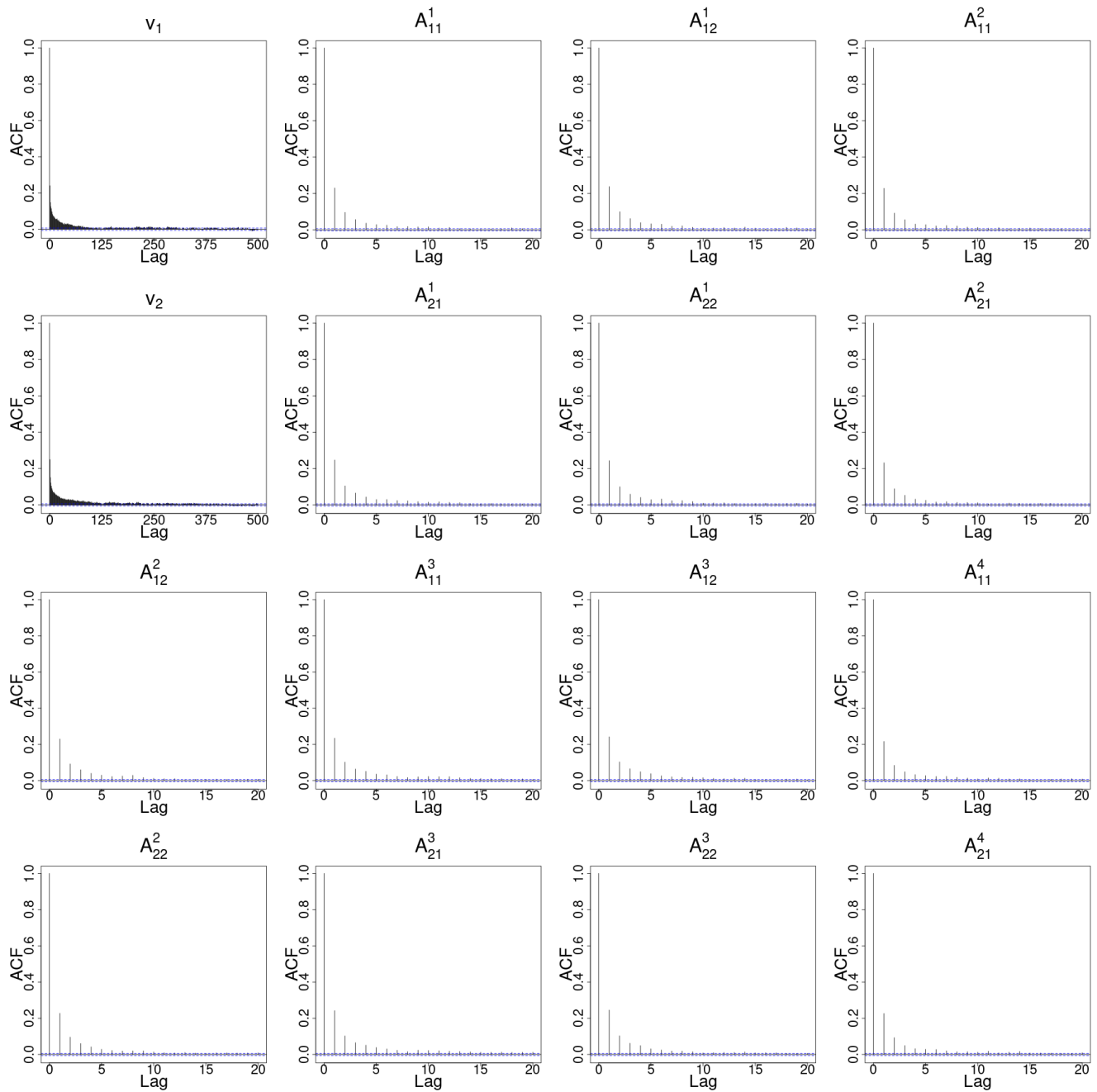


Figura 4.7: Función de autocorrelación del vector termino de interceptación y las matrices de coeficientes VAR.

Fuente: Elaboración propia



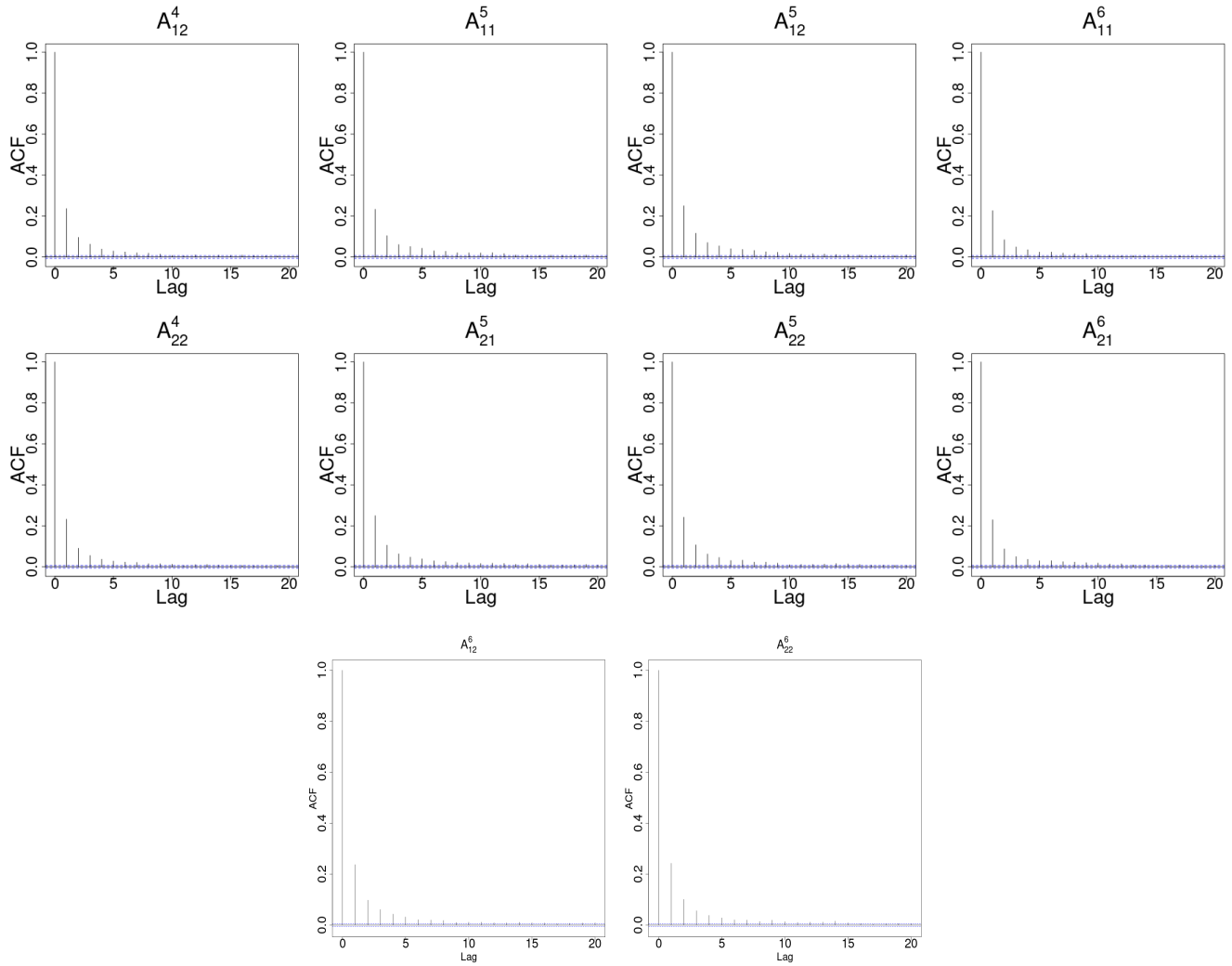


Figura 4.8: Función de autocorrelación de las matrices de coeficientes VAR.

Fuente: Elaboración propia

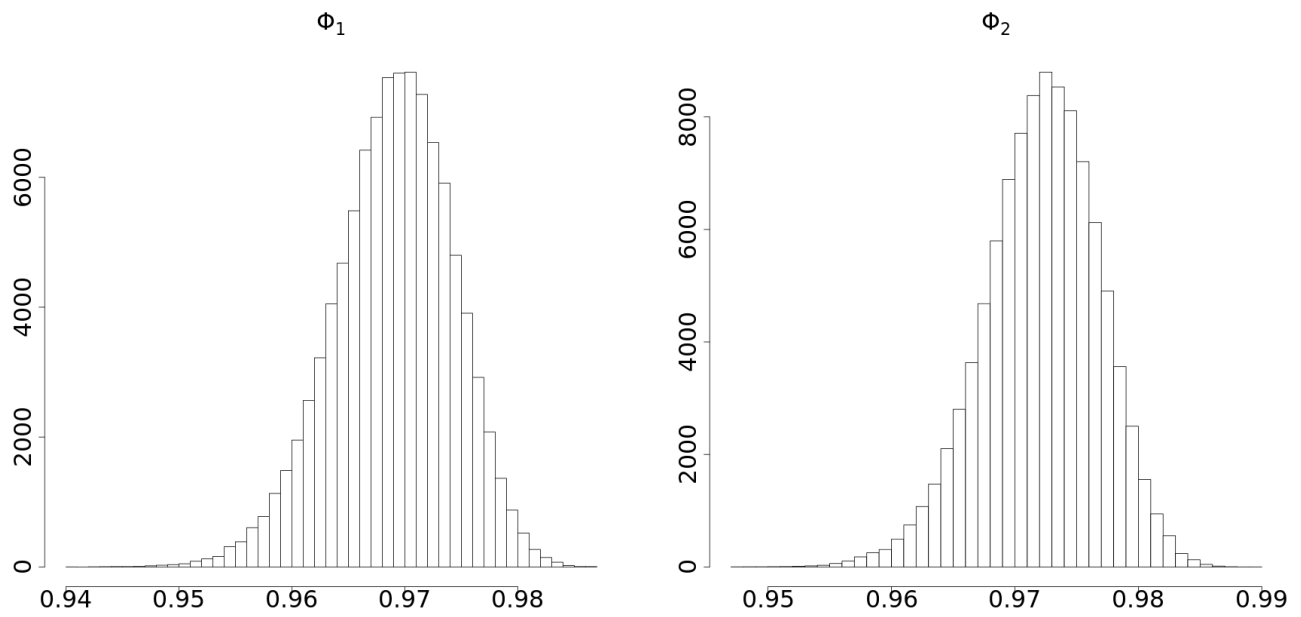


Figura 4.9: Histograma de  $\phi_1$  y  $\phi_2$  respectivamente.

Fuente: Elaboración propia

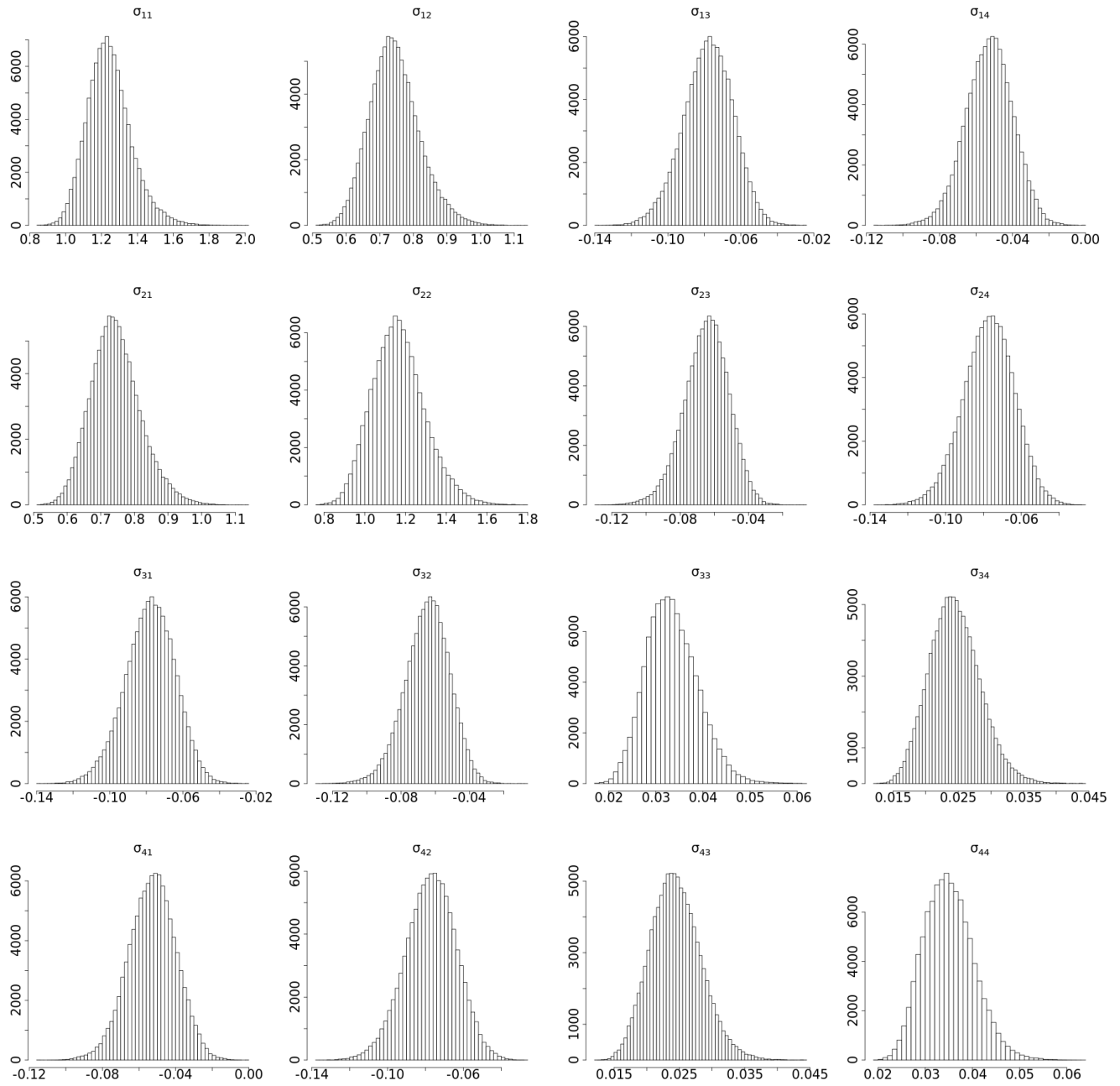


Figura 4.10: Histograma de los elementos de  $\Sigma$ .

Fuente: Elaboración propia

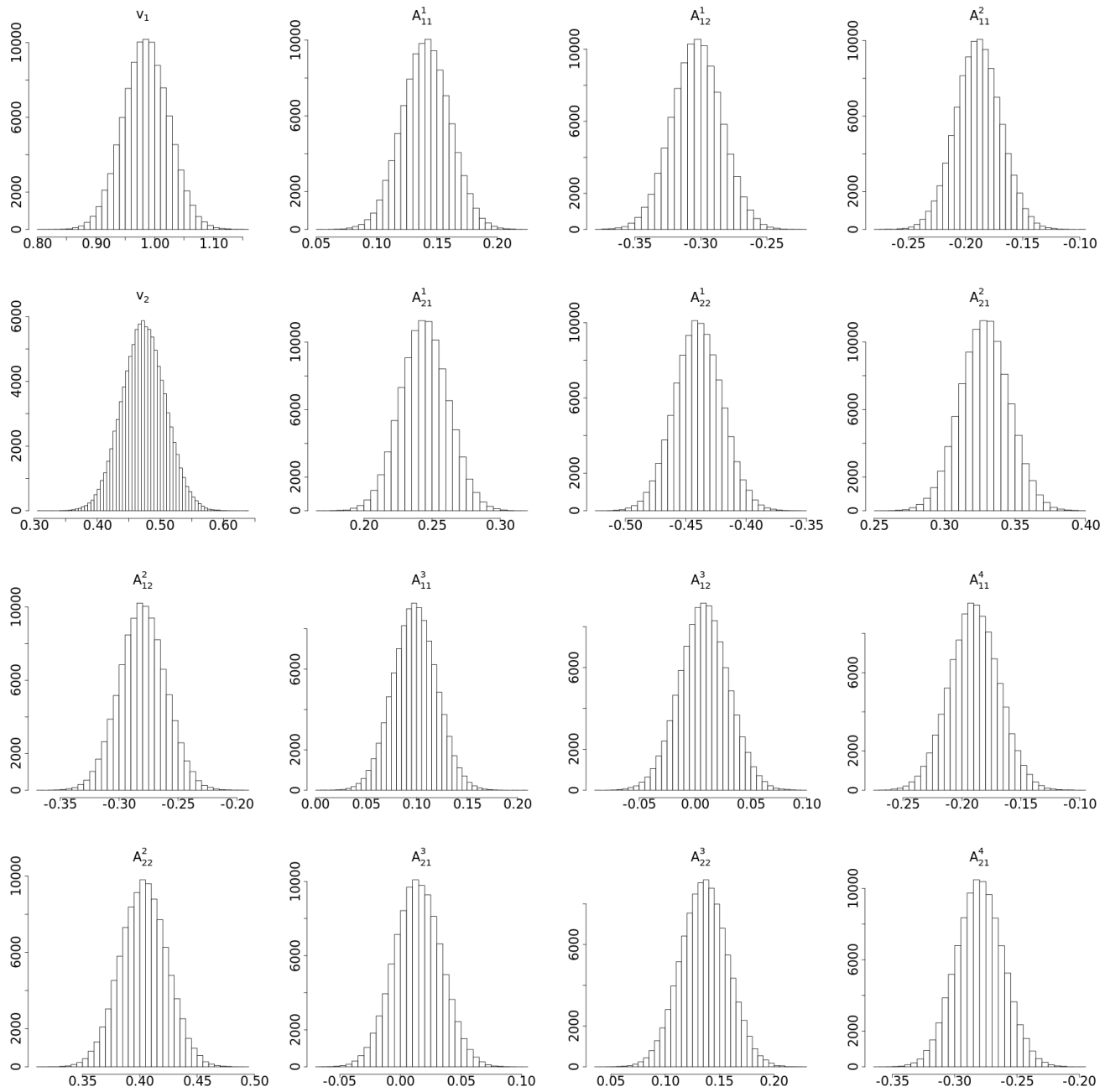


Figura 4.11: Histograma del vector termino de interceptación y las matrices de coeficientes VAR.  
 Fuente: Elaboración propia

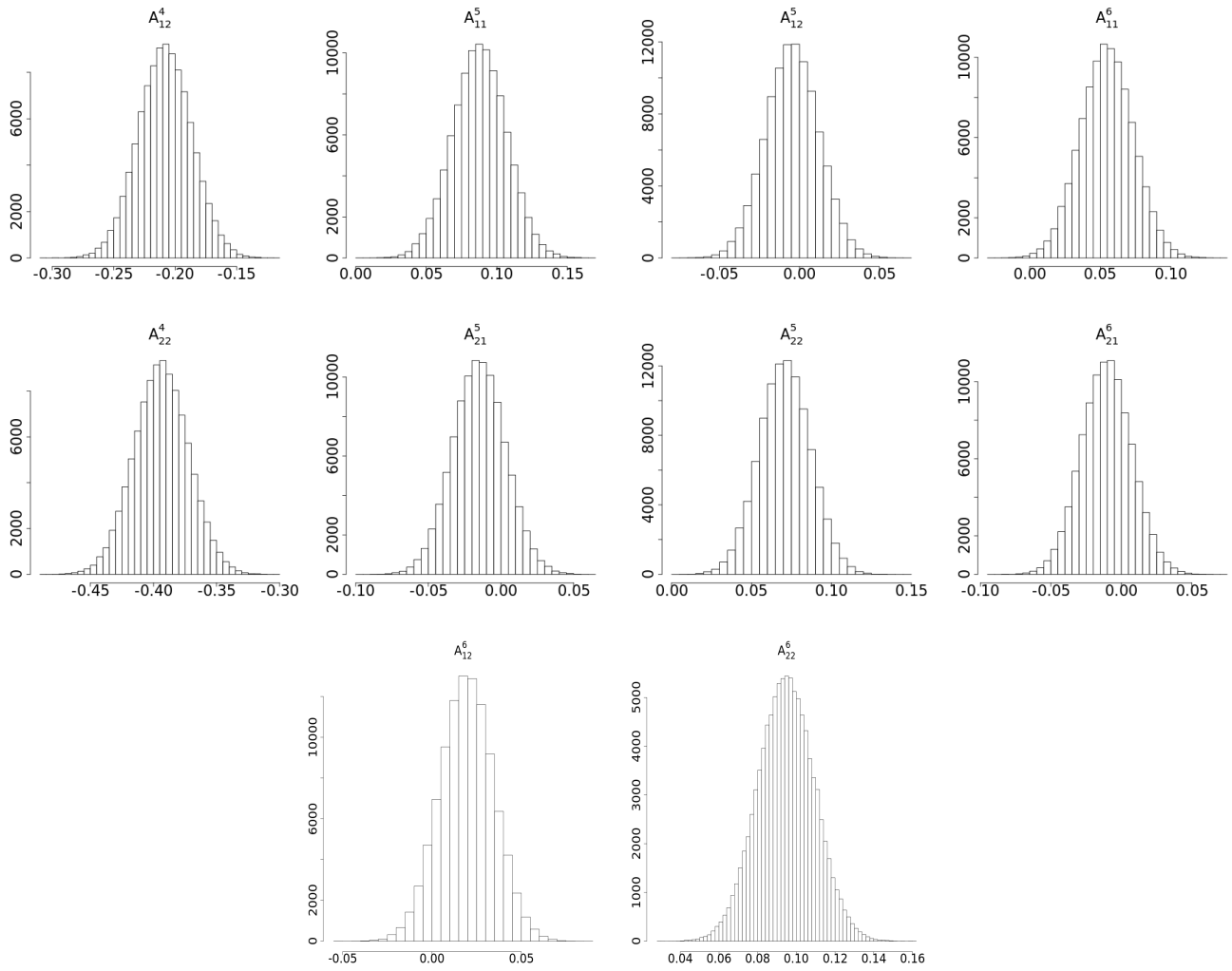


Figura 4.12: Histograma de las matrices de coeficientes VAR.

Fuente: Elaboración propia

## 4.6. Resultados Empíricos

En esta sección se aplica el modelo VAR-MSV con efecto de apalancamiento propuesto en la sección 4.2 a datos de series de tiempo bivariadas financieras consistiendo de precios de cierre de las acciones de Apple y Microsoft. Los hiperparámetros de las distribuciones priori  $\phi$ ,  $\Sigma$  y  $\beta$  son los que se usaron en la sección previa. Los datos son de frecuencia diaria excluyendo fines de semana y días festivos. Abarcan un periodo desde el 12 de enero del 2004 hasta el 13 de abril del 2022 consistiendo en 4597 observaciones y son mostrados en la figura 4.13. Se usa la ecuación (1.4) para convertir los precios a retornos logarítmicos. De ahora en adelante en este trabajo se les llamará solo retornos. Los retornos de Apple y Microsoft son mostrados en la figura 4.14 y se puede observar la presencia de unos pocos valores en valor absoluto muy grandes en relación con las demás observaciones. También parece indicar que la volatilidad varía con el tiempo y que parece ser relativamente alta en algunos periodos de tiempo y relativamente baja en otros periodos. Se pueden observar grupos de volatilidad especialmente el que se ubica entre los años 2007 y 2009 coincidiendo con la así llamada *Gran Recesión*. Otro grupo de volatilidad que resalta es el que se ubica entre 2019 y mediados del 2021 coincidiendo con la pandemia del coronavirus (COVID-19).

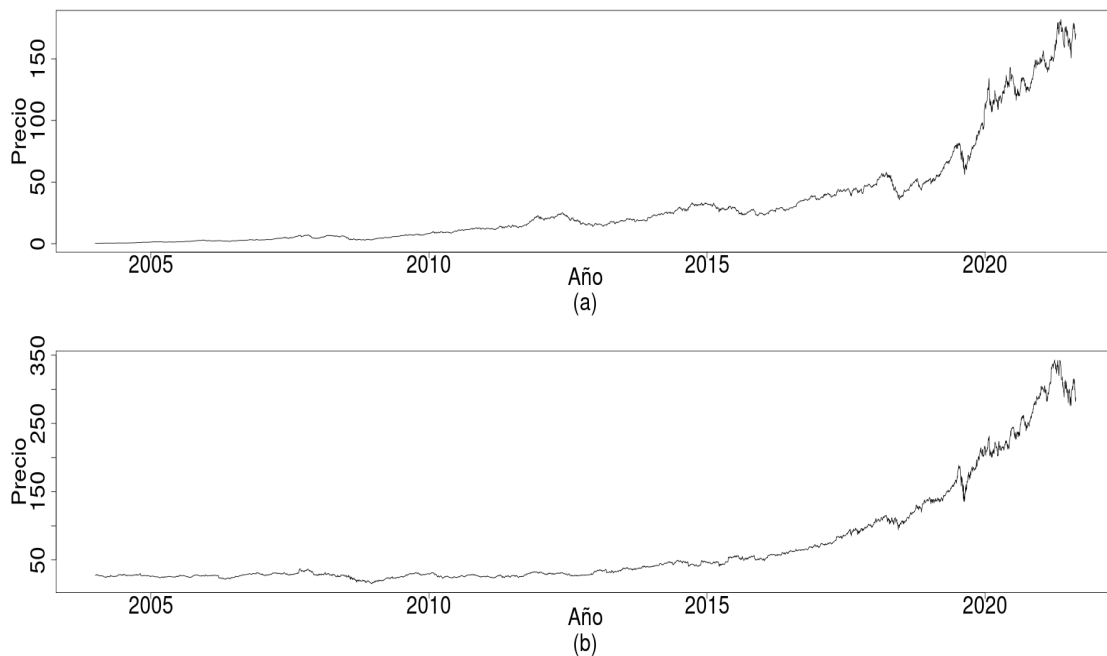


Figura 4.13: Gráficos temporales diarios de los precios de cierre desde el 12 de enero del 2004 hasta el 13 de abril del 2022: (a) Precios de cierre de Apple y (b) Precios de cierre de Microsoft.

**Fuente:** Elaboración propia

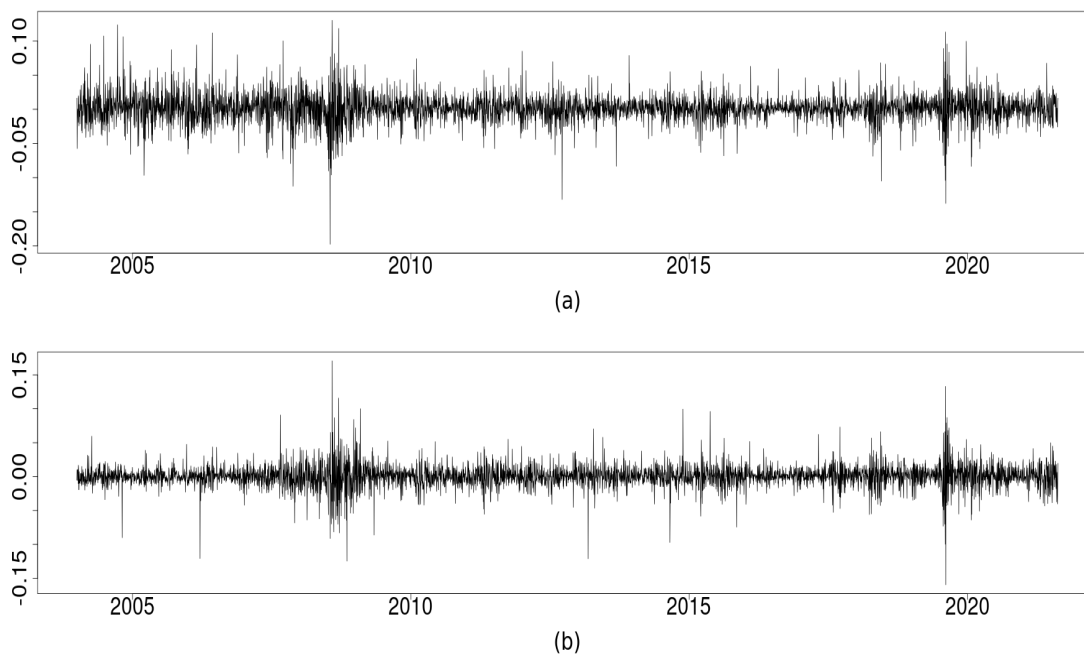


Figura 4.14: Gráficos temporales de los retornos: (a) Retornos de Apple y (b) Retornos de Microsoft. Fuente: Elaboración propia

La figura 4.15 muestra los histogramas de los retornos de Apple y Microsoft. A simple vista parece que ambas series de retornos tienen una distribución similar a la normal pero, al observar el gráfico  $Q - Q$  normal de la figura 4.16 indica que los retornos de ambas series tienen distribución de colas pesadas. En la parte superior de la figura 4.17 se pueden observar las funciones de correlación de los retornos de Apple y Microsoft. Se pueden observar rezagos significativos 4, 6, 7, 8, 9, 12, 15 y 16 aunque muy pequeños para los retornos de Apple mientras que para los retornos de Microsoft se observan rezagos significativos 1, 3, 4, 6, 8, 9, 12, 15 y 17 pero pequeños.

En la parte inferior de la figura 4.17 se puede observar que las funciones de correlación de los retornos al cuadrado de Apple y Microsoft son muy correlacionadas. EL ACF de los retornos al cuadrado de Apple comienza en 0.184 teniendo un máximo en 0.202 en el lag 10 mientras que el ACF de los retornos al cuadrado de Microsoft comienza en 0.28 siendo ese mismo valor el máximo en el rezago 1. Ambas ACF decrecen suavemente. Todas las observaciones anteriores muestran propiedades estadísticas de los retornos. Por último, la tabla 4.3 contiene estadísticas descriptivas de los retornos. Las estadísticas descriptivas de las dos series de retornos muestran medias de exceso de curtosis positiva implicando colas pesadas como se observó en el gráfico  $Q - Q$  normal. Además, ambas series de retornos tienen coeficientes de asimetría menores que cero.

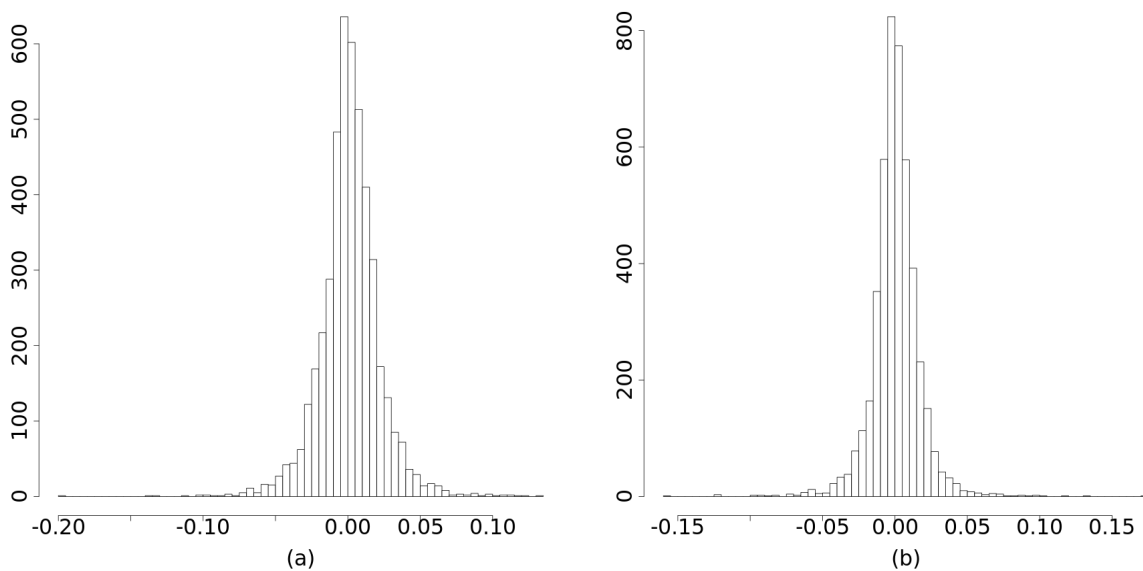


Figura 4.15: Histograma de los retornos: (a) Retornos de Apple y (b) Retornos de Microsoft.  
**Fuente: Elaboración propia**

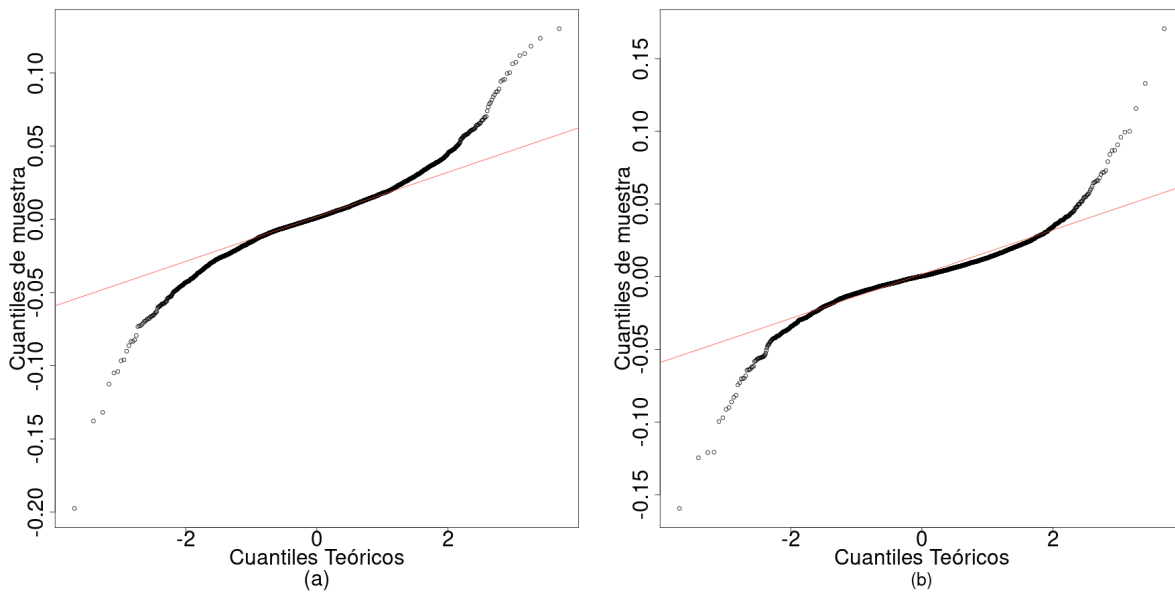


Figura 4.16: Gráfico  $Q - Q$  normal de los retornos: (a) Retornos de Apple y (b) Retornos de Microsoft.  
**Fuente: Elaboración propia**



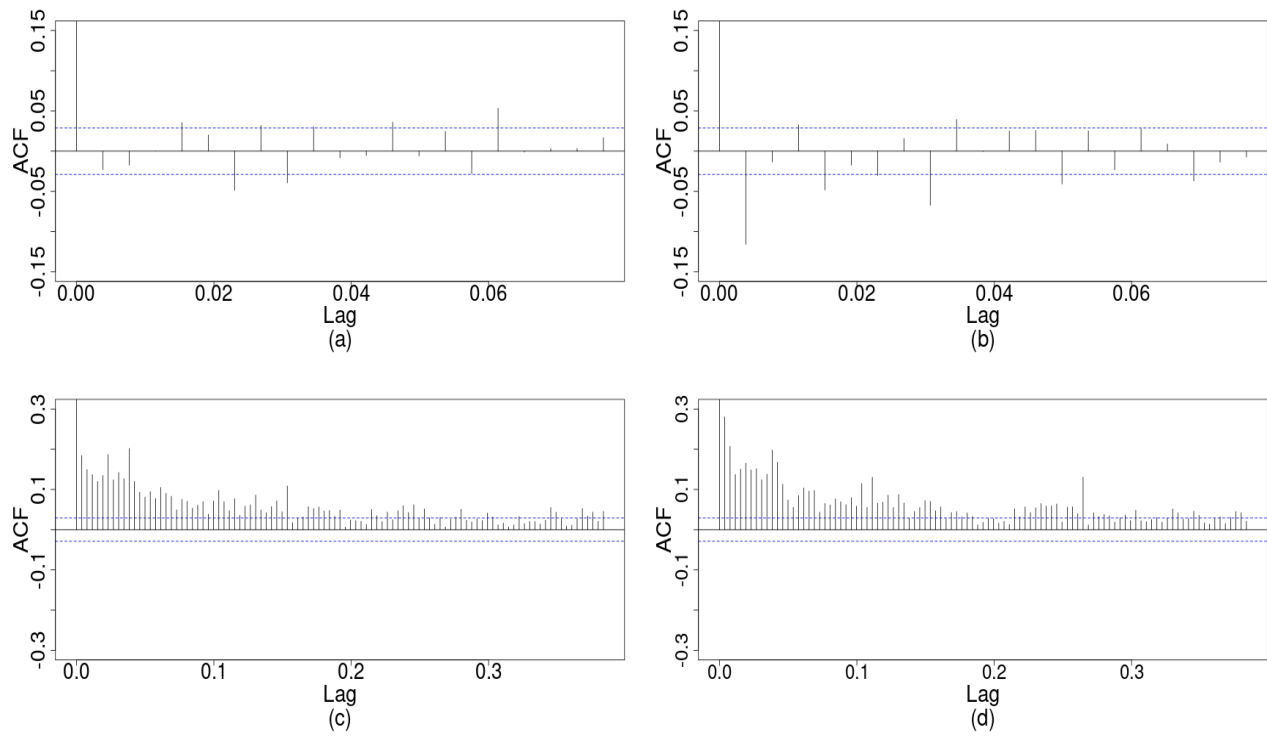


Figura 4.17: Función de autocorrelación de los retornos y los retornos al cuadrado: (a) Retornos de Apple, (b) Retornos de Microsoft, (c) Retornos al cuadrado de Apple y (d) Retornos al cuadrado de Microsoft.

**Fuente:** Elaboración propia

	Apple	Microsoft
Observaciones	4596	4596
Media	0.001305	0.000510
Mediana	0.001108	0.000359
Desviación estándar	0.021000	0.016872
Asimetría	-0.136485	-0.066780
Curtosis	5.861174	10.783503
Mínimo	-0.197470	-0.159453
Máximo	0.130194	0.170626

Tabla 4.3: Estadísticos descriptivos para los retornos de Apple y Microsoft.

Se estimaron tres configuraciones de modelos: VAR(4)-MSV, VAR(5)-MSV y VAR(6)-MSV. Para escoger el mejor modelo se usó el criterio DIC discutido en la sección 3.5. En la tabla 4.4 se muestran los valores DIC y sus respectivas desviaciones estándar de los modelos VAR(4)-MSV, VAR(5)-MSV y VAR(6)-MSV. El valor DIC más pequeño corresponde al modelo VAR(5)-MSV, por lo que indica que el modelo VAR(5)-MSV se ajusta mejor a los datos para mostrar su comportamiento y relaciones dinámicas.

Modelo	VAR(4)-MSV	VAR(5)-MSV	VAR(6)-MSV
DIC	-59213.95	-59218.18	-59200.73
Desviación estándar	0.1881723	0.2531091	0.2105574
DIC <sub>máx</sub>	-59215.54	-59220.17	-59202.14
DIC <sub>mín</sub>	-59212.89	-59216.9	-59199.2

Tabla 4.4: Medias de las muestras, errores estándar, valores más grandes y valores más pequeños DIC de los modelos VAR(4)-MSV, VAR(5)-MSV y VAR(6)-MSV.

Las figuras 4.18-4.21 muestran los gráficos de convergencia de los parámetros estimados. Las tablas 4.5-4.6 muestran la media posterior, desviación estándar e intervalos de credibilidad del 95 % de los elementos de la matriz  $\Phi$ , la matriz de covarianza  $\Sigma$ , el termino de intersección  $v$  y las matrices de coeficientes  $A_1, A_2, \dots, A_5$ .

Todos los intervalos de credibilidad contienen el cero a excepción de los intervalos de credibilidad de  $v_1, v_2, A_{12}^1$  y  $A_{11}^4$  mostrando evidencia que puede que en realidad las matrices de coeficientes sean igual a cero. Los estimadores de  $\phi_i$  no muestra persistencia en la volatilidad tan alta variando de 0.936 a 0.942.  $\rho_{ij,\eta\eta} = 0.74$  indica una correlación muy alta en  $\eta_{1t}$  y  $\eta_{2t}$ .  $\rho_{ij,\varepsilon\varepsilon} = 0.5016$  refleja una correlación que no es tan alta por lo que existe una relación lineal muy débil entre los retornos de  $y_{1t}$  (Apple) y  $y_{2t}$  (Microsoft). Los elementos de las matrices de coeficientes  $A_{12}^i, i = 1, \dots, 5$ ,

	Media	Desviación Estándar	Intervalo de Credibilidad del 95 %
$\phi_1$	0.94213	0.008771	[0.92380,0.95852]
$\phi_2$	0.93611	0.010165	[0.91506,0.95494]
$\sigma_{11}$	0.00030	0.000028	[0.000251,0.000363]
$\sigma_{12}$	0.00011	0.000010	[0.000096,0.000137]
$\sigma_{13}$	-0.00155	0.000257	[-0.002067,-0.001063]
$\sigma_{14}$	-0.00137	0.000262	[-0.001901,-0.000872]
$\sigma_{21}$	0.00011	0.000010	[0.000096,0.000137]
$\sigma_{22}$	0.00017	0.000015	[0.000145,0.000207]
$\sigma_{23}$	-0.00076	0.000192	[-0.001150,-0.000393]
$\sigma_{24}$	-0.00077	0.000202	[-0.001186,-0.000390]
$\sigma_{31}$	-0.00155	0.000257	[-0.002067,-0.001063]
$\sigma_{32}$	-0.00076	0.000192	[-0.001150,-0.000393]
$\sigma_{33}$	0.10832	0.014398	[0.082934,0.139478]
$\sigma_{34}$	0.08535	0.010334	[0.066795,0.107140]
$\sigma_{41}$	-0.00137	0.000262	[-0.001901,-0.000872]
$\sigma_{42}$	-0.00077	0.000202	[-0.001186,-0.000390]
$\sigma_{43}$	0.08535	0.010334	[0.066795,0.107140]
$\sigma_{44}$	0.12281	0.016507	[0.094076,0.158113]

Tabla 4.5: Media posterior, Desviación estándar e Intervalos de credibilidad del 95 % de los elementos de la matriz  $\Phi$  y la matriz de covarianza  $\Sigma$ .

son muy cercanos a cero por lo que parecen tener poco efecto condicional de  $y_{2,t-i}, i = 1, \dots, 5$ , en  $y_{1t}$  dado  $y_{1,t-i}, i = 1, \dots, 5$ , además,  $A_{11}^i, i = 1, \dots, 5$ , están más alejados de cero por lo que  $y_{1t}$  está más alimentada por sus propias observaciones pasadas que de las de  $y_{2t}$ . Por otro lado, los elementos de las matrices de coeficientes  $A_{21}^i, i = 1, \dots, 5$ , están un poco más alejados de cero por lo que es un poco más fuerte el efecto condicional de  $y_{1,t-i}, i = 1, \dots, 5$  en  $y_{2t}$  dado  $y_{2,t-i}, i = 1, \dots, 5$ .

Observe que los elementos de las matrices de coeficientes  $A_{22}^i, i = 1, \dots, 5$  son más cercanos a cero que  $A_{21}^i, i = 1, \dots, 5$ , por lo que  $y_{2t}$  parece que es más alimentado por las observaciones pasadas de  $y_{1t}$  que de las suyas propias. Se debe tener en cuenta que los retornos de Apple varían entre  $[-0.1978, 0.1301]$  y los retornos de Microsoft varían entre  $[-0.1594, 0.1706]$  y, dado que según la figura 4.14 la mayoría de los retornos están más cercanos a cero que de sus respectivos máximos y mínimos por lo que podría ser que los elementos de las matrices de coeficientes en realidad tengan un mayor impacto. Las correlaciones  $\rho_{11,\varepsilon\eta} = -0.2725$  y  $\rho_{22,\varepsilon\eta} = -0.1689$  son ambas negativas y sus respectivos intervalos de credibilidad del 95 % son  $[-0.3500, -0.1930]$  y  $[-0.2492, -0.0864]$ . Ninguno de los intervalos de confianza contiene al cero por lo que existe evidencia de efecto de apalancamiento en ambas series.

	Media	Desviación Estándar	Intervalo de Credibilidad del 95 %
$v_1$	0.00106	0.00023	[0.00060, 0.00153]
$A_{11}^1$	0.00769	0.01519	[-0.02217, 0.03755]
$A_{12}^1$	0.00037	0.00098	[-0.00160, 0.00239]
$A_{11}^2$	0.00796	0.01477	[-0.02094, 0.03690]
$A_{12}^2$	-0.00148	0.00096	[-0.00343, 0.00047]
$A_{11}^3$	-0.01133	0.01444	[-0.03969, 0.01695]
$A_{12}^3$	0.00015	0.00095	[-0.00169, 0.00217]
$A_{11}^4$	0.03509	0.01412	[0.00743, 0.06286]
$A_{12}^4$	0.00007	0.00093	[-0.00178, 0.00199]
$A_{11}^5$	0.01210	0.01388	[-0.01510, 0.03922]
$A_{12}^5$	-0.00017	0.00067	[-0.00157, 0.00112]
$v_2$	0.00060	0.00018	[0.00024, 0.00096]
$A_{21}^1$	-0.04562	0.01021	[-0.06561, -0.02551]
$A_{22}^1$	0.00126	0.00067	[-0.00011, 0.00261]
$A_{21}^2$	0.00175	0.01010	[-0.01792, 0.02153]
$A_{22}^2$	-0.00097	0.00065	[-0.00228, 0.00036]
$A_{21}^3$	-0.01494	0.00986	[-0.03431, 0.00437]
$A_{22}^3$	0.00040	0.00064	[-0.00084, 0.00172]
$A_{21}^4$	0.01148	0.00977	[-0.00760, 0.03065]
$A_{22}^4$	0.00090	0.00062	[-0.00033, 0.00217]
$A_{21}^5$	0.00183	0.00962	[-0.01703, 0.02061]
$A_{22}^5$	-0.00074	0.00044	[-0.00165, 0.00011]

Tabla 4.6: Media posterior, Desviación estándar e Intervalos de credibilidad del 95 % de los elementos del vector término intersección  $v$  y las matrices de coeficientes  $A_1, A_2, \dots, A_5$ .

Por otro lado, las correlaciones  $\rho_{12,\varepsilon\eta} = -0.2268$  y  $\rho_{21,\varepsilon\eta} = -0.1765$  también son negativas y sus respectivos intervalos de credibilidad del 95 % son  $[-0.3062, -0.1463]$  y  $[-0.2586, -0.092]$ . Ninguno de los intervalos de confianza contiene el cero por lo que existe evidencia de efecto de apalancamiento cruzado en ambas series. Note que  $\rho_{12,\varepsilon\eta} < \rho_{21,\varepsilon\eta}$  indicando asimetría. Esto significa que la volatilidad de la serie Microsoft está más influenciada por las disminuciones en los retornos de la serie de Apple mientras que esta última está menos sujeta a cambios en los retornos de Microsoft. Esto podría sugerir que los participantes del mercado no reaccionan tan bruscamente cuando una disminución en el rendimiento se limita a la serie de Microsoft, pero reaccionan de manera más sensible si se producen disminuciones en Apple. De todas las observaciones anteriores se resaltan las siguientes tres:

1. Los retornos  $y_{1t}$  correspondientes a Apple parece que están más alimentados por sus propios retornos pasados que de los retornos pasados de  $y_{2t}$  correspondientes a Microsoft.

2. Los retornos  $y_{2t}$  correspondientes a Microsoft parece que están más alimentados por los retornos pasados de  $y_{1t}$  correspondiente a Apple que de los suyos propias.
3. Los participantes del mercado no reaccionan tan bruscamente cuando una disminución en el rendimiento se limita a la serie de Microsoft, pero reaccionan de manera más sensible si se producen disminuciones en Apple.

Las tres observaciones anteriores parecen sugerir que el efecto de apalancamiento cruzado se ve reflejado en las matrices de coeficientes.

Las figuras 4.22-4.25 muestran las funciones de autocorrelación de los parámetros estimados. Los histogramas se presentan en las figuras 4.26-4.29.

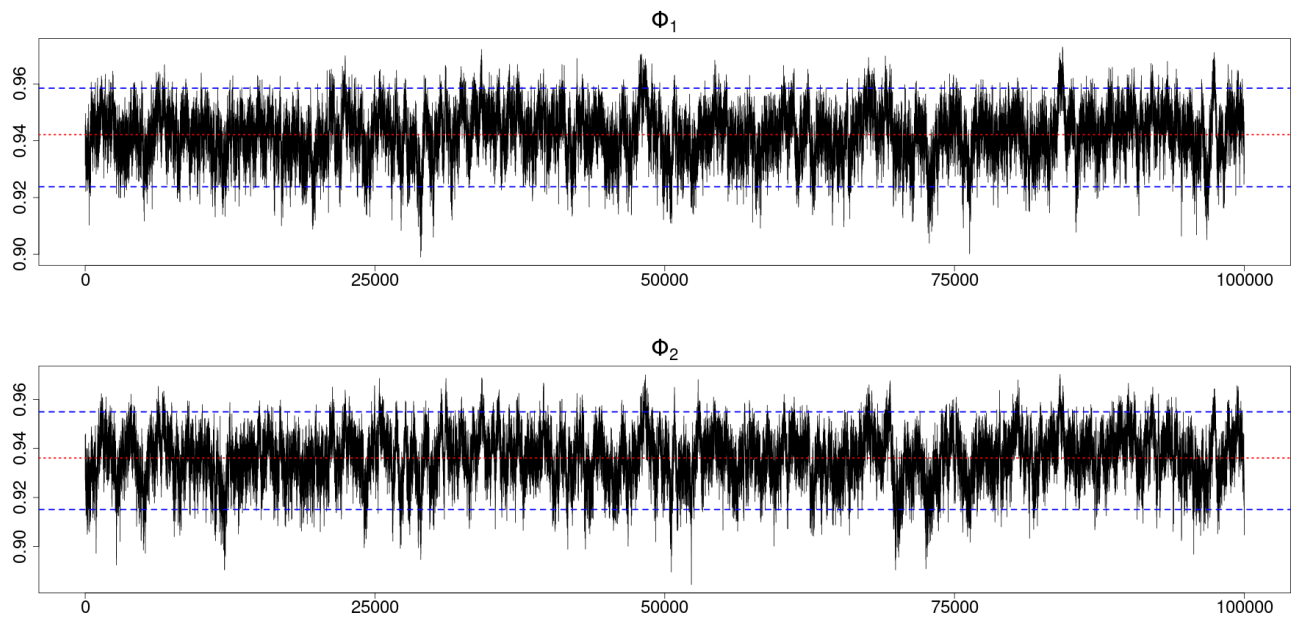


Figura 4.18: Gráfico de convergencia de las muestras de  $\phi_1$  y  $\phi_2$  respectivamente. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.

**Fuente: Elaboración propia**

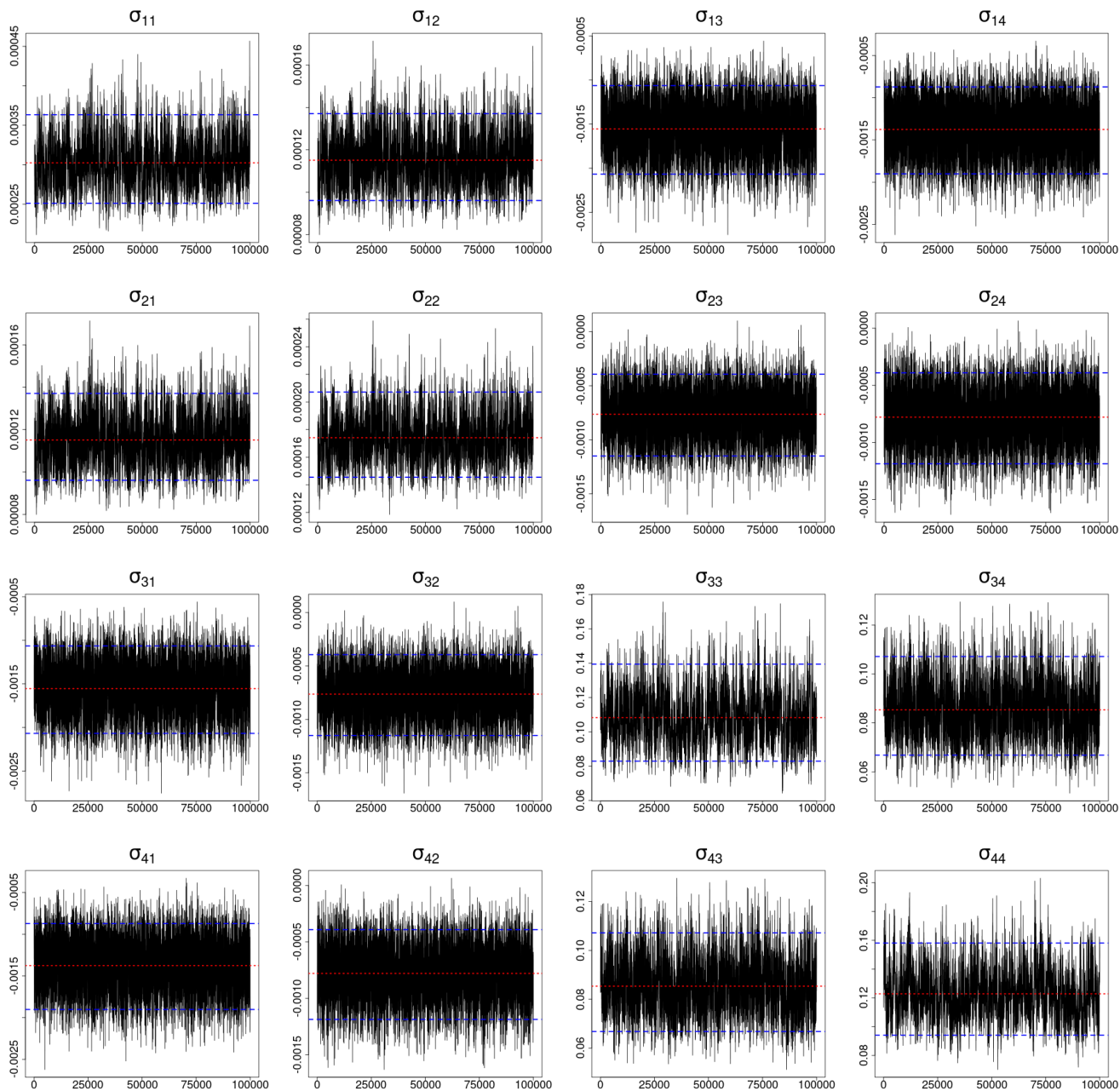


Figura 4.19: Gráfico de convergencia de las muestras de los elementos de la matriz de covarianza  $\Sigma$ . La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95%.

Fuente: Elaboración propia

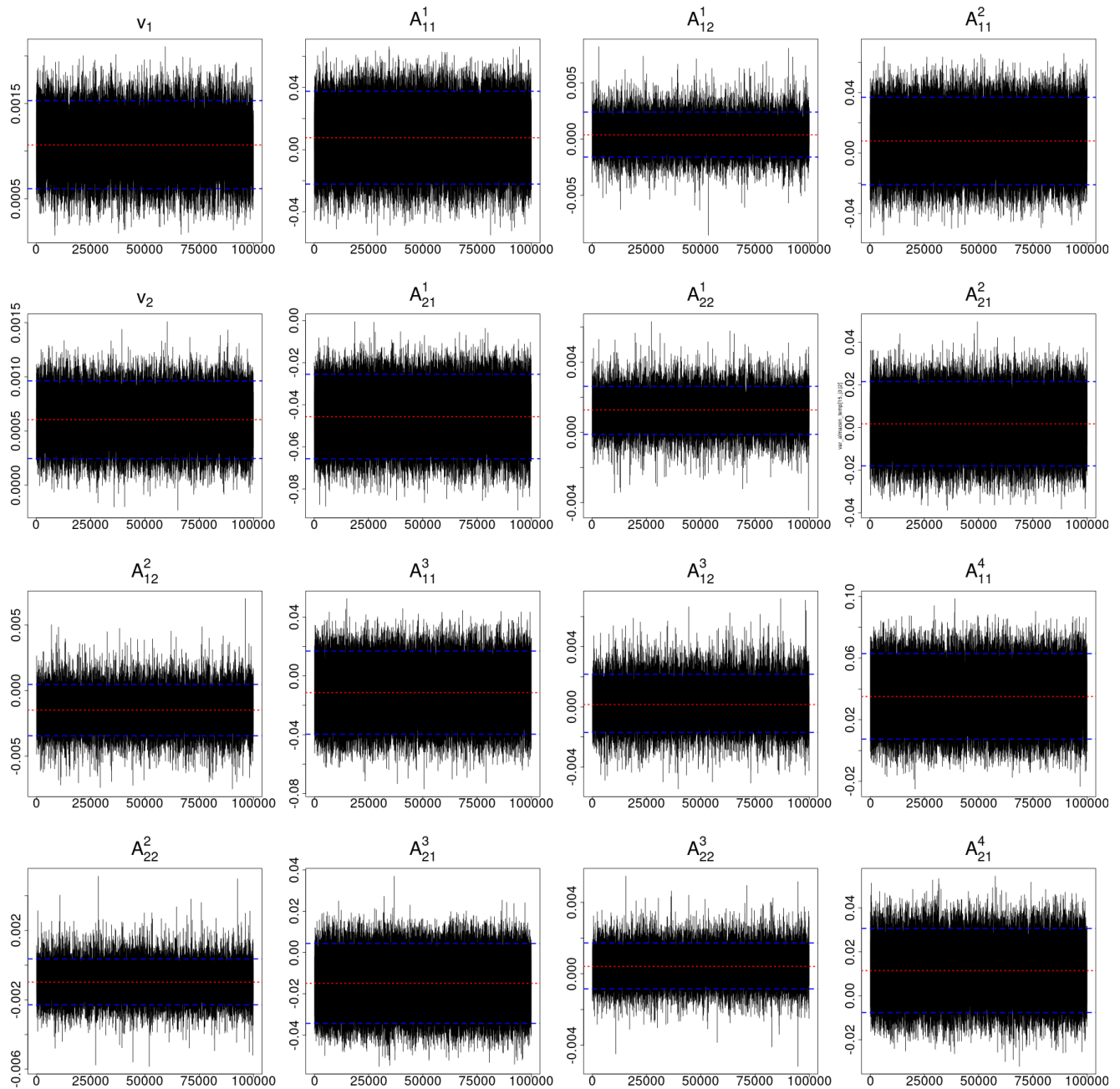


Figura 4.20: Gráfico de convergencia de las muestras de los elementos del vector término de intercepción y las matrices de coeficientes VAR. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.

Fuente: Elaboración propia

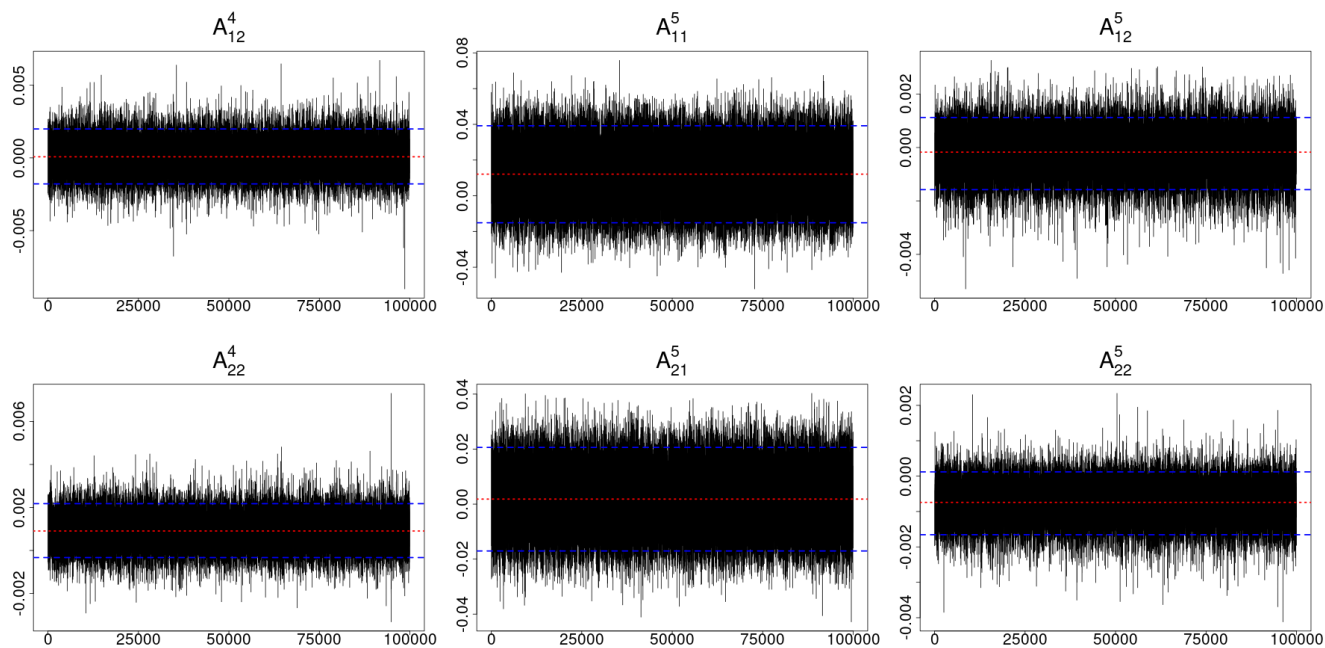


Figura 4.21: Gráfico de convergencia de las muestras de los elementos del vector término de intercepción y las matrices de coeficientes VAR. La línea roja es la media posterior de la muestra. Las líneas azules son el intervalo de credibilidad del 95 %.

Fuente: Elaboración propia

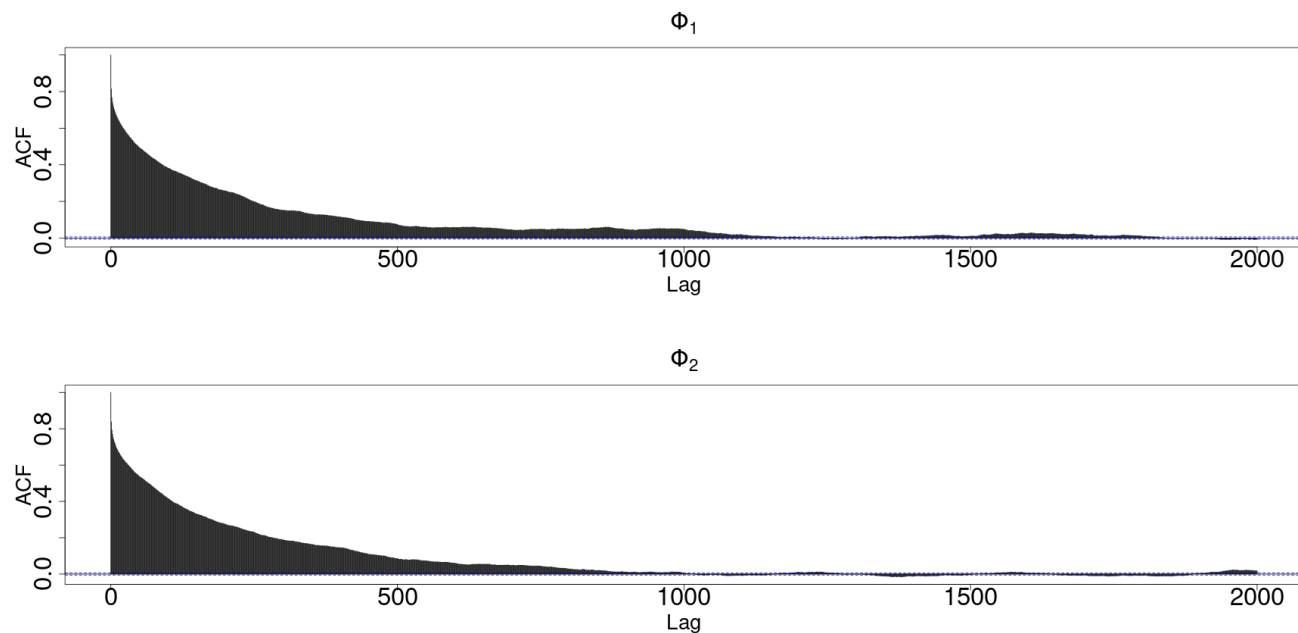


Figura 4.22: Función de autocorrelación de  $\phi_1$  y  $\phi_2$  respectivamente.

Fuente: Elaboración propia



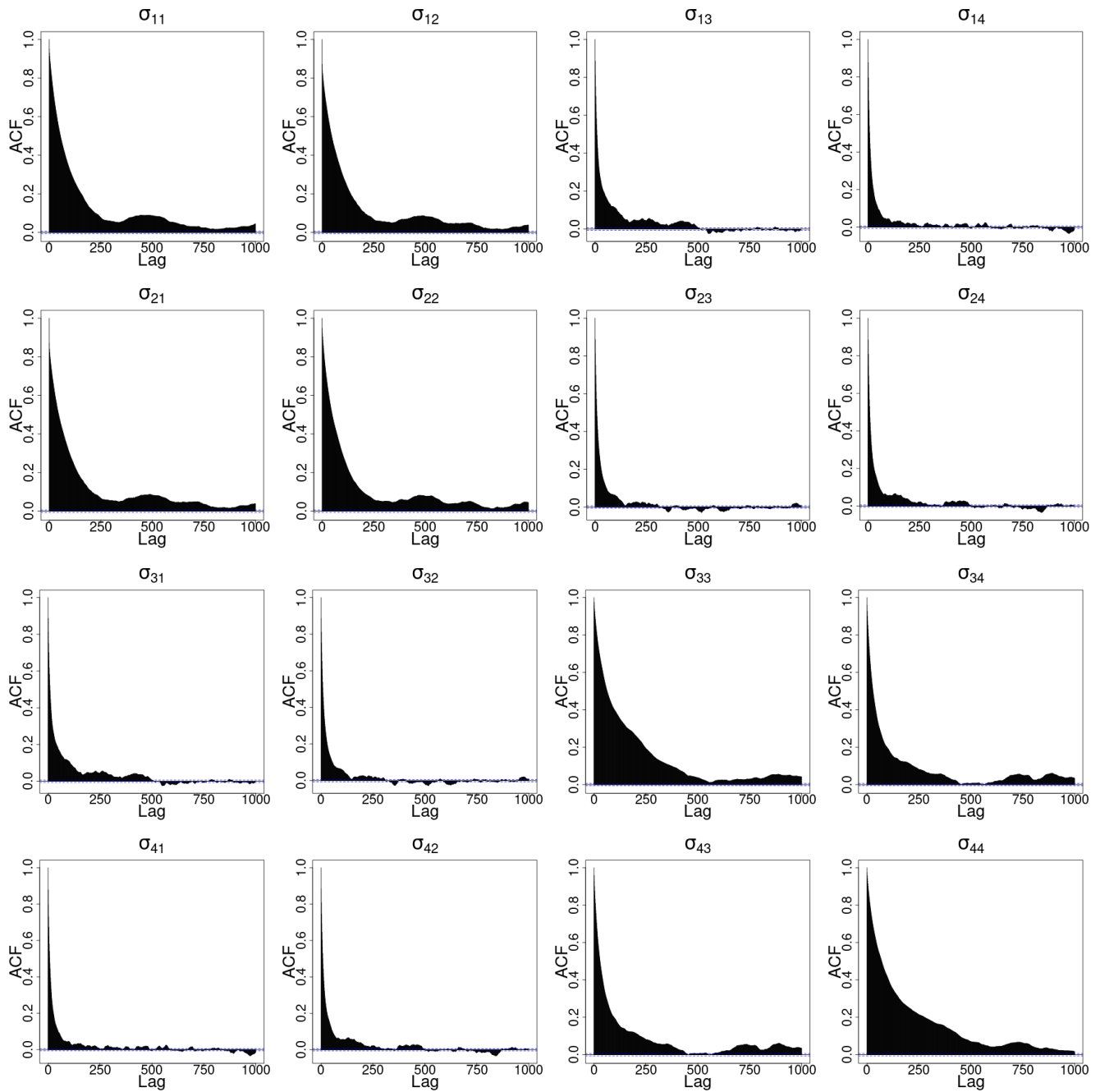


Figura 4.23: Función de autocorrelación de los elementos de la matriz de covarianza  $\Sigma$ .

Fuente: Elaboración propia

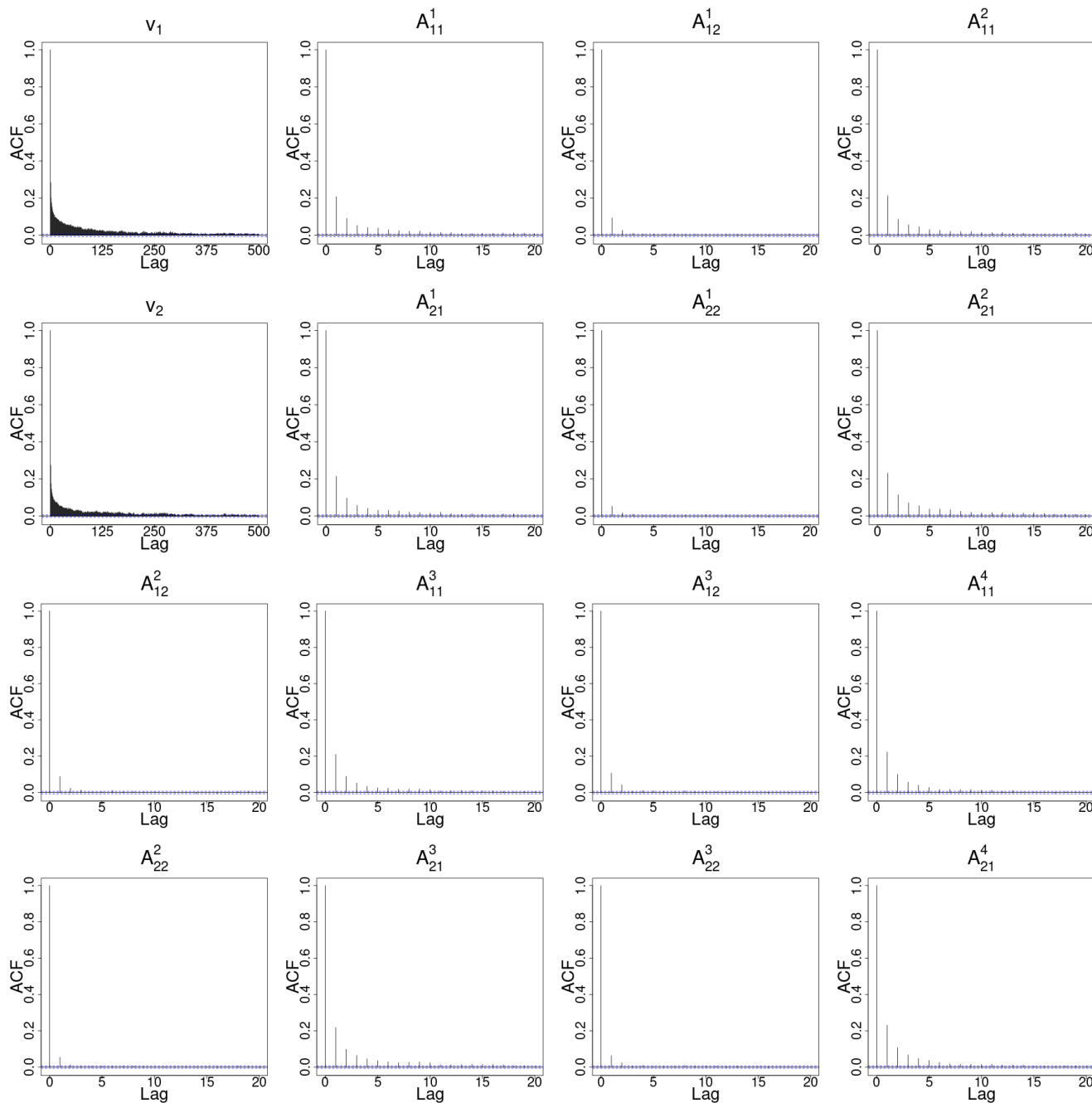


Figura 4.24: Función de autocorrelación de los elementos del vector término de intercepción y las matrices de coeficientes VAR.

Fuente: Elaboración propia

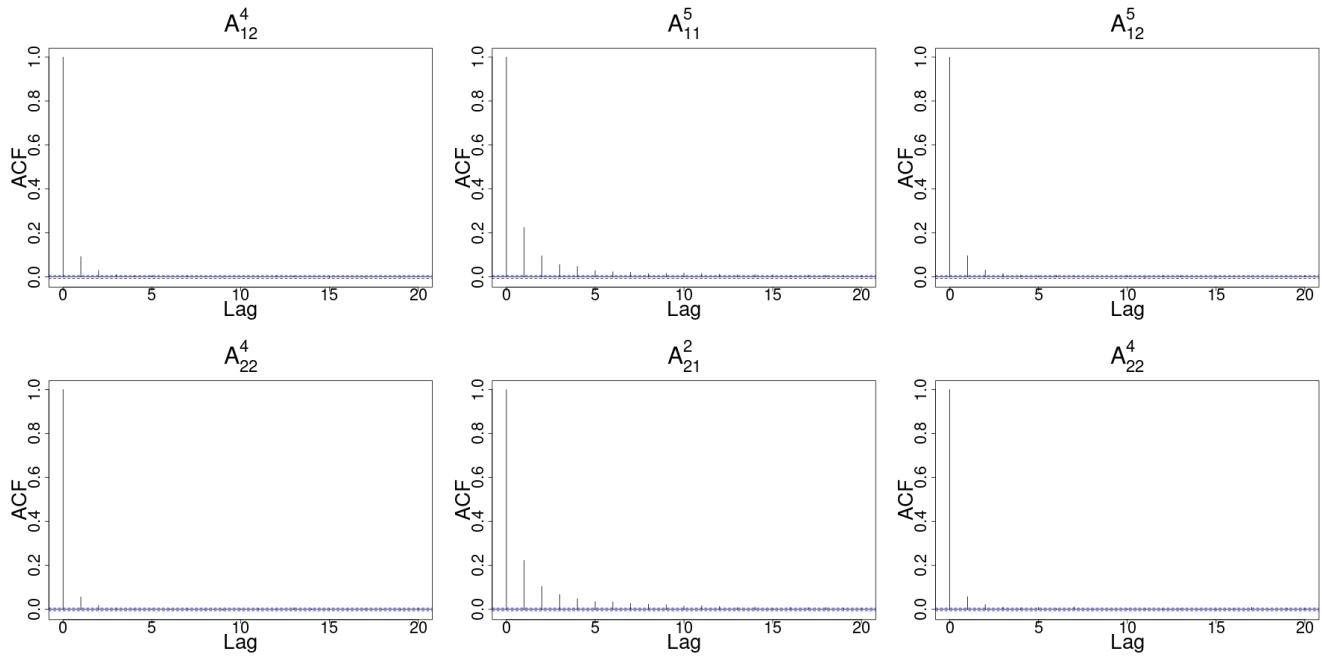


Figura 4.25: Función de autocorrelación de los elementos de las matrices de coeficientes VAR.  
**Fuente: Elaboración propia**

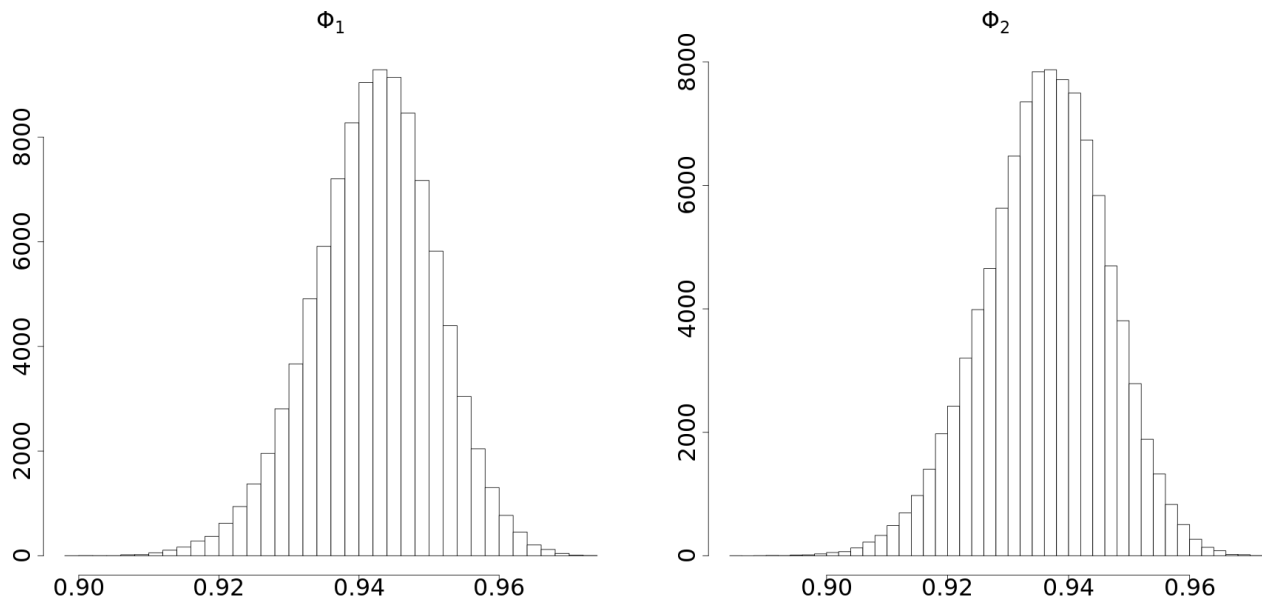


Figura 4.26: Histograma de  $\phi_1$  y  $\phi_2$  respectivamente.  
**Fuente: Elaboración propia**

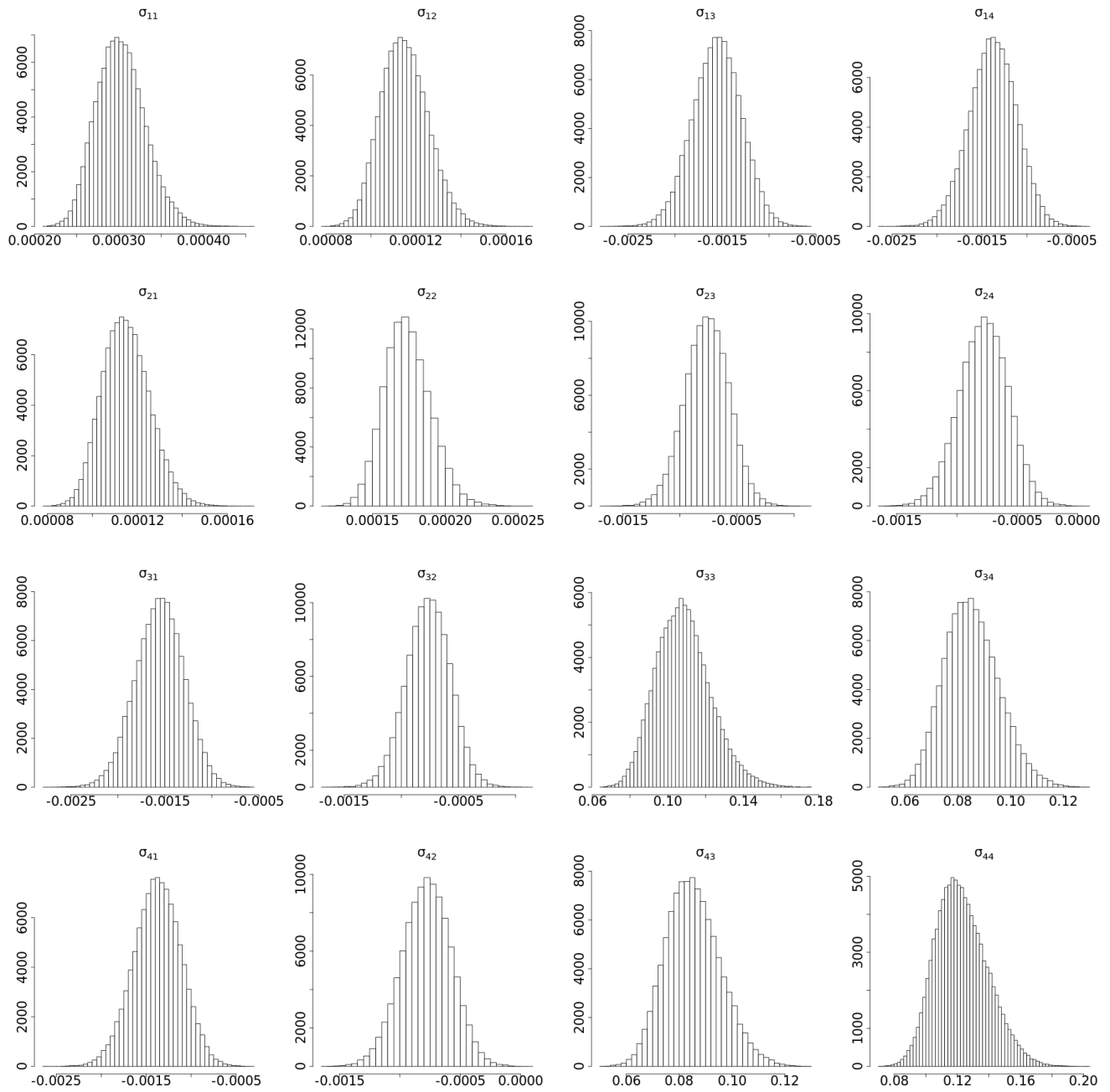


Figura 4.27: Histograma de los elementos de la matriz de covarianza  $\Sigma$ .  
 Fuente: Elaboración propia

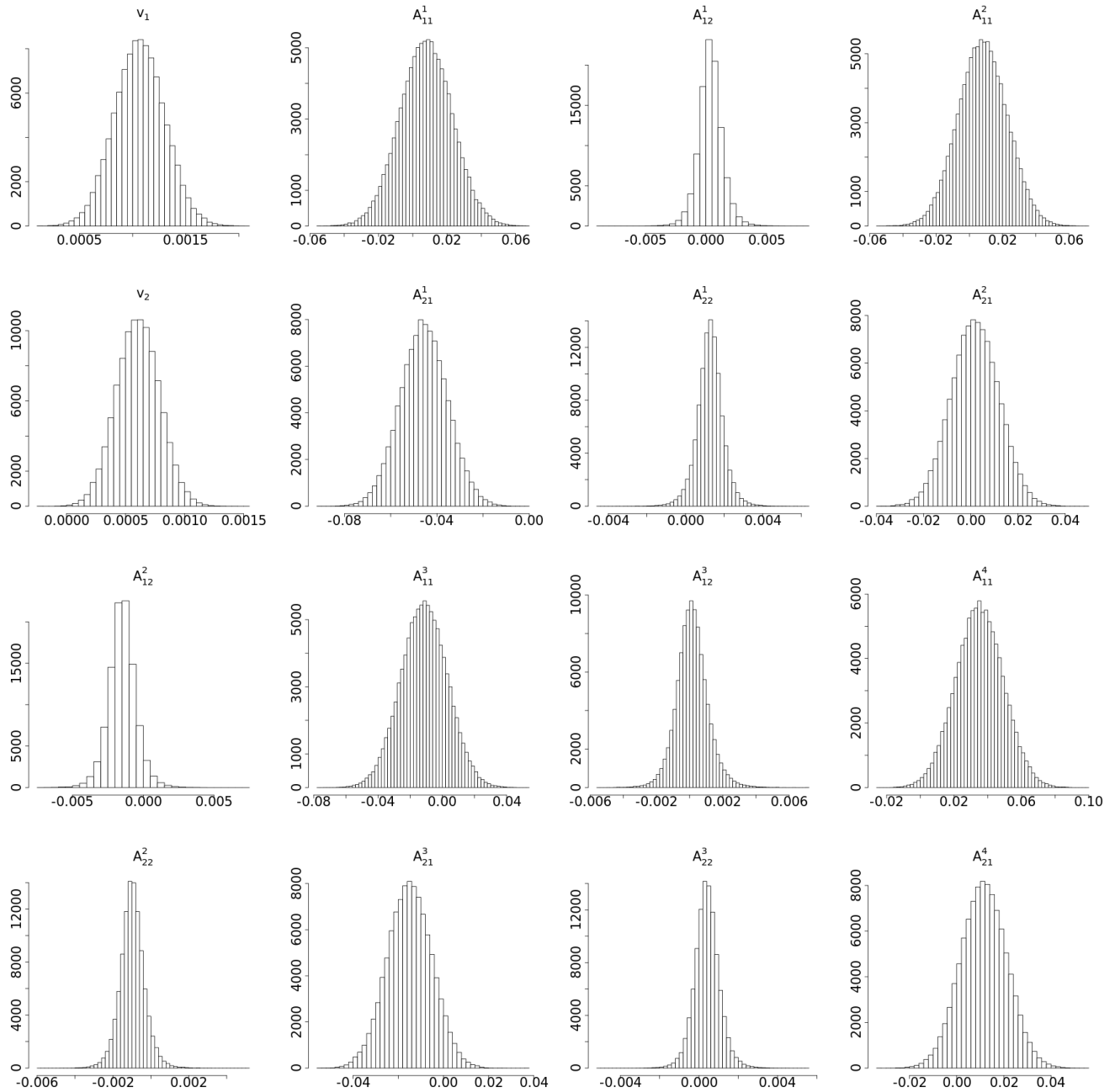


Figura 4.28: Histograma de los elementos del vector término de intercepción y las matrices de coeficientes VAR.

Fuente: Elaboración propia

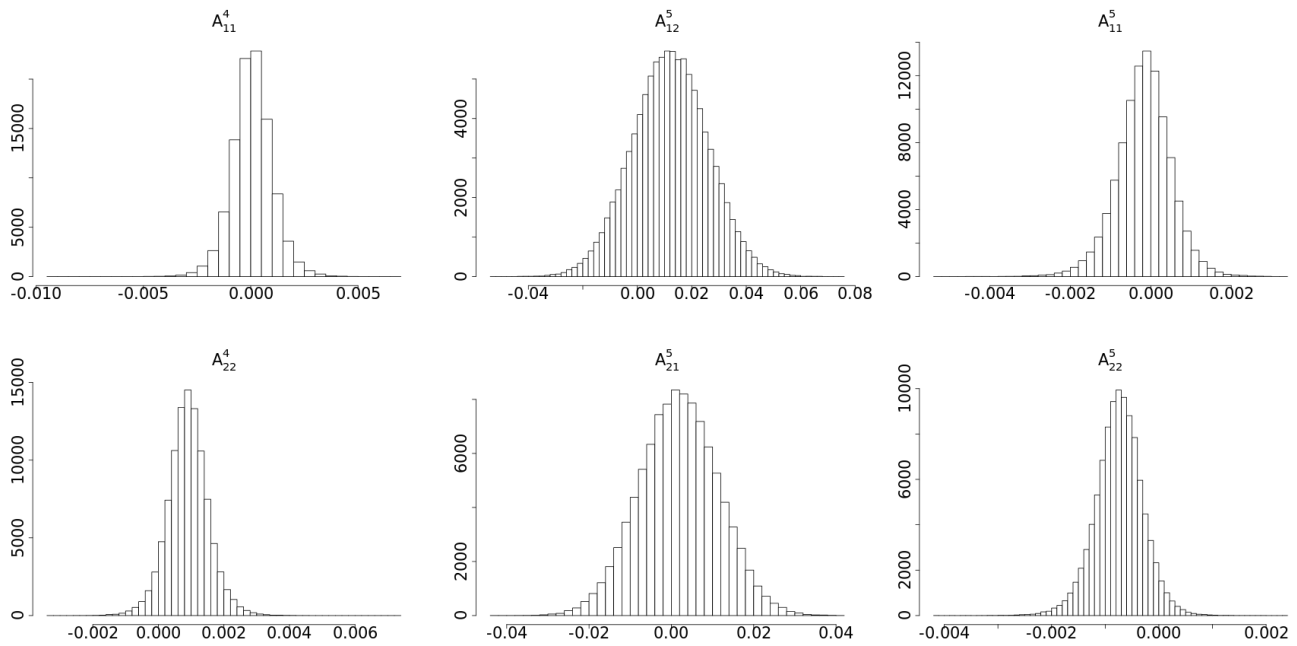


Figura 4.29: Histograma de los elementos de las matrices de coeficientes VAR.

Fuente: Elaboración propia

# Capítulo 5

## Conclusiones

En esta tesis se propuso un modelo VAR-MSV para modelar series de tiempo multivariadas. Se usó una configuración bayesiana para estimar los parámetros del modelo. La configuración de la volatilidad estocástica multivariada es con efecto de apalancamiento cruzado propuesto por Ishihara y Omori [53]. Los datos simulados reflejan lo preciso que es el método para estimar los parámetros, claro está que la precisión se debe a que el método muestreador de movimiento múltiple (multi move sampler) propuesto por Ishihara y Omori [53] es muy eficiente.

Se aplicó el modelo a series de tiempo bivariadas compuestas de los precios de Apple y Microsoft. Los gráficos de convergencia muestran que para cada uno de los parámetros estimados la trayectoria de la cadena exhibe el mismo comportamiento a través de las iteraciones lo que indica convergencia. Los gráficos de las funciones de autocorrelación de los elementos del parámetro  $\Sigma$  y  $\Phi$  muestran la correlación usual que se ve en la literatura.

Por otro lado, como se observó en la sección 4.5-4.6, las funciones de autocorrelación de los elementos de las matrices de coeficientes VAR muestran poca correlación. A diferencia de los elementos de las matrices de coeficientes, los gráficos de las funciones de autocorrelación de los términos de intersección alcanzan sus máximos en 0.24 para  $v_1$  y 0.249 para  $v_2$  en los datos simulados y, 0.284 para  $v_1$  y 0.273 para  $v_2$  en los resultados empíricos. Además, en ambos casos decrecen lentamente mostrando la alta dependencia de valores pasados. Esto puede ser debido a la priori de Litterman ya que las varianzas priori son infinitas para  $v_1$  y  $v_2$  reflejando que no se tiene ninguna conjetura a

priori para estos coeficientes. Los histogramas revelan que las cadenas de los parámetros estimados parecen tener distribución Gaussiana.

Los diagnósticos de convergencia muestran que el enfoque bayesiano para estimar los parámetros es eficiente pero también muestran que se puede mejorar, a saber, probando diferentes distribuciones priori para el parámetro  $\beta$ . Debido a la presencia de saltos, otra forma de mejorar el método propuesto es incorporando los errores-t de Ishihara y Omori [53], ya que la literatura sugiere que agregarlos ayuda a capturar mejor las propiedades de los retornos.

Los intervalos de credibilidad de los elementos de las matrices de coeficientes contienen el cero, por lo que existe la posibilidad de que los elementos de las matrices de coeficientes sean cero. En el caso de que los elementos de las matrices de coeficientes no sean cero, se puede observar que aparentemente el efecto de apalancamiento cruzado se ve reflejado en las matrices de coeficientes debido a que los retornos pasados de Apple alimentan sus propios retornos actuales y los retornos de Microsoft.

## 5.1. Trabajos futuros

En la realización de esta investigación se identificó los siguientes posibles trabajos futuros:

1. Probar diferentes distribuciones priori de la literatura para ver cual o cuales mejoran la inferencia del modelo propuesto.
2. Aplicar el modelo VAR-MSV del capítulo 4 a diferentes conjuntos de series de tiempo financieras, para encontrar evidencia empírica de que el efecto de apalancamiento cruzado se ve reflejado en las matrices de coeficientes estimadas, ya que con un solo conjunto de series de tiempo bivariadas no es suficiente evidencia.
3. Extender el modelo VAR-MSV para que las matrices de coeficientes junto con el término de intersección sean generados por su propios procesos estocásticos, a saber, un proceso autorregresivo AR(1), para observar si las relaciones dinámicas de los retornos tienen mas efecto condicional en algunos puntos del tiempo que en otros, por ejemplo esos puntos de



tiempo donde parece que hay más volatilidad. El proceso AR(1) de las matrices de coeficientes puede tener dos configuraciones:  $\beta_t = \beta_{t-1} + w_t$  y  $\beta_{t+1} = \beta_t + w_t$ , donde  $w_t$  es el término error. Los dos procesos AR(1) anteriores se pueden intentar correlacionar con la volatilidad de mañana. Se puede intentar correlacionar los dos procesos AR(1) anteriores con los choques de las variables endógenas. También se podría intentar correlacionar los tres choques. La distribución de los choques sería  $(w_t, \varepsilon_t, \eta_t) \sim N(0, \Sigma_{w\varepsilon\eta})$ .

4. Aplicar los diferentes métodos del punto (2) a diferentes conjuntos de datos para ver qué modelos brindan un mejor ajuste y observar como las diferentes configuraciones de los modelos capturan las propiedades de los datos.

# Bibliografía

- [1] AGUILAR, O., AND WEST, M. Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of business and economic statistics* 18 (2000), 338–357.
- [2] AHN, S. K. Distribution for residual autocovariances in multivariate autoregressive models with structured parametrization. *Biometrika* 75 (1988), 590–593.
- [3] ANDERSON, T. G., BOLLERSLEV, T., DIEBOLD, F. X., AND EBENS, H. The distribution of realized stock return volatility. *Journal of Financial Economics* 61 (2001), 43–76.
- [4] ANDERSON, T. G., BOLLERSLEV, T., DIEBOLD, F. X., AND LABYS, P. The distribution of realized exchange rate volatility. *Journal of American Statistical Association* 96 (2001), 42–55.
- [5] ASAI, M., AND MCALEER, M. Asymmetric multivariate stochastic volatility. *Econometric Reviews* 25(2-3) (2005), 453–473.
- [6] ASAI, M., MCALEER, M., AND YU, J. Multivariate stochastic volatility: a review. *Econometric reviews* 25 (2006), 145–175.
- [7] BARNDORFF-NIELSEN, O. E., AND SHEPHARD, N. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72 (2004), 885–925.
- [8] BAUWENS, L., HAFNER, C., AND LAURENT, S. *Handbook of volatility models and their applications*. Wiley Handbooks in Financial Engineering and Econometrics. John Wiley Sons, Inc, 2012.

- [9] BAUWENS, L., LAURENT, S., AND ROMBOUTS, J. V. Multivariate garch models: A survey. *Journal of Applied Econometrics* 21 (2006), 79–109.
- [10] BENATI, L. The great moderation in the united kingdom. *Journal of Money, Credit and Banking* 40(1) (2008), 121–147.
- [11] BLACK, F. Studies of stock market volatility changes. *Proceedings of the american statistical asociation, business and economic statistics* (1976), 177–181.
- [12] BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (1986), 307–327.
- [13] BOLLERSLEV, T., ENGLE, R. F., AND WOOLDRIDGE, J. M. A capital asset pricing model with time-varying covariances. *Journal of Political Economy* 96(1) (1988), 116–131.
- [14] BRANDT, P. T., AND WILLIAMS, J. T. *Multiple Time Series Models*. SAGE Publications, Inc., 2007.
- [15] BROTO, C., AND RUIZ, E. Estimation methods for stochastic volatility models: a survey. *Journal of economic surveys* 18 (2004), 613–649.
- [16] CAMPBELL, J., LO, A., AND MACKINKAY, A. *The econometrics of financial market*. Princeton, NJ: Princeton University Press, 1997.
- [17] CANOVA, F. *Methods for applied macroeconomic research*. Princeton, NJ: Princeton University Press, 2007.
- [18] CARTER, C., AND KOHN, R. On gibbs sampling for state space models. *Biometrika* 81 (1994), 541–553.
- [19] CHIB, S., NARDURI, F., AND SHEPHARD, N. Analysis of high dimensional multivariate stochastic volatility models. *Journal of econometrics* (2005). to appear in.
- [20] CHIU, C.-W. J., MUMTAZ, H., AND PINTÉR, G. Forecasting with var models: Fat tails and stochastic volatility. *International Journal of Forecasting* 33 (2017), 1124–1143. issue 4.
- [21] CHRISTIE, A. The stochastic behavior of common stock variances: value, leverage and interest, rate effects. *Journal of financial economics* 10 (1982), 407–432.

- [22] CLARCK, P. K. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41(1) (1973), 135–155.
- [23] CLARK, T. E. Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics* 29(3) (2011), 327–341.
- [24] CLARK, T. E., AND RAVAZZOLO, F. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics* 30(4) (2015), 551–575.
- [25] COGLEY, T. How fast can the new economy grow? a bayesian analysis of the evolution of trend growth. *Macroeconomics* 27 (2005), 179–207.
- [26] COGLEY, T., AND SARGENT, T. J. Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic Dynamics* 8 (2005), 262–302. issue 2.
- [27] COX, D., HINKLEY, D., AND BARNDORFF-NIELSEN, O. *Time Series Models: In econometrics, finance and other fields*. Monographs on Statistics and Applied Probability. Springer US, 2013.
- [28] D’AGOSTINO, A., GAMBETTI, L., AND GIANNONE, D. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* 28(1) (2013), 82–101.
- [29] DANIELSSON, J. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of econometrics* 64 (1994), 641–660.
- [30] DING, Q., HUANG, J., AND ZHANG, H. The time-varying effects of financial and geopolitical uncertainties on commodity market dynamics: A tvp-svar-sv analysis. *Resources Policy* 72 (2021).
- [31] DING, Z., GRANGER, C. W., AND ENGLE, R. F. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1(1) (1986), 83–106.
- [32] DOAN, T., LITTERMAN, R., AND SIMS, C. A. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3 (1984), 1–100.
- [33] DONALDSON, R. G., AND KAMSTRA, M. An artificial neural network-garch model for international stock return volatility. *Journal of Empirical Finance* 4(1) (1997), 17–46.

- [34] DOORNIK, J. *Ox: Object Oriented Matrix Programming*. Timberlake Consultants Press, London, 2006.
- [35] DUAN, J., GENTLE, J., AND HARDLE, W., Eds. *Handbook of computational finance*. DOI 10.1007/978-3-642-17254-0\_1. Springer-Verlag Berlin Heidelberg, 2012. Springer handbooks of computational statistics.
- [36] DURBIN, J., AND KOOPMAN, S. Monte carlo maximum likelihood estimation for non.gaussian state space models. *Biometrika* 84 (1997), 669–684.
- [37] DURBIN, J., AND KOOPMAN, S. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*. 89 (2002), 603–616.
- [38] ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflations. *Econometrica* 50 (1982), 987–1007.
- [39] ENGLE, R. F., AND BOLLERSLEV, T. Modelling the persistence of conditional variances. *Econometric Reviews* 5(1) (1986), 1–50.
- [40] ENGLE, R. F., AND KRONER, K. F. Multivariate simultaneous generalized arch. *Econometric Theory* 11(1) (1995), 122–150.
- [41] FORNARI, F., AND MELE, A. Sing-and volatility-switching arch models: Theory and applications to international stock market. *Journal of Applied Econometrics* 12(1) (1997), 49–65.
- [42] FRANCO, C., AND ZAKOIAN, J.-M. *GARCH Models: Structure, Statistical Inference and Financial Applications, 2nd Edition*. John Wiley & Sons, 2019.
- [43] FRENCH, K. R., SCHWERT, G. W., AND STAMBAUGH, R. F. Expected stock returns and volatility. *Journal of Financial Economics* 19 (1987), 3–29.
- [44] FULLER, W. *Introduction to statistical time series*. John Wiley, New York, 1976.
- [45] GALI, J., AND GAMBETTI, L. On the sources of the great moderation. *American Economic Journal: Macroeconomics* 1(1) (2009), 26–57.
- [46] GALLANT, A., AND TAUCHEN, G. Which moments to match? *Econometric theory* 12 (1996), 657–681.

- [47] GLOSTEN, L., JAGANNATHAN, R., AND RUNKLE, D. On the relation between the expected value and volatility of nominal excess returns on stocks. *Journal of finance* 46 (1993), 1779–1801.
- [48] GOURIÉOUX, C. *ARCH Models and Financial Applications*. New York: Springer, 1997.
- [49] HAMILTON, J. *Time Series Analysis*. Princeton University Press, 1994.
- [50] HARVEY, A., RUIZ, E., AND SHEPPARD, N. Multivariate stochastic variance models. *Review of economic studies* 61 (1994), 247–264.
- [51] HARVEY, A., AND SHEPHARD, N. Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of business and economic statistics* 14 (1996), 429–434.
- [52] HOSKING, J. Lagrange-multiplier test of multivariate time series model. *Journal of the Royal Statistical Society* 43, series B (1981), 219–230.
- [53] ISHIHARA, T., AND OMORI, Y. Efficient bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors. *Computational statistics and data analysis* 56(11) (2012), 3674–9473.
- [54] JACQUIER, E., POLSON, N., AND ROSSI, P. Models and prior distributions for multivariate stochastic volatility. *University of Chicago* (1995). unpublished paper.
- [55] JACQUIER, E., POLSON, N., AND ROSSI, P. Bayesian analysis of stochastic volatility models with fat tails and correlated errors. *Journal of econometrics* 122 (2004), 185–212.
- [56] JACQUIER, E., POLSON, N. G., AND ROSSI, P. Bayesian analysis of stochastic volatility models. *Journal of Business Economic Statistics* 12(4) (1994), 371–389.
- [57] JONG, P., AND SHEPHARD, N. The simulation smoother for time series models. *Biometrika*. 82 (1995), 339–350.
- [58] KADIYALA, K. R., AND KARLSSON, S. Numerical methods for estimation and inference in bayesian var models. *Journal of applied econometrics* 12 (1997), 99–132.
- [59] KIM, S., SHEPPARD, N., AND SCHIB, S. Stochastic volatility: likelihood inference and comparisson with arch models. *Review of economic studies* 65 (1998), 361–393.

- [60] KOOPMAN, S. Disturbance smoother for state space models. *Biometrika*. 80 (1993), 117–126.
- [61] LI, W., AND LAM, K. Modelling asymmetry in stock returns by a threshold arcg model. *The Statistician* 44 (1995), 333–341.
- [62] LI, W., AND MCLEOD, A. Distribution of the residual autocorrelations in multivariate time series models. *Journal of the Royal Statistical Society* 43, series B (1981), 231–239.
- [63] LIESENFELD, R., AND RICHARD, J.-F. Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *Journal of Empirical Finance* 10(4) (2003), 505–531.
- [64] LITTERMAN, R. Forecasting with bayesian vector autoregressions: Five years of experience. *Journal of Business and Economic Statistics*. 4 (1986), 25–38.
- [65] LUTKEPOHL, H. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.
- [66] MANN, H., AND WALD, A. On the statistical treatment of linear stochastic difference equations. *Econometrica* 11 (1943), 173–220.
- [67] MELINO, A., AND TURNBULL, S. M. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45(1-2) (1990), 239–265.
- [68] MUMTAZ, H. A generalized stochastic volatility in mean var. *Economice Letters* 173 (2018), 10–14.
- [69] MUMTAZ, H. A Generalised Stochastic Volatility in Mean VAR. An Updated Algorithm. Working Papers 908, Queen Mary University of London, School of Economics and Finance, July 2020.
- [70] NAKAJIMA, J. Time-varying parameter var model with stochastic volatility: An overview of methodology and empirical applications. *Monetary and Economic Studies* 29 (2011), 107–142.
- [71] NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59(2) (1991), 347–370.

- [72] NEWEY, W. K., AND MCFADDEN, D. *Large sample estimation and hypothesis testing*. Elsevier Science, 1994. Handbook of econometrics, vol. IV, capítulo 36.
- [73] NGUYEN, T. A new efficient algorithm for the analysis with non-linear and non-gaussian state space models. *University Amsterdam, Department of econometrics* (2007).
- [74] OMORI, Y., AND ISHIHARA, T. *Multivariate stochastic volatility models*. John Wiley & sons, 2012. Handbook of volatility models and their applications edited by L. Bauwens, C. Hofner and S. Laurent.
- [75] OMORI, Y., SHEPHARD, N., AND NAKAJIMA, J. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics* 140(2) (2007), 425–449.
- [76] OMORI, Y., AND WATANABE, T. Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics and Data Analysis*. 52(6) (2008), 2892–2910.
- [77] PITT, M., AND SHEPHARD, N. Time varying covariances: a factor stochastic volatility approach with discussion. *Bayesian statistics 6* (1999), 547–570. Bernardo J.M. and Berger J.O. and David A.P. and Smith A.F.M.
- [78] PRIMICERI, G. C. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies* 72 (2005), 821–852. issue 3.
- [79] QUINTANA, J., AND WEST, M. An analysis of international exchange rates using multivariate dlms. *The statistician* 36 (1987), 275–281.
- [80] RICHARD, J., AND ZHANG, W. Efficient high-dimensional importance sampling. *Journal of econometrics* 141 (2007), 1385–1411.
- [81] ROBERT ENGLE, V. K. N., AND ROTHSCHILD, M. Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics* 45 (1990), 213–237.
- [82] ROSS, S. The arbitrage theory of asset pricing. *Journal of economic theory* 13 (1976), 641–660.
- [83] SHEPHARD, N. *Stochastic Volatility: Selected Readings*. Oxford University Press, 2005.



- [84] SHEPHARD, N., AND K.PITT, M. *Time varying covariances: a factor stochastic volatility approach*, (edited by j.m. bernardo, j.o. berger, a.p. dawid and a.f.m smith) ed. Oxford University Press, Oxford, 1999, pp. 547–570.
- [85] SHEPHARD, N., AND PITT, M. Likelihood analysis of non-gaussian measurement time series. *Biometrika* 84 (1997), 653–667.
- [86] SIMS, C. A. Money, income, and causality. *American Economic Review* 62 (4) (1972), 540–552.
- [87] SIMS, C. A. The role of approximate prior restrictions in distributed lag estimation. *Journal of the American Statistical Association* 67 (337) (1972), 169–175.
- [88] SIMS, C. A. Comparison of interwar and postwar business cycles: Monetarism reconsidered. *American Economic Review*. 70 (2) (1980), 250–257.
- [89] SIMS, C. A. Macroeconomics and reality. *Econometrica* 48(1) (1980), 1–48.
- [90] SO, M., AND CHOI, C. A multivariate threshold stochastic volatility model. *Mathematics and computers in simulation (MATCOM)* 79 (2008), 306–317.
- [91] SO, M., LI, W., AND LAM, K. A threshold stochastic volatility model. *Journal of forecasting* 21 (2002), 473–500.
- [92] SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P., AND VAN DER LINDE, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 4 (2002), 583–639.
- [93] TAUCHEN, G. E., AND PITT, M. The price variability- volume relationship on speculative markets. *Econometrica* 51(2) (1983), 485–505.
- [94] TAYLOR, S. *Financial returns modelled by the product of two stochastic processes. A study of daily sugar prices 1961-79*. O.D. Anderson, Amsterdam, North Holland, 1982.
- [95] TAYLOR, S. J. Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* 4 (1994), 183–204.
- [96] TONG, H. *Threshold models un nonlinear time series analysis*. Springer-Verlag: New York, 1983.

- [97] TONG, H. *Nonlinear time series: a dynamical systems approach*. Oxford university press, 1990.
- [98] TONG, H., AND LIM, K. Threshold autoregression limit cycles and cyclical data (with discussion). *Journal of the royal statistical society* 42 (1980), 245–292.
- [99] TRIANTAFYLLOPOULOS, K. Time-varying vector autoregressive models with volatility. *Journal of Applied Statistics* 38(2) (2011), 369–382.
- [100] TSAY, R. S. *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons, 2013.
- [101] UHLIG, H. Bayesian vector autoregressions with stochastic volatility. *Econometrica* 65(1) (1984), 59–72.
- [102] VRONTOS, I. D., DELLAPORTAS, P., AND POLITIS, D. N. A full-factor multivariate garch model. *The Econometrics Journal* 6(2) (2003), 312–334.
- [103] YU, J. On leverage in a stochastic volatility model. *Journal of econometrics* 127 (2005), 165–178.
- [104] YU, J., AND MEYER, R. Multivariate stochastic volatility models; bayesian estimation and model comparison. *Econometric reviews* 25 (2006), 361–384.
- [105] ZAKOIAN, J.-M. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18(5) (1994), 931–955.