



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE CIENCIAS

Modelos de Volatilidad Multivariados

T E S I S

QUE PARA OBTENER EL TÍTULO DE:  
ACTUARIO

PRESENTA:  
ROBERTO BÁRCENAS CURTIS

DIRECTOR DE TESIS:  
DRA. RUTH SELENE FUENTES GARCÍA

2009





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

<p>1. Datos del alumno</p> <p>Bárceñas Curtis Roberto 56 44 42 08 Universidad Nacional Autónoma de México Facultad de Ciencias Actuaría 099035736</p>
<p>2. Datos del tutor</p> <p>Dra. Ruth Selene Fuentes García</p>
<p>3. Datos del sinodal 1</p> <p>Dra. Silvia Ruiz-Velasco Acosta</p>
<p>4. Datos del sinodal 2</p> <p>Mat. Margarita Elvira Chávez Cano</p>
<p>5. Datos del sinodal 3</p> <p>Dr. Alberto Contreras Cristan</p>
<p>6. Datos del sinodal 4</p> <p>Act. Miguel Angel Chong Rodriguez</p>
<p>7. Datos del trabajo escrito</p> <p>Modelos de Volatilidad Multivariados 119 p 2009</p>

# Agradecimientos

A mi tutora y a cada uno de mis sinodales por la revisión del trabajo doy mención especial.

Agradezco:

A la Dra. Ruth por la impartición de los cursos Series de Tiempo y Análisis Multivariado, motivando la inquietud de encontrar el tema ideal para relacionarlos. Su atención y supervisión han sido fundamentales para la realización de esta tesis.

A la Dra. Silvia las apreciables sugerencias y su ayuda para mejorar la presentación del texto en L<sup>A</sup>T<sub>E</sub>X.

Al Dr. Alberto por haberme dado la oportunidad de colaborar con él y todas las valiosas ideas que permitieron concluir los Capítulos 2 y 3.

A Miguel Angel por el tiempo invertido en la lectura y sus oportunas observaciones.

A la Dra. Margarita sus amables comentarios y correcciones en la escritura.

A mi familia agradezco todo.

A mis padres Hilario y Magdalena por su inagotable apoyo y amor, gracias.

A mis hermanos Elena y César su cálida compañía, con cariño.

*“La única razón para la existencia del tiempo  
es que no ocurra todo a la vez.”*

Albert Einstein

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Modelos de Volatilidad Multivariados . . . . .	1
1.2. Desarrollo de la tesis . . . . .	2
<b>2. Series de Tiempo</b>	<b>3</b>
2.1. Procesos Estacionarios . . . . .	3
2.1.1. Media, Autocovarianza y Autocorrelación Muestrales . . . . .	7
2.2. Procesos ARMA . . . . .	8
2.2.1. Procesos MA . . . . .	10
2.2.2. Procesos AR . . . . .	12
2.2.3. ARMA(1,1) . . . . .	15
2.3. Series No Estacionarias . . . . .	17
2.3.1. Caminata Aleatoria . . . . .	17
2.3.2. Procesos ARIMA . . . . .	18
<b>3. Modelos con Heteroscedasticidad Condicional</b>	<b>19</b>
3.1. Propiedades de la volatilidad . . . . .	19
3.1.1. Especificación del Modelo de Volatilidad . . . . .	22
3.2. ARCH . . . . .	24
3.2.1. ARCH(1) . . . . .	27
3.2.2. Efectos ARCH . . . . .	29
3.3. GARCH . . . . .	30

3.3.1.	ACF y PACF para procesos GARCH . . . . .	32
3.3.2.	GARCH(1,1) . . . . .	33
3.4.	Estimación de los modelos GARCH . . . . .	35
3.4.1.	Máxima y Cuasi Máxima Verosimilitud . . . . .	35
3.4.2.	Student-t . . . . .	37
3.4.3.	GED . . . . .	37
3.4.4.	Criterios de Información: AIC y BIC . . . . .	38
3.5.	ARCH-M . . . . .	39
3.6.	IGARCH . . . . .	40
3.7.	EGARCH . . . . .	42
3.7.1.	EGARCH(1,0) . . . . .	44
<b>4.</b>	<b>Modelos MGARCH</b>	<b>45</b>
4.1.	Generalizaciones directas GARCH . . . . .	47
4.1.1.	VEC . . . . .	47
4.1.2.	BEKK . . . . .	51
4.2.	Modelos de Factores . . . . .	53
4.2.1.	F-GARCH . . . . .	55
4.2.2.	O-GARCH . . . . .	58
4.2.3.	GO-GARCH . . . . .	59
4.2.4.	FF-GARCH . . . . .	60
4.2.5.	GOF-GARCH . . . . .	61
4.3.	Modelos de Correlación Condicional . . . . .	63
4.3.1.	Correlación Condicional Constante . . . . .	63
4.3.2.	Correlación Condicional Dinámica . . . . .	65
4.4.	Estimación MGARCH . . . . .	67
<b>5.</b>	<b>Aplicación MGARCH</b>	<b>69</b>
5.1.	Series de Tipo de Cambio . . . . .	69
5.1.1.	Análisis de los Datos . . . . .	71

---

5.2. Modelos Bivariados . . . . .	74
5.2.1. VEC Diagonal . . . . .	75
5.2.2. BEKK . . . . .	77
5.2.3. O-GARCH . . . . .	78
5.2.4. CCC . . . . .	81
5.2.5. Resultados . . . . .	84
<b>6. Conclusiones</b>	<b>85</b>
<b>A. Demostraciones</b>	<b>87</b>
<b>B. Análisis Multivariado</b>	<b>89</b>
<b>C. Tablas y Gráficas Adicionales en S-PLUS</b>	<b>99</b>
C.1. Series Univariadas . . . . .	99
C.1.1. Volatilidad . . . . .	101
C.2. Modelos GARCH Bivariados . . . . .	103
C.2.1. DVEC . . . . .	103
C.2.2. BEKK . . . . .	105
C.2.3. O-GARCH . . . . .	107
C.2.4. CCC . . . . .	110
<b>Bibliografía</b>	<b>117</b>

# Capítulo 1

## Introducción

### 1.1. Modelos de Volatilidad Multivariados

La *volatilidad* es una variable clave usada en diversas aplicaciones financieras, en ese contexto se entiende como la varianza condicional del rendimiento de un instrumento financiero dada la información disponible del pasado.

Engle [1982] en su modelo autoregresivo con heteroscedasticidad condicional (ARCH) propuso una forma sistemática para analizar la volatilidad, Bollerslev [1986] introdujo los procesos GARCH, refiriéndose a la generalización ARCH. De ellos se desprenden un gran número de variaciones, cuyo objetivo es pronosticar la volatilidad considerando ciertas condiciones específicas y asimetría, de las cuales las más influyentes han sido las primeras: ARCH en media (ARCH-M) de Engle, Lilien y Robins [1987], GARCH integrado (IGARCH) por Bollerslev y Engle [1986] y GARCH exponencial (EGARCH) por Nelson [1991].

Si bien los modelos anteriores son útiles para estudiar la volatilidad en el caso de rendimientos de un solo activo, cuando el interés se centra en observar su comportamiento en un portafolio de inversión es necesario estimar las correlaciones y covarianzas entre cada uno de los instrumentos que lo componen.

En la actualidad se ha encontrado evidencia empírica de que las volatilidades se mueven conjuntamente a través del tiempo en los diferentes mercados financieros, por ejemplo en el caso de las tasas de interés con el tipo de cambio en divisas. A partir de esta situación, el introducir la noción multivariada permite enriquecer los modelos conocidos y obtener mayor información que al trabajar con ajustes univariados por separado.

Además ante los acontecimientos ocurridos en el sector económico de los Estados Unidos durante los años 2007 y 2008, donde la volatilidad ha jugado un papel fundamental y su

repercusión es notable en la atmósfera global, se ha hecho evidente la necesidad de nuevas herramientas que mejoren la toma de decisiones en áreas tales como selección de portafolios, administración de riesgos y valuación de instrumentos financieros.

En el presente trabajo se abordarán los modelos de volatilidad multivariados, surgiendo naturalmente como una extensión de los procesos ARCH/GARCH para analizar series de datos financieros y económicos.

Así, para nosotros el desarrollo de los modelos multivariados GARCH (**MGARCH**) se hará principalmente en tres enfoques:

El primero referido a la generalización directa del modelo univariado GARCH al contexto multivariado, incluyendo los modelos **VEC**, VEC diagonal (**DVEC**), ambos de Bollerslev, Engle y Wooldridge [1988] y **BEKK** por Engle y Kroner [1995].

El segundo basado en la idea de factores, dentro del cual encontramos las aplicaciones de Diebold y Nerlove [1989], de Engle, Ng y Rothschild [1990] y los modelos: Factores GARCH (**F-GARCH**) por Lin [1992], Factores Completos GARCH (**FF-GARCH**) de Vrontos, Dellaportas y Pollitis [2003], GARCH Ortogonal (**O-GARCH**) de Alexander [2001], GARCH Ortogonal Generalizado (**GO-GARCH**) de van der Weide [2002] y el modelo de Factores GARCH Ortogonal Generalizado (**GOF-GARCH**) de Lanne y Saikkonen [2007].

Y un tercer enfoque que considera la descomposición a través de la matriz de correlación, teniéndose los modelos de Correlación Condicional Constante (**CCC**) de Bollerslev [1990] y Correlación Condicional Dinámica (**DCC**) por Engle [2002] y Tse y Tsui [2002].

## 1.2. Desarrollo de la tesis

En el Capítulo 2 se introducen los fundamentos de series de tiempo que serán empleados a lo largo del trabajo, remarcando los procesos estacionarios. Se incluyen los modelos autoregresivos (AR), de promedios móviles (MA) y la combinación de ambos (ARMA) con sus propiedades fundamentales.

La teoría de heteroscedasticidad condicional se desarrolla en el tercer capítulo, procurando ilustrar las ideas que motivaron su surgimiento y discutiendo sus especificaciones relevantes: ARCH, GARCH, ARCH-M, IGARCH y EGARCH.

El Capítulo 4 se centra en los modelos de volatilidad multivariados GARCH (MGARCH), de acuerdo a las referencias mencionadas anteriormente.

Finalmente en el Capítulo 5 se realiza la aplicación de un modelo MGARCH para un caso bivariado, considerando los tipos de cambio diarios Peso/Dólar y Peso/Euro del mercado de divisas desde Octubre de 2002 hasta Septiembre de 2008.

## Capítulo 2

# Series de Tiempo

### 2.1. Procesos Estacionarios

Una serie de tiempo es un conjunto de observaciones, generalmente ordenadas en el tiempo.

Diariamente en la vida cotidiana es posible encontrar diferentes ejemplos, tales como el precio de una acción en el mercado de valores o la tasa de inflación en la economía, aunque por supuesto también se presentan en fenómenos físicos, demográficos y en muchas otras áreas de estudio.

Introduciendo una notación matemática, una serie de tiempo puede representarse como una familia de variables  $x_t$ , cada una registrada en un tiempo específico  $t$ .

El análisis de series de tiempo tiene como objetivo el estudiar la dinámica del proceso involucrado, describir su comportamiento a través de cierta representación y realizar pronósticos para observaciones futuras.

Originalmente se ha dividido en dos orientaciones: del *dominio de las frecuencias*, basada en la descomposición espectral y empleando la representación de Fourier; y del *dominio en el tiempo*, que utiliza la función de autocorrelación y modelos paramétricos para describir la dependencia de la serie. La segunda será considerada en nuestro trabajo.

Por su naturaleza, las observaciones sucesivas generalmente no son independientes entre sí, pues en la mayoría de los casos debe tomarse en cuenta el orden en que fueron recolectadas. Ahora, si con los valores anteriores de la serie es posible predecir valores futuros de forma exacta, se dice que es *determinista*; pero si solamente parte del futuro proviene del conocimiento del pasado, es llamada *estocástica*.

Una serie de tiempo se dice *continua*, si las observaciones se tienen en todo momento, es

decir, sin pausas. En cambio se denomina *discreta* cuando las observaciones se realizan en tiempos definidos y separados a la misma distancia. En la práctica, la mayoría de las series solamente pueden observarse en tiempos determinados, sin embargo, se hace referencia a ellas como continuas, en el sentido del espacio donde toman sus valores. Específicamente, las series de rendimientos son consideradas continuas porque el espacio donde se encuentran son los números reales.

El primer paso para el tratamiento de series de tiempo será la selección adecuada de un modelo matemático para los datos, y con la finalidad de permitir el comportamiento impredecible de futuras observaciones, es natural suponer que cada  $x_t$  es la realización de cierta variable aleatoria  $X_t$ , requiriéndose la noción de proceso estocástico.

Un *proceso estocástico* es una colección de variables aleatorias  $\{X_t, t \in T\}$  definidas en un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ <sup>1</sup>. Así, la serie de tiempo es entonces la realización de una familia de variables aleatorias.

En este caso se sugiere modelar los datos como una trayectoria única  $\{x_t, t \in T_0\}$  de un proceso estocástico  $\{X_t, t \in T\}$   $T_0 \subseteq T$ , donde  $T$  es el conjunto de índices que representa puntos en el tiempo, regularmente  $\{0, \pm 1, \pm 2, \dots\}, \{1, 2, 3, \dots\}, [0, \infty]$  ó  $[-\infty, \infty]$ .

Nótese que para cada  $t \in T$ ,  $X_t(\cdot)$  es una función en  $\Omega$ ; por otra parte, para cada  $\omega$  fija en  $\Omega$ ,  $X(\omega)$  es de hecho una función en  $T$ .

Las funciones  $\{X(\omega), \omega \in \Omega\}$  en  $T$  se conocen como realizaciones o trayectorias del proceso  $\{X_t, t \in T\}$ , el cual en lo sucesivo se denotará solamente como  $\{X_t\}$ .

Sea  $\mathbf{T}$  el conjunto de todos los vectores  $\{\mathbf{t} = (t_1, \dots, t_n)' \in \mathbf{T}^n : t_1 < t_2 < \dots < t_n, n = 1, 2, \dots\}$ , entonces las funciones de distribución finito dimensionales de  $\{X_t\}$  son las funciones  $\{F_{\mathbf{t}}(\cdot), \mathbf{t} \in \mathbf{T}\}$  definidas para  $\mathbf{t} = (t_1, \dots, t_n)'$  por

$$F_{\mathbf{t}}(x) = \mathbb{P}(x_{t_1} \leq x_1, \dots, x_{t_n} \leq x_n), \quad x = (x_1, \dots, x_n)' \in \mathbb{R}^n.$$

Enseguida se presentarán algunas definiciones importantes del enfoque de series en el dominio del tiempo, las cuales servirán como piedra angular para el desarrollo de los siguientes capítulos.

**Definición 1 (Estacionariedad Estricta)** Un proceso  $\{X_t\}$  se dice estrictamente estacionario si las distribuciones conjuntas de  $(X_{t_1}, \dots, X_{t_k})'$  y  $(X_{t_1+h}, \dots, X_{t_k+h})'$  son iguales para todo  $k$  entero positivo y todos  $t_1, \dots, t_k, h \in \mathbb{Z}$ . Es equivalente a establecer que  $(X_1, \dots, X_k)'$  y  $(X_{1+h}, \dots, X_{k+h})'$  tienen la misma función de distribución conjunta para todos  $k, h$  enteros positivos.

<sup>1</sup>Un conjunto  $\Omega$ , una  $\sigma$ -álgebra  $\mathcal{F}$  de subconjuntos de  $\Omega$ , y una medida de probabilidad  $\mathbb{P}$ .

La definición de estacionariedad estricta implica que la estructura probabilística del proceso no cambia al correr el tiempo.

Cuando se tiene un número finito de variables aleatorias es útil calcular la matriz de covarianza para identificar la dependencia entre ellas, en el caso de las series de tiempo se requiere ampliar el concepto y lidiar con colecciones infinitas de variables aleatorias, encontrando la función de autocovarianza como tal extensión.

**Definición 2 (Función de Autocovarianza)** Si  $\{X_t\}$  es un proceso de tal forma que  $Var(X_t) < \infty$  para cada  $t \in T$ , entonces la función de autocovarianza  $\gamma_X(\cdot, \cdot)$  de  $\{X_t, t \in T\}$  se define por

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = \mathbb{E}\left[(X_r - \mathbb{E}[X_r])(X_s - \mathbb{E}[X_s])\right], \quad r, s \in T.$$

Dada la función de autocovarianza es posible relajar el supuesto de estacionariedad estricta, a una estacionariedad en un sentido más débil.

**Definición 3 (Estacionariedad de Segundo Orden)** Un proceso  $\{X_t\}$  con conjunto de índices  $T = \mathbb{Z}$ , se dice estacionario de segundo orden o débilmente estacionario si

1.  $\mathbb{E}|X_t|^2 < \infty, \forall t \in \mathbb{Z}$ ,
2.  $\mathbb{E}[X_t] = \mu, \forall t \in \mathbb{Z}$ ,
3.  $\gamma_X(r, s) = \gamma_X(r + t, s + t), \forall r, s, t \in \mathbb{Z}$ .

Es decir, un proceso será estacionario débilmente, si posee momentos de segundo orden finitos, y si su media y función de autocovarianza no dependen del tiempo.

Nótese que si  $\{X_t\}$  es un proceso estacionario de segundo orden entonces

$$\gamma_X(r, s) = \gamma_X(r - s, 0), \forall r, s \in \mathbb{Z},$$

permitiendo redefinir la función de autocovarianza de un proceso estacionario de segundo orden como función de una sola variable como

$$\gamma(h) \equiv \gamma_X(h, 0) = \text{Cov}(X_{t+h}, X_t), \forall t, h \in \mathbb{Z}.$$

La función  $\gamma(\cdot)$  se referirá a la función de autocovarianza del proceso  $\{X_t\}$  y  $\gamma(h)$  a su valor en el rezago  $h$ .

De la misma manera se define la función de autocorrelación del proceso  $\{X_t\}$  como la función cuyo valor en el rezago  $h$  es

$$\rho(h) \equiv \frac{\gamma(h)}{\gamma(0)} \quad \forall t, h \in \mathbb{Z}.$$

**Proposición 1** La función de autocovarianza  $\gamma(\cdot)$  tiene las siguientes propiedades:

1.  $\gamma(0) \geq 0$ ,
2.  $|\gamma(h)| \leq \gamma(0) \quad \forall h \in \mathbb{Z}$ ,
3.  $\gamma(h) = \gamma(-h), \forall h \in \mathbb{Z}$ .

*Demostración*

1. Claramente del hecho de que  $\gamma(0) = \text{Cov}(X_{t+0}, X_t) = \text{Var}(X_t) \geq 0$ .
2. Usando la desigualdad de Cauchy-Schwarz

$$|\gamma(h)| = |\text{Cov}(X_{t+h}, X_t)| \leq (\text{Var}(X_{t+h}))^{1/2} (\text{Var}(X_t))^{1/2}.$$

3. Es la propiedad de simetría, observando que

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

Así, por la propiedad 3, la función de autocovarianza se trata de una función par.

Por su parte la función de autocorrelación hereda esa propiedad, es decir,  $\rho(h) = \rho(-h)$ , cumpliendo además  $|\rho(h)| \leq 1$ .

**Definición 4 (Proceso Gaussiano)** El proceso  $\{X_t\}$  es un proceso Gaussiano si y sólo si las funciones de distribución finito dimensionales de  $\{X_t\}$  son todas normales multivariadas.

De lo anterior se pueden desprender las siguientes implicaciones:

- Un proceso estrictamente estacionario con momentos de segundo orden finitos es débilmente estacionario, pero, no necesariamente a la inversa.
- Considerando un proceso  $\{X_t\}$  estacionario de segundo orden Gaussiano, entonces se trata también de un proceso estacionario estrictamente, pues para todo  $n \in \{1, 2, \dots\}$  y para todos  $h, t_1, t_2, \dots \in \mathbb{Z}$ , los vectores aleatorios  $(X_{t_1}, \dots, X_{t_n})$  y  $(X_{t_1+h}, \dots, X_{t_n+h})$  tienen la misma media y varianza.

Asimismo, es conveniente definir el proceso de *ruido blanco*, que se encontrará de manera natural en los modelos a tratar más adelante.

**Definición 5 (Ruido Blanco)** Una sucesión  $\{W_t\}$  de variables aleatorias independientes e idénticamente distribuidas con media cero y varianza  $\sigma_w^2$ , se dice que es *ruido blanco* con media cero y varianza  $\sigma_w^2$ , escrito por  $\{W_t\} \sim \text{RB}(0, \sigma_w^2)$  y cuya función de autocovarianza es

$$\gamma(h) = \text{Cov}(W_{t+h}, W_t) = \begin{cases} \sigma_w^2 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases} .$$

### 2.1.1. Media, Autocovarianza y Autocorrelación Muestrales

Hemos visto que un proceso débilmente estacionario  $\{X_t\}$  se caracteriza, al menos desde el punto de vista de segundo orden, por su media y autocovarianza, por lo que la estimación de  $\mu$ ,  $\gamma(\cdot)$  y  $\rho(\cdot) = \frac{\gamma(\cdot)}{\gamma(0)}$  a partir de las observaciones  $x_1, \dots, x_n$  será de utilidad en la inferencia y particularmente en la tarea de construir un modelo apropiado para los datos.

El estimador de la media  $\mu$  de un proceso estacionario es la media muestral

$$\bar{x}_n = \frac{1}{n} \sum_{t=1}^n x_t, \text{ donde } n \text{ es el tamaño de la muestra.}$$

La función de autocovarianza muestral es

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x}_n)(x_t - \bar{x}_n).$$

De donde la función de autocorrelación muestral se define como

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x}_n)(x_t - \bar{x}_n)}{\sum_{t=1}^n (x_t - \bar{x}_n)^2}.$$

Ambos estimadores (de autocovarianza y autocorrelación) son sesgados, aún cuando  $1/n$  es reemplazado por el factor  $1/(n-h)$ , pero, bajo supuestos generales son cercanamente insesgados para tamaños de muestra grandes.

Para cada  $h \geq 1$ , la matriz de covarianza muestral  $k$  dimensional no negativa definida se establece como

$$\hat{\Gamma}_h = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(h-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(h-2) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}(h-1) & \hat{\gamma}(h-2) & \cdots & \hat{\gamma}(0) \end{bmatrix}.$$

Por consecuencia la matriz de autocorrelación muestral dada por  $\hat{R}_h = \hat{\Gamma}_h / \hat{\gamma}(0)$  es también no negativa definida.

Quenouille [1949] establece que para  $n$  grande, las funciones de autocorrelación muestrales de una sucesión de variables aleatorias  $Y_1, \dots, Y_n$  independientes e idénticamente distribuidas (i.i.d.) con varianza finita son aproximadamente i.i.d. con distribución  $N(0, 1/n)$ . Por lo tanto, si  $y_1, \dots, y_n$  es una realización de una sucesión i.i.d., considerando un rezago  $h$  fijo, el 95 % de las autocorrelaciones muestrales del mismo en  $k$  realizaciones, se encontrarán dentro de las bandas de confianza  $\pm 1.96/\sqrt{n}$ .

Entonces, para determinar si un conjunto muestral de  $n$  observaciones son no correlacionadas, Box y Pierce [1970] propusieron la prueba de *Portmanteau* al rezago  $m$  para probar la hipótesis nula  $H_0 : \rho(1) = \cdots = \rho(m) = 0$  contra la hipótesis alternativa  $H_a : \rho(i) \neq 0$  para alguna  $i \in \{1, \dots, m\}$ . El estadístico de prueba es

$$Q^*(m) = n \sum_{l=1}^m \hat{\rho}^2(l).$$

Ljung y Box [1978] modificaron el estadístico  $Q^*(m)$  en muestras finitas como

$$Q(m) = n(n+2) \sum_{l=1}^m \hat{\rho}^2(l)/(n-l).$$

En cada caso la regla de decisión es rechazar  $H_0$  al nivel de significancia  $\alpha$ , si el estadístico de prueba es mayor que el  $100(1-\alpha)$  percentil de una distribución Chi-cuadrada ( $\chi_\alpha^2$ ) con  $m$  grados de libertad. O en caso de contar con el p-value de la prueba, rechazar la hipótesis nula si el p-value es menor o igual que el nivel de significancia, por lo regular  $\alpha = 0.01$  ó  $\alpha = 0.05$ .

## 2.2. Procesos ARMA

Los procesos *autoregresivos de promedios móviles* o ARMA pertenecen a una clase importante de series de tiempo y desde su aparición han conformado una familia paramétrica de procesos estacionarios que desempeña un papel fundamental en la modelación de series

de datos, además de que su estructura lineal proporciona una forma práctica para poder realizar pronósticos.

**Definición 6 (ARMA( $p, q$ ))** El proceso  $\{X_t\}$ , es un proceso autoregresivo de promedios móviles de ordenes  $p$  y  $q$ , o bien ARMA( $p, q$ ), si  $\{X_t\}$  es estacionario (de segundo orden) y si para cada  $t$  se tiene

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q}, \quad (2.1)$$

con  $\{W_t\} \sim \text{RB}(0, \sigma_w^2)$ .

La ecuación 2.1 puede describirse de forma más compacta como

$$\phi(B)X_t = \theta(B)W_t, \quad (2.2)$$

donde  $\phi(\cdot)$  y  $\theta(\cdot)$  son los polinomios de grado  $p$  y  $q$

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

y

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q,$$

tales que no tienen raíces en común y  $B$  es el operador de retraso definido por  $B^j X_t = X_{t-j}$ . Generalmente los polinomios  $\phi(z)$  y  $\theta(z)$  son referidos como los polinomios autoregresivo y de promedios móviles respectivamente.

Como casos particulares de procesos ARMA, se encuentran los procesos de *promedios móviles* (MA( $q$ )) y *autoregresivos* (AR( $p$ )).

**Definición 7 (MA( $q$ ))** El proceso  $\{X_t\}$ , es un proceso de promedios móviles de orden  $q$ , MA( $q$ ), si

$$X_t = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q},$$

donde  $W_t \sim \text{RB}(0, \sigma_w^2)$  y  $\theta_1, \dots, \theta_q$  son constantes.

**Definición 8 (AR( $p$ ))**  $\{X_t\}$  se dice ser un proceso de autoregresivo de orden  $p$ , AR( $p$ ), si

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + W_t,$$

donde  $W_t \sim \text{RB}(0, \sigma_w^2)$  y  $\phi_1, \dots, \phi_p$  son constantes.

También son de mencionarse dos conceptos importantes relacionados con la expresión de los procesos ARMA, primero como un proceso de promedios móviles, y por otra parte como la representación del *ruido* en términos de un proceso autoregresivo. Tales son *causalidad* e *invertibilidad*.

**Definición 9 (Causalidad)** Un proceso ARMA( $p, q$ )  $\{X_t\}$  se dice que es causal (independiente del futuro) o una función causal de  $\{W_t\}$  si existen constantes  $\{\psi_j\}$  tales que  $\sum_{j=1}^{\infty} |\psi_j| < \infty$  y  $X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j} \quad \forall t$ , lo cual es equivalente a la condición

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0, \quad \forall |z| \leq 1.$$

La propiedad de causalidad, indica que el proceso  $\{X_t\}$  puede escribirse como una combinación lineal de ruido, esto quiere decir, que es estacionario.

**Definición 10 (Invertibilidad)** Un proceso ARMA( $p, q$ )  $\{X_t\}$  es invertible si existen constantes  $\{\pi_j\}$  tales que  $\sum_{j=1}^{\infty} |\pi_j| < \infty$  y  $W_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \quad \forall t$ , equivalentemente

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q \neq 0, \quad \forall |z| \leq 1.$$

Es semejante a causalidad, pero, ahora en el caso del proceso de ruido  $W_t$ , el cual podrá verse en términos de la parte autoregresiva y garantizará una estructura única para la autocovarianza del proceso.

### 2.2.1. Procesos MA

En esta parte se desarrollan los procesos de promedios móviles MA(1) y MA( $q$ ), encontrando las funciones de autocovarianza y autocorrelación de cada uno de ellos.

Considérese un proceso  $\{X_t\}$  de promedios móviles de orden 1, MA(1), representado por

$$X_t = W_t + \theta W_{t-1}, \quad W_t \sim \text{RB}(0, \sigma_w^2).$$

Tomando la esperanza del mismo, se tiene que

$$\mathbb{E}[X_t] = \mathbb{E}[W_t + \theta W_{t-1}] = \mathbb{E}[W_t] + \theta \mathbb{E}[W_{t-1}] = 0.$$

Usando este resultado su varianza es

$$\gamma(0) = \text{Var}[X_t] = \mathbb{E}[X_t^2] = (1 + \theta^2)\sigma_w^2.$$

Calculando la función de autocovarianza se obtiene

$$\begin{aligned}
\gamma(h) &= \text{Cov}(X_t, X_{t+h}) \\
&= \text{Cov}(W_t + \theta W_{t-1}, W_{t+h} + \theta W_{t+h-1}) \\
&= \mathbb{E}[(W_t + \theta W_{t-1})(W_{t+h} + \theta W_{t+h-1})] \\
&= \mathbb{E}[W_t W_{t+h} + \theta W_t W_{t+h-1} + \theta W_{t-1} W_{t+h} + \theta^2 W_{t-1} W_{t+h-1}],
\end{aligned}$$

la cual contará con un término sólo si  $h = 1$  ó  $h = -1$ .

Por lo tanto, dado que  $\gamma(h) = \gamma(-h)$ , la función de autocovarianza para un proceso MA(1) se puede expresar como

$$\gamma(h) = \begin{cases} (1 + \theta^2)\sigma_w^2 & \text{si } h = 0 \\ \theta\sigma_w^2 & \text{si } |h| = 1 \\ 0 & \text{si } |h| > 1 \end{cases}.$$

Basta dividir entre  $\gamma(0)$  para obtener su función de autocorrelación

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{si } |h| = 1 \\ 0 & \text{si } |h| > 1 \end{cases}.$$

Sea  $\{X_t\}$  un proceso de promedios móviles de orden  $q$ , MA( $q$ ), definido por

$$X_t = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q},$$

o bien

$$X_t = \sum_{j=0}^q \theta_j W_{t-j}, \quad \theta_0 = 1, \quad \text{con } W_t \sim \text{RB}(0, \sigma_w^2).$$

Se puede notar que  $\mathbb{E}[X_t] = 0$ .

La función de autocovarianza se obtiene de la siguiente manera

$$\begin{aligned}
\gamma(h) &= \text{Cov}(X_t, X_{t+h}) \\
&= \text{Cov}\left(\sum_{j=0}^q \theta_j W_{t-j}, \sum_{k=0}^q \theta_k W_{t+h-k}\right) \\
&= \mathbb{E}\left[\sum_{j=0}^q \theta_j W_{t-j} \sum_{k=0}^q \theta_k W_{t+h-k}\right] \\
&= \sum_{j=0}^q \sum_{k=0}^q \theta_j \theta_k \mathbb{E}[W_{t-j} W_{t+h-k}] \\
&= \sum_{j=0}^{q-h} \theta_j \theta_{j+h} \mathbb{E}[W_{t-j} W_{t-j}],
\end{aligned}$$

donde usando que la función de autocovarianza es una función par

$$\gamma(h) = \begin{cases} \sigma_w^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|} & \text{si } |h| \leq q \\ 0 & \text{si } |h| > q \end{cases}.$$

En particular  $\gamma(0) = \text{Var}(X_t) = \sigma_w^2(1 + \theta_1^2 + \dots + \theta_q^2)$  es la varianza del proceso.

Y por tanto, la función de autocorrelación del proceso MA( $q$ ) es

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}}{1 + \theta_1^2 + \dots + \theta_q^2} & \text{si } |h| \leq q \\ 0 & \text{si } |h| > q \end{cases} .$$

**Observación 1** La función de autocorrelación (ACF) para los procesos MA( $q$ ) se hace cero después del rezago  $q$ . De esta manera es conveniente para determinar el orden en el ajuste de series cuya ACF presenta un comportamiento similar al de procesos MA( $q$ ).

### 2.2.2. Procesos AR

Corresponde en esta parte calcular las funciones de autocovarianza y autocorrelación de los procesos autoregresivos de órdenes 1 y  $p$  respectivamente.

Una serie  $\{X_t\}$  sigue un proceso autoregresivo de orden 1, AR(1), si

$$X_t = \phi X_{t-1} + W_t \quad \text{ó} \quad (1 - \phi B)X_t = W_t,$$

donde  $BX_t = X_{t-1}$  es el operador de retraso y  $W_t \sim \text{RB}(0, \sigma_w^2)$ .

Considerando retrospectivamente el proceso AR(1) a través de sustituciones, se tiene

$$\begin{aligned} X_t &= \phi X_{t-1} + W_t = \phi(\phi X_{t-2} + W_{t-1}) + W_t \\ &= W_t + \phi W_{t-1} + \phi^2 W_{t-2} + \phi^3 X_{t-3} \\ &\vdots \\ &= \sum_{j=0}^{\infty} \phi^j W_{t-j}. \end{aligned}$$

Como  $\mathbb{E}[W_{t-j}] = 0 \quad \forall j$ , entonces también  $\mathbb{E}[X_t] = 0$ .

Ahora, nótese que

$$\begin{aligned} \gamma(0) &= \text{Var}[X_t] = \text{Var}\left[\sum_{j=0}^{\infty} \phi^j W_{t-j}\right] \\ &= \sum_{j=0}^{\infty} \phi^{2j} \text{Var}(W_{t-j}) = \sum_{j=0}^{\infty} \phi^{2j} \sigma_w^2 \\ &= (1 + \phi^2 + \phi^4 + \phi^6 + \dots) \sigma_w^2 \end{aligned}$$

forma una serie cuya convergencia se tiene solamente si  $|\phi| < 1$ . Esto es, la condición de estacionariedad del proceso AR(1) es  $|\phi| < 1$ .

En ese caso su varianza será  $\gamma(0) = \frac{\sigma_w^2}{1 - \phi^2}$ .

Luego, calcúlese

$$\begin{aligned}
\gamma(h) &= \text{Cov}(X_t, X_{t+h}) = \mathbb{E}[X_t X_{t+h}] \\
&= \mathbb{E}\left[\left(\sum_{j=0}^{\infty} \phi^j W_{t-j}\right)\left(\sum_{k=0}^{\infty} \phi^k W_{t+h-k}\right)\right] \\
&= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi^j \phi^k \mathbb{E}[W_{t-j} W_{t+h-k}] \\
&= \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \mathbb{E}[W_{t-j} W_{t-j}] \\
&= \phi^h \sigma_w^2 \sum_{j=0}^{\infty} \phi^{2j} \\
&= \frac{\sigma_w^2 \phi^h}{1-\phi^2}.
\end{aligned}$$

Por lo tanto, usando la propiedad de simetría, la función de autocovarianza del proceso AR(1) es

$$\gamma(h) = \frac{\sigma_w^2 \phi^{|h|}}{1-\phi^2}$$

y su función de autocorrelación para  $h = 0, \pm 1, \pm 2, \dots$ , es  $\rho(h) = \phi^{|h|}$ .

Considérese en general un proceso  $\{X_t\}$  autoregresivo de orden  $p$  o AR( $p$ )

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = W_t \quad \text{ó} \quad \phi(B)X_t = W_t,$$

donde  $W_t \sim \text{RB}(0, \sigma_w^2)$  y  $\phi(B)$  es el operador de retraso. Multiplicando ambos lados del modelo por  $X_{t-h}$  y tomando esperanza se tiene

$$\mathbb{E}[(X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p})X_{t-h}] = \mathbb{E}[X_{t-h} W_t].$$

Bajo condiciones de estacionariedad y para  $h = 0$  obtenemos

$$\gamma(0) - \phi_1 \gamma(1) - \dots - \phi_p \gamma(p) = \sigma_w^2,$$

pues

$$\begin{aligned}
\mathbb{E}[X_t W_t] &= \mathbb{E}[(\phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + W_t)W_t] \\
&= \phi_1 \mathbb{E}[X_{t-1} W_t] + \dots + \phi_p \mathbb{E}[X_{t-p} W_t] + \mathbb{E}[W_t^2] \\
&= \sigma_w^2.
\end{aligned}$$

Para  $h > 0$  obtenemos

$$\gamma(h) - \phi_1 \gamma(h-1) - \dots - \phi_p \gamma(h-p) = 0.$$

Dividiendo en ambos casos entre  $\gamma(0)$  y usando que  $\rho(h) = \rho(-h)$ , se tiene

$$\rho(h) - \phi_1\rho(h-1) - \dots - \phi_p\rho(h-p) = \begin{cases} \frac{\sigma_w^2}{\gamma(0)} & \text{si } h = 0 \\ 0 & \text{si } |h| > 0 \end{cases}.$$

Lo que permite establecer las ecuaciones de Yule-Walker, adecuadas para realizar estimaciones de los coeficientes en el ajuste de modelos autoregresivos (AR).

**Definición 11 (Ecuaciones de Yule-Walker)** El sistema de ecuaciones de Yule-Walker de orden  $p$  está determinado por

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(2-p) & \rho(1-p) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(3-p) & \rho(2-p) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho(p-1) & 1 & \rho(p-2) & \cdots & \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}.$$

**Observación 2** La función de autocorrelación del proceso  $\text{AR}(p)$  decae exponencialmente u oscilando cuando  $h \rightarrow \infty$ .

Enseguida veamos la función de autocorrelación parcial.

**Definición 12 (Función de Autocorrelación Parcial)** La función de autocorrelación parcial (PACF) es la función  $\alpha(\cdot)$  definida por las ecuaciones

$$\alpha(0) = 1, \quad \alpha(h) = \phi_{hh} \quad h \geq 1,$$

donde  $\phi_{hh}$  es la última componente de  $\phi(h) = \Gamma_h^{-1}\gamma(h)$ , con  $\Gamma_h = [\gamma(i-j)]_{i,j=1}^h$  y  $\gamma(h) = [\gamma(1), \gamma(2), \dots, \gamma(h)]'$ .

Una forma alternativa de definirla es considerando los coeficientes de las  $p$  regresiones lineales

$$\begin{aligned} X_t &= \phi_{01} + \phi_{11}X_{t-1} + e_{1t}, \\ X_t &= \phi_{02} + \phi_{12}X_{t-1} + \phi_{22}X_{t-2} + e_{2t}, \\ &\vdots \\ X_t &= \phi_{0p} + \phi_{1p}X_{t-1} + \dots + \phi_{pp}X_{t-p} + e_{pt}, \end{aligned}$$

donde la función de autocorrelación parcial (PACF) es  $\phi_{hh}$  ( $h = 1, \dots, p$ ).

**Observación 3** Se puede probar que la PACF de un proceso AR( $p$ ) se anula para rezagos de orden mayor o igual a  $p$ , y para los procesos MA( $q$ ) decae exponencialmente u oscilando cuando  $q \rightarrow \infty$ ; desempeñando un papel para los procesos autoregresivos, semejante al de la función de autocorrelación (ACF) para los procesos de promedios móviles.

Para algún conjunto de observaciones la función de autocorrelación parcial muestral  $\hat{\alpha}(h)$  está dada por

$$\hat{\alpha}(0) = 1, \quad \hat{\alpha}(h) = \hat{\phi}_{hh} \quad h \geq 1,$$

donde  $\hat{\phi}_{hh}$  es la última componente de  $\hat{\phi}(h) = \hat{\Gamma}_h^{-1} \hat{\gamma}(h)$ , con  $\hat{\Gamma}_h$  la matriz de autocovarianza muestral.

### 2.2.3. ARMA(1,1)

Una serie de tiempo  $\{X_t\}$  es un proceso ARMA(1,1), si es estacionario y satisface para cada  $t$

$$X_t - \phi X_{t-1} = W_t + \theta W_{t-1}, \quad \text{donde } \{W_t\} \sim \text{RB}(0, \sigma_w^2) \quad \text{y} \quad \phi + \theta \neq 0.$$

En primer lugar, veáse que

$$\begin{aligned} \text{Cov}(X_t, W_t) &= \mathbb{E}[X_t W_t] - \mathbb{E}[X_t] \mathbb{E}[W_t] \\ &= \mathbb{E}[(\phi X_{t-1} + W_t + \theta W_{t-1}) W_t] \\ &= \mathbb{E}[\phi X_{t-1} W_t + W_t^2 + \theta W_{t-1} W_t] \\ &= \mathbb{E}[W_t^2] = \sigma_w^2, \end{aligned} \tag{2.3}$$

ya que  $X_{t-1}, W_t$  son no correlacionados y además  $\mathbb{E}[W_t] = 0$ .

Luego, la varianza es

$$\begin{aligned} \text{Var}(X_t) &= \text{Var}(\phi X_{t-1}) + \text{Var}(W_t) + \text{Var}(\theta W_{t-1}) + 2\text{Cov}(\phi X_{t-1}, \theta W_{t-1}) \\ &= \phi^2 \text{Var}(X_{t-1}) + \sigma_w^2 + \theta^2 \sigma_w^2 + 2\phi\theta \text{Cov}(X_{t-1}, W_{t-1}) \\ &= \phi^2 \text{Var}(X_{t-1}) + \sigma_w^2 + \theta^2 \sigma_w^2 + 2\phi\theta \sigma_w^2. \end{aligned}$$

Usando 2.3 y que  $X_t$  se trata de un proceso estacionario i.e.  $\text{Var}(X_t) = \text{Var}(X_{t-1})$ , se tiene

$$(1 - \phi^2) \text{Var}(X_t) = (1 + \theta^2 + 2\phi\theta) \sigma_w^2,$$

y por lo tanto

$$\gamma(0) = \text{Var}(X_t) = \frac{1 + 2\phi\theta + \theta^2}{1 - \phi^2} \sigma_w^2.$$

Obteniendo  $\gamma(0)$  para un proceso ARMA(1,1) descrito por la ecuación

$$X_t = \phi X_{t-1} + W_t + \theta_{t-1},$$

multiplíquese ésta última por  $X_{t-h}$  y tómesese la esperanza

$$\mathbb{E}[X_t X_{t-h}] = \phi \mathbb{E}[X_{t-1} X_{t-h}] + \mathbb{E}[W_t X_{t-h}] + \theta \mathbb{E}[W_{t-1} X_{t-h}],$$

de donde

$$\gamma(h) = \phi \gamma(h-1) + \text{Cov}(W_t, X_{t-h}) + \theta \text{Cov}(W_{t-1}, X_{t-h}). \quad (2.4)$$

En particular para  $h = 1$  y por 2.3

$$\gamma(1) = \phi \gamma(0) + \theta \sigma_w^2,$$

y dado que se conoce  $\gamma(0)$ , es posible calcular explícitamente  $\gamma(1)$  como

$$\gamma(1) = \phi \frac{1 + 2\phi\theta + \theta^2}{1 - \phi^2} \sigma_w^2 + \theta \sigma_w^2 = \frac{(\phi + \theta)(1 + \phi\theta)}{1 - \phi^2} \sigma_w^2.$$

En la ecuación 2.4 para  $h > 1$ , la expresión

$$\gamma(h) = \phi \gamma(h-1)$$

permite llegar recursivamente a

$$\gamma(h) = \phi^{h-1} \gamma(1),$$

en donde dividiendo entre  $\gamma(0)$  se obtiene

$$\rho(h) = \phi^{h-1} \rho(1).$$

Conociendo

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2},$$

la función de autocorrelación para un proceso ARMA(1,1) es en general

$$\rho(h) = \phi^{h-1} \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2}.$$

## 2.3. Series No Estacionarias

En la última sección de este capítulo nuestro objetivo es solamente introducir una pequeña parte de la teoría de series no estacionarias, ubicadas en diferentes casos de estudio como series de precios, tasas de interés y divisas, principalmente por el hecho de que no existe un nivel fijo para ellas.

### 2.3.1. Caminata Aleatoria

Un proceso  $\{Y_t\}$  es una caminata aleatoria si

$$Y_t = Y_{t-1} + W_t, \quad W_t \sim \text{RB}(0, \sigma^2),$$

notando que en particular puede tratarse como un proceso AR(1) no estacionario, ya que el coeficiente de  $Y_{t-1}$  es la unidad. A pesar de su no estacionariedad, este tipo de proceso sirve para modelar el comportamiento de los precios, pues usualmente se suponen no predecibles o reversibles a la media, en el sentido que para todos los horizontes la predicción del precio es simplemente el valor de la serie en el origen del pronóstico.

Considerando la representación

$$\begin{aligned} Y_t &= Y_{t-1} + W_t = (Y_{t-2} + W_{t-1}) + W_t \\ &\vdots \\ &= W_t + W_{t-1} + W_{t-2} + \cdots = \sum_{i=0}^{\infty} \psi_i W_{t-i}, \end{aligned}$$

es claro que  $\psi_i = 1$  para toda  $i$ , por lo tanto el impacto de cualquier ruido o *shock* pasado en  $Y_t$  no decae en el tiempo y como consecuencia la serie recordará todos los *errores* anteriores. En economía ese comportamiento es conocido como el efecto permanente del

ruido en la serie y podrá observarse en la ACF muestral, cuando al incrementar el tamaño de la muestra, los valores se aproximarán cada vez más a la unidad.

### 2.3.2. Procesos ARIMA

Otro caso de series no estacionarias se encuentra en los procesos ARMA donde el polinomio autoregresivo cuenta con una raíz unitaria o muy cerca del círculo unitario. La generalización que incorpora esta clase de modelos son los Autoregresivos de Promedios Móviles Integrados (ARIMA).

**Definición 13 (ARIMA( $p,d,q$ ))** Si  $d$  es un entero no negativo, entonces  $\{X_t\}$  es un proceso ARIMA( $p,d,q$ ) si  $Y_t := (1 - B)^d X_t$  es un proceso ARMA( $p,q$ ) causal.

$\nabla X_t := (X_t - X_{t-1})$  es el operador diferencia, donde en general  $\nabla^d X_t := (1 - B)^d X_t$ .

Entonces el modelo ARIMA es un proceso  $\{X_t\}$  tal que diferenciado  $d$  veces produce un proceso estacionario y de la definición, satisface la ecuación

$$\phi^*(B)X_t \equiv \phi(B)(1 - B)^d X_t = \theta(B)W_t, \quad W_t \sim \text{RB}(0, \sigma^2),$$

donde  $\phi(z)$  y  $\theta(z)$  son los polinomios de grado  $p$  y  $q$  respectivamente,  $\phi(z) \neq 0$  para  $z \leq 1$  y el polinomio  $\phi^*(z)$  tiene un cero de orden  $d$  en  $z = 1$ .<sup>2</sup>

Precisamente debido a que el polinomio AR tiene una raíz unitaria, adoptan el nombre de *no estacionarios de raíz unitaria* y de la misma manera que la caminata aleatoria los errores tendrán un efecto permanente en la serie, siendo ésta de gran memoria.

El ejemplo más común es el proceso ARIMA( $p,1,q$ ):  $Y_t = X_t - X_{t-1} = (1 - B)X_t$ , que transforma la serie no estacionaria por una que sí lo es (el proceso diferenciado o también llamado cambio de  $X_t$ ).

---

<sup>2</sup>El proceso  $X_t$  por sí mismo es débilmente estacionario si y sólo si  $d = 0$ , es decir, si se trata de un ARMA( $p,q$ ).

## Capítulo 3

# Modelos con Heteroscedasticidad Condicional

### 3.1. Propiedades de la volatilidad

Aunque el riesgo es parte central en diversas áreas de las finanzas modernas, por ejemplo en teoría de portafolios se mide por las varianzas y covarianzas de los rendimientos, no fue sino hasta décadas recientes que investigadores en economía financiera comenzaron a utilizar explícitamente la variación en el tiempo en momentos de segundo orden y superiores para modelarlo, obteniendo resultados satisfactorios.

A partir de esa idea los modelos con heteroscedasticidad condicional se han vuelto importantes en el tratamiento de series de tiempo, en especial dentro del ámbito financiero donde la meta es analizar y pronosticar la *volatilidad*.

Su principal aportación en econometría ha sido el descubrimiento de que para algunas familias de modelos, lineales o no lineales, es posible especificar un proceso estocástico para los términos de error en series de rendimientos de instrumentos financieros (principalmente) y predecir el tamaño promedio de ellos. Esa es la esencia de la modelación de procesos Autoregresivos con Heteroscedasticidad Condicional (ARCH) introducida por Engle [1982].

En este capítulo se define volatilidad como la desviación estándar condicional de los rendimientos de un instrumento financiero, remarcando primero que, no puede calcularse directamente, teóricamente es variante en el tiempo y sólo puede inferirse, más no observarse; y segundo, su relevancia proviene de la existencia de una serie de aspectos verificados empíricamente, en la literatura económica conocidos como *hechos estilizados*, propios de las series de rendimientos, a saber:

- *Conglomerados de volatilidad*

La volatilidad puede ser baja para ciertos periodos de tiempo y alta para otros, ya que la mayoría de las series exhiben fases de tranquilidad relativa seguida por periodos de gran volatilidad. Este fenómeno es conocido como conglomerados de volatilidad y puede observarse claramente cuando los rendimientos se grafican en el tiempo.

- *Distribuciones de rendimientos con colas pesadas*

Mandelbrot [1963] documentó que las distribuciones de series de rendimientos en el caso de instrumentos financieros tendían a ser leptocúrticas, es decir, a tener colas pesadas o densas en los extremos.

Actualmente ambas propiedades son ampliamente aceptadas y parecen estar directamente relacionadas.

- *Correlación y co-movimientos en la volatilidad*

Puede decirse que los movimientos en el mercado reflejan las variaciones de cada uno de los activos financieros que lo conforman. De ahí que instrumentos volátiles sean más sensibles ante cambios en la volatilidad de los precios y generalmente tenderán a moverse hacia la misma dirección que lo haga el mercado.

- *Asimetría*

También se suele llamar *efecto palanca* y se debe a que la volatilidad parece reaccionar de manera diferente a grandes incrementos en los precios de los activos que a grandes caídas en ellos. Black [1976] observó la tendencia de los movimientos en los precios a ser negativamente correlacionados con cambios en la volatilidad. En otras palabras, a medida que el precio de un instrumento sube (baja) la volatilidad del mismo tenderá a bajar (subir). Se verá el modelo GARCH exponencial (EGARCH) permitiendo capturar esta característica.

- *Persistencia en la Varianza*

Algunas series de datos exhiben persistencia, en el sentido que la varianza condicional de segundo orden mantiene su importancia en el pronóstico para todos los horizontes, pero su representación univariada tiene una raíz unitaria. Veremos que la varianza condicional en el modelo GARCH integrado (IGARCH) es persistente, aunque el proceso no es débilmente estacionario.

Además de los hechos anteriores, consideramos importante distinguir entre la nociones de *homocedasticidad* y *heteroscedasticidad*.

Considérese el modelo de regresión lineal múltiple  $Y = x\beta + \epsilon$ , en donde  $Y$  es llamada variable dependiente o respuesta,  $x$  el conjunto de variables independientes o predictivas,  $\beta$  el vector de parámetros a estimar y  $\epsilon$  son los errores o residuales, de tal forma que

$$\beta = (\beta_0, \beta_1, \dots, \beta_p)' \quad p < n, \quad \epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)' \quad \text{y} \quad x = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}.$$

Si  $\epsilon_i \sim N(0, \sigma^2)$ , la condición de homocedasticidad se refiere al supuesto de varianza constante para  $\epsilon$ , con  $\mathbb{E}[\epsilon_i] = 0$  y  $\text{Var}[\epsilon_i] = \mathbb{E}[\epsilon_i^2] = \sigma^2$ . Por lo tanto el tamaño esperado del error es constante y no depende del tamaño de la variable  $x$ .

En cambio se dice que existe evidencia de heteroscedasticidad, cuando el tamaño esperado del error no es constante. Un ejemplo en la economía es el caso del ingreso y gasto familiar, ya que las personas estarán dispuestas a incrementar su consumo a medida que sus salarios se elevan, sin embargo, no lo harán en la misma cantidad.

Considerando series de tiempo, una manera de establecer un modelo de regresión de valores futuros con los valores del presente y pasado de una misma observación  $x$ , sería

$$x_{t+1} = \beta_0 + \beta_1 x_t + \cdots + \beta_n x_{t-n} + \epsilon_{t+1},$$

donde el término de error  $\epsilon_{t+1}$  es condicional a la información disponible, en este caso representada por los  $n$  valores pasados de la variable  $x$ . De esta forma podemos establecer que los términos de error son condicionalmente heteroscedásticos ya que varían durante el tiempo.

Recordemos que un proceso  $\{x_t\}$  es estrictamente estacionario si ninguna de sus distribuciones finito dimensionales depende del tiempo, mientras que se dice débilmente estacionario si sus momentos de segundo orden son todos constantes.

Lo último esencialmente significa que su media y varianza (no condicionales) son constantes y además no dependen del tiempo

$$\mu_t = \mu, \quad \sigma_t^2 = \sigma^2.$$

Ahora, basados en la información disponible al tiempo inmediato anterior  $t-1$ , denotada por  $F_{t-1}$ , se pueden definir la media y varianza condicionales

$$\mu_t = \mathbb{E}[x_t|F_{t-1}], \quad \sigma_t^2 = \text{Var}(x_t|F_{t-1}) = \mathbb{E}[(x_t - \mu_t)^2|F_{t-1}].$$

Observando que si la media condicional es constante, entonces la varianza no condicional será la esperanza de la varianza condicional i.e.  $\sigma^2 = \mathbb{E}[\text{Var}(x_t|F_{t-1})]$ , en caso contrario no necesariamente. Por lo tanto es posible que algún proceso  $\{x_t\}$  débilmente estacionario posea varianza condicional que cambia en el tiempo.

Para introducir el modelo que se utilizará para describir el comportamiento heterocedástico de la volatilidad representada por el error, y ya que nos enfocaremos a aplicaciones financieras, se hará uso de la siguiente notación a fin de facilitar la lectura del capítulo.

Sea  $P_t$  el precio de un activo o instrumento financiero al tiempo  $t$ , el rendimiento simple del instrumento se define como

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1,$$

de donde  $(1 + R_t) = \frac{P_t}{P_{t-1}}$  es conocido como el rendimiento neto.

Debido a ciertas propiedades estadísticas y a su relación con procesos dinámicos en la valuación de instrumentos financieros derivados, en la práctica se utiliza de manera más conveniente el log-rendimiento, definido como

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

En adelante se considerarán series de tiempo  $\{r_t\}$ , que se denotarán como  $r_t$  y a las cuales llamaremos *rendimientos*, entendiendo que se tratan de log-rendimientos.

### 3.1.1. Especificación del Modelo de Volatilidad

Se describe el modelo para una serie de tiempo univariada de rendimientos  $r_t$  como

$$r_t = \mu_t + \epsilon_t, \tag{3.1}$$

donde  $\mu_t$  es la media condicional y  $\epsilon_t$  el proceso de error.

En particular, los errores o *shocks*  $\epsilon_t$  desempeñan la parte impredecible de la serie y se suponen como un proceso de *innovación*, es decir, forman una serie de observaciones no correlacionadas con media condicional cero.

Siguiendo el tratamiento de algunos autores como Tsay [2005] se propone modelar  $\mu_t$  a través de un proceso ARMA( $p,q$ ) usando los fundamentos descritos en el capítulo anterior y llamando a ese ajuste *ecuación media* para  $r_t$ .

Suponiendo que la ecuación media es conocida, la varianza condicional de  $r_t$  es entonces

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

donde  $F_{t-1}$  denota al conjunto de información disponible al tiempo  $t - 1$ , usualmente conformada por todas las combinaciones lineales de los rendimientos pasados.

Los modelos con heteroscedasticidad condicional se referirán a la evolución de  $\sigma_t^2$ , y la manera en que lo haga, diferenciará a un modelo de otro. De hecho la ecuación asignada a  $\sigma_t^2$  es conocida como *varianza* o *ecuación de volatilidad*.

Se define un proceso  $\epsilon_t$  autoregresivo con heteroscedasticidad condicional (ARCH) de la forma

$$\epsilon_t = \sigma_t z_t,$$

con  $\sigma_t$  la desviación estándar condicional del proceso y  $z_t$  una sucesión de variables aleatorias independientes idénticamente distribuidas (i.i.d.) con media cero y varianza uno.

Por lo regular se supone a la distribución de  $z_t$  Gaussiana, sin embargo, al considerar observaciones de alta frecuencia (intradía), los hechos estilizados indican efectivamente que existen conglomerados de volatilidad, pero las distribuciones de rendimientos tienden a tener colas más pesadas que la distribución Normal. Por tanto se sugieren distribuciones condicionales alternativas que permitan capturar la característica leptocúrtica de los rendimientos, considerándose las distribuciones Student-t y la Distribución de Error Generalizada (GED).

Como la varianza no condicional del proceso  $r_t$  en general no coincide con la varianza no condicional del error, éste último no puede separarse del resto del modelo, pues sólo se tiene una realización de la serie de tiempo y el término de volatilidad está sujeto al modelo empleado para pronosticar los rendimientos esperados. Entonces el comportamiento del error dependerá del paradigma elegido para representar los datos, pudiéndose utilizar diferentes modelos para ajustarlos, cada uno con distintos niveles de precisión y una especificación de heteroscedasticidad única.

A continuación se presentan los modelos más relevantes para describir la volatilidad en el caso univariado.

### 3.2. ARCH

En 1982, con la intención de proveer una manera sistemática para modelar la volatilidad, Engle introdujo el modelo ARCH, tratándose originalmente de la primera representación con heteroscedasticidad condicional y cuyas ideas clave son:

- La serie de errores  $\epsilon_t$  es no correlacionada, pero dependiente.
- La varianza condicional de  $\epsilon_t$  es una función lineal de sus  $r$  valores cuadráticos pasados.

De tal forma el modelo ARCH de orden  $r$  o ARCH( $r$ ) es

$$\epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_r \epsilon_{t-r}^2, \quad (3.2)$$

donde  $z_t$  es una sucesión de variables aleatorias i.i.d. con media cero y varianza 1, y los parámetros deben satisfacer  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, r$ , para garantizar la positividad de la varianza condicional y que el modelo esté bien definido.

Análogamente  $\sigma_t^2$  tiene la representación

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i \epsilon_{t-i}^2 = \alpha_0 + \alpha(B) \epsilon_t,$$

con  $B$  el operador de retraso  $B^i \epsilon_t = \epsilon_{t-i}$ ,  $\alpha(z)$  el polinomio característico del proceso y se cumplen  $\alpha_0 > 0$  y  $\alpha_i \geq 0$ ,  $i = 1, \dots, r$ .

Para hallar una expresión de  $\epsilon_t$  semejante a la de procesos autoregresivos, defínase el proceso condicionalmente heteroscedástico  $\nu_t = \epsilon_t^2 - \sigma_t^2$ , permitiendo reescribir 3.2 como

$$\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_r \epsilon_{t-r}^2 + \nu_t,$$

que corresponde a un proceso AR de orden  $r$  para  $\epsilon_t^2$ , ya que  $\mathbb{E}[\nu_t | F_{t-1}] = 0$ .

Se puede probar que el proceso es débilmente estacionario, de la misma manera que para procesos autoregresivos, si las raíces del polinomio característico (en este caso formado por las  $\alpha$ 's) se encuentran fuera del círculo unitario. En tal situación la varianza no condicional es definida como

$$\text{Var}[\epsilon_t] \equiv \sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^r \alpha_i} \quad \text{con} \quad \sum_{i=1}^r \alpha_i < 1,$$

siendo la última una condición de regularidad que asegura la finitud de la varianza no condicional.

Para motivar la ventaja de emplear momentos condicionales, veamos los siguientes argumentos sin considerar por ahora el supuesto de heteroscedasticidad.

Sea  $r_t$  un proceso (estacionario) autoregresivo de orden uno con nivel  $\phi_0$

$$r_t = \phi_0 + \phi_1 r_{t-1} + \epsilon_t, \quad |\phi_1| < 1 \text{ y } \epsilon_t \sim \text{RB}(0, \sigma^2).$$

Se pueden calcular la esperanza y varianza condicionales dada la observación  $r_{t-1}$

$$\begin{aligned} \mathbb{E}[r_t | r_{t-1}] &= \mathbb{E}[\phi_0 + \phi_1 r_{t-1} + \epsilon_t | r_{t-1}] \\ &= \phi_0 + \phi_1 \mathbb{E}[r_{t-1} | r_{t-1}] + \mathbb{E}[\epsilon_t | r_{t-1}] \\ &= \phi_0 + \phi_1 r_{t-1} + \mathbb{E}[\epsilon_t] \\ &= \phi_0 + \phi_1 r_{t-1}, \end{aligned}$$

y

$$\begin{aligned} \text{Var}(r_t | r_{t-1}) &= \mathbb{E}[(r_t - \mathbb{E}[r_t | r_{t-1}])^2 | r_{t-1}] \\ &= \mathbb{E}[(r_t - \phi_0 - \phi_1 r_{t-1})^2 | r_{t-1}] \\ &= \mathbb{E}[\epsilon_t^2 | r_{t-1}] = \mathbb{E}[\epsilon_t^2] = \sigma^2. \end{aligned}$$

Para obtener la varianza no condicional, véase el proceso en forma retrospectiva

$$\begin{aligned} r_t &= \phi_0 + \phi_1 r_{t-1} + \epsilon_t = \phi_0 + \phi_1(\phi_0 + \phi_1 r_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &\vdots \\ &= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \phi_1^3 \epsilon_{t-3} + \dots = \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} + \sum_{j=1}^{\infty} \phi_0 \phi_1^j, \end{aligned}$$

donde los últimos términos serán productos de las constantes  $\phi$ 's. Por tanto

$$\begin{aligned} \text{Var}(r_t) &= \text{Var}(\epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \dots) \\ &= \text{Var}(\epsilon_t) + \phi_1^2 \text{Var}(\epsilon_{t-1}) + \phi_1^4 \text{Var}(\epsilon_{t-2}) + \dots \\ &= \sigma^2 + \phi_1^2 \sigma^2 + \phi_1^4 \sigma^2 + \dots \\ &= \frac{\sigma^2}{1 - \phi_1^2}. \end{aligned}$$

Pues tratándose de un proceso estacionario, se cumple  $|\phi_1| < 1$ .

De tal manera se concluye que  $\text{Var}[r_t|r_{t-1}] < \text{Var}[r_t]$ . Esto es, la varianza condicional es menor que la varianza no condicional del proceso.

Empleando este resultado en la realización de un pronóstico, nuevamente considérese el modelo estacionario AR(1) con nivel  $\phi_0$ , pudiéndose probar que la esperanza condicional de  $r_{t+1}$  dado  $r_t$  es  $\phi_0 + \phi_1 r_t$ . La varianza del error dado  $r_t$  para el pronóstico es entonces

$$\begin{aligned}\mathbb{E}[(r_{t+1} - \mathbb{E}(r_{t+1}|r_t))^2|r_t] &= \mathbb{E}[(r_{t+1} - \phi_0 - \phi_1 r_t)^2|r_t] \\ &= \mathbb{E}[\epsilon_{t+1}^2|r_t] = \mathbb{E}[\epsilon_{t+1}^2] = \sigma^2.\end{aligned}$$

Al contrario, utilizando un pronóstico no condicional sabemos que

$$\mathbb{E}[r_{t+1}] = \mathbb{E}[\phi_0 + \phi_1 r_t + \epsilon_t] = \phi_0 + \phi_1 \mathbb{E}[r_t],$$

donde suponiendo estacionariedad del proceso i.e.  $\mathbb{E}[r_{t+1}] = \mathbb{E}[r_t]$ , se tiene la llamada media a largo plazo

$$\mathbb{E}[r_{t+1}] = \frac{\phi_0}{1 - \phi_1}.$$

**Observación 4** En este caso la varianza del error para el pronóstico no condicional será

$$\mathbb{E}[(r_{t+1} - \mathbb{E}[r_{t+1}])^2] = \mathbb{E}\left[\left(\phi_0 + \phi_1 r_t + \epsilon_{t+1} - \frac{\phi_0}{1 - \phi_1}\right)^2\right] = \frac{\sigma^2}{1 - \phi_1^2}. \quad 3$$

Reafirmamos entonces que el pronóstico no condicional proporciona una varianza mayor que el pronóstico condicional.

Intuitivamente hemos observado que es de esperarse obtener mejores intervalos de predicción, permitiendo a la información disponible del pasado influir en la varianza de un pronóstico.

Ahora, ¿qué pasa si la varianza de  $\epsilon_t$  no es constante?, o sea, en el caso de heteroscedasticidad.

Consideremos a  $\epsilon_t$  el error del modelo AR(1) estacionario y supongamos que su varianza condicional no es constante. Usando los cuadrados de los errores pasados estimados una aproximación de la misma es

$$\hat{\epsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\epsilon}_{t-1}^2 + \alpha_2 \hat{\epsilon}_{t-2}^2 + \cdots + \alpha_r \hat{\epsilon}_{t-r}^2 + z_t, \quad \text{donde } z_t \sim \text{RB}. \quad (3.3)$$

---

<sup>3</sup>Demostración en el Apéndice A

Si la varianza condicional evoluciona de acuerdo al proceso dado en la ecuación anterior, que de hecho tiene la estructura de un proceso autoregresivo de orden  $r$  para  $\hat{\epsilon}_t^2$ , es posible pronosticarla como

$$\text{Var}(r_{t+1}|F_t) = \mathbb{E}[\hat{\epsilon}_{t+1}^2|F_t] = \alpha_0 + \alpha_1\hat{\epsilon}_t^2 + \alpha_2\hat{\epsilon}_{t-1}^2 + \cdots + \alpha_r\hat{\epsilon}_{t+1-r}^2,$$

donde  $F_t$  representa el conjunto de información disponible hasta el tiempo  $t$ .

De ahí la justificación para nombrarles modelos autoregresivos con heteroscedasticidad condicional (ARCH) y de los cuales Engle [1982] propuso el ejemplo

$$\epsilon_t = z_t \sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2}, \text{ donde } z_t \text{ son i.i.d. con media 0 y varianza 1.}$$

En síntesis el modelo ARCH posee las siguientes características:

- Establece que  $\epsilon_t$  conforma una serie de observaciones no correlacionadas, pero dependientes. Esa dependencia se da a través de una función ( $\sigma_t^2$ ) de sus valores pasados.
- Existe una tendencia a que grandes (pequeños) valores absolutos de la serie sean seguidos por valores grandes (pequeños) de signo no predecible. De manera natural introduce la condición de conglomerados de volatilidad.
- Si la distribución de la innovación  $z_t$  se supone invariante, la distribución de  $\epsilon_t$  tendrá colas más pesadas que la distribución de  $z_t$ .

Presentamos el ARCH de orden 1, estableciendo algunos resultados importantes.

### 3.2.1. ARCH(1)

En particular se trata de un proceso con heteroscedasticidad condicional multiplicativo

$$\epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2, \quad \text{donde } \alpha_0 > 0, \quad \alpha_1 \geq 0, \quad z_t \text{ son i.i.d. y } z_t \sim N(0, 1).$$

Encontramos tres aspectos a considerar.

Primero, la media no condicional de  $\epsilon_t$  es cero, pues

$$\mathbb{E}[\epsilon_t] = \mathbb{E}[\mathbb{E}[\epsilon_t|F_{t-1}]] = \mathbb{E}[\sigma_t \mathbb{E}[z_t|F_{t-1}]] = 0.$$

Segundo, considérese la varianza no condicional de  $\epsilon_t$

$$\begin{aligned}
\text{Var}[\epsilon_t] &= \mathbb{E}[\epsilon_t^2] = \mathbb{E}[\mathbb{E}[\epsilon_t^2|F_{t-1}]] \\
&= \mathbb{E}[\alpha_0 + \alpha_1\epsilon_{t-1}^2] = \alpha_0 + \alpha_1\mathbb{E}[\epsilon_{t-1}^2] \\
&= \alpha_0 + \alpha_1\text{Var}[\epsilon_{t-1}],
\end{aligned}$$

donde suponiendo que  $\epsilon_t$  se trata de un proceso estacionario, se cumple que la varianza  $\text{Var}[\epsilon_t] = \mathbb{E}[\epsilon_t^2] = \mathbb{E}[\epsilon_{t-1}^2] = \text{Var}[\epsilon_{t-1}]$ . Así que

$$\text{Var}[\epsilon_t] = \frac{\alpha_0}{1 - \alpha_1}.$$

Como  $\text{Var}[\epsilon_t] > 0$ , entonces necesariamente  $0 \leq \alpha_1 < 1$ .

Tercero, en algunos casos se requiere la existencia de momentos de orden mayor de  $\epsilon_t$ , entonces analicemos el comportamiento de las colas de la distribución. Bajo el supuesto de normalidad de  $z_t$ , un resultado para variables aleatorias gaussianas <sup>4</sup> indica

$$\mathbb{E}[\epsilon_t^4|F_{t-1}] = 3(\mathbb{E}[\epsilon_t^2|F_{t-1}])^2 = 3(\alpha_0 + \alpha_1\epsilon_{t-1}^2)^2.$$

Entonces consideremos

$$\begin{aligned}
\mathbb{E}[\epsilon_t^4] &= \mathbb{E}[\mathbb{E}[\epsilon_t^4|F_{t-1}]] \\
&= 3\mathbb{E}[(\alpha_0 + \alpha_1\epsilon_{t-1}^2)^2] \\
&= 3\mathbb{E}[\alpha_0^2 + 2\alpha_0\alpha_1\epsilon_{t-1}^2 + \alpha_1^2\epsilon_{t-1}^4] \\
&= 3\{\alpha_0^2 + 2\alpha_0\alpha_1\mathbb{E}[\epsilon_{t-1}^2] + \alpha_1^2\mathbb{E}[\epsilon_{t-1}^4]\},
\end{aligned}$$

de donde si  $\epsilon_t$  también es estacionario de cuarto orden i.e.  $\mathbb{E}[\epsilon_t^4] = \mathbb{E}[\epsilon_{t-1}^4]$ , se tiene

$$\begin{aligned}
\mathbb{E}[\epsilon_t^4] &= 3(\alpha_0^2 + 2\alpha_0\alpha_1\text{Var}[\epsilon_t] + \alpha_1^2\mathbb{E}[\epsilon_t^4]) \\
&= 3(\alpha_0^2 + 2\alpha_0\alpha_1\frac{\alpha_0}{1 - \alpha_1} + \alpha_1^2\mathbb{E}[\epsilon_t^4]),
\end{aligned}$$

y por lo tanto

$$\mathbb{E}[\epsilon_t^4] = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}.$$

Además, como el cuarto momento es positivo,  $\alpha_1$  debe satisfacer  $1 - 3\alpha_1^2 > 0$ , equivalente a  $0 \leq \alpha_1 < \frac{1}{\sqrt{3}}$ .

---

<sup>4</sup>Ver Mood, Graybill y Boes [1974]

Se puede calcular la curtosis no condicional

$$\begin{aligned} k(\epsilon_t) &= \mathbb{E}[\epsilon_t^4] / (\text{Var}[\epsilon_t])^2 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} / \left(\frac{\alpha_0}{1-\alpha_1}\right)^2 \\ &= \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \cdot \frac{(1-\alpha_1)^2}{\alpha_0^2} = 3 \frac{1-\alpha_1^2}{1-3\alpha_1^2}, \end{aligned}$$

donde sabiéndose que la curtosis de una función de distribución de probabilidad Normal es 3, en este caso la distribución de  $\epsilon_t$  tiene colas más pesadas que una Gaussiana, pues

$$k(\epsilon_t) = 3 \frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3. \quad (3.4)$$

Finalmente, se puede observar que la media condicional del proceso es cero

$$\mathbb{E}[\epsilon_t | F_{t-1}] = \mathbb{E}[\sigma_t z_t | F_{t-1}] = 0,$$

dado que  $z_t$ ,  $\epsilon_{t-1}$  son independientes y  $\mathbb{E}[z_t] = 0$ , siendo la varianza condicional

$$\mathbb{E}[\epsilon_t^2 | F_{t-1}] = \mathbb{E}[\alpha_0 + \alpha_1 \epsilon_{t-1}^2 | F_{t-1}] = \alpha_0 + \alpha_1 \epsilon_{t-1}^2,$$

y cuyo valor depende de la realización  $\epsilon_{t-1}$ . Si ésta última es grande, entonces la varianza condicional también lo será.

### 3.2.2. Efectos ARCH

De acuerdo a la presentación de Engle para los modelos ARCH, los mínimos cuadrados ordinarios son el procedimiento adecuado para probar si los errores son o no condicionalmente heteroscedásticos y como el modelo ARCH requiere procedimientos iterativos, antes de obtener la estimación, sería deseable realizar la prueba de Multiplicador de Lagrange (LM) ideal para estos casos.

Bajo la hipótesis nula  $\alpha_1 = \dots = \alpha_r = 0$ , puede calcularse el estadístico de prueba  $TR^2$ , donde  $R^2$  es el coeficiente de determinación (total de varianza explicada) de la regresión de  $\epsilon_t^2$  con sus valores rezagados  $\epsilon_{t-1}^2, \dots, \epsilon_{t-r}^2$  y  $T$  denota el tamaño de muestra de los errores. Esto es, si no existieran efectos ARCH, los valores estimados de  $\alpha_1 = \dots = \alpha_r$  deberían ser conjuntamente cero.

Dada la hipótesis de no efectos ARCH, el estadístico  $TR^2$  converge a una distribución Chi-cuadrada ( $\chi^2$ ) con  $r$  grados de libertad. Si  $TR^2$  es suficientemente grande, se rechazará la hipótesis nula, equivalente a rechazar la hipótesis de no efectos ARCH. Por el contrario, si

el estadístico es suficientemente pequeño, se puede concluir que no hay evidencia de efectos ARCH. A su vez la PACF de  $\epsilon_t^2$  puede utilizarse para determinar el orden de  $r$  donde el efecto ARCH es significativo, para posteriormente hacer la prueba correspondiente.

Una prueba alterna, pero asintóticamente equivalente, es someter a  $\epsilon_t^2$  a las pruebas estándar de dependencia basadas en la estructura de autocorrelación Portmanteau o Ljung-Box, rechazando efectos ARCH si no existiera correlación *seria* entre los errores al cuadrado, reflejada en valores muestrales de  $\rho(h)$  dentro de las bandas de confianza.

### 3.3. GARCH

Mientras que los modelos econométricos convencionales operan bajo el supuesto de varianza constante, hemos visto que los ARCH permiten a la varianza condicional cambiar en el tiempo como una función de los errores pasados dejando a la varianza no condicional invariante.

Ahora, ya que por lo común las aplicaciones empíricas requieren una gran cantidad de parámetros para describir de forma adecuada la evolución de la volatilidad de los rendimientos, sumado a que muchas veces para eliminar problemas en la obtención de varianza negativa se impone un número fijo de valores de rezago, parece inmediato para fines prácticos, extender la clase ARCH, de tal manera que se permitan memoria larga<sup>5</sup> y una estructura de rezagos más flexible. Bollerslev [1986] proporciona esa alternativa en la generalización ARCH (GARCH).

Un proceso  $\epsilon_t$  sigue un ARCH Generalizado o GARCH( $r, s$ ) si

$$\epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, \quad (3.5)$$

donde  $z_t \sim \text{i.i.d. } N(0, 1)$ ,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$   $i = 1, \dots, r$ ,  $\beta_j \geq 0$   $j = 1, \dots, s$ .<sup>6</sup>

Entendiendo que  $\alpha_i = 0$  para  $i > r$ ,  $\beta_j = 0$  para  $j > s$  y  $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) < 1$ , la última suposición implica que la varianza no condicional de  $\epsilon_t$  es finita, en tanto su varianza condicional  $\sigma_t^2$  evoluciona en el tiempo.

Usando los polinomios de retraso  $\alpha(B)$ ,  $\beta(B)$  y bajo las mismas condiciones en los índices y parámetros, una representación equivalente de  $\sigma_t^2$  en 3.5 es

$$\sigma_t^2 = \alpha_0 + \alpha(B)\epsilon_t^2 + \beta(B)\sigma_t^2. \quad (3.6)$$

<sup>5</sup>Ver Davidson [2004]

<sup>6</sup>Si  $s = 0$ , la ecuación 3.5 se reduce al modelo ARCH( $r$ )

Con  $\nu_t = \epsilon_t^2 - \sigma_t^2$  y reordenando términos, se puede representar  $\epsilon_t^2$  como un proceso autoregresivo de promedios móviles con nivel  $\alpha_0$

$$\epsilon_t^2 = \alpha_0 + [\alpha(B) + \beta(B)]\epsilon_t^2 + \nu_t - \beta(B)\nu_t,$$

hallando cierta analogía a la extensión de los procesos de series de tiempo AR al ARMA general, con una descripción más parsimoniosa en diferentes situaciones y donde

$$\mathbb{E}[\epsilon_t^2] = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i)}$$

es la varianza no condicional, siempre y cuando el denominador sea positivo.

De 3.6 también podemos escribir

$$(1 - \beta(B))\sigma_t^2 = \alpha_0 + \alpha(B)\epsilon_t^2,$$

donde si las raíces del polinomio  $\beta(z) = 1$  están fueran del círculo unitario,

$$\sigma_t^2 = \alpha_0(1 - \beta(1))^{-1} + \alpha(B)(1 - \beta(B))^{-1}\epsilon_t^2 = \alpha_0 \left(1 - \sum_{j=1}^s \beta_j\right)^{-1} + \sum_{i=1}^{\infty} \delta_i \epsilon_{t-i}^2$$

es un proceso infinito dimensional, referido por algunos autores como ARCH( $\infty$ ), debido a la dependencia de la varianza condicional en todos los errores pasados.

Las  $\delta_i$ 's provienen de la expansión en serie de Taylor de  $\delta(1) = \alpha(1)(1 - \beta(1))^{-1}$ ,

$$\delta_i = \begin{cases} \alpha_i + \sum_{j=1}^n \beta_j \delta_{i-j}, & i = 1, \dots, r, \\ \sum_{j=1}^n \beta_j \delta_{i-j}, & i = r + 1, \dots, \end{cases} \quad \text{donde } n = \min\{s, i - 1\}.$$

Nótese que si  $\beta(1) < 1$ ,  $\delta_i$  decrecerá para  $i > m$  con  $m = \max\{r, s\}$ , y si  $\alpha(1) < 1$  el peso de los errores más viejos será pequeño. Entonces si  $\delta(1) < 1$ , el proceso GARCH( $r, s$ ) podrá aproximarse con cierto grado de precisión por un ARCH( $Q$ ) estacionario, para un valor suficientemente grande de  $Q$ .

Las condiciones de estacionariedad se establecen en el siguiente teorema.

**Teorema 1** El proceso GARCH( $r, s$ ) definido en 3.5 es estacionario de segundo orden con  $\mathbb{E}(\epsilon_t) = 0$ ,  $\text{Var}(\epsilon_t) = \frac{\alpha_0}{1 - \alpha(1) - \beta(1)}$  y  $\text{Cov}(\epsilon_t, \epsilon_s) = 0$  para  $t \neq s$  si y sólo si  $\alpha(1) + \beta(1) < 1$ .<sup>7</sup>

<sup>7</sup>Demostración el Apéndice A

La diferencia entre los modelos ARCH y GARCH, radica en que para el primero la varianza condicional es simplemente una función lineal de los errores pasados, mientras que el proceso GARCH introduce las varianzas condicionales rezagadas, correspondiendo a algún tipo de mecanismo adaptativo de aprendizaje.

### 3.3.1. ACF y PACF para procesos GARCH

Las funciones de Autocorrelación (ACF) y Autocorrelación Parcial (PACF) son de gran utilidad en la identificación y prueba del comportamiento en el proceso de error al cuadrado, así como para el diagnóstico de la ecuación de varianza condicional del modelo GARCH.

Consideremos el proceso GARCH( $r,s$ ) definido por 3.5, suponiendo su momento de cuarto orden finito y siendo  $\nu_t = \epsilon_t^2 - \sigma_t^2 = (z_t^2 - 1)\sigma_t^2$ , con  $z_t \sim \text{i.i.d. } N(0, 1)$ , reescrito como

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^s \beta_j \epsilon_{t-j}^2 + \nu_t - \sum_{j=1}^s \beta_j \nu_{t-j},$$

donde multiplicando por  $\epsilon_{t-h}^2$  es posible obtener la función de autocovarianza para  $\epsilon_t^2$

$$\gamma(h) = \sum_{i=1}^r \alpha_i \gamma(h-i) + \sum_{i=1}^s \beta_i \gamma(h-i) = \sum_{i=1}^m \varphi_i \gamma(h-i), \quad h \geq s+1,$$

con  $m = \max\{r, s\}$ ,  $\varphi_i = \alpha_i + \beta_i$   $i = 1, \dots, m$ ,  $\alpha_i = 0$   $i > r$  y  $\beta_i = 0$   $i > s$ .

Dividiendo por  $\gamma(0)$  se tiene un análogo a las ecuaciones de Yule-Walker

$$\rho(h) = \sum_{i=1}^m \varphi_i \rho(h-i), \quad h \geq s+1. \quad (3.7)$$

Entonces las primeras  $s$  autocorrelaciones dependen directamente de los parámetros  $\alpha_1, \dots, \alpha_r$  y  $\beta_1, \dots, \beta_s$ , pero dadas  $\rho(s), \dots, \rho(s+1-m)$  la ecuación 3.7 determina la autocorrelación en ordenes superiores. Lo anterior es un resultado similar al de las autocorrelaciones para procesos ARMA.

Por otra parte, la función de autocorrelación parcial para  $\epsilon_t^2$  al rezago  $k$ , denotada por  $\phi_{kk}$ , se halla al resolver el conjunto de las  $k$  ecuaciones para  $\phi_{k1}, \dots, \phi_{kk}$ :

$$\rho(h) = \sum_{i=1}^k \phi_{ki} \rho(h-i), \quad h = 1, \dots, k.$$

Como el proceso  $\epsilon_t^2$  puede pensarse como un proceso autoregresivo, notamos que  $\phi_{kk}$  se anulará después del rezago  $r$  para procesos ARCH( $r$ ) i.e.

$$\phi_{kk} \begin{cases} \neq 0 & k \leq r \\ = 0 & k > r \end{cases},$$

mostrando un comportamiento idéntico para la función de autocorrelación parcial de un proceso AR( $r$ ). Se puede probar en general que la PACF para un proceso GARCH( $r,s$ ) no es cero, pero tiende a desaparecer conforme los rezagos crecen.

Sabemos que las funciones  $\rho(h)$  y  $\phi_{kk}$  no serán conocidas, entonces sus contrapartes muestrales  $\hat{\rho}(h)$  y  $\hat{\phi}_{kk}$  proporcionarán estimadores consistentes para ellas, que junto con su varianza asintótica bajo la hipótesis nula de no efectos ARCH, podrán usarse en la etapa preliminar de identificación del orden del modelo, siendo de gran ayuda para verificar el diagnóstico del mismo.

### 3.3.2. GARCH(1,1)

Los aspectos centrales de los modelos GARCH podemos observarlos en el GARCH(1,1)

$$\epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

donde  $z_t$  i.i.d.  $\sim N(0,1)$ ,  $0 \leq \alpha_1, \beta_1 < 1$ , y se cumple la condición  $(\alpha_1 + \beta_1) < 1$ .

La esperanza no condicional de  $\epsilon_t$  es cero. Notando que  $\mathbb{E}[\epsilon_t] = \mathbb{E}[\sigma_t z_t] = 0$ , pues  $\sigma_t$  no depende de la realización de  $z_t$  y  $\mathbb{E}[z_t] = 0$ , se sigue el resultado.

Como  $\epsilon_t^2 = z_t^2(\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)$ , la varianza no condicional del proceso es

$$\mathbb{E}[\epsilon_t^2] = \mathbb{E}[z_t^2(\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)]. \quad (3.8)$$

Sabiendo que  $\mathbb{E}[z_t^2] = 1$ , probemos que  $\mathbb{E}[\epsilon_{t-1}^2] = \mathbb{E}[\sigma_{t-1}^2]$ . Al seguir la ley de esperanzas iteradas  $\mathbb{E}[\epsilon_t^2] = \mathbb{E}[\mathbb{E}[\epsilon_t^2 | F_{t-1}]]$ , se desea ver que la esperanza (no condicional) de la varianza condicional es la varianza no condicional.

Retrasando la relación un periodo y escribiendo  $\mathbb{E}[\epsilon_{t-1}^2] = \mathbb{E}[\mathbb{E}[\epsilon_{t-1}^2 | F_{t-2}]]$ , se tiene  $\mathbb{E}[\epsilon_{t-1}^2] = \mathbb{E}[\sigma_{t-1}^2]$ , ya que  $\sigma_{t-1}^2$  es la esperanza condicional de  $\epsilon_{t-1}^2$  dado el conjunto de información  $F_{t-2}$ .

Por lo tanto sustituyendo en 3.8, obtenemos la igualdad

$$\mathbb{E}[\epsilon_t^2] = \alpha_0 + (\alpha_1 + \beta_1)\mathbb{E}[\epsilon_{t-1}^2],$$

que bajo condiciones de estacionariedad proporciona la varianza no condicional

$$\mathbb{E}[\epsilon_t^2] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}, \text{ satisfaciendo } (\alpha_1 + \beta_1) < 1.$$

Al considerar  $\mathbb{E}[\epsilon_t \epsilon_{t-j}] = \mathbb{E}[\sigma_t z_t \sigma_{t-j} z_{t-j}]$ , se sigue que la autocovarianza entre  $\epsilon_t$  y  $\epsilon_{t-j}$  es cero, ya que  $\sigma_t, z_{t-j}$  y  $\sigma_{t-j}$  no dependen del valor de  $z_t$  y  $\mathbb{E}[z_t] = 0$ , de donde además su autocorrelación será nula.

Pensado en la realización de pronósticos, veáse primero el proceso actualizado

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

por lo que la predicción a un paso será la esperanza condicional dada  $F_t$ , en este caso  $\mathbb{E}[\sigma_{t+1}^2 | F_t] = \alpha_0 + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2$ .

Luego, para obtener un pronóstico después de  $h$  pasos, notemos que  $\epsilon_t^2 = \sigma_t^2 z_t^2$ , entonces también  $\epsilon_{t+h}^2 = \sigma_{t+h}^2 z_{t+h}^2$ .

De  $\mathbb{E}[\epsilon_{t+h}^2 | F_t] = \mathbb{E}[\sigma_{t+h}^2 z_{t+h}^2 | F_t]$ , podemos concluir que

$$\mathbb{E}[\epsilon_{t+h}^2 | F_t] = \mathbb{E}[\sigma_{t+h}^2 | F_t], \quad (3.9)$$

pues por definición  $z_{t+h}$  es independiente de  $\sigma_{t+h}$  y  $\mathbb{E}[z_{t+h}^2 | F_t] = 1$ .

Actualizando la varianza condicional en  $h$  pasos obtenemos

$$\sigma_{t+h}^2 = \alpha_0 + \alpha_1 \epsilon_{t+h-1}^2 + \beta_1 \sigma_{t+h-1}^2,$$

donde tomando esperanza condicional se tiene

$$\mathbb{E}[\sigma_{t+h}^2 | F_t] = \alpha_0 + \alpha_1 \mathbb{E}[\epsilon_{t+h-1}^2 | F_t] + \beta_1 \mathbb{E}[\sigma_{t+h-1}^2 | F_t].$$

Por el resultado 3.9, se puede verificar que  $\mathbb{E}[\sigma_{t+h}^2 | F_t] = \alpha_0 + (\alpha_1 + \beta_1) \mathbb{E}[\sigma_{t+h-1}^2 | F_t]$ .

De esta relación en forma retrospectiva llegamos a la expresión

$$\mathbb{E}[\sigma_{t+h}^2 | F_t] = \alpha_0 [1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2 + \cdots + (\alpha_1 + \beta_1)^{h-1}] + (\alpha_1 + \beta_1)^h \sigma_t^2, \quad (3.10)$$

y si  $(\alpha_1 + \beta_1) < 1$ , el pronóstico condicional de  $\sigma_{t+h}^2$  convergerá al valor a largo plazo  $\alpha_0 / (1 - \alpha_1 - \beta_1)$  cuando  $h \rightarrow \infty$ .

Finalmente en forma análoga a 3.4, puede probarse que si  $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$ , la curtosis es

$$k(\epsilon_t) = \frac{\mathbb{E}[\epsilon_t^4]}{[\mathbb{E}[\epsilon_t^2]]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} > 3. \quad (3.11)$$

Por lo que también se obtiene una distribución con colas más pesadas que una Gaussiana.

## 3.4. Estimación de los modelos GARCH

### 3.4.1. Máxima y Cuasi Máxima Verosimilitud

La lógica de la inferencia estadística es interpretar la densidad de los datos como una función de parámetros condicional en un conjunto de resultados muestrales, conocida como *verosimilitud*.

Dado el modelo para la media y varianzas condicionales de la serie de rendimientos observada, se puede utilizar el enfoque de estimación maximizando la verosimilitud o de manera equivalente su logaritmo (*log-verosimilitud*) para ajustar los parámetros del paradigma que se ha especificado. Como la evaluación recursiva de ella es condicional en valores no observados, se produce que la estimación máximo verosímil no sea perfectamente exacta y los modelos GARCH sean definidos a partir de sus valores no condicionales esperados. Es en este punto donde se habla de una máxima verosimilitud aproximada, y en el caso de no normalidad llamada *pseudo máxima verosimilitud* o *cuasi máxima verosimilitud*.

Aunado a establecer la estimación por máxima verosimilitud (MV) o cuasi máxima verosimilitud (CMV), es necesario considerar un supuesto adicional sobre el proceso de innovación  $z_t$ , otorgándole una función de densidad de probabilidad estandarizada (con media cero y varianzas uno). El procedimiento inferirá el proceso de innovación por un filtro inverso, transformando el proceso observado  $\epsilon_t$  en un proceso no correlacionado  $z_t$ . La función log-verosimilitud usará las innovaciones estimadas para inferir las correspondientes varianzas condicionales  $\sigma_t^2$  vía sustituciones recursivas en el modelo de ecuaciones condicionales y evaluando la función log-verosimilitud objetivo apropiada.

Y dado que las observaciones  $\epsilon_t$  o sus estimados  $\hat{\epsilon}_t$  no son independientes, la especificación de la densidad conjunta tiene una interpretación como el producto de las condicionales y la correspondiente densidad marginal. Siendo  $\xi_{T-1}$  la sucesión de variables aleatorias de tal manera que contiene a  $\epsilon_0, \epsilon_1, \dots, \epsilon_{T-1}$  y asumiendo que  $\epsilon_0$  es una constante o proviene de una distribución conocida, la distribución conjunta de una muestra finita de un proceso GARCH es

$$\begin{aligned} f(\epsilon_1, \dots, \epsilon_T | \boldsymbol{\theta}) &= f(\epsilon_1 | \xi_{T-1}) f(\xi_{T-1}) \\ &= f(\epsilon_T | \xi_{T-1}) f(\epsilon_{T-1} | \xi_{T-2}) \cdots f(\epsilon_1 | \xi_0) f(\xi_0). \end{aligned}$$

Usando la hipótesis de normalidad, la función del log-verosimilitud condicional para la muestra  $\epsilon_T, \epsilon_{T-1}, \dots, \epsilon_1$  es

$$\log L(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^T \left[ \log(2\pi) + \log(\sigma_t^2) + \frac{\epsilon_t^2}{\sigma_t^2} \right], \quad (3.12)$$

donde  $T$  es el número de observaciones y  $\boldsymbol{\theta}$  el vector de los parámetros desconocidos en el modelo.

El estimador máximo verosímil es el vector de parámetros específico  $\hat{\boldsymbol{\theta}}$  que maximiza la función log-verosimilitud y a pesar de que no puede ser obtenido analíticamente, se calcula con rutinas iterativas de optimización.

Hemos mencionado que la función de verosimilitud en 3.12 se determina bajo normalidad condicional, en cambio ignorando ese supuesto en las innovaciones, la maximización de la función log-verosimilitud se justifica por la teoría de cuasi máxima verosimilitud (CMV).<sup>8</sup> De hecho la evidencia indica que para una gran variedad de procesos GARCH estacionarios, la consistencia y normalidad asintótica del estimador cuasi máximo verosímil han sido probadas.

Si la media y varianzas condicionales son correctamente especificadas y alguna distribución paramétrica para  $z_t$  puede asumirse razonablemente, métodos exactos MV pueden sustituirse por estimación CMV en términos de eficiencia, es decir, el estimador CMV será consistente con un grado de eficiencia relacionado directamente al nivel de desviación de la normalidad.

Lo anterior se debe a que empíricamente, la idea de fijar una distribución Gaussiana parece no adaptarse a datos reales, por ello varios autores proponen distribuciones alternativas para  $z_t$  de colas pesadas con el objetivo de capturar la curtosis observada de una forma más eficaz.

De otra forma, si se desea considerar innovaciones leptocúrticas incluso asimétricas, se tiene que modelar  $z_t$  con una función de probabilidad estandarizada que dependa de parámetros adicionales de forma que modifiquen la curtosis y el sesgo.

Ilustraremos enseguida la aplicación de distribuciones simétricas con un parámetro adicional modelando la curtosis, tales son la distribución Student-t y la Distribución de Error Generalizada (GED) estandarizadas.

---

<sup>8</sup>Ver Bollerslev y Wooldridge [1992]

### 3.4.2. Student-t

Una variable aleatoria  $z_t$  sigue una distribución Student-t estandarizada con  $\nu$  grados de libertad si su densidad es

$$f(z_t|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}} \left(1 + \frac{z_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad (3.13)$$

con  $\nu > 2$  y donde  $\Gamma(\cdot)$  es la función Gamma.<sup>9</sup>

En 3.13,  $\nu$  es un parámetro de forma y si  $\nu \rightarrow \infty$ , la distribución converge a la densidad Gaussiana.

En la estimación GARCH, la distribución Student-t otorga una log-verosimilitud

$$\begin{aligned} \log L(\boldsymbol{\theta}) = & T \left[ \log \Gamma\left(\frac{\nu+1}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \log[\pi(\nu-2)] \right] \\ & - \frac{1}{2} \sum_{t=1}^T \left[ (1+\nu) \log\left(1 + \frac{\epsilon_t^2}{(\nu-2)\sigma_t^2}\right) + \log(\sigma_t^2) \right], \end{aligned} \quad (3.14)$$

donde  $\nu$  son los grados de libertad. Por lo tanto en el vector  $\boldsymbol{\theta}$  existe una dependencia del parámetro  $\nu$ .

### 3.4.3. GED

Sea  $z_t$  es una variable aleatoria que sigue una Distribución de Error Generalizada (GED) estandarizada con parámetro de forma  $\nu$ . Su densidad está dada por

$$f(z_t|\boldsymbol{\theta}, \nu) = \frac{\nu}{2^{\frac{\nu+1}{\nu}} \Gamma(\frac{1}{\nu}) \lambda_\nu} \exp\left(-\frac{1}{2} \left|\frac{z_t}{\lambda_\nu}\right|^\nu\right), \quad (3.15)$$

donde

$$\lambda_\nu = \left(\frac{\Gamma(1/\nu) 2^{-2/\nu}}{\Gamma(3/\nu)}\right)^{\frac{1}{2}}, \quad \text{con } 0 < \nu \leq \infty.$$

Como casos particulares de la GED se tienen:

- Si  $\nu = 2$ , es igual a la distribución Normal estándar.

---

<sup>9</sup> $\Gamma(\alpha) = \int_0^\infty w^{\alpha-1} e^{-w} dw, \quad \alpha > 0.$

- En el caso  $\nu < 2$ , es leptocúrtica.
- Es la distribución de Laplace (doble exponencial) cuando  $\nu = 1$ .
- Para  $\nu \rightarrow \infty$ , se tiene la distribución uniforme en  $[-\sqrt{3}, \sqrt{3}]$ .

Considerando una GED estandarizada para  $z_t$ , la log-verosimilitud está dada por

$$\log L(\boldsymbol{\theta}) = \sum_{t=1}^T \left[ \log\left(\frac{\nu}{\lambda_\nu}\right) - \frac{1}{2} \left| \frac{\epsilon_t^2}{\sigma_t^2 \lambda_\nu} \right|^\nu - \left(1 + \frac{1}{\nu}\right) \log(2) - \log\Gamma\left(\frac{1}{\nu}\right) - \frac{1}{2} \log(\sigma_t^2) \right]. \quad (3.16)$$

Podemos establecer que las distribuciones anteriores cuentan con colas pesadas, así que el exceso de curtosis se incluye en ellas, sin embargo, nótese que siguen siendo simétricas.

#### 3.4.4. Criterios de Información: AIC y BIC

Para elegir el modelo apropiado, preservando la idea de parsimonia (menor número de parámetros), existen fundamentalmente dos criterios de información basados en la verosimilitud, siendo el objetivo considerar el modelo que minimice el criterio elegido.

El primero de ellos es el *Criterio de Información Akaike* (AIC), definido como

$$\text{AIC} = -\frac{2}{T} \log(MV) + \frac{2}{T} P. \quad (3.17)$$

El segundo, se trata del *Criterio de Información Bayesiano* (BIC) de Schwarz

$$\text{BIC} = -\frac{2}{T} \log(MV) + \frac{\log(T)}{T} P, \quad (3.18)$$

donde  $MV$  se refiere a la verosimilitud evaluada en el estimador máximo verosímil,  $T$  es el tamaño de muestra y  $P$  el número de parámetros.

En cada uno, la primera parte mide la bondad del ajuste a los datos y la segunda *penaliza* al modelo de acuerdo al número de parámetros incluidos.

Nótese que para el AIC, el parámetro de penalización es 2, mientras que en el BIC se emplea  $\log(T)$ . Entonces el BIC tenderá a seleccionar el modelo de orden menor cuando el tamaño de muestra sea moderado o grande.

Aunque se ha descrito la estimación para modelos GARCH, algunos otros, como los que se verán en las siguientes secciones podrán estimarse de la misma manera, solamente cambiarán la especificación del modelo y los supuestos distribucionales.

### 3.5. ARCH-M

Los modelos anteriores implican que el riesgo inherente de las actividades en los mercados financieros varía en el tiempo, y ya que en ellos se implementa una valuación riesgo-rendimiento de los instrumentos financieros, sería razonable estudiar las relaciones entre cambios en el riesgo y su repercusión en los rendimientos.

Para correlacionar variaciones en el riesgo, en este caso medidas por la volatilidad, Engle, Lilien y Robins [1987] ampliaron la noción del trabajo ARCH para que la media de la serie de rendimientos dependiera de su varianza condicional. A dicha clase de modelo se le llamó ARCH en media (ARCH-M) y particularmente se ha empleado para el estudio de instrumentos del mercado financiero.

Su idea básica es que los agentes con aversión al riesgo requieren una compensación por tener un instrumento riesgoso y dado que el riesgo de un instrumento puede medirse por la varianza de sus rendimientos, la prima de riesgo será una función creciente en la varianza condicional de los rendimientos.

Originalmente los autores describieron el ARCH-M como

$$Y_t = \mu_t + \epsilon_t, \quad (3.19)$$

donde  $Y_t$  es el exceso de rendimiento por tener a largo plazo un instrumento riesgoso en referencia a un periodo de otro activo,  $\mu_t$  es la prima de riesgo necesaria para inducir al agente adverso al riesgo a retener el instrumento en lugar de un periodo de un bono y  $\epsilon_t$  el error no pronosticable del exceso de rendimiento esperado por conservar a largo plazo el instrumento.

Para explicar 3.19, nótese que el exceso de rendimiento esperado por conservar a largo plazo el instrumento debe ser igual a la prima de riesgo i.e.  $\mathbb{E}[Y_t|F_{t-1}] = \mu_t$ . Así, suponen a la prima de riesgo como una función creciente de la varianza condicional de  $\epsilon_t$ , esto es, ante un incremento en la varianza condicional de los rendimientos, se tendrá el incremento necesario en la compensación para inducir al agente a retener el instrumento a largo plazo.

Sea  $\sigma_t^2$  la varianza condicional de  $\epsilon_t$ , entonces la prima de riesgo puede escribirse como

$$\mu_t = \beta + \delta\sigma_t^2, \quad \text{con } \beta, \delta \text{ constantes y } \sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i \epsilon_{t-i}^2. \quad (3.20)$$

Las ecuaciones 3.19 y 3.20 constituyen el modelo básico ARCH-M. De 3.20 la media condicional de  $Y_t$  depende de la varianza condicional, siendo  $\epsilon_t^2$  modelado por un proceso ARCH( $r$ ). Es de remarcar que si la varianza condicional es constante i.e.  $\alpha_i = 0$ , el modelo generará el caso tradicional de prima de riesgo constante.

Desde una perspectiva general, la modelación puede extenderse de tal manera que la varianza condicional siga un proceso GARCH, y a fin de relacionarlo con el modelo de volatilidad especificado en la ecuación 3.1, el GARCH en media (GARCH-M) puede pensarse como un modelo no lineal completo de los rendimientos (no solamente del comportamiento del error), modelados como una constante más un término proporcional a la varianza condicional. Usando la representación habitual, la dinámica de  $r_t$  bajo el modelo GARCH-M es  $r_t = \mu_t + \epsilon_t = \mu_t + \sigma_t z_t$ , donde la media y varianza condicionales son respectivamente

$$\mu_t = \beta + \delta \sigma_t^2 \quad \text{y} \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2,$$

con  $\beta$ ,  $\delta$  constantes y  $\epsilon_t^2$  sigue un proceso GARCH( $r, s$ ).

Obsérvese que un valor positivo de  $\delta$ , indicaría que el rendimiento estaría positivamente correlacionado con su volatilidad. Su existencia es una de las razones por las que algunas series históricas presentan correlaciones *serias*, por lo que el GARCH-M admite su presencia, introduciéndolas en el proceso  $\sigma_t^2$ .

### 3.6. IGARCH

Considereremos la siguiente representación del modelo GARCH, observando que se trata de un proceso autoregresivo de promedios móviles de ordenes  $\max(r, s)$  y  $s$  para  $\epsilon_t^2$

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) \epsilon_{t-i} + \nu_t - \sum_{j=1}^s \beta_j \nu_{t-j},$$

con  $\nu_t = \epsilon_t^2 - \sigma_t^2$ ,  $\alpha_i = 0$  para  $i > r$  y  $\beta_i = 0$  para  $i > s$ .

Para satisfacer la condición de estacionariedad debe cumplirse  $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) < 1$ , de tal forma el error en la varianza condicional tendrá un efecto decreciente para algún pronóstico ( $\sigma_{t+h}^2$ ) cuando  $h$  crezca y será asintóticamente no significativo.

Ahora, en la mayoría de los ajustes con datos de alta frecuencia se ha observado regularmente que la suma de los coeficientes  $\alpha$ 's y  $\beta$ 's es muy cercana a la unidad, por lo que el imponer la condición  $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) = 1$  nos lleva al proceso GARCH integrado (IGARCH).

Una expresión equivalente a la ecuación GARCH( $r, s$ ) es

$$[1 - \alpha(B) - \beta(B)]\epsilon_t^2 = \alpha_0 + [1 - \beta(B)][\epsilon_t^2 - \sigma_t^2], \quad (3.21)$$

donde si el polinomio  $1 - \alpha(1) - \beta(1) = 0$  tiene una raíz unitaria, o lo que es lo mismo  $(\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j) = 1$ , se tiene el modelo IGARCH( $r,s$ ) de Engle y Bollerslev [1986].

La ecuación 3.21 puede reescribirse como  $\varphi(B)(1 - B)\epsilon_t^2 = \alpha_0 + [1 - \beta(B)](\epsilon_t^2 - \sigma_t^2)$ , para  $\varphi(B) = [1 - \alpha(B) - \beta(B)](1 - B)^{-1}$  de orden  $\max\{r, s\} - 1$ .

Reordenando algunos términos en 3.21, podemos expresar a la varianza condicional como una función de los errores de la siguiente manera

$$\sigma_t^2 = \alpha_0[1 - \beta(1)]^{-1} + \{1 - \varphi(B)(1 - B)[1 - \beta(B)]^{-1}\}\epsilon_t^2,$$

recordando a la escritura del proceso GARCH como un ARCH de orden infinito

$$\sigma_t^2 = \alpha_0 \left(1 - \sum_{j=1}^s \beta_j\right)^{-1} + \sum_{i=1}^{\infty} \delta_i \epsilon_{t-i}^2,$$

con  $\delta_i$  los coeficientes de rezago dependientes no linealmente de  $\alpha_i$  y  $\beta_j$ .

Davidson [2004] define la *amplitud* como la medida de las variaciones en  $\sigma_t^2$ , dada por

$$S = \sum_{i=1}^{\infty} \delta_i,$$

indicando cuan grandes son los cambios en la varianza condicional y por tanto el orden de los momentos existentes. Luego introduce la *memoria* del modelo  $\rho$ , la cual determina la longitud, en el sentido del número de rezagos, donde los errores tenderán a disiparse en la volatilidad.

Volviendo al GARCH(1,1), sabemos es estacionario de segundo orden si  $\alpha_1 + \beta_1 < 1$ , lo que implica que su varianza no condicional existe y es finita. Así, se establece que el GARCH(1,1) tiene amplitud  $S = \alpha_1/(1 - \beta_1)$  y una memoria geométrica  $\rho = 1/\beta_1$ .

Considerando un caso particular del mismo, con  $\alpha_1 = 1 - \beta_1$ , se encuentra el modelo IGARCH(1,1)

$$\epsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \alpha_0 + (1 - \beta_1)\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2,$$

donde  $z_t \sim \text{i.i.d. } N(0, 1)$  y  $0 < \beta_1 < 1$ .

Cuyas características notables son:

- Su amplitud es  $S = 1$ .

- La varianza no condicional  $\alpha_0/(1 - \alpha_1 - \beta_1)$  no es finita, por lo tanto su segundo momento no existe y el proceso no es débilmente estacionario.
- Nelson [1990] prueba que si  $\mathbb{E}[\log(\beta_1 + \alpha_1 z_t^2)] < 0$ , el IGARCH(1,1) es estrictamente estacionario.

También de la ecuación 3.10, podemos ver que el pronóstico de la varianza condicional después de  $h$  pasos es  $\mathbb{E}[\sigma_{t+h}^2 | F_t] = h\alpha_0 + \sigma_t^2$ .

Excepto por el término de  $\alpha_0$ , es visible que el pronóstico al siguiente periodo es el valor de la varianza condicional actual, por ese motivo se dice que es persistente en la varianza.

Para ver el comportamiento de la memoria, usando el operador de retraso  $\sigma_{t-1}^2 = B\sigma_t^2$ , escribimos el IGARCH(1,1) como

$$\sigma_t^2 = \alpha_0 + (1 - \beta_1)\epsilon_{t-1}^2 + \beta_1 B\sigma_t^2,$$

donde resolviendo para  $\sigma_t^2$ , se tiene la solución

$$\sigma_t^2 = \alpha_0/(1 - \beta_1) + (1 - \beta_1) \sum_{i=0}^{\infty} \beta_1^i \epsilon_{t-1-i}^2.$$

Por lo tanto, la varianza condicional es una función que decae geométricamente en las realizaciones de los errores pasados.

### 3.7. EGARCH

Black [1976] encontró evidencia de correlaciones negativas entre los rendimientos y cambios en la volatilidad (*asimetría*), mostrando que una reducción en el valor de un instrumento incrementaba su riesgo, manifestado en un aumento de su volatilidad futura.

Actualmente es comprobable que la volatilidad tiende a subir en respuesta a “malas noticias”(rendimientos por debajo de lo esperado), y en cambio disminuye a causa de “buenas noticias”(rendimientos mejores de lo esperado).

En el contexto de los modelos GARCH, este hecho parece inexistente, pues se observa que carecen de la capacidad de capturar la asimetría, ya que sólo asumen a la varianza condicional como una función de las varianzas condicionales anteriores y las innovaciones cuadráticas pasadas, pero el signo de los rendimientos no juega ningún papel. Otra limitación proviene las restricciones impuestas en los parámetros para asegurar la no negatividad de la varianza condicional, muchas veces no respetadas por los coeficientes estimados y que coartan la dinámica del proceso de  $\sigma_t^2$ .

Lo anterior sugeriría la motivación a introducir un modelo donde la volatilidad responda asimétricamente a la positividad o negatividad de los rendimientos y además no exista problemática alguna en los parámetros. Nelson [1991] lo logró en su modelo GARCH exponencial (EGARCH), adoptando en él una forma distinta para asegurar que  $\sigma_t^2$  sea no negativa y haciendo que  $\log(\sigma_t^2)$  sea lineal en alguna función del tiempo junto con las  $z_t$ 's rezagadas.

Esto es, para una función adecuada  $g(\cdot)$ , se tiene que

$$\log(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(z_{t-k}), \quad (3.22)$$

con  $\{\alpha_t\}$ ,  $\{\beta_k\}$  sucesiones de números reales y  $\beta_1 \equiv 1$ .

Para capturar los efectos de la asimetría de rendimientos positivos y negativos,  $g(z_t)$  debe ser una función de la magnitud y el signo de  $z_t$ , por ejemplo, una combinación lineal de  $z_t$  y  $|z_t|$ . Nelson [1991] considera

$$g(z_t) = \theta z_t + \gamma [|z_t| - \mathbb{E}(|z_t|)],$$

donde  $\theta$ ,  $\gamma$  son números reales positivos y las componentes  $\theta z_t$  y  $\gamma [|z_t| - \mathbb{E}(|z_t|)]$  son i.i.d. con media cero continuas. Obsérvese que por construcción  $g(z_t)$  es una sucesión aleatoria i.i.d. con media cero, o sea, se trata de un proceso de innovación. Podemos decir que la primera de la función representa el efecto del signo y la segunda el efecto de la magnitud de los errores.

La asimetría se observa al reescribir

$$g(z_t) = \begin{cases} (\theta + \gamma)z_t - \gamma\mathbb{E}(|z_t|) & \text{si } z_t \geq 0 \\ (\theta - \gamma)z_t - \gamma\mathbb{E}(|z_t|) & \text{si } z_t < 0 \end{cases}.$$

Formalmente el modelo EGARCH( $r, s$ ) puede parametrizarse como

$$\epsilon_t = \sigma_t z_t, \quad \log(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 B + \dots + \beta_1 B^s}{1 - \alpha_1 B - \dots - \alpha_r B^r} g(z_{t-1}), \quad (3.23)$$

donde  $\alpha_0$  es una constante,  $B$  el operador de retraso tal que  $Bg(z_t) = g(z_{t-1})$  y los polinomios  $1 + \beta_1 B + \dots + \beta_1 B^s$  y  $1 - \alpha_1 B - \dots - \alpha_r B^r$  no tienen factores en común y sus raíces se encuentran fuera del círculo unitario.

El uso de la transformación  $\log(\cdot)$  en la varianza condicional precisamente asegura que  $\sigma_t^2$  será siempre positiva y la media no condicional de  $\log(\sigma_t^2)$  es  $\alpha_0 / (1 - \sum_{i=1}^r \alpha_i)$ .

Es importante establecer que  $\mathbb{E}(|z_t|)$  depende de la distribución no condicional adoptada para  $z_t$ . Para el caso de distribución Normal, es cierto que  $\mathbb{E}(|z_t|) = \sqrt{2/\pi}$ .

Empleando una distribución Student-t (simétrica) se tiene

$$\mathbb{E}(|z_t|) = \frac{2\Gamma(\frac{1+\nu}{2})\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)}.$$

Y con una Distribución de Error Generalizada (GED) el resultado es

$$\mathbb{E}(|z_t|) = 2^{1/\nu}\lambda_\nu \frac{\Gamma(2/\nu)}{\Gamma(1/\nu)},$$

recordando que los parámetros  $\nu$  y  $\lambda_\nu$  conciernen a la forma de la distribución.

### 3.7.1. EGARCH(1,0)

Considérese el modelo EGARCH(1,0)

$$\epsilon_t = \sigma_t z_t, \quad (1 - \alpha B)\log(\sigma_t^2) = \alpha_0 + g(z_{t-1}), \quad \text{con } z_t \sim \text{i.i.d. } N(0, 1).$$

En este caso suponiendo normalidad  $\mathbb{E}(|z_t|) = \sqrt{2/\pi}$ . Así que

$$(1 - \alpha B)\log(\sigma_t^2) = \begin{cases} \alpha_* + (\theta + \gamma)z_{t-1} & \text{si } z_{t-1} \geq 0 \\ \alpha_* + (\theta - \gamma)z_{t-1} & \text{si } z_{t-1} < 0 \end{cases}, \quad \text{con } \alpha_* = \alpha_0 - \sqrt{2/\pi}.$$

Para este modelo EGARCH, la varianza condicional evoluciona de una manera no lineal dependiendo del signo de  $\epsilon_{t-1}$ . Explícitamente

$$\sigma_t^2 = (\sigma_{t-1}^2)^\alpha \exp(\alpha_*) \begin{cases} \exp[(\theta + \gamma)\frac{\epsilon_{t-1}}{\sigma_{t-1}}] & \text{si } \epsilon_{t-1} \geq 0 \\ \exp[(\theta - \gamma)\frac{\epsilon_{t-1}}{\sigma_{t-1}}] & \text{si } \epsilon_{t-1} < 0 \end{cases},$$

donde los coeficientes  $(\theta + \gamma)$  y  $(\theta - \gamma)$  muestran la asimetría en respuesta a valores positivos y negativos de  $\epsilon_{t-1}$ .

## Capítulo 4

# Modelos MGARCH

Distintos elementos, como la globalización económica, el desarrollo de las comunicaciones e introducción de nuevas tecnologías, han acelerado la integración de los mercados financieros mundiales durante la última década. A su vez los movimientos en los precios de un instrumento financiero repercuten rápidamente en las transacciones de muchos activos, razón por la que son más dependientes entre sí a como eran antes. Por lo tanto ya no sólo deben considerarse individualmente, sino en conjunto para obtener un mejor entendimiento de la estructura de las finanzas globales.

El hecho que un mercado puede *dirigir* a otro bajo ciertas circunstancias, pero la relación puede revertirse en algunas situaciones, hará fundamental (para un inversionista o institución) el conocimiento de la interrelación entre las bolsas de valores y economías, ya que la dinámica entre ellos jugará un papel relevante en el manejo de los riesgos.

Hemos visto que el análisis y predicción de la dependencia temporal en los momentos de segundo orden de los rendimientos de instrumentos se ha convertido en un tema muy importante en econometría y administración financiera. Reconocer estos aspectos a través de la modelación multivariada abre la posibilidad para tomar mejores decisiones en áreas como modelos de valuación y selección de portafolios, cobertura en futuros y estimación del Valor en Riesgo (VaR).

Existe una literatura extensa de modelos univariados ARCH concernientes a la varianza dinámica en el tiempo, asimetría y densidades colas pesadas; pero hasta ahora contados artículos se refieren a su noción multivariada.<sup>10</sup>

Comenzaremos presentando el modelo multivariado que se supone en lo sucesivo.

---

<sup>10</sup>Se sugiere consultar el Apéndice B, donde se encontrará una revisión de Teoría de Matrices y Análisis Multivariado.

### Especificación multivariada

Considérese un vector  $\mathbf{r}_t = \{r_{i,t}\}$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ , del proceso de rendimientos, de tal forma que en cada momento  $t$  puede representarse como

$$\mathbf{r}_t = \boldsymbol{\mu}_t(\boldsymbol{\theta}) + \boldsymbol{\epsilon}_t,$$

donde  $\boldsymbol{\mu}_t(\boldsymbol{\theta})$  es el vector de medias condicionales y el término  $\boldsymbol{\epsilon}_t$  se escribe como

$$\boldsymbol{\epsilon}_t = \mathbf{H}_t^{1/2}(\boldsymbol{\theta})\mathbf{z}_t,$$

con  $\mathbf{H}_t^{1/2}(\boldsymbol{\theta})$  una matriz positiva definida y notando que en ambos existe una dependencia del vector de parámetros desconocidos  $\boldsymbol{\theta}$ .

Además se asume que el vector aleatorio  $\mathbf{z}_t$  tiene los primeros dos momentos:

$$\mathbb{E}[\mathbf{z}_t] = 0, \quad \text{Var}(\mathbf{z}_t) = \mathbf{I}_N.$$

Para explicar la naturaleza de  $\mathbf{H}_t^{1/2}$ , veamos que condicionando en el conjunto de información disponible al tiempo inmediato anterior ( $F_{t-1}$ ) y omitiendo por un momento la dependencia en  $\boldsymbol{\theta}$ , podemos escribir

$$\text{Var}(\mathbf{r}_t|F_{t-1}) = \text{Var}_{t-1}(\mathbf{r}_t) = \text{Var}_{t-1}(\boldsymbol{\epsilon}_t) = \mathbf{H}_t^{1/2}\text{Var}_{t-1}(\mathbf{z}_t)(\mathbf{H}_t^{1/2})' = \mathbf{H}_t.$$

Entonces  $\mathbf{H}_t^{1/2}$  es alguna matriz positiva definida, tal que  $\mathbf{H}_t$  es la varianza condicional del proceso  $\mathbf{r}_t$ . Por ejemplo,  $\mathbf{H}_t^{1/2}$  puede obtenerse a través de la factorización de Cholesky de  $\mathbf{H}_t$ .

En este capítulo se revisarán diferentes especificaciones para la matriz  $\mathbf{H}_t$ , teniendo como objetivos fundamentales en la modelación:

- Determinar las condiciones en los parámetros de  $\boldsymbol{\theta}$  para asegurar que la matriz  $\mathbf{H}_t$  sea positiva definida.
- Mantener una estimación factible, reduciendo el número de parámetros, pero otorgando flexibilidad a la dinámica de  $\mathbf{H}_t$ .
- Encontrar la matriz de varianza no condicional  $\boldsymbol{\Sigma}$ , tal que  $\boldsymbol{\Sigma} = \text{Var}(\mathbf{r}_t) = \mathbb{E}[\mathbf{H}_t]$ , estableciendo las condiciones de estacionariedad débil para el modelo.

El tratamiento de las formulaciones paramétricas se divide en las siguientes secciones en tres categorías:

1. Modelos VEC y BEKK, que pueden verse como generalizaciones directas del modelos univariados GARCH al caso multivariado.
2. Modelos de factores, basados en decomposiciones de la matriz de heteroscedasticidad condicional y contando con diferentes especificaciones.
3. Modelos de correlación condicional, las cuales que emplean una clase de combinaciones no lineales de procesos GARCH univariados, y donde precisamente la correlación condicional primero es vista desde una perspectiva constante, para luego introducir en ella cierta dinámica.

## 4.1. Generalizaciones directas GARCH

### 4.1.1. VEC

Una especificación general propuesta por Bollerslev, Engle y Wooldridge [1988] para la matriz de varianza  $\mathbf{H}_t$  es el modelo **VEC**, donde cada elemento de  $\mathbf{H}_t$  es una función lineal de los errores cuadráticos rezagados, productos cruzados de errores rezagados y sus valores pasados.

El modelo general de órdenes  $(r, s)$  puede representarse como

$$\text{vech}(\mathbf{H}_t) = \mathbf{C} + \sum_{i=1}^r \mathbf{A}_i \text{vech}(\boldsymbol{\epsilon}_{t-i} \boldsymbol{\epsilon}'_{t-i}) + \sum_{j=1}^s \mathbf{B}_j \text{vech}(\mathbf{H}_{t-j}), \quad (4.1)$$

donde  $\mathbf{C}$  es un vector de dimensión  $N(N+1)/2 \times 1$ ,  $\mathbf{A}_i$  y  $\mathbf{B}_j$  matrices de parámetros ambas con dimensión  $N(N+1)/2 \times N(N+1)/2$  y  $\text{vech}(\cdot)$  es el operador que vectoriza la parte diagonal inferior de una matriz simétrica.

Vista como una función de un rezago de  $\boldsymbol{\epsilon}_t$  y un valor pasado de sí misma (como en los GARCH),  $\mathbf{H}_t$  tiene su forma simplificada en el VEC(1,1)

$$\text{vech}(\mathbf{H}_t) = \mathbf{C} + \mathbf{A} \text{vech}(\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}'_{t-1}) + \mathbf{B} \text{vech}(\mathbf{H}_{t-1}), \quad (4.2)$$

con  $\mathbf{C}$ ,  $\mathbf{A}$  y  $\mathbf{B}$  (matrices de parámetros) definidas como antes.

En el caso bivariado se puede escribir

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix},$$

de donde cada ecuación será respectivamente

$$\begin{aligned} h_{11,t} &= c_1 + a_{11}\epsilon_{1,t-1}^2 + a_{12}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{13}\epsilon_{2,t-1}^2 + b_{11}h_{11,t-1} + b_{12}h_{12,t-1} + b_{13}h_{22,t-1}, \\ h_{12,t} &= c_2 + a_{21}\epsilon_{1,t-1}^2 + a_{22}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{23}\epsilon_{2,t-1}^2 + b_{21}h_{11,t-1} + b_{22}h_{12,t-1} + b_{23}h_{22,t-1}, \\ h_{22,t} &= c_3 + a_{31}\epsilon_{1,t-1}^2 + a_{32}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{33}\epsilon_{2,t-1}^2 + b_{31}h_{11,t-1} + b_{32}h_{12,t-1} + b_{33}h_{22,t-1}. \end{aligned}$$

El modelo proporciona la ventaja en el sentido de interpretación, pero a la vez tiene la dificultad de garantizar que  $\mathbf{H}_t$  sea positiva definida sin imponer condiciones restrictivas para los parámetros. Gouriéroux [1997] lo muestra para el caso ARCH, es decir, sin incluir rezagos de  $\mathbf{H}_t$ .

Ahora, en la versión no restringida existen interacciones entre todos los elementos de la matriz de varianzas condicionales obteniendo  $(r+s)[N(N+1)/2]^2 + N(N+1)/2$  parámetros, el cual es un gran número a menos que  $N$  sea pequeña. Lo que lleva en la práctica a utilizarlo sólo para caso bivariados.

Un modelo VEC(1,1) trivariado incluyendo la modelación de la media condicional con un vector autoregresivo requerirá la estimación de 90 parámetros. En cambio parece que dicho ajuste no mejora los GARCH(1,1) parsimoniosos, probablemente por el hecho de que al permitir a cada elemento de  $\mathbf{H}_t$  a ser una función de cada elemento de  $\mathbf{H}_{t-1}$ , dentro del vector  $\text{vech}(\epsilon_{t-1}\epsilon'_{t-1})$  se producen estimadores *ruidosos*, en el sentido que no proporcionan mayor información.

Para reducir los parámetros, Bollerslev, Engle y Wooldridge simplifican el modelo, haciendo que las matrices  $\mathbf{A}_i$  y  $\mathbf{B}_j$  sean diagonales, llamándolo VEC diagonal (**DVEC**). Esto facilita la estimación ya que cada ecuación se puede obtener por separado, teniéndose que existen  $(r+s+1)N(N+1)/2$  parámetros y cada varianza  $h_{ii,t}$  depende únicamente de su propio error cuadrado pasado  $\epsilon_{i,t-1}^2$  y su rezago anterior  $h_{ii,t-1}$ . También la covarianza  $h_{ij}$  contiene el producto cruzado pasado  $\epsilon_{i,t-1}\epsilon_{j,t-1}$  y el rezago  $h_{ij}$ . De tal forma no existe la interacción entre las diferentes varianzas y covarianzas condicionales, pero se tiene un modelo con un menor número de parámetros.

El modelo DVEC(1,1) bivariado puede escribirse como

$$\begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix},$$

o bien

$$\begin{aligned} h_{11,t} &= c_1 + a_{11}\epsilon_{1,t-1}^2 + b_{11}h_{11,t-1}, \\ h_{12,t} &= c_2 + a_{22}\epsilon_{1,t-1}\epsilon_{2,t-1} + b_{22}h_{12,t-1}, \\ h_{22,t} &= c_3 + a_{33}\epsilon_{2,t-1}^2 + b_{33}h_{22,t-1}. \end{aligned}$$

Observamos que cada elemento de  $\mathbf{H}_t$  depende solamente de su valor pasado y el correspondiente producto de los errores, donde cada ecuación es del tipo GARCH(1,1).

Para verificar las condiciones de estacionariedad, considérese el VEC(1,1) reescrito como

$$h_t = \mathbf{C} + \mathbf{A}\eta_{t-1} + \mathbf{B}h_{t-1}, \quad (4.3)$$

donde  $h_t = \text{vech}(\mathbf{H}_t)$ ,  $\eta_t = \text{vech}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')$ ,  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{C}$  son las matrices de parámetros. Si  $w_t = \eta_t - h_t$ , la ecuación 4.3 puede expresarse como

$$\eta_t = \mathbf{C} + (\mathbf{A} + \mathbf{B})\eta_{t-1} + w_t - \mathbf{B}w_{t-1},$$

tratándose de un (vector) ARMA(1,1) para  $\eta_t$ .

Como resultado, el modelo VEC(1,1) es estacionario si los eigenvalores de  $\mathbf{A} + \mathbf{B}$  son menores en módulo que la unidad. En tal caso su matriz de varianza es

$$\text{vech}(\boldsymbol{\Sigma}) = \mathbb{E}[\text{vech}(\mathbf{H}_t)] = (I_{N^*} - \mathbf{A} - \mathbf{B})^{-1}\mathbf{C},$$

donde  $N^* = N(N + 1)/2$ .

En general, el modelo 4.1 es estacionario de segundo orden si y sólo si todos los eigenvalores de  $\sum_{i=1}^r \mathbf{A}_i + \sum_{j=1}^s \mathbf{B}_j$  son menores que uno en módulo. Usando este hecho se dan las condiciones de positividad para el modelo VEC.<sup>11</sup>

Por definición de estacionariedad

$$\begin{aligned} h_{ii,t-1} &= \mathbb{E}_{t-2}[\epsilon_{i,t-1}^2], \quad i = 1, 2, \\ h_{12,t-1} &= \mathbb{E}_{t-2}[\epsilon_{1,t-1}\epsilon_{2,t-2}], \end{aligned}$$

ahora, escríbanse

<sup>11</sup>Ver Engle y Kroner [1995] y Gouriéroux [1997]

$$\begin{aligned}
h_{11,t} &= c_1 + \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{11} & \frac{a_{12}}{2} \\ \frac{a_{12}}{2} & a_{13} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\
&\quad + \mathbb{E}_{t-2} \left[ \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} b_{11} & \frac{b_{12}}{2} \\ \frac{b_{12}}{2} & b_{13} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \right], \\
h_{12,t} &= c_2 + \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{21} & \frac{a_{22}}{2} \\ \frac{a_{22}}{2} & a_{23} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\
&\quad + \mathbb{E}_{t-2} \left[ \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} b_{21} & \frac{b_{22}}{2} \\ \frac{b_{22}}{2} & b_{23} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \right], \\
h_{22,t} &= c_3 + \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{31} & \frac{a_{32}}{2} \\ \frac{a_{32}}{2} & a_{33} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\
&\quad + \mathbb{E}_{t-2} \left[ \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} b_{31} & \frac{b_{32}}{2} \\ \frac{b_{32}}{2} & b_{33} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \right].
\end{aligned}$$

Uniendo cada una de las componentes de  $\mathbf{H}_t$  se tiene

$$\begin{aligned}
\mathbf{H}_t &= \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t-1} & 0 \\ \epsilon_{2,t-1} & 0 \\ 0 & \epsilon_{1,t-1} \\ 0 & \epsilon_{2,t-1} \end{pmatrix}' \begin{pmatrix} a_{11} & \frac{a_{12}}{2} & a_{21} & \frac{a_{22}}{2} \\ \frac{a_{12}}{2} & a_{13} & \frac{a_{22}}{2} & a_{23} \\ a_{21} & \frac{a_{22}}{2} & a_{31} & \frac{a_{32}}{2} \\ \frac{a_{22}}{2} & a_{23} & \frac{a_{32}}{2} & a_{33} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} & 0 \\ \epsilon_{2,t-1} & 0 \\ 0 & \epsilon_{1,t-1} \\ 0 & \epsilon_{2,t-1} \end{pmatrix} \\
&\quad + \mathbb{E}_{t-2} \left[ \begin{pmatrix} \epsilon_{1,t-1} & 0 \\ \epsilon_{2,t-1} & 0 \\ 0 & \epsilon_{1,t-1} \\ 0 & \epsilon_{2,t-1} \end{pmatrix}' \begin{pmatrix} b_{11} & \frac{b_{12}}{2} & b_{21} & \frac{b_{22}}{2} \\ \frac{b_{12}}{2} & b_{13} & \frac{b_{22}}{2} & b_{23} \\ b_{21} & \frac{b_{22}}{2} & b_{31} & \frac{b_{32}}{2} \\ \frac{b_{22}}{2} & b_{23} & \frac{b_{32}}{2} & b_{33} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} & 0 \\ \epsilon_{2,t-1} & 0 \\ 0 & \epsilon_{1,t-1} \\ 0 & \epsilon_{2,t-1} \end{pmatrix} \right].
\end{aligned}$$

Denotando por  $\tilde{\mathbf{A}}$  y  $\tilde{\mathbf{B}}$  a la matrices de parámetros escritas arriba, una matriz general para  $\mathbf{H}_t$  en el modelo VEC(1,1) es

$$\mathbf{H}_t = \mathbf{C} + (\mathbf{I}_N \otimes \epsilon'_{t-1}) \tilde{\mathbf{A}} (\mathbf{I}_N \otimes \epsilon_{t-1}) + \mathbb{E}_{t-2} [(\mathbf{I}_N \otimes \epsilon'_{t-1}) \tilde{\mathbf{B}} (\mathbf{I}_N \otimes \epsilon_{t-1})],$$

con  $\otimes$  el producto de Kronecker. De donde las condiciones suficientes para la positividad de  $\mathbf{H}_t$  son:

- $\mathbf{C} > 0$ ,  $\tilde{\mathbf{A}} \geq 0$ ,  $\tilde{\mathbf{B}} \geq 0$ . Ya que algunas restricciones no son fáciles de imponer, por ejemplo la asignación de valores iniciales, de entrada no se exigen, pero pueden verificarse después de la estimación no restringida.

El gran número de parámetros desconocidos es el obstáculo para estimar los modelos VEC, incluso después de imponer varias restricciones, por lo que es difícil emplearlos cuando existen más de 3 o 4 series a analizar. En el último capítulo para nuestro ejemplo bivariado se realiza la estimación del modelo VEC diagonal con el programa estadístico S-PLUS.

#### 4.1.2. BEKK

Una nueva parametrización para  $\mathbf{H}_t$  es propuesta por Engle y Kroner [1995] en el modelo BEKK, acrónimo de Baba-Engle-Kraft-Kroner, quienes trabajaron previamente para desarrollarlo. Su atractivo es que las matrices de varianza condicional son positivo definidas por construcción y se reduce el número de parámetros.

El modelo BEKK de ordenes  $r$  y  $s$ ,  $\text{BEKK}(r,s,K)$ , lo describe

$$\mathbf{H}_t = \mathbf{C}^{*'} \mathbf{C}^* + \sum_{i=1}^r \sum_{k=1}^K \mathbf{A}_{ki}^{*'} (\epsilon_{t-i} \epsilon'_{t-i}) \mathbf{A}_{ki}^* + \sum_{j=1}^s \sum_{k=1}^K \mathbf{B}_{kj}^{*'} \mathbf{H}_{t-j} \mathbf{B}_{kj}^*, \quad (4.4)$$

donde  $\mathbf{C}^*$  es una matriz triangular inferior,  $\mathbf{A}_{ki}^*$ ,  $\mathbf{B}_{kj}^*$  son matrices de parámetros y el término  $K$  determina la generalidad del modelo. La imposición de  $\mathbf{C}^*$  a ser triangular inferior es para propósitos de identificación, de tal manera que la descomposición  $\mathbf{C}^{*'} \mathbf{C}^*$  vista como el producto de dos matrices triangulares asegura que  $\mathbf{H}_t$  sea positiva definida.

Si bien esta especificación permite dependencia dinámica entre la volatilidad de las series, los parámetros en  $\mathbf{A}_{ik}^*$  y  $\mathbf{B}_{jk}^*$  no tienen interpretación directa de los valores rezagados en los errores. También, el número de parámetros necesarios es  $(r+s)KN^2 + N(N+1)/2$ , que conforme  $r$  y  $s$  se incrementen, será bastante grande.

Para  $K > 1$  existen otras parametrizaciones que arrojan la misma representación, sin embargo, Engle y Kroner [1995] establecen los supuestos para eliminar la redundancia de representaciones observacionales equivalentes y para evitar estructuras semejantes dan condiciones suficientes para identificar modelos BEKK con  $K = 1$ . Tales son que las matrices  $\mathbf{A}_{ki}^*$ ,  $\mathbf{B}_{kj}^*$  y los elementos de la diagonal de  $\mathbf{C}^*$  sea todos positivos.

Considerando  $r$  y  $s$  ambos como uno, se tiene el modelo  $\text{BEKK}(1,1,K)$

$$\mathbf{H}_t = \mathbf{C}^{*'} \mathbf{C}^* + \sum_{k=1}^K \mathbf{A}_k^{*'} (\epsilon_{t-1} \epsilon'_{t-1}) \mathbf{A}_k^* + \sum_{k=1}^K \mathbf{B}_k^{*'} \mathbf{H}_{t-1} \mathbf{B}_k^*, \quad (4.5)$$

análogamente  $\mathbf{C}^*$ ,  $\mathbf{A}_k^*$  y  $\mathbf{B}_k^*$  son matrices de parámetros,  $\mathbf{C}^*$  es triangular inferior y el índice  $K$  determina la generalidad del modelo.

Si  $K = 1$ , la representación del caso bivariado es

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{11}^* & 0 \\ c_{21}^* & c_{22}^* \end{bmatrix} \begin{bmatrix} c_{11}^* & c_{21}^* \\ 0 & c_{22}^* \end{bmatrix} + \begin{bmatrix} a_{11}^* & a_{21}^* \\ a_{12}^* & a_{22}^* \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{1,t-1}\epsilon_{2,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \\ + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} + \begin{bmatrix} b_{11}^* & b_{21}^* \\ b_{12}^* & b_{22}^* \end{bmatrix} \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} b_{11}^* & b_{12}^* \\ b_{21}^* & b_{22}^* \end{bmatrix}.$$

Requiriendo 11 parámetros en lugar de 21 del VEC(1,1) y con la misma estructura lineal, aunque sin la explicación de los parámetros. Al incrementar  $K$  se tendrá una especificación manejable, por ejemplo si  $N = K = 2$  aparecerán 19 parámetros que comparados con el VEC son un número cercano, con la diferencia de no representar directamente el impacto de los distintos términos de rezago de  $\mathbf{H}_t$ .

A fin de reducir la parametrización es posible imponer que  $\mathbf{A}_{ki}^*$ ,  $\mathbf{B}_{kj}^*$  sean diagonales, llevando al BEKK diagonal, donde nuevamente podrá verse como un DVEC, pero con la desventaja de no garantizar que la matriz de varianza condicional sea positiva definida.

Un resultado importante implica que cada BEKK representa un único modelo VEC y genera matrices de varianza condicionales positivo definidas. Engle y Kroner [1995] proporcionan condiciones suficientes para que los modelos VEC y BEKK sean equivalentes y también proponen un Teorema de representación estableciendo la igualdad de modelos VEC diagonales (con matrices de varianza condicional positivo definidas) y modelos BEKK diagonales en general. Cuando el número de parámetros en el modelo BEKK es menor que al correspondiente número en el VEC, la parametrización BEKK impone restricciones que los hacen diferentes, pero incrementando la generalidad a través de  $K$  se llegará al mismo modelo que usando un VEC.

Para las condiciones de estacionariedad, es posible establecerlas para  $K = 1$  y luego generalizar dicho resultado.

Considérese el modelo BEKK(1,1,1)

$$\mathbf{H}_t = \mathbf{C}^{*'} \mathbf{C}^* + \mathbf{A}^{*'} (\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}') \mathbf{A}^* + \mathbf{B}^{*'} \mathbf{H}_{t-1} \mathbf{B}^*, \quad (4.6)$$

escrito como un VEC (aplicando en 4.5 el operador  $\text{vec}$  que vectoriza una matriz)<sup>12</sup> de la forma

$$\text{vec}(\mathbf{H}_t) = \text{vec}(\mathbf{C}^{*'} \mathbf{C}^*) + (\mathbf{A}^* \otimes \mathbf{A}^*)' \text{vec}(\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}') + (\mathbf{B}^* \otimes \mathbf{B}^*)' \text{vec}(\mathbf{H}_{t-1}). \quad (4.7)$$

Por tanto el modelo BEKK es débilmente estacionario si todos los eigenvalores de  $(\mathbf{A}^* \otimes \mathbf{A}^*) + (\mathbf{B}^* \otimes \mathbf{B}^*)$  son menores que uno en módulo.

<sup>12</sup>Usando la propiedad  $\text{vec}(ABC) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(B)$

Entonces su matriz de varianza no condicional  $\Sigma = \mathbb{E}[\mathbf{H}_t]$ , está dada por

$$\text{vec}(\Sigma) = \left( \mathbf{I}_{N^2} - (\mathbf{A}^* \otimes \mathbf{A}^*)' + (\mathbf{B}^* \otimes \mathbf{B}^*)' \right)^{-1} \text{vec}(\mathbf{C}^* \mathbf{C}^{*\prime}). \quad (4.8)$$

Por su estructura y dada la positividad de la matriz  $\mathbf{H}_t$ , encontramos diversas aplicaciones que logran la estimación del BEKK, en el programa estadístico R existe un paquete llamado `bekk`. Nosotros empleamos la función disponible en S-PLUS ajustándolo de manera favorable.

## 4.2. Modelos de Factores

Ciertamente su idea proviene de la teoría económica, Ross [1976] en su publicación “*Teoría de Valuación en Arbitraje*” propone que los rendimientos se generan por un número común de componentes o factores fundamentales no observados.

Los modelos de factores referidos a la especificación MGARCH, pueden ser considerados como variables heteroscedásticas, donde cada componente del proceso se expresa como una función lineal de algunos procesos, llamados *factores*, más un proceso de ruido. Suponen que las observaciones son generadas de factores subyacentes condicionalmente heteroscedásticos y poseen una estructura GARCH, encontrando las ventajas de superar las dificultades de estimación y restricciones en los parámetros, además de utilizar una estructura dinámica común para  $\mathbf{H}_t$ . Como resultando se obtienen modelos menos parametrizados y una reducción de la dimensionalidad cuando el número de factores es menor respecto al vector original de observaciones.

Diebold y Nerlove [1989] consideran el modelo de un factor (latente) ARCH

$$\boldsymbol{\epsilon}_t = \boldsymbol{\lambda} f_t + \mathbf{e}_t,$$

donde  $\boldsymbol{\lambda}$  es el vector de parámetros que refleja la sensibilidad del proceso para el factor  $f_t$  y  $\mathbf{e}_t$  el vector de errores específicos no correlacionados. Remarcando en este caso que los procesos  $f_t$  y  $\mathbf{e}_t$  son independientes.

De tal manera  $f_t$  resume el fenómeno heteroscedástico como

$$f_t | f_{t-1} \sim N(0, \sigma_t^2), \quad \text{con} \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i f_{t-i}^2,$$

y  $\mathbf{e}_t$  es el proceso de ruido cuyos momentos son

$$\mathbb{E}[\mathbf{e}_t | \mathbf{e}_{t-1}] = 0 \quad \text{y} \quad \text{Var}(\mathbf{e}_t | \mathbf{e}_{t-1}) = \mathbf{\Omega}.$$

Por tanto la matriz de varianza condicional del proceso  $\boldsymbol{\epsilon}_t$  es

$$\mathbf{H}_t = \text{Var}_{t-1}(\boldsymbol{\epsilon}_t) = \sigma_t^2 \boldsymbol{\lambda} \boldsymbol{\lambda}' + \mathbf{\Omega}.$$

Observemos que esta formulación nos recuerda a los modelos de factores en análisis multivariado, donde  $\sigma_t^2 \boldsymbol{\lambda} \boldsymbol{\lambda}'$  puede pensarse como la varianza explicada (*comunalidad*) por el factor  $f_t$  y  $\mathbf{\Omega}$  representa la varianza específica.

Si denotamos  $\boldsymbol{\epsilon}_t$  al proceso de error, el modelo se puede generalizar considerando  $K$  factores subyacentes de la forma

$$\boldsymbol{\epsilon}_t = \mathbf{\Lambda} \mathbf{F}_t + \mathbf{e}_t,$$

o escrito explícitamente

$$\begin{pmatrix} \epsilon_{1,t} \\ \vdots \\ \epsilon_{N,t} \end{pmatrix} = \begin{pmatrix} \lambda_{1,1} & \cdots & \lambda_{1,K} \\ \vdots & \ddots & \vdots \\ \lambda_{N,1} & \cdots & \lambda_{N,K} \end{pmatrix} \begin{pmatrix} f_{1,t} \\ \vdots \\ f_{K,t} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ \vdots \\ e_{N,t} \end{pmatrix},$$

donde  $\mathbf{F}_t$  y  $\mathbf{e}_t$  son dos procesos Gaussianos independientes

$$\mathbf{F}_t | \mathbf{F}_{t-1} \sim \text{N}(0, \boldsymbol{\Sigma}_t) \quad \text{y} \quad \mathbf{e}_t | \mathbf{e}_{t-1} \sim \text{N}(0, \mathbf{\Omega}).$$

En términos económicos, los factores comunes representan influencias generales que tienden a afectar a las series, en tanto  $\mathbf{\Lambda}$  refleja el impacto de los mismos. Los factores *únicos*  $e_{i,t}$  son la componentes no correlacionadas en términos de ruido y juegan el papel de riesgo no sistemático.

Así, el proceso  $\boldsymbol{\epsilon}_t$  es tal que

$$\boldsymbol{\epsilon}_t | \mathbf{F}_{t-1}, \mathbf{e}_{t-1} \sim \text{N}(0, \mathbf{\Lambda} \boldsymbol{\Sigma}_t \mathbf{\Lambda}' + \mathbf{\Omega}),$$

donde  $\boldsymbol{\Sigma}_t$  contiene a las varianzas condicionales de los factores  $f_{k,t}$ ,  $k = 1, \dots, K$ , cada uno representado por un ARCH( $r$ )

$$\sigma_{k,t}^2 = \alpha_0 + \sum_{i=1}^r \alpha_i f_{k,t-i}^2.$$

Engle, Ng y Rothschild [1990] piensan que los movimientos de los rendimientos son dirigidos por un pequeño número de variables comunes y definen una estructura para la matriz de varianza condicional  $\mathbf{H}_t$ , suponiendo que se compone por una parte constante y otra variante en el tiempo generada por  $K$  factores subyacentes  $\mathbf{f}_{k,t}$ , con  $k = 1, \dots, K$  y  $K < N$ , sin la condición de ser necesariamente no correlacionados. De esta manera proponen

$$\mathbf{H}_t = \mathbf{\Omega} + \sum_{k=1}^K \boldsymbol{\lambda}_k \boldsymbol{\lambda}_k' \mathbf{h}_{k,t}, \quad (4.9)$$

donde  $\mathbf{\Omega}$  es una matriz no negativa definida,  $\boldsymbol{\lambda}_k$  son vectores de pesos linealmente independientes y  $\mathbf{h}_{k,t}$  cuenta con una estructura GARCH(1,1) para los factores  $\mathbf{f}_{k,t} = \boldsymbol{\gamma}_k' \boldsymbol{\epsilon}_t$  de la siguiente manera

$$\mathbf{h}_{k,t} = \omega_k + \theta_k^2 \mathbf{f}_{k,t-1}^2 + \phi_k^2 \mathbf{h}_{k,t-1},$$

con  $\omega_k$ ,  $\theta_k$ ,  $\phi_k$  valores escalares y  $\boldsymbol{\gamma}_k$  un vector de ponderaciones (regularmente usado en portafolios de inversión).

Su objetivo es que el número de factores  $K$  sea mucho menor que  $N$ , lo que hará al modelo factible incluso para un gran número de instrumentos. Esta idea es usada por Bollerslev y Engle [1993] para modelar la persistencia común en las varianzas condicionales.

#### 4.2.1. F-GARCH

El modelo  $K$ -Factores GARCH especifica que la parte variante en el tiempo de la matriz de varianza heteroscedástica es una función de  $K$  combinaciones lineales de variables aleatorias, donde  $K$  no es mayor al número de variables, reduciendo el rango de la matriz de variante en el tiempo y la dimensión del espacio de parámetros. Su aportación se encuentra en la varianza transmitida por los factores a los conglomerados de volatilidad a través de diferentes variables.

A manera de introducir el GARCH de factores de primer orden o F-GARCH(1,1,K), Lin [1992] lo describe como un caso particular del BEKK(1,1,K) donde las matrices  $\mathbf{A}_k^*$  y  $\mathbf{B}_k^*$  tienen rango uno, con los mismos eigenvectores izquierdos y derechos,  $\boldsymbol{\lambda}_k$  y  $\boldsymbol{\omega}_k$ . Formalmente

$$\mathbf{A}_k^* = \alpha_k \boldsymbol{\omega}_k \boldsymbol{\lambda}_k', \quad \mathbf{B}_k^* = \beta_k \boldsymbol{\omega}_k \boldsymbol{\lambda}_k',$$

donde  $\alpha_k$  y  $\beta_k$  son escalares,  $\boldsymbol{\omega}_k$  y  $\boldsymbol{\lambda}_k$  son vectores que satisfacen para cada  $k = 1, \dots, K$

$$\omega'_k \lambda_i = \begin{cases} 0 & \text{para } k \neq i \\ 1 & \text{para } k = i \end{cases} \quad \text{con} \quad \sum_{n=1}^N \omega_{kn} = 1,$$

siendo ésta última una restricción de identificación.

Definiendo  $C^{*'} C^* = \Omega$  en la ecuación BEKK 4.5, se tiene

$$\mathbf{H}_t = \Omega + \sum_{k=1}^K \alpha_k^2 \lambda_k \omega'_k \epsilon_{t-1} \epsilon'_{t-1} \omega_k \lambda'_k + \sum_{k=1}^K \beta_k^2 \lambda_k \omega'_k \mathbf{H}_{t-1} \omega_k \lambda'_k. \quad (4.10)$$

Así, el modelo de  $K$  factores GARCH implica que la parte variante en el tiempo tiene rango reducido  $K$ , pero la matriz de varianza condicional  $\mathbf{H}_t$  sigue siendo de rango completo, debido a que la parte invariante al tiempo  $\Omega$  es positiva definida. Se dice que  $\lambda_k$  es la carga del  $k$ -ésimo factor denotado por  $\mathbf{f}_{k,t} = \omega'_k \epsilon_t$ .

Expresando 4.10 en forma compacta tenemos

$$\mathbf{H}_t = \Omega + \sum_{k=1}^K \lambda_k \lambda'_k (\alpha_k^2 \omega'_k \epsilon_{t-1} \epsilon'_{t-1} \omega_k + \beta_k^2 \omega'_k \mathbf{H}_{t-1} \omega_k). \quad (4.11)$$

Una interpretación es que los factores exhiben varianzas cambiantes en el tiempo y errores específicos homocedásticos, llevando a la estructura de la ecuación 4.11. En ella el factor común  $k$  es medido por una combinación lineal de los errores, dada por  $\omega'_k \epsilon_t$ , con  $\omega_k$  un vector de ponderaciones del factor y la  $k$ -ésima carga del factor  $\lambda_k$  representa la magnitud del error en respuesta a cambios del factor.

Cuando  $K = 1$  obtenemos el modelo de un factor, donde se define la varianza condicional del factor  $\mathbf{f}_t = \omega' \epsilon_t$  como un proceso GARCH(1,1) de la forma

$$\mathbf{h}_t^* = \omega^* + \alpha^2 \mathbf{f}_{t-1}^2 + \beta^2 \mathbf{h}_{t-1}^*,$$

con las siguientes condiciones  $\omega^* = \omega \Omega \omega'$  y  $\mathbf{h}_t^* = \omega \mathbf{H}_t \omega'$ .

Usando 4.11, la varianza condicional de  $\epsilon_t$  depende de la varianza condicional del factor  $\mathbf{f}_t$ , pues reescribiendo

$$\begin{aligned}
\mathbf{H}_t &= \mathbf{\Omega} + \boldsymbol{\lambda}\boldsymbol{\lambda}'(\alpha^2\boldsymbol{\omega}'\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}\boldsymbol{\omega} + \beta^2\boldsymbol{\omega}'\mathbf{H}_{t-1}\boldsymbol{\omega}) \\
&= \mathbf{\Omega} - \boldsymbol{\lambda}\boldsymbol{\lambda}'(\boldsymbol{\omega}\boldsymbol{\Omega}\boldsymbol{\omega}') + \boldsymbol{\lambda}\boldsymbol{\lambda}'(\boldsymbol{\omega}\boldsymbol{\Omega}\boldsymbol{\omega}') + \boldsymbol{\lambda}\boldsymbol{\lambda}'(\alpha^2\boldsymbol{\omega}'\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}\boldsymbol{\omega} + \beta^2\boldsymbol{\omega}'\mathbf{H}_{t-1}\boldsymbol{\omega}) \\
&= \mathbf{\Omega} - \boldsymbol{\lambda}\boldsymbol{\lambda}'(\boldsymbol{\omega}\boldsymbol{\Omega}\boldsymbol{\omega}') + \boldsymbol{\lambda}\boldsymbol{\lambda}'(\boldsymbol{\omega}\boldsymbol{\Omega}\boldsymbol{\omega}' + \alpha^2\boldsymbol{\omega}'\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}\boldsymbol{\omega} + \beta^2\boldsymbol{\omega}'\mathbf{H}_{t-1}\boldsymbol{\omega}) \\
&= \mathbf{\Omega} - \boldsymbol{\lambda}\boldsymbol{\lambda}'\boldsymbol{\omega}^* + \boldsymbol{\lambda}\boldsymbol{\lambda}'(\boldsymbol{\omega}^* + \alpha^2\mathbf{f}_{t-1}^2 + \beta^2\mathbf{h}_{t-1}^*) \\
&= \mathbf{\Omega}^* + \boldsymbol{\lambda}\boldsymbol{\lambda}'\mathbf{h}_t^*,
\end{aligned}$$

donde  $\mathbf{\Omega}^* = \mathbf{\Omega} - \boldsymbol{\lambda}\boldsymbol{\lambda}'\boldsymbol{\omega}^*$ . Ya que la especificación univariada GARCH para el factor puede reemplazarse por alguna otra, por ejemplo IGARCH o EGARCH, el modelo es bastante flexible.

Usando la representación habitual  $\mathbf{r}_t - \boldsymbol{\mu}_t = \boldsymbol{\epsilon}_t = \boldsymbol{\lambda}\mathbf{f}_t + \mathbf{e}_t$ , bajo los supuestos:

- El factor común  $\mathbf{f}_t$  y el error específico  $\mathbf{e}_t$  son no correlacionados, cada uno con varianza condicional

$$\mathbf{h}_t^* = \text{Var}_{t-1}(\mathbf{f}_t) = \boldsymbol{\omega}^* + \alpha^2\mathbf{f}_{t-1}^2 + \beta^2\mathbf{h}_{t-1}^* \quad \text{y} \quad \text{Var}_{t-1}(\mathbf{e}_t) = \mathbf{\Omega}^*,$$

al igual que como para los GARCH(1,1) la estacionariedad ocurre si  $(\alpha^2 + \beta^2) < 1$ , se sigue que la matriz de varianza condicional

$$\mathbf{H}_t = \text{Var}_{t-1}(\boldsymbol{\epsilon}_t) = \mathbf{\Omega}^* + \boldsymbol{\lambda}\boldsymbol{\lambda}'\mathbf{h}_t^* \quad (4.12)$$

tiene una estructura muy parecida a la de la matriz de varianza adoptada por Engle, Ng y Rothschild [1989].

Extendiendo la representación de  $\mathbf{H}_t$  para dos factores, el modelo F-GARCH(1,1,2) es

$$\mathbf{H}_t = \mathbf{\Omega} + \lambda_1\boldsymbol{\lambda}'_1(\alpha_1^2\boldsymbol{\omega}'_1\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}\boldsymbol{\omega}_1 + \beta_1^2\boldsymbol{\omega}'_1\mathbf{H}_{t-1}\boldsymbol{\omega}_1) + \lambda_2\boldsymbol{\lambda}'_2(\alpha_2^2\boldsymbol{\omega}'_2\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}\boldsymbol{\omega}_2 + \beta_2^2\boldsymbol{\omega}'_2\mathbf{H}_{t-1}\boldsymbol{\omega}_2).$$

En estos modelos  $\mathbf{H}_t$  obedece la dinámica determinada por el elemento común  $\mathbf{H}_{t-1}$  y los factores son generalmente correlacionados. Si así no lo fueran, genuinamente representarían diferentes componentes comunes manejando a la serie. De esta consideración surgen modelos de factores no correlacionados, donde la serie  $\boldsymbol{\epsilon}_t$  se supone relacionada a una serie  $\mathbf{f}_t$  de factores no observados  $\mathbf{f}_t$ , y a través de la transformación  $\mathbf{W}\mathbf{f}_t$  no correlacionados, tratándose  $\mathbf{W}$  de una matriz no singular.

El uso de esta noción reduce potencialmente el número de factores respecto a los enfoques anteriores, los cuales nuevamente seguirán procesos GARCH. La diferencia entre los modelos radicará en la especificación de la transformación  $\mathbf{W}$  y de manera importante, en como el número de factores heteroscedásticos será menor al número de instrumentos.

### 4.2.2. O-GARCH

En el modelo GARCH Ortogonal (O-GARCH) se asume que los datos observados son generados por una transformación ortogonal de  $N$  (o un número menor) procesos GARCH univariados. Donde la transformación lineal es la matriz ortogonal de eigenvalores de la matriz de varianza no condicional de los errores estandarizados.

El modelo O-GARCH(1,1,m) se define como

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t = \mathbf{V}^{1/2} \mathbf{u}_t, \quad \mathbf{u}_t = \mathbf{W} \mathbf{f}_t, \quad (4.13)$$

donde  $\mathbf{V} = \text{diag}(v_1, \dots, v_N)$  con  $v_i$  la varianza de  $\epsilon_{i,t}$  y  $\mathbf{W}$  una matriz dada por

$$\mathbf{W} = \mathbf{P}_m \text{diag}(\lambda_1^{1/2}, \dots, \lambda_m^{1/2})$$

con  $\lambda_1 \geq \dots \geq \lambda_m$  los  $m$  eigenvalores más grandes de la matriz de correlación muestral y  $\mathbf{P}_m$  la matriz de eigenvectores asociados (mutuamente ortogonales).

Ahora, el vector  $\mathbf{f}_t = (f_{1,t}, \dots, f_{m,t})'$  representa un proceso aleatorio con los dos primeros momentos condicionales

$$\mathbb{E}_{t-1}[\mathbf{f}_t] = \mathbf{0} \quad \text{y} \quad \boldsymbol{\Sigma}_t = \text{Var}_{t-1}(\mathbf{f}_t) = \text{diag}(\sigma_{f_{1,t}}^2, \dots, \sigma_{f_{m,t}}^2),$$

donde cada  $f_{i,t}$  sigue GARCH(1,1) i.e.

$$\sigma_{f_{i,t}}^2 = (1 - \alpha_i - \beta_i) + \alpha_i f_{i,t-1}^2 + \beta_i \sigma_{f_{i,t-1}}^2, \quad i = 1, \dots, m.$$

Entonces la matriz de varianza condicional del proceso  $\boldsymbol{\epsilon}_t$  es

$$\mathbf{H}_t = \text{Var}(\boldsymbol{\epsilon}_t) = \mathbf{V}^{1/2} \mathbf{V}_t \mathbf{V}^{1/2}, \quad (4.14)$$

donde  $\mathbf{V}_t = \text{Var}_{t-1}(\mathbf{u}_t) = \mathbf{W} \boldsymbol{\Sigma}_t \mathbf{W}'$  es la varianza condicional de los errores estandarizados. Los parámetros del modelo son  $\mathbf{V}$  y  $\mathbf{W}$ , en tanto los parámetros de los factores GARCH las  $\alpha_i$ 's y  $\beta_i$ 's, siendo en total  $N(N+5)/2$  si  $m = N$ . En la práctica  $\mathbf{V}$  y  $\mathbf{W}$  se reemplazan por sus análogas muestrales y para elegir  $m$  se usa el análisis de componentes principales aplicado a los residuales muestrales estandarizados  $\hat{\mathbf{u}}_t$ .

Alexander [2001] considera la transformación en componentes principales, donde los procesos GARCH son precisamente las componentes, teniendo la ventajas de proporcionar una reducción en la dimensión y conservar la variabilidad de las series originales de rendimientos. Esto es, dado un vector que contenga varias series de rendimientos, si

aplicando las componentes principales, se tiene que una proporción aceptable de la varianza es alcanzada, por ejemplo, con las dos primeras componentes, bastará realizar dos ajustes GARCH para modelar la matriz de varianza condicional.

De 4.14,  $\Sigma_t$  es una matriz diagonal que representa las varianzas condicionales de las componentes principales y  $\mathbf{V}_t$  es la matriz dinámica en el tiempo del sistema original aproximada por  $\mathbf{V}_t = \mathbf{W}\Sigma_t\mathbf{W}$ , donde  $\mathbf{W}$  contiene a las ponderaciones o cargas de las componentes.

En particular es un modelo GARCH Ortogonal, donde las varianzas condicionales para cada componente de la matriz diagonal  $\Sigma_t$  se estiman con modelos GARCH univariados e.g. considerando un GARCH(1,1) el modelo es

$$\sigma_{i,t}^2 = \omega + \alpha\xi_{i,t-1}^2 + \beta\sigma_{i,t-1}^2,$$

con  $\xi_{i,t}$  la  $i$ -ésima componente principal y  $\sigma_{i,t}^2$  su varianza condicional. El coeficiente  $\alpha$  tiene la explicación de *reacción* a movimientos del mercado y  $\beta$  mide la *persistencia* en la volatilidad.

La representación en 4.14, proporciona una matriz positiva definida en cada  $t$ , incluso cuando el número de componentes sea menor al número de variables del sistema. Su precisión dependerá de emplear las componentes suficientes para explicar la mayor proporción de varianza en el sistema.

### 4.2.3. GO-GARCH

El autor van der Weide [2002] extiende la caracterización anterior en su modelo GARCH Ortogonal Generalizado GO-GARCH. En este caso los factores no correlacionados tienen varianzas condicionales unitarias, esto es,  $\mathbb{E}[\mathbf{f}_t\mathbf{f}_t'] = \mathbf{I}_m$  y construye el mapeo lineal  $\mathbf{W}$ , usando la descomposición de la varianza no condicional  $\mathbb{E}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'] = \mathbf{W}\mathbf{W}'$ .

Definiendo cada componente como en el O-GARCH, con  $m = N$  y  $\mathbf{W}$  una matriz no singular de parámetros, la matriz de correlación condicional implícita de  $\boldsymbol{\epsilon}_t$  puede ser expresada en el GO-GARCH como

$$\mathbf{R}_t = \mathbf{J}_t^{-1}\mathbf{V}_t\mathbf{J}_t^{-1}, \quad \text{donde } \mathbf{J}_t = (\mathbf{V}_t \odot \mathbf{I}_m)^{1/2} \quad \text{y} \quad \mathbf{V}_t = \mathbf{W}\Sigma_t\mathbf{W}', \quad (4.15)$$

con  $\odot$  el producto de Hadamard<sup>13</sup>, bajo la parametrización  $\mathbf{W} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{U}$ , cuya matriz  $\mathbf{U}$  es ortogonal y las matrices  $\mathbf{P}$ ,  $\boldsymbol{\Lambda}$  contienen a los eigenvectores y eigenvalores de  $\mathbb{E}[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t']$ .

<sup>13</sup>Ver Apéndice B

De la misma manera supone que los errores son manejados por algunos factores no correlacionados, ligados a  $\boldsymbol{\epsilon}_t$  a través de un mapeo lineal  $\mathbf{f}_t = \mathbf{W}^{-1}\boldsymbol{\epsilon}_t$ , y para identificarlos necesita solamente la invertibilidad de la matriz asociada  $\mathbf{W}$ .

Las matrices  $\mathbf{P}$  y  $\mathbf{A}$  se estiman de información no condicional, sustituyéndolas en un primer paso por sus matrices muestrales. Los factores  $\mathbf{f}_t$  son condicionalmente heteroscedásticos y siguen una dinámica del tipo GARCH. En particular cuando  $m = N$ , la elección  $\mathbf{U} = \mathbf{I}_N$  corresponde al modelo O-GARCH, pero se distingue porque  $\mathbf{W}$  no requiere ser ortogonal, solamente tiene que ser una matriz cuadrada e invertible. Los parámetros faltantes son estimados conjuntamente con los coeficientes GARCH en un segundo paso.

Ahora, la matriz diagonal de varianza condicional de  $\mathbf{f}_t$  se puede definir como

$$\boldsymbol{\Sigma}_t = (\mathbf{I}_N - \mathbf{A} - \mathbf{B}) + \mathbf{A} \odot \mathbf{f}_{t-1}\mathbf{f}'_{t-1} + \mathbf{B}\boldsymbol{\Sigma}_{t-1}, \quad (4.16)$$

donde  $\mathbf{A}$  y  $\mathbf{B}$  son matrices diagonales de parámetros. La forma del término constante impone la restricción mencionada  $\mathbb{E}[\mathbf{f}_t\mathbf{f}'_t] = \mathbf{I}_m$  y la estacionariedad de los modelos con factores no correlacionados se asegura si los elementos de la diagonal de  $\mathbf{A} + \mathbf{B}$  son menores que uno. Además la matriz de varianza condicional de  $\boldsymbol{\epsilon}_t$  se expresa como

$$\mathbf{H}_t = \mathbf{W}\boldsymbol{\Sigma}_t\mathbf{W}' = \sum_{k=1}^N \boldsymbol{\omega}_{(k)}\boldsymbol{\omega}'_{(k)}\sigma_{k,t}^2, \quad (4.17)$$

siendo  $\boldsymbol{\omega}_{(k)}$  las columnas de la matriz  $\mathbf{W}$  y  $\sigma_{k,t}^2$  los elementos de la diagonal de  $\boldsymbol{\Sigma}_t$ .

Notemos que las ecuaciones 4.9 y 4.17 son parecidas, sin embargo, la diferencia se encuentra en el hecho de que para la segunda los factores son no correlacionados. Entonces, en el modelo GO-GARCH no es posible tener menos factores que instrumentos.

#### 4.2.4. FF-GARCH

El modelo GARCH multivariado de factores completos GARCH (Full Factor GARCH o FF-GARCH) de Vrontos, Dellaportas y Politis [2003] se define a través del proceso

$$\mathbf{r}_t - \boldsymbol{\mu}_t = \boldsymbol{\epsilon}_t = \mathbf{W}\mathbf{f}_t, \quad (4.18)$$

con la matriz de covarianza condicional

$$\mathbf{H}_t = \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}' = \mathbf{W}\boldsymbol{\Sigma}_t^{1/2}\boldsymbol{\Sigma}_t^{1/2}\mathbf{W}'. \quad (4.19)$$

En este caso  $\mathbf{W}$  es una matriz (invertible) triangular inferior de parámetros con la diagonal conformada de unos y  $\mathbf{\Sigma}_t$  es la matriz diagonal dada por

$$\mathbf{\Sigma}_t = \text{diag}(\sigma_{1,t}^2, \dots, \sigma_{N,t}^2),$$

con  $\sigma_{i,t}^2$  la varianza condicional del  $i$ -ésimo factor  $f_{i,t}$  pudiéndose definir por separado como algún proceso univariado, por lo regular GARCH(1,1).

Usa el resultado que una matriz positiva definida puede descomponerse como el producto de una matriz triangular por su transpuesta (*descomposición de Cholesky*), de hecho la descomposición es única si los elementos de la diagonal se restringen a ser positivos. Por este motivo en el modelo se considera  $\mathbf{W}$  triangular, con elementos  $w_{ij} = 0$  para  $j > 0$  y  $w_{ii} > 0$  para  $i = 1, \dots, N$ . Con el objetivo reducir el número de parámetros, una imposición natural es hacer  $w_{ii} = 1$  para  $i = 1, \dots, N$ .

Por construcción  $\mathbf{H}_t$  es siempre positiva definida, y aunque su estructura depende del orden de la serie  $\boldsymbol{\epsilon}_t$ , puede aplicarse a series de gran dimensión, obteniendo un número de parámetros relativamente pequeño.

#### 4.2.5. GOF-GARCH

Lanne y Saikkonen [2007] en el modelo de factores ortogonal generalizado GARCH (GOF-GARCH) siguen el tratamiento propuesto por van der Weide [2002], considerando los factores no correlacionados con varianzas condicionales unitarias i.e.  $\mathbb{E}[\mathbf{f}_t \mathbf{f}_t'] = \mathbf{I}_m$ .

Recordando que el error se representa a través de la transformación  $\boldsymbol{\epsilon}_t = \mathbf{W} \mathbf{f}_t$ , se define la matriz de varianza (no condicional) de  $\boldsymbol{\epsilon}_t$  como

$$\begin{aligned} \mathbf{\Sigma} &= \mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'] = \mathbb{E}[\mathbf{W} \mathbf{f}_t (\mathbf{W} \mathbf{f}_t)'] \\ &= \mathbb{E}[\mathbf{W} \mathbf{f}_t \mathbf{f}_t' \mathbf{W}'] = \mathbf{W} \mathbb{E}[\mathbf{f}_t \mathbf{f}_t'] \mathbf{W}' \\ &= \mathbf{W} \mathbf{W}' = \mathbf{C} \mathbf{C}'. \end{aligned}$$

Ahora, de la descomposición espectral  $\mathbf{\Sigma} = \mathbf{P} \mathbf{\Lambda}^{1/2} \mathbf{P}'$ , sugieren expresar el mapeo  $\mathbf{W}$  a través la descomposición  $\mathbf{W} = \mathbf{C} \mathbf{U}$ , con  $\mathbf{C}$  una matriz simétrica positiva definida y  $\mathbf{U}$  una matriz ortogonal.

Entonces  $\mathbf{C}$  es la raíz cuadrada de la matriz de varianza, de tal forma

$$\mathbf{C} = \mathbf{\Sigma}^{1/2} = \mathbf{P} \mathbf{\Lambda}^{1/2} \mathbf{P}',$$

de donde  $\mathbf{U} = \mathbf{\Sigma}^{-1/2} \mathbf{W}$ .

Las columnas de  $\mathbf{P}$  son los eigenvectores de  $\mathbf{\Sigma}$  y la matriz diagonal  $\mathbf{\Lambda}$  contiene a sus eigenvalores, necesitando para obtenerlas solamente información no condicional, pero para calcular  $\mathbf{U}$  sí se requiere el uso de información condicional.

Es importante notar que en este modelo algunos factores pueden ser condicionalmente homocedásticos y para mantener la parsimonia se permite al modelo no sólo incluir componentes sistemáticos de riesgo, sino también específicos.

En otras palabras, existiendo  $K \leq N$  factores heteroscedásticos, el resto  $N - K$  serán factores homocedásticos. Sin pérdida de generalidad supónganse los primeros  $K$  elementos de  $\mathbf{f}_t$  heteroscedásticos, entonces los últimos  $N - K$  elementos diagonales de  $\mathbf{A}$  y  $\mathbf{B}$  en 4.16 serán cero, resultando en 4.17 de la forma

$$\begin{aligned} \mathbf{H}_t &= \sum_{k=1}^K \boldsymbol{\omega}_{(k)} \boldsymbol{\omega}'_{(k)} \sigma_{k,t}^2 + \sum_{k=K+1}^N \boldsymbol{\omega}_{(k)} \boldsymbol{\omega}'_{(k)} \\ &= \sum_{k=1}^K \boldsymbol{\omega}_{(k)} \boldsymbol{\omega}'_{(k)} \sigma_{k,t}^2 + \mathbf{\Omega}^*. \end{aligned} \tag{4.20}$$

A pesar de ser similar a 4.9, tiene dos diferencias relevantes:

- Los factores son no correlacionados.
- El papel de  $\mathbf{\Omega}^*$  es distinto, aquí tiene rango reducido en relación con el número de factores heteroscedásticos, quienes también tienen que ver con la matriz de varianza no condicional de  $\boldsymbol{\epsilon}_t$ .

Por lo tanto el GOF-GARCH puede verse como la combinación de los modelos de factores (con un número reducido de factores heteroscedásticos) y los modelos ortogonales, donde su estimación relativamente fácil radica en la ortogonalidad de los factores, teniendo las respectivas ventajas de cada enfoque.

## Estacionariedad

Nótese que los modelos ortogonales son casos particulares de los modelos F-GARCH, a su vez representados como BEKK bajo ciertas condiciones, pudiéndose concluir que sus propiedades pueden seguirse de ése modelo. Entonces, en particular los modelos O-GARCH y GO-GARCH son estacionarios si cada uno de los  $m$  procesos GARCH univariados son estacionarios.

### 4.3. Modelos de Correlación Condicional

En estos modelos, además de representar a las varianzas condicionales, se incluye una dinámica en las correlaciones, llevando a cierta flexibilidad en la modelación de las varianzas (no necesariamente la misma para cada componente), pues por ejemplo se puede usar un GARCH(1,1) estándar para cierto componente y un EGARCH para otro. La finalidad será la especificación de la matriz de correlación condicional, de tal manera que sea positiva definida.

Usando la descomposición de la matriz de varianza condicional en desviaciones estándar y correlaciones condicionales,  $\mathbf{H}_t$  puede escribirse como

$$\begin{aligned}\mathbf{H}_t &= \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t, \quad \text{donde} \\ \mathbf{D}_t &= \text{diag}(h_{11,t}^{1/2}, \dots, h_{NN,t}^{1/2}), \\ \mathbf{R}_t &= (\rho_{ij,t}), \quad \text{con } \rho_{ii,t} = 1, \quad i = 1, \dots, N.\end{aligned}\tag{4.21}$$

La matriz de correlación condicional es  $\mathbf{R}_t$  y  $h_{ii,t}$  se define como un GARCH univariado. Las covarianzas condicionales se obtienen multiplicando la correlación condicional por las desviaciones estándar condicionales, esto es

$$h_{ij,t} = \rho_{ij,t} \sqrt{h_{ii,t} h_{jj,t}} \quad \forall i \neq j,\tag{4.22}$$

donde la positividad de  $\mathbf{H}_t$  se seguirá de la positividad de  $\mathbf{R}_t$  y de cada varianza condicional.

Para ciertos casos, las condiciones de positividad pueden imponerse fácilmente y la estimación se realiza de manera eficiente. Por lo general se condiciona a cada varianza condicional a ser función de sus propios rezagos y de sus errores cuadrados pasados, manteniendo en equilibrio el número de parámetros con respecto a los modelos BEKK y de Factores, compensando los coeficientes ahorrados en la modelación de las varianzas, con parámetros adicionales para la dinámica de las correlaciones.

#### 4.3.1. Correlación Condicional Constante

Conocido como CCC, se trata del primer modelo multivariado de correlación y se incluye en los otros modelos de correlación condicional. En él, la correlación condicional es invariante en el tiempo, así que la matriz de varianza condicional se define como

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t.\tag{4.23}$$

Entonces por la definición anterior,  $\mathbf{R}_t$  permanece constante, de la siguiente manera  $\mathbf{R}_t = \mathbf{R} = (\rho_{ij})$ , con  $\rho_{ii} = 1 \quad \forall i = 1, \dots, N$ , y es una matriz simétrica positiva definida.

Las covarianzas condicionales serán proporcionales al producto de la correlación con sus correspondientes desviaciones estándar como en 4.22, ayudando a reducir el número de parámetros desconocidos y simplificar la estimación. Entonces se tiene que

$$\mathbf{D}_t = \text{diag}(h_{11,t}^{1/2}, \dots, h_{NN,t}^{1/2}),$$

donde  $h_{ii,t}$  puede definirse como algún modelo GARCH univariado. En el modelo original CCC, Bollerslev [1990] especifica un GARCH(1,1) para cada varianza condicional en  $\mathbf{D}_t$

$$h_{ii,t} = \alpha_{i0} + \alpha_i \epsilon_{i,t-1}^2 + \beta_i h_{ii,t-1}, \quad i = 1, \dots, N. \quad (4.24)$$

Nótese que los procesos no necesariamente deben tener el mismo número de rezagos y pueden seguir diferentes especificaciones.

En particular, el caso bivariado puede representarse por las ecuaciones

$$\begin{aligned} h_{11,t} &= \alpha_{10} + \alpha_1 \epsilon_{1,t-1}^2 + \beta_1 h_{11,t-1}, \\ h_{22,t} &= \alpha_{20} + \alpha_2 \epsilon_{2,t-1}^2 + \beta_2 h_{22,t-1}. \end{aligned}$$

Hemos visto que los modelos VEC y BEKK permiten a cada elemento de  $\mathbf{H}_t$  relacionarse con cada elemento de  $\mathbf{H}_{t-i}$  y  $\epsilon_t \epsilon_t'$ , ahora en el CCC se restringe la dinámica de la varianzas condicionales para que sean funciones únicamente de sus propios rezagos y realizaciones pasadas. El CCC contiene  $N(N+5)/2$  parámetros y  $\mathbf{H}_t$  es positiva definida si y sólo si todas las  $N$  varianzas condicionales son positivas, de donde también  $\mathbf{R}$  será positiva definida. Las varianzas no condicionales se obtienen como en el caso univariado, sin embargo, las covarianzas no condicionales son difíciles de calcular debido a la no linealidad de  $\mathbf{H}_t$ .

Es cierto que proporciona una parametrización parsimoniosa de las varianzas otorgando estimadores menos ruidosos, pero empíricamente el supuesto de correlaciones constantes es restrictivo. Por ello, conservando la descomposición previa, pero haciendo la matriz de correlación variante en el tiempo se tienen los modelos dinámicos.

En la referencia de Tsay [2005] encontramos otra definición del modelo CCC para el caso bivariado a través de la representación matricial

$$\begin{bmatrix} h_{11,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} \alpha_{10} \\ \alpha_{20} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{22,t-1} \end{bmatrix}.$$

Un poco distinto de la definición de Bollerslev, ya que se permite una dependencia de la volatilidad de los errores  $\epsilon_{1,t-1}$  y  $\epsilon_{2,t-1}$ . Eliminando dicha relación se observan dos casos:

- Si  $\alpha_{12} = \beta_{12} = 0$ , la volatilidad de  $\epsilon_{1,t-1}$  no dependerá de la volatilidad pasada de  $\epsilon_{2,t-1}$ .
- Si  $\alpha_{21} = \beta_{21} = 0$ , la volatilidad de  $\epsilon_{2,t-1}$  no dependerá de la volatilidad pasada de  $\epsilon_{1,t-1}$ .

Al final se emplearán los modelos CCC propuestos, primero bajo el supuesto de no dependencia en los errores cruzados pasados, al que llamaremos CCC de *un elemento* y posteriormente con el modelo completo dada la interacción entre varianzas y errores, que denominaremos de *dos elementos*.

### 4.3.2. Correlación Condicional Dinámica

Respecto a los modelos de correlación constante, difieren de la estimación numérica, puesto que la matriz debe ser invertible para toda  $t$  durante cada iteración e introducen el objetivo de imponer la positividad de la matriz de correlación condicional.

Dada la interpretación se tienen dos modelos principales.

#### DCC Engle

Engle [2002] propone una generalización del CCC proporcionando una dinámica para la matriz de correlación condicional.

De la misma manera define la matriz de varianza condicional como

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t,$$

donde  $\mathbf{D}_t$  es la matriz diagonal cuadrada cuyos elementos siguen procesos GARCH univariados y  $\mathbf{R}_t$  es la matriz de correlación condicional cambiante en el tiempo definida por

$$\mathbf{R}_t = \text{diag}(q_{11,t}^{-1/2}, \dots, q_{NN,t}^{-1/2}) \mathbf{Q}_t \text{diag}(q_{11,t}^{-1/2}, \dots, q_{NN,t}^{-1/2}), \quad (4.25)$$

donde  $\mathbf{Q}_t$  es una matriz simétrica que sigue la dinámica

$$\mathbf{Q}_t = (1 - \theta_1 - \theta_2) \mathbf{S} + \theta_1 \mathbf{u}_{t-1} \mathbf{u}'_{t-1} + \theta_2 \mathbf{Q}_{t-1}. \quad (4.26)$$

En la ecuación anterior,  $u_{i,t} = \epsilon_{i,t}/\sqrt{h_{ii,t}}$  son los errores estandarizados y  $\mathbf{S}$  puede ser una matriz simétrica positiva definida de parámetros o representar la matriz de varianza muestral de los errores estandarizados. Además  $\theta_1$  y  $\theta_2$  son parámetros escalares, el primero positivo y el segundo no negativo, tales que  $(\theta_1 + \theta_2) < 1$ . Observamos entonces que los términos de  $\mathbf{S}$  pueden calcularse o alternativamente emplear las varianzas muestrales empíricas para hacer la estimación más simple.

### DCC Tse y Tsui

Tse y Tsui [2002] introducen un modelo parecido, cuya distinción radica en la dinámica de  $\mathbf{R}_t$ . Para ellos la matriz de correlación es

$$\mathbf{R}_t = (1 - \theta_1 - \theta_2)\mathbf{R} + \theta_1\mathbf{\Psi}_{t-1} + \theta_2\mathbf{R}_{t-1}, \quad (4.27)$$

donde  $\mathbf{R} = (\rho_{ij})$  es una matriz (de correlación) simétrica positiva definida invariante en el tiempo con  $\rho_{ii} = 1$ ,  $\theta_1, \theta_2$  parámetros no negativos que cumplen la condición  $(\theta_1 + \theta_2) < 1$  y  $\mathbf{\Psi}_{t-1}$  la matriz de correlación muestral de  $\epsilon_\tau$ ,  $\tau = t - M, t - M + 1, \dots, t - 1$ , donde el elemento  $(i, j)$  esta dado por

$$\Psi_{ij,t-1} = \frac{\sum_{m=1}^M u_{i,t-m}u_{j,t-m}}{\sqrt{(\sum_{m=1}^M u_{i,t-m}^2)(\sum_{m=1}^M u_{j,t-m}^2)}}, \quad \text{con } u_{i,t} = \frac{\epsilon_{i,t}}{\sqrt{h_{ii,t}}}.$$

Una condición necesaria para garantizar la positividad de  $\mathbf{\Psi}_{t-1}$  y por ende la de  $\mathbf{R}_t$ , es que  $M > N$ . Por 4.27,  $\mathbf{R}_t$  puede pensarse como un promedio ponderado de las matrices  $\mathbf{R}$ ,  $\mathbf{\Psi}_{t-1}$  y  $\mathbf{R}_{t-1}$ , de donde se sigue que será positiva si alguno de sus tres componentes lo es también. Los modelos difieren en la manera de modelar las correlaciones condicionales y la forma de emplear la información contenida en los productos de los errores rezagados.

La formulación DCC permite una modelación aceptable de la varianza permitiendo que alguna estructura GARCH gobierne la dinámica de los componentes. Estos modelos proporcionarán las varianzas condicionales  $h_{ii,t-1}$  que se emplearán para estandarizar los errores rezagados  $\epsilon_{ii,t-1}$  y calcular su correlación muestral.

Se observa que  $\theta_1$  y  $\theta_2$  son escalares para que  $\mathbf{R}_t$  sea positiva definida y una matriz de correlación. Por lo tanto todas las correlaciones seguirán la misma dinámica, ahorrando algunos parámetros respecto a los VEC y BEKK. En ambos casos, si las varianzas condicionales siguen un GARCH(1,1) se tendrán en total  $(N + 1)(N + 4)/2$  parámetros.

## 4.4. Estimación MGARCH

Como se estableció al inicio del capítulo, la media y varianza condicionales  $\boldsymbol{\mu}_t$  y  $\mathbf{H}_t$  dependen de un vector de parámetros  $\boldsymbol{\theta}$ , cuya estimación envuelve la maximización de la función de verosimilitud y que bajo el supuesto adicional de una distribución i.i.d. para el proceso de innovación estandarizado  $\mathbf{z}_t$  puede verse como una función de cuasi verosimilitud. Denotando a esa densidad como  $g(\mathbf{z}_t|\boldsymbol{\zeta})$ , con  $\boldsymbol{\zeta}$  un vector de parámetros de ruido, el objetivo será maximizar la verosimilitud muestral  $L_T(\boldsymbol{\theta}, \boldsymbol{\zeta})$  para las  $T$  observaciones (condicional en valores iniciales  $\boldsymbol{\mu}_0$  y  $\mathbf{H}_0$  con respecto al vector completo de parámetros).

La función de log-verosimilitud es

$$L_T(\boldsymbol{\theta}, \boldsymbol{\zeta}) = \sum_{i=1}^T \log f(\mathbf{r}_t|\boldsymbol{\theta}, \boldsymbol{\zeta}, \mathbf{I}_{t-1}), \quad (4.28)$$

con  $f(\mathbf{r}_t|\boldsymbol{\theta}, \boldsymbol{\zeta}, \mathbf{I}_{t-1}) = |\mathbf{H}_t^{-1/2}|g(\mathbf{H}_t^{-1/2}(\mathbf{r}_t - \boldsymbol{\mu}_t); \boldsymbol{\zeta})$ , donde la dependencia con respecto a  $\boldsymbol{\theta}$  ocurre a través de  $\boldsymbol{\mu}_t$  y  $\mathbf{H}_t$ .

En algunos casos se asume  $\mathbf{z}_t \sim N(0, \mathbf{I}_N)$ , para hacer el vector  $\boldsymbol{\zeta}$  vacío. De tal manera

$$L_T(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^T \left( \log |\mathbf{H}_t| + (\mathbf{r}_t - \boldsymbol{\mu}_t)' \mathbf{H}_t^{-1} (\mathbf{r}_t - \boldsymbol{\mu}_t) \right) \quad (4.29)$$

es la log-verosimilitud Gaussiana que proporciona el estimador cuasi máximo verosímil consistente para  $\boldsymbol{\theta}$ , incluso si la densidad verdadera no es  $N(0, \mathbf{I}_N)$ , para ello basta especificar correctamente a  $\boldsymbol{\mu}_t$  y  $\mathbf{H}_t$ .

En el caso de series financieras, en particular, de log-rendimientos, se ha observado el aspecto poco realista de la normalidad, por ello y debido a la importancia de suponer la densidad correcta con la finalidad de evitar la inconsistencia de los estimadores, encontramos distribuciones multivariadas alternas propuestas: Student-t, Student-t sesgada, mezcla de dos densidades Gaussianas y la Distribución Hiperbólica Generalizada, logrando capturar el exceso de curtosis y sesgo condicionales.



## Capítulo 5

# Aplicación MGARCH

### 5.1. Series de Tipo de Cambio

Las divisas son las monedas u otros activos financieros utilizados para el intercambio internacional. De esta manera entendemos mercados de divisas, como los lugares en los que se compran y venden las monedas de los distintos países y donde se lleva a cabo el cambio de la moneda nacional (el peso) por la moneda de los países con los que se mantienen relaciones económicas, las cuales pueden ser exportaciones, importaciones y turismo, originándose un conjunto de oferta y demanda de moneda extranjera a por moneda nacional y siendo este tipo de transacciones las que determinan el precio o el tipo de cambio del peso frente a las monedas foráneas.

El Banco de México en nuestro país es el banco central, y tiene por objetivos: proveer a la economía del país de moneda nacional y procurar la estabilidad del poder adquisitivo de dicha moneda. Por lo tanto no es ajeno al desempeño del tipo de cambio y sus indicadores. Es cierto, la paridad cambiaria es fijada libremente por el mercado; pero también existen factores asociados a las actividades económicas y financieras que pueden interpretarse en alzas o caídas de la moneda.

En el contexto actual observamos una tendencia de debilitamiento del peso respecto a las monedas de referencia globales dólar y euro, y aunque se puede mencionar a la volatilidad del mercado como una causa de este patrón, no es la única. La reducción en el precio del petróleo, menores exportaciones del país, disminución de remesas de migrantes mexicanos, baja inversión extranjera, todas ellas reflejadas en la falta de liquidez o escasez de divisas (dólar y euro) son factores que derivan el encarecimiento de éstas últimas. Si la oferta de dólares y euros es menor a la demanda, hará que la cotización de esas monedas extranjeras se aprecie, en detrimento del peso mexicano (que se deprecia), tal y como sucede con el precio de cualquier mercancía dentro de un mercado de productos.

Tampoco olvidemos que la presencia de fenómenos especulativos, explicada en gran parte por la elevada percepción de riesgo relacionada con el desempeño de los mercados en el entorno global, pone en desventaja a los mercados llamados *emergentes*, puesto que la lectura de los financieros internacionales los considera puertos menos seguros. De allí que no sólo México sino muchas naciones en el orbe ubicados en la categoría de economías en desarrollo experimenten bajas en el precios de su moneda. Gran parte de los inversionistas prefieren refugiarse en mercados y activos más seguros, por ejemplo los mercados europeos e instrumentos sólidos, como los bonos del Tesoro de Estados Unidos, originando un alza en el euro y el dólar.

Un punto importante propuesto es que no todo debe dejarse a la deriva de los mercados y sean ellos mismos lo que se ajusten y recuperen su equilibrio. Las herramientas y tecnologías con las que contamos pueden favorecer al entedimiento de los fenómenos económicos y financieros, y como es sabido, tener mayor información resulta en una mejor toma de decisiones.

Finalmente para la aplicación de los modelos de volatilidad multivariados MGARCH, se emplearán los tipos de cambio diarios Peso/Dólar y Peso/Euro desde Octubre de 2002 hasta Septiembre de 2008, con información publicada por el Banco de México en su página de internet [www.banxico.org.mx](http://www.banxico.org.mx). Se ha elegido dicho periodo porque se pretende analizar el periodo de recuperación después del colapso de las economías mundiales ocurrido a finales de 2001, y asimismo, observar el efecto de la crisis mundial ocurrida a partir del año 2007 con sus consecuencias más que visibles en la actualidad.

Es de interés considerar las series de tipo de cambio por dos motivos:

Primero, porque el tipo de cambio forma parte de los procesos fundamentales de intercambio de bienes y servicios, íntimamente ligado con las economías externas, en el caso de nuestro país principalmente con Estados Unidos.

Segundo, se dice que una moneda fuerte es indicativo de una economía saludable, en el sentido del crecimiento y estabilidad de las actividades económicas del país. Entonces debe ser una tarea, el establecer las políticas monetarias que favorezcan ese ambiente, y aunque no es posible fijar la paridad cambiaria, sí se pueden tomar medidas preventivas ante condiciones adversas que disminuyan el valor del peso.

Nuestro análisis inicial hace una revisión de los datos y sus características estadísticas con el fin de ejemplificar los que hemos llamado *hechos estilizados*. Observando esos supuestos, se ajustaron modelos MGARCH bivariados en cuatro especificaciones (DVEC, BEKK, OGARCH y CCC) en el programa estadístico S-PLUS, empleando las funciones disponibles en el módulo `finmetrics`.<sup>14</sup>

---

<sup>14</sup>En el Apéndice C se encuentran las funciones y salidas utilizadas.

## 5.1.1. Análisis de los Datos

En la figura 5.1 se muestra las series de tiempo de los tipos de cambio considerados diariamente desde Octubre de 2002 a Septiembre de 2008.

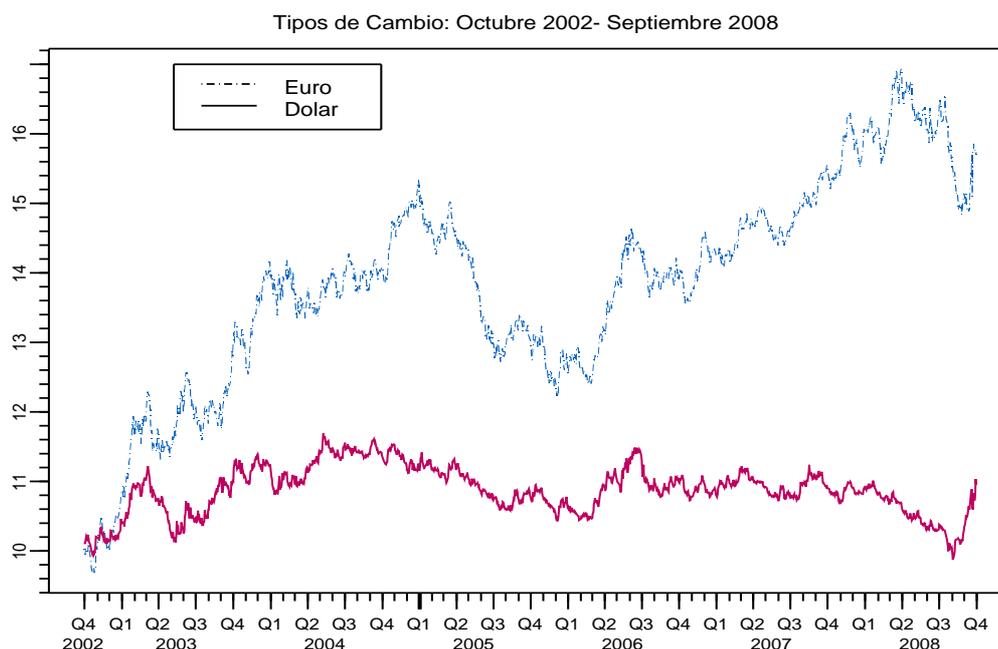


Figura 5.1: Series Originales

Claramente las series comparten una serie de movimientos, aunque los cambios en el tipo de cambio dólar se presentan en menor proporción al euro, ambas mantienen un cierto nivel. A partir de 2002 el tipo de cambio Peso/Euro incluso era menor al del dólar, fue a partir de 2003 cuando comenzó a tener alzas importantes. Observamos que desde el segundo trimestre de 2007 y hasta mediados de 2008 se da un fenómeno inverso, mientras el euro mantenía una tendencia creciente, el dólar iba a la baja. Es a finales de 2008 cuando los dos tienen un repunte atribuido a los estragos de la crisis económica en los mercados mundiales.

Recordando que es conveniente utilizar log-rendimientos, se realizó la transformación  $r_t = 100 \cdot \log(P_t/P_{t-1})$ , donde  $P_t$  es el tipo de cambio al tiempo  $t$ . Decidimos emplear porcentajes con la meta de contar con una medida explicativa de los movimientos diarios (ganancias o pérdidas).

Las características más importantes de las series de log-rendimientos en sus estadísticas

básicas son la cercanía de la media a cero y curtosis superior a 3, cuyo exceso podría ser un indicativo de no normalidad. En la Figura 5.2 graficamos las series de log-rendimientos.

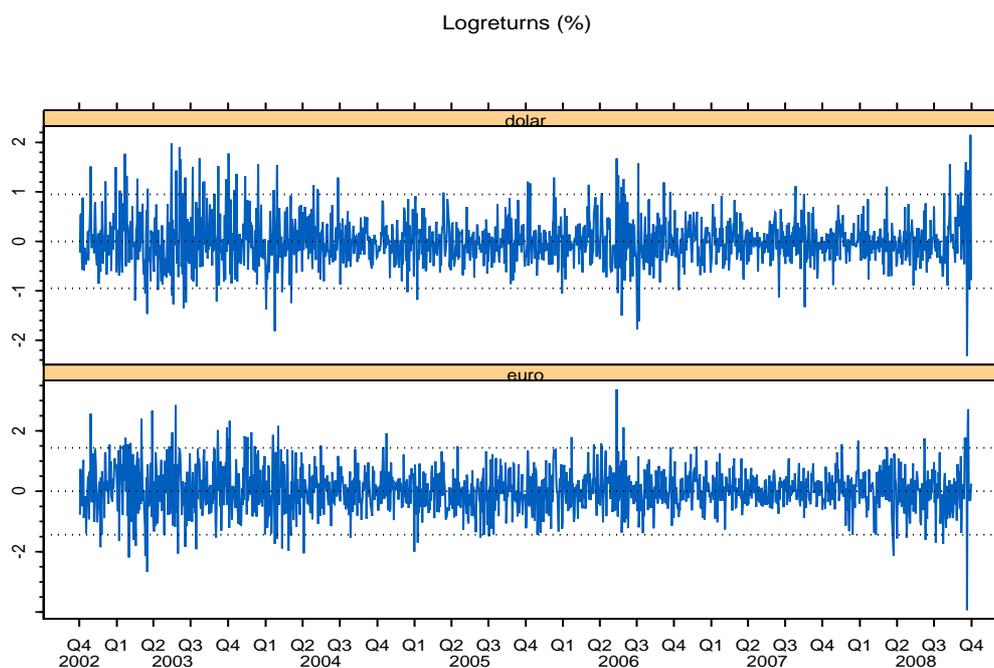


Figura 5.2: Series de log-rendimientos

Se confirma la existencia de conglomerados de volatilidad y movimientos conjuntos de ambas series, mostrando de cierta manera su relación. Luego, se aprecian cambios de mayor frecuencia y magnitud en la serie dólar, entonces en ese sentido podemos pensar que es más *volátil*. También son remarcarse las etapas donde se generan grandes *saltos* de volatilidad: entre el segundo y tercer trimestre de 2003, en el segundo trimestre de 2006 y al final del tercer trimestre de 2008.

De la evidencia empírica, generalmente se ha encontrado que las series de rendimientos son no correlacionadas, hecho que puede observarse en las gráficas de las funciones de Autocorrelación y Autocorrelación Parcial para cada serie (Figura 5.3).

Confirmando este resultado, la prueba de autocorrelación Ljung-Box proporciona los estadísticos  $Q(12)=17.4317$  para la serie dólar y  $Q(12)=5.67$  para la serie euro, de tal forma que comparando con el cuantil 0.95 de la distribución Chi-cuadrada con 12 grados de libertad  $\chi_{(12)}=21.02$ , no se rechaza en ningún caso la hipótesis de no correlación.

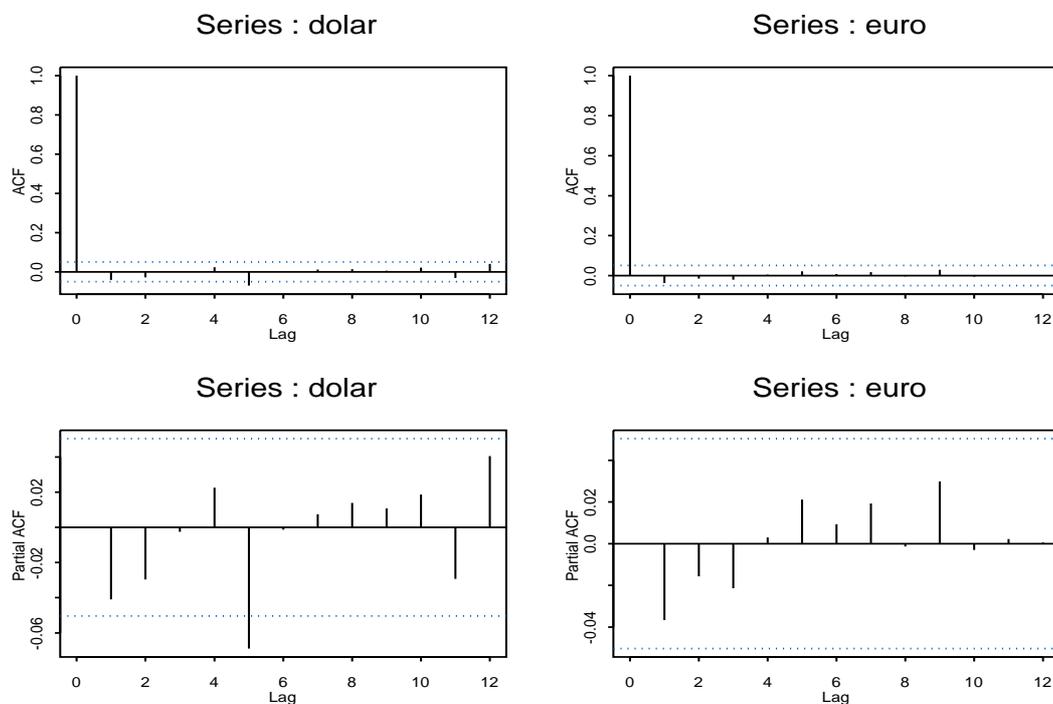


Figura 5.3: Series: ACF y PACF

Para la sustentar o no la normalidad se considera la prueba Jarque-Bera, encontrando un p-value cero para cada una, por lo que se rechaza en los dos casos la distribución Gaussiana.

A pesar de que nos encontramos con observaciones no correlacionados, la prueba gráfica de relaciones ARCH que utiliza las series elevadas al cuadrado, permite percibir la dependencia de segundo orden en los errores. Las funciones de autocorrelación de las series al cuadrado en la Figura 5.4 sostienen dicha idea, así que utilizando la prueba Ljung-Box hemos obtenido primero el estadístico  $Q(12)=199.93$  para la serie dólar al cuadrado y  $Q(12)=155.59$  para la serie euro al cuadrado. Entonces podemos intuir que las series cuadráticas son correlacionadas.

Formalmente la presencia de efectos ARCH puede encontrarse con la prueba de Multiplicador de Lagrange, entonces aplicándola directamente a las serie dólar y euro, otorga un p-value cero en ambas. Concluyendo que los efectos ARCH no pueden rechazarse en ningún caso.

## 5.2. Modelos Bivariados

### Multivariate Series : dolar.euro^2

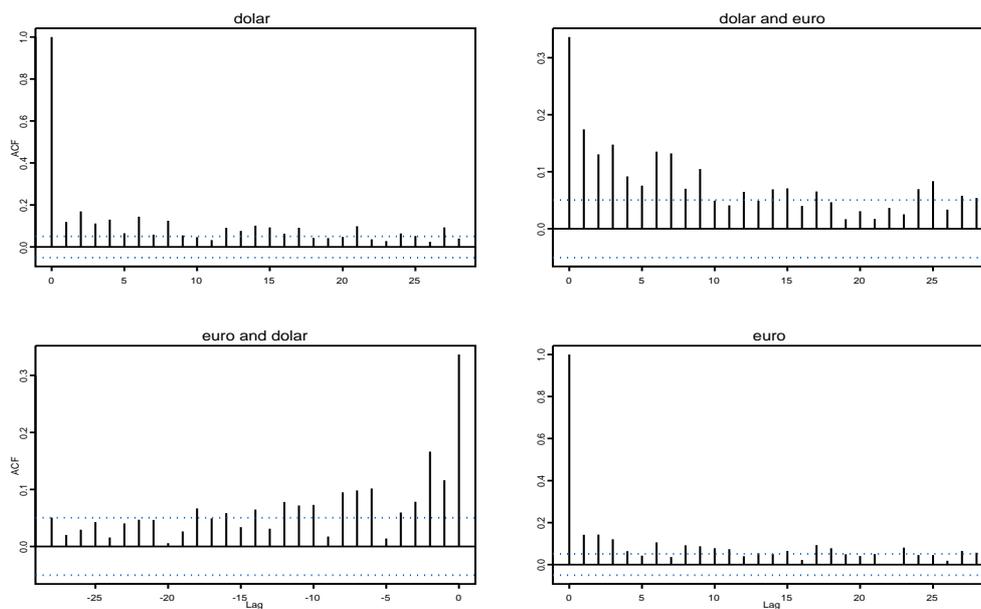


Figura 5.4: ACF Serie Bivariada al Cuadrado

Se ha verificado individualmente en las series a analizar la presencia de componentes ARCH. Ahora, vistas conjuntamente como un vector bivariado, en la Figura 5.4 se muestran las autocorrelaciones y las autocorrelaciones cruzadas para las series de segundo orden y se observa su significancia.

De la especificación  $\mathbf{r}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t$ , se considerará a la media condicional constante, enfocándonos entonces en la modelación del error  $\boldsymbol{\epsilon}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t$ , a través de la matriz de varianzas condicional  $\mathbf{H}_t$ . Comparativamente cada uno de los modelos se ajustó usando una distribución condicional Gaussiana para  $\mathbf{z}_t$  y a causa de la no normalidad de los residuales, enseguida se planteó una distribución Student-t, observando cambios mínimos en los coeficientes estimados, pero sí una ganancia en los criterios de información, reflejada en la verosimilitud y en el cumplimiento de los supuestos.

En esta sección se presentan solamente los ajustes bivariados bajo la distribución Student-t para el proceso de innovación, con el objetivo de capturar el efecto leptocúrtico de la series.

### 5.2.1. VEC Diagonal

La primera extensión a los modelos multivariados se encuentra en el modelo VEC, definido por una función lineal de productos cruzados de los errores rezagados y las matrices de varianzas condicionales pasadas. Se motivan por la naturaleza simétrica de la matriz heteroscedástica condicional y aligeran la notación aplicando el operador  $\text{vech}(\cdot)$  a la matriz de varianzas condicionales, guardando sólo la información de interés.

Su fuerza radica en la simplicidad en la modelación de las varianzas y covarianzas condicionales, pero enfrenta la problemática de garantizar que la matriz  $\mathbf{H}_t$  sea positiva definida, llevando a restricciones muchas veces difíciles de cumplir en la estimación.

Aprovechando los parsimoniosos y útiles GARCH(1,1), se seguirá el modelo de orden uno, es decir, la dependencia de sólo un vector de error cuadrado pasado y la matriz de varianzas condicionales inmediata anterior. Y aunado a que en términos de estimación el encontrar los parámetros involucrados es una tarea exhaustiva, las matrices de parámetros serán diagonales.

Considerando el modelo DVEC(1,1), se ha obtenido la estimación

$$\begin{aligned} h_{11,t} &= 0.0030 + 0.0580\epsilon_{1,t-1}^2 + 0.9286h_{11,t-1}, \\ h_{12,t} &= 0.0008 + 0.0194\epsilon_{1,t-1}\epsilon_{2,t-1} + 0.9708h_{12,t-1}, \\ h_{22,t} &= 0.0029 + 0.0323\epsilon_{2,t-1}^2 + 0.9625h_{22,t-1}, \end{aligned}$$

donde cada uno de los parámetros tiene significancia al 5%, es decir, se rechaza que sean cero.

Se tiene una dependencia alta en los valores de la varianzas y covarianzas condicionales anteriores, indicador de persistencia en la volatilidad.

Los residuales estandarizados estimados se definen como  $\hat{\mathbf{u}}_{i,t} = \hat{\epsilon}_{i,t} / \sqrt{\hat{h}_{ii,t}}$  y siguen los supuestos:

- Son una serie no correlacionada.
- Los residuales estandarizados cuadrados son no correlacionados, carecen de efectos ARCH y se ajustan al QQ-Plot teórico de la distribución estandarizada adoptada.

Para nuestro ejemplo es importante revisar los supuestos respecto a ellos. En el primer caso de la prueba Ljung-Box tenemos los estadísticos  $Q(12)=12.99$  para la serie dólar y  $Q(12)=22.53$  para la serie euro, de donde encontramos presencia de correlación en los residuales estandarizados en el ajuste de la serie euro.

Los residuales estandarizados al cuadrado se comportan de manera adecuada pues la prueba de Ljung-Box proporciona los estadísticos  $Q(12)=6.48$  para la serie dólar y

$Q(12)=20.14$ , rechazando autocorrelación y la prueba de Multiplicador de Lagrange también rechaza la existencia de efectos ARCH al obtener un  $p\text{-value}=0.8839$  en la serie dólar y un  $p\text{-value}=0.0519$  para la serie euro. La Figura C.2 muestra la no correlación y de acuerdo al QQ-Plot (Figura C.3) los residuales se ajustan satisfactoriamente a la distribución Student-t.<sup>15</sup>

En la Figura 5.5 encontramos las volatilidades de las series de rendimientos y la correlación condicional.

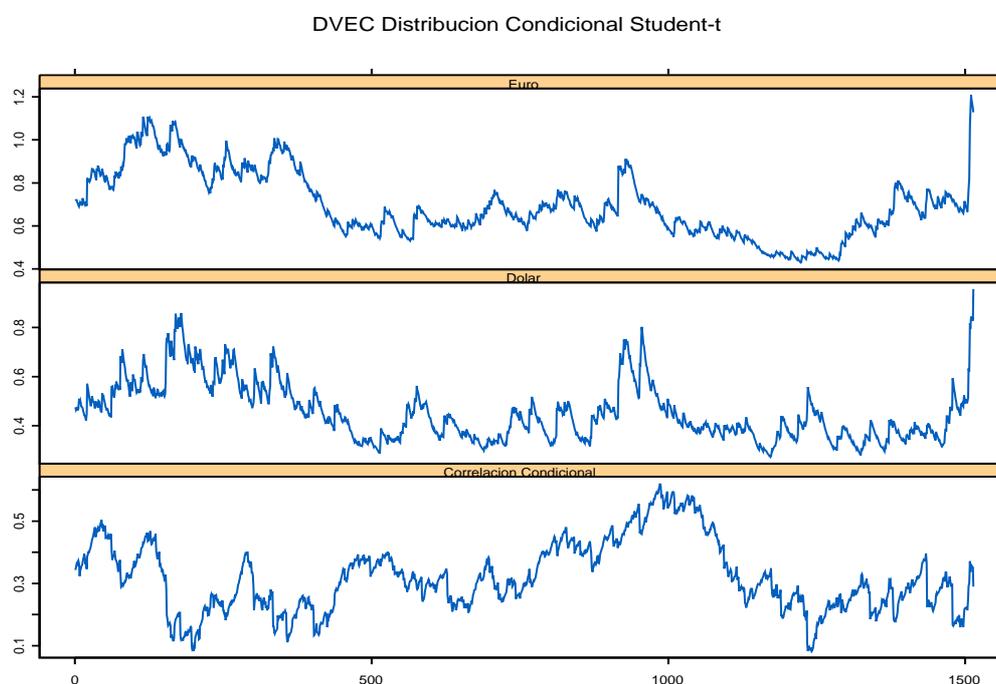


Figura 5.5: Volatilidades DVEC (Student-t)

Las series muestran gran actividad de la volatilidad, con varios lapsos de altas y bajas (conglomerados de volatilidad). Efectivamente de la escala, es claro que la serie euro tiene un rango de volatilidad más amplio, pero nuevamente la serie dólar es más volátil en el sentido de los cambios, que aunque son más pequeños se presentan con mayor frecuencia.

La tercera parte de la gráfica es de gran utilidad, refleja la correlación condicional e indica que se encuentran altamente relacionadas, además permite remarcar el comportamiento de correlación negativa en el periodo que suponíamos entre 2007 y 2008.

<sup>15</sup>Ver Apéndice C

## 5.2.2. BEKK

El siguiente modelo ajustado es el BEKK, que por definición proporciona una matriz de varianza condicional positiva definida y facilitará la estimación haciéndola más eficiente.

Su formulación es

$$\mathbf{H}_t = \mathbf{C}^{*'} \mathbf{C}^* + \sum_{k=1}^K \mathbf{A}_k^{*'} (\epsilon_{t-1} \epsilon_{t-1}') \mathbf{A}_k^* + \sum_{k=1}^K \mathbf{B}_k^{*'} \mathbf{H}_{t-1} \mathbf{B}_k^*.$$

Por su generalidad en el valor del índice  $K$ , parece ser bastante flexible, pero por la rapidez con la que crece el número de parámetros, se sugiere utilizar el valor de  $K=1$  e igual número en el caso de los rezagos.

Consideramos entonces el BEKK(1,1,1), es decir, el modelo de orden uno, obteniendo la matriz de varianza condicional

$$\begin{aligned} \begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} &= \begin{bmatrix} 0.0522 & 0 \\ -0.0111 & 0.0448 \end{bmatrix} \begin{bmatrix} 0.0522 & -0.0111 \\ 0 & 0.0448 \end{bmatrix} \\ &+ \begin{bmatrix} 0.2199 & -0.0618 \\ 0.0081 & 0.1677 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1} \epsilon_{2,t-1} \\ \epsilon_{1,t-1} \epsilon_{2,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} 0.2199 & 0.0081 \\ -0.0618 & 0.1677 \end{bmatrix} \\ &+ \begin{bmatrix} 0.9660 & 0.0088 \\ 0.0047 & 0.9837 \end{bmatrix} \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} 0.9660 & 0.0047 \\ 0.0088 & 0.9837 \end{bmatrix}. \end{aligned}$$

La especificación proporciona 11 parámetros, el cual es un número cercano al del modelo VEC diagonal de orden uno, sin embargo, existen algunas observaciones:

- Los parámetros no tienen interpretación directa con respecto a la varianza y los rezagos que incluyen. Por ejemplo, los valores negativos en la matriz  $\mathbf{A}^*$ , no necesariamente quieren decir que exista alguna relación en los errores cruzados, o sea, en el producto  $\epsilon_{1,t-1} \epsilon_{2,t-1}$ .
- Los residuales estandarizados presentan autocorrelación para la serie euro, ya que se obtuvo para ellos  $Q(12)=22.39$  al realizar la prueba Ljung-Box. Comparado con el valor  $\chi_{(12)}=21.02$ , el estadístico  $Q$  es mayor.
- También los residuales estandarizados al cuadrado de la serie euro son autocorrelacionados (Figura C.4) y usando la prueba de multiplicador de Lagrange se rechaza la hipótesis de no efectos ARCH.<sup>16</sup>

<sup>16</sup>Ver Apéndice C

La representación de las volatilidades (Figura 5.6) es muy parecida a la del modelo VEC, incluso la correlación condicional da una explicación adecuada de los movimientos, permitiendo distinguir los periodos de correlación positiva y negativa.

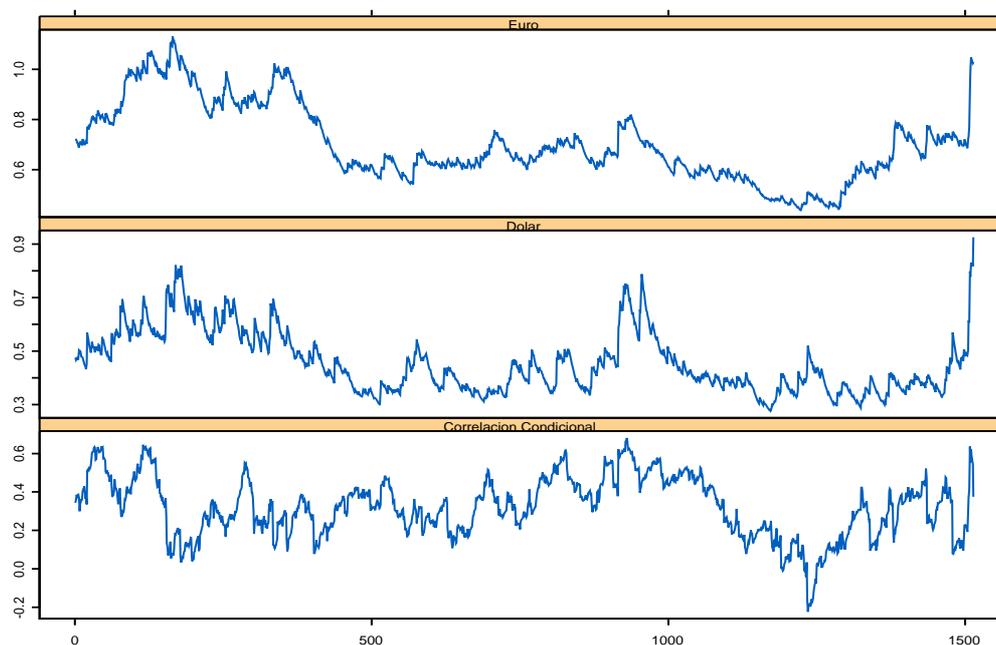


Figura 5.6: Volatilidades BEKK (Student-t)

### 5.2.3. O-GARCH

Basados en la descomposición de la matriz de varianza condicional en factores heteroscedásticos, encontramos en especial el modelo GARCH ortogonal (O-GARCH).

Su idea es bastante práctica, ya que previo a realizar el ajuste, se realiza el análisis de componentes principales a las series de rendimientos; observando la proporción de varianza explicada por cada componente, se decidirá cuántos modelos (univariados por cada componente) GARCH se requieren para describir a la matriz de varianza condicional. En el caso bivariado, tal vez no sea necesario limitar el empleo de modelos, ya que seguramente los dos ajustes a las componentes representarán el comportamiento heteroscedástico de  $H_t$  transformada en sus eigenvalores y eigenvectores.

En el cuadro 5.1 se tiene el análisis de componentes principales, encontrando el eigenvalor asociado a la primera y segunda componentes, así como la proporción de varianza explicada por la mismas.

Componente	Eigenvalor	Proporción de Varianza
Comp. 1	0.5564	0.7513
Comp. 2	0.1841	0.2486

Cuadro 5.1: Componentes Principales

Nótese que con la primera componente principal se alcanza un poco más de 75% de la varianza e incluyendo a la segunda componente se contará con el total de la variabilidad. Calculando las cargas o pesos de las variables en cada una de las componente, en el siguiente cuadro se muestran las ponderaciones.

	Comp. 1	Comp. 2
dólar	0.334	0.943
euro	0.943	-0.334

Cuadro 5.2: Cargas en las Componentes

Hemos encontrado las combinaciones lineales de las variables que maximizan la estructura de varianza, entonces el siguiente paso es estimar los modelo GARCH. Siendo consistentes con la noción de parsimonia y explicación en los parámetros, ajustamos dos GARCH(1,1) de la forma  $\sigma_t^2 = \omega + \alpha\xi_{t-1}^2 + \beta\sigma_{t-1}^2$ .

Véase el Cuadro 5.3 encontrando los coeficientes estimados ajustando un modelo por cada componente.

Coeficientes	Comp. 1	Comp. 2
$\omega$	0.0024	0.0042
$\alpha$	0.0348	0.0758
$\beta$	0.9614	0.9032

Cuadro 5.3: Ajuste OGARCH

De acuerdo a ellos, se observa que la reacción a la volatilidad del mercado ( $\alpha$ 's) no es muy alta, es decir, las componentes reaccionan lentamente ante los movimientos de los errores. En cambio la persistencia en la volatilidad ( $\beta$ 's) tiene un peso mayor y probablemente desempeñara un efecto continuo si se desean realizar pronósticos.

Los supuestos de los residuales estandarizados se cumplen, excepto por la autocorrelación de los residuales estandarizados para la segunda componente, donde la prueba Ljung-Box otorga un estadístico  $Q(12)=53.319$  bastante grande y un p-value por debajo de 0.05, rechazando la hipótesis de no correlación al 95 %.

Los residuales estandarizados al cuadrado no tienen algún inconveniente, son no correlacionados (sus estadísticos así lo muestran) y no presentan dependencia de segundo orden, en otras palabras, se rechazan efectos ARCH en ellos (Figura C.7). Además su ajuste al QQ-Plot de la distribución Student-t es bueno (Figura C.7) y mejora notablemente respecto a la distribución Gaussiana.

En la Figura 5.7 encontramos un refinamiento de las desviaciones estándar condicionales y la correlación condicional, observando un suavizamiento en los movimientos y de donde es posible establecer co-movimientos e intervalos de correlación negativa. Siendo más cuidadosos inclusive se pueden separar las correlaciones por periodos trimestrales observando un patrón (conglomerados de volatilidad).

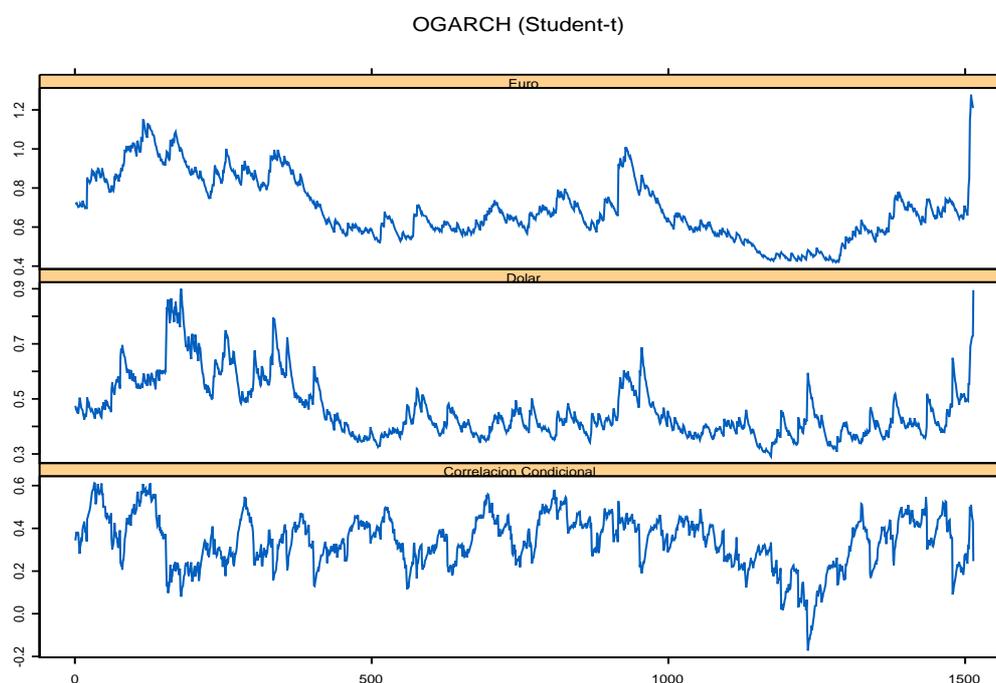


Figura 5.7: Volatilidad OGARCH (Student-t)

## 5.2.4. CCC

Consideremos la descomposición de la matriz de varianzas condicional  $\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$ , de tal forma que  $\mathbf{R}_t$  es la matriz de correlación condicional y  $\mathbf{D}_t$  es la matriz diagonal que contiene las desviaciones estándar condicionales. En la representación anterior si  $\mathbf{R}_t$  es invariante en el tiempo, es decir,  $\mathbf{R}_t = \mathbf{R}$ , encontramos el modelo de Correlación Condicional Constante (CCC).

Su metodología consiste en fijar la correlación condicional y estimar las varianzas condicionales a través de modelos univariados, ya sean los GARCH empleados generalmente o las diferentes especificaciones tratadas en el Capítulo 3 (IGARCH, GARCH-M y EGARCH).

Finalmente se implementa el caso de correlación condicional constante, teniendo disponible en S-PLUS la aplicación CCC y se ha dividido el ajuste de los datos en dos casos.

**CCC: Un elemento**

Primero adoptamos el modelo CCC considerando la especificación inicial de Bollerslev [1990] donde cada varianzas condicional sigue un GARCH(1,1) y no existe interacción entre las varianzas condicionales pasadas, con la estimación

$$\begin{bmatrix} h_{11,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} 0.0026 \\ 0.0028 \end{bmatrix} + \begin{bmatrix} 0.0548 & \cdot \\ \cdot & 0.0330 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} 0.9340 & \cdot \\ \cdot & 0.9622 \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{22,t-1} \end{bmatrix}.$$

En el cuadro 5.4 se tiene la matriz de correlación condicional constante estimada.

	dólar	euro
dólar	1.0000	0.3283
euro	0.3283	1.0000

Cuadro 5.4: CCC

El valor de la correlación es razonable de acuerdo a periodos en los que se mueven conjuntamente las series y su error estándar es 0.0258. Entonces, dado el valor 0.3283 como correlación condicional, se ajustaron por separado modelos GARCH(1,1) para cada serie de rendimientos.

Interpretamos los coeficientes como en modelos univariados siendo cada uno de ellos significativo al 5% y donde el peso de la varianzas condicional pasada es notable. Los criterios de información, tanto el de Akaike como el Bayesiano son menores respecto a los del modelo O-GARCH.

El modelo tiene una falla en los residuales estandarizados de la serie euro, pues con el estadístico  $Q(12)=22.90$ , notamos que son correlacionados. En tanto los residuales estandarizados al cuadrado preservan no correlaciones *serias* y no efectos ARCH. En la Figura C.9 observamos el aporte de la distribución Student-t, ya que el QQ-Plot de los residuales luce mejor al considerar ésta distribución en lugar de la Gaussiana.

La Figura 5.8 cuenta con las dos gráficas de volatilidad y muestra los aspectos mencionados anteriormente, salvo que ahora hemos perdido la contribución de la correlación condicional variante en el tiempo.

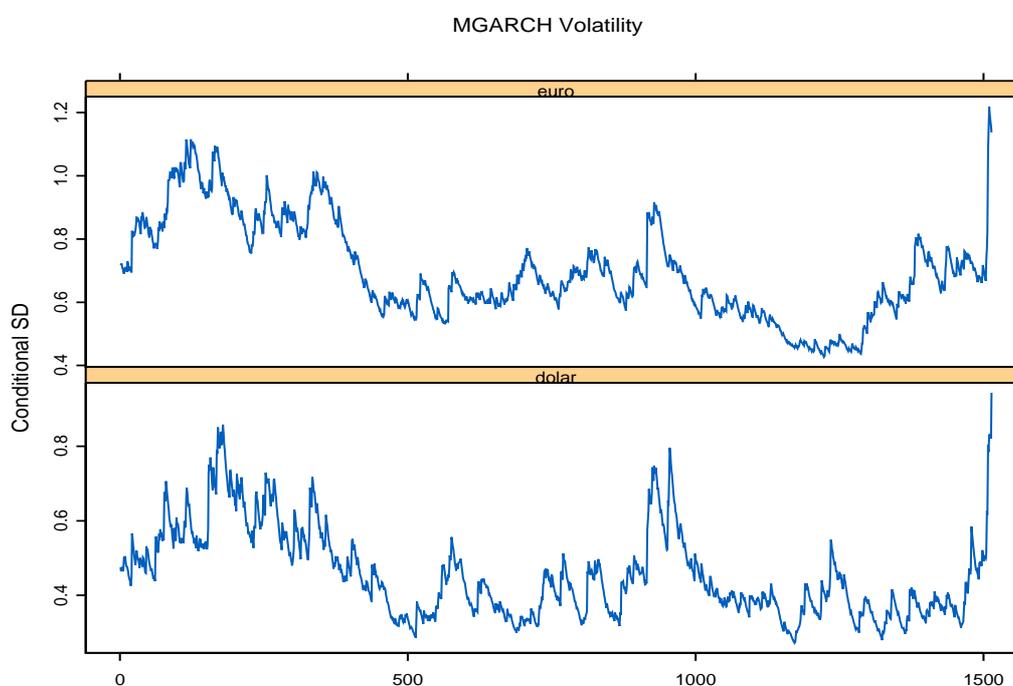


Figura 5.8: Volatilidad CCC (Student-t Un elemento)

### CCC: Dos Elementos

Por otra parte ajustamos un modelo CCC, ahora permitiendo varianzas condicionales se relacionen, de tal manera que la correlación siga siendo constante, pero exista en las volatilidades una contribución de los dos errores cuadráticos de rezago y diferentes varianzas condicionales pasadas.

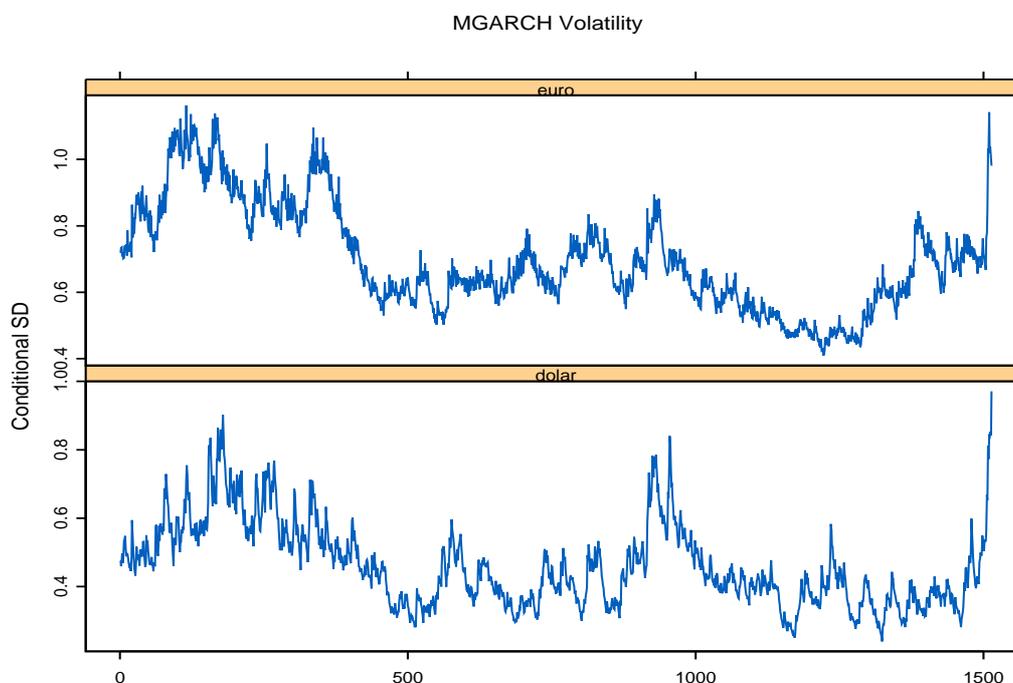


Figura 5.9: Volatilidad CCC (Student-t Dos elementos)

Veamos las varianzas condicionales estimadas

$$\begin{bmatrix} h_{11,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} 0.0007 \\ 0.0079 \end{bmatrix} + \begin{bmatrix} 0.0367 & 0.0561 \\ 0.0279 & 0.0412 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} 0.9523 & 0.8200 \\ -0.4433 & 0.9596 \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{22,t-1} \end{bmatrix}.$$

Se han introducido nuevos parámetros en las ecuaciones de cada  $h_{ii,t}$  que dan cuenta de la relación entre los errores y varianzas condicionales pasadas. Por ejemplo, en la segunda ecuación existe un valor negativo para  $h_{11,t-1}$ , que bien puede interpretarse como una relación inversa entre las volatilidades  $h_{22,t}$  y  $h_{11,t}$ .

Como en este modelo CCC la correlación condicional constante es 0.3298, notamos que disminuye ligeramente respecto al caso anterior, en cambio no supera la deficiencia en los residuales estandarizados, pues siguen siendo correlacionados; aunque los residuales estandarizados al cuadrado continúan cumpliendo los supuestos (Figuras C.10 y C.11).

En la Figura 5.9 se tiene la gráfica de las volatilidades.

### 5.2.5. Resultados

Los objetivos de nuestra aplicación han sido analizar de la dependencia temporal de las paridades Peso/Dólar y Peso/Euro y establecer las variaciones en la volatilidad ocurridas históricamente durante el periodo de observación, además de obtener información adicional para explicación de fenómenos económicos y así poder proporcionar una interpretación a los movimientos en los tipos de cambio.

De los modelos considerados, la parametrización **DVEC** otorga una buena interpretación de los estimadores, ya que los coeficientes involucrados en la matriz de varianza condicional representan relaciones directas de las volatilidades y co-volatilidades. Notamos sin embargo, que los residuales del tipo de cambio euro son correlacionados, hecho que puede atribuirse a algún comportamiento no estacionario de la serie y el cual podría solucionarse incluyendo otro rezago en el ajuste, bajo el costo de obtener nuevos parámetros que incluso en lugar de mejorarlo, puedan introducir *ruido* no deseable.

Respecto al modelo **BEKK**, podríamos decir que se trata de una modelo teórico cuya bondad es proporcionar una matriz de varianza condicional positiva definida, pero a la vez pierde la explicación de los coeficientes estimados. Del ejemplo, al no contar con una interpretación de los parámetros y puesto que los residuales son correlacionados para las dos series, pensamos que no es el modelo más conveniente para nuestros datos.

El caso práctico nos ha permitido descubrir el **GARCH ortogonal**, que describe de manera adecuada la relación de las volatilidades en los tipo de cambio y logra fundamentar el modelo a través de la teoría de componentes principales, dando una interpretación conveniente de los parámetros. Referente a la estimación, proporciona las ventajas de asegurar que la matriz de covarianza condicional sea positiva definida sin imponer restricciones severas y usando el mismo modelo paramétrico para cada una de las series. También en él, las correlaciones condicionales son más estables, debido a que el ruido en las componentes puede medirse y si se desea puede ignorarse, tomando en cuenta sólo la información necesaria.

Basados en los modelos anteriores, la imposición de correlación condicional constante es un poco irrealista aunque posiblemente en ciertos periodos podría adoptarse. En el modelo **CCC** se produce que en la mayoría del tiempo las observaciones compartan un patrón semejante, en cambio la diferencia entre los movimientos de las volatilidades dólar y euro para determinados etapas contradice esa suposición, pues para esos lapsos podría pensarse que la correlación condicional es variante y además tiene signo negativo.

Ocurre lo mismo para el segundo **CCC**, pero la ganancia al incluir los que hemos llamado *dos elementos*, es obtener un refinamiento en las volatilidades, percibiendo pequeñas variaciones de forma más precisa. En el estudio de series de alta frecuencia cuya correlación condicional permanezca en cierto nivel, este modelo será bastante adecuado para modelar la volatilidad.

## Capítulo 6

# Conclusiones

Motivado por la visita de Robert Engle, premio Nobel en Ciencias Económicas 2003, a la Facultad de Ciencias UNAM en septiembre de 2008, se ha escrito esta tesis que trata de modelos de volatilidad multivariados (MGARCH).

Ellos se encuentran dentro de la disciplina estadística de series de tiempo, por lo que se han desarrollado los elementos necesarios para el entendimiento de datos ordenados en el tiempo. Los procesos a observar son series financieras de precios, centrándonos en series de rendimientos. Un concepto clave es estacionariedad de segundo orden, implicando a la media y varianza de una serie a no depender del tiempo. Esto es, no importa el lapso donde se observe la serie, los dos primeros momentos permanecerán invariantes.

Los mencionados anteriormente son llamados momentos no condicionales, ahora, utilizando la información disponible (precios históricos), se definen los momentos condicionales: media y varianza. Del segundo se desprende el concepto de volatilidad, formalmente definida como la raíz cuadrada de la tasa a la que la varianza condicional de los rendimientos crece o decrece en el horizonte de tiempo.

El tratamiento de la volatilidad tiene diferentes enfoques, el más estudiado es considerando modelos autoregresivos con heteroscedasticidad condicional (ARCH).

El modelo ARCH de Engle [1982] representa un valioso aporte a la economía, sin precedentes, y es una pieza medular de la econometría contemporánea. De él se derivan un sinnúmero de modelaciones con diferentes especificaciones, cuya esencia radica en los hechos estilizados de la volatilidad:

- *Conglomerados de volatilidad.* Grandes cambios suelen ser seguidos por grandes cambios, mientras que cambios pequeños generalmente tienden a seguir otros cambios pequeños.

- *Asimetría.* La volatilidad parece reaccionar de manera distinta ante incrementos o caídas en los precios de los instrumentos.

La parametrización univariada GARCH ha sido ampliamente estudiada, por lo que nuestro objetivo ha sido presentar una revisión bibliográfica del paradigma MGARCH y concluir con la aplicación de los modelos DVEC, BEKK, OGARCH y CCC para el caso bivariado en el programa estadístico S-PLUS.

Notemos que los modelos contienen una cantidad considerable de parámetros, dependiendo de la especificación adoptada, así que la tarea de realizar las estimaciones no es sencilla. De la accesibilidad a programas estadísticos únicamente se han presentado las especificaciones anteriores, ya que S-PLUS contiene las rutinas especiales para los ajustes con la función `mgarch`. En ella aparecen las opciones `dvec`, `bekk`, `prcomp.garch` y `ccc`, presentadas en el Capítulo 5. Descartamos la realización de pronósticos debido a que las herramientas especializadas para se implementarlos se encuentran en desarrollo y seguramente pronto se contará con funciones que permitan obtener predicciones con un grado de precisión aceptable.

Para llevar a cabo la aplicación no encontramos inconvenientes numéricos en los datos que impidieran la estimación, claro, las series tienen un comportamiento regular y sus movimientos son muy similares. Sería interesante ampliar el problema e incluir más series en la modelación.

En particular, para las volatilidades de las series de tipo de cambio consideradas en nuestro ejemplo se distinguen en primer lugar los hechos estilizados comunes en rendimientos, observando que la distribución de ellos difiere de la Normal a causa de su exceso de curtosis. De esta manera, la novedad implementada en los modelos fue la utilización de una distribución Student-t en el proceso de innovación.

El resultado fue una mejora en el cumplimiento de los supuestos para los residuales estandarizados y en la obtención de la verosimilitud, reflejada en una disminución de los criterios de información AIC y BIC.

Podemos decir que las volatilidades de los tipos de cambio Peso/Dólar y Peso/Euro son altamente correlacionadas con varios lapsos de movimientos comunes, pero también se localiza un periodo de correlación negativa, que responde a un comportamiento estable en las bolsas de valores pertenecientes a la Comunidad Económica Europea y la baja ocurrida en el mercado financiero de Estados Unidos, causada por su crisis hipotecaria.

Creemos que nuestro trabajo es sólo un primer paso en el tratamiento multivariado de series de tiempo y seguramente será una referencia para futura investigación.

## Apéndice A

# Demostraciones

### Observación 4 Capítulo 3

Desarrollando la expresión y bajo condiciones de estacionariedad ( $|\phi_1| < 1$ ) se sigue que el error cuadrático medio o varianza del pronóstico no condicional es

$$\begin{aligned}\mathbb{E}\left[(r_{t+1} - \mathbb{E}[r_{t+1}])^2\right] &= \mathbb{E}\left[\left(\phi_0 + \phi_1 r_t + \epsilon_{t+1} - \frac{\phi_0}{1 - \phi_1}\right)^2\right] \\ &= \mathbb{E}\left[\left(\phi_0 + \phi_1(\phi_0 + \phi_1 r_{t-1} + \epsilon_t) + \epsilon_{t+1} - \frac{\phi_0}{1 - \phi_1}\right)^2\right] \\ &= \mathbb{E}\left[\left(\epsilon_{t+1} + \phi_1 \epsilon_t + \phi_0 + \phi_0 \phi_1 + \phi_1^2 r_{t-1} - \frac{\phi_0}{1 - \phi_1}\right)^2\right] \\ &= \mathbb{E}\left[\left(\epsilon_{t+1} + \phi_1 \epsilon_t + \phi_0 + \phi_0 \phi_1 + \phi_1^2(\phi_0 + \phi_1 r_{t-2} + \epsilon_{t-1}) - \frac{\phi_0}{1 - \phi_1}\right)^2\right] \\ &\vdots \\ &= \mathbb{E}\left[\left(\epsilon_{t+1} + \phi_1 \epsilon_t + \phi_1^2 \epsilon_{t-1} + \dots + \phi_0(1 + \phi_1 + \phi_1^2 + \dots) - \frac{\phi_0}{1 - \phi_1}\right)^2\right] \\ &= \mathbb{E}\left[(\epsilon_{t+1} + \phi_1 \epsilon_t + \phi_1^2 \epsilon_{t-1} + \dots)^2\right] \\ &= \mathbb{E}[\epsilon_{t+1}^2 + \phi_1^2 \epsilon_t^2 + \phi_1^4 \epsilon_{t-1}^2 + \phi_1^6 \epsilon_{t-2}^2 + \dots] \\ &= \mathbb{E}[\epsilon_{t+1}^2] + \phi_1^2 \mathbb{E}[\epsilon_t^2] + \phi_1^4 \mathbb{E}[\epsilon_{t-1}^2] + \phi_1^6 \mathbb{E}[\epsilon_{t-2}^2] + \dots \\ &= \sigma^2 + \phi_1^2 \sigma^2 + \phi_1^4 \sigma^2 + \phi_1^6 \sigma^2 + \dots = \frac{\sigma^2}{1 - \phi_1^2} .\end{aligned}$$

**Teorema 1 Capítulo 3**

Consideremos la representación del modelo GARCH como un ARMA( $r,s$ ) para  $\epsilon_t^2$

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + \nu_t - \sum_{j=1}^s \beta_j \nu_{t-j}.$$

Tomando esperanza en ambos lados se obtiene

$$\mathbb{E}[\epsilon_t^2] = \alpha_0 + \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) \mathbb{E}[\epsilon_{t-i}^2].$$

Entonces, si las raíces del polinomio característico  $1 - \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) B^i$  se encuentran estrictamente fuera del círculo unitario, se tiene que el proceso  $\mathbb{E}[\epsilon_t^2]$  converge y es (asintóticamente) estacionario de segundo orden, lo que implica  $\alpha(1) + \beta(1) < 1$ .

Si ocurriera que  $\alpha(1) + \beta(1)$  fuera mayor o igual a 1, ciertamente  $1 - \alpha(1) + \beta(1) \leq 1$ , existiendo una raíz del polinomio característico entre cero y uno.

Por otra parte supongamos  $\alpha(1) + \beta(1) < 1$ , y sea la raíz  $z$  del polinomio característico cuyo módulo es estrictamente menor a la unidad, teniéndose

$$1 = \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) z^i = \left| \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) z^i \right| \leq \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) |z|^i \leq \alpha(1) + \beta(1) < 1,$$

llegamos a una contradicción.

Por lo tanto se verifica que

$$\alpha(1) + \beta(1) = \sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1,$$

es una condición necesaria para la estacionariedad del proceso.

## Apéndice B

# Análisis Multivariado

### Teoría de Matrices

Una matriz  $A_{m \times n}$ , donde  $m \times n$  es la dimensión se denotará por  $A$  y puede escribirse

$$A \equiv [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Observando que los términos  $a_{ii}$  son los elementos de la diagonal y cuando  $m = n$  la matriz se dice cuadrada. La matriz de dimensión  $m \times 1$  se dice vector columna  $m$ -dimensional.

Si  $a_{ij} = 0$  para  $i \neq j$  y  $m = n$ , entonces se habla de una matriz diagonal denotada por  $\text{diag}[a_{ii}]$ . Además si  $a_{ii} = 1 \forall i \in \{1, \dots, m\}$ , entonces se tiene la matriz identidad  $I_m$  de dimensión  $m \times m$ .

En la matriz  $A$ , cambiando los renglones por columnas, se obtiene la matriz transpuesta  $A'$ , de tal manera

$$A' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nm} \end{bmatrix},$$

de donde  $a'_{ij} = a_{ji}$  y  $(A')' = A$ . Si  $A' = A$ , se dice que  $A$  es una matriz simétrica.

Sean  $A = [a_{ij}]_{m \times n}$  y  $B = [b_{ij}]_{p \times q}$  matrices de dimensiones dadas en el subíndice, y  $c$  un número real, se tienen las operaciones

- Suma:  $A + B = [a_{ij} + b_{ij}]_{m \times n}$  si  $m = p$  y  $n = q$ .
- Multiplicación Escalar:  $cA = [ca_{ij}]_{m \times n}$ .
- Multiplicación:  $AB = [\sum_{k=1}^n a_{ik}c_{kj}]_{m \times q}$  siempre que  $n = p$ .

Una matriz cuadrada  $A$  es no singular o invertible, si existe una única matriz tal que  $AA^{-1} = I_m$ . En este caso  $A^{-1}$  es llamada matriz *inversa* de  $A$ . Además para una matriz no singular se cumple  $(A^{-1})' = (A')^{-1}$ .

La traza de una matriz  $A$  es la suma de los elementos de su diagonal, de tal manera  $tr(A) = \sum_{i=1}^m a_{ii}$ .

Se puede verificar que

- $tr(A + C) = tr(A) + tr(C)$
- $tr(A) = tr(A')$
- $tr(AC) = tr(CA)$

Un vector  $m$ -dimensional  $x$  es un *eigenvector* con *eigenvalor*  $\lambda$  de la matriz  $A$  si se tiene que  $Ax = \lambda x$ . La matriz  $A$  es no singular si y sólo si todos sus eigenvalores son diferentes de cero.

Si  $\lambda_i$ ,  $i \in \{1, \dots, m\}$  son los eigenvalores de la matriz  $A$ , entonces puede probarse que

$$tr(A) = \sum_i^m \lambda_i,$$

y su **rango** será el número de eigenvalores diferentes de cero de la matriz simétrica  $AA'$ . Se dice que es de rango completo si así lo son todos sus eigenvalores, y en caso contrario es de rango reducido.

También el determinante de  $A$  puede definirse como  $|A| = \prod_{i=1}^m \lambda_i$ .

Una matriz cuadrada  $A$  se dice positiva definida si

- Es simétrica y  $b'Ab > 0$  para cada vector  $b' = (b_1, \dots, b_m)$  diferente de cero.

De la misma manera se dice que  $A$  es una matriz no negativa definida si para cada vector  $b' = (b_1, \dots, b_m)$  se tiene que  $b'Ab \geq 0$ .

## Descomposición Espectral

Si  $A$  es una matriz simétrica, entonces

$$A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \cdots + \lambda_p e_p e_p',$$

donde  $e_i e_i' = 1$ ,  $e_i e_j' = 0$  si  $i \neq j$  y  $\lambda_i$  son los eigenvalores y  $e_i$  los correspondientes eigenvectores de  $A$ .

Análogamente

- Si  $\lambda_i > 0 \forall i$ , entonces  $A$  es positiva definida.
- Si  $\lambda_i = 0$  para alguna  $i$ , se dice que  $A$  es no negativa definida.

Para una matriz  $A$  positiva definida se establece la *descomposición espectral* de  $A$  como

$$A = \sum_{i=1}^p \lambda_i e_i e_i' = P \Omega P', \quad (\text{B.1})$$

donde  $\Omega$  es una matriz diagonal consistente de todos los eigenvalores de  $A$ , esto es,  $\Omega = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  y  $P$  es una matriz de dimensión  $m \times m$  que contiene a los eigenvectores de  $A$ . Regularmente se ordenan los eigenvalores como  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$  y los eigenvectores  $e_1, \dots, e_m$ .

## Descomposición de Cholesky

Para una matriz simétrica  $A$ , existe una matriz triangular inferior  $L$  con elementos en la diagonal 1 y una matriz diagonal  $G$  tales que  $A = LGL'$ .

Si  $A$  es positiva definida, los elementos de la diagonal de  $G$  son positivos, en ese caso

$$A = L\sqrt{G}\sqrt{G}L' = (L\sqrt{G})(\sqrt{G}L)', \quad (\text{B.2})$$

donde  $L\sqrt{G}$  es otra vez una matriz triangular inferior y la raíz cuadrada es tomada elemento por elemento.

La ecuación B.2 es llamada *descomposición de Cholesky*, y muestra que una matriz positiva definida  $A$  puede diagonalizarse como

$$L^{-1}A(L')^{-1} = L^{-1}A(L^{-1})' = G.$$

### Vectorización, Producto de Kronecker y de Hadamard

Sea la matriz  $A$  de dimensión  $m \times n$ , escrita por  $A = [a_1, \dots, a_n]$  donde  $a_i$  representa la  $i$ -ésima columna, la vectorización de  $A$  se define como

$$\text{vec}(A) = [a'_1, a'_2, \dots, a'_m],$$

siendo un vector de dimensión  $mn \times 1$ .

Para dos matrices  $A_{m \times n}$  y  $C_{p \times q}$ , el producto de *Kronecker* es

$$A \otimes C = \begin{bmatrix} a_{11}C & a_{12}C & \cdots & a_{1n}C \\ a_{21}C & a_{22}C & \cdots & a_{2n}C \\ \vdots & \vdots & & \vdots \\ a_{m1}C & a_{m2}C & \cdots & a_{mn}C \end{bmatrix},$$

resultando una matriz de dimensión  $mp \times nq$  y con las propiedades

- $(A \otimes C)' = A' \otimes C'$ .
- $A \otimes (C + D) = A \otimes C + A \otimes D$ .
- $(A \otimes C)(F \otimes G) = (AF) \otimes (CG)$ .
- Si  $A$  y  $C$  son invertibles, entonces  $(A \otimes C)^{-1} = A^{-1} \otimes C^{-1}$ .

Entre los resultados importantes de las operaciones vectorización y producto de Kronecker se encuentran:

- $\text{vec}(A + C) = \text{vec}(A) + \text{vec}(C)$ .
- $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ .
- $\text{tr}(AC) = \text{vec}(C)'\text{vec}(A) = \text{vec}(A)'\text{vec}(C)$ .

Para una matriz cuadrada simétrica  $A = [a_{ij}]_{k \times k}$ , la vectorización de los elementos de su región triangular inferior se define con el operador

$$\text{vech}(A) = (a'_{1.}, a'_{2*}, \dots, a'_{k*})',$$

donde  $a_{1.}$  es la primera columna de  $A$  y  $a_{i*} = (a_{ii}, a_{i+1,i}, \dots, a_{ki})'$  es un vector  $(k - i + 1)$ -dimensional. Obsérvese entonces que la dimensión de  $\text{vech}(A)$  es  $\frac{k(k+1)}{2}$ .

Si  $A = (a_{ij})$  y  $B = (b_{ij})$  son ambas matrices  $m \times n$ , entonces el producto de *Hadamard*, denotado por  $A \odot B$ , es la matriz  $m \times n$  que contiene los productos elemento a elemento  $(a_{ij}b_{ij})$  en cada entrada.

### Observaciones Multivariadas

El conjunto de datos multivariados  $\mathbf{X}$  puede representarse como

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n \times p},$$

donde  $x_{jk}$  es la medición de la variable  $k$ -ésima en la observación  $j$ .

### Estadísticas Descriptivas

Dada la matriz de datos multivariados pueden calcularse para cada variable

1. Media Muestral:  $\bar{x}_k = \frac{1}{n} \sum_{j=1}^n x_{jk}$ ,  $k = 1, \dots, p$ .
2. Varianza Muestral:  $s_k^2 = \frac{1}{n} \sum_{j=1}^n (x_{jk} - \bar{x}_k)^2$  ó  $s_k^2 = \frac{1}{n-1} \sum_{j=1}^n (x_{jk} - \bar{x}_k)^2$ , con  $k = 1, \dots, p$ .
3. Covarianza Muestral:  $s_{ik} = \frac{1}{n} \sum_{j=1}^n (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)$ ,  $i, k = 1, \dots, p$ .
4. Coeficientes de Correlación Muestrales:

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_i^2} \sqrt{s_k^2}} = \frac{\sum_{j=1}^n (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)}{\sqrt{\sum_{j=1}^n (x_{ji} - \bar{x}_i)^2} \times \sqrt{\sum_{j=1}^n (x_{jk} - \bar{x}_k)^2}}.$$

Donde  $r_{ik}$  es conocido como el coeficiente de correlación de *Pearson*, y cuyas características son que mide la asociación lineal entre las variables, es invariante ante la escala y cumple  $-1 \leq r_{ik} \leq 1$ .

La representación matricial de ellos es

1. Vector de Medias Muestrales:  $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)'$ .
2. Matriz de Covarianza Muestral:  $\mathbf{S}_n = [s_{ik}]$ .
3. Matriz de Correlación Muestral:  $\mathbf{R}_n = [r_{ik}]$ .

### Matrices y Vectores Aleatorios

Se dice que  $\mathbf{X} = [X_{ij}]$  es una matriz aleatoria si  $X_{ij}$  es una variable aleatoria, en cuyo caso su esperanza es  $\mathbb{E}[\mathbf{X}] = [\mathbb{E}[X_{ij}]]$ .

#### §Propiedades

- $\mathbb{E}[\mathbf{X} + \mathbf{Y}] = \mathbb{E}[\mathbf{X}] + \mathbb{E}[\mathbf{Y}]$ .
- $\mathbb{E}[\mathbf{A}\mathbf{X}\mathbf{B}] = \mathbf{A}\mathbb{E}[\mathbf{X}]\mathbf{B}$ , con  $\mathbf{A}$  y  $\mathbf{B}$  matrices constantes.

Sean  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}_x$  y  $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}_x$ , la media y matriz de covarianza de  $\mathbf{X}$  respectivamente. Entonces para algún vector constante  $n$ -dimensional  $\mathbf{c}$ , se tienen

- $\mathbb{E}[\mathbf{c}'\mathbf{X}] = \mathbf{c}\mathbb{E}[\mathbf{X}] = \mathbf{c}\boldsymbol{\mu}_x$ .
- $\text{Var}(\mathbf{Y}) = \mathbf{c}'\boldsymbol{\Sigma}_x\mathbf{c}$ .
- $\mathbb{E}[\mathbf{X}\mathbf{X}'] = \boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x'$ .

Una combinación de variables aleatorias en  $\mathbf{X}$  puede escribirse  $\mathbf{Y} = \mathbf{c}'\mathbf{X}$ , donde  $\mathbf{c}$  es un vector constante  $p$ -dimensional. Por lo anterior la media y varianza de  $\mathbf{Y}$  son

- $\mathbb{E}[\mathbf{Y}] = \mathbf{c}'\boldsymbol{\mu}_x$ .
- $\text{Var}(\mathbf{Y}) = \mathbf{c}'\boldsymbol{\Sigma}_x\mathbf{c}$ .

### Muestras aleatorias

Sea  $\mathbf{X}_1, \dots, \mathbf{X}_n$  una muestra aleatoria de una distribución conjunta que tiene vector de media  $\boldsymbol{\mu}$  y matriz de covarianza  $\boldsymbol{\Sigma}_x$ . Se definen sus estimadores

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i, \quad \mathbf{S}_n = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$$

Pudiéndose probar que

$$\mathbb{E}[\bar{\mathbf{X}}] = \boldsymbol{\mu}, \quad \text{Cov}(\bar{\mathbf{X}}) = \frac{1}{n} \boldsymbol{\Sigma}_x \quad \text{y} \quad \mathbb{E}[\mathbf{S}_n] = \frac{n-1}{n} \boldsymbol{\Sigma}_x.$$

## Normal Multivariada

Si  $\mathbf{X}$  es un vector aleatorio normal multivariado (N-dimensional) con media  $\boldsymbol{\mu}$  y matriz de varianza  $\boldsymbol{\Sigma}$ , tiene una densidad

$$f(\mathbf{X}) = (2\pi)^{-N/2} |\det \boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})\right]. \quad (\text{B.3})$$

Se denota  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  y puede escribirse como  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{T}\boldsymbol{\xi}$ , con  $\boldsymbol{\xi} \sim N(0, \mathbf{I}_N)$ .

Lo anterior se verifica ya que  $\boldsymbol{\Sigma}$  es simétrica, entonces  $\boldsymbol{\Sigma} = \mathbf{T}\mathbf{T}'$ .

Por lo tanto  $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu}$  y  $\text{Var}(\mathbf{X}) = \text{Var}(\mathbf{T}\boldsymbol{\xi}) = \mathbf{T}\text{Var}(\boldsymbol{\xi})\mathbf{T}' = \mathbf{T}\mathbf{T}'$ .

### §Propiedades

- Sea  $\mathbf{c}'$  un vector constante, entonces  $\mathbf{c}'\mathbf{X} \sim N(\mathbf{c}'\boldsymbol{\mu}, \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c})$ , esto es, combinaciones lineales de  $\mathbf{X}$  tienen una distribución normal (univariada).
- Dado  $\mathbf{X} = (X_1, \dots, X_N)$  sus distribuciones marginales  $X_i$  son normales.

## Student-t Multivariada

Esta densidad denotada por  $ST(0, \mathbf{I}_N; \nu)$  es

$$g(\mathbf{z}_t | \nu) = \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})(\pi(\nu-2))^{N/2}} \left(1 + \frac{\mathbf{z}_t \mathbf{z}_t'}{\nu-2}\right)^{-\frac{\nu+N}{2}}, \quad (\text{B.4})$$

con  $\Gamma(\cdot)$  la función Gamma.

Se impone  $\nu > 2$  y aunque  $\mathbf{z}_t$  es no correlacionado, es decir,  $\mathbb{E}[\mathbf{z}_t] = 0$  y  $\text{Var}(\mathbf{z}_t) = \mathbf{I}_N$ , sus elementos no son independientes.

### §Casos particulares

- Cuando  $\nu \rightarrow \infty$ , la distribución tiende a la normal estándar multivariada.
- Para  $\nu \rightarrow 2$ , las colas de la distribución se vuelven cada vez más pesadas.

## Componentes Principales

El análisis de Componentes Principales concierne a la explicación de la estructura de varianza-covarianza de un conjunto de variables a través de combinaciones lineales de dichas

variables. Sus objetivos esencialmente son la reducción de la dimensión y una interpretación más simplificada.

Considerando al conjunto de variables aleatorias  $X_1, \dots, X_p$ , las componentes principales buscan seleccionar un nuevo sistema coordinado rotando el sistema original con  $X_1, \dots, X_p$  como ejes, que representen las direcciones con máxima variabilidad y proporcionen una descripción más simple y parsimoniosa de la estructura de covarianza.

Sea  $\mathbf{X} = (X_1, \dots, X_p)'$  un vector aleatorio y denótese su matriz de covarianza por  $\Sigma$  cuyos eigenvalores son  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Dada  $Y_i = \mathbf{a}_i' \mathbf{X} = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$  una combinación lineal de  $\mathbf{X}$ , se tiene que

$$\text{Var}(Y_i) = \text{Var}(\mathbf{a}_i' \mathbf{X}) = \mathbf{a}_i' \Sigma \mathbf{a}_i \quad i = 1, \dots, p. \quad (\text{B.5})$$

De manera general para  $p$  cuyas combinaciones lineales

$$\text{Cov}(Y_i, Y_j) = \mathbf{a}_i' \Sigma \mathbf{a}_j \quad i, j = 1, \dots, p,$$

las componentes principales son las combinaciones lineales  $Y_1, \dots, Y_p$  que son no correlacionadas y cuya varianza en B.5 es lo más grande posible.

- La primera componente principal de  $\mathbf{X}$  es la combinación lineal  $Y_1 = \mathbf{a}_1' \mathbf{X}$  que maximiza  $\text{Var}(Y_1)$  sujeto a la condición  $\mathbf{a}_1' \mathbf{a}_1' = 1$ .
- La segunda componente principal de  $\mathbf{X}$  es la combinación lineal  $Y_2 = \mathbf{a}_2' \mathbf{X}$  que maximiza  $\text{Var}(Y_2)$  sujeto a la condición  $\mathbf{a}_2' \mathbf{a}_2' = 1$ . No correlacionada con  $Y_1$ .
- La  $i$ -ésima componente principal de  $\mathbf{X}$  es la combinación lineal  $Y_i = \mathbf{a}_i' \mathbf{X}$  que maximiza  $\text{Var}(Y_i)$  sujeto a la condición  $\mathbf{a}_i' \mathbf{a}_i' = 1$ . No correlacionada con  $Y_j$  para  $j \neq i$ .

Dado que la matriz de covarianza  $\Sigma$  es no negativa definida, puede representarse a través de su descomposición espectral. Sean las parejas de eigenvalores y eigenvectores  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$  tales que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ .

La  $i$ -ésima componente principal de  $\mathbf{X}$  es  $Y_i = \mathbf{e}_i' \mathbf{X} = e_{i1}X_1 + \dots + e_{ip}X_p \quad i = 1, \dots, p$  donde

$$\begin{aligned} \text{Var}(Y_i) &= \mathbf{e}_i' \Sigma \mathbf{e}_i = \lambda_i \quad i = 1, \dots, p, \\ \text{Cov}(Y_i, Y_j) &= \mathbf{e}_i' \Sigma \mathbf{e}_j = 0, \quad i \neq j. \end{aligned}$$

Si algunos eigenvalores  $\lambda_i$  son iguales, las elecciones de los correspondientes eigenvectores no son únicas.

Obsérvese que las componentes otorgan la misma varianza que las variables originales

$$\sum_{i=1}^p \text{Var}(X_i) = \text{tr}(\boldsymbol{\Sigma}) = \sum_{i=1}^p \lambda_i = \sum_{i=1}^p \text{Var}(Y_i).$$

Así, la proporción de varianza total explicada por la  $i$ -ésima componente principal es

$$\frac{\text{Var}Y_i}{\sum_{i=1}^p \text{Var}(Y_i)} = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i}.$$

Cuando la matriz de correlación  $\rho$  de  $\mathbf{X}$  es utilizada para determinar las componentes principales, dado que  $\text{tr}(\rho) = p$ , la proporción total de varianza explicada por la  $i$ -ésima componente será  $\lambda_i/p$ .

## Modelos de Factores

Este análisis busca combinaciones lineales de las variables llamadas *factores*, que representan cantidades subyacentes fundamentales expresadas en las variables observadas. Su idea es que un número pequeño de factores pueden explicar varias mediciones en un conjunto observacional. Los factores pueden pensarse como variables latentes no observadas que determinan patrones en las observaciones y deberán inferirse a través de los datos.

Sea  $\mathbf{X} = (X_1, \dots, X_p)'$  un vector aleatorio  $p$ -dimensional con media  $\boldsymbol{\mu}$  y matriz de covarianza  $\boldsymbol{\Sigma}$ .  $\mathbf{X}$  sigue un modelo de factores si

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon}, \quad (\text{B.6})$$

donde  $\mathbf{F} = (F_1, \dots, F_m)'$  es un vector de factores comunes con  $m \leq p$ ,  $\mathbf{L} = [l_{ij}]_{p \times m}$  es una matriz de cargas de los factores ( $l_{ij}$  es la carga de la  $i$ -ésima variable en el  $j$ -ésimo factor común) y  $\boldsymbol{\epsilon}$  un vector de errores específicos.

Una característica clave del modelo es que ninguno de los elementos a excepción de  $\mathbf{X}$  es observable.

El modelo B.6 se dice ortogonal si

1.  $\mathbb{E}[\mathbf{F}] = \mathbf{0}$  y  $\text{Cov}(\mathbf{F}) = \mathbf{I}_m$ , con  $\mathbf{I}_m$  la matriz identidad.
2.  $\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}$  y  $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} = \text{diag}\{\psi_1, \dots, \psi_p\}$ .
3.  $\mathbf{F}$  y  $\boldsymbol{\epsilon}$  son independientes i.e.  $\text{Cov}(\mathbf{F}, \boldsymbol{\epsilon}) = \mathbb{E}[\mathbf{F}\boldsymbol{\epsilon}'] = \mathbf{0}$ .

Dadas las condiciones anteriores se verifica que

$$\begin{aligned}\text{Cov}(\mathbf{X}) &= \mathbb{E}(\mathbf{LF} + \boldsymbol{\epsilon})(\mathbf{LF} + \boldsymbol{\epsilon})' \\ &= \mathbf{L}\mathbb{E}[\mathbf{FF}']\mathbf{L}' + \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']\mathbf{L}' + \mathbf{L}\mathbb{E}[\mathbf{F}\boldsymbol{\epsilon}'] + \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] \\ &= \mathbf{LL}' + \boldsymbol{\Psi}.\end{aligned}$$

Por lo tanto

$$\begin{aligned}\text{Var}(X_i) &= l_{i1}^2 + \cdots + l_{im}^2 + \psi_i, \\ \text{Cov}(X_i, X_j) &= l_{i1}l_{j1} + \cdots + l_{im}l_{jm},\end{aligned}$$

la primera igualdad nos indica que la varianza puede verse como

$$\sigma_{ii} = h_i^2 + \psi_i, \quad \text{con } h_i^2 = l_{i1}^2 + \cdots + l_{im}^2.$$

La expresión  $h_i^2$  es llamada *comunalidad* e indica la porción de varianza contribuida por los  $m$  factores, mientras que  $\psi_i$  es la *varianza específica*. Además se prueba que  $\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}'] = \mathbf{L}\mathbb{E}[\mathbf{FF}'] + \mathbb{E}[\boldsymbol{\epsilon}\mathbf{F}'] = \mathbf{L}$ , entonces se tiene que  $\text{Cov}(X_i, F_j) = l_{ij}$ .

Para  $m > 1$ , sea  $\mathbf{T}$  alguna matriz  $m \times m$  ortogonal i.e.  $\mathbf{TT}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$ . Entonces

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LTT}'\mathbf{F} + \boldsymbol{\epsilon} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\epsilon}, \quad \text{donde } \mathbf{L}^* = \mathbf{LT} \text{ y } \mathbf{F}^* = \mathbf{T}'\mathbf{F},$$

es también un modelo de factores ortogonal, pues se cumplen

$$\mathbb{E}[\mathbf{F}^*] = \mathbf{0} \text{ y } \text{Cov}(\mathbf{F}^*) = \mathbf{T}'\text{Cov}(\mathbf{F})(\mathbf{T}) = \mathbf{T}'(\mathbf{T}) = \mathbf{I}.$$

Esta falta de identificabilidad permite realizar elecciones de  $\mathbf{T}$  conocidas como *rotaciones*, a fin de encontrar factores comunes con interpretaciones adecuadas de acuerdo al contexto. Así,  $\mathbf{F}^* = \mathbf{T}'\mathbf{F}$  es una rotación en el espacio  $m$ -dimensional, y por ejemplo en algunas aplicaciones donde se requiera que los factores sean correlacionados, podrán emplearse rotaciones (oblicuas) para encontrarlos.

Los modelos de factores pueden estimarse de dos maneras:

1. La primera usa el análisis de componentes principales, pero no requiere el supuesto de normalidad ni la preespecificación del número de factores comunes. Su ventaja es que puede aplicarse a ambas, la matriz de covarianza o matriz de correlación; sin embargo, su solución será generalmente una aproximación.
2. El segundo método es la máxima verosimilitud que usa una densidad normal y requiere especificar el número factores comunes que se utilizarán.

## Apéndice C

# Tablas y Gráficas Adicionales en S-PLUS

### C.1. Series Univariadas

§Estadísticas Básicas

```
> summaryStats(dolar)
```

Sample Quantiles:

min	1Q	median	3Q	max
-2.313	-0.27	-0.02727	0.2409	2.153

Sample Moments:

mean	std	skewness	kurtosis
0.005168	0.475	0.4413	5.093

```
> summaryStats(euro)
```

Sample Quantiles:

min	1Q	median	3Q	max
-3.934	-0.3959	0.008909	0.4313	3.377

Sample Moments:

mean	std	skewness	kurtosis
0.02978	0.7175	0.1288	4.529

Number of Observations: 1514

## §Pruebas de Correlación y Normalidad

```
> autocorTest(dolar, lag = 12)
Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation
```

Test Statistics:

```
Test Stat 17.4317
p.value 0.1341
```

Dist. under Null: chi-square with 12 degrees of freedom

```
> normalTest(dolar, method = "jb")
Test for Normality: Jarque-Bera
Null Hypothesis: data is normally distributed
```

Test Statistics:

```
Test Stat 325.4873
p.value 0.0000
```

Dist. under Null: chi-square with 2 degrees of freedom

```
> autocorTest(euro, lag = 12)
Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation
```

Test Statistics:

```
Test Stat 5.6716
p.value 0.9317
```

Dist. under Null: chi-square with 12 degrees of freedom

```
> normalTest(euro, method = "jb")
Test for Normality: Jarque-Bera
Null Hypothesis: data is normally distributed
```

Test Statistics:

```
Test Stat 151.7447
  p.value  0.0000
```

Dist. under Null: chi-square with 2 degrees of freedom

### C.1.1. Volatilidad

§Series al Cuadrado y Efectos ARCH

```
> autocorTest((dolar)^2, lag.n = 12)
Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation
```

Test Statistics:

```
Test Stat 199.9338
  p.value  0.0000
```

Dist. under Null: chi-square with 12 degrees of freedom  
Total Observ.: 1514

```
> archTest(dolar, lag = 12)
Test for ARCH Effects: LM Test
Null Hypothesis: no ARCH effects
```

Test Statistics:

```
Test Stat 120.7415
  p.value  0.0000
```

Dist. under Null: chi-square with 12 degrees of freedom

```
> autocorTest((euro)^2, lag.n = 12)
Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation
```

Test Statistics:

```
Test Stat 155.5931
p.value 0.0000
```

```
Dist. under Null: chi-square with 12 degrees of freedom
Total Observ.: 1514
```

```
> archTest(euro, lag = 12)
Test for ARCH Effects: LM Test
Null Hypothesis: no ARCH effects
```

```
Test Statistics:
```

```
Test Stat 96.2235
p.value 0.0000
```

```
Dist. under Null: chi-square with 12 degrees of freedom
```

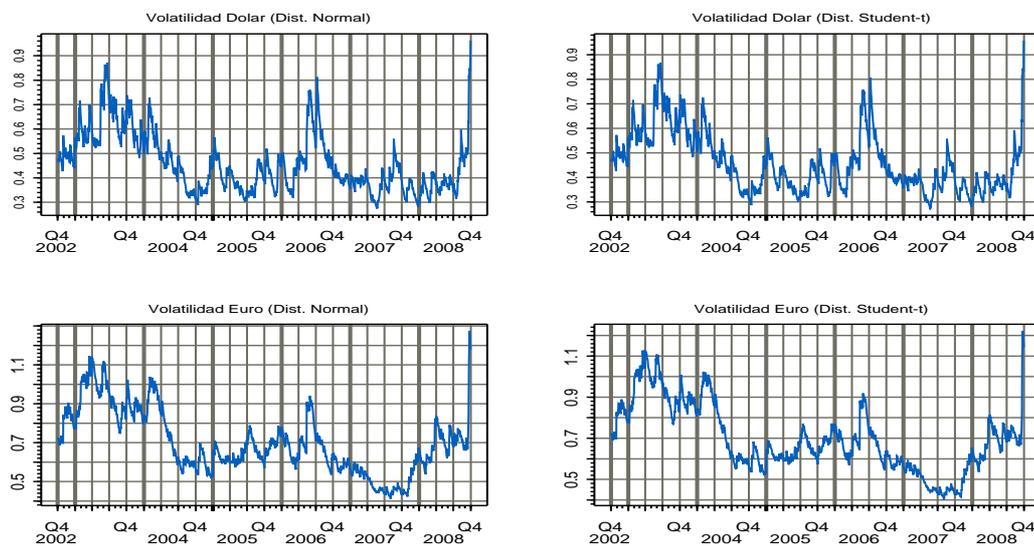


Figura C.1: Volatilidad Series Univariadas

## C.2. Modelos GARCH Bivariados

### C.2.1. DVEC

#### DVEC Student-t

```
> dolar.euro.dvec.t = mgarch(dolar.euro ~ 1, ~ dvec(1, 1), cond.dist = "t")
> summary(dolar.euro.dvec.t)
```

```
Conditional Distribution: t
with estimated parameter 10.18084 and standard error 1.703558
```

```
-----
Estimated Coefficients:
```

```
-----
              Value Std.Error   t value   Pr(>|t|)
      C(1) -0.0165420 0.0102400   -1.6154 1.064e-001
      C(2)  0.0125500 0.0160408    0.7824 4.341e-001
      A(1, 1) 0.0030227 0.0012103    2.4975 1.261e-002
      A(2, 1) 0.0008698 0.0004878    1.7833 7.474e-002
      A(2, 2) 0.0029038 0.0017149    1.6933 9.061e-002
      ARCH(1; 1, 1) 0.0580941 0.0109778    5.2920 1.388e-007
      ARCH(1; 2, 1) 0.0194046 0.0061064    3.1777 1.514e-003
      ARCH(1; 2, 2) 0.0323120 0.0081214    3.9786 7.261e-005
      GARCH(1; 1, 1) 0.9286913 0.0134413   69.0925 0.000e+000
      GARCH(1; 2, 1) 0.9708619 0.0090279  107.5405 0.000e+000
      GARCH(1; 2, 2) 0.9625502 0.0096873   99.3624 0.000e+000
-----
```

```
AIC(12) = 4731.81
```

```
BIC(12) = 4795.68
```

```
Normality Test:
```

```
-----
              Jarque-Bera   P-value Shapiro-Wilk   P-value
dolar           98.63 0.00000000    0.9848 1.522e-011
euro            19.87 0.00004839    0.9964 1.355e-003
-----
```

```
Ljung-Box test for standardized residuals:
```

```
-----
              Statistic P-value Chi^2-d.f.
dolar           12.99 0.36983           12
-----
```

```
euro      22.53 0.03195      12
```

Ljung-Box test for squared standardized residuals:

```
-----
Statistic P-value Chi^2-d.f.
dolar      6.48 0.89000      12
euro      20.14 0.06447      12
```

Lagrange multiplier test:

```
-----
TR^2 P-value F-stat P-value
dolar  6.582 0.88395  0.601  0.9240
euro  20.895 0.05194  1.926  0.1084
```

### ACF of Squared Std. Residuals

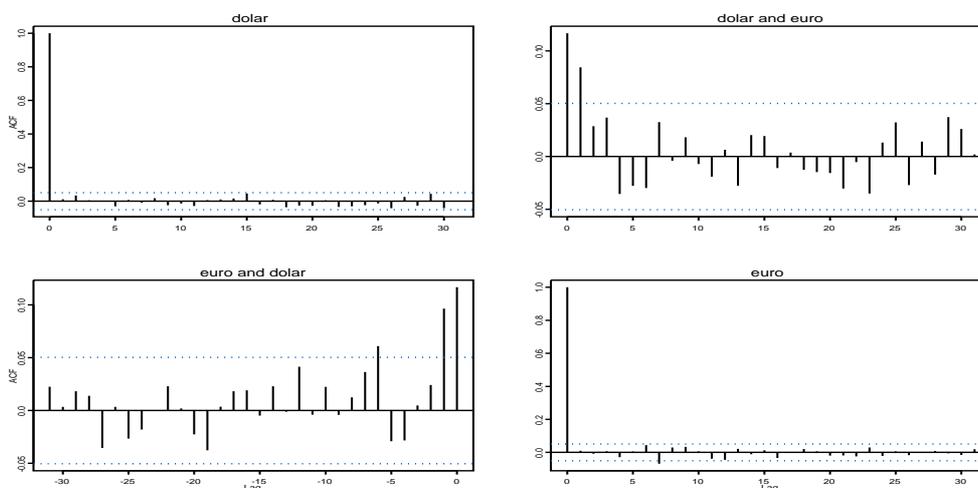


Figura C.2: ACF DVEC (Student-t)

### §Comparación DVEC (Student-t y Normal)

```
>compare.mgarch(dolar.euro.dvec.t, dolar.euro.dvec)
      dolar.euro.dvec.t dolar.euro.dvec
AIC           4732           4790
BIC           4796           4849
```

Likelihood                      -2354                      -2384

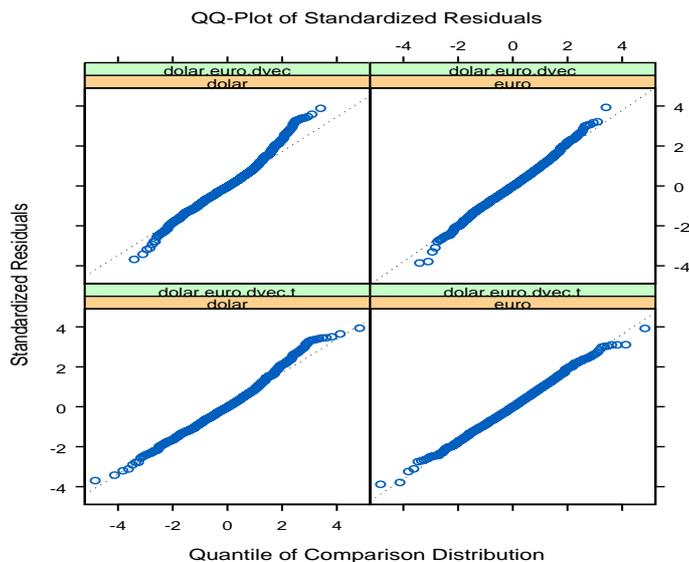


Figura C.3: QQ Plot DVEC (Normal y Student-t)

### C.2.2. BEKK

#### BEKK Student-t

```
> dolar.euro.bekk.t = mgarch(dolar.euro ~ 1, ~ bekk(1, 1), cond.dist = "t")
> summary(dolar.euro.bekk.t)
```

Conditional Distribution: t  
with estimated parameter 9.847082 and standard error 1.612853

-----  
Estimated Coefficients:

	Value	Std.Error	t value	Pr(> t )
C(1)	-0.015673	0.010238	-1.5310	1.260e-001
C(2)	0.013211	0.016082	0.8214	4.115e-001
A(1, 1)	0.052264	0.010864	4.8107	1.655e-006
A(2, 1)	-0.011122	0.019830	-0.5609	5.750e-001

A(2, 2)	0.044836	0.016817	2.6662	7.753e-003
ARCH(1; 1, 1)	0.219967	0.026388	8.3359	2.220e-016
ARCH(1; 2, 1)	-0.061854	0.037827	-1.6352	1.022e-001
ARCH(1; 1, 2)	0.008130	0.018159	0.4477	6.544e-001
ARCH(1; 2, 2)	0.167789	0.021281	7.8846	5.995e-015
GARCH(1; 1, 1)	0.966085	0.007877	122.6436	0.000e+000
GARCH(1; 2, 1)	0.008859	0.009924	0.8927	3.722e-001
GARCH(1; 1, 2)	0.004769	0.004801	0.9934	3.207e-001
GARCH(1; 2, 2)	0.983764	0.005085	193.4465	0.000e+000

-----

AIC(14) = 4744.355

BIC(14) = 4818.87

Normality Test:

-----

	Jarque-Bera	P-value	Shapiro-Wilk	P-value
dolar	96.09	0.000e+000	0.9846	1.184e-011
euro	31.06	1.798e-007	0.9949	5.052e-005

Ljung-Box test for standardized residuals:

-----

	Statistic	P-value	Chi <sup>2</sup> -d.f.
dolar	12.91	0.37575	12
euro	22.39	0.03335	12

Ljung-Box test for squared standardized residuals:

-----

	Statistic	P-value	Chi <sup>2</sup> -d.f.
dolar	7.342	0.83420	12
euro	23.508	0.02371	12

Lagrange multiplier test:

-----

	TR <sup>2</sup>	P-value	F-stat	P-value
dolar	7.502	0.82273	0.6854	0.85928
euro	25.183	0.01398	2.3284	0.05651

§Comparación BEKK (Student-t y Normal)

> compare.mgarch(dolar.euro.bekk.t, dolar.euro.bekk)

## ACF of Squared Std. Residuals

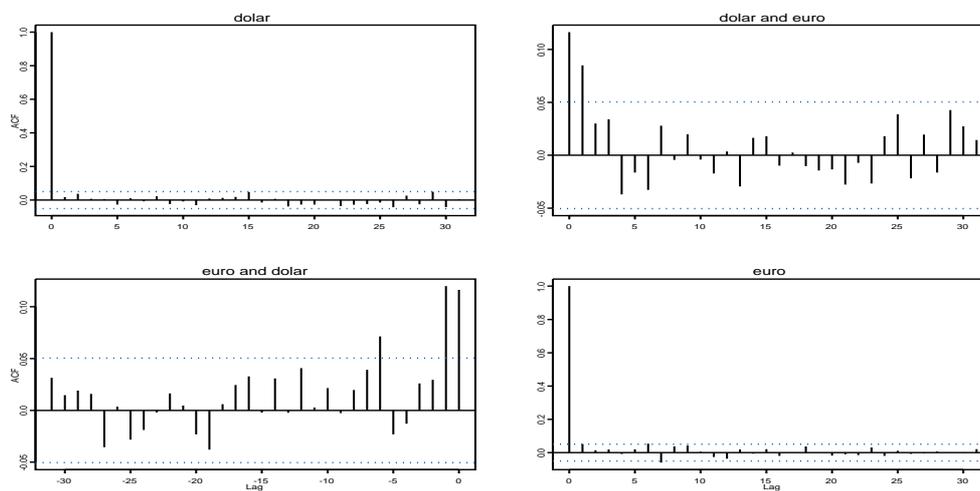


Figura C.4: ACF BEKK (Student-t)

	dolar.euro.bekk.t	dolar.euro.bekk
AIC	4744	4807
BIC	4819	4877
Likelihood	-2358	-2391

## C.2.3. O-GARCH

## Componentes Principales

```
> dolar.euro.pc=princomp(dolar.euro)
```

```
Importance of components:
```

```

                                Comp.1   Comp.2
Standard deviation 0.7456471 0.4289772
Proportion of Variance 0.7513261 0.2486739
Cumulative Proportion 0.7513261 1.0000000
```

```
> dolar.euro.pc$loadings
  Comp.1 Comp.2
```

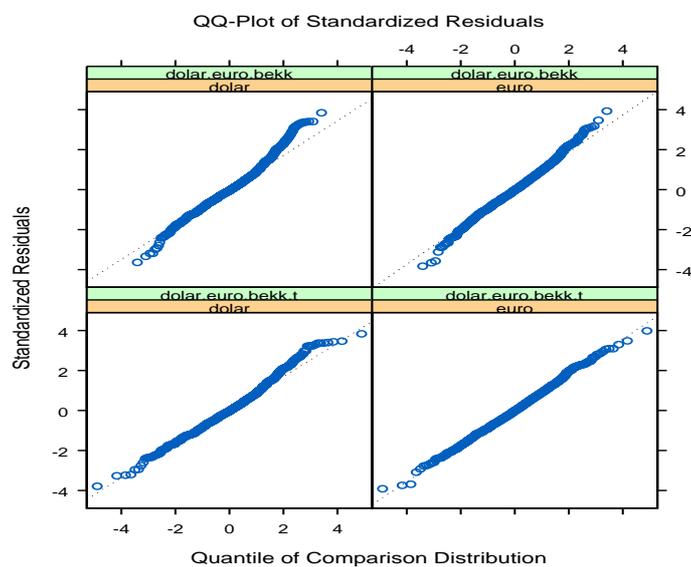


Figura C.5: QQ Plot BEKK (Normal y Student-t)

```
dolar 0.334 0.943
euro 0.943 -0.334
```

```
> dolar.euro.pc$correlations
      Comp.1   Comp.2
dolar 0.5243350 0.8515121
euro 0.9798567 -0.1997020
```

### OGARCH Student-t

```
> ogarch.t = mgarch(dolar.euro ~ 1, ~ prcomp.garch(1, 1), trace = F, cond.dist="t")
> summary(ogarch.t)
```

```
Conditional Distribution: t
with estimated parameter 10.48786 and standard error 1.774811
```

```
-----
Estimated Coefficients:
```

```
-----
Value Std.Error t value Pr(>|t|)
```

```

          C(1)  0.013473  0.016640   0.8097  0.41824628
          C(2) -0.016071  0.009499  -1.6917  0.09090104
    A(1, 1)  0.002461  0.001717   1.4337  0.15185561
    A(2, 2)  0.004216  0.001822   2.3136  0.02082508
  ARCH(1; 1, 1)  0.034809  0.008424   4.1321  0.00003791
  ARCH(1; 2, 2)  0.075894  0.017277   4.3927  0.00001197
  GARCH(1; 1, 1)  0.961445  0.009655  99.5768  0.00000000
  GARCH(1; 2, 2)  0.903215  0.022798  39.6187  0.00000000

```

Eigenvectors: (orthonormal transform matrix):

```

-----
          dolar    euro
dolar  0.3339  0.9426
euro   0.9426 -0.3339

```

Eigenvalues:

```

-----
[1]  0.5564  0.1841
-----

```

AIC(9) = 4745.726

BIC(9) = 4793.629

Normality Test:

```

-----
          Jarque-Bera    P-value Shapiro-Wilk    P-value
PRCOMP1          85.53  0.000e+000          0.9915  1.184e-007
PRCOMP2          31.61  1.369e-007          0.9954  1.340e-004

```

Ljung-Box test for standardized residuals:

```

-----
          Statistic    P-value Chi^2-d.f.
PRCOMP1          6.986  8.585e-001          12
PRCOMP2         53.319  3.614e-007          12

```

Ljung-Box test for squared standardized residuals:

```

-----
          Statistic P-value Chi^2-d.f.
PRCOMP1          17.011  0.1492          12
PRCOMP2           8.954  0.7069          12

```

Lagrange multiplier test:

```
-----
          TR^2 P-value F-stat P-value
PRCOMP1 17.209 0.1419 1.5826 0.1973
PRCOMP2  9.512 0.6587 0.8702 0.6815
```

### ACF of Squared Std. Residuals

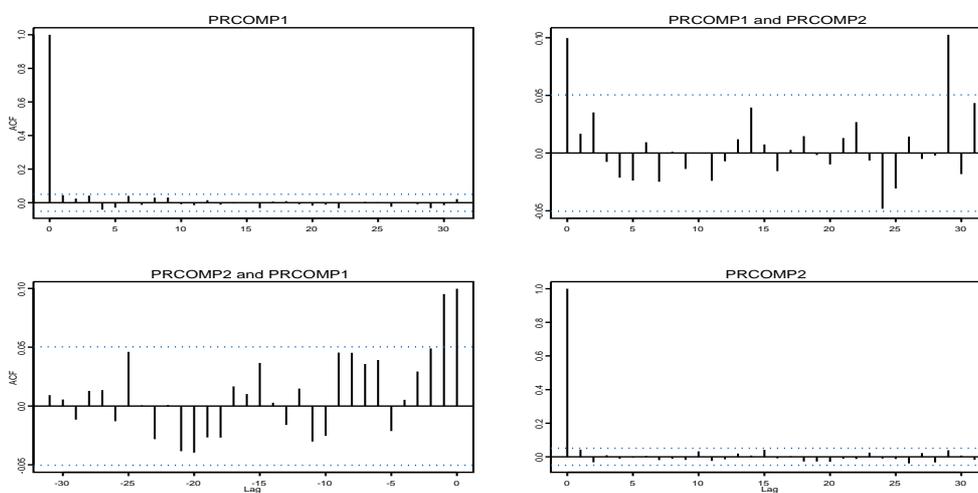


Figura C.6: ACF OGARCH (Student-t)

#### §Comparación OGARCH (Student-t y Normal)

```
> compare.mgarch(ogarch.t, ogarch)
          ogarch.t ogarch
AIC         4746  4802
BIC         4794  4845
Likelihood  -2364 -2393
```

### C.2.4. CCC

#### CCC: Un elemento

##### §Student-t

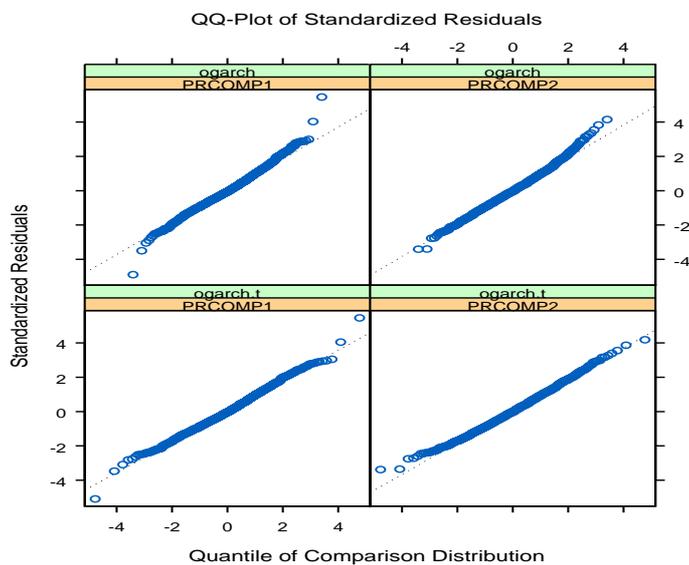


Figura C.7: QQ-Plot OGARCH (Normal y Student-t)

```
>dolar.euro.ccc.t=mgarch(dolar.euro~1,~ccc(1, 1),trace=F,cond.dist="t")
>summary(dolar.euro.ccc.t)
```

Conditional Distribution: t  
with estimated parameter 10.12064 and standard error 1.674487

-----  
Estimated Coefficients:

	Value	Std.Error	t value	Pr(> t )
C(1)	-0.016398	0.010358	-1.5831	1.136e-001
C(2)	0.015318	0.016091	0.9519	3.413e-001
A(1, 1)	0.002605	0.001147	2.2705	2.332e-002
A(2, 2)	0.002847	0.001734	1.6415	1.009e-001
ARCH(1; 1, 1)	0.054899	0.010935	5.0204	5.768e-007
ARCH(1; 2, 2)	0.033042	0.008566	3.8574	1.194e-004
GARCH(1; 1, 1)	0.934091	0.013329	70.0791	0.000e+000
GARCH(1; 2, 2)	0.962225	0.010008	96.1474	0.000e+000

-----  
Estimated Conditional Constant Correlation Matrix:

```

      dolar  euro
dolar 1.0000 0.3283
euro 0.3283 1.0000

```

Standard Errors:

```

      dolar  euro
dolar      NA 0.02588
euro 0.02588      NA

```

```

-----
AIC(9) = 4744.81
BIC(9) = 4792.712

```

Normality Test:

```

-----
      Jarque-Bera  P-value Shapiro-Wilk  P-value
dolar      98.18 0.00000000      0.9848 1.576e-011
euro      21.73 0.00001909      0.9964 1.165e-003

```

Ljung-Box test for standardized residuals:

```

-----
      Statistic P-value Chi^2-d.f.
dolar      13.03 0.36697      12
euro      22.90 0.02858      12

```

Ljung-Box test for squared standardized residuals:

```

-----
      Statistic P-value Chi^2-d.f.
dolar      6.61 0.8823      12
euro      17.28 0.1395      12

```

Lagrange multiplier test:

```

-----
      TR^2 P-value F-stat P-value
dolar 6.719 0.8756 0.6136 0.9155
euro 17.970 0.1166 1.6534 0.1739

```

§Comparación CCC Un elemento (Student-t y Normal)

```

> compare.mgarch(dolar.euro.ccc.t, dolar.euro.ccc)
      dolar.euro.ccc.t dolar.euro.ccc

```

## ACF of Squared Std. Residuals

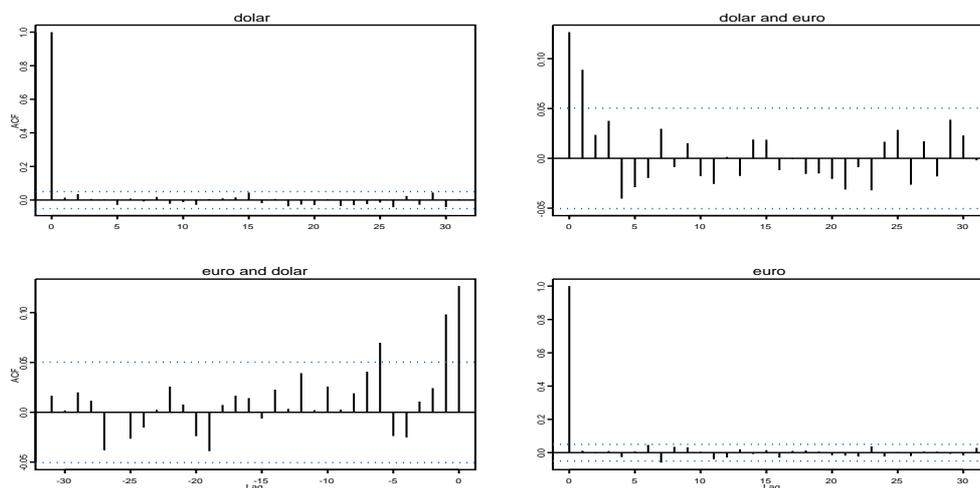


Figura C.8: ACF CCC Un elemento (Student-t)

AIC	4745	4806
BIC	4793	4849
Likelihood	-2363	-2395

## CCC: Dos Elementos

§Student-t

```
>dolar.euro.ccc2.t=mgarch(dolar.euro~1,~ccc.two.comp(1,1),trace=F,
  cond.dist="t")
>summary(dolar.euro.ccc2.t)
```

```
Conditional Distribution: t
  with estimated parameter 10.50048 and standard error 1.879685
```

-----  
 Estimated Coefficients:  
 -----

	Value	Std.Error	t value	Pr(> t )
C(1)	-0.021243	0.010297	-2.0629	0.0392889

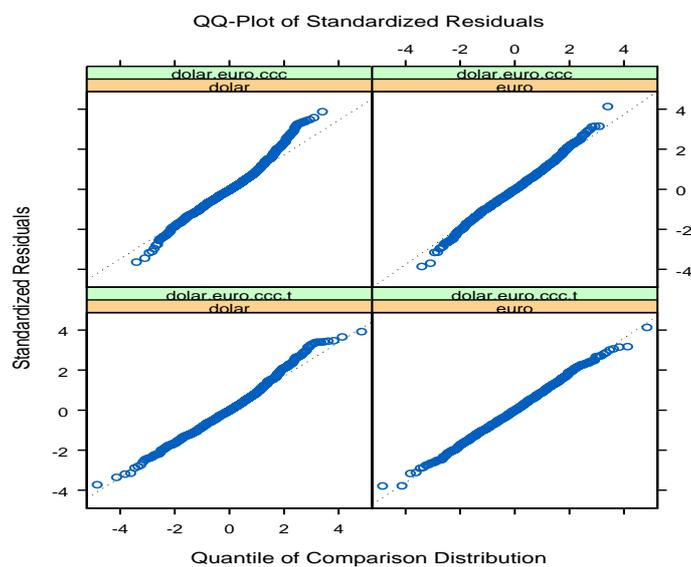


Figura C.9: QQ-Plot CCC Un elemento (Normal y Student-t)

C(2)	0.014276	0.016089	0.8873	0.3750542
A(1, 1)	0.000758	0.001628	0.4656	0.6415389
A(2, 2)	0.007943	0.006707	1.1844	0.2364580
ARCH(1; 1, 1)	0.036761	0.059347	0.6194	0.5357288
ARCH(1; 2, 2)	0.041294	0.011533	3.5805	0.0003539
ARCH(2; 1, 1)	0.056123	0.057685	0.9729	0.3307455
ARCH(2; 2, 2)	0.027910	0.034844	0.8010	0.4232585
GARCH(1; 1, 1)	0.952366	0.040014	23.8008	0.0000000
GARCH(1; 2, 2)	0.959642	0.011470	83.6659	0.0000000
GARCH(2; 1, 1)	0.820082	0.230405	3.5593	0.0003834
GARCH(2; 2, 2)	-0.443310	0.893066	-0.4964	0.6196909

Estimated Conditional Constant Correlation Matrix:

```
-----
      dolar  euro
dolar 1.0000 0.3298
euro  0.3298 1.0000
```

Standard Errors:

```
      dolar  euro
```

```
dolar      NA 0.02611
euro 0.02611      NA
```

```
-----
AIC(13) = 4745.93
BIC(13) = 4815.122
```

Normality Test:

```
-----
          Jarque-Bera   P-value Shapiro-Wilk   P-value
dolar      77.08 0.00000000      0.9863 8.795e-011
euro       19.84 0.00004925      0.9964 1.333e-003
```

Ljung-Box test for standardized residuals:

```
-----
          Statistic P-value Chi^2-d.f.
dolar      13.10 0.36161      12
euro       23.59 0.02313      12
```

Ljung-Box test for squared standardized residuals:

```
-----
          Statistic P-value Chi^2-d.f.
dolar      5.719 0.9296      12
euro       16.954 0.1513      12
```

Lagrange multiplier test:

```
-----
          TR^2 P-value F-stat P-value
dolar  5.772 0.9271 0.5267 0.9645
euro  17.471 0.1327 1.6070 0.1889
```

§Comparación CCC Dos elementos (Student-t y Normal)

```
> compare.mgarch(dolar.euro.ccc2.t, dolar.euro.ccc2)
          dolar.euro.ccc2.t dolar.euro.ccc2
          AIC              -23143          -23088
          BIC              -23074          -23025
Likelihood              11584           11556
```

ACF of Squared Std. Residuals

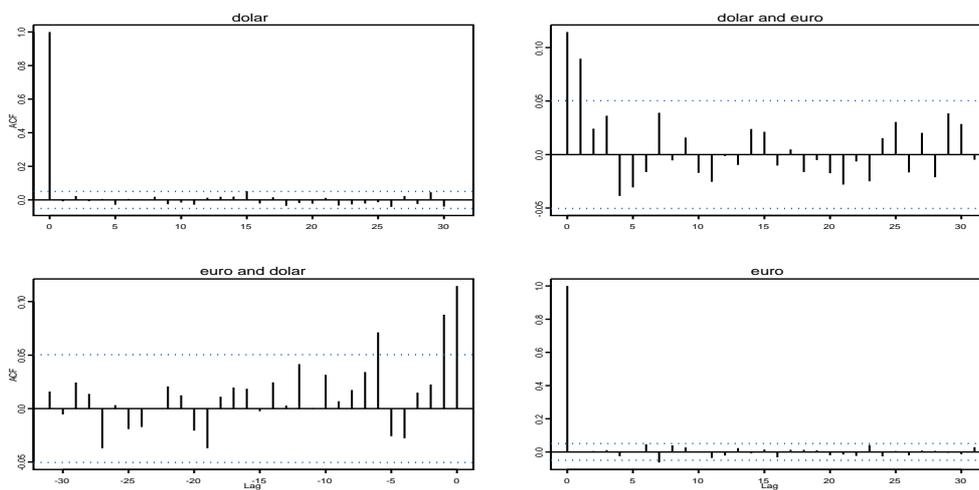


Figura C.10: ACF CCC (Student-t Dos elementos)

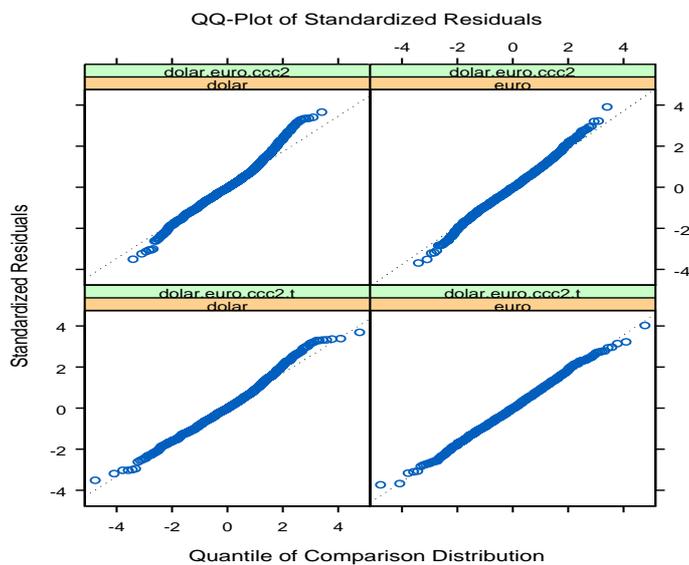


Figura C.11: QQ-Plot (CCC Dos elementos)

# Bibliografía

- [1] Alexander C. [2001] Orthogonal GARCH. *Mastering Risk Volume 2*. Pearson, 21-38.
- [2] Bauwens L.; Laurent S.; Rombouts J. [2006] Multivariate GARCH Models: a Survey. *Journal of Applied Econometrics*, **21**:79–109.
- [3] Black F. [1976] Studies in stock price volatility changes. Proceedings of the 1976 Business Meeting of the Business and Economic Section. *American Statistical Association*, pages 177–181.
- [4] Bollerslev T. [1986] Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**:307–327.
- [5] Bollerslev T.; Engle R.; Wooldridge J. [1988] A capital asset pricing model with time varying covariances. *Journal of Political Economy*, **96**:116–131.
- [6] Bollerslev T. [1990] Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, **72**:498–505.
- [7] Bollerslev T.; Chou R.; Kroner K. [1992] ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, **52**:5–59.
- [8] Bollerslev T.; Wooldridge J. [1992] Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, **11**:143–172.
- [9] Bollerslev T.; Engle R. [1993] Common persistence in conditional variances. *Econometrica*, **61**:167–186.
- [10] Bollerslev T.; Engle R.; Nelson D. [1994] ARCH models. In *Handbook of Econometrics*, IV:2959–3038.
- [11] Box G.; Pierce D. [1970] Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association*, **65**:1509–1526.

- [12] Brockwell P.; Davis R. [1991]. *Time Series: Theory and Methods*, 2nd Ed. Springer Verlag: New York.
- [13] Brockwell P.; Davis R. [1996]. *Introduction to Time Series and Forecasting*. Springer: New York.
- [14] Campbell J.; Lo A.; MacKinlay A. [1997]. *The Econometrics of Financial Markets*. Princeton University Press: New Jersey.
- [15] Chatfield C. [2004]. *The Analysis of Time Series: An Introduction*, 6th Ed. Chapman & Hall.
- [16] Davidson J. [2004] Moments and memory properties of linear conditional heteroscedasticity models. *Journal of Business and Economic Statistics*, **22**:16–29.
- [17] Diebold F.; Nerlove M. [1989] The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics*, **4**:1–21.
- [18] Enders W. [2004]. *Applied econometric time series*, 2nd Ed. Wiley: New Jersey.
- [19] Engle R. [1982] Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**:987–1007.
- [20] Engle R.; Lilien D.; Robins R. [1987] Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, **55**:391–407.
- [21] Engle R.; Ng V.; Rothschild M. [1990] Asset pricing with a factor ARCH covariance structure: empirical estimates for Treasury Bills. *Journal of Econometrics*, **45**:213–238.
- [22] Engle R.; Kroner K. [1995] Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, **11**:122–150.
- [23] Engle R. [2002] Dynamic Conditional Correlation: a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, **20**:339–350.
- [24] Gouriéroux C. [1997]. *ARCH models and financial applications*. Springer Verlag: New York.
- [25] Lanne M.; Saikkonen P. [2007] A Multivariate Generalized Orthogonal Factor GARCH Model. *Journal of Business and Economic Statistics*, **25**:61–75.
- [26] Lin W. [1992] Alternative estimators for Factor GARCH models A Monte Carlo comparison. *Journal of Applied Econometrics*, **7**:259–279.
- [27] Ljung G.; Box G. [1978] On a measure of lack of fit in time series models. *Biometrika*, **65**:297–303.

- 
- [28] Mandelbrot B. [1963] The variation of certain speculative prices. *Journal of Business*, **36**:394–419.
- [29] Mood A.; Graybill F.; Boes D. [1974]. *Introduction to the Theory of Statistics*. McGraw-Hill.
- [30] Nelson D. [1990] Stationarity and persistence in the GARCH model. *Econometric Theory*, **6**:318–334.
- [31] Nelson D. [1991] Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, **59**:347–370.
- [32] Palandri A. [2004] Multivariate GARCH Models: Inference and Evaluation. Working paper, Department of Economics, Duke University.
- [33] Peña D.; Tiao G; Tsay R. [2001]. *A course in time series analysis*. John Wiley & Sons: New York.
- [34] Quenouille M. [1949] Approximate tests of correlation in time series. *Journal of the Royal Statistical Society*, **11**:68–84.
- [35] Ross S. [1976] The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13**:341–360.
- [36] Silvennoinen A.; Teräsvirta T. [2008] Multivariate GARCH models. To appear in *Handbook of Financial Time Series*. New York Springer.
- [37] Tsay R. [2005]. *Analysis of financial time series*, 2nd Ed. John Wiley: New York.
- [38] Tse Y.; Tsui A. [2002] A Multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics*, **20**:351–362.
- [39] van der Weide R. [2002] Generalized orthogonal GARCH-A multivariate GARCH model. *Journal of Applied Econometrics*, **17**:549–564.
- [40] Vrontos I.; Dellaportas P.; Politis D. [2003] A full-factor multivariate GARCH model. *Econometrics Journal*, **6**:312–334.
- [41] Zivot E.; Wang J. [2002]. *Modelling financial time series with S-Plus*, 1st. Ed. Springer Verlag.