



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**ESTIMACIÓN DE MEDIDAS DE RIESGO DE
MERCADO EN SERIES FINANCIERAS MEXICANAS:
UN ESTUDIO COMPARATIVO ENTRE PROCESOS
GARCH CON INNOVACIONES GAUSSIANAS, T-
STUDENT Y PARETO GENERALIZADA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A:

ALBERTO SAAVEDRA ESPINOSA



**DIRECTORA DE TESIS:
DRA. MARÍA ASUNCIÓN BEGOÑA FERNÁNDEZ
FERNÁNDEZ**

2016

Ciudad Universitaria, CDMX



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.

Hoja de datos del jurado

1. Datos del alumno:
 - Apellido paterno: Saavedra
 - Apellido materno: Espinosa
 - Nombre: Alberto
 - Teléfono: 5694-7399
 - Universidad: Universidad Nacional Autónoma de México
 - Facultad: Facultad de Ciencias
 - Carrera: Actuaría
 - Número de Cuenta: 308100550
2. Datos de la tutora:
 - Grado: Dra.
 - Nombres: María Asunción Begoña
 - Apellido paterno: Fernández
 - Apellido materno: Fernández
3. Datos del sinodal 1
 - Grado: Dra.
 - Nombre: Ana
 - Apellido paterno: Meda
 - Apellido materno: Guardiola
4. Datos del sinodal 2
 - Grado: Dr.
 - Nombre: Yuri
 - Apellido paterno: Salazar
 - Apellido materno: Flores
5. Datos del sinodal 3
 - Grado: M en F.
 - Nombres: María Antonieta
 - Apellido paterno: Campa
 - Apellido materno: Rojas
6. Datos del sinodal 4
 - Grado: Dr.
 - Nombre: Serafin
 - Apellido paterno: Martínez
 - Apellido materno: Jaramillo
7. Datos del trabajo escrito
 - Título: Estimación de medidas de riesgo de mercado en series financieras mexicanas: un estudio comparativo entre procesos GARCH con innovaciones Gaussianas, t-Student y Pareto Generalizada.
 - Número de páginas: 280 p.
 - Año: 2016

Índice general

Introducción	5
1. Conceptos básicos	13
1.1. Factores de riesgo y distribuciones de pérdida	14
1.1.1. Definiciones generales	14
1.1.2. Distribuciones de pérdida condicional y no condicional	15
1.2. Medidas de riesgo	16
1.2.1. Valor en Riesgo (<i>VaR</i>)	18
1.2.2. <i>Expected Shortfall</i> (ES)	23
1.2.3. Medidas de riesgo coherentes	27
1.3. Evaluando sistemas de cuantificación de riesgos: <i>backtesting</i>	33
2. Series de tiempo financieras	37
2.1. Bases para el análisis de series de tiempo	38
2.1.1. Definiciones básicas	38
2.1.2. Procesos ARMA	43
2.1.3. Metodología básica para el análisis estadístico de series de tiempo	46
2.2. Modelos de volatilidad estocástica	52
2.2.1. Procesos ARCH	52
2.2.2. Procesos GARCH	59
2.2.3. Ajustando modelos ARCH y GARCH a datos de origen financiero	63
2.3. Medición del riesgo vía modelos de volatilidad estocástica . . .	68
2.3.1. Pronósticos de volatilidad	69
2.3.2. Estimación de medidas de riesgo condicionales	71
2.3.3. <i>Backtesting</i>	72
3. Teoría de Valores Extremos	75
3.1. Máximos: propiedades de convergencia	76

3.1.1.	El Teorema de Fisher-Tippet y la Distribución Generalizada de Valores Extremos	77
3.1.2.	Dominio de Atracción del Máximo	83
3.2.	Modelos para excesos sobre un umbral	87
3.2.1.	El Teorema de Gnedenko-Balkema-Pickands-de Haan y la Distribución Pareto Generalizada	87
3.2.2.	El método POT	94
3.2.3.	Estimación de los parámetros de la Distribución Pareto Generalizada	102
4.	Potenciando el GARCH vía la Teoría de Valores Extremos	105
4.1.	Metodología	106
4.2.	El modelo híbrido GARCH-DPG	111
4.3.	<i>Backtesting</i>	117
5.	Aplicación al caso mexicano	121
5.1.	Tipo de Cambio FIX	124
5.1.1.	Análisis descriptivo	127
5.1.2.	Modelando la dinámica del FIX	132
5.1.3.	Modelando las innovaciones del FIX	149
5.1.4.	<i>Backtesting</i>	164
5.2.	Índice de Precios y Cotizaciones	179
5.2.1.	Análisis descriptivo	181
5.2.2.	Modelando la dinámica del IPC	185
5.2.3.	Modelando las innovaciones del IPC	201
5.2.4.	<i>Backtesting</i>	215
	Conclusiones	227
	Apéndices	233
A.	Teoría estadística complementaria	233
A.1.	Teorema del Límite Central	233
A.2.	Teorema de Glivenko-Cantelli	234
A.3.	P-Valor	235
A.4.	T-Valor	238
B.	Desarrollo de pruebas de hipótesis	239
B.1.	Pruebas Portmanteau	239
B.2.	Criterio de información de Akaike	241
B.3.	<i>Test</i> binomial	243
B.4.	Prueba <i>bootstrap</i>	245

<i>Índice general</i>	3
-----------------------	---

C. Códigos en R	247
------------------------	------------

C.1. Análisis exploratorios	247
---------------------------------------	-----

C.2. Modelando la dinámica de la volatilidad de una serie de tiempo	249
---	-----

C.3. Modelando las innovaciones de un proceso GARCH	254
---	-----

C.4. Estimación de medidas de riesgo	257
--	-----

C.5. <i>Backtests</i> del <i>VaR</i> y <i>Expected Shortfall</i> condicionales	264
--	-----

Bibliografía	275
---------------------	------------

Introducción

La sociedad moderna confía en el correcto funcionamiento del sistema financiero y tiene un interés colectivo en su estabilidad. El proceso regulatorio que culminó en los Acuerdos de Basilea para el mundo bancario, o las Directivas de Solvencia para el asegurador, ha estado fuertemente motivado por el *riesgo*¹ sistémico; es decir, el peligro de que problemas en una sola institución financiera perturben el adecuado funcionamiento de todo el sistema financiero. En consecuencia, la sociedad ve de forma positiva la administración del riesgo y confía a los reguladores la tarea de forjar un marco que salvaguarde sus intereses.

Actualmente, dicho marco regulatorio toma en consideración un importante conjunto de factores: pruebas de estrés, requerimientos de capital mínimo, procedimientos de supervisión regulatoria, disciplina de mercado, requerimientos de liquidez, entre otros.

Cada uno de los factores que comprenden el marco que regula la industria financiera cuenta con características particulares y retos importantes, de naturaleza tanto cuantitativa como cualitativa. Esta diversidad permite que distintas áreas del conocimiento puedan atacar los problemas de la industria financiera.

Desde el punto de vista de la probabilidad y la estadística, uno de los desafíos más importantes, e interesantes, es el de desarrollar metodologías que permitan *cuantificar* de la forma más precisa posible ciertos riesgos financieros. Dicho desafío va de la mano con uno de los factores del marco regulatorio de la industria financiera: la determinación de requerimientos de capital mínimo.

En efecto, la idea detrás de un requerimiento de capital es que la cantidad de dinero que se le requiere mantener a una institución por ley (capital re-

¹ La Real Academia Española define riesgo como “una contingencia o proximidad de un daño”. El diccionario Oxford en inglés como “peligro, una posibilidad de malas consecuencias, pérdida o exposición a la mala fortuna”.

gulatorio) sea lo más cercana posible al *capital económico* de dicha firma, el cual suele definirse como el nivel de capital que es requerido para cubrir las pérdidas de una institución con una cierta probabilidad o nivel de confianza.

La necesidad de contar con metodologías precisas para medir el riesgo de una institución financiera surge a partir de la idealización anterior: mientras más exactas sean las metodologías de cuantificación de riesgos, el capital regulatorio podrá ser más sensible a los riesgos de una firma y, por tanto, ser más parecido a su capital económico.

Ahora bien, es claro que una institución financiera puede estar sujeta a muchos tipos de riesgo, algunos intrínsecos de la industria y otros aplicables a prácticamente cualquier empresa. No obstante, las cargas de capital regulatorio suelen centrar su atención en la primera clase de riesgo: *los riesgos financieros*.²

Dentro de los riesgos financieros, las principales categorías de riesgo bien podrían ser las siguientes:

- *Riesgo de mercado*: el riesgo de un cambio en el valor de una posición financiera debido a cambios en el valor de sus componentes subyacentes, tales como precios de bonos o acciones, tipos de cambio, tasas de interés, etc.
- *Riesgo de crédito*: el riesgo de no recibir pagos de vuelta sobre inversiones excepcionales, tales como préstamos o bonos, debido al incumplimiento del deudor o emisor.
- *Riesgo operativo*: el riesgo de tener pérdidas como resultado de procesos internos fallidos o inadecuados, errores humanos, fallas de sistemas o por eventos externos.
- *Riesgo de liquidez*: el riesgo derivado de la falta de comerciabilidad de una inversión que no puede ser comprada o vendida suficientemente rápido como para evitar o minimizar una pérdida.

Dichas categorías de riesgo son tan importantes en la industria financiera que algunos acuerdos internacionales, como los del Comité de Basilea, han publicado recomendaciones generales para su manejo. Más aún, para las primeras tres categorías de riesgo el Comité de Basilea también plantea requerimientos

² En un contexto financiero, podríamos definir el riesgo como “cualquier evento o acción que pueda dificultar que una institución logre sus objetivos o ejecute sus estrategias” o, de forma alternativa, “la verosimilitud cuantificable de pérdidas o de rendimientos menores a los esperados”.

de capital específicos para cubrir las pérdidas ocasionadas por su materialización.

Por lo tanto, dada la importancia de contar con metodologías precisas para cuantificar riesgos, en este trabajo nos centraremos en el estudio de una técnica para medir el riesgo financiero mejor conocido en la industria bancaria: el riesgo de mercado.

Para cuantificar dicho riesgo nos auxiliaremos de las matemáticas; específicamente, de las ramas de la Teoría de la Probabilidad y la Estadística. En consecuencia, nuestra meta será estimar, a través de un modelo matemático sólido, cantidades que reflejen el riesgo de mercado que acarree algún activo financiero.

El enfoque a la medición del riesgo de mercado que tomaremos tiene dos componentes importantes. La primera de ellas es su objeto de estudio: pérdidas diarias originadas por movimientos en el valor un activo riesgoso. Entre otras cosas, esto quiere decir que la intención del modelo que utilizaremos será describir el comportamiento de los cambios diarios en el valor de algún activo con riesgo, a fin de predecir de forma precisa tales variaciones.

De esta manera, dada la información a analizar en este trabajo (series de rendimientos financieros), la clase de modelo que utilizaremos para describir su comportamiento dará un tratamiento especializado a una de sus componentes más importantes: la *volatilidad*.

De forma más específica, aprovecharemos el éxito probado de los modelos de volatilidad estocástica para describir series de rendimientos financieros; en particular, el de los modelos GARCH (*Generalized Autoregressive Conditional Heteroscedasticity*).³ En efecto, trabajos como [1] han mostrado que un modelo GARCH apropiadamente calibrado puede capturar de forma adecuada la dinámica de la volatilidad de una serie de rendimientos.

No obstante, el enfoque utilizado en [1] al trabajar con el GARCH permitió utilizar dichos modelos únicamente para describir (y predecir) pérdidas que ocurren con una probabilidad moderada, como $\alpha = 0.95$.⁴ En contraste, en

³ El GARCH es una generalización del modelo de volatilidad estocástica propuesto en 1982 por Robert Engle (premio Nobel de Economía en 2003), el proceso ARCH (*Autoregressive Conditional Heteroscedasticity*). El modelo fue propuesto en la revista *Econometrica*; específicamente, en un artículo titulado *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*.

⁴ Para el lector familiarizado con el GARCH, el enfoque utilizado en [1] consistió en modelar las innovaciones de dichos procesos mediante distribuciones Gaussianas y *t*-Student. Al lector que explora por primera vez esta clase de modelos, lo referimos al Capítulo 2 de

el presente trabajo estamos interesados tanto en pérdidas moderadas como en aquellas que por su magnitud y frecuencia podrían considerarse raras o *extremas*.

Es bajo la consideración anterior que surge la segunda componente de nuestro enfoque a la medición del riesgo de mercado, la cual busca tomar en consideración una realidad importante en el mundo de los riesgos financieros: *los extremos importan*.

En efecto, al cuantificar riesgos financieros, hoy en día es una necesidad tomar en consideración resultados inesperados, anormales o extremos en lugar de resultados esperados, normales o promedio, los cuales son el principal punto de atención de muchas aplicaciones clásicas de la Teoría de la Probabilidad y la Estadística a la cuantificación de riesgos.

El hecho anterior es consistente con el punto de vista regulatorio expresado por Alan Greenspan⁵ en la conferencia de investigación conjunta entre Bancos Centrales de 1995:

“From the point of view of the risk manager, inappropriate use of the normal distribution can lead to an understatement of risk, which must be balanced against the significant advantage of simplification. From the central bank’s corner, the consequences are even more serious because we often need to concentrate on the left tail of the distribution in formulating lender-of-last-resort policies. Improving the characterization of the distribution of extreme values is of paramount importance”

Por lo tanto, el enfoque que utilizaremos también hará uso de una rama de la Teoría de la Probabilidad y la Estadística especializada en describir eventos raros: *la Teoría de Valores Extremos*.

Ahora bien, aunque en otros trabajos, como [2], se han utilizado técnicas de la Teoría de Valores Extremos (e.g. el método *Peaks Over Threshold* o el de Máximos de Bloque) para medir el riesgo de mercado, nuestro enfoque contrasta con dichas aplicaciones en el siguiente modo: nosotros utilizaremos la Teoría de Valores Extremos para complementar nuestro modelo GARCH; específicamente, utilizaremos el método POT para modelar las innovaciones de dicha clase de procesos.⁶ Por lo tanto, nosotros aplicaremos las técnicas

este trabajo, donde podrá familiarizarse con el tema.

⁵ Alan Greenspan es un Doctor en Economía de origen estadounidense. Greenspan fue presidente de la Reserva Federal de los Estados Unidos de América desde el año 1987 hasta 2006.

⁶ Para conocer a detalle el método de Máximos de Bloque referimos al lector a [2]. Para

de valores extremos a una serie aproximadamente i.i.d., y no directamente a una serie de tiempo, como en [2].

Cabe mencionar que el enfoque anteriormente descrito para combinar el GARCH y el método POT en un solo modelo fue originalmente propuesto por Alexander McNeil y Rüdiger Frey, en su extraordinaria contribución *Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach*, [3].

Así, en este trabajo exploraremos cómo funciona el enfoque anterior en datos mexicanos; específicamente, en las series de rendimientos logarítmicos diarios históricos del Tipo de Cambio FIX (USD/MXN) y del Índice de Precios y Cotizaciones (principal indicador del desempeño del mercado mexicano de valores). Estas series fueron elegidas para realizar este trabajo por dos razones: i) su relevancia para el sistema financiero mexicano; y ii) a fin de contar con diversidad en los datos a analizar (en este caso, tenemos un índice bursátil y un tipo de cambio).

De este modo, realizaremos la estimación de dos importantes medidas de riesgo para dichas series históricas: el *VaR* y el *Expected Shortfall*. Más aún, desarrollaremos una serie de pruebas (*backtests*) que nos permitirán evaluar el desempeño de tales estimaciones; esto, para momentos tanto de crisis financieras (i.e. periodos de alta volatilidad en los mercados; donde los extremos ocurren) como de relativa calma (periodos donde la volatilidad no es tan grande).

Adicionalmente, ilustraremos cómo atacar algunos problemas prácticos que suelen aparecer al utilizar procesos GARCH en series de rendimientos durante largos periodos de tiempo. En particular, exploramos cómo se comporta la metodología descrita previamente cuando se le complementa con un método empírico de pronóstico de volatilidad (el *Exponential Weighted Moving Average*; EWMA) en los días en que alguno de parámetros del GARCH ajustado a los datos bajo consideración es no significativo (que si bien son raros, pueden llegar a presentarse en aplicaciones que estudian series durante amplias ventanas temporales).

Al realizar la aplicación previamente descrita también utilizaremos modelos GARCH con innovaciones Normal y *t*-Student, a fin de comparar los resultados obtenidos con aquellos de trabajos similares, como el [1]. No obstante, cabe mencionar que nuestra aplicación permitirá estimar cuantiles y déficits esperados a niveles de confianza considerablemente superiores a $\alpha = 0.95$.

conocer sobre el método POT, referimos al lector al Capítulo 3 de este trabajo.

Por lo tanto, nuestros objetivos serán dos: i) evaluar si la metodología que combina la potencia del Método POT para modelar extremos con el dinamismo de los modelos GARCH para modelar volatilidad, es capaz de entregar estimaciones razonables del *VaR* y *Expected Shortfall* condicionales a altos niveles de confianza en series financieras mexicanas y; ii) comparar los resultados de dicho enfoque con aquellos entregados por la metodología que consiste en modelar las innovaciones de un GARCH ajustado a una serie financiera mediante una distribución Normal o *t*-Student (i.e. un enfoque como el de [1]).

Ahora bien, al explorar la literatura relevante sobre este tipo de aplicaciones encontramos que en [4] se realiza un ejercicio similar al nuestro; no obstante, dicho trabajo se enfoca en índices bursátiles norteamericanos y en los precios de activos tales como el petróleo y el maíz en épocas de crisis. En otras palabras, [4] no analiza ninguna serie de origen mexicano.

Por otro lado, en [5], [6] y [7] se utiliza un GARCH en conjunto con el Método POT para analizar, entre otros activos, una serie financiera mexicana: el IPC. Sin embargo, dichas aplicaciones se enfocan únicamente en la estimación y evaluación del *VaR*; es decir, se deja de lado el análisis del *Expected Shortfall*, la cual es una medida de riesgo igualmente importante y que cuenta con la ventaja de ser una medida de riesgo *coherente* (véase la Sección 1.2.3).

Asimismo, en [8] se utiliza un GARCH y otra pieza de la TVE, el Método de Máximos de Bloque, para analizar otra serie financiera de origen mexicano: el Tipo de Cambio FIX. No obstante, dicho trabajo también centra su atención únicamente en el *VaR*; es decir, se deja fuera del análisis a otras series financieras mexicanas y al *Expected Shortfall*.

De este modo, nuestro trabajo se separa de los previamente mencionados en diversos sentidos. En primer lugar, explora qué tan adecuada es la metodología para la elección de un umbral óptimo sugerida en [3] para datos mexicanos. La elección del umbral es un tema de gran relevancia en las aplicaciones de extremos, por lo que investigar la adecuación de dicho método en series mexicanas tiene una gran importancia: muestra la presencia de un hecho empírico importante en datos nacionales, el cual puede ser relevante en análisis posteriores de los mismos.

Nuestra aplicación también contrasta con trabajos previos en el hecho de que analiza la eficacia en datos mexicanos de los modelos GARCH complementados por el Método POT tanto en periodos de crisis (a diferencia de [5] y [6]) como de relativa calma (en contraste con [4], [8] y [7]).

Finalmente, hasta donde sabemos, este es el primer análisis de esta naturale-

za que centra su atención tanto en el IPC como en el Tipo de Cambio FIX; es decir, es el primer trabajo que aplica y evalúa una metodología GARCH combinada con el Método POT en el tipo de cambio más relevante para México. Más aún, este es el primer trabajo que estima y evalúa las estimaciones del *Expected Shortfall* a altos niveles de confianza para series mexicanas.

La estructura de esta tesis se describe a continuación. En el primer Capítulo, introduciremos un marco teórico general en el que daremos las definiciones y nociones básicas necesarias para estudiar la cuantificación del riesgo de mercado. En el segundo Capítulo, tomaremos un enfoque especializado en modelos para series de tiempo y estudiaremos las bases de los procesos de volatilidad estocástica más relevantes en este trabajo, los procesos GARCH. En el tercer Capítulo estudiaremos los elementos de la Teoría de Valores Extremos esenciales para nuestra aplicación; es decir, aquellos necesarios para poder aplicar el método POT a una serie de datos i.i.d. En el cuarto Capítulo, estudiaremos cómo combinar las técnicas aprendidas en los Capítulos 2 y 3 para medir el riesgo de mercado de un activo financiero. En el quinto Capítulo expondremos nuestra aplicación en series financieras mexicanas. En el último Capítulo, presentaremos las conclusiones del trabajo, analizaremos las implicaciones de los resultados obtenidos y propondremos algunas ideas para posibles trabajos posteriores.

Cabe mencionar que todos los gráficos, pruebas y cálculos realizados en esta tesis fueron elaborados a través del software estadístico R.

Capítulo 1

Conceptos básicos

En este primer Capítulo estudiaremos algunos conceptos esenciales en la cuantificación de riesgos. Específicamente, introduciremos un marco probabilístico teórico para modelar el riesgo financiero; en particular el riesgo de mercado. Daremos definiciones probabilistas de medida de riesgo, factores de riesgo, ganancia y pérdida, entre otras.

También estudiaremos dos enfoques generales para modelar el riesgo de mercado: el condicional y el no condicional. Específicamente, analizaremos las propiedades de dichos enfoques y discutiremos cuál podría resultar más ante distintos escenarios o necesidades que pueden surgir al cuantificar el riesgo de mercado.

Por otro lado, una vez definida la noción de medida de riesgo y discutidos los dos enfoques anteriores para cuantificar el riesgo de mercado, nuestro estudio se centrará en dos medidas muy importantes en la industria financiera: el Valor en Riesgo y el Déficit Esperado (ES, por sus siglas en inglés: *Expected Shortfall*). Nuestro estudio de tales medidas mostrará diversas ventajas y desventajas de dichas cantidades, en un ámbito tanto práctico como teórico.

Terminaremos este Capítulo con el estudio de una metodología estándar para evaluar el desempeño de un sistema de cuantificación de riesgo de mercado: el *backtesting*. Dichas técnica nos será de gran utilidad en los Capítulos 2, 4 y 5, donde estudiaremos cómo utilizarla para alcanzar los objetivos planteados en la introducción de este trabajo.

Finalmente, cabe mencionar que el material de este Capítulo está basado en el contenido de [9], [10] y [11]; no obstante, a lo largo de cada Sección pueden encontrarse algunas referencias adicionales.

1.1. Factores de riesgo y distribuciones de pérdida

En esta Sección introduciremos el lenguaje y marco probabilístico básico que utilizaremos a lo largo de todo este trabajo. Comenzaremos con una serie de definiciones generales que nos ayudarán a dar certeza a diversos aspectos importantes en la cuantificación del riesgo de mercado.

Después, estudiaremos los enfoques condicional y no condicional para medir el riesgo de mercado. Dichos enfoques se caracterizarán por el tipo de función de distribución analizada en cada uno de ellos y nos permitirán estudiar pérdidas a corto y largo plazo, según se requiera.

1.1.1. Definiciones generales

Representamos lo incierto de los estados futuros del mundo mediante un espacio de probabilidad (Ω, \mathcal{F}, P) , que será el dominio de todas las variables aleatorias (v.a.'s) que definamos en este trabajo.

Así, consideremos un activo dado, este podría ser una acción, un bono, un tipo de cambio, un derivado o una tasa de interés, entre otros. Denotamos el valor del activo al tiempo s como $V(s)$ y suponemos que la v.a. $V(s)$ es observable al tiempo s . Para un horizonte de tiempo dado Δ , tal como 1 o 10 días, la pérdida del activo sobre el periodo $[s, s + \Delta]$ está dada por

$$L_{[s, s+\Delta]} = -(V(s + \Delta) - V(s))$$

En un contexto de portafolios, activos como los anteriormente mencionados suelen ser denominados *factores de riesgo* del portafolio, pues cambios en su valor pueden afectar el valor del portafolio que los contiene.

Ahora bien, normalmente se suele tomar la hipótesis de que $L_{[s, s+\Delta]}$ es observable al tiempo $s + \Delta$, pero en general representa una v.a. el tiempo s .

Haremos distinción entre la distribución condicional de $L_{[s, s+\Delta]}$ (aquella que se condiciona a toda la información disponible al tiempo s , incluyendo s) y la no condicional.

Observación 1.1. *Como buscamos una distribución de pérdidas hacemos que éstas se reflejen en cantidades positivas y las ganancias en negativas,*

esto permite la aplicación de técnicas actuariales, las cuales se centran en el análisis de variables aleatorias positivas, en la administración de riesgos.

Debido a que la mayor parte del tiempo utilizaremos un horizonte de tiempo fijo Δ , será conveniente medir el tiempo en unidades de Δ e introducir una notación de series de tiempo; pasamos de un proceso general $Y(s)$ a una serie de tiempo $\{Y_t\}_{t \in \mathbb{N}}$ con $Y_t := Y(t\Delta)$. Utilizando esta notación la pérdida de nuestro activo puede escribirse como:

$$L_{t+1} := L_{t\Delta, (t+1)\Delta} = -(V_{t+1} - V_t) \quad (1.1)$$

En la administración de riesgo de mercado normalmente se trabaja con modelos financieros donde el tiempo calendario s está medido en años y las tasas de interés y volatilidades son citadas en bases anualizadas. Entonces, si estuviéramos interesados en pérdidas diarias haríamos $\Delta = \frac{1}{365}$ o $\Delta \approx \frac{1}{250}$; la última convención es principalmente utilizada en mercados de derivados pues hay aproximadamente 250 días hábiles bancarios al año. Las variables aleatorias V_t y V_{t+1} representan entonces el valor del activo en los días t y $t + 1$ respectivamente y L_{t+1} representa la pérdida del día t al $t + 1$.

1.1.2. Distribuciones de pérdida condicional y no condicional

En administración de riesgos frecuentemente debemos decidir si estamos interesados en la distribución condicional o no condicional de las pérdidas de un activo. Ambas son importantes en esta disciplina, pero es importante estar conscientes de las diferencias entre ambos conceptos.

Las diferencias entre la distribución condicional de las pérdidas y la no condicional están fuertemente ligadas a las propiedades de serie de tiempo de dichas pérdidas. Supongamos que $\{L_t\}_{t \in \mathbb{N}}$ forma una serie de tiempo estacionaria con *distribución estacionaria* F_L en \mathbb{R} . Esencialmente, esto quiere decir que la distribución de $\{L_t\}_{t \in \mathbb{N}}$ es invariante ante cambios en el tiempo; la mayoría de los modelos de series de tiempo utilizados en la práctica para modelar cambios en el valor de un activo satisfacen esta propiedad.

Ahora, fijemos un punto t en el tiempo (momento actual), y denotemos por \mathcal{F}_t la sigma álgebra que representa la información públicamente disponible al tiempo t . Normalmente $\mathcal{F}_t = \sigma(\{L_s, s \leq t\})$, la σ -álgebra generada por los cambios en el valor de nuestro activo en el pasado y hasta ahora (comúnmente llamada la historia hasta el tiempo t e incluyendo t). Denotemos por $F_{L_{t+1}|\mathcal{F}_t}$

la distribución condicional de L_{t+1} dada la información actual, \mathcal{F}_t . En la mayoría de los modelos de series de tiempo estacionarias relevantes en administración de riesgos $F_{L_{t+1}|\mathcal{F}_t}$ no es igual a F_L , la distribución estacionaria de la serie de tiempo.

Un ejemplo importante de esto es provisto por los modelos GARCH (véase el Capítulo 2). En esta clase de modelos, la varianza de la distribución condicional de L_{t+1} es una función de las pérdidas pasadas y, posiblemente, de sus propios valores anteriores. Por otro lado, si $\{L_t\}_{t \in \mathbb{N}}$ es una serie de variables aleatorias i.i.d., tendremos la identidad obvia $F_{L_{t+1}|\mathcal{F}_t} = F_L$.

Las distribuciones condicionales son particularmente relevantes en la administración de riesgo de mercado, hecho que estudiaremos con detalle en este trabajo. En contraste, las distribuciones no condicionales de pérdida son de particular interés si el horizonte de tiempo sobre el que deseamos medir nuestras pérdidas es relativamente grande (como suele ser el caso en la administración de riesgo de crédito o en el sector asegurador).

A las técnicas de administración de riesgos basadas en la distribución condicional de pérdida se les suele llamar administración *condicional* o *dinámica* de riesgos, mientras que a las basadas en la distribución no condicional se les suele conocer como administración *estática* de riesgos.

1.2. Medidas de riesgo

La mayoría de las medidas de riesgo modernas son cantidades estadísticas que describen la distribución condicional o no condicional de pérdida de un activo, sobre un horizonte de tiempo predeterminado Δ . Algunos ejemplos de este tipo de medidas son el *VaR*, la desviación estándar o el *Expected Shortfall*.⁷ Es problemático confiar en una estadística particular para describir el riesgo contenido en toda una distribución. Sin embargo, la perspectiva de que la distribución de pérdida, como un todo, da una imagen precisa del riesgo en

⁷ La desviación estándar es una de las medidas de riesgo más antiguas. Ésta cobró un papel relevante en la administración de riesgos a partir de que Harry Markowitz la utilizara como medida de riesgo es su teoría de portafolios eficientes. Su impacto fue tal que a partir de entonces se trató de generalizar la noción de desviación estándar, a través de la definición de las *medidas de riesgo de desviación* (*deviation risk measures*, en inglés), las cuales comparten algunas características con las medidas de riesgo coherentes. No obstante, la desviación estándar no es una medida de riesgo coherente (pues no satisface uno de los axiomas de coherencia: invarianza bajo traslaciones) y tiene algunas deficiencias como medida de riesgo (e.g. no distingue entre desviaciones positivas y negativas de la media).

un activo es bastante razonable y cuenta con muchas ventajas:

- Es natural basar la medición del riesgo en las distribuciones de las pérdidas, pues estas son el objeto central de la administración de riesgos.
- El concepto de una distribución de pérdida tiene sentido en todos los niveles de agregación, desde un portafolio con un solo instrumento hasta la posición entera de una institución financiera.
- Si se estima apropiadamente, la distribución de pérdida refleja diferencias entre posiciones largas y cortas, además de efectos de diversificación.
- Las distribuciones de pérdida pueden ser comparadas directamente entre portafolios.

Por ejemplo, hace perfecto sentido comparar las distribuciones de pérdida de un portafolio de préstamos con la de un libro de derivados, si al menos el horizonte temporal, Δ , es el mismo para ambos casos.

Existen dos grandes problemas al trabajar con distribuciones de pérdida. Primero, cualquier estimador de la distribución de pérdida estará basado en información pasada. Si las leyes que gobiernan los mercados cambian, dicha información es de uso limitado para predecir el riesgo futuro. El segundo problema relacionado con este enfoque es práctico. Incluso en un ambiente estacionario, es difícil estimar la distribución de pérdida de forma precisa, particularmente en un contexto de portafolios grandes; por tanto, muchos sistemas de administración de riesgos aparentemente sofisticados están basados en modelos estadísticos relativamente burdos para la distribución de pérdida (incluyendo, por ejemplo, hipótesis insostenibles de normalidad).

Sin embargo, los puntos anteriores no son un argumento en contra del uso de distribuciones de pérdida. Más bien, llama a hacer mejoras en los métodos para estimar distribuciones de pérdida y, por supuesto, a utilizar con prudencia los modelos de administración de riesgos basados en la estimación de distribuciones de pérdida.

Con las consideraciones anteriores en mente, en las siguientes Secciones aterrizaremos el concepto de medida de riesgo, mediante el estudio de dos importantes medidas de riesgo basadas en distribuciones de pérdida: el *VaR* y el *Expected Shortfall*.

1.2.1. Valor en Riesgo (*VaR*)

El *VaR* es quizá la medida de riesgo más utilizada en instituciones financieras y se ha logrado abrir paso en el marco de adecuación de capital del Comité de Basilea, por lo que merece una discusión y análisis extensos. En esta Sección introducimos el *VaR* y discutimos algunos problemas que tiene esta medida en la práctica. Más adelante se discutirán los problemas que tiene el *VaR* en un sentido teórico, desde el punto de vista de las medidas de riesgo coherentes.

Consideremos un activo riesgoso y un horizonte de tiempo fijo Δ , y denotemos por $F_L(l) = P(L \leq l)$ la correspondiente función de distribución de pérdida. En esta Sección no hacemos distinción entre distribución condicional o no condicional; más bien, suponemos que la elección de ésta ha sido hecha en el preámbulo del análisis y que F_L representa la distribución de interés.

De esta forma, buscamos definir una estadística basada en F_L que mida la severidad del riesgo que carga nuestro activo sobre el periodo de tiempo Δ . Un candidato obvio es la máxima pérdida posible de nuestro activo, es decir el $\inf\{l \in \mathbb{R} | F_L(l) = 1\}$; por desgracia, en la mayoría de los modelos de interés en administración de riesgos el soporte de F_L no es acotado, por lo que dicha estadística termina por ser infinito. Además, al utilizar esta estadística hacemos caso omiso de cualquier información probabilística contenida en F_L . Es por esto que en lugar de centrarnos en la “máxima” pérdida, nos concentramos en la “máxima pérdida que no es excedida con un nivel alto de probabilidad”, el también llamado nivel de confianza.

Definición 1.2. (*Valor en Riesgo.*) Dado $\alpha \in (0, 1)$ (el nivel de confianza), el Valor en Riesgo (*VaR*) de nuestro activo está dado por

$$VaR_\alpha = \inf\{l \in \mathbb{R} | P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} | F_L(l) \geq \alpha\}$$

Así, en términos probabilísticos, el *VaR* es simplemente un cuantil de la distribución de pérdida. Valores típicos de α son $\alpha = .95$ o $\alpha = .99$. En la administración de riesgo de mercado se suele utilizar Δ como 1 o 10 días, mientras que en riesgo de crédito suele utilizarse como 1 año. Notemos que por su propia definición, una desventaja importante del *VaR* a un nivel de confianza α es que no nos da información sobre la severidad de las pérdidas que ocurren con probabilidad menor que $1 - \alpha$.

La Figura 1.1 ilustra la noción del *VaR*. Se muestra la función de densidad de probabilidad de una distribución de pérdida, con una línea sólida vertical

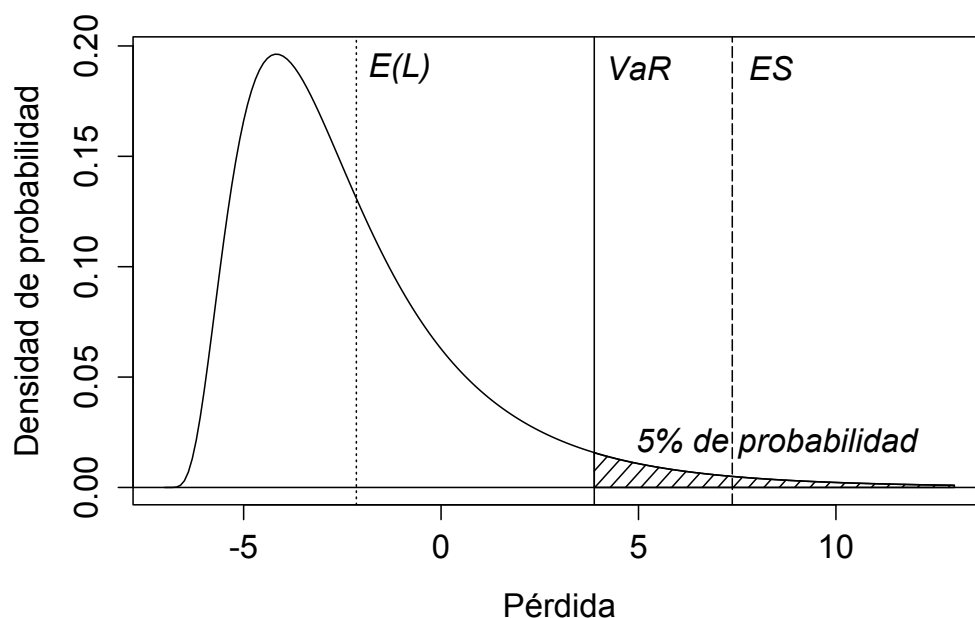


Figura 1.1: Un ejemplo de una distribución de pérdida con el VaR al 95% de confianza marcado con una línea horizontal; la pérdida promedio se muestra con una línea vertical punteada y el *Expected Shortfall* al 95% con una línea de guiones.

marcada en el VaR al 95% de confianza. Notemos que la esperanza de la pérdida es negativa ($E(L) = -2.15$), indicando que esperamos tener una ganancia, pero la cola derecha de la distribución es bastante más larga que la izquierda. El VaR al 95% de confianza es aproximadamente 3.9, lo cual indica que existe una probabilidad de 0.05 de que perdamos al menos esta cantidad; en la Figura 1.1 también se ilustra la noción del *Expected Shortfall*, la cual estudiaremos detalladamente más adelante.

Con lo anterior en mente es claro que los cuantiles juegan un papel importante en la administración de riesgos, por lo que requerimos una definición precisa de ellos.

Definición 1.3. Inversa Generalizada y Función Cuantil.

i) Dada una función creciente $T : \mathbb{R} \rightarrow \mathbb{R}$, la inversa generalizada de T

está definida como

$$T^{\leftarrow}(y) := \inf\{x \in \mathbb{R} | T(x) \geq y\}$$

ii) Dada una función de distribución F , la inversa generalizada de F es llamada la función cuantil de F . Para $\alpha \in (0, 1)$ el cuantil α de F está dado por

$$q_{\alpha}(F) := F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} | F(x) \geq \alpha\}$$

Para una variable aleatoria X solemos utilizar la notación $q_{\alpha}(X) := q_{\alpha}(F)$. Si F es continua y estrictamente creciente tenemos simplemente $F^{\leftarrow}(\alpha) = F^{-1}(\alpha)$. No obstante, para calcular cuantiles en casos más generales podemos utilizar el siguiente criterio:

Lema 1.4. *Un punto $x_0 \in \mathbb{R}$ es el cuantil α de una función de distribución $F \Leftrightarrow F(x_0) \geq \alpha$ y $F(x) < \alpha \quad \forall x < x_0$.*

Demostración. La demostración del lema se sigue inmediatamente de la definición de inversa generalizada y de la continuidad por la derecha de F . \square

Una vez con un criterio bien definido para encontrar cuantiles, veamos un ejemplo de cómo llevar a cabo esta tarea para un par de distribuciones muy conocidas.

Ejemplo 1.5. VaR para las distribuciones de pérdida Gaussiana y t -Student. *Supongamos que la función de distribución de la pérdida F_L es normal con media μ y varianza σ^2 . Sea $\alpha \in (0, 1)$. Entonces*

$$\text{VaR}_{\alpha} = \mu + \sigma\Phi^{-1}(\alpha), \quad (1.2)$$

donde Φ denota la función de distribución de una normal estándar y $\Phi^{-1}(\alpha)$ denota el cuantil α de Φ . La prueba es sencilla: sabemos que podemos escribir a L como $L = \mu + \sigma Z$, con Z una variable aleatoria normal estándar. Ahora, como tenemos una variable aleatoria continua, F_L es estrictamente creciente por lo que utilizando el Lema 1.4 y la definición de VaR tenemos que $\text{VaR}_{\alpha} = \mu + \sigma\Phi^{-1}(\alpha)$. Solo resta mostrar que $F_L(\text{VaR}_{\alpha}) = \alpha$. Tenemos

$$P(L \leq \text{VaR}_{\alpha}) = P\left(\frac{L - \mu}{\sigma} \leq \Phi^{-1}(\alpha)\right) = \Phi(\Phi^{-1}(\alpha)) = \alpha$$

Por supuesto, un resultado similar al anterior puede ser obtenido para cualquier familia que tenga parámetros de localización y escala; otro ejemplo útil

es la distribución de pérdida *t-Student*. Supongamos que nuestra pérdida L es tal que $(L - \mu)/\sigma$ tiene una distribución *t* estándar con ν grados de libertad; denotemos este modelo de pérdida como $L \sim t(\nu, \mu, \sigma^2)$ y notemos que sus momentos están dados por $E(L) = \mu$ y $Var(L) = \nu\sigma^2/(\nu - 2)$ cuando $\nu > 2$, por lo que σ no es la desviación estándar de la distribución. Así, con un procedimiento análogo al utilizado con la distribución normal obtenemos

$$VaR_\alpha = \mu + \sigma t_\nu^{-1}(\alpha) \quad (1.3)$$

Donde t_ν denota la función de distribución de una *t* estándar, la cual está disponible en la mayoría de los paquetes de cómputo estadísticos junto con su inversa.

Comentarios adicionales sobre el *VaR*

No subaditividad. El *VaR* ha sido fundamentalmente criticado como medida de riesgo por tener pobres propiedades de agregación. En [11] se discute que el *VaR* no es una medida de riesgo coherente al no ser subaditiva, propiedad que los autores de tal trabajo consideran que cualquier medida de riesgo razonable debería de tener.

En este momento no ahondaremos mucho en la subaditividad del *VaR*; no obstante, notemos que lo anterior tiene como consecuencia directa que si tenemos dos distribuciones de pérdida, F_1 y F_2 , para dos activos y denotamos la distribución de la pérdida total, $L = L_1 + L_2$, de un portafolio conformado por los dos activos por F_L , no necesariamente tendremos que $q_\alpha(F_L) \leq q_\alpha(F_{L_1}) + q_\alpha(F_{L_2})$, por lo que el *VaR* del portafolio no necesariamente está acotado superiormente por la suma de los *VaR* de los activos individuales. Esto contradice nuestra noción de que debería haber un beneficio por la diversificación obtenida al unificar las dos posiciones; también quiere decir que una desagregación de la administración de riesgos utilizando el *VaR* es difícil, pues no podemos estar seguros de que al sumar los *VaR* de distintos portafolios obtengamos una cota superior para el riesgo de toda una firma.

Riesgo de Modelo. En la práctica, a los números *VaR* se les da con frecuencia una interpretación errónea e incluso peligrosa. Decir que el *VaR* diario de un activo particular a un nivel de confianza $\alpha = .99$ es igual a l suele interpretarse como “con probabilidad .99 la pérdida del activo será menor que l ”.

Esta interpretación es errónea por dos razones. Para empezar, nuestra distribución de pérdida está sujeta a errores de estimación y al problema del riesgo de modelo.

El riesgo de modelo puede ser definido como el riesgo de que una institución financiera incurra en pérdidas porque sus modelos de administración de riesgos están mal especificados o erróneamente calibrados, o porque algunas de las hipótesis de estos modelos no se sostienen en la práctica. Por ejemplo, podríamos trabajar con una distribución normal para modelar pérdidas cuando en realidad la verdadera distribución de dichas pérdidas tiene colas pesadas, o bien, podríamos fallar al identificar *clusters* de volatilidad o dependencia en las colas al modelar la distribución de diversos activos a la vez.

Puesto que cualquier modelo financiero es solo una simplificación, y por lo tanto una manera imperfecta, de representar el mundo económico y la manera en que agentes económicos llevan a cabo inversiones, es justo decir que cualquier modelo de administración de riesgos está en algún grado sujeto al riesgo de modelo. Por supuesto, estos problemas son más pronunciados si tratamos de estimar el *VaR* a un nivel muy alto de confianza como $\alpha = 0.9997$, como se nos podría solicitar hacer en la determinación de capital económico.

Elección de los Parámetros del VaR. Ya hemos discutido que al trabajar con medidas de riesgo basadas en una distribución de pérdida debemos siempre elegir un horizonte de tiempo apropiado. En el caso del *VaR* tenemos además otro parámetro, α , asociado al nivel de confianza. Por supuesto, no hay valores óptimos para estos parámetros, pero existen algunas consideraciones que pueden influir en su elección.

El horizonte de administración de riesgos Δ debería reflejar el periodo de tiempo que una institución financiera esté comprometida a conservar su portafolio. Este periodo de tiempo es afectado por restricciones contractuales y legales, y por consideraciones de liquidez. Además, este horizonte de tiempo típicamente variará entre mercados; al elegir un horizonte de administración de riesgos para toda una empresa, una institución financiera o corporación no tiene mucha más opción que elegir el horizonte apropiado para el mercado en el que lleve a cabo la mayor parte de sus actividades. Por ejemplo, las compañías de seguros están típicamente limitadas a conservar su portafolio de reclamaciones por un año, es decir, que en este tiempo no pueden modificar sustancialmente su portafolio ni renegociar las primas que reciben. Por lo tanto, el horizonte global de administración de riesgos de estas instituciones debería ser de un año, lo que ocasionaría su horizonte de administración de riesgo de mercado también sea de un año; aun cuando los valores típicos de estos horizontes temporales suelen ser de apenas 1 o 10 días.

Por otro lado, existen otras consideraciones más bien prácticas que sugieren que Δ debería ser relativamente pequeño: la suposición de que un porta-

folio permanezca sin cambios durante este horizonte se sostiene con mayor facilidad para Δ pequeño. Más aún, la calibración y evaluación de modelos estadísticos para pérdidas $\{L_t\}_{t \in \mathbb{N}}$ resultan más sencillos si Δ es pequeño, pues esto suele traducirse en tener más datos a nuestra disposición para tales fines.

En cuanto a la elección del nivel de confianza α , nuevamente es difícil dar una recomendación general pues distintos niveles de α son apropiados para diferentes propósitos. Afortunadamente, una vez que tenemos una estimación de la distribución de pérdida es fácil calcular cuantiles para distintos niveles de confianza a la vez. Para fines de adecuación de capital, un alto nivel de confianza es utilizado con el propósito de tener un margen de seguridad suficiente. Por ejemplo, el Comité de Basilea propone utilizar el *VaR* a un nivel de confianza de .99 y Δ igual a 10 días para la administración de riesgo de mercado. Con el objetivo de fijar límites a los *traders*, un banco típicamente tomaría $\alpha = 0.95$ y Δ igual a 1 día.

1.2.2. *Expected Shortfall* (ES)

El propósito de esta Sección es introducir la noción del *Expected Shortfall*, el cual es un importante estadístico de la distribución de pérdida que frecuentemente es utilizado como medida de riesgo en finanzas, seguros y administración de riesgos. Para ello, al igual que en la Sección anterior, supondremos que una cierta distribución de pérdida F_L ha sido ajustada en el preámbulo del análisis.

El *Expected Shortfall* o déficit esperado está íntimamente relacionado con el *VaR*. No obstante, esta medida de riesgo puede llegar a ser preferida sobre el *VaR* debido a que supera varias de sus deficiencias conceptuales (relacionadas con subaditividad).

Definición 1.6. *Expected Shortfall.* Para una pérdida L con $E(|L|) < \infty$ y función de distribución F_L , el *Expected Shortfall* a un nivel de confianza $\alpha \in (0, 1)$ está definido como

$$ES_\alpha := \frac{1}{1 - \alpha} \int_\alpha^1 q_u(F_L) du, \quad (1.4)$$

donde $q_u(F_L) = F^{\leftarrow}(u)$ es la función cuantil de F_L .

El *Expected Shortfall* y el *VaR* guardan entonces la siguiente relación:

$$ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 VaR_u(L) du.$$

Así, en lugar de fijar un nivel de confianza α , promediamos los VaR sobre todos los niveles de confianza $u \geq \alpha$ y así “miramos más profundamente dentro de la cola” de la distribución de pérdida. Obviamente, el ES_α depende solo de la distribución de L y tenemos que $ES_\alpha \geq VaR_\alpha$.

Para distribuciones de pérdida continuas podemos derivar una expresión aún más intuitiva que nos permite ver que el ES_α puede ser interpretado como la pérdida esperada en caso de que el VaR sea excedido.

Lema 1.7. *Para una pérdida integrable L con función de distribución continua F_L y cualquier $\alpha \in (0, 1)$, tenemos*

$$ES_\alpha = \frac{E(L; L \geq q_\alpha(L))}{1-\alpha} = E(L|L \geq VaR_\alpha), \quad (1.5)$$

donde utilizamos la notación $E(X; A) := E(XI_A)$ para una variable aleatoria integrable genérica X y un conjunto genérico $A \in \mathcal{F}$.

Demostración. Sea U una variable aleatoria con distribución uniforme en el intervalo $[0, 1]$. Sabemos entonces que la variable aleatoria $F_L^\leftarrow(U)$ tiene función de distribución F_L . Debemos mostrar que $E(L; L \geq q_\alpha(L)) = \int_\alpha^1 F_L^\leftarrow(u) du$, pues de este modo, al dividir la igualdad entre $1 - \alpha = P(L > q_\alpha(L))$ tendríamos nuestro resultado (gracias a la definición de esperanza condicional). Entonces,

$$\begin{aligned} E(L; L \geq q_\alpha(L)) &= E(F_L^\leftarrow(U); F_L^\leftarrow(U) \geq F_L^\leftarrow(\alpha)) \\ &= E(F_L^\leftarrow(U); U \geq \alpha) \\ &= \int_\alpha^1 F_L^\leftarrow(u) du \\ &= \int_\alpha^1 q_u(F_L) du \end{aligned}$$

La primera igualdad de arriba es obtenida al aplicar la definición de función cuantil a $q_\alpha(L)$ y utilizar el hecho de que $F_L(U) \stackrel{d}{=} L$; la segunda se da gracias a que F_L^\leftarrow es estrictamente creciente (pues F_L es continua); la tercera igualdad se obtiene a aplicar el *lema del estadístico inconsciente*; para la última igualdad simplemente aplicamos nuevamente la definición de función cuantil y tenemos nuestro resultado. \square

Observación 1.8. *Para una pérdida cuya función de distribución F_L sea discontinua la fórmula (1.5) no es verdadera para toda α . En su lugar, tenemos la siguiente expresión un tanto más complicada*

$$ES_\alpha = \frac{1}{1-\alpha} [E(L; L \geq q_\alpha) + q_\alpha(1-\alpha - P(L \geq q_\alpha))] \quad (1.6)$$

Podemos encontrar una prueba de esto en la proposición 3.2 de [10].

Ahora utilizamos el Lema 1.7 para calcular el *Expected Shortfall* de dos distribuciones continuas muy utilizadas en administración de riesgos.

Ejemplo 1.9. *Expected Shortfall para una distribución de pérdida Gaussiana.* *Supongamos que la distribución de pérdida es normal con media μ y varianza σ^2 . Sea $\alpha \in (0, 1)$. Entonces,*

$$ES_\alpha = \mu + \frac{\sigma\phi(\Phi^{-1}(\alpha))}{1-\alpha}, \quad (1.7)$$

donde ϕ es la función de densidad de una distribución normal estándar. La prueba de esto es relativamente sencilla.

*Recordemos nuevamente que si $L \sim N(\mu, \sigma^2)$, entonces podemos escribirla como $L = \mu + \sigma Z$, con $Z \sim N(0, 1)$ (de donde obviamente tenemos que $Z = \frac{L-\mu}{\sigma}$). Con esto en mente, y con ayuda de la expresión dada por el Lema 1.7 para el *Expected Shortfall*, tenemos que:*

$$ES_\alpha = E(L|L \geq q_\alpha(L)) = \mu + \sigma E(Z|Z \geq q_\alpha(Z));$$

*por lo tanto, basta con calcular el *Expected Shortfall* para la variable aleatoria Z para conocer el *ES* de L . Entonces, realicemos el cálculo del *ES* de Z :*

$$ES_\alpha(Z) = \frac{1}{1-\alpha} \int_{\Phi^{-1}(\alpha)}^{\infty} l\phi(l) dl = \frac{1}{1-\alpha} [-\phi(l)]_{\Phi^{-1}(\alpha)}^{\infty} = \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

Finalmente, con esto podemos concluir que

$$ES_\alpha(L) = \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

Ejemplo 1.10. *Expected Shortfall para una distribución de pérdida t-Student.* Supongamos que la pérdida L es tal que $\tilde{L} = \frac{L-\mu}{\sigma}$ tiene distribución t -Student estandarizada con ν grados de libertad, como en el Ejemplo 1.5. Supongamos también que $\nu > 1$. Siguiendo el mismo razonamiento que en el ejemplo anterior, y que de hecho es aplicable a cualquier distribución que forme parte de una familia con parámetros de localización y escala, tenemos que $ES_\alpha(L) = \mu + \sigma ES_\alpha(\tilde{L})$. El *Expected Shortfall* de la distribución t -Student estándar puede calcularse fácilmente vía integración directa y resulta ser:

$$ES_\alpha(\tilde{L}) = \frac{g_\nu(t_\nu^{-1}(\alpha))}{1-\alpha} \left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu-1} \right), \quad (1.8)$$

donde t_ν denota la función de distribución de la t -Student estandarizada y g_ν su función de densidad. Finalmente, el *ES* de la distribución t -Student resulta:

$$ES_\alpha(L) = \mu + \sigma \left[\frac{g_\nu(t_\nu^{-1}(\alpha))}{1-\alpha} \left(\frac{\nu + (t_\nu^{-1}(\alpha))^2}{\nu-1} \right) \right].$$

El siguiente Lema nos da una especie de Ley de los Grandes Números para el *Expected Shortfall* en términos de estadísticas de orden.

Lema 1.11. Para una sucesión $\{L_i\}_{i \in \mathbb{N}}$ de variables aleatorias *i.i.d.* con función de distribución F_L , tenemos que

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^{\lfloor n(1-\alpha) \rfloor} L_{i,n}}{\lfloor n(1-\alpha) \rfloor} = ES_\alpha, \quad c.s., \quad (1.9)$$

donde $L_{1,n} \geq L_{2,n} \geq \dots \geq L_{n,n}$ representan las estadísticas de orden de L_1, L_2, \dots, L_n y $\lfloor n(1-\alpha) \rfloor$ es el entero más grande que no excede $n(1-\alpha)$.

Este Lema nos da una forma de estimar el *Expected Shortfall* cuando tenemos muestras de tamaño considerable y $n(1-\alpha)$ es un número relativamente grande. Sin embargo, este no es generalmente el caso en la práctica, excepto quizá cuando un enfoque Montecarlo de estimación de riesgo es utilizado. Una prueba del Lema 1.11 puede ser encontrada en la proposición 4.1 de [10].

Cabe mencionar que, puesto que el ES_α puede ser pensado como un promedio de todas las pérdidas que son mayores o iguales al VaR_α , éste es sensible a la severidad de las pérdidas que excedan al VaR_α .

Observación 1.12. *Es posible derivar resultados asintóticos del ratio Short-fall-Cuantil ES_α/VaR_α para $\alpha \rightarrow 1$. Para la distribución normal tenemos que $\lim_{\alpha \rightarrow 1} ES_\alpha/VaR_\alpha = 1$; para la distribución t con $\nu > 1$ grados de libertad tenemos que $\lim_{\alpha \rightarrow 1} ES_\alpha/VaR_\alpha = \nu/(\nu - 1) > 1$. Esto muestra que para una distribución de cola pesada la diferencia el ES y el VaR es más pronunciada que para la distribución normal.*

Con la observación anterior hemos concluido nuestro estudio inicial del VaR y el *Expected Shortfall*. Pasaremos entonces a estudiar un concepto fundamental en el estudio de las medidas de riesgo: las medidas de riesgo coherentes. Las cuales, nos dan un criterio para reconocer medidas de riesgo con buenas propiedades teóricas y que por tanto podrían ser de utilidad en la práctica.

1.2.3. Medidas de riesgo coherentes

Se ha mencionado en diferentes ocasiones a lo largo de este trabajo que a pesar de que el VaR es quizá la medida de riesgo más ampliamente utilizada en la actualidad, esta carece de algunas propiedades que podríamos considerar *deseables* en toda medida de riesgo razonable. Es por esto que en esta Sección introducimos la noción de medida de riesgo *coherente*. Este concepto hace referencia a medidas de riesgo que satisfacen todas las propiedades que los expertos en administración de riesgos estiman necesarias para considerar que una medida de riesgo es razonable.

A continuación, introducimos un poco de notación para definir lo que es una medida de riesgo coherente; después, exponemos un ejemplo que muestra que el VaR no cumple las condiciones necesarias para considerarse dentro del conjunto de tales medidas; finalmente, demostramos que el *Expected Shortfall* sí cumple con dichas condiciones y que puede, por tanto, considerarse como una medida de riesgo coherente.

¿Qué es una medida de riesgo coherente?

Sea $\mathcal{L}^0(\Omega, \mathcal{F}, P)$ el conjunto de todas las variables aleatorias definidas sobre (Ω, P) tales que $P(X < \infty) = 1$. Definamos \mathcal{M} como el conjunto de variables aleatorias que representan las pérdidas de algún portafolio de inversión sobre un horizonte de tiempo fijo Δ . Supondremos que \mathcal{M} es un cono convexo, esto es, si tenemos que $L_1, L_2 \in \mathcal{M}$, entonces, $L_1 + L_2 \in \mathcal{M}$ y $\lambda L_1 \in \mathcal{M}$, si $\lambda > 0$. Adicionalmente, diremos que ρ es una medida de riesgo si es una función real-valuada definida de \mathcal{M} a \mathbb{R} . Así, la cantidad $\rho(L)$ puede ser interpretada

como el capital que es necesario agregar a una posición para que el perfil de riesgo de la inversión sea aceptable ante una entidad reguladora.

Con las definiciones y nociones anteriores estamos en posibilidad de definir lo que es una medida de riesgo coherente.

Definición 1.13. Medida de Riesgo Coherente. Diremos que una medida de riesgo ρ es una medida de riesgo coherente si satisface los siguientes axiomas:

1. *Invarianza bajo Traslaciones.* Para toda $L \in \mathcal{M}$, $l \in \mathbb{R}$, se cumple que:

$$\rho(L + l) = l + \rho(L).$$

2. *Subaditividad.* Para toda $L_1, L_2 \in \mathcal{M}$, se tiene:

$$\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2).$$

3. *Homogeneidad Positiva.* Para toda $\lambda > 0$, $L \in \mathcal{M}$, se satisface:

$$\rho(\lambda L) = \lambda \rho(L).$$

4. *Monotonicidad.* $\forall L_1, L_2 \in \mathcal{M}$ tales que $P(L_1 \leq L_2) = 1$, se cumple que:

$$\rho(L_1) \leq \rho(L_2).$$

Podemos notar entonces que el axioma 1 nos dice que si sumamos o restamos una cantidad determinística l a una posición que conlleva una pérdida L , nuestros requerimientos de capital se modifican acorde a tal cantidad.

El axioma 2 es el más debatido de todos los axiomas de medidas de riesgo coherentes, quizá porque es justo en este donde el *VaR* falla y por el cual se le excluye del conjunto de las medidas coherentes de riesgo. Este axioma refleja las bondades de la diversificación en los portafolios, sosteniendo que, si un portafolio está más diversificado que otro, este debería acarrear menos riesgo a su tenedor que el primero.

El axioma 3 tiene sentido casi de forma inmediata si el axioma 2 se sostiene. La subaditividad implica que, para toda $n \in \mathbb{N}$

$$\rho(nL) = \rho(L + \cdots + L) \leq n\rho(L);$$

puesto que si no hay neteo o diversificación entre las pérdidas en este portafolio, es natural que se requiera que la relación de igualdad se sostenga en la ecuación anterior.

Finalmente, podemos observar que el axioma 4 resulta obvio desde un punto de vista económico: posiciones que llevan a pérdidas más grandes sin importar lo que pase en el mundo requieren más capital de riesgo.

El VaR no es una medida de riesgo coherente

Una vez definida y analizada la definición de una medida de riesgo coherente conviene preguntarnos: ¿Qué medidas de riesgo estudiadas hasta este momento son coherentes? Naturalmente, una de las primeras medidas que viene a nuestra mente es el VaR , puesto que, como ya se ha mencionado, esta es quizá la medida de riesgo más utilizada en la actualidad en el sector financiero.

Sin embargo, a pesar del amplio uso y popularidad que tiene el VaR , esta resulta no ser una medida de riesgo coherente. A continuación, demostramos que el VaR satisface 3 de los 4 axiomas de coherencia de las medidas de riesgo, y terminamos esta Sección con un ejemplo donde se muestra que el VaR no cumple uno de ellos: la subaditividad.

Para probar que el VaR es invariante bajo traslaciones basta con recordar su definición y algunas propiedades del ínfimo de un conjunto de números reales. Sean L la pérdida de algún portafolio sobre un horizonte fijo de tiempo Δ y $l \in \mathbb{R}$. Realizando los cálculos de manera directa tenemos que:

$$\begin{aligned} VaR_{\alpha}^{L+l} &= \inf\{x \in \mathbb{R} : P(L+l > x) \leq 1 - \alpha\} \\ &= \inf\{x \in \mathbb{R} : P(L > x-l) \leq 1 - \alpha\}, \quad \text{sea } y = x-l \\ &= \inf\{y+l \in \mathbb{R} : P(L > y) \leq 1 - \alpha\} \\ &= \inf\{y \in \mathbb{R} : P(L > y) \leq 1 - \alpha\} + l \\ &= VaR_{\alpha}^L + l \end{aligned}$$

Y por lo tanto, el VaR satisface el axioma 1 de coherencia.

La demostración de que el VaR es una medida con homogeneidad positiva también puede obtenerse mediante cálculo directo, sea $\lambda \in \mathbb{R}^+$, entonces, tenemos que:

$$\begin{aligned} VaR_{\alpha}^{\lambda L} &= \inf\{x \in \mathbb{R} : P(\lambda L > x) \leq 1 - \alpha\} \\ &= \inf\{x \in \mathbb{R} : P(L > x/\lambda) \leq 1 - \alpha\}, \quad \text{sea } y = x/\lambda \\ &= \inf\{\lambda y \in \mathbb{R} : P(L > y) \leq 1 - \alpha\} \\ &= \lambda \inf\{y \in \mathbb{R} : P(L > y) \leq 1 - \alpha\} \\ &= \lambda VaR_{\alpha}^L \end{aligned}$$

Y, de este modo, tenemos que el VaR satisface el axioma 3 de coherencia.

El último de los axiomas de coherencia que satisface el VaR es el de monotonicidad, para demostrar este hecho tomemos dos pérdidas L_1 y L_2 , tales que $P(L_1 \leq L_2) = 1$ y notemos que este hecho nos basta para tener que

$$\begin{aligned} VaR_\alpha^{L_1} &= \inf\{x \in \mathbb{R} : P(L_1 > x) \leq 1 - \alpha\} \\ &\leq \inf\{x \in \mathbb{R} : P(L_2 > x) \leq 1 - \alpha\}, \quad \text{pues } P(L_1 \leq L_2) = 1 \\ &= VaR_\alpha^{L_2} \end{aligned}$$

Por lo tanto, el VaR satisface el axioma 4 de coherencia.

Ahora, veremos que el VaR no satisface el axioma 2 de coherencia. Para esto, exponemos un ejemplo en el cual no se sostiene la subaditividad para el VaR de dos portafolios.

Ejemplo 1.14. El VaR no es Subaditivo. Consideremos un portafolio de $d = 100$ bonos corporativos “fallables”, es decir, que estos bonos tienen asociada una probabilidad positiva de caer en incumplimiento de pago. Supongamos que los incumplimientos de los distintos bonos son independientes entre sí y que la probabilidad de incumplimiento es idéntica para todos ellos e igual a 2%. Fijemos el precio actual de los bonos en \$100. Si no hay incumplimiento, un bono paga al tiempo $t + 1$ (donde t se mide en, digamos, años) un monto de \$105, y \$0 en cualquier otro caso. Por lo tanto L_i , la pérdida asociada al bono i , es de -5 si no hay incumplimiento y de 100 si lo hay. Sea Y_i la indicadora de incumplimiento del bono i

$$Y_i = \begin{cases} 0, & \text{si el bono } i \text{ no incumple en } [t, t+1] \\ 1, & \text{si el bono } i \text{ sí incumple en } [t, t+1] \end{cases}$$

De este modo $L_i = 100Y_i - 5(1 - Y_i) = 105Y_i - 5$. Por lo tanto, las L_i s forman una sucesión de variables aleatorias *i.i.d.* con $P(L_i = -5) = 0.98$ y $P(L_i = 100) = 0.02$. Comparemos los siguientes dos portafolios, ambos con un valor actual de \$10,000 (es decir, $V_t = 10,000$):

- *Portafolio A: Completamente concentrado. Está constituido por 100 unidades del bono 1.*
- *Portafolio B: Completamente diversificado. Está constituido por 1 unidad de cada uno de los bonos.*

Considerando lo anterior, la intuición económica sugiere que el portafolio B es menos riesgoso y que por tanto debería tener un VaR menor. Sin embargo, veamos qué pasa al calcular el VaR al 95% de confianza para ambos portafolios.

La pérdida del portafolio A está dada por $L_A = 100L_1$, entonces, por la homogeneidad positiva del VaR, tenemos que $VaR_{0.95}(L_A) = 100VaR_{0.95}(L_1)$. Ahora, como $P(L_1 \leq -5) = 0.98 > 0.95$ y $P(L_1 < l) = 0$ para toda $l < -5$, tenemos que $VaR_{0.95}(L_1) = -5$, y por lo tanto que $VaR_{0.95}(L_A) = -500$. Lo que significa que aun después de retirar \$500 del capital de riesgo de este portafolio, dicho capital seguiría siendo aceptable para un administrador de riesgos que trabaje con un VaR al 95 % de confianza.

Por otro lado, para el portafolio B tenemos que $L_B = 105 \sum_{i=1}^{100} Y_i - 500$. Por lo tanto, debido a la homogeneidad positiva e invarianza bajo traslaciones del VaR, tenemos que $VaR_{0.95}(L_B) = 105VaR_{0.95}(\sum_{i=1}^{100} Y_i) - 500$. Sea $M = \sum_{i=1}^{100} Y_i$, entonces $M \sim Binomial(100, 0.02)$. Así, tenemos que $P(M \leq 5) \approx 0.984 > 0.95$ y $P(M \leq 4) \approx 0.949$, por lo que $VaR_{0.95}(M) = 5$. Por lo tanto, $VaR_{0.95}(L_B) = 105(5) - 500 = 525 - 500 = 25$. Es decir, para el portafolio B se necesitaría un capital del \$25 para satisfacer a un regulador que trabaje con un VaR al 95 % de confianza. En consecuencia, claramente tenemos que $VaR_{0.95}(L_B) > VaR_{0.95}(L_A)$.

Con lo anterior, tenemos un ejemplo de que pueden existir ocasiones en que medir el riesgo con el VaR puede llevar a resultados sin sentido; esta medida de riesgo puede no reflejar los beneficios de la diversificación sobre un portafolio. Más aún, el Ejemplo 1.14 muestra que el VaR no es subaditivo en general pues, para cualquier medida de riesgo coherente ρ que dependa únicamente de la distribución de L , tenemos que

$$\rho\left(\sum_{i=1}^{100} L_i\right) \leq \sum_{i=1}^{100} \rho(L_i) = 100\rho(L_1) = \rho(100L_1).$$

Por lo tanto *cualquier* medida de riesgo coherente, que dependa solo de la distribución de pérdida, inducirá a un mayor requerimiento de capital para el portafolio A que para el portafolio B.

Cabe mencionar que es posible construir ejemplos como el anterior gracias al uso de distribuciones muy sesgadas en ellos; es en esta característica donde surge el problema del VaR con la subaditividad. Sin embargo, si se tienen hipótesis un poco más restrictivas sobre la distribución de la pérdida de un portafolio, como que sea parte de la familia de distribuciones elípticas, se puede probar que el VaR puede llegar a ser subaditivo.⁸

⁸ Véase el Capítulo 3 de [9].

El *Expected Shortfall* es una medida de riesgo coherente

Para terminar nuestra breve discusión sobre medidas de riesgo coherentes demostraremos que nuestra segunda medida de riesgo favorita, el *Expected Shortfall*, sí cumple con los cuatro axiomas de coherencia de las medidas de riesgo.

Comenzaremos probando que el *Expected Shortfall* es invariante bajo traslaciones. Para ello tomemos la distribución de alguna pérdida L y un número real l . Podemos entonces verificar vía cálculo directo, y utilizando la definición del *Expected Shortfall*, que:

$$\begin{aligned} ES_\alpha^{L+l} &= \frac{1}{1-\alpha} \left[\int_\alpha^1 VaR_u^{L+l} du \right] \\ &= \frac{1}{1-\alpha} \left[\int_\alpha^1 VaR_u^L du + \int_\alpha^1 l du \right], \text{ invarianza bajo traslaciones del } VaR \\ &= \frac{1}{1-\alpha} \int_\alpha^1 VaR_u^L du + l \\ &= ES_\alpha^L + l. \end{aligned}$$

Por lo tanto, el *Expected Shortfall* cumple el axioma 1 de coherencia.

Para verificar que el *Expected Shortfall* cumple el axioma de homogeneidad positiva nos apoyamos del hecho de que el VaR cumple con esta característica, sea $\lambda \in \mathbb{R}^+$, tenemos que:

$$\begin{aligned} ES_\alpha^{\lambda L} &= \frac{1}{1-\alpha} \left[\int_\alpha^1 VaR_u^{\lambda L} du \right] \\ &= \frac{1}{1-\alpha} \left[\int_\alpha^1 \lambda VaR_u^L du \right], \text{ por la homogeneidad positiva del } VaR \\ &= \frac{\lambda}{1-\alpha} \left[\int_\alpha^1 VaR_u^L du \right] \\ &= \lambda ES_\alpha^L. \end{aligned}$$

Y así, tenemos que el *Expected Shortfall* cumple con el axioma 2 de coherencia. Para probar que el *Expected Shortfall* es monótono tomemos dos pérdidas, L_1 y L_2 , tales que $P(L_1 \leq L_2) = 1$, tenemos entonces que:

$$\begin{aligned} ES_\alpha^{L_1} &= \frac{1}{1-\alpha} \left[\int_\alpha^1 VaR_u^{L_1} du \right] \\ &\leq \frac{1}{1-\alpha} \left[\int_\alpha^1 VaR_u^{L_2} du \right], \text{ pues el } VaR \text{ es monótono} \\ &= ES_\alpha^{L_2}. \end{aligned}$$

Y, de este modo, el *Expected Shortfall* cumple también el axioma de coherencia 4.

Para demostrar que el *Expected Shortfall* es subaditivo tomamos toda una nueva metodología; pues en esta ocasión no podemos apoyarnos en las propiedades del VaR .

Consideremos una sucesión genérica de variables aleatorias L_1, \dots, L_n con estadísticas de orden $L_{1,n} \geq L_{2,n} \geq \dots \geq L_{n,n}$ y notemos que para un m arbitrario tal que $1 \leq m \leq n$ se cumple que

$$\sum_{i=1}^m L_{i,m} = \sup\{L_{i_1} + \dots + L_{i_m} : 1 \leq i_1 < \dots < i_m \leq n\}.$$

Ahora consideremos 2 variables aleatorias L y \tilde{L} con función de distribución conjunta F y una sucesión de vectores aleatorios bivariados $(L_1, \tilde{L}_1), (L_2, \tilde{L}_2), \dots, (L_n, \tilde{L}_n)$ con la misma función de distribución F . Escribiendo $(L + \tilde{L})_i := L_i + \tilde{L}_i$ y $(L + \tilde{L})_{i,n}$, para las estadísticas de orden de $(L + \tilde{L})_1, \dots, (L + \tilde{L})_n$ observamos que se debe cumplir que

$$\begin{aligned} \sum_{i=1}^m (L + \tilde{L})_{i,n} &= \sup\{(L + \tilde{L})_{i_1} + \dots + (L + \tilde{L})_{i_m} : 1 \leq i_1 < \dots < i_m \leq m\} \\ &\leq \sup\{L_{i_1} + \dots + L_{i_m} : 1 \leq i_1 < \dots < i_m \leq m\} \\ &\quad + \sup\{\tilde{L}_{i_1} + \dots + \tilde{L}_{i_m} : 1 \leq i_1 < \dots < i_m \leq m\} \\ &= \sum_{i=1}^m L_{i,n} + \sum_{i=1}^m \tilde{L}_{i,n}. \end{aligned}$$

Donde la desigualdad de arriba se obtiene gracias a las propiedades del supremo de la “suma” de dos conjuntos. Haciendo $m = [n(1 - \alpha)]$ y dejando $n \rightarrow \infty$ obtenemos, por el Lema 1.11, que

$$ES_\alpha^{L+\tilde{L}} \leq ES_\alpha^L + ES_\alpha^{\tilde{L}}.$$

De esta manera, queda demostrado que el *Expected Shortfall* cumple el axioma 2 de coherencia y, por lo tanto, con la definición de una Medida de Riesgo Coherente.

1.3. Evaluando sistemas de cuantificación de riesgos: *backtesting*

En esta Sección discutiremos una metodología fundamental en la administración de riesgo de mercado: el *backtesting*. Dicha metodología ofrece un alternativa general para monitorear y evaluar el desempeño de sistemas de medición de riesgo de mercado.

En las secciones anteriores hemos considerado métodos estándar para estimar medidas de riesgo al tiempo t para la distribución de la pérdida sobre el siguiente periodo de tiempo. Cuando este procedimiento es implementado de manera continua, tenemos la posibilidad de monitorear el desempeño de los

métodos y de comparar su desempeño relativo. Dicho proceso de monitoreo es conocido como *backtesting*.

Supongamos que hacemos estimaciones para el *VaR* y el *Expected Shortfall* para 1 y h periodos de tiempo. Denotemos las verdaderas medidas de riesgo para un periodo como VaR_α^t y ES_α^t , y los de h periodos como $VaR_\alpha^{t,h}$ y $ES_\alpha^{t,h}$. Estas medidas pueden ser condicionales o no condicionales. Al tiempo $t+1$ tenemos la oportunidad de comparar nuestras estimaciones de 1 periodo con lo que en realidad sucedió y al tiempo $t+h$ lo mismo para nuestras estimaciones de h periodos.

Por definición de *VaR* (y suponiendo una distribución continua para la pérdida) tenemos que $P(L_{t+h} > VaR_\alpha^{t,h}) = 1 - \alpha$, por lo que la probabilidad de una violación (exceso) al *VaR* es de $1 - \alpha$. En la práctica, las medidas de riesgo tienen que ser estimadas a partir de datos, por lo que introducimos una notación de indicadores de violaciones a las estimaciones del *VaR*:

$$\widehat{I}_{t+1} := I_{\{L_{t+1} > \widehat{VaR}_\alpha^t\}}, \quad \widehat{I}_{t+h} := I_{\{L_{t+h} > \widehat{VaR}_\alpha^{t,h}\}}. \quad (1.10)$$

Esperamos que, si nuestro método de estimación es razonable, estas indicadores se comporten como variables aleatorias Bernoulli con probabilidad de éxito (violación) cercana a $1 - \alpha$. Si realizamos múltiples comparaciones de las predicciones del *VaR* y las correspondientes pérdidas realizadas (que en realidad sucedieron), entonces esperamos que la proporción de ocasiones en las que el *VaR* es excedido sea de $1 - \alpha$, aproximadamente.

En situaciones más específicas podemos decir más. Por ejemplo, si realizamos estimaciones de un *VaR* condicional para 1 periodo utilizando un enfoque dinámico, entonces esperamos que las indicadores de violación \widehat{I}_t se comporten como variables aleatorias i.i.d. Bernoulli($1 - \alpha$); el número de excesos a estas cantidades durante m periodos debería comportarse como una variable aleatoria Binomial($m, 1 - \alpha$), si nuestro método de estimación es razonable. Esto será discutido más adelante, en el Capítulo correspondiente a series de tiempo.

También nos gustaría poder probar la eficacia de nuestra estimación del *Expected Shortfall*. Considerando, por simplicidad, el estimador para un periodo del *Expected Shortfall*, se sigue del Lema 1.7 que para una distribución de pérdida continua la identidad

$$E[(L_{t+1} - ES_\alpha^t)I_{\{L_{t+1} > VaR_\alpha^t\}}] = 0,$$

se satisface. Esto sugiere que fijemos nuestra atención en la discrepancia $L_{t+1} - \widehat{ES}_\alpha^t$ en los días en los que el *VaR* es excedido. Si nuestras estimaciones

son razonables, estos datos deberían provenir de una distribución con media cero. Más adelante, en el Capítulo de series de tiempo de este trabajo, con hipótesis adicionales de modelo, analizaremos esta idea con mayor detalle.

Capítulo 2

Series de tiempo financieras

En este Capítulo estudiaremos distintos modelos de series de tiempo para series de cambios (variaciones) en el valor de algún activo financiero; en particular, estudiaremos modelos univariados para series de rendimientos logarítmicos de precios, tasas, tipos de cambio, etc.

De esta manera, en la primera Sección del Capítulo documentaremos algunos conceptos fundamentales en el análisis de series de tiempo, tales como estacionariedad, correlaciones y su estimación, procesos de ruido blanco y procesos ARMA (procesos Auto-Regresivos de Medias Móviles).

Posteriormente, una vez sentadas las bases del estudio de series de tiempo, centraremos nuestra atención en los procesos ARCH y GARCH (Procesos Auto-Regresivos de Heterocedasticidad Condicional Generalizados) univariados, los cuales capturan el comportamiento de un fenómeno fundamental en el estudio de series de tiempo financieras: la volatilidad. Específicamente, analizaremos diversos enfoques para ajustar este tipo de procesos a datos financieros; es decir, estudiaremos diversas maneras de estimar los parámetros de los modelos ARCH y GARCH a partir de datos.

Avanzaremos después a estudiar cómo medir el riesgo de mercado acarreado por las series modeladas a través de modelos ARCH o GARCH. Así, analizaremos cómo realizar pronósticos de volatilidad a partir del ajuste de dicha clase de procesos, lo cual nos permitirá atacar un problema fundamental para nuestros propósitos: la estimación medidas de riesgo condicionales a partir del ajuste de un modelo dinámico de series de tiempo a datos de origen financiero. Las medidas de riesgo de nuestro interés serán el *VaR* y el *Expected Shortfall*.

De esta manera, concluiremos este Capítulo estudiando cómo evaluar la calidad de las estimaciones de medidas de riesgo condicionales obtenidas a partir de modelos ARCH y GARCH; dichas evaluaciones se realizarán con base en una metodología estudiada en el Capítulo anterior: *backtesting*.

El contenido de este Capítulo se basó en [9], [12] y [13]. Bibliografía adicional se cita a lo largo del Capítulo, en las secciones pertinentes.

2.1. Bases para el análisis de series de tiempo

Esta Sección provee un breve resumen de los elementos esenciales que deben tenerse en consideración al analizar datos que conformen series de tiempo univariadas; en particular, se documentan diversas definiciones básicas para entender los modelos probabilísticos fundamentales de series de tiempo, se estudia una familia fundamental en el análisis básico de series de tiempo (los procesos ARMA) y se expone estructura de una metodología básica de análisis de series temporales.

2.1.1. Definiciones básicas

Un modelo de series de tiempo para las variaciones en el valor de un activo es un proceso estocástico $\{X_t\}_{t \in \mathbb{Z}}$, es decir, una familia de variables aleatorias indexadas en los enteros y definidas sobre algún espacio de probabilidad (Ω, \mathcal{F}, P) .

Momentos de una serie de tiempo. Suponiendo que existen, definimos la *función de media* $\mu(t)$ y la *función de autocovarianza* de $\{X_t\}_{t \in \mathbb{Z}}$, $\gamma(t, s)$, como:

$$\mu(t) = E(X_t), \quad t \in \mathbb{Z}$$

$$\gamma(t, s) = E((X_t - \mu(t))(X_s - \mu(s))), \quad t, s \in \mathbb{Z}.$$

De la definición anterior se sigue que la función de autocovarianza satisface $\gamma(t, s) = \gamma(s, t)$ para toda t, s y que, en particular, $\gamma(t, t) = \text{Var}(X_t)$.

Estacionariedad. Generalmente, los procesos que consideremos en la práctica serán estacionarios, en alguno o ambos de los siguientes dos sentidos:

Definición 2.1. Estacionariedad Estricta. Decimos que la serie de tiempo $\{X_t\}_{t \in \mathbb{Z}}$ es estrictamente estacionaria si

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+k}, \dots, X_{t_n+k})$$

Para toda $t_1, \dots, t_n, k \in \mathbb{Z}$ y para toda $n \in \mathbb{N}$.

Definición 2.2. Estacionariedad en Covarianza. La serie de tiempo $\{X_t\}_{t \in \mathbb{Z}}$ es estacionaria en covarianza (o estacionaria de segundo orden o débilmente estacionaria) si sus primeros dos momentos existen y satisfacen:

$$\mu(t) = \mu, \quad t \in \mathbb{Z}$$

$$\gamma(t, s) = \gamma(t+k, s+k), \quad t, s, k \in \mathbb{Z}.$$

Estas dos definiciones tratan de formalizar matemáticamente la noción de que el comportamiento de una serie de tiempo es similar en cualquier época en pudiéramos observarla. Cambios sistemáticos de media, varianza o covarianzas entre observaciones equidistantes son inconsistentes con el concepto de estacionariedad.

Con las definiciones anteriores, el hecho de que una serie de tiempo estrictamente estacionaria con varianza finita (segundo momento finito) es estacionaria en covarianza puede ser verificado de forma sencilla; no obstante, es importante notar que pueden existir procesos con varianza infinita que sean estrictamente estacionarios pero no estacionarios en covarianza. Es decir, el hecho de que un proceso sea estrictamente estacionario, sin hipótesis adicionales, no garantiza que tal proceso sea estacionario en covarianza; la hipótesis adicional de que éste tenga segundo momento finito es necesaria.

Autocorrelación en series de tiempo estacionarias. La definición de estacionariedad en covarianza implica que para toda s, t tenemos $\gamma(t-s, 0) = \gamma(t, s) = \gamma(s, t) = \gamma(s-t, 0)$, por lo que la covarianza entre X_t y X_s sólo depende de su separación temporal $|t-s|$, que es conocida como el retraso (*lag*). Por lo tanto, para procesos estacionarios en covarianza podemos escribir la función de autocovarianza como una función de una sola variable:

$$\gamma(h) := \gamma(h, 0), \quad h \in \mathbb{Z}.$$

Notando que $\gamma(0) = \text{Var}(X_t)$, para toda t , podemos definir ahora la función de autocorrelación para procesos estacionarios en covarianza.

Definición 2.3. Función de Autocorrelación. La función de autocorrelación (ACF por sus siglas en inglés) $\rho(h)$ de un proceso estacionario en covarianza $\{X_t\}_{t \in \mathbb{Z}}$ es:

$$\rho(h) = \rho(X_h, X_0) = \gamma(h)/\gamma(0), \quad \forall h \in \mathbb{Z}$$

Hablamos de la autocorrelación o *correlación serial* $\rho(h)$ para un retraso de h periodos de tiempo. En el análisis clásico de series de tiempo, el estudio en conjunto de correlaciones seriales y sus análogos empíricos, estimados de los datos, son el principal objeto de interés. Al estudio de las autocorrelaciones se le conoce como *análisis en el dominio del tiempo*.

Autocorrelación parcial en series de tiempo estacionarias. La función de autocorrelación nos brinda información sobre la asociación entre valores de una serie de tiempo separados por k periodos; por ejemplo X_t y X_{t-k} . Sin embargo, dicha función no toma en consideración que dicha asociación puede ser atribuida a correlaciones entre X_t y X_{t-k} y los valores intermedios entre ellos y no necesariamente a una asociación directa entre dichas variables. Es decir, no se considera hasta qué punto la asociación entre X_t y X_{t-k} puede atribuirse más al hecho de que ambos estén asociados con los valores $X_{t-1}, \dots, X_{t-k+1}$ y no a que ellos estén directamente correlacionados. La función de autocorrelación parcial toma este hecho en consideración, y nos permite calcular la asociación entre X_t y X_{t-k} eliminando las posibles contribuciones de las variables intermedias a dicha correlación.

Definición 2.4. Función de Autocorrelación Parcial. La función de autocorrelación parcial (PACF por el inglés *Partial Autocorrelation Function*) se define como el parámetro $\varphi_{h,h}$ en la regresión

$$\tilde{X}_t = \varphi_{h,1}\tilde{X}_{t-1} + \varphi_{h,2}\tilde{X}_{t-2} + \dots + \varphi_{h,h}\tilde{X}_{t-h} + Z_t, \quad (2.1)$$

donde $\tilde{X}_t := X_t - \mu(t)$ y Z_t es independiente de \tilde{X}_{t-i} , para $i \geq 1$.

Notemos que la regresión (2.1) puede ser reescrita como $X_t = \varphi_{h,0} + \varphi_{h,1}X_{t-1} + \varphi_{h,2}X_{t-2} + \dots + \varphi_{h,h}X_{t-h} + Z_t$, donde $\varphi_{h,0} := \mu(t)(1 - \varphi_{h,1} - \varphi_{h,2} - \dots - \varphi_{h,h})$. Esto nos permite ver claramente que $\varphi_{h,h}$ representa, de hecho, el efecto parcial (o *ceteris paribus*) de X_{t-h} sobre X_t .

Por lo tanto $\varphi_{h,h}$ es una medida del grado de asociación lineal entre dos componentes cualesquiera de $\{X_t\}_{t \in \mathbb{Z}}$ separados entre sí por h observaciones (por ejemplo, X_t y X_{t-h}) que no es debida a la correlación entre cada uno de ellos y todos los componentes de $\{X_t\}_{t \in \mathbb{Z}}$ que se encuentran entre ambos ($X_{t-1}, \dots, X_{t-h+1}$).

Observación 2.5. A partir de la regresión (2.1) puede comprobarse que $\varphi_{h,h} = |\mathbf{A}_h|/|\mathbf{B}_h|$, donde

$$\mathbf{A}_h = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(h-2) & \rho(1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(h-3) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(h-4) & \rho(3) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \cdots & \rho(1) & \rho(h) \end{pmatrix},$$

$$\mathbf{B}_h = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(h-2) & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(h-3) & \rho(h-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(h-4) & \rho(h-3) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \cdots & \rho(1) & 1 \end{pmatrix},$$

de forma que los coeficientes de autocorrelación parcial pueden ser calculados en función de los coeficientes de autocorrelación simple de un proceso estacionario en covarianza. De este modo, por ejemplo, para $k = 1, 2$ tenemos:

$$\varphi_{1,1} = \rho(1), \quad \varphi_{2,2} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$$

Procesos Ruido Blanco. Los bloques de construcción básicos para la creación de modelos de series de tiempo útiles son procesos estacionarios sin correlación serial (no autocorrelacionados); los cuales son conocidos como procesos *ruido blanco* y están definidos como sigue:

Definición 2.6. Ruido Blanco. $\{X_t\}_{t \in \mathbb{Z}}$ es un proceso Ruido Blanco si es estacionario en covarianza y tiene función de autocorrelación

$$\gamma(h) = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

Un proceso Ruido Blanco centrado para tener media cero y con varianza $\sigma^2 = \text{Var}(X_t)$ será denotado $RB(0, \sigma^2)$. Un ejemplo simple de un proceso Ruido Blanco es el de una serie de variables aleatorias i.i.d. con varianza finita; dichos procesos son conocidos como *Ruido Blanco Estricto*.

Definición 2.7. Ruido Blanco Estricto. $\{X_t\}_{t \in \mathbb{Z}}$ es un proceso Ruido Blanco Estricto si es una serie de variables aleatorias i.i.d. con varianza finita.

Un proceso Ruido Blanco Estricto (RBE) centrado para tener media cero y con varianza σ^2 será denotado $RBE(0, \sigma^2)$. Aunque el RBE es el proceso ruido más fácil de entender, este no es el único ruido que utilizaremos. Más tarde veremos que los procesos ARCH y GARCH estacionarios en covarianza son, de hecho, ruidos blancos.

Diferencia de Martingalas. Un concepto adicional de ruido que utilizaremos, en particular cuando discutamos volatilidad y procesos GARCH, es el de una sucesión con la propiedad de *diferencia de martingalas*. Para discutir este concepto, tomaremos la hipótesis adicional de que la serie $\{X_t\}_{t \in \mathbb{Z}}$ es adaptada a alguna *filtración* $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ que representa la *acumulación de información en el tiempo*. La σ -álgebra \mathcal{F}_t representa la información disponible al tiempo t y, típicamente, esta será la información contenida en los valores pasados y presente de la serie de tiempo misma $\{X_s\}_{s \leq t}$, a la cual nos referimos como la *historia* hasta el tiempo t y denotamos por $\mathcal{F}_t = \sigma(\{X_s, s \leq t\})$; la filtración correspondiente es conocida como la *filtración natural*.

Para una sucesión de variables aleatorias con la propiedad de diferencia de martingalas, la esperanza del siguiente valor de la sucesión, dada la información actual, es siempre cero. Se suele decir que las diferencias de martingalas modelan nuestras ganancias en ensayos consecutivos de un *juego justo*.

Definición 2.8. Diferencia de Martingalas. Decimos que la serie de tiempo $\{X_t\}_{t \in \mathbb{Z}}$ tiene la propiedad de *diferencia de martingalas* (o que es una *diferencia de martingalas*) con respecto a la filtración $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ si $E(X_t) < \infty$, X_t es \mathcal{F}_t -medible (adaptada) y

$$E(X_t | \mathcal{F}_{t-1}) = 0$$

Nótese que de la propiedad anterior se sigue que la esperanza no condicional de este proceso también es cero:

$$E(E(X_t | \mathcal{F}_{t-1})) = 0.$$

Más aún, si $E(X_t^2) < \infty$ para toda t , entonces las autocovarianzas satisfacen que:

$$\gamma(t, s) = E(X_t X_s) = \begin{cases} E(E(X_t X_s | \mathcal{F}_{s-1})) = E(X_t E(X_s | \mathcal{F}_{s-1})) = 0, & t < s \\ E(E(X_s X_t | \mathcal{F}_{t-1})) = E(X_s E(X_t | \mathcal{F}_{t-1})) = 0, & s < t \end{cases}$$

Por lo tanto, una diferencia de martingalas con varianza finita tiene media y covarianza cero. Con lo anterior en mente, vemos que si la varianza de una diferencia de martingalas es constante para toda t , dicho proceso es un ruido blanco.

2.1.2. Procesos ARMA

La familia de los procesos ARMA clásicos es ampliamente utilizada en muchas aplicaciones tradicionales de series de tiempo. Los modelos ARMA son procesos estacionarios en covarianza que son construidos utilizando procesos ruido blanco como bloque básico. Utilizando la convención general de notación en literatura sobre este tipo de procesos, de aquí en adelante denotaremos el ruido blanco por $\{\epsilon_t\}_{t \in \mathbb{Z}}$ y al ruido blanco *estricto* por $\{Z_t\}_{t \in \mathbb{Z}}$.

Los procesos ARMA consisten de dos partes, una parte autoregresiva (AR) y una parte media móvil (MA). Estos procesos son usualmente denotados como $\text{ARMA}(p, q)$, donde p es el orden la parte autoregresiva y q es el orden de la parte media móvil. Si p o q toman el valor cero, los procesos resultantes ARMA se suelen denotar simplemente por $\text{MA}(q)$ y $\text{AR}(p)$, respectivamente.

Definición 2.9. Procesos ARMA. Sea $\{\epsilon_t\}_{t \in \mathbb{Z}}$ un $\text{RB}(0, \sigma_\epsilon^2)$. El proceso $\{X_t\}_{t \in \mathbb{Z}}$ es un proceso $\text{ARMA}(p, q)$ con media cero si es estacionario en covarianza y satisface las ecuaciones

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \quad \forall t \in \mathbb{Z}, \quad (2.2)$$

$\{X_t\}_{t \in \mathbb{Z}}$ es un proceso $\text{ARMA}(p, q)$ con media μ si el proceso centrado $\{X_t - \mu\}_{t \in \mathbb{Z}}$ es un proceso $\text{ARMA}(p, q)$ con media cero.

De acuerdo con la definición anterior, no existe tal cosa como un proceso ARMA que no sea estacionario en covarianza. Si el proceso es estrictamente estacionario o no dependerá de la naturaleza exacta del proceso ruido blanco que lo conduce, $\{\epsilon_t\}_{t \in \mathbb{Z}}$, también conocido como proceso de *innovaciones*, proceso de choques o proceso de errores. Si las innovaciones son i.i.d., o forman un proceso estrictamente estacionario ellas mismas, entonces el proceso ARMA también será estrictamente estacionario.

Para todo fin práctico, podemos restringir nuestro estudio de los procesos ARMA a los procesos ARMA *causales*. Con lo anterior, nos referimos a procesos que satisfacen la ecuación (2.2) que tienen una representación de la forma:

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (2.3)$$

Donde los ψ_i son coeficientes que deben satisfacer

$$\sum_{i=0}^{\infty} |\psi_i| < \infty \quad (2.4)$$

Observación 2.10. *La también llamada condición de sumabilidad absoluta en (2.4) es una condición técnica que asegura que $E|X_t| < \infty$. Esto garantiza que la suma infinita (2.3) converja absolutamente, casi seguramente. Lo cual significa que ambas $\sum_{i=0}^{\infty} |\psi_i| |\epsilon_{t-1}|$ y $X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ son finitas con probabilidad 1.⁹ Adicionalmente, notemos que si un proceso ARMA es causal, este puede ser escrito como un proceso MA(∞).*

La siguiente proposición garantiza que todos los procesos ARMA causales son estacionarios en covarianza, a la vez que nos da la forma de su función de autocorrelación.

Proposición 2.11. *Cualquier proceso que satisfaga (2.3) y (2.4) es estacionario en covarianza y tiene función de autocorrelación*

$$\rho(h) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}}{\sum_{i=0}^{\infty} \psi_i^2} \quad (2.5)$$

Demostración. Para toda t tenemos que $E(X_t) = 0$ y $Var(X_t) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty$, debido a (2.4). Más aún, las autocovarianzas están dadas por

$$cov(X_t, X_{t+h}) = E \left(\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \epsilon_{t+h-j} \right).$$

Puesto que $\{\epsilon_t\}_{t \in \mathbb{Z}}$ es un ruido blanco, se sigue que $E(\epsilon_{t-i} \epsilon_{t+h-j}) \neq 0 \iff j = i + h$ y por lo tanto que:

$$\gamma(h) = cov(X_t, X_{t+h}) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}, \quad |h| \in \mathbb{Z}, \quad (2.6)$$

la cual depende solo del retraso h y no de t . La función de autocorrelación se sigue de forma sencilla. \square

⁹ La demostración de este hecho puede consultarse en [12].

Observación 2.12. Teoría los Procesos ARMA generales. En el caso del proceso ARMA general de la Definición 2.9, el problema de si este tiene o no una representación causal de la forma (2.3) es resuelto mediante el estudio de dos polinomios en el plano complejo que están dados en términos de los parámetros del modelo ARMA como

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

$$\tilde{\theta}(z) = 1 + \theta_1 z + \dots + \theta_q z^q.$$

Provisto que $\tilde{\phi}(z)$ y $\tilde{\theta}(z)$ no tienen raíces comunes, entonces el proceso ARMA es un proceso causal que satisface (2.3) y (2.4) si y sólo si $\tilde{\phi}(z)$ no tiene raíces en el círculo unitario $|z| \leq 1$. Los coeficientes ψ_i en la representación (2.3) están determinados por la ecuación

$$\sum_{i=0}^{\infty} \psi_i z^i = \frac{\tilde{\theta}(z)}{\tilde{\phi}(z)}, \quad |z| < 1.$$

Invertibilidad. Un proceso ARMA(p, q) se denomina invertible si existen constantes π_j tales que $\sum_{j=1}^{\infty} \pi_j < \infty$ y

$$\epsilon_t = \sum_{j=1}^{\infty} \pi_j X_{t-j}. \quad (2.7)$$

Que un proceso ARMA sea invertible puede interpretarse como que la información de las innovaciones pasadas y actuales es equivalente a la información de nuestras observaciones pasadas y actuales (por lo que podemos recuperar una a partir de la otra).

Los modelos que ajustamos en la práctica a datos reales son soluciones tanto causales como invertibles de las ecuaciones que definen los procesos ARMA. Esto se debe a que si un proceso ARMA satisface dichas condiciones, entonces el desarrollo de dicho proceso depende únicamente de su propia historia, por lo que al estudiar su pasado, podríamos encontrar una forma razonable de predecir su comportamiento en el futuro.¹⁰

Modelos para esperanza condicional. Consideremos un modelo ARMA invertible general con media distinta de cero. Para lo que viene después, será útil

¹⁰ Para conocer más sobre las implicaciones de la invertibilidad y causalidad de un proceso ARMA, referimos al lector a [12], donde podrá encontrar una amplia discusión del tema.

observar que podemos escribir tales modelos como

$$X_t = \mu_t + \epsilon_t, \quad \mu_t = \mu + \sum_{i=1}^p \phi_i (X_{t-i} - \mu) + \sum_{j=1}^q \theta_j \epsilon_{t-j} \quad (2.8)$$

Puesto que hemos supuesto invertibilidad, el término ϵ_{t-j} , y por tanto μ_t , puede ser escrito en términos del pasado infinito del proceso hasta el tiempo $t-1$; se dice que μ_t es *medible* con respecto a $\mathcal{F}_{t-1} = \sigma(\{X_s, s \leq t-1\})$.

Si tomamos la hipótesis de que el ruido blanco $\{\epsilon_t\}_{t \in \mathbb{Z}}$ es una diferencia de martingalas con respecto a $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$, entonces $E(X_t | \mathcal{F}_{t-1}) = \mu_t$. En otras palabras, un proceso ARMA como este puede pensarse como colocar una estructura particular en la esperanza condicional μ_t del proceso. Se verá más adelante que los procesos ARCH y GARCH pueden ser vistos como colocar una estructura en la varianza condicional $Var(X_t | \mathcal{F}_{t-1})$.

2.1.3. Metodología básica para el análisis estadístico de series de tiempo

En esta Sección describiremos las etapas de un análisis general de datos de origen financiero a través de modelos de series de tiempo. Dicha descripción estará enfocada al análisis de series temporales vía procesos ARMA; no obstante, la estructura de la metodología puede ser utilizada para ajustar otro tipo de modelos de series de tiempo.

Análisis preliminar. Los datos son graficados y la plausibilidad de un modelo estacionario es considerada. Puesto que en este trabajo nos concentramos en series con valores de diferencias de logaritmos, supondremos que, a lo más, una manipulación menor de nuestros datos es requerida. El análisis clásico de series de tiempo cuenta con diversas técnicas para remover *tendencias y estacionalidades* de datos “no estacionarios”; estas técnicas son discutidas con profundidad en diversa bibliografía, incluyendo [14]. Así, si bien es cierto que determinados tipos de series financieras efectivamente muestran patrones estacionales, como series de ingresos y egresos, supondremos que tales efectos son relativamente menores en las series de rendimientos diarios o semanales, que son la base para los métodos de administración de riesgos.

Claramente, la hipótesis de estacionariedad resulta más cuestionable si consideramos ventanas temporales grandes, o en las cuales han tomado lugar cambios de política económica excepcionales. Aunque los mercados cambian constantemente, siempre habrá una tensión entre nuestro deseo de utilizar

la información más reciente posible y nuestra necesidad de incluir suficientes datos para tener precisión en la estimación estadística. Si medio año de información, cinco años o 10 años son apropiados dependerá de la situación. Por lo tanto, realizar varios análisis con distintas ventanas de tiempo e investigar la sensibilidad de la inferencia estadística a la cantidad de información es, ciertamente, una buena idea.

Análisis en el dominio del tiempo. Una vez definida la información a utilizar, es momento de realizar un análisis en el dominio del tiempo de los datos; este análisis involucra el cálculo de estimadores empíricos de autocovarianzas y autocorrelaciones a partir de los datos para realizar inferencia sobre la estructura de dependencia serial que éstos guardan entre sí. Aplicando correlogramas, correlogramas parciales y pruebas portmanteau, como la de Ljung-Box, a los datos crudos y sus valores absolutos (o al cuadrado), la hipótesis de que estos formen un RBE puede ser evaluada.¹¹ Si dicha hipótesis no puede ser rechazada para los datos en consideración, entonces el análisis formal de series de tiempo termina y el ajuste simple de una distribución podría ser utilizado en lugar de un modelaje dinámico.

Para series de rendimientos diarios esperamos rechazar rápidamente la hipótesis de RBE. A pesar del hecho de que los correlogramas de datos crudos pueden mostrar poca evidencia de una correlación serial, es posible que los correlogramas de los valores absolutos o al cuadrado de los datos muestren evidencia de una fuerte dependencia serial. En otras palabras, los datos podrían soportar la hipótesis de ser un ruido blanco, pero no la de ser un ruido blanco estricto. En tal caso, los modelos ARMA no son requeridos, pero los modelos de la siguiente Sección podrían ser útiles.

No obstante, si los correlogramas proveen evidencia de patrones de correlación serial producidos por procesos ARMA, entonces podríamos intentar ajustar un modelo ARMA a nuestros datos.

Ajuste del modelo. Un enfoque tradicional para ajustar modelos ARMA consiste en *identificar el orden* de un proceso ARMA adecuado para nuestros datos; para lo cual se suele utilizar la información provista por el correlograma y el correlograma parcial. Específicamente, esto se realiza con base en el comportamiento teórico de la ACF y la PACF de los modelos ARMA. Por ejemplo, un punto de corte de la ACF en el retraso $h = q$ podría considerarse evidencia de un comportamiento MA(q); en contraste, un corte de la PACF en el retraso $h = p$ podría interpretarse como evidencia de un comportamiento AR(p).

¹¹ Los detalles de la prueba de Ljung-Box pueden ser consultados en el Apéndice B.

Simulación Procesos ARMA

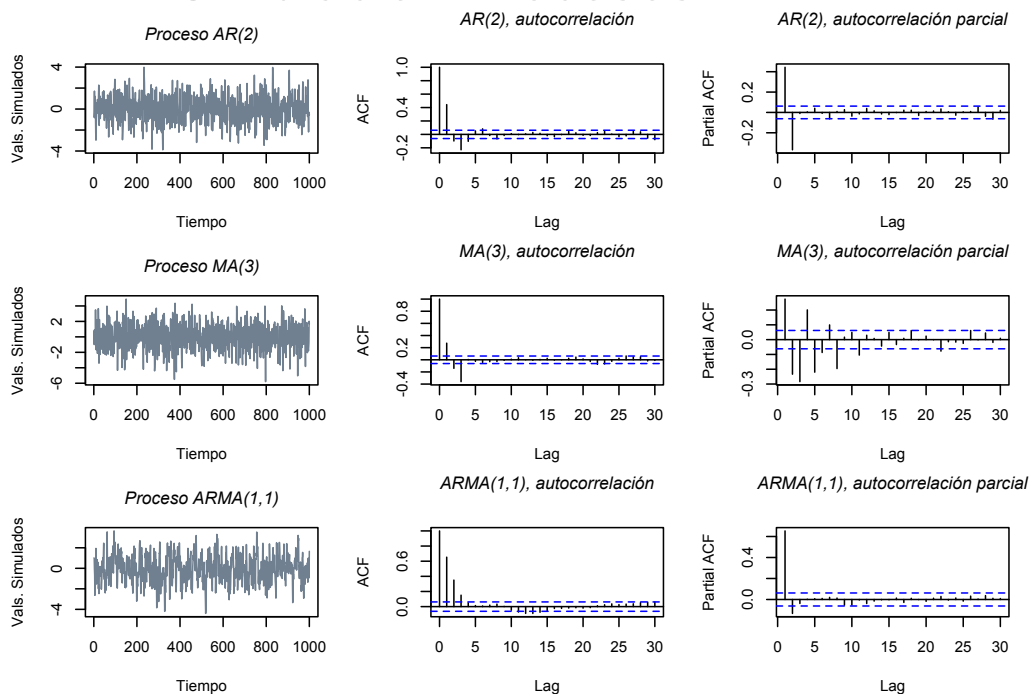


Figura 2.1: Ejemplos de Procesos ARMA. En el primer renglón (a) se tiene un modelo AR(2) con $\phi = 0.6, -0.4$; en el segundo renglón (b) se muestra un proceso MA(3) con $\theta = -0.7, 0.3, -0.9$; en el último renglón (c) se tiene un proceso ARMA(1,1) con $\phi = 0.5$ y $\theta = 0.2$. En todos los casos se pueden observar también las respectivas funciones de autocorrelación y autocorrelación parcial.

Con dicha metodología en mente, en la Tabla 2.1 se sintetiza el comportamiento teórico de las funciones de autocorrelación y autocorrelación parcial de modelos AR, MA y ARMA. Adicionalmente, en la Figura 2.1 se ilustra el comportamiento de dichos procesos; específicamente, se muestran las gráficas de las simulaciones de tres procesos ARMA (AR(2), MA(3) y ARMA(1,1)) y el comportamiento teórico de sus respectivas ACF y PACF.

Al trabajar con un conjunto de datos, uno podría estimar la ACF y PACF de dicha muestra y analizar si las gráficas de dichas funciones siguen alguno de los comportamientos descritos en la Tabla 2.1. De ser así, el ajuste de un modelo ARMA a los datos podría resultar razonable.

Una vez elegido el modelo a ajustar, existe una gran diversidad de métodos de ajuste, como método de máxima verosimilitud (condicional) estudiado

Modelos ARMA: Características teóricas de la ACF y PACF		
	ACF $\rho(h)$	PACF $\varphi_{h,h}$
AR(p)	Funciones sinusoidales o exponenciales amortiguadas.	Se vuelve cero para $h > p$.
MA(q)	Se vuelve cero para $h > q$.	Dominada por funciones sinusoidales o exponenciales amortiguadas.
ARMA(p, q)	Funciones exponenciales o sinusoidales.	Dominada por funciones sinusoidales o exponenciales amortiguadas.

Tabla 2.1: Comportamiento teórico de la función de autocorrelación y la función de autocorrelación parcial para modelo AR, MA y ARMA.

más adelante.

Análisis de residuales y comparación de modelos. Recordemos la representación (2.8) de un proceso ARMA causal e invertible y supongamos que hemos ajustado tal proceso y estimado los parámetros ϕ_i y θ_j . Los residuales son los valores inferidos $\hat{\epsilon}_t$ para las innovaciones no observadas ϵ_t y son calculados recursivamente de los datos y el modelo ajustado como

$$\hat{\epsilon}_t = X_t - \hat{\mu}_t, \quad \hat{\mu}_t = \hat{\mu} + \sum_{i=1}^p \hat{\phi}_i (X_{t-i} - \hat{\mu}) + \sum_{j=1}^q \hat{\theta}_j \hat{\epsilon}_{t-j} \quad (2.9)$$

donde los valores $\hat{\mu}_t$ algunas veces son conocidos como los *valores ajustados*. Obviamente, tenemos un problema al calcular los primeros pocos valores de $\hat{\epsilon}_t$ debido a la finitud de nuestra muestra y la naturaleza infinita de las recursiones en (2.9). Una de las muchas soluciones posibles a dicho problema podría ser fijar $\epsilon_{-\hat{q}+1} = \epsilon_{-\hat{q}+2} \cdots = \hat{\epsilon}_0 = 0$ y $X_{-p+1} = X_{-p+2} = \cdots = X_0 = \bar{X}$ y entonces utilizar (2.9) para $t = 1, \dots, n$, donde n es el tamaño de nuestra muestra. Puesto que los primeros pocos valores estarán influenciados por estos valores iniciales, estos podrían ser ignorados en análisis posteriores.

Los residuales $\{\hat{\epsilon}_t\}$ deberían comportarse como una realización de un proceso ruido blanco, puesto que esta es la hipótesis de nuestro modelo para las innovaciones. Dicha hipótesis puede ser evaluada al construir el correlograma y correlograma parcial de $\{\hat{\epsilon}_t\}$. Si aún hay evidencia de correlación o correlación parcial serial en los residuos, ello sugiere que no se ha encontrado todavía un modelo ARMA apropiado. Más aún, podemos utilizar pruebas portman-teau para verificar formalmente que los residuales se comportan como una

realización de un ruido blanco estricto. Si los residuales se comportan como un RBE, entonces no se requiere modelación adicional de series de tiempo; si se comportan como un RB, pero no un RBE, entonces los modelos de volatilidad de la siguiente Sección pueden ser utilizados.

Usualmente es posible encontrar más de un modelo ARMA razonable para nuestros datos, y técnicas formales de comparación de modelos pueden ser requeridas para elegir un mejor modelo global. El criterio de selección de Akaike descrito en el Apéndice B puede ser utilizado, o una de las diversas variantes de dicho criterio que son preferidas para series de tiempo.

Predicción. Existen muchos enfoques para pronosticar o realizar predicciones de series de tiempo; nosotros sintetizaremos dos de ellos, los cuales pueden extenderse de forma sencilla al caso de los modelos GARCH. La primera estrategia hace uso del modelo ARMA ajustado y es conocida como el *enfoque Box-Jenkins*. La segunda estrategia es un enfoque libre de modelo para pronosticar conocido como *suavizamiento exponencial*, el cual está relacionado con la técnica de promedios móviles exponenciales ponderados para predecir volatilidad que estudiaremos más adelante.

Predicción utilizando modelos ARMA. Consideremos un modelo ARMA y su representación (2.8). Sea \mathcal{F}_t la historia del proceso hasta el tiempo t , incluyendo t , y supongamos que las innovaciones $\{\epsilon_t\}_{t \in \mathbb{Z}}$ tienen la propiedad de diferencia de martingalas respecto a $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$.

Para nuestro problema de predicción, será conveniente denotar nuestra muestra de n observaciones por X_{t-n+1}, \dots, X_t . Supondremos que éstas son realizaciones de variables aleatorias que siguen un modelo ARMA particular. Nuestro objetivo es predecir X_{t+1} o de forma más general X_{t+h} , y denotamos nuestra predicción por $P_t X_{t+h}$. El método que describimos supone que tenemos acceso a la historia infinita del proceso hasta el tiempo t y deriva un fórmula que es entonces aproximada por nuestra muestra finita.

Como predictor de X_{t+h} , utilizamos la esperanza condicional $E(X_{t+h} | \mathcal{F}_t)$. Entre todas las predicciones $P_t X_{t+h}$ basadas en la historia infinita del proceso hasta el tiempo t , este predictor minimiza el error cuadrático medio de predicción $E((X_{t+h} - P_t X_{t+h})^2)$.

La idea básica es que, para $h \geq 1$, la predicción $E(X_{t+h} | \mathcal{F}_t)$ sea evaluada recursivamente en términos de $E(X_{t+h-1} | \mathcal{F}_t)$. Utilizamos el hecho de que $E(\epsilon_t | \mathcal{F}_t) = 0$ (la propiedad de diferencia de martingalas de las innovaciones) y que las variables aleatorias $\{X_s\}_{s \leq t}$ y $\{\epsilon_s\}_{s \leq t}$ son “conocidas” al tiempo t . La hipótesis de invertibilidad (2.7) asegura que la innovación ϵ_t puede ser reescrita como una función de la historia infinita del proceso $\{X_s\}_{s \leq t}$. Para

ilustrar este enfoque será suficiente considerar un modelo ARMA(1,1), la generalización a modelos ARMA(p, q) se sigue fácilmente.

Ejemplo 2.13. Predicción para el modelo ARMA(1,1). Supongamos que un modelo ARMA(1,1) de la forma (2.8) ha sido ajustado a un conjunto de información y que sus parámetros μ , ϕ_1 y θ_1 han sido determinados. Nuestra predicción para X_{t+1} es:

$$E(X_{t+1} | \mathcal{F}_t) = \mu_{t+1} = \phi_1(X_t - \mu) + \theta_1\epsilon_t,$$

puesto que $E(\epsilon_{t+1} | \mathcal{F}_t) = 0$. Para predicciones a dos pasos obtenemos:

$$\begin{aligned} E(X_{t+2} | \mathcal{F}_t) &= E(\mu_{t+2} | \mathcal{F}_t) = \mu + \phi_1(E(X_{t+1} | \mathcal{F}_t) - \mu) \\ &= \mu + \phi_1^2(X_t - \mu) + \phi_1\theta_1\epsilon_t \end{aligned}$$

y por lo tanto en general tenemos que:

$$E(X_{t+h} | \mathcal{F}_t) = \mu + \phi_1^h(X_t - \mu) + \phi_1\theta_1\epsilon_t.$$

Sin conocer todos los valores históricos $\{X_s\}_{s \leq t}$ este predictor no puede ser evaluado de forma exacta, pero puede ser aproximado de forma precisa si n es razonablemente grande. La forma más sencilla de hacer esto es sustituir el residual del modelo $\hat{\epsilon}_t$, calculado de (2.9), en lugar de ϵ_t . Notemos que $\lim_{h \rightarrow \infty} E(X_{t+h} | \mathcal{F}_t) = \mu$, casi seguramente, por lo que la predicción converge a la media no condicional del proceso para horizontes de tiempo mayores.

Suavizamiento exponencial. Esta es una técnica popular utilizada tanto para predicción en series de tiempo como para estimación de tendencias. Bajo este enfoque no suponemos necesariamente que los datos provienen de un modelo estacionario, aunque sí suponemos que no existe una componente determinística estacional en el modelo. El método es menos apropiado, en general, para series de rendimientos con cambios frecuentes de signo y resulta mejor para series de valores o precios no diferenciadas. Adicionalmente, esta técnica forma la base de una metodología muy común para predicción de volatilidad que se estudia más adelante.

De esta manera, supongamos que nuestros datos representan realizaciones de variables aleatorias Y_{t-n+1}, \dots, Y_t , sin hacer referencia a algún modelo paramétrico en particular. Como pronóstico para Y_{t+1} utilizamos una predicción de la forma:

$$P_t Y_{t+1} = \sum_{i=0}^{n-1} \lambda(1-\lambda)^i Y_{t-i}, \quad \text{donde } 0 < \lambda < 1.$$

Así, ponderamos la información de la más reciente a la más lejana con una sucesión exponencialmente decreciente que suma casi uno. Por lo tanto, podemos calcular de forma sencilla que:

$$\begin{aligned} P_t Y_{t+1} &= \sum_{i=0}^{n-1} \lambda(1-\lambda)^i Y_{t-i} = \lambda Y_t + (1-\lambda) \sum_{j=0}^{n-2} \lambda(1-\lambda)^j Y_{t-1-j} \\ &= \lambda Y_t + (1-\lambda) P_{t-1} Y_t, \end{aligned} \tag{2.10}$$

por lo que la predicción al tiempo t es obtenida de la predicción al tiempo $t-1$ a través de un simple sistema recursivo. La elección de λ es subjetiva; mientras más grande sea su valor, más peso es colocado en las observaciones más recientes. No obstante, es posible utilizar diversos estudios de validación empírica para determinar un valor de λ que de buenos resultados.

2.2. Modelos de volatilidad estocástica

En esta Sección estudiaremos los modelos más importantes para series de tiempo financieras univariadas de este trabajo; aquellos capaces de capturar la dinámica de la volatilidad de una serie de tiempo financiera. Dicho estudio seguirá, a grandes rasgos, la siguiente estructura.

Primeramente, estudiaremos las definiciones de los modelos ARCH (Auto-Regresivos de Heterocedasticidad Condicional) y GARCH (ARCH Generalizados), y analizaremos diversas propiedades matemáticas de dicha clase de procesos.

Después, estudiaremos cómo utilizar modelos ARCH y GARCH en la práctica; específicamente, discutiremos cómo identificar datos que podrían ser modelados dicha clase de procesos, cómo estimar los parámetros de estos modelos a partir de datos históricos y cómo elegir el mejor proceso ARCH o GARCH para un conjunto de datos.

2.2.1. Procesos ARCH

Definición 2.14. Procesos ARCH. Sea $\{Z_t\}_{t \in \mathbb{Z}}$ un RBE(0,1). El proceso $\{X_t\}_{t \in \mathbb{Z}}$ es un proceso ARCH(p) si es estrictamente estacionario y satisface, para toda $t \in \mathbb{Z}$ y algún proceso que solo tome valores estrictamente positivos $\{\sigma_t\}_{t \in \mathbb{Z}}$, las ecuaciones:

$$X_t = \sigma_t Z_t \tag{2.11}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 \quad (2.12)$$

Donde $\alpha_0 > 0$ y $\alpha_i \geq 0$, $1 \leq i \leq p$.

Nuevamente denotamos por \mathcal{F}_t a la σ -álgebra que representa la historia del proceso hasta el tiempo t , de modo que $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ sea la filtración natural del proceso. Claramente, la construcción de (2.12) garantiza que σ_t es medible respecto a \mathcal{F}_{t-1} . Esto nos permite calcular, provisto que $E|X_t| < \infty$, que

$$E(X_t | \mathcal{F}_{t-1}) = E(\sigma_t Z_t | \mathcal{F}_{t-1}) = \sigma_t E(Z_t | \mathcal{F}_{t-1}) = 0 \quad (2.13)$$

De modo que el proceso ARCH tiene la propiedad de diferencia de martingalas con respecto a $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$. Por lo tanto, si el proceso es estacionario en covarianza, entonces simplemente es un ruido blanco, como discutiremos más adelante.

Observación 2.15. *Notemos que la independencia entre Z_t y \mathcal{F}_t que supusimos arriba se sigue del hecho de que un proceso ARCH tiene que ser causal, es decir, las ecuaciones (2.11) y (2.12) deben tener una solución de la forma $X_t = f(Z_t, Z_{t-1}, \dots)$ para alguna f , por lo que Z_t es independiente de los valores anteriores del proceso. Esto contrasta con los procesos ARMA, que pueden tener soluciones no causales.*

Si simplemente suponemos que el proceso es un ruido blanco estacionario en covarianza (hecho para el cual daremos condiciones más adelante), entonces $E(X_t^2) < \infty$ y

$$\text{Var}(X_t | \mathcal{F}_{t-1}) = E(\sigma_t^2 Z_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 \text{Var}(Z_t | \mathcal{F}_{t-1}) = \sigma_t^2.$$

Por lo tanto, este modelo tiene la interesante propiedad de que su desviación estándar condicional, o *volatilidad*, es una función continuamente cambiante de los valores anteriores del proceso elevados al cuadrado. Si uno o más de $|X_{t-1}|, \dots, |X_{t-p}|$ son particularmente *grandes*, entonces X_t está efectivamente dado por una distribución con una varianza más grande y podría en sí mismo ser grande; en este sentido, el modelo genera *clusters* (cúmulos) de volatilidad. El nombre ARCH se refiere a esta estructura: el modelo es Auto-Regresivo pues X_t claramente depende de los anteriores X_{t-i} , y condicionalmente *heterocedástico*, pues la varianza condicional del proceso cambia continuamente.

La distribución de las innovaciones $\{Z_t\}_{t \in \mathbb{Z}}$ puede ser, en principio, cualquier distribución con media cero y varianza 1. Para fines de ajuste estadístico

Simulación ARCH(1)

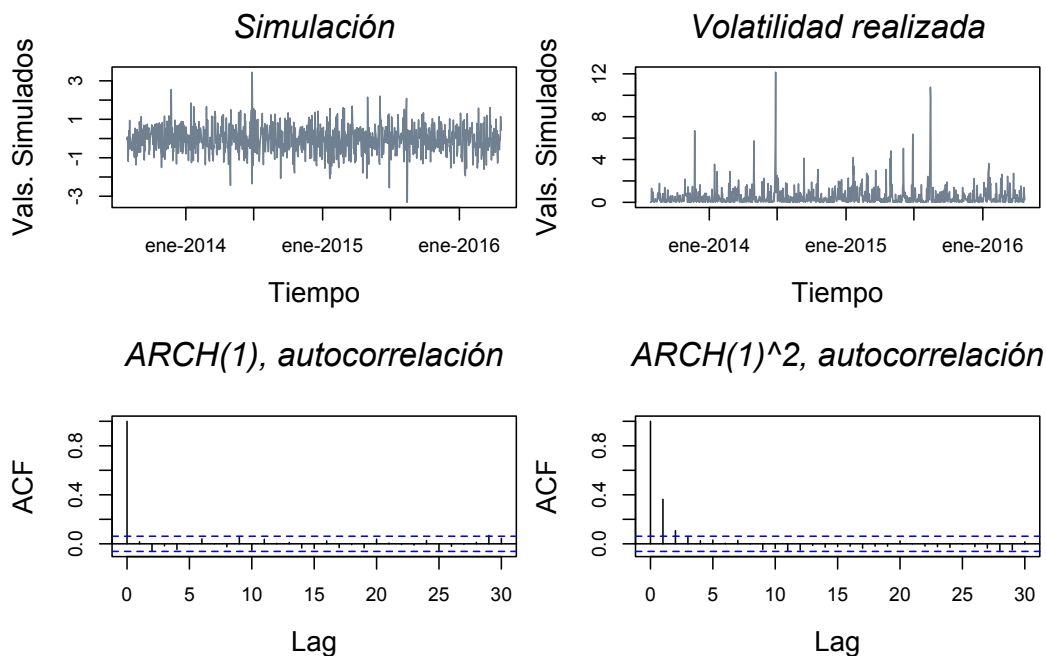


Figura 2.2: Ejemplo de un proceso ARCH(1) cuyos valores al cuadrado siguen un proceso AR(1). En el primer cuadrante se observa la simulación del proceso; en el segundo, la volatilidad realizada del mismo; en el tercero la función de autocorrelación de los valores simulados crudos; en el cuarto, la función de autocorrelación de los valores simulados al cuadrado.

podríamos o no especificar la distribución, dependiendo de si estamos implementando un método de ajuste no paramétrico, de máxima verosimilitud (MV) o de quasi-máxima verosimilitud (QMV). Para el enfoque MV las elecciones más comunes son innovaciones con una distribución normal estándar o t escaladas. Para la segunda queremos decir $Z_t \sim t_1(\nu, 0, (\nu - 2)/\nu)$, de modo que la varianza de la distribución sea 1. Mantendremos estas elecciones en mente cuando discutamos propiedades adicionales de los modelos ARCH y GARCH.

El modelo ARCH(1). En el resto de esta Sección analizaremos algunas propiedades del modelo ARCH(1). Estas propiedades se extienden a toda la familia de modelos ARCH y GARCH, pero son introducidas de forma más sencilla en el caso más simple ($p = 1$). En la Figura 2.2, se puede observar la simulación de un proceso ARCH(1) con innovaciones Gaussianas con la realización del proceso de volatilidad correspondiente. Utilizando $X_t^2 = \sigma_t^2 Z_t^2$

y (2.12) en el caso $p = 1$, deducimos que el cuadrado del proceso ARCH(1) satisface:

$$X_t^2 = \alpha_0 Z_t^2 + \alpha_1 Z_t^2 X_{t-1}^2. \quad (2.14)$$

Un análisis matemático detallado del proceso ARCH(1) involucra el estudio de la ecuación (2.14), la cual es una *ecuación estocástica recurrente* (EER). Como en el caso de los modelos ARMA, nos gustaría saber cuándo es que esta ecuación tiene soluciones estacionarias expresadas en términos de la historia infinita de las innovaciones, es decir, soluciones de la forma $X_t^2 = f(Z_{t-1}, Z_{t-2}, \dots)$.

Para los modelos ARCH debemos distinguir cuidadosamente entre soluciones estrictamente estacionarias y soluciones estacionarias en covarianza; es posible tener modelos ARCH(1) con varianza infinita, que obviamente no son estacionarios en covarianza.

Ecuaciones Estocásticas Recurrentes. La teoría requerida para analizar relaciones estocásticas recurrentes de la forma de (2.14) está fuera del alcance y propósitos de este trabajo. No obstante, a continuación se proveen breves notas para indicar las ideas involucradas en dicha clase de estudios. Los análisis siguientes están basados en ideas de [15] y [16].

La ecuación (2.14) es un ejemplo particular de una clase de ecuaciones recurrentes de la forma

$$Y_t = A_t Y_{t-1} + B_t, \quad (2.15)$$

donde $\{A_t\}_{t \in \mathbb{Z}}$ y $\{B_t\}_{t \in \mathbb{Z}}$ son sucesiones de variables aleatorias i.i.d. Condiciones suficientes para tener una solución son:

$$E(\max\{0, \ln |B_t|\}) < \infty, \quad y \quad E(\ln |A_t|) < 0. \quad (2.16)$$

La solución única está dada por:

$$Y_t = B_t + \sum_{i=1}^{\infty} B_{t-i} \prod_{j=0}^{i-1} A_{t-j}, \quad (2.17)$$

donde la suma converge absolutamente, casi seguramente.

Podemos generar algo de intuición sobre las condiciones en (2.16) y la forma de la solución (2.17) realizando iteraciones sobre la ecuación (2.15) k veces:

$$Y_t = A_t(A_{t-1}Y_{t-2} + B_{t-1}) + B_t = B_t + \sum_{i=1}^k B_{t-i} \prod_{j=1}^{i-1} A_{t-j} + Y_{t-k-1} \prod_{i=0}^k A_{t-i}$$

Las condiciones en (2.16) aseguran que el término de en medio de la última igualdad converja y que el término final desaparezca. En particular notemos que

$$\frac{1}{k+1} \sum_{i=0}^k \ln |A_{t-i}| \xrightarrow{c.s.} E(\ln |A_t|) < 0,$$

por la Ley Fuerte de los Grandes Números. De modo que:

$$\prod_{i=0}^k |A_{t-i}| = \exp \left(\sum_{i=0}^k \ln |A_{t-i}| \right) \xrightarrow{c.s.} 0,$$

lo cual muestra la importancia de la condición de que $E(\ln |A_t|) < 0$. La solución (2.17) a la EER es un proceso estrictamente estacionario (siendo una función de las variables aleatorias i.i.d. $\{A_s, B_s\}_{s \leq t}$), y la condición de que $E(\ln |A_t|) < 0$ resulta ser la clave para la estacionariedad estricta de los modelos ARCH y GARCH.

Estacionariedad del modelo ARCH(1). El modelo ARCH(1) al cuadrado en (2.14) es una EER de la forma (2.15) con $A_t = \alpha_1 Z_t^2$ y $B_t = \alpha_0 Z_t^2$. Por lo que las condiciones (2.16) se traducen en los requerimientos $E(\ln^+ |\alpha_0 Z_t^2|) < \infty$ (donde $\ln^+ = \max\{0, \ln(x)\}$), la cual es automáticamente cierta para el proceso ARCH(1) tal como lo hemos definido, y $E(\ln(\alpha_1 Z_t^2)) < 0$. Esta es la condición para una solución estrictamente estacionaria de las ecuaciones del proceso ARCH(1) y puede ser probado que, de hecho, esta es una condición necesaria y suficiente para tener estacionariedad estricta. La solución de la ecuación (2.14) toma la forma:

$$X_t^2 = \alpha_0 \sum_{i=0}^{\infty} \alpha_1^i \prod_{j=0}^i Z_{t-j}^2. \quad (2.18)$$

Si las innovaciones $\{Z_t\}_{t \in \mathbb{Z}}$ son normales estándar, entonces la condición para una solución estrictamente estacionaria es aproximadamente $\alpha_1 < 3.562$; quizá de forma sorprendente, si las innovaciones tienen una distribución t escalada con 4 grados de libertad y varianza 1, la condición es $\alpha_1 < 5.437$. La estacionariedad estricta depende de la distribución de las innovaciones, pero la estacionariedad en covarianza no; la condición necesaria y suficiente para la estacionariedad en covarianza es siempre que $\alpha_1 < 1$, como probamos a continuación.

Proposición 2.16. *El Proceso ARCH(1) es un proceso ruido blanco estacionario en covarianza si y sólo si $\alpha_1 < 1$. La varianza del proceso estacionario en covarianza está dada por $\alpha_0/(1 - \alpha_1)$.*

Demostración. Suponiendo que el proceso es estacionario en covarianza, se sigue de (2.14) y de que $E(Z_t^2) = 1$ que

$$\sigma_x^2 = E(X_t^2) = \alpha_0 + \alpha_1 E(X_{t-1}^2) = \alpha_0 + \alpha_1 \sigma_x^2;$$

despejando tenemos que $\sigma_x^2 = \alpha_0/(1 - \alpha_1)$, y debemos tener $\alpha_1 < 1$.

Inversamente, si $\alpha_1 < 1$, entonces, por la desigualdad de Jensen,

$$E(\ln(\alpha_1 Z_t^2)) \leq \ln(E(\alpha_1 Z_t^2)) = \ln(\alpha_1) < 0,$$

y podemos utilizar (2.18) para calcular que

$$E(X_t^2) = \alpha_0 \sum_{i=0}^{\infty} \alpha_1^i = \frac{\alpha_0}{1 - \alpha_1}.$$

El proceso $\{X_t\}_{t \in \mathbb{Z}}$ es una diferencia de martingalas con un segundo momento finito que no depende del tiempo. Por lo tanto, el proceso es un Ruido Blanco. \square

Sobre la Distribución Estacionaria de X_t . Es claro de (2.18) que la distribución de $\{X_t\}$ en un modelo ARCH(1) guarda una relación complicada con la distribución de las innovaciones $\{Z_t\}$. Incluso si la distribución de las innovaciones es Gaussiana, la distribución de la serie de tiempo no es Gaussiana, sino más bien una distribución leptocúrtica, con colas que decaen más lentamente que las de una distribución normal.

Enunciamos ahora una proposición que nos da condiciones para la existencia de los momentos de un modelo ARCH(1). Nuevamente, estas condiciones quedan en términos de las innovaciones del proceso, lo que muestra nuevamente la importancia que juegan éstas dentro de los modelos ARCH y GARCH.

Proposición 2.17. *Para $m \geq 1$, el proceso ARCH(1) estrictamente estacionario tiene momentos finitos de orden $2m$ si y sólo si $E(Z_t^{2m}) < \infty$ y $\alpha_1 < (E(Z_t^{2m}))^{-1/m}$.*

Demostración. Reescribimos (2.18) como $X_t^2 = Z_t^2 \sum_{i=0}^{\infty} Y_{t,i}$ para variables aleatorias positivas $Y_{t,i} = \alpha_0 \alpha_1 \prod_{j=1}^i Z_{t-j}^2, i \geq 1$, y $Y_{t,0} = \alpha_0$. Para $m \geq 1$ se tienen las siguientes desigualdades (la última, es consecuencia de la desigualdad de Minkowski):

$$E(Y_{t,1}^m) + E(Y_{t,2}^m) \leq E((Y_{t,1} + Y_{t,2})^m) \leq ((E(Y_{t,1}^m))^{1/m} + (E(Y_{t,2}^m))^{1/m})^m.$$

Puesto que

$$E(X_t^{2m}) = E(Z_t^{2m}) E\left(\left(\sum_{i=0}^{\infty} Y_{t,i}\right)^m\right),$$

se sigue que

$$E(Z_t^{2m}) \sum_{i=0}^{\infty} E(Y_{t,i}^m) \leq E(X_t^{2m}) \leq E(Z_t^{2m}) \left(\sum_{i=0}^{\infty} (E(Y_{t,i}^m))^{1/m}\right)^m.$$

Como $E(Y_{t,i}^m) = \alpha_0^m \alpha_1^{im} (E(Z_t^{2m}))^i$, puede deducirse que las tres cantidades son finitas si y solo si $E(Z_t^{2m}) < \infty$ y $\alpha_1^m E(Z_t^{2m}) < 1$. \square

La proposición anterior nos dice que, por ejemplo, para un cuarto momento finito ($m = 2$) requerimos $\alpha < 1/\sqrt{3}$ en el caso de las innovaciones Gaussianas y $\alpha < 1/\sqrt{6}$ en el caso de innovaciones t con seis grados de libertad; para innovaciones t con 4 grados de libertad, el cuarto momento está indefinido.

Suponiendo la existencia de un cuarto momento finito, es sencillo calcular su valor y también el de la curtosis del proceso. Elevamos al cuadrado ambos lado de (2.14), tomamos esperanza de ambos lado y resolvemos para $E(X_t^4)$ para obtener

$$E(X_t^4) = \frac{\alpha_0 E(Z_t^4) (1 - \alpha_1^2)}{(1 - \alpha_1)^2 (1 - \alpha_1^2 E(Z_t^4))}.$$

La curtosis de la distribución estacionara del proceso puede entonces ser calculada:

$$\kappa_X = \frac{E(X_t^4)}{E(X_t^2)^2} = \frac{\kappa_Z (1 - \alpha_1^2)}{(1 - \alpha_1^2 \kappa_Z)},$$

donde $\kappa_Z = E(Z_t^4)$ denota la curtosis de las innovaciones. Claramente, cuando $\kappa_Z > 1$, la curtosis de la distribución estacionaria es mayor en comparación con la de la distribución de las innovaciones; para innovaciones Gaussianas o t se tiene que $\kappa_X > 3$, por lo que la distribución estacionaria es leptocúrtica.

Paralelos con el proceso AR(1). Centramos ahora nuestra atención en la estructura de dependencia serial del cuadrado de la serie en el caso en el que tenemos estacionariedad en covarianza ($\alpha_1 < 1$). Escribimos el proceso al cuadrado como

$$X_t^2 = \sigma_t^2 Z_t^2 = \sigma_t^2 + \sigma_t^2 (Z_t^2 - 1) \quad (2.19)$$

Sea $V_t = \sigma_t^2(Z_t^2 - 1)$, notemos que $\{V_t\}_{t \in \mathbb{Z}}$ forma una serie con la propiedad de ser una diferencia de martingalas pues $E|V_t| < \infty$ (lo cual se debe a que X_t es estacionario en covarianza) y $E(V_t | \mathcal{F}_{t-1}) = \sigma_t E(Z_{t-1}^2 - 1) = 0$ (pues σ_t es \mathcal{F}_{t-1} -medible, Z_t es independiente de \mathcal{F}_{t-1} y $E(Z_{t-1}^2) = 1$). Ahora reescribimos (2.19) como $X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + V_t$, y observamos que esto se parece mucho a un proceso AR(1) para X_t^2 , a excepción de que V_t no es necesariamente un ruido blanco. Si restringimos nuestra atención a procesos donde $E(X_t^4) < \infty$, entonces V_t tiene un segundo momento finito y constante y es un proceso ruido blanco. De acuerdo con esta hipótesis, según la Definición 2.9, X_t^2 es un proceso AR(1) de la forma:

$$\left(X_t^2 - \frac{\alpha_0}{1 - \alpha_1} \right) = \alpha_1 \left(X_{t-1}^2 - \frac{\alpha_0}{1 - \alpha_1} \right) + V_t$$

Este proceso tiene media $\alpha_0/(1 - \alpha_1)$ y su función de autocorrelación está dada por $\rho(h) = \alpha_1^{|h|}$, $h \in \mathbb{Z}$.

2.2.2. Procesos GARCH

Definición 2.18. *Procesos GARCH.* Sea $\{Z_t\}_{t \in \mathbb{Z}}$ un RBE(0,1). El proceso $\{X_t\}_{t \in \mathbb{Z}}$ es un proceso GARCH(p, q) si es estrictamente estacionario y satisface, para toda $t \in \mathbb{Z}$ y algún proceso que solo tome valores estrictamente positivos $\{\sigma_t\}_{t \in \mathbb{Z}}$, las ecuaciones

$$X_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-1}^2 \quad (2.20)$$

Donde $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1, \dots, p$ y $\beta_j \geq 0, j = 1, \dots, q$.

Los procesos GARCH son procesos ARCH generalizados en el sentido de que se permite que la volatilidad al cuadrado $\{\sigma_t^2\}_{t \in \mathbb{Z}}$ dependa de los cuadrados tanto de volatilidades como de valores anteriores del proceso.

El modelo GARCH(1,1). En la práctica, los modelos GARCH de orden bajo son los procesos más ampliamente utilizados; por tal motivo, en esta Sección centraremos nuestra atención en el modelo GARCH(1,1). En este modelo, periodos de alta volatilidad tienden a ser *persistentes*, puesto que $|X_t|$ tiene oportunidad de ser grande si *cualquiera* entre $|X_{t-1}|$ y σ_{t-1} fue grande; el mismo efecto puede ser logrado en modelos ARCH de orden alto, sin embargo, los modelos GARCH de bajo orden logran el mismo efecto de una forma más parsimoniosa.

Estacionariedad. Se sigue de (2.20) que para un modelo GARCH(1,1) tenemos

$$\sigma_t^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1) \sigma_{t-1}^2, \quad (2.21)$$

la cual es de nuevo una EER de la forma $Y_t = A_t Y_{t-1} + B_t$ como en (2.15). Esta vez tenemos una EER para $Y_t = \sigma_t^2$ en vez de X_t^2 , pero su análisis se sigue fácilmente de aquel realizado para el caso del modelo ARCH(1).

La condición de $E(\ln |A_t|) < 0$ para una solución estrictamente estacionaria de (2.15) se traduce a la condición $E(\ln(\alpha_1 Z_t^2 + \beta_1)) < 0$ para (2.21) y la solución general (2.17) se convierte en

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=1}^i (\alpha_1 Z_{t-j}^2 + \beta_1). \quad (2.22)$$

Si $\{\sigma_t\}_{t \in \mathbb{Z}}$ es un proceso estrictamente estacionario, entonces también lo es $\{X_t\}_{t \in \mathbb{Z}}$, pues $X_t = \sigma_t Z_t$ y $\{Z_t\}_{t \in \mathbb{Z}}$ es simplemente un ruido blanco estricto. La solución de las ecuaciones que definen el GARCH(1,1) entonces es

$$X_t = Z_t \sqrt{\alpha_0 \left(1 + \sum_{i=1}^{\infty} \prod_{j=1}^i (\alpha_1 Z_{t-j}^2 + \beta_1) \right)}, \quad (2.23)$$

y podemos utilizar esto para derivar la condición de estacionariedad en covarianza.

Proposición 2.19. *El Proceso GARCH(1,1) es un Proceso ruido blanco estacionario en covarianza si sólo si $\alpha_1 + \beta_1 < 1$. La varianza del proceso estacionario en covarianza está dada por $\alpha_0 / (1 - \alpha_1 - \beta_1)$.*

Demostración. La demostración de esta proposición es completamente análoga a la de la Proposición 2.16 \square

Cuartos momentos y curtosis. Utilizando un enfoque similar al de la Proposición 2.17, podemos utilizar (2.23) para derivar condiciones para la existencia de momentos más altos de un proceso GARCH(1,1) estacionario en covarianza. Para la existencia de un cuarto momento, una condición necesaria y suficiente es que $E((\alpha_1 Z_t^2 + \beta_1)^2) < 1$, o, de forma alternativa, que

$$(\alpha_1 + \beta_1)^2 < 1 - (\kappa_Z - 1) \alpha_1^2.$$

Suponiendo que la condición anterior se satisface, podemos calcular el cuarto momento y la curtosis de X_t . Elevamos al cuadrado ambos lados de (2.21) y

tomamos esperanza para obtener

$$E(\sigma_t^4) = \alpha_0^2 + (\alpha_1^2 \kappa_Z + \beta_1^2 + 2\alpha_1 \beta_1) E(\sigma_t^4) + 2\alpha_0(\alpha_1 + \beta_1) E(\sigma_t^2).$$

Resolviendo la ecuación anterior para $E(\sigma_t^4)$, teniendo en mente que $E(\sigma_t^2) = E(X_t^2) = \alpha_0/(1 - \alpha_1 - \beta_1)$, y fijando $E(X_t^4) = \kappa_Z E(\sigma_t^4)$, obtenemos:

$$E(\sigma_t^4) = \frac{\alpha_0^2 \kappa_Z (1 - (\alpha_1 + \beta_1)^2)}{(1 - \alpha_1 - \beta_1)^2 (1 - \alpha_1^2 \kappa_Z - \beta_1^2 - 2\alpha_1 \beta_1)},$$

de donde se sigue que

$$\kappa_X = \frac{\kappa_Z (1 - (\alpha_1 + \beta_1)^2)}{(1 - (\alpha_1 + \beta_1)^2 - (\kappa_Z - 1)\alpha_1^2)}.$$

Nuevamente es claro que la curtosis de X_t es más grande que la de Z_t siempre que $\kappa_Z > 1$, como en el caso de innovaciones Gaussianas o t escaladas.

Paralelos con el Proceso ARMA(1,1). Utilizando la misma representación que en la ecuación (2.19), el proceso GARCH(1,1) estacionario en covarianza puede ser escrito como

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + V_t,$$

donde V_t es una diferencia de martingalas, dada por $V_t = \sigma_t^2(Z_t^2 - 1)$. Ya que $\sigma_{t-1}^2 = X_{t-1}^2 - V_{t-1}$, podríamos escribir

$$X_t^2 = \alpha_0 + (\alpha_1 + \beta_1) X_{t-1}^2 - \beta_1 V_{t-1} + V_t, \quad (2.24)$$

expresión que comienza a parecerse a un proceso ARMA(1,1) para X_t^2 . Si suponemos además que $E(X_t^4) < \infty$, entonces, recordando que $\alpha_1 + \beta_1 < 1$, tenemos formalmente que

$$\left(X_t^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \right) = (\alpha_1 + \beta_1) \left(X_{t-1}^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \right) - \beta_1 V_{t-1} + V_t$$

es un proceso ARMA(1,1). La Figura 2.3 muestra una simulación de un proceso GARCH(1,1) con cuarto momento finito cuyos valores al cuadrado siguen un proceso ARMA(1,1)

El modelo GARCH(p,q). Modelos ARCH y GARCH de órdenes más altos tienen el mismo comportamiento general que el ARCH(1) y el GARCH(1,1), pero su análisis matemático se vuelve más tedioso. La condición necesaria y

Simulación GARCH(1,1)

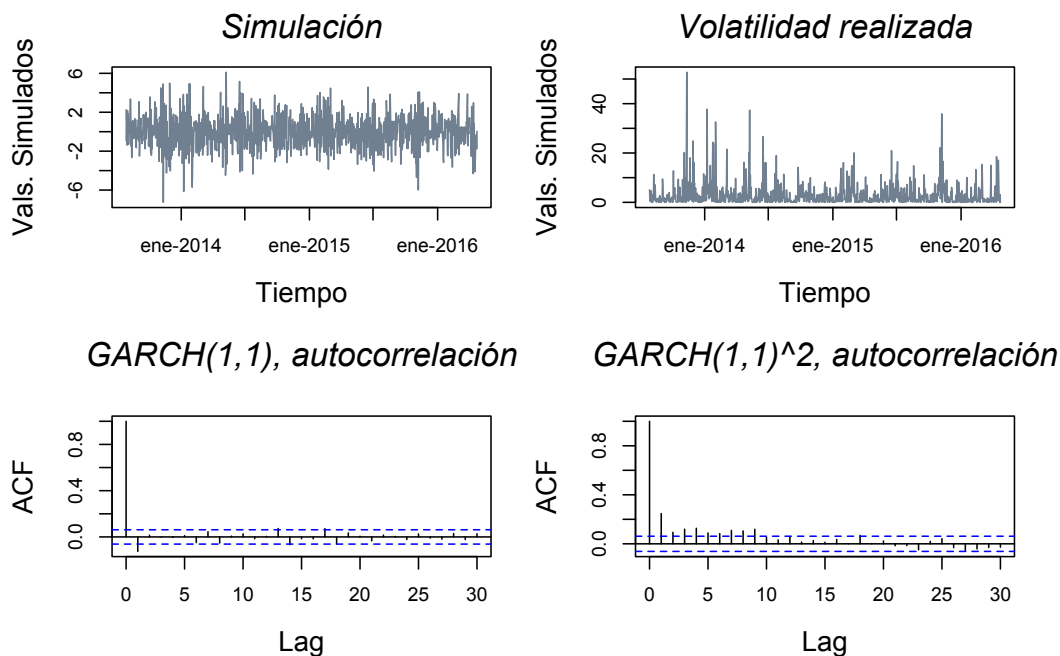


Figura 2.3: Ejemplo de un proceso GARCH(1,1) cuyos valores al cuadrado siguen un proceso ARMA(1,1). En el primer cuadrante se observa la simulación del proceso; en el segundo, la volatilidad realizada del mismo; en el tercero la función de autocorrelación de los valores simulados crudos; en el cuarto, la función de autocorrelación de los valores simulados al cuadrado.

suficiente para tener soluciones estrictamente estacionarias de las EERs que los definen es: $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$.¹²

El cuadrado de un Proceso GARCH(p, q) tiene la estructura

$$X_t^2 = \alpha_0 + \sum_{i=1}^{\max\{p,q\}} (\alpha_i + \beta_i) X_{t-i}^2 - \sum_{j=1}^q \beta_j V_{t-j} + V_t,$$

donde $\alpha_i = 0$ para $i = p+1, \dots, q$ si $q > p$ o $\beta_j = 0$ para $j = q+1, \dots, p$ si $p > q$. Esto tiene la forma de un proceso ARMA($\max(p, q), q$) y es formalmente tal si $E(X_t^4) < \infty$.

¹² Los detalles sobre esto pueden consultarse en [17].

2.2.3. Ajustando modelos ARCH y GARCH a datos de origen financiero

Una vez analizadas las principales propiedades de los procesos ARCH y GARCH, el siguiente problema a resolver es práctico. Específicamente, a fin de poder utilizar esta clase de modelos en la práctica debemos ser capaces de estimar sus parámetros a partir de un conjunto de información. Por lo tanto, a continuación estudiamos distintos enfoques para obtener dichas estimaciones y poder así ajustar estos modelos a datos reales.

Construyendo la Verosimilitud. En la práctica, el enfoque más utilizado para ajustar modelos GARCH a datos es la Máxima Verosimilitud. Estudiaremos el ajuste de los modelos ARCH(1) y el GARCH(1,1); los ajustes del ARCH(p) y GARCH(p, q) se siguen de estos análisis de forma sencilla.

Para los modelos ARCH(1) y GARCH(1,1) supongamos que tenemos $n + 1$ valores o datos X_0, X_1, \dots, X_n , podemos escribir su función de densidad como:

$$f_{X_0, X_1, \dots, X_n}(x_0, x_1, \dots, x_n) = f_{X_0}(x_0) \prod_{t=1}^n f_{X_t | X_{t-1}, \dots, X_0}(x_t | x_{t-1}, \dots, x_0). \quad (2.25)$$

Para el proceso ARCH(1) puro, que es Markoviano de primer orden, las densidades condicionales $f_{X_t | X_{t-1}, \dots, X_0}$ en (2.25) dependen del proceso sólo a través del valores de σ_t o, equivalentemente, X_{t-1} . Entonces, la densidad condicional puede ser calculada como

$$f_{X_t | X_{t-1}, \dots, X_0}(x_t | x_{t-1}, \dots, x_0) = f_{X_t | X_{t-1}}(x_t | x_{t-1}) = \frac{1}{\sigma_t} g\left(\frac{x_t}{\sigma_t}\right), \quad (2.26)$$

donde $\sigma_t = (\alpha_0 + \alpha_1 X_{t-1}^2)^{1/2}$ y $g(z)$ denota la función de densidad de las innovaciones $\{Z_t\}_{t \in \mathbb{Z}}$. Recordemos que ésta debe tener media cero y varianza 1, y que elecciones típicas serían densidades normales estándar o t escaladas.

Sin embargo, no se conoce la densidad marginal f_{X_0} en (2.25) de una forma manejable y cerrada para los modelos ARCH y GARCH, lo cual origina un problema para basar una verosimilitud en (2.25). La solución empleada en la práctica es construir la *verosimilitud condicional* dado X_0 , la cual es calculada como:

$$f_{X_1, X_2, \dots, X_n | X_0}(x_1, x_2, \dots, x_n | x_0) = \prod_{t=1}^n f_{X_t | X_{t-1}, \dots, X_0}(x_t | x_{t-1}, \dots, x_0). \quad (2.27)$$

Para el ARCH(1) tenemos entonces de (2.26) que

$$L(\alpha_0, \alpha_1; \mathbf{X}) = f_{X_1, X_2, \dots, X_n | X_0}(X_1, X_2, \dots, X_n | X_0) = \prod_{t=1}^n \frac{1}{\sigma_t} g\left(\frac{X_t}{\sigma_t}\right),$$

con $\sigma_t = (\alpha_0 + \alpha_1 X_{t-1}^2)^{1/2}$. Para un modelo ARCH(p) utilizaríamos argumentos análogos para escribir una verosimilitud condicionada a las primeras p observaciones.

En el modelo GARCH(1,1) σ_t es definida recursivamente en términos de σ_{t-1} y aquí, en lugar de utilizar (2.26), construimos la densidad conjunta de X_1, X_2, \dots, X_n condicionada a los valores de ambos, X_0 y σ_0 :

$$f_{X_1, \dots, X_n | X_0, \sigma_0}(x_1, \dots, x_n | x_0, \sigma_0) = \prod_{t=1}^n f_{X_t | X_{t-1}, \dots, X_0, \sigma_0}(x_t | x_{t-1}, \dots, x_0, \sigma_0).$$

Las densidades condicionales $f_{X_t | X_{t-1}, \dots, X_0, \sigma_0}(x_t | x_{t-1}, \dots, x_0, \sigma_0)$ dependen del pasado sólo a través del valor de σ_t , el cual está dado recursivamente por $\sigma_0, X_0, \dots, X_{t-1}$ utilizando $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$. Esto nos da la verosimilitud condicional

$$L(\alpha_0, \alpha_1, \beta_1; \mathbf{X}) = \prod_{t=1}^n \frac{1}{\sigma_t} g\left(\frac{X_t}{\sigma_t}\right), \quad \sigma_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2}$$

El problema sigue siendo que el valor de σ_0^2 no es verdaderamente observado, y esto usualmente se soluciona eligiendo un valor inicial, como la varianza muestral de X_0, X_1, \dots, X_n o simplemente cero.

Para un modelo GARCH(p, q) supondríamos que tenemos $n+p$ datos $X_{-p+1}, \dots, X_0, \dots, X_n$, y evaluaríamos la verosimilitud condicionada a los valores (observados) de X_{-p+1}, \dots, X_0 así como en los valores (no observados) de $\sigma_{-q+1}, \dots, \sigma_0$, cuyos valores iniciales son fijados como se describió arriba. Por ejemplo, si $p = 1$ y $q = 3$, requeriríamos valores para σ_{-2}, σ_{-1} y σ_0 .

Derivando estimaciones de parámetros. Consideremos una log-verosimilitud de la forma

$$\ln(L(\boldsymbol{\theta}; \mathbf{X})) = \sum_{t=1}^n l_t(\boldsymbol{\theta}), \quad (2.28)$$

donde $l_t(\boldsymbol{\theta})$ representa la contribución a la log-verosimilitud de la t -ésima observación.

El estimador máximo verosímil $\hat{\boldsymbol{\theta}}$ maximiza la log-verosimilitud (condicional) (2.28) y, siendo un máximo local, resuelve las ecuaciones

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln(L(\boldsymbol{\theta}, \mathbf{X})) = \sum_{t=1}^n \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \quad (2.29)$$

donde al lado izquierdo de la ecuación también se le conoce como *vector de conteos* de la verosimilitud condicional. Las ecuaciones en (2.29) normalmente son resueltas con procedimientos Newton-Raphson modificados. Para modelos GARCH, un método ampliamente utilizado es el BHHH de Bendt, Hall, Hall y Hausmann.

Al describir el comportamiento de los estimadores de los parámetros en los siguientes párrafos consideramos dos situaciones. En la primera situación suponemos que el modelo ajustado fue *correctamente especificado*, de modo que los datos son realmente generados por un modelo de series de tiempo con la parte dinámica e innovaciones supuestas. Describimos el comportamiento de los Estimadores Máximo Verosímiles (EMV) bajo esta idealización.

En la segunda situación, suponemos que la forma dinámica fue correctamente ajustada pero que las innovaciones son erróneamente supuestas como Gaussianas. Al procedimiento de ajuste del modelo bajo esta mala especificación se le conoce como *Quasi-Máxima Verosimilitud* y a los estimadores obtenidos como Estimadores Quasi-Máximo Verosímiles (EQMV). Esencialmente, la verosimilitud Gaussiana es tratada como función objetivo a ser maximizada en lugar de la verosimilitud apropiada; nuestra intuición sugiere que esto podría seguir dando estimaciones razonables de parámetros, y este resulta ser el caso bajo hipótesis apropiadas para la verdadera distribución de las innovaciones.

Propiedades de los Estimadores Máximo Verosímiles (EMVs). En esta Sección se obtienen resultados para el comportamiento asintótico de los estimadores máximo verosímiles para los modelo ARCH y GARCH, el cual termina siendo muy similar al que se tiene para EMVs en el caso de variables aleatorias i.i.d.; sin embargo, debemos tener claro que los resultados obtenidos para este caso no se siguen del caso i.i.d., sino que se trabajaron separadamente de forma rigurosa. Si se desea recordar los resultados asintóticos clásicos para EMVs puede consultarse el libro de [18].

Para un modelo GARCH(p, q) puro con innovaciones Gaussianas (suponiendo que el modelo ha sido correctamente especificado) se puede probar que

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N_{p+q+1}(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta})^{-1}),$$

donde

$$\mathbf{I}(\boldsymbol{\theta}) = E \left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) = -E \left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \quad (2.30)$$

es la matriz de información de Fisher que surge de cada observación. Así, tenemos estimadores consistentes y asintóticamente normales de los parámetros del GARCH. En la práctica, la matriz de información esperada $\mathbf{I}(\boldsymbol{\theta})^{-1}$ es aproximada por una matriz de información *observada*, y aquí podríamos tomar la matriz de información observada proveniente de cualquiera de las formas equivalentes para la matriz de información (2.30), esto es:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) \quad \text{o} \quad \bar{\mathbf{J}}(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{t=1}^n \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \quad (2.31)$$

donde se dice que la primer matriz está en forma de *producto exterior* y la segunda en forma *Hessiana*. Estas matrices son estimadas al evaluarlas en los estimadores máximo verosímiles para obtener $\bar{\mathbf{I}}(\hat{\boldsymbol{\theta}})$ o $\bar{\mathbf{J}}(\hat{\boldsymbol{\theta}})$. En la práctica, esto se lleva a cabo diferenciando numéricamente en primer y segundo orden la log-verosimilitud en el EMV y las matrices necesarias son obtenidas como subproductos del procedimiento BHHH para obtener estimaciones de los parámetros.

Si el modelo está bien especificado, deberíamos observar que $\bar{\mathbf{I}}(\hat{\boldsymbol{\theta}})$ y $\bar{\mathbf{J}}(\hat{\boldsymbol{\theta}})$ son muy similares, pues son estimadores basados en dos expresiones distintas para la misma matriz de información de Fisher. En la práctica, también podríamos estimar $\bar{\mathbf{I}}(\boldsymbol{\theta})$ por $\bar{\mathbf{J}}(\hat{\boldsymbol{\theta}}) \bar{\mathbf{I}}(\hat{\boldsymbol{\theta}})^{-1} \bar{\mathbf{J}}(\hat{\boldsymbol{\theta}})$; esto nos anticipa al llamado *estimador sándwich* utilizado en el procedimiento QMV.

Propiedades de los Estimadores Quasi-Máximo Verosímiles (EQMV). Bajo este enfoque tomamos un modelo GARCH(p, q) con innovaciones no Gaussianas pero intentamos estimar los parámetros del proceso maximizando la verosimilitud de un GARCH(p, q) con innovaciones Gaussianas. Siguiendo esta metodología, aún obtenemos estimadores consistentes de los parámetros del modelo y si la verdadera distribución de las innovaciones tiene cuarto momento finito, obtenemos nuevamente normalidad asintótica; sin embargo, la forma de la matriz asintótica de covarianzas cambia.

Ahora distinguimos entre las matrices $\bar{\mathbf{I}}(\boldsymbol{\theta})$ y $\bar{\mathbf{J}}(\boldsymbol{\theta})$, dadas por

$$\bar{\mathbf{I}}(\boldsymbol{\theta}) = E \left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right), \quad \bar{\mathbf{J}}(\boldsymbol{\theta}) = -E \left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)$$

donde la esperanza es ahora tomada con respecto al verdadero modelo (no el Gaussiano). Las matrices $\bar{\mathbf{I}}(\boldsymbol{\theta})$ y $\bar{\mathbf{J}}(\boldsymbol{\theta})$ difieren en general (a menos que el modelo Gaussiano sea correcto). Se puede probar que

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}) \xrightarrow{d} N_{p+q+1}(\mathbf{0}, \bar{\mathbf{J}}(\boldsymbol{\theta})^{-1} \bar{\mathbf{I}}(\boldsymbol{\theta}) \bar{\mathbf{J}}(\boldsymbol{\theta})^{-1}) \quad (2.32)$$

y se dice que la matriz de covarianza está en forma de sándwich; dicha matriz puede ser estimada por $\bar{\mathbf{J}}(\hat{\boldsymbol{\theta}})^{-1} \bar{\mathbf{I}}(\hat{\boldsymbol{\theta}}) \bar{\mathbf{J}}(\hat{\boldsymbol{\theta}})^{-1}$, donde $\bar{\mathbf{I}}(\hat{\boldsymbol{\theta}})$ y $\bar{\mathbf{J}}(\hat{\boldsymbol{\theta}})$ están definidas como en (2.31). Si la revisión del modelo sugiere que la dinámica de nuestra serie ha sido descrita adecuadamente por el modelo GARCH, pero la hipótesis Gaussiana parece dudosa, entonces los errores estándar de los estimadores de los parámetros deben basarse en la estimación de esta matriz de covarianza.

Calidad de ajuste del modelo. Como en los modelos ARMA, es común evaluar la calidad de ajuste de los modelos GARCH utilizando residuales. Consideremos un modelo general GARCH(p, q) de la forma $X_t = \sigma_t Z_t$, con σ_t como en la Definición 2.18. A los residuales de este modelo se les suele llamar *residuos estandarizados*. Dichos residuales son realizaciones reconstruidas del RBE que conduce el modelo GARCH utilizado, y son calculados como:

$$\hat{Z}_t = \frac{X_t}{\hat{\sigma}_t}, \quad \hat{\sigma}_t = \hat{\alpha}_0 + \sum_{i=1}^p \hat{\alpha}_i \epsilon_{t-i}^2 + \sum_{j=1}^q \hat{\beta}_j \hat{\sigma}_{t-j}. \quad (2.33)$$

No obstante, para utilizar las fórmulas anteriores necesitamos algunos valores iniciales, y una solución es fijar los valores iniciales requeridos para ϵ_t iguales a cero y los valores iniciales requeridos de la volatilidad $\hat{\sigma}_t$ iguales a la varianza muestral o cero. Como los primeros valores estarán influenciados por valores iniciales fijados, podríamos ignorarlos en análisis posteriores.

Ahora bien, si el modelo GARCH(p, q) ha logrado capturar apropiadamente la dinámica de la serie estudiada, los residuos estandarizados deberían comportarse como un ruido blanco estricto. Dicha hipótesis podría ser corroborada construyendo los correlogramas de los residuos estandarizados crudos (y de sus valores absolutos o al cuadrado) y aplicándoles pruebas *portmanteau* (como la de Ljung-Box) para ruidos blancos estrictos.

Suponiendo que la hipótesis de RBE no es rechazada, de modo que la parte dinámica de la serie ha sido capturada de forma satisfactoria, la validez de la distribución de las innovaciones utilizada en el ajuste por máxima verosimilitud puede también ser investigada utilizando gráficos QQ y pruebas de bondad de ajuste para las distribuciones normal estándar o t escalada. Si la

verosimilitud Gaussiana hace un buen trabajo capturando la parte dinámica del modelo (las Z_t son un RBE), pero los residuales estandarizados no se comportan como observaciones Gaussianas i.i.d., entonces la filosofía de ajuste Quasi-Máximo Verosímil puede ser adoptada y los errores estándar pueden ser estimados utilizando el estimador sándwich implicado arriba en (2.32).

Esto abre la posibilidad de realizar análisis en dos etapas en el ajuste de este tipo de modelos: primero, la parte dinámica es estimada utilizando métodos QMV y después la distribución de las innovaciones es modelada utilizando los residuales estandarizados de la parte dinámica como datos. En la segunda etapa podríamos considerar utilizar modelos de colas más pesadas que el Gaussiano que a su vez permitan algo de asimetría en las innovaciones. Este procedimiento nos lleva a tener más transparencia en la construcción del modelo y nos permite separar la tarea de modelar la volatilidad (que se logra una vez que se ha modelado X_t) y los choques que manejan el proceso (las Z_t); más adelante, en el Capítulo 4, analizamos de forma profunda esta posibilidad, complementándola con metodologías de la Teoría de Valores Extremos.

2.3. Medición del riesgo vía modelos de volatilidad estocástica

En esta Sección estudiaremos cómo atacar un problema fundamental en la administración de riesgo de mercado: la estimación de medidas de riesgo condicionales, como el *VaR* y el *Expected Shortfall*.

En nuestro caso, dicho estudio se realizará con base en los modelos de volatilidad estocástica estudiados en la Sección anterior. De esta manera, nuestro estudio también discutirá cómo realizar evaluaciones de tales estimaciones; es decir, analizaremos una metodología de *backtesting* específica para estimaciones condicionales del *VaR* y el *Expected Shortfall* obtenidas a partir de procesos ARCH y GARCH.

Ahora bien, a fin de realizar estimaciones como las que deseamos, debemos ser capaces de pronosticar un fenómeno fundamental en el análisis de series de tiempo: la volatilidad. Por lo tanto, a continuación estudiaremos un par de metodologías que nos permiten realizar dicho pronóstico.

2.3.1. Pronósticos de volatilidad

El primer método de predicción de volatilidad que estudiaremos está basado en un modelo tipo GARCH, mientras que el segundo utiliza promedios móviles exponenciales para realizar el pronóstico deseado.

Predicción de volatilidad basada en el GARCH. Supongamos que los datos de los rendimientos X_{t-n+1}, \dots, X_t siguen un modelo particular de la familia GARCH. Queremos pronosticar volatilidad futura, es decir, predecir el valor de σ_{t+h} para $h \geq 1$. Esto está íntimamente relacionado con predecir X_{t+h}^2 . Por el momento, tomaremos la hipótesis de que tenemos acceso a la historia infinita del proceso hasta el tiempo t , representada por $\mathcal{F}_t = \sigma(\{X_s, s \leq t\})$, luego adaptaremos nuestra fórmula de predicción para tomar en cuenta el tamaño finito de la muestra.

En lo que sigue, supondremos que el modelo GARCH ha sido ajustado y sus parámetros estimados. Presentamos los cálculos para modelos simples a partir de los cuales se siguen fácilmente los casos de modelos más generales (en los desarrollos siguientes, omitiremos la notación $\hat{\theta}$ para los parámetros estimados).

Ejemplo 2.20. Predicción en el modelo GARCH(1,1). *Supongamos que utilizamos un modelo GARCH(1,1) puro de acuerdo con la Definición 2.18 y que dicho modelo es estacionario en covarianza, por lo que $E(X_t^2) = E(\sigma_t^2) < \infty$. Dado que $\{X_t\}_{t \in \mathbb{Z}}$ es una diferencia de martingalas, las predicciones óptimas de X_{t+h} son cero pues*

$$E(X_{t+h} | \mathcal{F}_{t+h-1}) = E(\sigma_{t+h} Z_{t+h} | \mathcal{F}_{t+h-1}) = \sigma_{t+h} E(Z_{t+h}) = 0.$$

Una predicción natural de X_{t+1}^2 basada en \mathcal{F}_t es su esperanza condicional dada por

$$E(X_{t+1}^2 | \mathcal{F}_t) = \sigma_{t+1}^2 = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \sigma_t^2,$$

y si $E(X_t^4) < \infty$, entonces este es el predictor óptimo en un sentido de mínimos cuadrados. Notemos que la predicción de la variable aleatoria X_t^2 basada en la información \mathcal{F}_t es el valor de σ_{t+1}^2 , que es conocida al tiempo t , siendo una función de la historia infinita del proceso (es decir, el proceso $\{\sigma_t\}_{t \in \mathbb{Z}}$ es previsible).

En la práctica, debemos hacer una aproximación basados en esta fórmula pues la serie infinita de valores que nos permitiría calcular σ_t^2 no está disponible. Un enfoque natural en aplicaciones es aproximar σ_t^2 con un estimador de la volatilidad al cuadrado, $\hat{\sigma}_t^2$, calculado de las ecuaciones de residuales en

(2.33). Nuestro pronóstico aproximado de X_{t+1}^2 también funciona como una estimación de la volatilidad al cuadrado al tiempo $t + 1$ y está dado por

$$\hat{\sigma}_{t+1}^2 = \hat{E}(X_{t+1}^2 | \mathcal{F}_t) = \alpha_0 + \alpha_1 X_t^2 + \beta_1 + \hat{\sigma}_t \quad (2.34)$$

Por lo tanto, la ecuación (2.34) puede ser pensada como un método recursivo para predecir la volatilidad de un paso adelante.

Cuando miramos $h > 1$ pasos adelante dada la información hasta el tiempo t , ambos X_{t+h}^2 y σ_{t+h}^2 , son variables aleatorias. Sus predicciones coinciden y son

$$\begin{aligned} E(X_{t+h}^2 | \mathcal{F}_t) &= E(\sigma_{t+h}^2 | \mathcal{F}_t) \\ &= \alpha_0 + \alpha_1 E(X_{t+h-1}^2 | \mathcal{F}_t) + \beta_1 E(\sigma_{t+h-1}^2 | \mathcal{F}_t) \\ &= \alpha_0 + (\alpha_1 + \beta_1) E(X_{t+h-1}^2 | \mathcal{F}_t) \end{aligned}$$

por lo que la fórmula general es

$$E(X_{t+h}^2 | \mathcal{F}_t) = \alpha_0 \sum_{i=0}^{h-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{h-1} (\alpha_1 X_t^2 + \beta_1 \sigma_t^2),$$

y obtenemos una fórmula para la práctica al sustituir un estimador de la volatilidad al cuadrado σ_t^2 como antes. Observemos que a medida que $h \rightarrow \infty$ tenemos que $E(\sigma_{t+h}^2 | \mathcal{F}_t) \rightarrow \alpha_0 / (1 - \alpha_1 - \beta_1)$, casi seguramente, por lo que la predicción de la volatilidad al cuadrado converge a la varianza no condicional del proceso.

Suavizamiento exponencial para volatilidad. Supongamos que creemos que nuestros datos de rendimientos siguen alguna clase de modelo de series de tiempo en el cual la volatilidad (la desviación estándar condicional) está definida, pero no deseamos especificar el modelo paramétrico exacto.

Podemos aplicar un suavizamiento exponencial como en (2.10) a las observaciones al cuadrado para obtener un procedimiento que sigue el esquema de actualización

$$P_t X_{t+1}^2 = \lambda X_t^2 + (1 - \lambda) P_{t-1} X_t^2 \quad (2.35)$$

para algún valor $\lambda \in (0, 1)$.

Puesto que esperamos que X_{t+1}^2 y σ_{t+1}^2 coincidan, podríamos considerar lo anterior como un suavizamiento exponencial para los valores no observados de la volatilidad al cuadrado. Podríamos definir un esquema recursivo para la volatilidad de un paso adelante como

$$\hat{\sigma}_{t+1}^2 = \lambda X_t^2 + (1 - \lambda) \hat{\sigma}_t^2, \quad \lambda \in (0, 1). \quad (2.36)$$

Esta es la idea esencial del enfoque de promedios móviles exponenciales ponderados (EWMA por sus siglas en inglés) para pronósticos de volatilidad.

2.3.2. Estimación de medidas de riesgo condicionales

Una vez contamos con metodologías para pronosticar volatilidad, volvemos al problema discutido en Capítulos anteriores: la medición del riesgo de forma condicional. Consideremos la situación donde deseamos medir el riesgo para una inversión en un activo, índice o moneda. En el día t , el valor de la posición es V_t y el log-rendimiento para el siguiente día es X_{t+1} , de modo que la pérdida para el día siguiente es $L_{t+1} = l_{[t]}(X_{t+1}) \approx -V_t X_{t+1}$. Requerimos estimaciones de medidas de riesgo, tales como el *VaR* y el *Expected Shortfall*, para la distribución condicional $F_{L_{t+1}|\mathcal{F}_t}$, donde $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$. Fijamos $V_t = 1$ para toda t y escribimos $L_t = -X_t$, de modo que $\{L_t\}_{t \in \mathbb{Z}}$ es el proceso de los log-rendimientos negativos.

Para calcular medidas de riesgo condicionales tomamos la siguiente hipótesis:

Hipótesis 2.21. *El proceso de pérdidas es adaptado a la filtración $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ y sigue un modelo estacionario de la forma $L_t = \sigma_t Z_t$, donde σ_t es \mathcal{F}_{t-1} -medibles y las $\{Z_t\}_{t \in \mathbb{Z}}$ son innovaciones que siguen un $RBE(0,1)$.*

Notemos que la Hipótesis 2.21 implica que para un modelo GARCH general se tiene que

$$F_{L_{t+1}|\mathcal{F}_t}(l) = P(\sigma_{t+1} Z_{t+1} \leq l) = G\left(\frac{l}{\sigma_{t+1}}\right) \quad (2.37)$$

donde G es la función de distribución de las innovaciones del proceso.

Por lo tanto, el cálculo de medidas de riesgo para la distribución de pérdida condicional para un periodo en este tipo de modelos se puede lograr a través el cálculo de medidas de riesgo para la distribución de las innovaciones, G . Utilizando el enfoque seguido en los Ejemplos 1.5 y 1.9 tenemos:

$$VaR_\alpha^t = \sigma_{t+1} q_\alpha(Z), \quad (2.38)$$

$$ES_\alpha^t = \sigma_{t+1} ES_\alpha(Z), \quad (2.39)$$

donde Z es una variable aleatoria genérica con función de distribución G . En general, para estimar las medidas de riesgo (2.38) y (2.39), necesitamos estimaciones de σ_{t+1} , la desviación estándar condicional para el proceso de pérdida. También requerimos el cuantil y déficit esperado de las innovaciones con función de distribución G . En general $q_\alpha(Z)$ y $ES_\alpha(Z)$ dependen de los parámetros de la distribución de las innovaciones. Para el caso Gaussiano tenemos que $q_\alpha(Z) = \phi^{-1}(\alpha)$ y $ES_\alpha(Z) = \phi(\Phi^{-1}(\alpha))/(1 - \alpha)$. En el caso

de innovaciones t , éstas tendrían que ser escaladas por el factor $\sqrt{(\nu - 2)/\nu}$ para tener varianza 1.

Estrategias concretas de estimación que podríamos adoptar, en orden decreciente de sofisticación, incluyen las siguientes:

1. Ajustar un modelo GARCH con una distribución de innovación apropiada a los datos L_{t-n+1}, \dots, L_t por máxima verosimilitud y utilizar la metodología de predicción discutida en las secciones anteriores para estimar σ_{t+1} . Cualesquiera demás parámetros de la distribución de las innovaciones serán estimados simultáneamente en el ajuste del modelo.
2. Ajustar un modelo GARCH por Quasi-Máxima Verosimilitud y utilizar la metodología de predicción como en la estrategia 1 para estimar σ_{t+1} . En un segundo paso, por separado utilizamos los residuales (estandarizados) del modelo para encontrar una distribución de innovación apropiada y estimar sus parámetros.
3. Utilizar un suavizamiento exponencial de volatilidad para estimar σ_{t+1} . Esta técnica, en conjunto con una hipótesis de innovaciones Gaussianas, es esencialmente el método *RiskMetrics*. En lugar tomar la hipótesis de innovaciones Gaussianas, podríamos estandarizar cada una de las pérdidas L_{t-n+1}, \dots, L_t dividiéndolas entre las estimaciones de la volatilidad $\hat{\sigma}_{t-n+1}, \dots, \hat{\sigma}_t$ calculadas utilizando el suavizamiento exponencial. Esto nos daría un conjunto de residuales, a partir del cual podríamos estimar la distribución de las innovaciones como en la estrategia 2.

2.3.3. *Backtesting*

En esta Sección estudiaremos metodologías específicas para evaluar el desempeño de estimaciones de medidas de riesgo condicionales. Es decir, analizaremos metodologías para evaluar el comportamiento histórico del VaR_α^y y ES_α^t obtenidos a partir del ajuste de un modelo GARCH.

Backtesting para el VaR. Utilizando la notación de la Sección 1.3 primero observemos que si definimos las variables indicadoras $I_{t+1} := I_{\{L_{t+1} > VaR_\alpha^t\}}$, indicando violaciones (excesos) a los cuantiles de la distribución condicional de pérdida, entonces el proceso $\{I_t\}_{t \in \mathbb{Z}}$, es un proceso de variables aleatorias Bernoulli con probabilidad de éxito (es decir, exceso) $1 - \alpha$. Esta propiedad es, ciertamente, verdadera bajo nuestra Hipótesis 2.21, puesto que tenemos $I_{t+1} = I_{\{L_{t+1} > VaR_\alpha^t\}} = I_{\{Z_{t+1} > q_\alpha(Z)\}}$ y las innovaciones $\{Z_t\}_{t \in \mathbb{Z}}$ son ellas mismas i.i.d. Sin embargo, esto también es cierto de forma más general, como

nos dice el siguiente lema.

Lema 2.22. *Sea $\{Y_t\}_{t \in \mathbb{Z}}$ una sucesión de variables indicadoras Bernoulli adaptadas a la filtración $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ que satisfacen $E(Y_t | \mathcal{F}_{t-1}) = p > 0$, $\forall t \in \mathbb{Z}$. Entonces, $\{Y_t\}_{t \in \mathbb{Z}}$ es un proceso de variables Bernoulli i.i.d.*

Demostración. El proceso $\{Y_t - p\}_{t \in \mathbb{Z}}$ tiene la propiedad de diferencia de martingalas pues

$$E(Y_t - p | \mathcal{F}_{t-1}) = E(Y_t | \mathcal{F}_{t-1}) - p = p - p = 0.$$

Más aún, $Var(Y_t - p | \mathcal{F}_{t-1}) = E[E((Y_t - p)^2 | \mathcal{F}_{t-1})] = p(1 - p)$. Así, $\{Y_t - p\}_{t \in \mathbb{Z}}$, y por tanto $\{Y_t\}_{t \in \mathbb{Z}}$, son procesos ruido blanco de variables no correlacionadas. Es bien sabido que variables aleatorias idénticamente distribuidas Bernoulli y no correlacionadas son i.i.d. \square

En la práctica, hacemos estimaciones condicionales del VaR un paso adelante, \widehat{VaR}_α^t , y consideramos las indicadoras de violación

$$\hat{I}_{t+1} := I_{\{L_{t+1} > \widehat{VaR}_\alpha^t\}}. \quad (2.40)$$

Si tenemos éxito al estimar cuantiles condicionales, esperaríamos que las indicadoras de violación empíricas se comportaran como realizaciones de ensayos Bernoulli i.i.d. con probabilidad de éxito $1 - \alpha$.

Para verificar que las indicadoras de violación al VaR un paso adelante soporten dicha hipótesis se les puede aplicar una prueba binomial de dos colas estándar. Ciertamente, si calculamos estimaciones del VaR_α^t para los tiempos $t = 1, \dots, m$, esperaríamos que $\sum_{i=1}^m \hat{I}_i \sim Binomial(m, 1 - \alpha)$. Así, dicha prueba verificaría que el número de violaciones obtenido es, en promedio, igual al esperado para el nivel de confianza con el que estemos trabajando. En consecuencia, rechazar la hipótesis nula de dicha prueba sugeriría cualquiera, subestimación o sobreestimación sistemática del VaR .

El *backtesting* para estimaciones condicionales del VaR para la distribución de pérdida sobre h periodos es más complicado. Para utilizar ideas como las anteriores tendríamos que basar nuestros *backtests* en periodos de tiempo que no se traslaparan. De lo contrario, crearíamos series de variables aleatorias Bernoulli dependientes, para las cuales la inferencia formal sería difícil.

Backtesting para el Expected Shortfall. Comenzamos observando que ES_α^t es el *Expected Shortfall* de la distribución (continua) de pérdida condicional $F_{L_{t+1} | \mathcal{F}_t}$ y definimos $S_{t+1} := (L_{t+1} - ES_\alpha^t)I_{t+1}$, entonces, para un proceso de pérdida arbitrario $\{L_t\}_{t \in \mathbb{Z}}$ el proceso $\{S_t\}_{t \in \mathbb{Z}}$ forma una serie con la propiedad

de diferencia de martingalas, por lo que satisface $E(S_{t+1} | \mathcal{F}_t) = 0$. Más aún, bajo nuestra Hipótesis 2.21 y utilizando (2.38) y (2.39), tenemos:

$$S_{t+1} = \sigma_{t+1}(Z_{t+1} - ES_\alpha(Z))I_{\{Z_{t+1} > q_\alpha(Z)\}},$$

la cual toma la forma de una sucesión de variables de innovación i.i.d. con media cero escalada por la volatilidad. Esto sugiere que, en la práctica, cuando las medidas de riesgo y la volatilidad son estimadas, podríamos formar *residuales de violación* de la forma

$$\hat{R}_{t+1} := \frac{\hat{S}_{t+1}}{\hat{\sigma}_{t+1}}, \quad \hat{S}_{t+1} = (L_{t+1} - ES_\alpha^t)\hat{I}_{t+1}, \quad (2.41)$$

donde \hat{I}_{t+1} es la indicadora de exceso definida en (2.40). Esperamos que éstos se comporten como realizaciones de variables i.i.d. de una distribución con media cero y un átomo de masa de probabilidad de tamaño α en cero. Para probar el comportamiento de media cero podríamos realizar una prueba *bootstrap* sobre los residuales de violación distintos de cero, la cual no haría ninguna suposición sobre su distribución.

Capítulo 3

Teoría de Valores Extremos

En este Capítulo estudiaremos los resultados más importantes de la Teoría de Valores Extremos (TVE). La TVE es una rama de la estadística especializada en el estudio del comportamiento de los eventos asociados a las colas de una distribución (los valores más grandes o pequeños de una variable aleatoria). La TVE tiene aplicaciones en muchas áreas, principalmente en ciencias ambientales, finanzas y seguros.

Cabe mencionar que el principal objetivo de la TVE es predecir la ocurrencia de eventos raros fuera del rango de información conocida (terremotos de fuertes magnitudes, o caídas estrepitosas de un índice o tipo de cambio). Los modelos de la TVE cuentan con dos características fundamentales para cumplir con este objetivo que los vuelen muy atractivos para la estimación de colas: están basados en una fuerte teoría estadística y ofrecen una forma paramétrica de la cola de una distribución. Por lo tanto, estos métodos permiten cierta extrapolación más allá del rango de los datos, lo que permite la obtención de medidas de riesgo extremo.

Existen dos clases principales de modelos para valores extremos. Los modelos más tradicionales son los modelos de *máximos de bloque*; estos son modelos para las observaciones más grandes recolectadas de grandes muestras de datos idénticamente distribuidos.

Por otro lado, un grupo de modelos más moderno y poderoso es el de los modelos para *excesos sobre un umbral*. Estos son modelos para observaciones grandes que exceden algún nivel alto dado y generalmente son considerados más útiles para aplicaciones prácticas debido a su uso más eficiente de la información (muchas veces limitada) sobre valores extremos. Por lo tanto, nuestro estudio de aplicaciones para extremos se centrará en esta clase de

modelos. Adicionalmente, debido a su potencia y practicidad, estos modelos serán de gran relevancia en el Capítulo 4, donde estudiaremos cómo combinarlos con aquellos estudiados en el Capítulo 2.

Ahora bien, a lo largo de este Capítulo estudiaremos los resultados teóricos más importantes en el análisis de valores extremos. Así mismo, estudiaremos cómo ajustar modelos para excesos sobre un umbral a datos; es decir, la estimación de los parámetros de dicha clase de modelos. Nuestro objetivo final será estimar medidas de riesgo extremo (*VaR* y *Expected Shortfall*), a partir los modelos de para excesos ajustados a un conjunto de datos.

Cabe mencionar que los modelos de la TVE constituyen un enfoque no condicional (o estático) de la administración de riesgos. Esto se debe a que bajo dichas metodologías se modela la distribución no condicional del conjunto de datos analizado.

Por lo tanto, y en contraste con las metodologías del Capítulo anterior, en este caso tomamos un enfoque a largo plazo y procedemos a asignar una magnitud a un evento raro específico, tal como la mayor pérdida esperada para un periodo de 5 años (e.g. el tamaño de la máxima pérdida diaria que ocurre en promedio cada 5 años). Este tipo de información puede ser de interés para un administrador de riesgos que desea realizar un análisis de escenarios para ganar intuición sobre la escala del peor caso posible.

El contenido de este Capítulo está tomado principalmente de [2], [9] y [19]. Otras referencias importantes son [20] y [21].

3.1. Máximos: propiedades de convergencia

La mayoría de los resultados relevantes, para la industria financiera y aseguradora, obtenidos de la teoría clásica de probabilidad se basan en sumas de la forma $S_n = \sum_{i=1}^n X_i$. El resultado más importante en este sentido es el Teorema del Límite Central (véase Apéndice A), que nos dice que, para valores de n suficientemente grandes, la suma S_n de observaciones independientes X_1, X_2, \dots, X_n de una v.a. X (las cuales pueden también ser interpretadas como una realización de n variables aleatorias i.i.d.) tiene una distribución aproximadamente normal con media $n\mu_X$ y varianza $n\sigma_X^2$.

No obstante, utilizar modelos basados en sumas funciona de forma adecuada solamente cuando nos interesan los promedios de fenómenos aleatorios; no podemos pretender ampliar el campo de aplicación de las sumas si lo que nos

interesa son eventos extremos, como es nuestro caso.

Para estudiar eventos extremos existe una herramienta que es mucho más adecuada, desde un punto de vista tanto intuitivo como teórico. Esta herramienta se basa en el máximo, M_n , de n v.a.i.i.d. con función de distribución F :

$$M_n = \max\{X_1, X_2, \dots, X_n\}.$$

Intuitivamente, el máximo M_n representa la observación más extrema dentro de un conjunto de n observaciones.

Con estas dos ideas en mente, en las secciones siguientes estudiaremos cómo derivar resultados similares a aquellos entregados por el Teorema del Límite Central para sumas de variables aleatorias i.i.d. pero para el caso de los máximos de dichas variables.

En lo que sigue, trabajaremos con una sucesión de variables aleatorias i.i.d., $\{X_n\}_{n \in \mathbb{Z}}$, con función de distribución F , las cuales serán interpretadas como pérdidas o ganancias. Nuevamente asociaremos un signo positivo a pérdidas y uno negativo a ganancias.

3.1.1. El Teorema de Fisher-Tippet y la Distribución Generalizada de Valores Extremos

En esta Sección estudiaremos uno de los resultados más importantes en la TVE: el Teorema de Fischer-Tippet. Este teorema especifica las tres posibles formas de las distribuciones límite para máximos apropiadamente normalizados de sucesiones de variables aleatorias i.i.d. Por este motivo, estos tres tipos de distribuciones se conocen como distribuciones de valores extremos. Estudiaremos también que podemos identificar estas distribuciones con una clase de distribuciones que posee una cierta condición de estabilidad: las llamadas distribuciones máx-estables.

Ahora bien, puesto que los valores extremos de una distribución ocurren “cerca” del extremo superior de su soporte, la conducta de M_n debe estar, intuitivamente, relacionada con la cola derecha de F cerca de su extremo derecho. Denotaremos por $\bar{F} = 1 - F$ a la cola derecha de la distribución F y por

$$x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$$

al extremo derecho de la distribución F .

Comenzaremos entonces dando las definiciones de tres distribuciones que resultarán muy importantes para el estudio de la convergencia de máximos apropiadamente normalizados.

Definición 3.1. Distribución Fréchet. La función de distribución de una variable aleatoria Fréchet, para $\alpha > 0$, está dada por

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0 \end{cases} \quad (3.1)$$

Definición 3.2. Distribución Gumbel. La función de distribución de una variable aleatoria Gumbel está dada por

$$\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R} \quad (3.2)$$

Definición 3.3. Distribución Weibull. La función de distribución de una variable aleatoria Weibull, para $\alpha > 0$, está dada por

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \quad (3.3)$$

De la definición de estas tres distribuciones de valores extremos es fácil notar que la Fréchet corresponde a un modelo con cota inferior finita y cota superior infinita; la Weibull corresponde a un modelo con cota superior finita y cota inferior infinita; mientras que la Gumbel es no acotada por ambos lados. En la Figura 3.1 podemos observar la gráfica de las funciones de distribución de las tres distribuciones de valores extremos.

La siguiente proposición nos muestra que existe una relación cercana entre las distribuciones Fréchet, Weibull y Gumbel.

Proposición 3.4. Sea X una variable aleatoria. Entonces $X \sim \Phi_\alpha(x) \iff \ln(X^\alpha) \sim \Lambda(x) \iff -\frac{1}{X} \sim \Psi_\alpha(x)$.

Demostración. $X \sim \Phi_\alpha(x) \iff P(X \leq x) = \exp(-x^{-\alpha})$ para $x > 0$ y $P(X \leq x) = 0$ para $x \leq 0 \iff P(-\frac{1}{X} \leq x) = P(X \geq -\frac{1}{x}) = 1 - P(X \leq -\frac{1}{x}) = 1 - 0 = 1$ para $x \geq 0 \iff P(\ln(X^\alpha) \leq x) = P(X^\alpha \leq \exp(x)) = P(X \leq (\exp(x))^{1/\alpha}) = (\exp\{-\exp(x)\})^{-\alpha} = \exp\{-\exp[x(-\alpha)(1/\alpha)]\} = \exp\{-\exp(-x)\} = \Lambda(x)$. \square

Definición 3.5. Distribuciones del mismo tipo. Dos variables aleatorias Z y Y son del mismo tipo si y sólo si $Z \stackrel{d}{=} aY + b$, para $a > 0$ y $b \in \mathbb{R}$. En términos de sus funciones de distribución, F_Z y F_Y , esto significa que $F_Z(x) = F_Y\left(\frac{x-b}{a}\right)$.

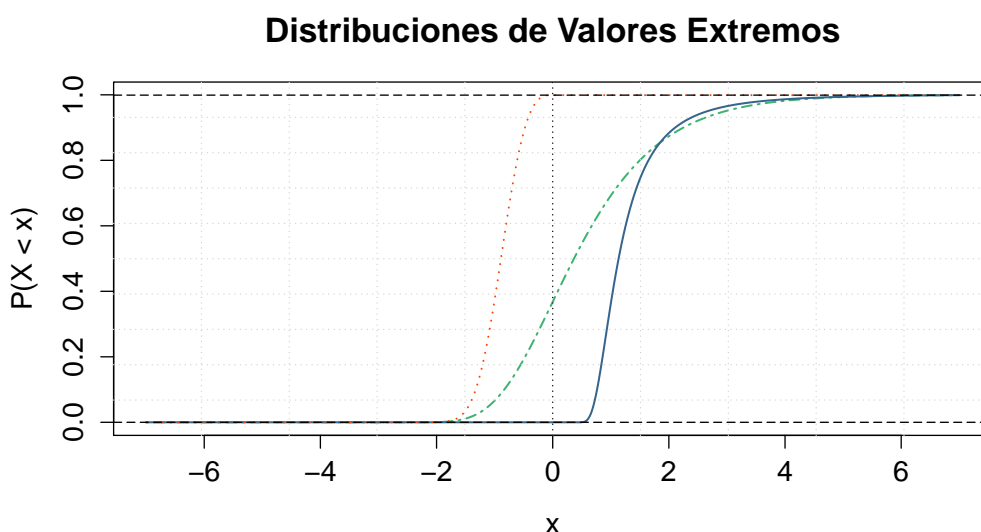


Figura 3.1: Gráficas de las funciones de distribución de las distribuciones de valores extremos. Se muestran: Fréchet (línea azul sólida), Gumbel (línea verde de puntos y guiones) y Weibull (línea roja de puntos); para los casos Fréchet y Weibull se utilizó $\alpha = 3$.

Esto es, dos variables del mismo tipo tienen la misma función de distribución, excepto por posibles cambios en los parámetros de localización y escala.

Una propiedad muy relevante de las distribuciones Fréchet, Weibull y Gumbel es que, cuando consideramos sus máximos, obtenemos que estos son del mismo tipo que la distribución original. Este hecho se formaliza en la siguiente definición.

Definición 3.6. Distribuciones máx-estables. Una variable aleatoria no degenerada X (su función de distribución)¹³ es máx-estable si, para X, X_1, X_2, \dots, X_n i.i.d. y para constantes adecuadas $c_n > 0$ y $d_n \in \mathbb{R}$ satisface

$$\max\{X_1, X_2, \dots, X_n\} \stackrel{d}{=} c_n X + d_n, \quad n \geq 2. \quad (3.4)$$

Ejemplo 3.7. Propiedad máx-estable de la función de distribución Fréchet. Sean X_1, \dots, X_n variables aleatorias independientes con distribu-

¹³ Que una función de distribución sea no degenerada quiere decir que ésta no se concentra en un solo punto.

ción Fréchet Φ_α . Entonces,

$$\begin{aligned} P(M_n \leq x) &= \Phi_\alpha^n(x) &&= (\exp(-x^{-\alpha}))^n \\ &= (\exp(-x^{-\alpha})n) &&= (\exp(-xn^{-1/\alpha})^{-\alpha}) \\ &&&= \Phi_\alpha(n^{-1/\alpha}x). \end{aligned}$$

Por lo tanto, la función de distribución Fréchet es máx-estable.

Ejemplo 3.8. Propiedad máx-estable de la función de distribución Weibull. Consideremos X_1, \dots, X_n variables aleatorias independientes con distribución Weibull Ψ_α . El máximo de estas variables aleatorias i.i.d. satisface:

$$\begin{aligned} P(M_n \leq x) &= \Psi_\alpha^n(x) &&= (\exp(-(-x)^\alpha))^n \\ &= (\exp(-(-x)^\alpha)n) &&= \exp(-(xn^{1/\alpha})^\alpha) \\ &&&= \Psi_\alpha(n^{1/\alpha}x), \end{aligned}$$

de donde es claro que la función de distribución Weibull es máx-estable.

Ejemplo 3.9. Propiedad máx-estable de la función de distribución Gumbel. Supongamos que tenemos X_1, \dots, X_n variables aleatorias independientes con distribución Gumbel Λ . Entonces,

$$\begin{aligned} P(M_n \leq x) &= \Lambda^n(x) &&= (\exp(-e^{-x}))^n \\ &= (\exp(-e^{-x})n) &&= \exp(-e^{-(x+\ln(n))}) \\ &&&= \Lambda(x - \ln(n)). \end{aligned}$$

Así, la función de distribución Gumbel también es máx-estable.

Al introducir las distribuciones Fréchet, Weibull y Gumbel se mencionó que estas resultarían de gran importancia para el estudio de la convergencia de máximos apropiadamente normalizados de variables aleatorias i.i.d. Esta afirmación está parcialmente justificada por el hecho de que las tres distribuciones son máx-estables. Sin embargo, más allá de esto, las distribuciones Fréchet, Weibull y Gumbel aparecen como límites de máximos apropiadamente normalizados de variables aleatorias i.i.d. Los siguientes ejemplos ilustran este hecho.

Ejemplo 3.10. Convergencia de máximos de la distribución exponencial. Sean X_1, \dots, X_n variables aleatorias independientes con distribución exponencial con parámetro λ . Entonces, tomando en cuenta que $P(M_n \leq x) = F^n(x) = (1 - e^{-\lambda x})^n$, tenemos:

$$P\left(M_n \leq \frac{x + \log(n)}{\lambda}\right) = (1 - e^{-\lambda(1/\lambda)(x+\log(n))})^n = \left(1 - \frac{e^{-x}}{n}\right) \rightarrow \exp(-e^{-x}).$$

Por lo tanto, la distribución límite es precisamente la distribución Gumbel Λ .

Ejemplo 3.11. Convergencia de máximos de la distribución uniforme estándar. Supongamos que tenemos X_1, \dots, X_n variables aleatorias independientes con distribución Uniforme en el intervalo $(0, 1)$. Puesto que $P(M_n \leq x) = F^n(x) = x^n$, tenemos:

$$P\left(M_n \leq \frac{x}{n} + 1\right) = \left(\frac{x}{n} + 1\right)^n \rightarrow \exp(-(-x)).$$

Así, la distribución límite corresponde a una distribución Weibull con parámetro $\alpha = 1$, es decir, Ψ_1 .

Ejemplo 3.12. Convergencia de máximos de la distribución Pareto. Sean X_1, \dots, X_n variables aleatorias independientes con distribución Pareto. Utilizando la forma de la función de distribución de dichas variables aleatorias podemos calcular que:

$$\begin{aligned} P(M_n \leq (Kn/\tau)^{1/\alpha}) &= F^n((Kn)^{1/\alpha}x) = [1 - K((Kn)^{1/\alpha}x)^\alpha]^n \\ &= \left(1 - \frac{1}{x^\alpha n}\right)^n \rightarrow \exp(-x^{-\alpha}). \end{aligned}$$

Por lo tanto, la distribución límite obtenida en este ejemplo es precisamente una distribución Fréchet con parámetro α , Φ_α .

Estos tres ejemplos nos han mostrado que podemos obtener a las distribuciones Fréchet, Weibull y Gumbel como límites de máximos apropiadamente normalizados. Surge entonces una pregunta: ¿es posible obtener alguna otra distribución límite no degenerada para máximos apropiadamente normalizados? El siguiente Teorema, que data de 1928, nos otorga la respuesta. Por esta razón es uno de los resultados más importantes de la Teoría de Valores Extremos.

Teorema 3.13. Fisher-Tippett Sean X_1, X_2, \dots, X_n variables aleatoria independientes e idénticamente distribuidas con función de distribución F , y sean $\{a_n\}_{n \in \mathbb{N}}$ y $\{b_n\}_{n \in \mathbb{N}}$ constantes tales que para alguna distribución límite H no degenerada se tiene

$$F^n(a_n x + b_n) = P\left(\frac{M_n - b_n}{a_n}\right) \xrightarrow{d} H(x), \quad x \in \mathbb{R} \quad (3.5)$$

Entonces H es del tipo de alguna de las tres siguientes distribuciones: la Fréchet Φ_α , la Weibull Ψ_α o la Gumbel Λ .

Demostración. La prueba de este importante resultado no se expone en este trabajo debido a la longitud de la misma; en su lugar, referimos al lector

a [2], donde podrá encontrar una demostración sencilla y detallada de este Teorema. \square

Esencialmente, el Teorema de Fisher-Tippet nos dice que si tenemos máximos apropiadamente normalizados que convergen en distribución a un límite no degenerado, entonces este límite debe ser una distribución de valores extremos. En este sentido, el Teorema de Fisher-Tippet nos proporciona un resultado para máximos similar al que ofrece el Teorema del Límite Central para sumas.

Mencionamos antes que las distribuciones Fréchet Φ_α , Weibull Ψ_α y Gumbel Λ (esto es, las tres posibles distribuciones límites no degeneradas para máximos apropiadamente normalizados de variables aleatorias i.i.d.) son máx-estables; es decir, satisfacen que sus máximos son del mismo tipo que la distribución original. El siguiente Teorema nos dice que no existe ningún otro tipo de distribución máx-estable que no sea degenerada.

Teorema 3.14. Propiedad límite de las distribuciones máx-estables.

La clase de distribuciones máx-estables coincide con la clase de las posibles distribuciones límite no degeneradas para máximos de variables aleatorias i.i.d., con la normalización apropiada.

Demostración. Una prueba detallada de este resultado puede encontrarse en [2]. \square

Definiremos ahora una función de distribución que nos permitirá reescribir en una sola expresión a las funciones de distribución Fréchet, Weibull y Gumbel.

Definición 3.15. Distribución Generalizada de Valores Extremos (DGVE). *La función de distribución de valores extremos generalizada está dada por*

$$H_\xi(x) = \begin{cases} \exp\left(-(1 + \xi x)^{-\frac{1}{\xi}}\right) & \text{si } \xi \neq 0 \\ \exp(-e^{-x}) & \text{si } \xi = 0 \end{cases} \quad (3.6)$$

donde $(1 + \xi x) > 0$ y ξ es el parámetro de forma.

El nombre de DGVE proviene del hecho de que la Fréchet, la Weibull y la Gumbel son, de acuerdo con el Teorema de Fisher-Tippet, los tres tipos de distribuciones de valores extremos, las cuales pueden obtenerse de la expresión (3.6) para la DGVE de la siguiente forma:

- $\xi > 0$: Fréchet, $H_\xi\left(\frac{x-1}{\xi}\right) = \Phi_{\frac{1}{\xi}}(x)$.
- $\xi = 0$: Gumbel, $H_0(x) = \Lambda(x)$.

- $\xi < 0$: Weibull, $H_\xi\left(\frac{-(x+1)}{\xi}\right) = \Psi_{-\frac{1}{\xi}}(x)$.

Observación 3.16. La parametrización dada por la DGVE es continua en ξ .

Demostración. En efecto,

$$\lim_{\xi \uparrow 0} \exp\left(-\left(1 + \xi x\right)^{-\frac{1}{\xi}}\right) = \exp(-e^{-x}) = \lim_{\xi \downarrow 0} \exp\left(-\left(1 + \xi x\right)^{-\frac{1}{\xi}}\right).$$

□

También pueden introducirse parámetros de localización μ y escala $\sigma > 0$ en la DGVE. Para ello, se define

$$H_{\xi, \mu, \sigma} := H_\xi\left(\frac{x - \mu}{\sigma}\right).$$

Así, se tiene que $H_{\xi, \mu, \sigma}$ es del mismo tipo que H_ξ .

Utilizando la definición de la DGVE podemos enunciar el Teorema de Fisher-Tippet de dos forma alternativas: Sean X_1, X_2, \dots, X_n variables aleatorias i.i.d. con función de distribución F , y sean $\{a_n\}_{n \in \mathbb{N}}$ y $\{b_n\}_{n \in \mathbb{N}}$ constantes tales que para alguna distribución límite H no degenerada se tiene la relación (3.5). Entonces:

- H es una distribución máx-estable.
- H es del tipo de una distribución GVE, H_ξ , para algún $\xi \in \mathbb{R}$.

Tenemos entonces que de la misma manera que en el Teorema del Límite Central la distribución Normal aparece como límite asintótico para la suma de variables aleatorias i.i.d., sin importar cuál era su distribución original, en el Teorema de Fisher-Tippet las tres distribuciones de valores extremos o, equivalentemente, las tres distribuciones máx-estables (la Fréchet Φ_α , la Weibull Ψ_α y la Gumbel Λ) aparecen como los únicos límites no degenerados posibles para los máximos M_n de variables aleatorias i.i.d., bajo la normalización adecuada.

3.1.2. Dominio de Atracción del Máximo

Ahora que sabemos que las distribuciones de valores extremos son las únicas distribuciones límite para máximos apropiadamente normalizados de variables aleatorias i.i.d. (por el Teorema de Fischer-Tippet), queremos saber bajo

qué condiciones sobre F los máximos M_n convergen débilmente a una distribución de valores extremos H .

Definición 3.17. Dominio de Atracción del Máximo (DAM). Decimos que una variable aleatoria X (su función de distribución F) pertenece al Dominio de Atracción del Máximo (DAM) de la distribución de valores extremos H si existen constantes $c_n > 0$, $d_n \in \mathbb{R}$ tales que $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$. Denotaremos esto $X \in DAM(H)$ ($F \in DAM(H)$).

Esto quiere decir que reunimos en una sola clase a todas las distribuciones F cuyos máximos normalizados tienen el mismo tipo de distribución límite. Para la mayoría de las aplicaciones, es suficiente notar que esencialmente todas las distribuciones continuas comunes están en el $DAM(H_\xi)$ para algún valor de ξ .

El caso Fréchet. Las distribuciones que conducen a la distribución límite Fréchet $H_\xi(x)$ para $\xi > 0$ tienen una caracterización particularmente elegante que involucra funciones de *variación lenta* y *variación regular*.

Definición 3.18. Funciones de variación lenta y variación regular.

i) Una función L positiva y Lebesgue-medible en $(0, \infty)$ es de *variación lenta* en ∞ si

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0.$$

ii) Una función h positiva y Lebesgue-medible en $(0, \infty)$ es de *variación regular* en ∞ con índice $\rho \in \mathbb{R}$ si

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\rho, \quad t > 0.$$

Las funciones de variación lenta son funciones que, en comparación con funciones de potencias, cambian de forma de manera relativamente lenta para x grande; un ejemplo lo constituye la función logaritmo $L(x) = \ln(x)$.

Las funciones de variación regular son funciones que pueden ser representadas por funciones de potencias multiplicadas por funciones de variación lenta, es decir $h(x) = x^\rho L(x)$.

Teorema 3.19. DAM de la distribución Fréchet (Gnedenko). Para $\xi > 0$,

$$F \in DAM(H_\xi) \iff \bar{F}(x) = x^{-1/\xi} L(x) \quad (3.7)$$

para alguna función L de variación lenta en ∞ .

Demostración. Una prueba detallada de este resultado puede encontrarse en [19]. \square

Esto significa que las distribuciones cuyos máximos normalizados apropiadamente dan lugar al caso Fréchet son distribuciones con colas que son funciones de variación regular con un índice de variación negativo. Esto quiere decir que sus colas esencialmente decaen como una función de potencias; la tasa de decaimiento $\alpha = 1/\xi$ es comúnmente conocida como el *índice de cola* de la distribución.

Estas son las distribuciones más estudiadas en la TVE y son de particular interés en aplicaciones financieras porque son distribuciones de cola pesada con momentos de altos órdenes finitos. Si X es una variable aleatoria no negativa cuya función de distribución F es un elemento de $DAM(H_\xi)$ para $\xi > 0$, entonces puede ser demostrado que $E(X^k) = \infty$ para $k > 1/\xi$. Si, para algún $\epsilon > 0$ pequeño, la distribución pertenece a $DAM(H_{(1/2)+\epsilon})$, entonces dicha distribución no tiene varianza finita; si la distribución pertenece a $DAM(H_{(1/4)+\epsilon})$, entonces es una distribución con cuarto momento infinito.

Ejemplos de distribuciones que pertenecen al DAM de la distribución Fréchet incluyen a la propia distribución Fréchet, así como las distribuciones Gamma inversa, t -Student, log-gamma, F y Burr.

El caso Gumbel. La caracterización de las distribuciones en esta clase es más complicada que la del caso Fréchet. En el Ejemplo 3.10 vimos que la distribución exponencial está en el DAM de la distribución Gumbel y, más generalmente, puede decirse que las distribuciones en esta clase tienen colas que tienen esencialmente un decaimiento exponencial. Una variable aleatoria positiva que toma valores en los reales con una función de distribución en $DAM(H_0)$ tiene momentos finitos de cualquier orden, es decir, $E(X^k) < \infty$ para cada $k > 0$. Los detalles sobre esta caracterización de las distribuciones en esta clase pueden encontrarse en [19].

Sin embargo, hay una gran variedad en las colas de las distribuciones en esta clase; así, por ejemplo, las distribuciones normal y log-normal se encuentran ambas en el DAM de la distribución Gumbel. La distribución normal tiene colas ligeras (Véase la Sección 3.1.4 de [9]), pero la distribución log-normal tiene colas mucho más pesadas y podríamos necesitar recolectar muchas observaciones de la distribución log-normal antes de poder distinguir el comportamiento de sus colas de aquel del de una distribución en la clase Fréchet.

En el modelaje financiero comúnmente se asume de forma errónea que los

únicos modelos interesantes para rendimientos financieros son las distribuciones con colas de potencias de la clase Fréchet. La clase Gumbel también es interesante porque contiene muchas distribuciones con colas mucho más pesadas que las de una distribución normal, aún si estas no son funciones de variación regular. Ejemplos de esto son las distribuciones hiperbólicas e hiperbólicas generalizadas (con la excepción del caso especial de frontera, que es la t -Student).

Otras distribuciones en $DAM(H_0)$ incluyen la gamma, ji-cuadrada, la propia distribución Gumbel y las distribuciones Benktander tipo I y II.

El caso Weibull. Este es quizá el caso menos importante para modelaje financiero, al menos en el área de riesgo de mercado, puesto que las distribuciones en esta clase tienen extremos derechos finitos. Aunque la mayoría de todas las pérdidas potenciales en seguros y finanzas están, en la práctica, acotadas, normalmente se siguen prefiriendo modelos que tienen un soporte infinito para el modelaje de pérdidas. Una excepción podría estar en el área de riesgo de crédito, donde distribuciones sobre el intervalo unitario $[0, 1]$ son muy útiles. Una caracterización de la clase Weibull es como sigue:

Teorema 3.20. *DAM de la distribución Weibull, Gnedenko.* Para $\xi < 0$,

$$F \in DAM(H_{1/\xi}) \iff x_F < \infty \text{ y } \bar{F}(x_F - x^{-1}) = x^{1/\xi}L(x)$$

para alguna función L de variación lenta en ∞ .

Demostración. Una demostración detallada de este Teorema puede ser consultada en [19]. \square

Puede ser demostrado (Véase pág. 137 de [19]) que una distribución beta con densidad $f_{\alpha,\beta}$ está en el $DAM(H_{-1/\beta})$. Esto incluye el caso especial de la distribución uniforme para $\beta = \alpha = 1$.

Con la consideración anterior hemos terminado nuestro estudio inicial de la teoría para máximos de variables aleatorias. No obstante, cabe mencionar que existen metodologías que permiten utilizar la Distribución Generalizada de Valores Extremos en la práctica, para modelar los máximos normalizados de pérdidas.

Un ejemplo de tales metodologías es el *Método de Máximos de Bloque*, el cual nos ayuda a estimar los parámetros de la DGVE a partir de un conjunto de datos. Una vez ajustada, es posible estimar los cuantiles de la DGVE para generar medidas de riesgo extremo a partir del método de máximos de bloque.

En este trabajo no profundizamos más en tales metodologías, no obstante, referimos al lector a [2]. Ahí, podrá estudiar a profundidad el método de máximos de bloque, así como consultar un par de aplicaciones del mismo a datos reales.

3.2. Modelos para excesos sobre un umbral

Un grupo de modelos más moderno que el de bloques es el de los modelos de Picos Sobre el Umbral (POT por sus siglas en inglés; *Peaks Over Threshold*), que son modelos para todas las observaciones (grandes) que exceden un umbral *alto*. Los modelos POT se consideran más útiles que los de bloques para aplicaciones prácticas en general, debido a que utilizan los datos de valores extremos, a menudo limitados, de forma más eficiente.

Dentro de la clase de los modelos POT pueden distinguirse dos tipos de análisis: los modelos semi-paramétricos (construidos alrededor del estimador de Hill y sus similares) y los modelos completamente paramétricos. Si bien ambos enfoques están justificados teóricamente y son útiles en la práctica cuando se utilizan de forma correcta, en este trabajo nos centraremos únicamente en los modelos paramétricos. A través de dichos modelos, es posible obtener fórmulas paramétricas sencillas de medidas de riesgo extremo, para las cuales es relativamente fácil dar estimaciones del error estadístico utilizando técnicas de inferencia máximo verosímil.

3.2.1. El Teorema de Gnedenko-Balkema-Pickands-de Haan y la Distribución Pareto Generalizada

En esta Sección daremos una breve introducción a la Distribución Pareto Generalizada (DPG), la cual resultará ser un modelo natural para los excesos por encima de un umbral alto u . Además, veremos que la DPG está íntimamente relacionada con las condiciones del Dominio de Atracción del Máximo de la Distribución Generalizada de Valores Extremos ($DAM(H_\xi)$) que analizamos en la Sección anterior.

Precisemos, antes que nada, la noción de exceso.

Definición 3.21. *Función de Distribución de Exceso.* Sea X una variable aleatoria con función de distribución F y extremos derecho x_F . Para

u fijo, $u < x_F$, decimos que

$$F_u(x) = P(X - u \leq x \mid X > u), \quad x \geq 0 \quad (3.8)$$

es la función de distribución de exceso sobre el umbral u de la variable aleatoria X .

Definición 3.22. Función Media de Exceso. La función media de exceso de una variable aleatoria X con función de distribución F está dada por

$$e(u) = E(X - u \mid X > u). \quad (3.9)$$

Los excesos por encima de un umbral u son fundamentales en muchos campos. Por ejemplo, en el lenguaje utilizado en estadísticas médicas F_u se conoce como la función de distribución de *exceso de vida* o de *vida residual*; en el contexto de seguros, se utiliza el término de función de distribución de *exceso de pérdida* para referirse a F_u .

El cálculo de la función media de exceso $e(u)$ puede realizarse utilizando la definición de $e(u)$ e integración por partes. Supongamos que X es una variable aleatoria positiva con función de distribución F y esperanza finita. Entonces

$$\begin{aligned} e(u) &= \int_u^{x_F} (x - u) dF(x) / \bar{F}(u) \\ &= \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x) dx, \quad x > 0. \end{aligned} \quad (3.10)$$

Si F es continua,

$$\bar{F}(x) = \frac{e(0)}{e(x)} \exp \left\{ - \int_0^x \frac{1}{e(u)} du \right\}, \quad x > 0. \quad (3.11)$$

De la expresión (3.11) se sigue que una función de distribución continua queda determinada de manera única por su función media de exceso.

Presentaremos ahora una nueva familia de funciones de distribución, la cual, como veremos más adelante, está muy relacionada con los excesos por encima de un umbral alto.

Definición 3.23. Distribución Pareto Generalizada (DPG). La función de distribución G_ξ definida como

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - e^{-x} & \text{si } \xi = 0 \end{cases}, \quad (3.12)$$

donde

$$\begin{aligned} x &\geq 0 && \text{si } \xi \geq 0 \\ 0 \leq x &\leq -1/\xi && \text{si } \xi < 0 \end{aligned}$$

se conoce como la *Distribución Pareto Generalizada (DPG)*. Si se reemplaza el argumento x por $(x - \nu)/\beta$ para $\nu \in \mathbb{R}$ y $\beta > 0$ se obtiene (haciendo los cambios de soporte pertinentes) la familia $G_{\xi, \nu, \beta}$ con parámetros de localización y escala; también nos referimos a $G_{\xi, \nu, \beta}$ como *DPG*.

Igual que en el caso de la distribución GVE H_0 , podemos interpretar a G_0 como el límite de G_ξ cuando $\xi \rightarrow 0$.

Más adelante veremos que la función de distribución $G_{\xi, 0, \beta}$ juega un papel importante en la práctica, puesto que el ajuste de DPGs es uno de los conceptos más importantes en la estadística de eventos extremos. Abusando un poco de la notación, la denotaremos por:

$$G_{\xi, \beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad x \in D(\xi, \beta)$$

donde

$$D(\xi, \beta) = \begin{cases} [0, \infty) & \text{si } \xi \geq 0 \\ (0, -\beta/\xi) & \text{si } \xi < 0. \end{cases} \quad (3.13)$$

Así, cuando digamos que X tiene una DPG con parámetros ξ y β , deberá entenderse que X tiene función de distribución $G_{\xi, \beta}$.

En la Figura 3.2 se presentan gráficas de la DPG para tres valores distintos de ξ ; obsérvese la diferencia en las escalas. En todos los casos se tiene $\beta = 1$ y $\nu = 0$.

Los siguientes ejemplos nos darán una idea intuitiva de la importancia de la DPG en el contexto de los excesos sobre un umbral alto u .

Ejemplo 3.24. Distribución exponencial. Tenemos $\bar{F}(x) = e^{-\lambda x}$, con $x, \lambda > 0$. Consideremos la siguiente expresión:

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \frac{e^{-\lambda[u+xa(u)]}}{e^{-\lambda u}} = e^{-\lambda xa(u)}.$$

Si ahora tomamos $a(u) = \lambda^{-1}$, entonces se obtiene

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = e^{-x}.$$

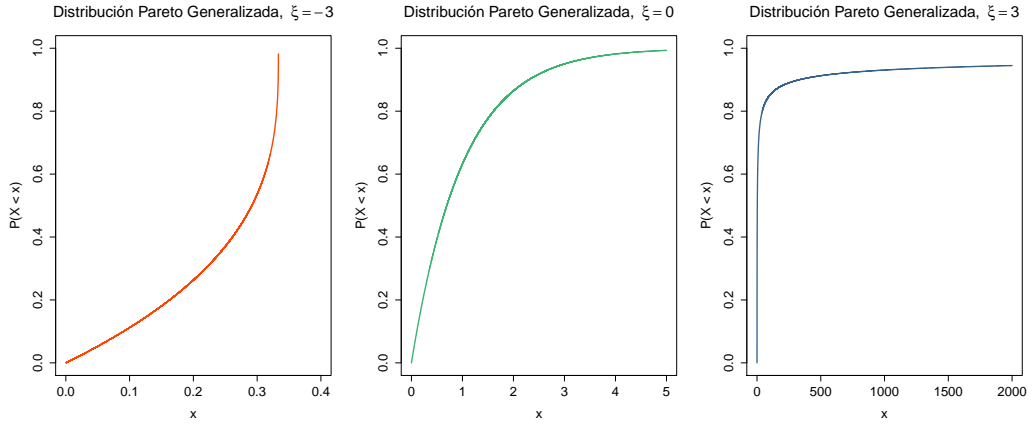


Figura 3.2: Ejemplos de la gráfica de Distribuciones Pareto Generalizadas para distintos valores del parámetro de forma ξ . Se muestra: DPG con $\xi = -3$ (izquierda), DPG con $\xi = 0$ (centro) y DPG con $\xi = 3$ (derecha); en todos los casos se utilizó $\beta = 1$ y $\nu = 0$.

Obsérvese que este es precisamente el valor de la cola de la DPG en el caso en que $\xi = 0$.

Ejemplo 3.25. Distribución Uniforme estándar. Se tiene $\bar{F}(x) = 1 - x$, con $0 < x < 1$. Entonces:

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \frac{1 - (u + xa(u))}{1 - u} = 1 - \frac{a(u)}{1 - u}x.$$

Sea $a(u) = 1 - u$, así

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = 1 - x.$$

Este resultado corresponde a la cola de una DPG con $\xi = -1$.

Ejemplo 3.26. Distribución Pareto. En este caso, $\bar{F}(x) = Kx^{-\alpha}$, con $K, \alpha > 0$, $x \geq K^{-1/\alpha}$. Evaluamos:

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \frac{K(u + xa(u))^{-\alpha}}{Ku^{-\alpha}} = \left(\frac{u + xa(u)}{u}\right)^{-\alpha} = \left(1 + x\frac{a(u)}{u}\right)^{-\alpha}.$$

Si hacemos $a(u) = u/\alpha$ y $\alpha = 1/\xi$, se obtiene

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \left(1 + \frac{x}{\alpha}\right)^{-\alpha} = (1 + \xi x)^{-1/\xi}.$$

Nótese que esto es la cola de una DPG donde $\xi > 0$.

Los resultados obtenidos en estos ejemplos no son casualidad. Habíamos visto antes, en el Ejemplo 3.10, que la función de distribución exponencial está en el $DAM(\Lambda)$, lo cual, hablando en términos de la distribución GVE H_ξ , puede expresarse diciendo que la exponencial está en el $DAM(H_0)$. También vimos, en el Ejemplo 3.12, que la función de distribución Pareto está en el $DAM(\Phi_\alpha)$ para $\alpha > 0$, lo cual puede reformularse diciendo que la función de distribución Pareto está en el $DAM(H_\xi)$ para $\xi > 0$. Por último, de acuerdo con el Ejemplo 3.11, la función de distribución Uniforme estándar está en el $DAM(\Psi_1)$, es decir, en el $DAM(H_{-1})$.

El siguiente Teorema, que uno de los resultados fundamentales en la TVE, nos confirma la relación existente entre el $DAM(H_\xi)$ y la DPG.

Teorema 3.27. Gnedenko-Balkema-Pickands-de Haan. Sea $\xi \in \mathbb{R}$. Son equivalentes:

- a) $F \in DAM(H_\xi)$
- b) Existe una función positiva y medible $a(\cdot)$ tal que para $1 + \xi x > 0$ se tiene que

$$\lim_{u \uparrow x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \bar{G}_\xi(x) = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ e^{-x} & \text{si } \xi = 0. \end{cases} \quad (3.14)$$

- c) Para $x, y > 0$, $y \neq 1$,

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & \text{si } \xi \neq 0 \\ \frac{\ln(x)}{\ln(y)} & \text{si } \xi = 0 \end{cases}, \quad (3.15)$$

donde $U(t) = F^{\leftarrow}(1 - t^{-1})$, $t > 0$ para F^{\leftarrow} la función cuantil de una función de distribución F .

Demostración. La demostración de este resultado puede ser consultada en [2]. □

Observación 3.28. La condición expresada en (3.14) tiene una interpretación probabilística muy útil e interesante: sea X una variable aleatoria con función de distribución $F \in DAM(H_\xi)$, entonces, (3.14) puede reformularse de la siguiente forma:

$$\lim_{x \uparrow x_F} P \left(\frac{X - u}{a(u)} > x \mid X > u \right) = \begin{cases} (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ e^{-x} & \text{si } \xi = 0 \end{cases} \quad (3.16)$$

De modo que (3.16) da una aproximación (en escala) a la distribución de los excesos sobre un umbral alto u , donde el factor de escala es $a(u)$. Esta interpretación es crucial para muchas de las aplicaciones de la TVE que veremos más adelante. Esta importante relación puede reescribirse del siguiente modo:

$$\lim_{u \uparrow x_F} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0.$$

El Teorema anterior nos ha hecho ver la importancia de la DPG. El siguiente resultado menciona algunas de sus propiedades básicas, las cuales serán ampliamente utilizadas más adelante.

Teorema 3.29. Propiedades de la Distribución Pareto Generalizada. Supongamos que X tiene una DPG con parámetros ξ y β . Entonces:

a) Para cada $\xi \in \mathbb{R}$, $F \in DAM(H_\xi)$ si y sólo si

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0,$$

para alguna función positiva β .

b) Supongamos que $x_i \in D(\xi, \beta)$, $i = 1, 2$; entonces

$$\frac{\bar{G}_{\xi, \beta}(x_1 + x_2)}{\bar{G}_{\xi, \beta}(x_1)} = \bar{G}_{\xi, \beta + \xi x_1}(x_2). \quad (3.17)$$

c) Supongamos que X tiene DPG con parámetros $\xi < 1$ y β . Entonces, para $u < x_F$,

$$e(u) = E[X - u | X > u] = \frac{\beta + \xi u}{1 - \xi}, \quad \beta + \xi u > 0.$$

Demostración. Una demostración de este hecho puede ser consultada en la página 165 de [19]. \square

Observación 3.30. La propiedad b) arriba mencionada puede reformularse como sigue: la clase de las DPGs es cerrada con respecto a cambios en el umbral. Esto puede verse de la siguiente forma: el lado izquierdo de la expresión (3.17) es la probabilidad condicional de que, dado que nuestra variable aleatoria subyacente excede x_1 , también exceda el umbral $x_1 + x_2$. El lado derecho de (3.17) nos dice que esta probabilidad es nuevamente de tipo Pareto Generalizada. Esta propiedad de cerradura es muy importante en reaseguro, donde las DPGs se utilizan en los contratos de exceso de pérdida.

Observación 3.31. *La propiedad a) nos sugiere a la DPG como una aproximación adecuada de la función de distribución de exceso, F_u , para u grande. Este resultado puede reformularse de la siguiente manera: para alguna función β , que deberá estimarse de los datos, se tiene:*

$$\bar{F}_u(x) = P(X - u > x | X > u) \approx G_{\xi, u, \beta(u)}(x).$$

De forma alternativa, puede considerarse, para $x > u$, que:

$$P(X > x | X > u) \approx \bar{G}_{\xi, u, \beta(u)}(x).$$

En ambos casos, u debe tomarse suficientemente grande.

Observación 3.32. *Por otro lado, la propiedad a) también nos permite ver de forma clara la relación que guardan la DPG y la DGVE: si la distribución de los máximos apropiadamente normalizados de una distribución converge a H_ξ , entonces la distribución de sus excesos a un umbral alto, u , convergerá a una $G_{\xi, u, \beta(u)}(x)$ y viceversa. Nótese que la DPG y la DGVE tienen el mismo parámetro de forma, ξ .*

Observación 3.33. *Juntas, las propiedades a) y c) nos proporcionan un buen método gráfico para elegir el umbral u suficientemente grande como para justificar una aproximación de la función de distribución de exceso mediante una DPG: dada una muestra i.i.d. X_1, \dots, X_n , construimos la función media de exceso empírica $e_n(u)$ como una versión muestral de la función media de exceso $e(u)$. Por c) sabemos que la función media de exceso de una DPG es lineal, de forma que buscamos una región de valores de u donde la gráfica de $e_n(u)$ sea aproximadamente lineal. Para una u en esa región parece razonable aproximar F_u mediante una DPG. Este enfoque se utilizará en los Capítulos siguientes para ajustar DPGs a excesos por encima de umbrales altos.*

Hemos visto que cuando $\xi > 0$ la DPG es de cola pesada, por lo que este es el caso más relevante (más no el único de interés) para la administración de riesgos. Mientras que la distribución normal tiene momentos de todos los órdenes, una distribución de cola pesada no tiene un conjunto de momentos. En el caso de la DPG con $\xi > 0$, tenemos que $E[X^k]$ es infinita para $k \geq 1/\xi$. Cuando $\xi = 1/2$, la DPG es una distribución con segundo momento (varianza) infinito; cuando $\xi = 1/4$, la DPG tiene cuarto momento infinito. Algunos tipos de datos de grandes reclamaciones de seguros sugieren un segundo momento infinito; de la misma manera, ciertos rendimientos de mercado indican una distribución con cuarto momento infinito. La distribución normal no puede modelar estos fenómenos, pero la DPG se utiliza precisamente para capturar este tipo de conducta.

Ahora bien, recapitulando el material estudiado hasta este momento tenemos que: hemos visto que la DPG $G_{\xi,\beta}$, con $\xi \in \mathbb{R}$ y $\beta > 0$, se obtiene como la distribución límite de los excesos de una variable aleatoria (bajo escala) por encima de un umbral. Adicionalmente, estudiamos algunas de las propiedades probabilísticas de este tipo de distribuciones generalizadas; propiedades serán de gran importancia para el análisis estadístico de eventos extremos. Por otro lado, el Teorema 3.27 nos dio las bases teóricas para poder estimar la cola y altos cuantiles para este tipo de datos. Más aún, la parte *b*) de ese mismo Teorema nos guio directamente a la definición de la DPG en la cual se basa el método de Picos Sobre el Umbral (POT, *Peaks Over Thresholds*) para aproximar la función de distribución de exceso.

De esta manera, a continuación estudiaremos de forma profunda el Método POT. Dicha metodología nos permitirá utilizar la Distribución Pareto Generalizada para estimar medidas de riesgo extremo a partir de un conjunto de datos.

3.2.2. El método POT

La meta del método POT es obtener estimadores de ES_α y VaR_α , para α grande (es decir, $\alpha \in (0.95, 1)$). Para ello, utiliza el Teorema de Gnedenko-Balkema-Pickands-de Haan estudiado en la Sección anterior para aproximar la distribución de exceso F_u por encima de un umbral determinado u mediante una Distribución Pareto Generalizada. Una vez que se tiene el estimador de la distribución de exceso, resulta sencillo reformular las expresiones probabilísticas de nuestro interés en términos de la cola \bar{F} de la distribución. De esta forma, puesto que el VaR_α es un alto cuantil de la distribución, éste puede obtenerse invirtiendo la fórmula para el estimador de colas que obtenemos. Así, podemos también estimar el valor de ES_α utilizando la relación entre éste y el VaR_α :

$$ES_\alpha = VaR_\alpha + E[X - VaR_\alpha | X > VaR_\alpha].$$

Estos pasos se explican detalladamente a continuación.

Supongamos que X, X_1, \dots, X_n son i.i.d. con función de distribución $F \in DAM(H_\xi)$ para algún $\xi \in \mathbb{R}$.

En general, supondremos que la distribución F subyacente es una distribución con extremo derecho infinito, es decir, que permite la posibilidad de que ocurran pérdidas arbitrariamente grandes, aun cuando atribuye probabilidades insignificantes a los resultados irrazonablemente grandes, como lo hacen

la distribución normal y la distribución t -Student. Sin embargo, también puede darse el caso de que se utilice F con extremo derecho finito. Un ejemplo de esto es la distribución beta en el intervalo $[0, 1]$, que asigna probabilidad cero a los resultados mayores que 1 y que podría usarse, por ejemplo, como la distribución de pérdidas de crédito expresadas en proporción a la exposición.

Empezamos por elegir un umbral u y denotamos

$$N_u = \text{card}\{i : i = 1, \dots, n, \quad X_i > u\}$$

el número de excesos de u por parte de X_1, \dots, X_n . Denotemos a los correspondientes excesos Y_1, \dots, Y_{N_u} .

Anteriormente (ver Definición 3.21) definimos la distribución de pérdidas excesivas por encima de un umbral alto u mediante

$$F_u(y) = P(X - u \leq y \mid X > u) = P(Y \leq y \mid X > u), \quad 0 \leq y < x_F - u. \quad (3.18)$$

La distribución de exceso representa la probabilidad de que una pérdida exceda el umbral u por a lo más una cantidad y , dado que sabemos que ha excedido el umbral. Es muy útil notar que la distribución de exceso F_u puede escribirse en términos de la distribución F subyacente:

$$\bar{F}(u + y) = \bar{F}(u)\bar{F}_u(y). \quad (3.19)$$

Recordemos ahora la definición de la Distribución Pareto Generalizada (véase Definición 3.23). El Teorema Gnedenko-Balkema-Pickands-de Haan nos dice que si X_1, \dots, X_n son i.i.d. con función de distribución $F \in \text{DAM}(H_\xi)$, $\xi \in \mathbb{R}$, entonces existe una función $\beta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ tal que

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0. \quad (3.20)$$

Basándonos en este resultado límite para $\bar{F}_u(y)$, podemos pensar en la siguiente aproximación para valores grandes de u :

$$\bar{F}_u(y) \approx \bar{G}_{\xi, \beta(u)}(y). \quad (3.21)$$

Es importante notar que β es una función del umbral u . En la práctica, u debe tomarse suficientemente grande. Dado un valor (grande) de u , ξ y $\beta = \beta(u)$ se estiman a partir de los datos de excesos, por lo que los estimadores resultantes dependen de u .

La relación (3.19) nos deja ver un método para estimar la cola de F mediante la estimación, por separado, de $\bar{F}_u(y)$ y $\bar{F}(u)$. Un estimador natural para $\bar{F}(u)$ está dado por la función de distribución empírica

$$\left(\widehat{\bar{F}(u)}\right) = \bar{F}_n(u) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i > u\}} = \frac{N_u}{n}.$$

Esto está justificado por el Teorema de Glivenko-Cantelli (ver Apéndice A), que nos dice que la función de distribución empírica es una aproximación uniforme a la verdadera función de distribución.

Por otra parte, la aproximación de la distribución de exceso mediante una DPG dada en la relación (3.21) (la cual podemos utilizar porque u es grande) motiva un estimador de la forma

$$\left(\hat{\bar{F}}_u(y)\right) \approx \bar{G}_{\hat{\xi}, \hat{\beta}}(y) \quad (3.22)$$

para $\hat{\xi} = \hat{\xi}_{N_u}$ y $\hat{\beta} = \hat{\beta}_{N_u}$ apropiados. Más adelante estudiaremos un método para estimar los parámetros de la DPG.

Una pregunta que surge en este momento es: ¿por qué no utilizamos el estimador empírico (o, dicho de otra manera, el método de simulación histórica) para estimar toda la cola de F , es decir, $F(x)$ para toda $x \geq u$? La respuesta es que la simulación histórica es un método bastante pobre en la cola de la distribución, pues los datos se vuelven escasos.¹⁴ Al fijar el umbral en u estamos considerando que se tienen suficientes observaciones que exceden a u como para permitir que obtengamos un estimador razonable de $\bar{F}(u)$ a través de simulación histórica, pero para niveles más altos el método es muy poco confiable.

Al sustituir nuestro estimador de $\bar{F}(u)$ obtenido por simulación histórica, junto con el estimador sugerido por (3.22), se obtiene el estimador de la cola $\bar{F}(u + y)$:

$$\left(\hat{\bar{F}}(u + y)\right) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{y}{\hat{\beta}}\right)^{-1/\hat{\xi}}, \quad y > 0. \quad (3.23)$$

¹⁴ En el enfoque de simulación histórica, el estimador de la distribución de pérdida de un activo está dado simplemente por la distribución empírica de las pérdidas y ganancias pasadas del activo. Este enfoque está motivado por el Teorema de Glivenko-Cantelli, el cual nos dice que dicho estimador de la función de distribución es consistente. Los detalles sobre el Teorema de Glivenko-Cantelli pueden consultarse en el Apéndice A.

Este estimador puede verse como un tipo de estimador por simulación histórica aumentado con TVE y puede construirse siempre que creamos que los datos vienen de una distribución común, aunque sus propiedades estadísticas se entienden mejor en las situaciones en que los datos se suponen independientes o sólo “levemente dependientes”.

Es fácil notar que $\hat{F}(u + y)$ es también una DPG con el mismo parámetro de forma ξ , pero con parámetro de escala $\hat{\beta} = \beta(1 - F(u))^\xi$ y parámetro de localización $\tilde{\nu} = \nu - \hat{\beta} [(1 - F(u))^{-\xi} - 1] / \xi$.

Lo que se está haciendo es extrapolar al área en que los datos se vuelven más escasos y son, por tanto, una guía poco confiable para su distribución.

En los casos Fréchet y Gumbel ($\xi \geq 0$), la restricción para el dominio en (3.23) es $y \geq 0$, enfatizando el hecho de que estimamos \bar{F} en la cola superior (o cola derecha; asociada a pérdidas en nuestro caso).

Ahora, dada una probabilidad $\alpha > F(u)$, lo único que necesitamos para obtener el estimador \widehat{VaR}_α del VaR_α es invertir la fórmula (3.23) del estimador de colas. Así,

$$\frac{\hat{F}(u + y)}{\frac{N_u}{n}} = \left(1 + \hat{\xi} \frac{y}{\hat{\beta}}\right)^{-1/\hat{\xi}},$$

de donde

$$\frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} \hat{F}(u + y) \right)^{-\hat{\xi}} - 1 \right) = y. \quad (3.24)$$

Sea $x = u + y$. Entonces, al hacer $\hat{F}(u + y) = \hat{F}(x) = 1 - \alpha$ se tiene que $x = y + u = \widehat{VaR}_\alpha$, y al sustituirlo en (3.24) se obtiene la fórmula de estimación del VaR_α :

$$\widehat{VaR}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u} (1 - \alpha) \right)^{-\hat{\xi}} - 1 \right). \quad (3.25)$$

Más aún, si $\hat{\xi} < 0$, un estimador del extremo derecho x_F de F está dado por

$$x_F = u - \frac{\hat{\beta}}{\hat{\xi}}.$$

Esta relación se obtiene haciendo $\hat{x}_F = \hat{x}_1$ (es decir, $\alpha = 1$) en la relación (3.25).

Como dijimos antes, la meta del método POT es la obtención de un estimador de la Esperanza Condicional de la Cola (ES_α). Tenemos, además, que ES_α y VaR_α están relacionados por la siguiente expresión

$$ES_\alpha = VaR_\alpha + E[X - VaR_\alpha | X > VaR_\alpha].$$

Es fácil notar que el término $E[X - VaR_\alpha | X > VaR_\alpha]$ es simplemente la esperanza de la distribución de exceso $F_{VaR_\alpha}(y)$ por encima del umbral VaR_α . Además, nuestro modelo (3.21) para la distribución de exceso por encima del umbral u tiene la propiedad de que, para umbrales más altos, VaR_α con $\alpha > F(u)$, la distribución de exceso por encima de este umbral más alto es también una DPG con el mismo parámetro de forma, aunque con escala distinta. Esto se demostró en el Teorema 3.27, y un argumento alternativo es: si suponemos que los excesos sobre el umbral u tienen distribución DPG, es decir, $(X - u | X > u) \sim G_{\xi, \beta}$, basta notar que para $VaR_\alpha > u$ podemos escribir

$$X - VaR_\alpha | X > VaR_\alpha = (X - u) - (VaR_\alpha - u) | (X - u) > (VaR_\alpha - u),$$

de donde es fácil ver que

$$F_{VaR_\alpha}(y) \approx G_{\xi, \beta + \xi(VaR_\alpha - u)}(y). \quad (3.26)$$

La ecuación (3.26) nos proporciona un modelo explícito para las pérdidas de exceso por encima del VaR_α . Con este modelo podemos calcular muchas de las características de este tipo de pérdidas; por ejemplo, si notamos que (para $\xi < 1$) la función media de exceso de la distribución dada en (3.26) es (utilizando la definición de función media de exceso para una DPG e integración por partes)

$$\frac{\beta + \xi(VaR_\alpha - u)}{1 - \xi},$$

podemos calcular el estimador de ES_α . Se obtiene que

$$\begin{aligned} \widehat{ES}_\alpha &= \widehat{VaR}_\alpha + e(\widehat{VaR}_\alpha) \\ &= \widehat{VaR}_\alpha + \frac{\hat{\beta} + \hat{\xi}(\widehat{VaR}_\alpha - u)}{1 - \hat{\xi}} \\ &= \frac{\widehat{VaR}_\alpha}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}} \end{aligned}$$

Reacomodando, tenemos

$$\frac{\widehat{ES}_\alpha}{\widehat{VaR}_\alpha} = \frac{1}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{(1 - \hat{\xi})\widehat{VaR}_\alpha}. \quad (3.27)$$

Vale la pena examinar esta razón de forma detenida en los casos en que la distribución subyacente tiene un extremo derecho x_F infinito. En este caso la razón queda mayormente determinada por el factor $1/(1 - \hat{\xi})$. El segundo término del lado derecho de (3.27) se vuelve insignificanemente pequeño conforme la probabilidad α se va acercando más y más a 1. Esta observación asintótica subraya la importancia del parámetro de forma ξ en la estimación de la cola, pues es este parámetro el que determina cómo difieren nuestras dos medidas de riesgo en la regiones extremas de la distribución de pérdida.

Ahora que obtuvimos un estimador de ES_α , debemos subrayar que nunca debe hacerse un análisis para un único valor de u . De hecho, es recomendable hacer gráficas para comparar las estimaciones resultantes a través de un amplio rango de valores de u . La mayoría de las gráficas exhibirán una conducta de este tipo: gran variabilidad para valores grandes de u (pocas observaciones) contra sesgo para valores pequeños de u (muchas observaciones pero, al mismo tiempo, puede no ser válida la aproximación (3.21)).

Existen técnicas gráficas que pueden ayudarnos a elegir un valor adecuado para u . Una de estas herramientas es la gráfica de la función media de exceso, la cual describimos a continuación.

Primeramente, recordemos de la Definición 3.22 que para una variable aleatoria X con extremo derecho x_F , su función media de exceso está dada por

$$e(u) = E[X - u | X > u], \quad 0 \leq u \leq x_F.$$

Para cada u fijo, decimos que la cantidad $e(u)$ es el exceso medio sobre el umbral u .

La función media de exceso describe el exceso esperado de un umbral u , dado que sabemos que éste ha sido excedido.

Podemos hacer un análisis gráfico de la conducta de la cola utilizando la función muestral media de exceso $e_n(u)$, que es un estimador empírico de la función media de exceso dada en (3.9). La función muestral media de exceso se define como

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)_+}{\sum_{i=1}^n I_{\{X_i > u\}}}, \quad (3.28)$$

es decir, es la suma de los excesos por encima del umbral u dividida entre el número de puntos que exceden este umbral.

Supongamos que tenemos la muestra X_1, \dots, X_n . Como de costumbre, ordenamos esta muestra para obtener los estadísticos de orden $X_{1,n} > \dots > X_{n,n}$. Se procede entonces a hacer la gráfica de la función media de exceso, dada por

$$\{(u, e_n(u)) \mid X_{n,n} < u < X_{1,n}\}. \quad (3.29)$$

La interpretación de la gráfica de $e_n(u)$ es la siguiente: si los puntos muestran una tendencia hacia arriba, entonces hay señales de conducta de cola pesada. Los datos con distribución exponencial darían una recta aproximadamente horizontal; mientras que los datos de una distribución de cola corta mostrarían una tendencia hacia abajo. En particular, si la gráfica empírica (3.29) parece seguir una línea razonablemente recta con pendiente positiva por encima de un cierto valor de u , entonces podemos suponer que los datos siguen, en el área de la cola y por encima del valor del umbral u , una DPG con parámetro de forma ξ positivo.

Notemos que, puesto que cerca de la mayor observación $X_{1,n}$ son muy pocos los datos que forman parte del cálculo de $e_n(u)$, existe un aumento en la variabilidad (sensibilidad a cambios en los datos) cerca del extremo superior de la gráfica. En ocasiones conviene detener el proceso de graficación en $X_{4,n}$ o $X_{3,n}$ para evitar esta conducta errática.

En la Figura 3.3 podemos observar el comportamiento teórico de la gráfica de función muestral media de exceso para seis distribuciones: beta, normal, log-normal, exponencial, t -Student y Pareto.

Tenemos entonces que, para la utilización del método POT, necesitamos:

- Un umbral u suficientemente alto.
- Valores estimados $\hat{\xi}$ y $\hat{\beta}$ para la DPG.

Una vez comentadas las dificultades para elegir un umbral “óptimo”, a continuación, presentamos uno de los métodos más conocidos para obtener estimaciones de los parámetros de la DPG.

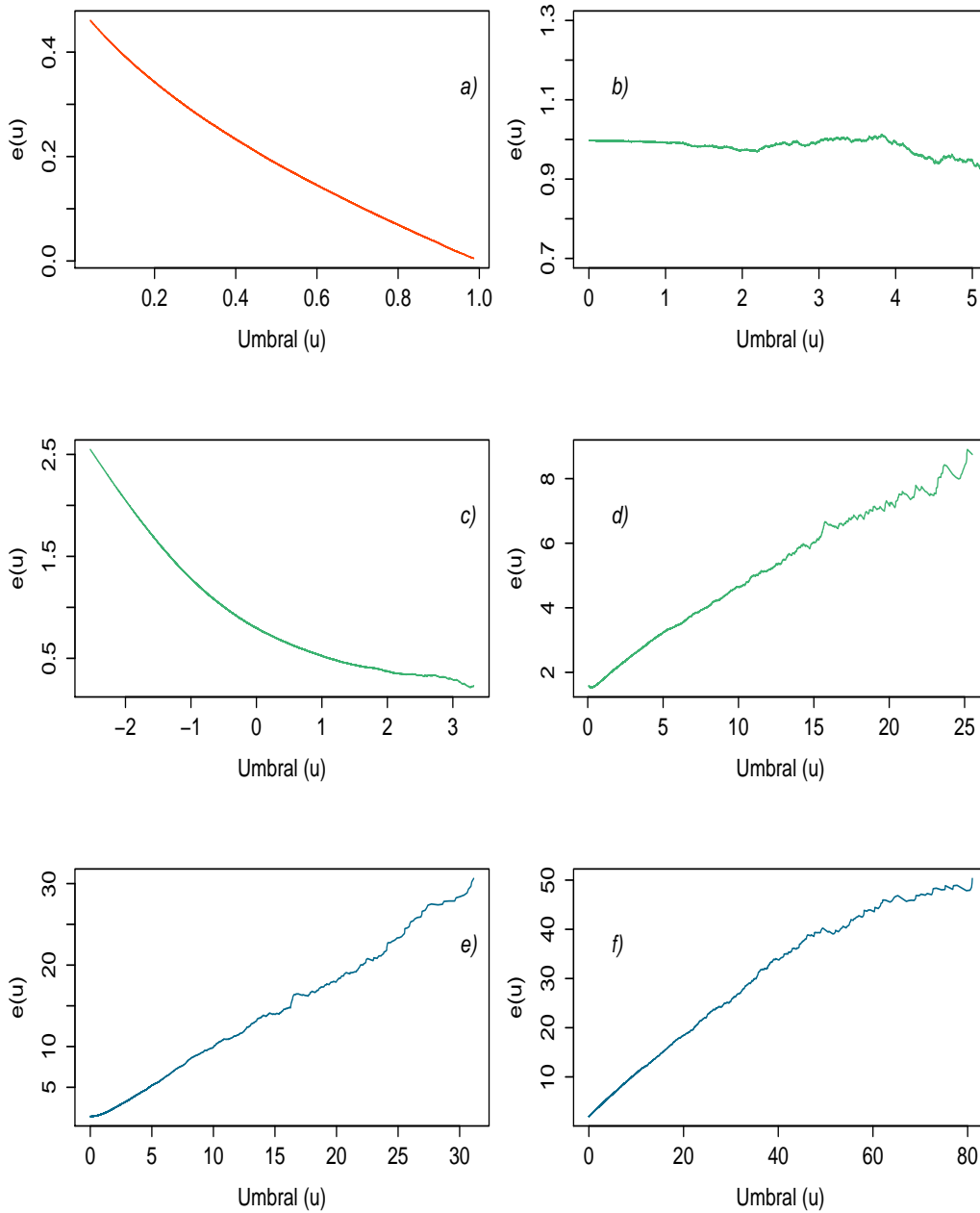


Figura 3.3: Ejemplos del comportamiento teórico de la función muestral media de exceso. Se muestra la gráfica para las distribuciones: *a)* $Beta(2, 2)$, *b)* $exp(1)$, *c)* $Normal(0, 1)$, *d)* $Lognormal(0, 1)$, *e)* $t - Student(2)$, *f)* $Pareto(2, 1)$. Las gráficas se construyeron a partir de datos simulados de cada distribución (cada muestra fue de tamaño $n = 100,000$ observaciones). En todos los casos se detuvo el proceso de graficación en $X_{50,100,000}$ para eliminar el exceso de variabilidad en la parte final de las gráficas.

3.2.3. Estimación de los parámetros de la Distribución Pareto Generalizada

En la relación (3.22) se hace necesaria la obtención de valores estimados de los parámetros de la DPG con la que se aproximará la función de distribución de exceso F_u por encima del umbral u . Uno de los métodos para obtener dichos valores estimados es utilizando el Estimador de Máxima Verosimilitud, el cual se presenta en seguida.

Partimos de datos $X = (X_1, \dots, X_n)$ i.i.d. con función de distribución F . Suponemos que F es una DPG con parámetros ξ y β , de manera que su densidad f es

$$f(x) = \frac{1}{\beta} \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}-1}, \quad x \in D(\xi, \beta),$$

donde $D(\xi, \beta)$ está dado en (3.13).

La función logarítmica de la verosimilitud es igual a

$$l((\xi, \beta); X) = -n \ln(\beta) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\beta} X_i\right).$$

Nótese que los argumentos de esta función deben satisfacer la restricción $X_i \in D(\xi, \beta)$. Ahora, las ecuaciones de verosimilitud pueden derivarse y resolverse numéricamente, dando como resultado los EMVs $\hat{\xi}_n$ y $\hat{\beta}_n$. El método funciona bien si $\xi > -1/2$ y, en este caso, puede probarse que

$$n^{1/2} \left(\hat{\xi}_n - \xi, \frac{\hat{\beta}_n}{\beta} - 1 \right) \xrightarrow{d} N(0, M^{-1}), \quad n \rightarrow \infty$$

donde

$$M^{-1} = (1 + \xi) \begin{pmatrix} 1 + \xi & -1 \\ -1 & 2 \end{pmatrix}$$

y $N(\mu, \Sigma)$ denota una distribución normal bivariada cuya media es el vector μ y con matriz de covarianzas Σ . Se dan las propiedades usuales del EMV, es decir, la eficiencia y consistencia.

Debido a la relación (3.21), resulta más realista suponer un modelo DPG para los excesos Y_1, \dots, Y_N , donde $N = N_u$ es independiente de las Y_i . Las

ecuaciones condicionales de verosimilitud que resultan se resuelven más fácilmente mediante la reparametrización $(\xi, \beta) \rightarrow (\xi, \tau)$, donde $\tau = -\xi/\beta$. Esto nos da la solución

$$\hat{\xi} = \hat{\xi}(\tau) = N^{-1} \sum_{i=1}^n \ln(1 - \tau Y_i),$$

donde τ satisface

$$h(\tau) = \frac{1}{\tau} + \frac{1}{N} \left(\frac{1}{\hat{\xi}(\tau)} + 1 \right) \sum_{i=1}^n \frac{Y_i}{1 - \tau Y_i} = 0.$$

La función $h(\tau)$, definida para $\tau \in (-\infty, \max\{Y_1, \dots, Y_N\})$ es continua en 0.

De esta forma, tenemos ya un método, llamado Picos Sobre el Umbral (POT), que nos permite calcular estimaciones de dos importantes medidas de riesgo (VaR_α y ES_α), utilizando una aproximación de la función de excesos F_u mediante una DPG.

Capítulo 4

Potenciando el GARCH vía la Teoría de Valores Extremos

En Capítulos anteriores estudiamos una serie de metodologías para medir el riesgo de mercado; considerando un enfoque tanto condicional (dinámico) como no condicional (estático). Esquemáticamente, los enfoques existentes para estimar la distribución de pérdida de un activo pueden ser divididos en tres grupos: el método no paramétrico de simulación histórica; los métodos completamente paramétricos basados en un modelo econométrico para la dinámica de la volatilidad y la hipótesis de normalidad condicional (e.g. *RiskMetrics* de J.P. Morgan y la mayoría de los modelos de las familias ARCH y GARCH); y, finalmente, los métodos basados en la Teoría de Valores Extremos. Estos enfoques tienen diversas ventajas y desventajas para un administrador de riesgo de mercado, las cuales resumimos a continuación.

En el enfoque de simulación histórica, como mencionamos antes, el estimador de la distribución de pérdida de un activo está dado simplemente por la distribución empírica de las pérdidas y ganancias pasadas del activo. Por lo tanto, el método es fácil de implementar y evita el uso de hipótesis adicionales sobre la forma de la distribución de pérdida. Sin embargo, el método sufre de serias desventajas. Los cuantiles extremos son notoriamente difíciles de estimar, pues la extrapolación más allá de las observaciones pasadas resulta imposible y las estimaciones de cuantiles extremos dentro de la muestra tienden a ser muy ineficientes (el estimador está sujeto a una varianza muy grande). Adicionalmente, si buscamos mitigar estos problemas considerando muestras grandes el método es incapaz de distinguir entre periodos de alta y baja volatilidad.

Por su lado, los modelos econométricos para la dinámica de la volatilidad que suponen normalidad condicional entregan estimaciones del *VaR* o *Expected Shortfall* que reflejan panorama actual de volatilidad. La principal debilidad de este enfoque es que la hipótesis de normalidad condicional suele no sostenerse para datos financieros reales. Por lo tanto, los modelos basados en normalidad condicional no son apropiados para estimar altos cuantiles de la distribución de pérdida.¹⁵

Finalmente, los modelos basados en la Teoría de Valores Extremos tienen dos características que los hacen atractivos para la estimación de colas: están basados en una sólida teoría estadística y ofrecen una forma paramétrica para la cola de una distribución. Por lo tanto, estos métodos permiten cierta extrapolación más allá del rango de los datos. Sin embargo, ninguno de los métodos con base en la TVE estudiados hasta el momento entrega estimaciones del *VaR* o *Expected Shortfall* que reflejen el panorama actual de volatilidad. Dada la heterocedasticidad condicional presente en la mayoría de los datos de origen financiero, la cual está bien documentada por el considerable éxito en la práctica de los modelos de las familias ARCH y GARCH, se considera que esta es la mayor desventaja de cualquier estimador del *VaR* o *Expected Shortfall* obtenido a través de metodologías derivadas de la TVE.

En consecuencia, y con el propósito de superar las desventajas de cada uno de los métodos mencionados, en este Capítulo exploramos cómo combinar ideas de los tres enfoques. El objetivo será lograr que las virtudes de cada técnica subsanen las desventajas identificadas en otras metodologías.

El material presentado en las siguientes secciones está tomado principalmente del trabajo de Alexander McNeil y Rüdiger Frey, “*Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach*”, [3]. En el cual se expone una metodología que muestra cómo lograr los objetivos anteriores de una forma eficiente y relativamente sencilla.

4.1. Metodología

En esta Sección presentaremos la metodología mencionada en la parte final de la introducción de este Capítulo. Comenzaremos por describir a grandes rasgos dicha metodología. Después, procederemos a introducir el lenguaje y notación necesarios para su desarrollo. Finalmente, enunciaremos algunas ventajas de dicha metodología sobre los enfoques que suponen distribuciones

¹⁵ Para consultar ejemplos sobre este hecho, véase [22].

condicionales Normales o t -Student y expondremos los pasos a seguir para su implementación.

En pocas palabras, la metodología discutida en este Capítulo consta de los pasos en seguida descritos: dado un conjunto de datos (típicamente diferencias logarítmicas de precios de activos, tasas de interés, tipos de cambio, etc., en donde asociamos un signo positivo a pérdidas y uno negativo a ganancias) utilizamos modelos tipo GARCH bajo un enfoque quasi-máximo verosímil para obtener estimaciones condicionales de la volatilidad.

En caso de que el ajuste logrado y las estimaciones obtenidas sean satisfactorios, pasamos a centrar nuestra atención en los residuales (estandarizados) del modelo GARCH; los cuales, deberían formar, al menos aproximadamente, una serie i.i.d. Más aún, pruebas estadísticas y diversos análisis exploratorios empíricos han confirmado dichos términos de error suelen exhibir colas pesadas.

De esta forma, una vez calculadas las versiones muestrales de sus residuos estandarizados utilizamos simulación histórica (para la parte central de la distribución) y métodos de umbrales de la TVE (para las colas) para estimar su distribución; nótese que la aplicación de estos métodos es facilitada por la (aproximada) independencia a lo largo del tiempo de los residuales.

De este modo, un estimador de la distribución condicional de los rendimientos puede ser fácilmente construido a partir de la distribución estimada de los residuales y las estimaciones de volatilidad condicional. Así, dicho estimador rescata dos hechos estilizados exhibidos por la mayoría de las series de rendimientos financieros: volatilidad estocástica y colas pesadas de las distribuciones condicionales de los rendimientos sobre periodos cortos de tiempo.

Procederemos entonces a formalizar las ideas arriba expresadas. Sea $\{X_t\}_{t=1}^{\infty}$ una serie de tiempo estacionaria que representa observaciones diarias de log-rendimientos del precio de un activo financiero. Supondremos que la dinámica de X_t está dada por

$$X_t = \sigma_t Z_t, \quad (4.1)$$

donde las innovaciones Z_t son un proceso Ruido Blanco Estricto (es decir, está conformado por variables aleatorias independientes e idénticamente distribuidas) con una función de distribución $F_Z(z)$ con media cero y varianza unitaria. Supondremos que σ_t es medible con respecto a \mathcal{F}_{t-1} , la información sobre el proceso de rendimientos disponible hasta el tiempo $t - 1$.

Adicionalmente, para $h > 0$, denotaremos por $F_{X_{t+1}, + \dots + X_{t+h} | \mathcal{F}_t}$ la función de distribución predictiva del rendimiento sobre los próximos h días, dado el

conocimiento sobre los rendimientos hasta t e incluyendo el día t .

Al aplicar la metodología descrita al inicio de esta Sección estimaremos medidas de riesgo (cuantiles y déficits esperados) en las colas tanto de la distribución predictiva como de la distribución de innovaciones. Notemos que la primera de ellas es una distribución de pérdida condicional mientras que la segunda es de naturaleza no condicional.

Para $0 \leq \alpha \leq 1$, un cuantil condicional de la distribución $F_{X_{t+1}, \dots, X_{t+h} | \mathcal{F}_{t+1}}$ será denotado por

$$VaR_\alpha^{t,h} = \inf\{x \in \mathbb{R} : F_{X_{t+1}, \dots, X_{t+h} | \mathcal{F}_{t+1}}(x) \leq 1\},$$

y un cuantil no condicional por

$$q_\alpha(Z) = VaR_\alpha(Z) = \inf\{z \in \mathbb{R} : F_Z(z) > \alpha\}.$$

Por otro lado, denotaremos un déficit esperado condicional por

$$ES_\alpha^{t,h} = E \left[\sum_{j=1}^h X_{t+j} \mid \sum_{j=1}^h X_{t+j} > VaR_\alpha^{t,h} \right],$$

mientras que el déficit esperado no condicional será denotado como

$$ES_\alpha(Z) = E[Z \mid Z > q_\alpha(Z)],$$

donde hemos aumentado la notación de Capítulos anteriores para hacer énfasis sobre la distribución a la cual se le está calculando el déficit esperado.

La meta final de la metodología expuesta en este Capítulo será estimar cuantiles y déficits esperados para la distribución predictiva 1 paso adelante, los cuales denotaremos respectivamente por VaR_α^t y ES_α^t . Puesto que, como vimos en el Capítulo 2,

$$\begin{aligned} F_{X_{t+1} | \mathcal{F}_t}(x) &= P(\sigma_{t+1} Z_{t+1} \leq x | \mathcal{F}_t) \\ &= F_Z\left(\frac{x}{\sigma_t}\right), \end{aligned}$$

estas medidas de riesgo se reducen a

$$VaR_\alpha^t = \sigma_{t+1} q_\alpha(Z), \tag{4.2}$$

$$ES_\alpha^t = \sigma_{t+1} ES_\alpha(Z), \tag{4.3}$$

donde $q_\alpha(Z)$ y $ES_\alpha(Z)$ denotan medidas de riesgo asociadas a la distribución de Z_t , las cuales, por hipótesis, no dependen de t .

Ahora bien, es común que al estimar tanto el VaR_α^t como el ES_α^t con modelos GARCH se suponga que la distribución de las innovaciones de dichos procesos es Normal estándar; de modo que los cuantiles y déficits esperados de las innovaciones están simplemente dados por $q_\alpha(Z) = \Phi^{-1}(\alpha)$ y $ES_\alpha = \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}$, donde Φ es la función de distribución de una v.a. Normal estándar.

Como vimos en el Capítulo 3, un modelo GARCH con innovaciones Gaussianas puede ser ajustado por máxima verosimilitud y σ_{t+1} puede ser estimada utilizando pronósticos de X_t un paso adelante. De esta manera, tanto VaR_α^t como ES_α^t son fácilmente construidos utilizando (4.2) y (4.3). Esto es cercano en espíritu al enfoque recomendado en *RiskMetrics* (véase [23]), pero tanto nuestros descubrimientos empíricos (véase Capítulo 5) como los de otros autores (véase [3]) concluyen que este enfoque a menudo subestima cuantiles y déficits esperados condicionales para $\alpha > 0.95$; la distribución de las innovaciones parece ser generalmente de cola más pesada, o más leptocúrtica, que la Normal.

Otro enfoque ampliamente recurrido al utilizar modelos GARCH para estimar medidas de riesgo condicionales consiste en suponer que las innovaciones tienen una distribución leptocúrtica, tal como una distribución t -Student (escalada para tener varianza 1). Bajo este enfoque, $q_\alpha(Z)$ y $ES_\alpha(Z)$ pueden ser calculados siguiendo la metodología del Ejemplo 1.10, por lo que tanto el VaR_α^t como el ES_α^t pueden nuevamente ser estimados de forma sencilla siguiendo (4.2) y (4.3).

Modelos GARCH con innovaciones t -Student pueden también ser ajustados por máxima verosimilitud y el parámetro adicional ν puede ser estimado. Más adelante, veremos que este enfoque puede ser visto como un caso especial de la metodología estudiada en este Capítulo; sin embargo, este caso particular rinde resultados satisfactorios siempre y cuando las colas positiva y negativa de la distribución de las innovaciones sean (aproximadamente) iguales.

En contraste con los enfoques previamente descritos, la metodología discutida en este Capítulo hace suposiciones mínimas sobre la distribución de las innovaciones y se concentra en el modelaje de sus colas utilizando la Teoría de Valores Extremos. Para ello, utilizamos el enfoque quasi-máximo verosímil de ajuste de modelos GARCH estudiado en el Capítulo 2. De esta manera, los pasos a seguir en la metodología aquí estudiada serían:

1. Elegimos y ajustamos el mejor modelo de la clase (4.1) para nuestros datos de rendimientos siguiendo un enfoque quasi-máximo verosímil (QMV); sin tomar ninguna hipótesis sobre $F_Z(z)$.

2. Calculamos los residuos (estandarizados) del modelo elegido y evaluamos su calidad de ajuste. En caso de que ésta sea satisfactoria, pasamos al paso 3 de la metodología.
3. Estimamos σ_{t+1} con base en el modelo ajustado.
4. Consideramos los residuos calculados en el paso 2 como una realización de un proceso Ruido Blanco Estricto y modelamos la cola de $F_Z(z)$ vía la Teoría de Valores Extremos; específicamente, aplicamos el Método de Picos sobre el umbral estudiado en el Capítulo 3 a las innovaciones del modelo GARCH elegido. Después, utilizamos el modelo Pareto Generalizado ajustado para estimar $q_\alpha(Z)$ y $ES_\alpha(Z)$.
5. Finalmente, con base en (4.2) y (4.3), utilizamos el pronóstico de volatilidad y las medidas de riesgo calculados en los pasos 3 y 4 respectivamente para estimar Var_α^t y ES_α^t .

En cuanto al paso 1 de la metodología anterior, en la literatura econométrica se han propuesto muchos modelos distintos para la dinámica de la volatilidad, incluyendo modelos de las familias ARCH/GARCH (en [9]), procesos HARCH (en [24]) y otras clases de modelos para volatilidad estocástica (en [25]). En este trabajo, modelaremos la dinámica de la volatilidad por medio de un proceso GARCH y, en particular, ilustraremos el desarrollo de la metodología anterior con un parsimonioso modelo GARCH(1, 1). No obstante, el enfoque puede extenderse de forma sencilla a modelos GARCH de órdenes más altos.

Con lo anterior en mente, en las siguientes secciones profundizaremos más en los detalles prácticos de la metodología descrita previamente. Primeramente, discutiremos cómo estimar los parámetros y medidas de riesgo involucrados en la aplicación de la metodología; a partir de datos (series de rendimientos) de origen financiero. Después, estudiaremos cómo evaluar el desempeño histórico de dicha técnica; es decir discutiremos cómo implementar un *backtest* de las estimaciones de medidas riesgo condicionales (Var_α^t y ES_α^t) obtenidas mediante su aplicación.

4.2. El modelo híbrido GARCH-DPG: estimación de parámetros y medidas de riesgo

Para fines de estimación y predicción fijamos una memoria constante de n observaciones. De esta forma, al final de cada día t nuestra información consistirá de los últimos n log-rendimientos $(x_{t-n+1}, \dots, x_{t-1}, x_t)$. Consideraremos que estas observaciones son una realización de un proceso GARCH(1, 1). Por lo tanto, la varianza condicional de la serie está dada por:

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_t^2 + \beta_1 \sigma_t^2, \quad (4.4)$$

donde $\alpha_0, \alpha_1, \beta_1 > 0$.¹⁶

El modelo es ajustado siguiendo un enfoque quasi-máximo verosímil; es decir, maximizamos la verosimilitud de un modelo GARCH(1, 1) con innovaciones Gaussianas para obtener valores estimados de los parámetros $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$.

En el Capítulo 2 discutimos que a pesar de que esta metodología consiste de ajustar un modelo siguiendo una hipótesis distribucional en la que no necesariamente creemos, la metodología QMV entrega estimaciones razonables de parámetros siempre que se cumplan ciertas condiciones técnicas.¹⁷

Así, una vez contamos con valores los valores estimados $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$, podemos calcular estimaciones condicionales de la desviación estándar de la serie $(\hat{\sigma}_{t-n+1}, \dots, \hat{\sigma}_t)$ a partir de las ecuaciones dadas en (4.4). Recordemos que para esto es necesario sustituir ciertos valores iniciales sensibles en dichas ecuaciones, como se discutió en la Sección 2.2.3.

Podemos entonces pasar a calcular los residuales del modelo. Dichos residuales pueden ser utilizados para evaluar qué tan adecuado resultó el modelo para nuestros datos y como información (observaciones) para la etapa número 4 de la metodología estudiada. Recordemos que estos residuales son calculados como

$$(z_{t-n+1}, \dots, z_t) = \left(\frac{x_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \dots, \frac{x_t}{\hat{\sigma}_t} \right)$$

¹⁶ Aunque en esta Sección utilizamos un modelo GARCH(1,1) para modelar la dinámica de la serie de tiempo estudiada, existen diversos criterios estadísticos que nos pueden dar una guía para elegir el mejor modelo GARCH para un conjunto de datos (e.g. el criterio de Akaike descrito en el Apéndice B). Esta clase de criterios son de gran utilidad en el paso 1 de la metodología de este Capítulo.

¹⁷ Véase la Sección 2.2.3 de este trabajo. Información adicional sobre estimación de parámetros QMV puede también ser encontrada en el Capítulo 4 de [26].

y que, en caso de que el modelo GARCH haya conseguido modelar adecuadamente la dinámica de la volatilidad, éstos deberían comportarse como una muestra de observaciones i.i.d.

La hipótesis i.i.d. hecha sobre los residuales estandarizados del GARCH puede evaluarse como se describió en la Sección 2.2.3; es decir, a través de correlogramas, correlogramas parciales y pruebas Portmanteau como la de Ljung-Box, las cuales pueden aplicarse a los datos crudos y un par de transformaciones de ellos. Cabe mencionar que mientras que para los datos crudos esperamos rechazar rápidamente la hipótesis i.i.d., ésta debería resultar más plausible para los residuales estandarizados del modelo GARCH.

Si estamos satisfechos con la calidad de ajuste del modelo GARCH, podemos terminar la etapa número 2 de la metodología y proceder a calcular pronósticos de la volatilidad de nuestra serie para el día $t + 1$:

$$\hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_t^2 + \beta_1 \hat{\sigma}_t^2.$$

No obstante, recordemos que la fórmula anterior, estudiada en Capítulo 2, fue desarrollada con base en un proceso GARCH(1,1) estacionario en covarianza. En consecuencia, para poder utilizarla es necesario que el proceso GARCH ajustado cumpla dicha hipótesis. Una forma de verificar este hecho, según la Proposición 2.19, es comprobar que los valores estimados de los parámetros del GARCH satisfagan la condición $\alpha_1 + \beta_1 < 1$. Una vez realizado este pronóstico de volatilidad, podemos dar por concluida la etapa 3 de la metodología.

Continuamos entonces con la fase número 4: modelar las innovaciones del GARCH. Para esto, podríamos construir primeramente un gráfico QQ, a fin de evaluar de forma gráfica la plausibilidad de que un modelo Normal ajuste apropiadamente a las innovaciones. No obstante, en la mayoría de los casos esperaríamos encontrar fuerte evidencia en contra de esta hipótesis distribucional; para nuestros propósitos, el ajuste en el área de las colas será de particular interés.¹⁸

En caso de que la hipótesis de normalidad sea rechazada, procederemos a utilizar las técnicas aprendidas en el Capítulo 3 sobre modelos de la Teoría de Valores Extremos para modelar la cola de la distribución de dichas innovaciones; en particular, utilizaremos los modelos para excesos sobre el umbral estudiados en la segunda parte dicho Capítulo.

¹⁸ Adicionalmente, podrían aplicarse pruebas de hipótesis como la de Anderson-Darling, Shapiro-Wilk o Jarque-Bera, entre otras, a los residuos estandarizados para evaluar la hipótesis de normalidad en las innovaciones.

De este modo, fijamos un umbral alto u y suponemos que los excesos de los residuales sobre este umbral siguen una Distribución Pareto Generalizada con función de distribución:

$$G_{\xi}(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - e^{-x} & \text{si } \xi = 0 \end{cases},$$

donde

$$\begin{aligned} x &\geq 0 && \text{si } \xi \geq 0 \\ 0 \leq x &\leq -1/\xi && \text{si } \xi < 0. \end{aligned}$$

Sabemos, gracias al material estudiado en el Capítulo 3, que la elección de esta distribución para modelar los excesos de los residuales sobre u está teóricamente justificada por el Teorema de Gnedenko-Balkema-Pickands-de Haan (ver Teorema 3.27).

Adicionalmente, sabemos que al modelar los excesos de los residuales sobre u mediante una DPG tenemos una gran flexibilidad para tratar con lo pesado de las colas de la distribución de las innovaciones, y que el valor estimado del parámetro de forma ξ de la DPG nos dará información sobre lo pesado de las mismas.

En pocas palabras: el caso $\xi > 0$ surgirá si las innovaciones tienen colas pesadas que decaen como funciones de potencias. El caso $\xi = 0$ hará sentido cuando las colas de la distribución de las innovaciones decaigan siguiendo, aproximadamente, una ley de exponencial; no obstante habrá gran diversidad en lo pesado de las colas en estos casos (recordemos que las distribuciones Normal y log-normal caen ambas en este escenario). El último caso, en el que $\xi < 0$, se podría dar si las observaciones de las innovaciones dieran pie a la hipótesis de provenir de una distribución de cola corta (con extremo derecho x_F finito), aunque este caso es el menos probable, dada la naturaleza de los datos analizados.

Supondremos entonces que la cola de la distribución que subyace a las innovaciones del GARCH comienza en el umbral u . De nuestra muestra de n puntos, un número aleatorio $N = N_u$ excederá este umbral. Si suponemos que estos N excesos son una muestra de datos i.i.d. que siguen una Distribución Pareto Generalizada podemos proceder a estimar los parámetros de dicha distribución utilizando la metodología estudiada en la Sección 3.2.3; esto nos permitiría obtener los valores estimados $\hat{\xi}$ y $\hat{\beta}$.

Para elegir un valor inicial de u podrían aplicarse a los residuales técnicas gráficas como la descrita en la parte final de la Sección 3.2.2 del Capítulo 4. Sin embargo, recordemos que siempre será conveniente realizar análisis

para distintos valores de u , de forma que sea posible encontrar aquel valor de u que brinde la mayor calidad de ajuste de la DPG a la cola de nuestras innovaciones; esto se explora con mayor profundidad en el Capítulo 5, con datos a la mano que permitan ilustrar de mejor forma estas ideas.

Consideremos ahora la siguiente igualdad, estudiada ya en el Capítulo 3, para puntos $x > u$ en la cola de F

$$1 - F(x) = (1 - F(u))(1 - F_u(x - u)) \quad (4.5)$$

Si estimamos primer término del lado derecho de la ecuación (4.5) utilizando la proporción aleatoria N/n (es decir, vía simulación histórica), y si estimamos el segundo término aproximando la distribución de exceso F_u vía una DPG ajustada por máxima verosimilitud (es decir, vía la TVE), obtenemos el estimador

$$\widehat{F}(x) = 1 - \frac{N}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}},$$

para $x > 0$.¹⁹

No obstante, para llevar la metodología de este Capítulo a la práctica modificamos ligeramente el procedimiento sugerido por la fórmula anterior y fijamos el número de datos en la cola para ser $N = k$, donde $k \ll n$. Esto nos da un umbral aleatorio en la $(k + 1)$ -ésima estadística de orden.

De esta manera, sean $z_{1,n} \geq z_{2,n} \cdots > z_{n,n}$ los residuos estandarizados ordenados, entonces la DPG con parámetros ξ y β es ajustada a los datos $(z_{1,n} - z_{k+1,n}, \dots, z_{k,n} - z_{k+1,n})$, las cantidades de exceso sobre el umbral de todos los residuales que excedieron el umbral. El estimador de la cola para $F_Z(z)$ toma entonces la forma

$$\widehat{F}_Z(z) = 1 - \frac{k}{n} \left(1 + \hat{\xi} \frac{z - z_{k+1,n}}{\hat{\beta}} \right)^{-1/\hat{\xi}}. \quad (4.6)$$

Para $q > 1 - k/n$ podemos invertir esta fórmula para obtener

$$\hat{q}_\alpha(Z) = \hat{q}_{\alpha,k}(Z) = z_{k+1,n} + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{1 - \alpha}{k/n} \right)^{-1/\hat{\xi}} - 1 \right); \quad (4.7)$$

donde utilizamos la notación $\hat{q}_{\alpha,k}(Z)$ para enfatizar la dependencia del estimador de la elección de k .

¹⁹ Las propiedades de este estimador pueden ser consultadas en el Capítulo 3 de este trabajo. Una referencia adicional es [27].

Ahora bien, recordemos que como medida de riesgo el VaR tiene ciertas deficiencias teóricas. Específicamente, en el Capítulo 1 mencionamos que el VaR es comúnmente criticado como medida de riesgo por no cumplir con todas las características de una *medida de riesgo coherente* (véase la Sección 1.2.3); en particular, el VaR cuenta con la desventaja de no ser una medida de riesgo subaditiva, lo que vuelve imposible basar todo un sistema de riesgos exclusivamente en el VaR . Adicionalmente, el VaR solo brinda una cota superior para las pérdidas que ocurren con una cierta frecuencia; dicha medida de riesgo no nos dice nada sobre el tamaño potencial de una pérdida dado que ésta ha excedido dicha cota.

No obstante, el *Expected Shortfall* es una medida de riesgo que supera dichas deficiencias: el *Expected Shortfall* es una medida de riesgo coherente que nos brinda información sobre el tamaño de una pérdida, dado que ésta ha excedido el VaR . En consecuencia, a continuación estudiaremos cómo estimar dicha medida de riesgo a partir de la metodología estudiada en este Capítulo.

Así, recordemos de la ecuación (4.3) que el *Expected Shortfall* condicional (a un día) está dado por

$$ES_{\alpha}^t = \sigma_{t+1} E[Z | Z > q_{\alpha}(Z)].$$

Por lo tanto, para estimar esta medida de riesgo requerimos de una estimación del *Expected Shortfall* para la distribución de las innovaciones $E[Z | Z > q_{\alpha}(Z)]$. Para una variable aleatoria W con distribución DPG con parámetros $\xi < 1$ y β estudiamos en la Sección 3.2.1 que se satisface

$$E[W | W > w] = \frac{w + \beta}{1 - \xi}, \quad (4.8)$$

donde $\beta + w\xi > 0$.

Supongamos entonces nuevamente que los excesos de las innovaciones sobre el umbral u tienen exactamente esta distribución, es decir, $(Z - u | Z > u) \sim G_{\xi, \beta}$. Notemos que para $q_{\alpha}(Z) > u$ podemos escribir:

$$Z - q_{\alpha}(Z) | Z > q_{\alpha}(Z) = (Z - u) - (q_{\alpha}(Z) - u) | (Z - u) > (q_{\alpha}(Z) - u),$$

de modo que puede ser demostrado (véase [19]) que

$$(Z - q_{\alpha}(Z) | Z > q_{\alpha}(Z)) \sim G_{\xi, \beta + \xi(q_{\alpha}(Z) - u)}, \quad (4.9)$$

por lo que los excesos sobre el umbral $q_{\alpha}(Z)$ también tienen distribución DPG con el mismo parámetro de forma ξ y distinto parámetro de escala. Podemos

entonces utilizar la ecuación (4.8) para obtener

$$E[Z | Z > q_\alpha(Z)] = q_\alpha(Z) \left(\frac{1}{1 - \xi} + \frac{\beta + \xi u}{(1 - \xi)q_\alpha(Z)} \right). \quad (4.10)$$

El *Expected Shortfall* condicional en (4.10) es estimado de forma análoga que el estimador del *VaR* en (4.7): reemplazando los parámetros desconocidos por sus valores estimados vía máxima verosimilitud y reemplazando u por $z_{k+1,n}$ y $q_\alpha(Z)$ por el estimador (4.7). Esto nos da el estimador del *Expected Shortfall* condicional:

$$\widehat{ES}_\alpha^t = \hat{\sigma}_{t+1} \hat{q}_\alpha(Z) \left(\frac{1}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} z_{k+1,n}}{1 - \hat{\xi} \hat{q}_\alpha(Z)} \right). \quad (4.11)$$

Ahora bien, mencionamos antes que el enfoque que consiste de modelar la distribución de las innovaciones por medio de una distribución t -Student puede ser visto como un caso particular de la metodología estudiada en este Capítulo. En efecto, ello se debe a que la distribución t -Student es un ejemplo de una distribución con colas pesadas, es decir, una distribución cuya distribución de excesos límite es DPG con $\xi > 0$.

En el Capítulo 3, en el Teorema 3.19, estudiamos que Gnedenko caracterizó todas estas distribuciones, bajo la particularidad de tener colas de la forma

$$1 - F(x) = x^{-1/\xi} L(x), \quad (4.12)$$

donde $L(x)$ es una función de variación lenta en ∞ y ξ es el parámetro positivo de la distribución DPG límite. De esta forma, para la distribución t -Student con ν grados de libertad puede demostrarse que su cola satisface

$$1 - F(x) \sim \frac{\nu^{(\nu-2)/2}}{\text{Beta}(1/2, \nu/2)} x^{-\nu}, \quad (4.13)$$

de manera que la distribución t -Student provee un ejemplo muy simple de una distribución simétrica que pertenece a la clase definida por Gnedenko a través de (4.12); el Dominio de Atracción del Máximo de la distribución Fréchet. Adicionalmente (véase [28]), se puede probar que el valor de ξ corresponde al recíproco de los grados de libertad de la distribución t -Student original.

No obstante, a pesar de que el modelo GARCH- t -Student puede ser visto como un caso particular de la metodología introducida en este Capítulo, debe notarse que esta última propuesta tiene la ventaja de ofrecer un modelo

explícito para cada tipo de cola que pueda presentar la distribución de las innovaciones (pesada, ligera o corta).

Adicionalmente, al estimar dos parámetros, ξ y β , para modelar únicamente el área de la cola de la distribución (recordemos que la parte central se modela vía simulación histórica), se posibilita obtener un mejor ajuste en general de toda la distribución de las innovaciones y, en consecuencia, puede esperarse que el modelo GARCH-DPG ajuste de mejor forma a los datos originales que modelos GARCH-Normal o GARCH- t -Student, entre otros similares. Más adelante, en el Capítulo 5, evaluaremos esta expectativa en datos financieros de origen mexicano.

4.3. *Backtesting*

Procederemos ahora a analizar cómo realizar una evaluación histórica de las estimaciones del VaR_α^t y ES_α^t entregadas por la metodología descrita en este Capítulo.

Evaluaremos el desempeño del VaR_α^t y del ES_α^t sobre una serie histórica x_1, \dots, x_m , donde $m \gg n$; con base en una memoria de n días, en los días $t \in T = \{n, \dots, m-1\}$. Esto quiere decir que si trabajáramos, por ejemplo, con una ventana temporal de $n = 1,000$ días bancarios, para cada predicción del VaR_α^t y ES_α^t utilizaríamos aproximadamente 4 años de información diaria.

Ahora bien, este punto vale la pena notar que en un *backtest* largo existe un parámetro de la DPG cuya calibración se dificulta. Específicamente, la determinación cuidadosa del umbral, u , para aplicar el método POT a las innovaciones del GARCH. En efecto, en una rutina automatizada resulta imposible detenerse a examinar en cada ocasión herramientas como la gráfica de la función muestral media de exceso para elegir un valor apropiado para k y, en consecuencia, para u . Por lo tanto, para evaluar la metodología de este Capítulo debe desarrollarse una técnica que permita elegir un valor para u de forma automatizada, o bien alguna clase de evidencia empírica que nos permita fijar este parámetro en un valor apropiado para dicho propósito.

Una opción para llevar a cabo una elección de u como la antes descrita consiste en realizar alguna clase de *estudio de simulación* que nos permita elegir un valor de u que funcione a lo largo de todo nuestro *backtest*. Sin embargo, por el momento no exploraremos más esta posibilidad, en su lugar, referimos al lector al Capítulo 5 de este trabajo, donde se ilustra una manera de resolver este problema.

Para llevar a cabo nuestro *backtest* del VaR_α^t , en cada día $t \in T$ ajustamos un nuevo modelo GARCH(1,1) a las observaciones de pérdidas correspondientes, calculamos los residuos estandarizados del modelo y determinamos un nuevo estimador de la distribución de las innovaciones vía el método POT. De esta manera, utilizamos la fórmula dada en (4.2) y con base en dicho modelo ajustado a los rendimientos estimamos el VaR_α^t . Después, comparamos \widehat{VaR}_α^t con x_{t+1} para los valores de α que se consideren apropiados; en nuestro caso, estaremos interesados en $\alpha \geq 0.95$. Diremos que una violación ocurrió siempre que $x_{t+1} > \widehat{VaR}_\alpha^t$.

Es posible desarrollar un *test* binomial con base en el número de violaciones para evaluar el desempeño de nuestras estimaciones del VaR_α^t . Tomando como hipótesis la dinámica de la ecuación (4.1), la indicadora de violación al tiempo $t \in T$ es una variable aleatoria Bernoulli

$$I_t = I_{\{X_{t+1} > VaR_\alpha^t\}} = I_{\{Z_{t+1} > q_\alpha(Z)\}} \sim \text{Bernoulli}(1 - \alpha).$$

Más aún, I_t e I_s son independientes para $s, t \in T$ y $s \neq t$, puesto que Z_t y Z_s son independientes. Por lo tanto,

$$\sum_{t \in T} I_t \sim \text{Binomial}(\text{card}(T), 1 - \alpha),$$

es decir, el número total de violaciones tiene una distribución binomial bajo el modelo propuesto.

De esta forma, bajo la hipótesis nula de que el modelo estima correctamente los cuantiles condicionales de la serie estudiada, la versión empírica de esta estadística $\sum_{t \in T} I_{\{x_{t+1} > \widehat{VaR}_\alpha^t\}}$ proviene de la distribución $\text{Binomial}(\text{card}(T), 1 - \alpha)$.

Por lo tanto, podemos realizar una prueba binomial de dos colas para contrastar dicha hipótesis nula contra la alternativa de que el método acarree un error sistemático en la estimación del VaR_α^t que conlleve a tener muy pocas o demasiadas violaciones; así, un p-valor menor al nivel de significancia con el que se trabaje (típicamente, 0.05) en la prueba binomial será interpretado como evidencia en contra de la hipótesis nula.

Ahora bien, también es posible desarrollar un *backtest* para evaluar nuestras estimaciones del ES_α^t . Dicho *backtest* es similar a aquel desarrollado para violaciones de cuantiles en el caso del VaR_α^t y nos permitirá investigar si el modelo propuesto en la metodología desarrollada en este Capítulo entrega estimaciones razonables de ES_α^t .

Esta vez estamos interesados en el tamaño de la diferencia entre X_{t+1} y ES_α^t dado el evento de una violación al cuantil VaR_α^t . Definimos los residuales

$$R_{t+1} = \frac{X_{t+1} - ES_\alpha^t}{\sigma_{t+1}} = Z_{t+1} - E[Z | Z > q_\alpha(Z)].$$

Es claro que bajo nuestro modelo (4.1) estos residuales son i.i.d. y que, condicionados al evento $\{X_{t+1} > VaR_\alpha^t\}$ o equivalentemente $\{Z_{t+1} > q_\alpha(Z)\}$, tienen un valor esperado igual a cero.

Así, supongamos nuevamente que realizamos nuestro *backtest* sobre los días en el conjunto T . Podemos formar versiones empíricas de estos residuales en los días en que ocurre una violación a los cuantiles correspondientes, es decir, en los días en que $x_{t+1} > \widehat{VaR}_\alpha^t$. Estos residuales serán llamados residuales de exceso y los denotaremos por

$$\left\{ r_{t+1} : t \in T, x_{t+1} > \widehat{VaR}_\alpha^t \right\}, \text{ donde } r_{t+1} = \frac{x_{t+1} - \widehat{ES}_\alpha^t}{\hat{\sigma}_{t+1}},$$

donde \widehat{ES}_α^t es una estimación condicional del déficit esperado.

De esta manera, bajo la hipótesis nula de que estimamos de forma correcta la dinámica del proceso de pérdidas (σ_{t+1}) y el primer momento de la distribución truncada de las innovaciones ($E[Z | Z > q_\alpha(Z)]$), los residuales de excesos deberían comportarse como una muestra de observaciones i.i.d. con media cero.

Para probar la hipótesis de media cero podemos utilizar una prueba *bootstrap* que no haga ninguna suposición sobre la distribución que subyace a los residuales de exceso.²⁰ Específicamente, puede realizarse una prueba de dos colas que contraste nuestra hipótesis nula de que los residuales de exceso tengan una media estadísticamente igual a cero contra la alternativa de que dichos residuales tengan una media distinta de cero; lo que significaría que el *Expected Shortfall* condicional fue sistemáticamente subestimado o sobrestimado.

Con la idea anterior hemos terminado de documentar el procedimiento sugerido para evaluar la metodología presentada en este Capítulo. Por lo tanto, a fin de ejemplificar e ilustrar el uso dicha metodología en la práctica, en nuestro siguiente y último Capítulo echaremos mano de todo lo estudiado hasta el momento y realizaremos un análisis completo de un par de series financieras de origen mexicano. Nuestro objetivo final será contrastar el desempeño

²⁰ Los detalles técnicos de este tipo de pruebas, así como el desarrollo de aquella utilizada en el Capítulo 5 de este trabajo, pueden ser consultados en el Apéndice B.

de las medidas de riesgo condicionales obtenidas siguiendo la metodología de este Capítulo contra aquel de las estimaciones que se obtendrían al seguir el enfoque de modelar las innovaciones de un GARCH ajustado a datos financieros mediante una distribución Normal o *t*-Student.

Capítulo 5

Aplicación en series financieras de origen mexicano: FIX e IPC

En este último Capítulo se llevará a la práctica la metodología de cuantificación de riesgo de mercado estudiada en el Capítulo 4. Dicha aplicación se realizará sobre un par de series de tiempo financieras de origen mexicano: las series de log-rendimientos históricos diarios del Tipo de Cambio FIX y del Índice de Precios y Cotizaciones (IPC).

Como se mencionó al inicio de este trabajo, las aplicaciones presentadas en este Capítulo pueden considerarse una continuación de los análisis realizados en [2] y [1]. En efecto, en [1] se investigó la capacidad de los modelos GARCH para medir el riesgo de mercado de series financieras mexicanas. No obstante, dicho trabajo se limitó a modelar las innovaciones del GARCH con distribuciones Normal y t -Student y estimar el $VaR_{0.95}^t$. Por otro lado, en [2] se exploró cómo medir el riesgo de mercado en series financieras a través de la TVE; específicamente, utilizando los métodos de máximos de bloque y POT y estimaciones no condicionales tanto del VaR como del *Expected Shortfall*. Sin embargo, en [2] las metodologías anteriores se aplicaron de forma directa a las series estudiadas, lo cual llevó a la necesidad de analizar cómo adaptarlas al caso de series de tiempo estacionarias (pues ambos métodos fueron desarrollados originalmente para trabajar con muestras i.i.d.).

En cambio, el enfoque que tomaremos utiliza un GARCH para modelar la volatilidad de las series estudiadas y aplica el método POT a las innovaciones de dicho proceso. De este modo, reflejamos el panorama de volatilidad presente en el mercado al momento de estimar el riesgo de cada serie, lo cual contrasta con el enfoque no condicional de [2]. Más aún, provisto que el

GARCH estime de forma adecuada dicha volatilidad, nuestro enfoque aplica el método POT a una serie i.i.d., por lo que no hay necesidad de adaptarlo como en [2]. Por otro lado, al introducir una modelación especializada en colas al análisis de innovaciones, tenemos la posibilidad de estimar medidas de riesgo extremo condicionales (i.e. a niveles de confianza considerablemente superiores a 0.95), lo cual contrasta con los resultados de [1].

Ahora bien, dado que la metodología aplicada en este Capítulo constituye un enfoque condicional de administración de riesgo de mercado, centraremos nuestros esfuerzos en obtener estimaciones del Var_{α}^t y ES_{α}^t . Nuestros objetivos finales serán dos: i) evaluar el desempeño de las estimaciones del Var_{α}^t y ES_{α}^t entregadas por la metodología del Capítulo 4 y, ii) compararlo contra aquel de las estimaciones de dichas medidas de riesgo obtenidas bajo el enfoque que considera modelar las innovaciones de un modelo GARCH (o ARCH) mediante distribuciones Gaussianas o t -Student.

En consecuencia, para cada caso de estudio presentado en este Capítulo se desarrollarán las siguientes actividades:

Primeramente, se realizará un análisis inicial de la serie de tiempo considerando solo una parte de los datos disponibles.²¹ Dicho análisis comenzará con un estudio descriptivo de la serie, el cual estará basado en gráficos tales como el histograma, el diagrama de caja y brazos, el QQ-Plot y la gráfica. Estos gráficos nos permitirán tener un primer acercamiento a nuestros datos, pues nos proporcionarán información estadística general sobre ellos; dado nuestro objetivo, el comportamiento de éstos en el área de las colas de su distribución será de particular interés, pues es justo allí donde realizaremos las estimaciones de nuestras medidas de riesgo.

Después, en el segundo paso de nuestro análisis inicial procederemos a examinar nuestros datos bajo un enfoque dinámico de series de tiempo. Esto es, evaluaremos la plausibilidad de ajustar un modelo de las familias ARCH y GARCH a nuestros datos.²² Luego, en caso de que el ajuste de dicha clase de modelos resulte razonable, procederemos a seleccionar los modelos que parezcan ser más apropiados para nuestra serie (dadas sus características de dependencia serial).

²¹ Realizar un análisis inicial considerando solo una porción de la información disponible tiene sentido dada nuestra hipótesis de que los procesos que subyacen a las series financieras a estudiar son estacionarios; hipótesis que se discute con detalle en el Capítulo 2. En nuestro caso, dichos análisis iniciales se realizarán con base en, aproximadamente, los últimos 4 años de información disponible de cada serie de tiempo.

²² A pesar de que en [1] se analizó este hecho, en el presente trabajo se vuelve a investigar esta posibilidad, debido a que las series contemplan datos más actuales.

Una vez contemos con diversos candidatos, elegiremos el mejor modelo para la serie (vía ciertos criterios estadísticos especializados para la elección de modelos²³) y evaluaremos su calidad de ajuste (a través de las herramientas descritas en el Capítulo 2, vía residuales); en caso de ser necesario, realizaremos este procedimiento repetidamente, hasta obtener un modelo final para la dinámica de la serie estudiada. Cuando logremos elegir un modelo final para la serie, pronosticaremos su volatilidad a partir de dicho modelo.

Posteriormente, centraremos nuestra atención en los residuos estandarizados del modelo elegido; es decir, en las innovaciones del modelo ARCH o GARCH ajustado. Específicamente, modelaremos cuidadosamente la distribución de dichas innovaciones; particularmente el área de las colas. El objetivo de dicho análisis será obtener estimaciones del *VaR* y *Expected Shortfall* para la distribución de las innovaciones. Haremos esto siguiendo distintos enfoques:

1. Aplicando el Método POT (estudiado en el Capítulo 3) a las innovaciones de cada serie; a fin de ajustar una Distribución Pareto Generalizada a la cola (derecha) de su distribución.
2. Ajustando (vía máxima verosimilitud) una distribución Normal y *t*-Student a dichas innovaciones.

Una vez logremos calibrar los tres modelos anteriores, procederemos a estimar nuestras medidas de riesgo condicionales (VaR_α^t y ES_α^t), siguiendo las fórmulas derivadas en los Capítulos 2 y 3:

$$VaR_\alpha^t = \sigma_{t+1}q_\alpha(Z), \quad ES_\alpha^t = \sigma_{t+1}ES_\alpha(Z). \quad (5.1)$$

Para utilizar las fórmulas dadas en (5.1) utilizaremos el pronóstico de volatilidad obtenido en la parte inicial de nuestro análisis y las medidas de riesgo para las innovaciones calculadas bajo los enfoques 1 y 2 anteriores.

Por lo tanto, el procedimiento anterior nos brindará tres estimaciones distintas del VaR_α^t y ES_α^t , una por cada modelo para las innovaciones: GARCH-DPG, GARCH-Normal y GARCH-*t*-Student.

Finalmente, una vez ilustrado el camino para obtener las tres estimaciones condicionales de medidas de riesgo anteriores, pasaremos a evaluar su desempeño histórico, vía *backtests*. En estas evaluaciones, consideraremos toda la información disponible de la serie de tiempo; es decir, la totalidad de la base de datos a nuestra disposición y no solo una parte de ésta.

²³ Véase Apéndice B.

Cabe mencionar que en [1] también se utilizaron *backtests* para evaluar el desempeño de estimaciones de medidas de riesgo; no obstante, en dicho trabajo los *backtests* se realizaron solo a lo largo de una ventana temporal de tres meses. Nuestro enfoque de evaluación contrastará con la práctica anterior pues se realizan *backtests* considerablemente más largos, lo cual está fuertemente motivado por el primero de nuestros objetivos: probar la eficacia de la metodología del Capítulo 4 para estimar medidas de riesgo a altos niveles de confianza.²⁴

Dado que nuestras pruebas nos permitirán examinar qué tan propenso es cada uno de los tres modelos considerados a subestimar o sobrestimar los verdaderos valores de VaR_{α}^t y ES_{α}^t , éstas nos permitirán elegir cuál es el mejor de ellos para cuantificar el riesgo de mercado de cada serie, permitiéndonos también alcanzar el segundo de objetivos finales.

Todos los análisis presentados en las secciones siguientes fueron realizados utilizando el *software* estadístico R. El código de las funciones desarrolladas para la realización de este trabajo puede ser consultado en el Apéndice C.

5.1. Tipo de Cambio FIX

El Tipo de Cambio (FIX) es un tipo de cambio flexible (determinado libremente por fuerzas de mercado) de pesos mexicanos contra dólares estadounidenses (USD/MXN). Este tipo de cambio es determinado por el Banco de México con base en un promedio de cotizaciones del mercado de cambios al mayoreo para operaciones liquidables el segundo día hábil bancario siguiente y que son obtenidas de plataformas de transacción cambiaria y otros medios electrónicos con representatividad en el mercado de cambios.

El Banco de México da a conocer el FIX a partir de las 12:00 horas de todos los días hábiles bancarios, éste se publica en el Diario Oficial de la Federación (DOF) un día hábil bancario después de la fecha de determinación y es utilizado para solventar obligaciones denominadas en dólares liquidables en la República Mexicana al día siguiente de la publicación en el DOF.

²⁴ Este hecho implica que los resultados presentados en las secciones finales de cada uno de los casos de estudio presentados en este trabajo no sean directamente comparables con aquellos de [1]. No obstante, al probar nuestra metodología en los datos utilizados en [1] (los log-rendimientos diarios del FIX e IPC de enero a marzo de 2013) se observó que, para ventanas temporales relativamente pequeñas y tiempos de relativa calma en los mercados, ambas técnicas ofrecen buenas estimaciones del $VaR_{0.95}^t$.

En nuestro caso, la serie estudiada corresponde a los log-rendimientos de las cotizaciones diarias históricas de este tipo de cambio. Específicamente, si la cotización al tiempo t del peso frente al dólar es de S_t pesos mexicanos por dólar estadounidense, entonces el log-rendimiento o rendimiento logarítmico estará dado por

$$x_t = \ln \left(\frac{S_t}{S_{t-1}} \right),$$

el cual, para movimientos pequeños de S_t , es aproximadamente igual al rendimiento relativo $(S_t - S_{t-1})/S_{t-1}$. En este análisis, las pérdidas son interpretadas como depreciaciones del peso respecto al dólar y se les asociará un log-rendimiento positivo.

La serie completa de log-rendimientos del FIX considerada en este trabajo abarca cotizaciones del 02 de enero de 1996 al 12 de febrero de 2016. La gráfica de esta serie de tiempo puede ser observada en la Figura 5.1.

Recordando que el enfoque de administración de riesgo de mercado que implementaremos es condicional, nuestro objetivo será modelar pérdidas diarias futuras de la serie del FIX. En otras palabras, modelaremos la función de distribución condicional del log-rendimiento del FIX al tiempo $t + 1$, con base en la información disponible hasta el tiempo t : $F_{X_{t+1}|\mathcal{F}_t}$.

Por otro lado, debido a que nuestro enfoque de administración de riesgo de mercado supone cierto comportamiento homogéneo a largo plazo de la serie de tiempo a estudiar (estacionariedad), podemos estudiar solo una parte de ésta para conocer la naturaleza del proceso estocástico que la genera. En consecuencia, los análisis y ajustes de las secciones siguientes se realizarán solo sobre una ventana de $n = 1,000$ observaciones diarias de log-rendimientos del FIX (la información disponible al tiempo t); después, utilizaremos el modelo seleccionado para dicha ventana inicial en otras ventanas temporales de la misma longitud.²⁵

Bajo las consideraciones anteriores, nuestro estudio inicial será realizado con observaciones diarias del 17 de febrero de 2012 al 12 de febrero de 2016. No obstante, en la Sección de *backtests*, utilizaremos la totalidad de nuestra serie para evaluar el desempeño de las estimaciones de medidas de riesgo obtenidas.

²⁵ Esto quiere decir que, por ejemplo, si se decide que un proceso GARCH(1,1) es el más adecuado para modelar nuestra serie inicial de 1,000 observaciones, resultaría plausible ajustar después, de forma directa, un GARCH(1,1) al FIX en otra ventana temporal, siempre que ésta tenga una longitud de 1,000 datos. Claramente, los parámetros de cada modelo podrían diferir, pero la naturaleza de ambos procesos sería la misma: GARCH(1,1).

Tipo de Cambio FIX

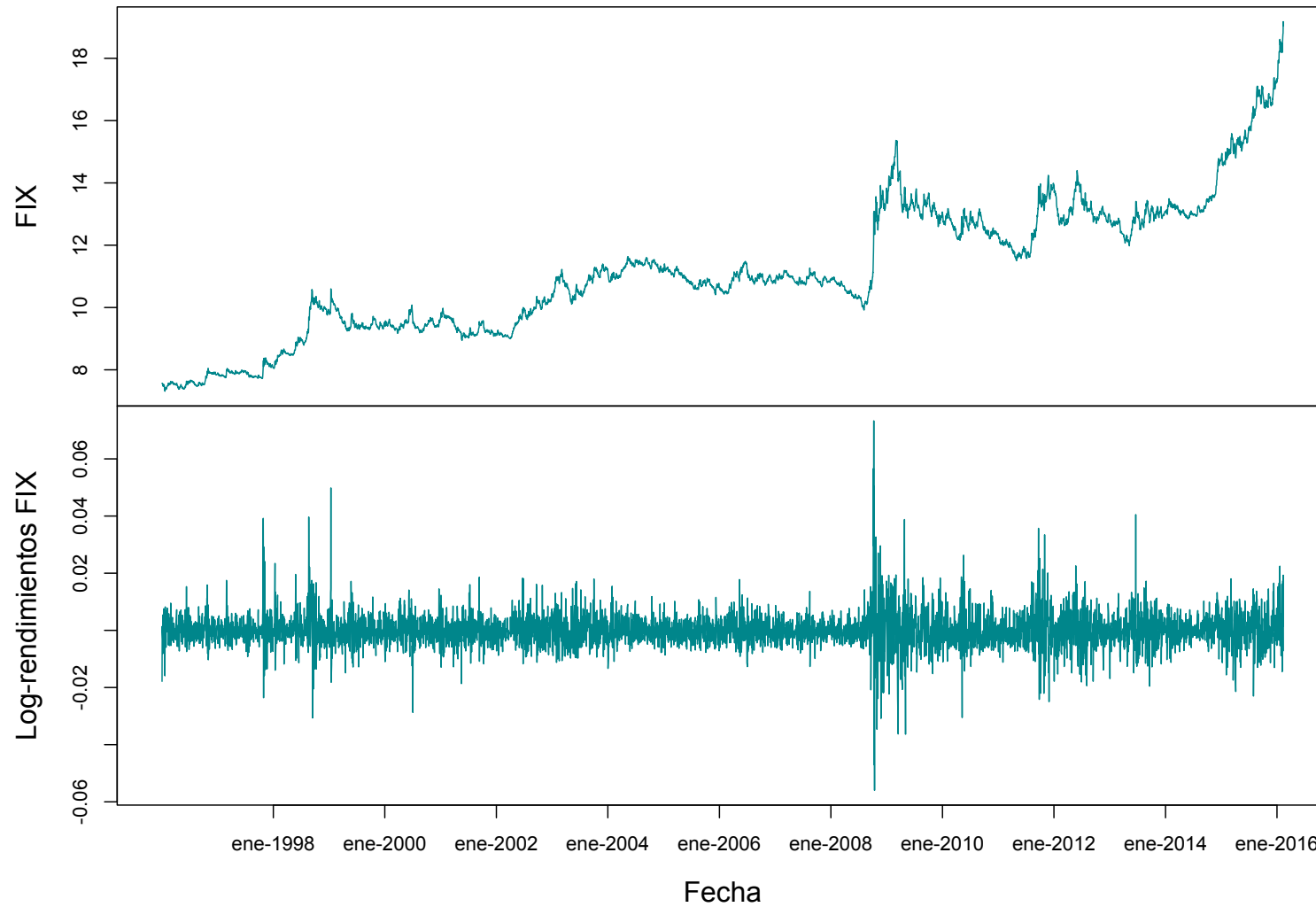


Figura 5.1: Tipo de Cambio FIX (arriba) y depreciaciones y apreciaciones del Peso Mexicano respecto al dólar estadounidense (abajo); las series van de enero 1996 a febrero de 2016.

5.1.1. Análisis descriptivo

Comenzamos con el análisis descriptivo general de este conjunto de datos. Dicho análisis consiste en la elaboración e inspección de diversas medidas y herramientas gráficas, las cuales nos permiten resumir la información contenida dentro de nuestros datos y obtener con ello una buena idea del comportamiento de los mismos.

Comenzaremos por examinar ciertas cantidades estadísticas descriptivas: media, curtosis de exceso, varianza, valor máximo y mínimo, entre otras.²⁶

La Tabla 5.1 agrupa dichas medidas, las cuales nos permiten notar lo siguiente: el valor positivo de la curtosis de exceso nos da la primera pieza de evidencia de que la serie del FIX podría tener colas pesadas; dicho valor nos indica que éstas pueden considerarse más pesadas que las colas de una distribución normal. Por otro lado, la pérdida (depreciación) más grande entre el 17 de febrero de 2012 y el 12 de febrero de 2016 fue mayor, en valor absoluto, que la ganancia (apreciación) más grande, lo que nos da las primeras señales de que, probablemente, la distribución de nuestros datos presente una cola derecha (asociada a pérdidas) más larga que la izquierda (asociada a ganancias). Adicionalmente, podemos ver que la mediana de nuestros datos es positiva, lo que significa que en más de la mitad de las ocasiones en este periodo el peso se depreció frente al dólar. Finalmente, en cada día durante nuestro periodo de estudio el peso perdió, en promedio, valor frente al dólar, lo cual se ve reflejado en una media positiva.

Este análisis nos dice que el tipo de cambio FIX presentó una tendencia importante al alza a lo largo de nuestra ventana temporal inicial: en más de la mitad de las ocasiones, el peso perdió valor frente al dólar y dichas pérdidas no pudieron ser compensadas con las apreciaciones presentadas durante el mencionado periodo. Asimismo, nos revela que existe la posibilidad de que esta serie tienda a presentar valores extremos de forma relativamente frecuente, los cuales podrían estar, en su mayoría, asociados a depreciaciones, dado el signo de la media de la serie y la diferencia en magnitud entre la apreciación y depreciación más grandes del FIX en este periodo.

²⁶ Dado que la curtosis de cualquier distribución normal univariada es igual a 3, muchas veces se compara la curtosis de una muestra con dicho valor. Para ello, se calcula la medida conocida como curtosis de exceso, la cual está definida como $\kappa = \frac{\mu_4}{\sigma^4} - 3$. De este modo, si la curtosis de exceso es mayor que cero, se tiene evidencia de colas pesadas en dicha muestra; si es igual (o muy parecida) a cero, entonces se tiene evidencia de que las colas de la distribución de la muestra son tan pesadas como las de una distribución normal; finalmente, si la curtosis de exceso es menor que cero, se tiene evidencia de que la distribución de la muestra tiene colas cortas.

Log-Rendimientos FIX: Medidas descriptivas					
Mínimo	Máximo	Mediana	Media	Varianza	Ex. Curtosis
-0.0230	0.0405	0.0003	0.0004	3.8×10^{-5}	2.5434

Tabla 5.1: Medidas de Tendencia Central y Valores Importantes de los Log-Rendimientos del Tipo de Cambio FIX entre febrero de 2012 y febrero de 2016.

Para profundizar más este primer acercamiento a nuestros datos, a continuación estudiaremos una serie de gráficos de estadística descriptiva. Dichos gráficos pueden utilizarse para explorar información de cualquier naturaleza, no necesariamente aquellos que formen una serie de tiempo, por lo que no nos brindarán información especializada sobre el comportamiento dinámico de nuestra serie. No obstante, su análisis es importante en esta etapa de nuestro estudio, pues nuestro primer objetivo es conocer lo más a fondo posible nuestros datos; ello nos permitirá ganar intuición sobre los resultados que obtengamos en análisis posteriores, los cuales serán más especializados.

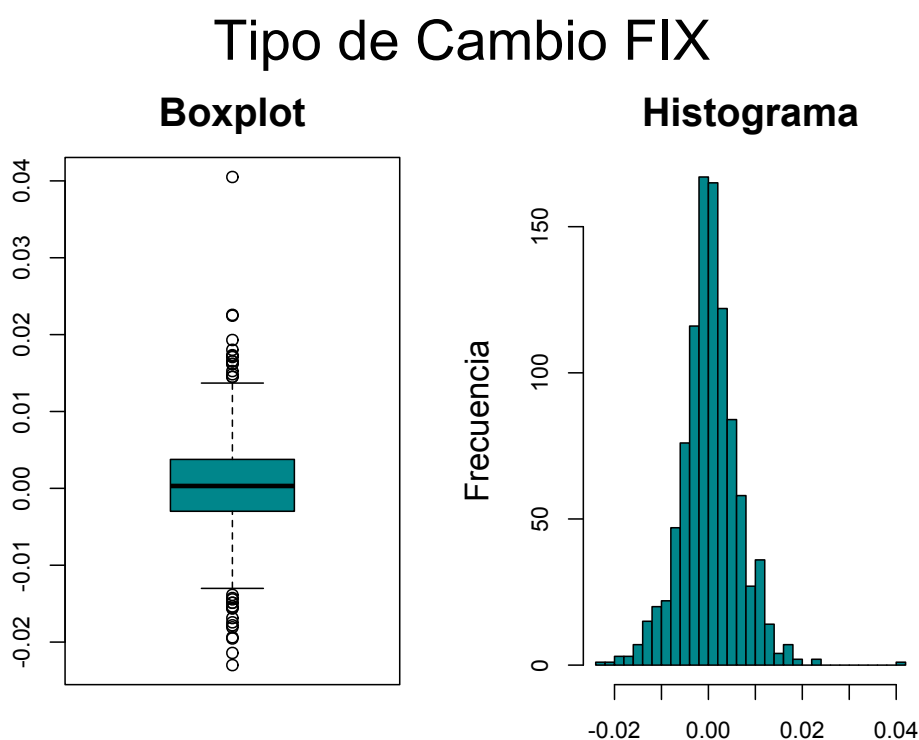


Figura 5.2: Histograma (izquierda) y *boxplot* (derecha) de los log-rendimientos del Tipo de Cambio FIX de 2012 a 2016.

Histograma y *boxplot*

En la Figura 5.2 podemos ver el Histograma y *boxplot* (diagrama de caja y brazos) del FIX. Al analizar este par de gráficas obtenemos evidencia adicional de la presencia de una cola derecha más larga que la izquierda en la distribución de estos log-rendimientos. Primeramente, un simple vistazo a la gráfica del histograma del FIX nos permite ver una cola derecha que abarca un rango considerablemente más largo que el de su contraparte izquierda.

Por otro lado, el *boxplot* nos muestra que esta distribución se encuentra centrada en cero (su media y mediana se encuentran muy cercanas a este valor) y que es aproximadamente simétrica respecto dicho valor. Adicionalmente, el *boxplot* también nos muestra que existen 3 pérdidas en esta serie que exceden el valor 0.02, mientras que solo 2 ganancias exceden dicha cantidad. Finalmente, también podemos notar el gran número de datos que están por encima y por debajo de los brazos del *boxplot*, este hecho aporta evidencia adicional de un dato de vital importancia para nuestros futuros análisis: esta distribución parece tener colas pesadas.

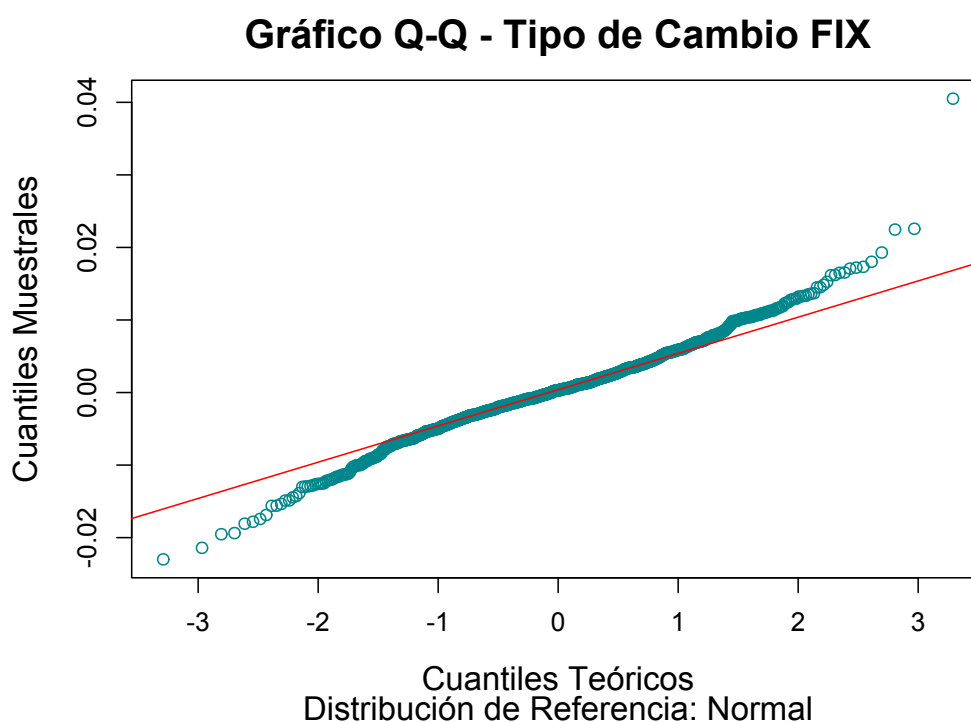


Figura 5.3: Gráfico QQ de los log-rendimientos del Tipo de Cambio FIX contra una distribución de referencia Normal. Los parámetros de la distribución Normal se obtuvieron al ajustar dicho modelo a los log-rendimientos del FIX vía Máxima Verosimilitud.

QQ-Plot

Dada la evidencia inicial sobre posibles colas pesadas en la distribución de nuestros datos, una pregunta natural es ¿qué tan pesadas son dichas colas? Una buena alternativa para investigar este hecho es comparar las colas de la distribución empírica de nuestros datos contra alguna distribución de referencia conocida. En otras palabras, realizar un QQ-Plot de nuestros datos contra una distribución cuyo comportamiento en el área de las colas sea conocido.

Una buena alternativa para llevar a cabo dicho análisis es la distribución Normal. Gracias a lo estudiado en el Capítulo 3, sabemos que las colas de esta distribución decaen siguiendo una ley exponencial y que dicha distribución no tiene colas pesadas (pues sus momentos de cualquier orden están bien definidos y son finitos). Por lo tanto, si apreciamos que las colas de la distribución empírica de nuestros datos son más pesadas que las de la distribución

normal, tendremos evidencia adicional del hecho que buscamos.²⁷

En la Figura 5.3 podemos observar dicho QQ-Plot; los parámetros de la distribución de Normal fueron obtenidos al ajustar un modelo Gaussiano a los log-rendimientos del FIX, vía el método de Máxima Verosimilitud. Al observar esta Figura podemos concluir que las colas del FIX parecen comportarse como sospechábamos: son más pesadas que las de una distribución Normal. Esto quiere decir que es probable que las colas de nuestra serie decaigan siguiendo una ley de potencias y no una ley exponencial, según lo estudiado en el Capítulo 3.

Gráfica del FIX

Una vez hemos realizado un análisis estadístico general de nuestros datos conviene recordar la naturaleza de los mismos: son observaciones de un fenómeno indexadas por un conjunto de fechas, es decir, nuestros datos conforman una serie de tiempo. Al graficar nuestras observaciones bajo esta consideración podemos observar, en la Figura 5.4, que éstas parecen haber tenido periodos de estrés: las observaciones más extremas tendieron a ubicarse cercanas entre sí, a lo largo de ciertos periodos de tiempo bien definidos. Ejemplos de este tipo de periodos de estrés pueden ubicarse, por ejemplo, entre abril y diciembre de 2013 o desde julio de 2015 a febrero 2016.

La presencia de periodos de estrés en la serie del FIX es la primera evidencia que tenemos sobre la posible no independencia ni *homocedasticidad* de dichas observaciones, pues ello indica que rendimientos de gran magnitud no son independientes entre sí y que la variabilidad del FIX tiende a cambiar a lo largo del tiempo. Sin embargo, por ahora no profundizaremos más en el análisis o manejo de este par de hechos, dejaremos tales estudios para las secciones siguientes, donde los trataremos de forma especializada.

Para terminar nuestro análisis preliminar, recapitulemos la información obtenida hasta el momento: la distribución de nuestros datos presenta evidencia de colas pesadas y parece ser aproximadamente simétrica respecto al cero; hay evidencia gráfica que señala que nuestras observaciones probablemente no sean independientes; la evidencia gráfica también apunta a que la varianza del proceso que subyace al FIX puede no ser constante a lo largo del tiempo: existen periodos de estrés bien delimitados en la serie.

²⁷ En el QQ-Plot esto se aprecia si la gráfica resultante se asemeja a una letra “S” volteada. Ello indica que los cuantiles de los datos en el área de las colas son mayores, en valor absoluto, que los correspondientes cuantiles de la distribución normal.

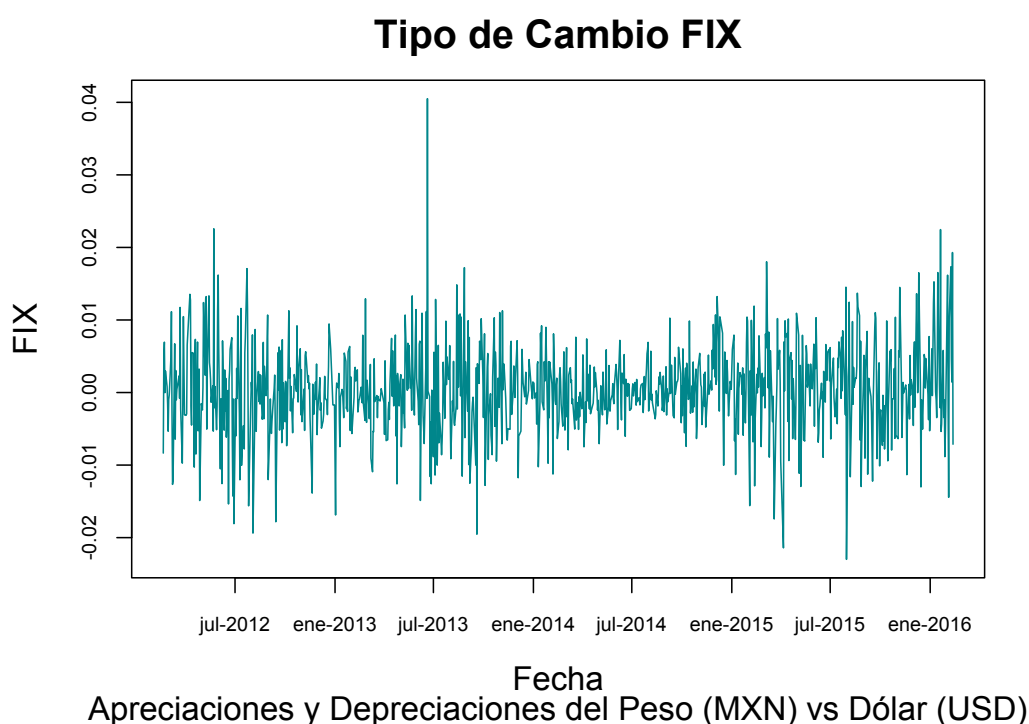


Figura 5.4: Depreciaciones y Apreciaciones del Peso Mexicano respecto al Dólar Estadounidense de febrero de 2012 a febrero de 2016.

Con este primer acercamiento a la serie del FIX podemos proceder a aplicar la metodología de administración de riesgo de mercado estudiada en el Capítulo anterior a dicha serie de tiempo. Comenzaremos por modelar la dinámica de la serie del FIX. Para ello, en la siguiente Sección evaluaremos de forma cuidadosa la plausibilidad de modelar dicha dinámica por medio de procesos de las familias ARCH y GARCH.

5.1.2. Modelando la dinámica del FIX

En esta Sección ilustraremos cómo realizar el ajuste de un modelo de las familias ARCH y GARCH a una serie de tiempo financiera; a este procedimiento también se le denomina “modelar la dinámica de la serie de tiempo”.

Naturalmente, el primer paso en este tipo de análisis es investigar si el ajuste de tales procesos resulta razonable para modelar el conjunto de datos con el que estamos trabajando. Por lo tanto, a continuación realizaremos una serie de pruebas que nos ayudarán a evaluar la plausibilidad de que la serie del

FIX pueda ser modelada vía un proceso ARCH o GARCH.

Hechos estilizados de series con volatilidad estocástica

Para investigar si un modelo ARCH o GARCH podría ajustar satisfactoriamente a al FIX, debemos realizar un análisis exploratorio más especializado que el de la Sección 5.1.1. Dicho análisis consiste en verificar la presencia de un conjunto de hechos estilizados²⁸ en dicha serie de tiempo:

1. Rendimientos que no son i.i.d., aunque presentan una correlación serial leve.
2. Las series del valor absoluto y cuadrado de los rendimientos muestran una correlación serial muy fuerte.
3. Los rendimientos esperados condicionales (con toda la información disponible hasta, e incluyendo, el tiempo t) son cercanos a cero.
4. Volatilidad aparentemente cambiante a lo largo del tiempo.
5. Series de rendimientos leptocúrticas o con colas pesadas.
6. Los rendimientos extremos aparecen en *clusters*.

Comencemos entonces por investigar la presencia de este conjunto de características en la serie del FIX. Primeramente, notemos que la información recopilada a través de nuestro análisis exploratorio inicial puede ayudarnos a identificar la presencia de algunos de estos hechos estilizados. En particular, con ayuda de la curtosis de exceso, el Gráfico QQ y el *boxplot* e histograma del FIX podemos notar que el hecho estilizado número 5 está presente en dicha serie de tiempo. Por otro lado, gracias al estudio de la gráfica del FIX como serie de tiempo, contamos con evidencia inicial de los hechos 4 y 6; no obstante, para obtener evidencia definitiva sobre la presencia de estos dos hechos estilizados en el FIX, es necesario explorar dicha serie de forma más profunda.

Para iniciar nuestro análisis exploratorio especializado, recordemos que si una muestra consiste de datos independientes, entonces los valores de su función de autocorrelación muestral deberían ser estadísticamente cero, para cualquier retraso que consideremos. Adicionalmente, si aplicáramos una transformación a dicha muestra independiente, los datos resultantes volverían a

²⁸ En este contexto, estas características son hechos empíricos, identificados por la industria y la academia, comúnmente presentes en series de tiempo financieras que son modeladas satisfactoriamente por procesos GARCH o ARCH.

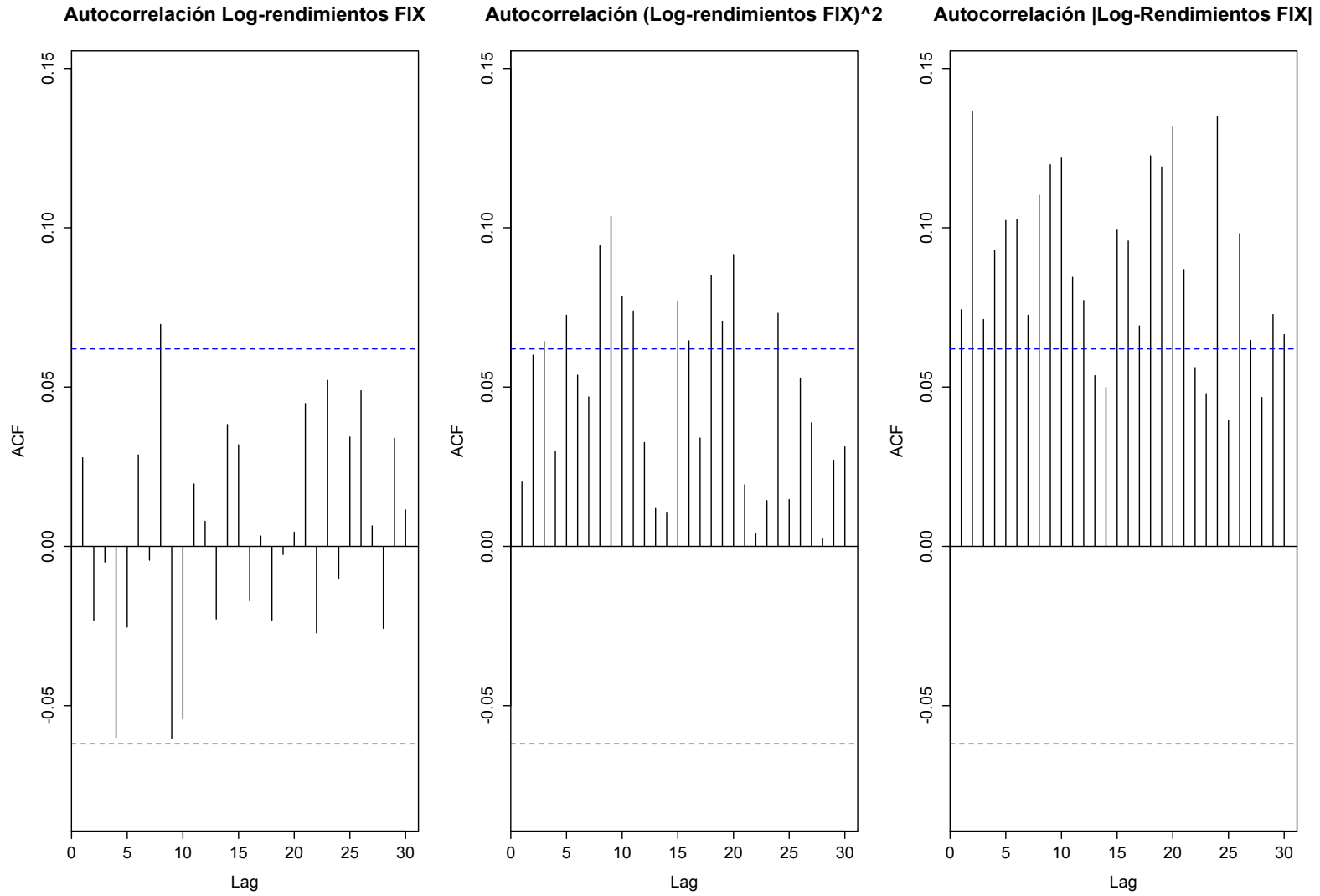


Figura 5.5: Autocorrelación del FIX. Se muestran las gráficas de la función de autocorrelación muestral para los datos crudos (izquierda), sus cuadrados (centro) y valores absolutos (derecha).

Prueba de Ljung-Box en el FIX: p-valores			
Datos	$h = \ln(n)$	$h = 20$	$h = 30$
FIX original	0.4867	0.3190	0.2877
FIX al cuadrado	0.0067	5.61×10^{-10}	4.66×10^{-09}
FIX en valor absoluto	1.917×10^{-11}	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

Tabla 5.2: P-valores de las pruebas de Ljung-Box aplicadas a las observaciones del Tipo de Cambio FIX crudas, al cuadrado y en valor absoluto.

ser independientes. La idea para investigar el hecho estilizado número 1 se basa en las observaciones anteriores: graficaremos la función de autocorrelación muestral de los datos del FIX crudos, al cuadrado y en valor absoluto.

En la Figura 5.5 podemos observar estas tres gráficas de autocorrelación. Dicha Figura, nos permite ver que, en efecto, las autocorrelaciones de nuestros rendimientos crudos, en los retrasos considerados, son estadísticamente iguales a cero; no obstante, para las series de los cuadrados y valores absolutos del FIX se observa un panorama distinto: las gráficas de autocorrelación de estas series muestran varios valores estadísticamente distintos de cero, lo cual aporta evidencia en contra de que los datos del FIX sean independientes.

Para investigar de forma más profunda la independencia de los datos del FIX, podemos aplicar una prueba de Ljung-Box²⁹ a dicha serie, así como a sus valores al cuadrado y en valor absoluto. Para el FIX, aplicamos la prueba de Ljung-Box con base en tres retrasos distintos, los resultados pueden ser consultados en la Tabla 5.2. Dicha Tabla nos permite concluir que, en efecto, las observaciones del FIX no son independientes: a pesar de no rechazar la hipótesis de independencia para los datos crudos, las pruebas aplicadas a los datos transformados rechazaron fuertemente dicha hipótesis para todos los retrasos considerados. De esta forma, podemos concluir que el hecho estilizado número 1 se encuentra presente en esta serie de tiempo.

Por otro lado, nuestra elección de funciones para probar la independencia de nuestros datos claramente no fue arbitraria: al observar una correlación serial fuerte en las series de los cuadrados y valores absolutos del FIX, también hemos documentado la presencia del hecho estilizado número 2 en dicha serie de tiempo.

Ahora bien, en el Capítulo 2 definimos la volatilidad como la desviación estándar de los rendimientos de un activo financiero, por lo tanto, para es-

²⁹ Para detalles sobre esta prueba, así como sobre la elección de los retrasos utilizados en la misma, véase el Apéndice B.

tudiarla es conveniente fijar nuestra atención en los cambios de magnitud de nuestras observaciones a lo largo del tiempo. Con esto en mente, y para investigar de forma más profunda sobre los hechos estilizados 4 y 6, observaremos las gráficas de las series de los cuadrados y valores absolutos del FIX. Estudiar dichas gráficas es conveniente pues a través de ellas podemos observar de una forma más clara la variabilidad de una serie de tiempo; en otras palabras, al analizarlas podemos obtener información sobre la volatilidad de forma más directa.

En la Figura 5.6, podemos observar las gráficas de las series del cuadrado y valor absoluto del FIX. Dichas gráficas nos muestran de forma clara que los datos del FIX sí presentan periodos de estrés bien definidos, por ejemplo: de febrero a octubre de 2012; de mayo a diciembre de 2013; de mayo de 2014 a mayo de 2015 y; desde julio de 2015 al final de nuestra ventana temporal en febrero de 2016. En todos estos periodos existieron rendimientos de grandes magnitudes (quizá con signos diferentes) que se presentaron juntos y que al terminar dieron lugar a épocas de relativa calma, antes de que iniciara el siguiente periodo de estrés.

El análisis anterior documenta la presencia del hecho estilizado número 4 en la serie del FIX: al existir periodos de estrés y relativa calma bien delimitados queda establecido que la volatilidad o variabilidad del FIX tiende a cambiar a lo largo del tiempo. Por otro lado, el hecho de que los rendimientos de magnitud grande se presenten juntos es evidencia de que nuestra serie tiende a formar *clusters* de rendimientos extremos, permitiéndonos concluir que el hecho estilizado número 6 también se encuentra presente en el FIX.

Ahora bien, notemos que el hecho de que existan *clusters* de volatilidad en una serie de tiempo financiera es coherente: si un instrumento presenta rendimientos extremos (cualquiera que sea su signo), es de esperarse que los participantes del mercado tomen acciones (vender, comprar, etc.) para sacar provecho (maximizar ganancias o minimizar pérdidas) del comportamiento de dicho instrumento. De este modo, al existir en este escenario súbitas posturas de compra y venta del activo en cuestión, el precio de este instrumento sufrirá cambios considerables, los cuales estarán presentes hasta que el activo llegue nuevamente a un precio de equilibrio, dando por terminado el periodo de alta volatilidad (*cluster*).

Finalmente, investigaremos la presencia del hecho estilizado número 3 en la serie del FIX. Para ello, notemos que si intentáramos utilizar únicamente rendimientos crudos pasados para predecir rendimientos futuros, necesariamente tendríamos que apoyar nuestras predicciones en lo que nos señala la función de autocorrelación, pues esta nos da una idea de cómo influyen, en

promedio, rendimientos pasados a los siguientes.

Sin embargo, como notamos en el hecho estilizado número 1, la correlación serial de nuestra serie es muy baja, por lo que la información exclusivamente de rendimientos crudos anteriores no nos dice mucho sobre cómo podrían comportarse los rendimientos futuros. Este hecho, en conjunto con la forma de nuestra serie (centrada en cero) sugieren que nuestra mejor predicción sobre los rendimientos de días siguientes, basados únicamente en información de residuos crudos, es que éstos serán cero, dando evidencia del hecho estilizado número 3 en nuestra serie.

De este modo, hemos terminado de investigar la presencia de los hechos estilizados de series financieras en la serie de log-rendimientos del Tipo de Cambio FIX, concluyendo que todos ellos se encuentran presentes en ella. Claramente, estas características representan la esencia de nuestros datos, por lo que una vez identificadas debemos proponer un proceso que las trate de forma tan especializada como sea posible, a fin modelar de forma apropiada el FIX: ello nos permitirá obtener un ajuste satisfactorio, que eventualmente tenga un desempeño óptimo en un sistema de administración de riesgo de mercado. Afortunadamente, gracias al material estudiado en el Capítulo 2, conocemos una clase de modelo que trata de forma especializada este tipo de series financieras: los procesos de las familias ARCH y GARCH.

Ajustando un modelo de volatilidad estocástica al FIX

Dada la evidencia de que un proceso de las familias ARCH o GARCH podría modelar de forma razonable al FIX, el siguiente paso a seguir es investigar qué proceso ARCH o GARCH puede ajustarse de mejor forma a dicha serie de tiempo. En otras palabras, determinar el orden del modelo ARCH o GARCH que subyace a nuestros datos.

Para alcanzar dicho objetivo, recordemos que en el Capítulo 2 analizamos una importante relación de los procesos ARCH y GARCH con los procesos AR y ARMA. Específicamente, estudiamos que bajo ciertas hipótesis sobre dichos procesos, el cuadrado de un proceso ARCH(p) o GARCH(p, q) podía ser reescrito como un proceso AR(p) o ARMA($\max\{p, q\}, q$), respectivamente.

Por otro lado, en el Capítulo 2 también estudiamos ciertas herramientas que nos permiten determinar de forma práctica el orden de un modelo ARMA. En particular, estudiamos cómo utilizar las gráficas de las funciones de autocorrelación y autocorrelación parcial empíricas para determinar el orden del modelo ARMA que subyace a un conjunto de datos.

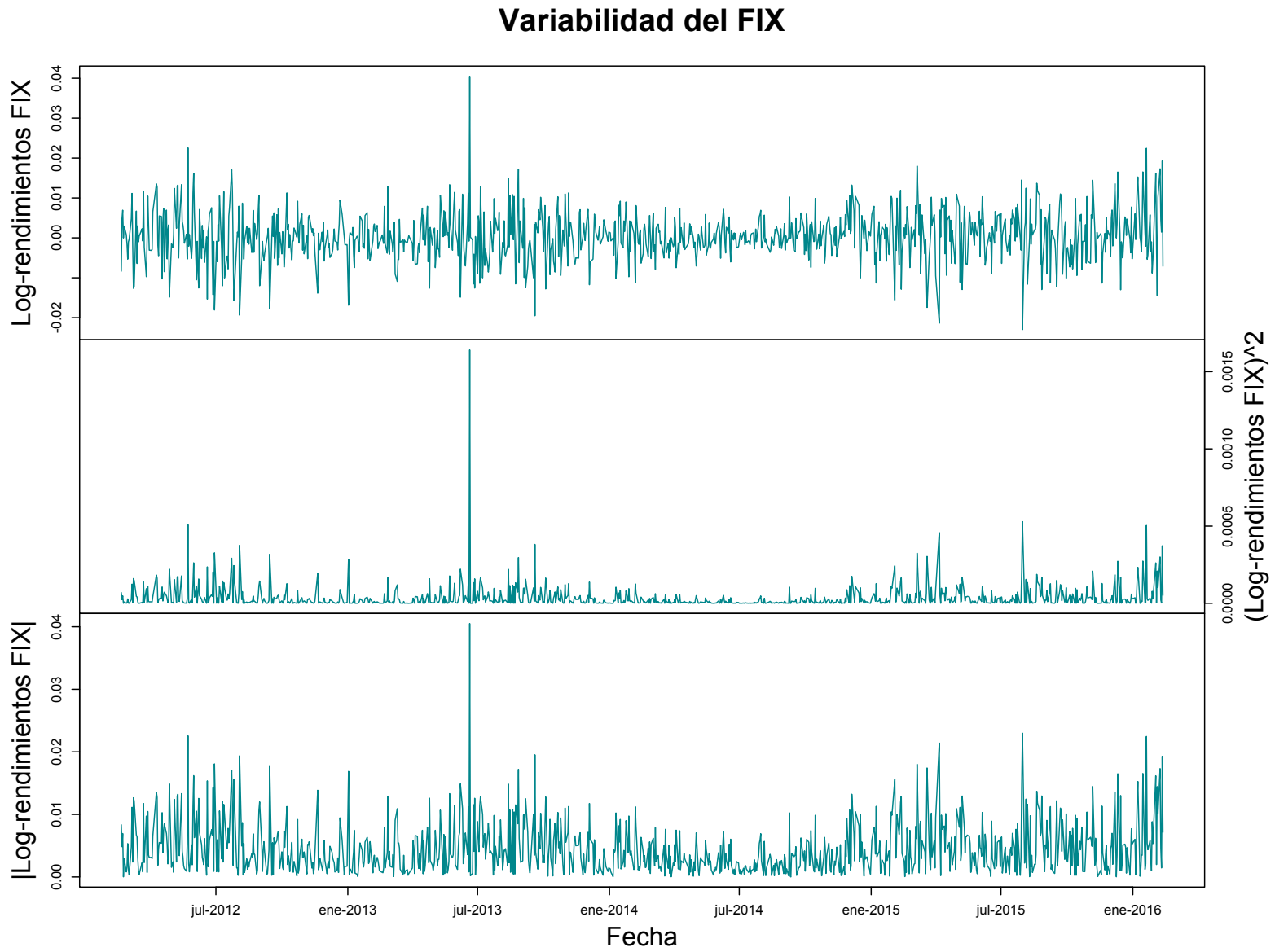


Figura 5.6: Gráficas del Tipo de Cambio FIX. Observaciones crudas (arriba), al cuadrado (en medio) y en valor absoluto (abajo).

(Log-rendimientos FIX)²

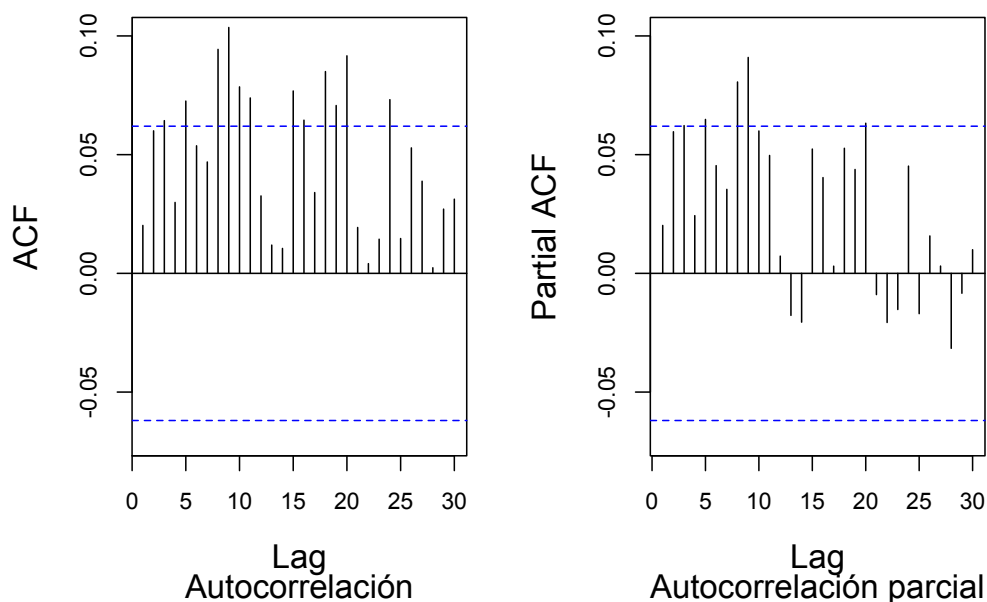


Figura 5.7: Autocorrelación (izquierda) y autocorrelación parcial (derecha) de los log-rendimientos al cuadrado del Tipo de Cambio FIX.

Claramente, podemos combinar las ideas anteriores para determinar el orden del proceso ARCH o GARCH que subyace al FIX. La idea es la siguiente: bajo la hipótesis de que las observaciones del FIX siguen un proceso ARCH o GARCH, sus cuadrados deben seguir un proceso AR o ARMA, respectivamente. Por lo tanto, podemos analizar las gráficas de las funciones de autocorrelación y autocorrelación parcial de las observaciones del FIX al cuadrado para determinar el orden del proceso ARCH o GARCH que subyace al FIX.

De este modo, en la Figura 5.7 podemos observar las gráficas del ACF y PACF de las observaciones del FIX al cuadrado. Analizando este par de gráficas podemos notar que:

1. La gráfica de autocorrelación parcial presenta sus valores más grandes en los retrasos $h = 8$ y $h = 9$. Ello nos indica que ajustar modelos ARCH(p) con $p = 1, \dots, 9$ podría resultar razonable para nuestros datos. No obstante, ello no sería práctico debido a que ajustar modelos de órdenes tan altos implica la estimación de muchos parámetros y

Log-Rendimientos FIX: AICC modelos ARCH(p)					
p	1	2	3	4	5
AICC	-7.321731	-7.332685	-7.321877	-7.295953	-7.274392

Tabla 5.3: AICC de los modelos ARCH(p) ajustados a la serie del Tipo de Cambio FIX.

podría llevarnos a *sobreajustar* nuestro modelo.³⁰ Por lo tanto, tomaremos un enfoque más parsimonioso y ajustaremos modelos ARCH(p) con $p = 1, 2, 3, 4, 5$.

- Por otro lado, la gráfica de autocorrelación también tiene sus valores más alto en $h = 8$ y $h = 9$, por lo que podríamos considerar ajustar al FIX modelos GARCH(p, q) con $q = 1, \dots, 9$. Sin embargo, si ajustamos modelos de órdenes tan altos de nueva cuenta correríamos el riesgo de sobreajustar nuestro modelo para la dinámica del FIX. En consecuencia, en esta ocasión también adoptaremos un enfoque parsimonioso y ajustaremos al FIX modelo GARCH(p, q) donde $q = 1, 2, 3, 4, 5$.

Dado que el análisis anterior nos proveyó varias opciones de modelos que podrían resultar plausibles para nuestros datos, el siguiente paso a seguir es elegir el mejor de ellos. Nuestro enfoque para seleccionar un modelo entre nuestras las distintas opciones estará basado en el criterio de Información de Akaike (AICC por sus siglas en inglés).³¹

Dividiremos este análisis por familia de procesos y comenzaremos por elegir el mejor modelo ARCH para nuestros datos. En la Tabla 5.3 podemos ver el AICC de los modelos ARCH(p), $p = 1, 2, 3, 4, 5$, ajustados a la serie del FIX.³² En dicha Tabla podemos apreciar que el AICC señala que el mejor modelo ARCH para la serie del FIX es el ARCH(2).³³

³⁰ Sobreajustar un modelo quiere decir que se han estimado demasiados parámetros (es decir, el modelo se ha vuelto demasiado complejo) para explicar el comportamiento de un conjunto de datos. En otras palabras, el modelo se ha vuelto tan elaborado que comienza a describir errores puramente aleatorios en la información, en lugar de explicar la relación subyacente entre las observaciones que la componen. Un modelo sobreajustado típicamente tiene un mal desempeño al realizar pronósticos, pues puede exagerar fluctuaciones menores en la información.

³¹ Para información adicional sobre este criterio, véase el Apéndice B.

³² Dado que nuestro objetivo actual de modelar la dinámica del FIX, en el ajuste de estos modelos ARCH y GARCH utilizamos un enfoque quasi-máximo verosímil; es decir, el ajuste en dos etapas del ARCH y el GARCH estudiado en el Capítulo 2.

³³ El criterio de selección considera que el mejor modelo para un cierto conjunto de datos es aquel con el AICC más pequeño. Véase Apéndice B.

Significancia de parámetros, modelo ARCH(2)		
Parámetro	Estimación puntual	P-valor
α_0	2.628×10^{-05}	$< 2 \times 10^{-16}$
α_1	0.1181	0.0147
α_2	0.2260	1.34×10^{-05}

Tabla 5.4: Significancia estadística de los parámetros del modelo ARCH(2) ajustado a la serie del Tipo de Cambio FIX.

Log-Rendimientos FIX: AICC modelos GARCH(p, q)					
p/q	1	2	3	4	5
1	-7.418316	-7.396137	-7.369946	-7.339676	-7.305329
2	-7.396262	-7.370549	-7.342804	-7.309593	-7.271087
3	-7.370037	-7.340216	-7.309381	-7.272666	-7.230057
4	-7.33972	-7.306778	-7.271002	-7.230037	-7.183298
5	-7.305895	-7.268266	-7.228429	-7.183311	-7.132408

Tabla 5.5: AICC de los modelos GARCH(p, q) ajustados a la serie del Tipo de Cambio FIX.

Antes de tomar el proceso ARCH(2) como modelo ARCH final para el FIX debemos comprobar que los parámetros del modelo ajustado al FIX sean estadísticamente significativos. En la Tabla 5.4, podemos ver los valores estimados y p-valores de la prueba t de significancia estadística de estos parámetros. Los valores contenidos en dicha Tabla indican que los parámetros del modelo son estadísticamente significativos y que el modelo ARCH que subyace a nuestros datos es estacionario en covarianza (pues $\alpha_1 + \alpha_2 = 0.3441 < 1$). Por lo tanto, tomaremos el proceso ARCH(2) como modelo ARCH final para la serie del FIX.

Procedemos ahora a seleccionar el mejor modelo GARCH para el FIX. Nuevamente, utilizaremos el AICC para elegir el mejor modelo entre todas las propuestas sugeridas por nuestro análisis de las gráficas del ACF y PACF del FIX al cuadrado.

En la Tabla 5.5 podemos observar el AICC de los modelos GARCH(p, q), $p = 1, 2, 3, 4, 5$ y $q = 1, 2, 3, 4, 5$, ajustados a la serie del FIX. Los valores contenidos en dicha Tabla señalan que el mejor modelo GARCH para nuestros datos es el GARCH(1, 1).

El siguiente paso consiste en comprobar que los parámetros del modelo GARCH(1,1) ajustado al FIX sean estadísticamente significativos. En la Tabla 5.6 podemos ver los valores estimados y p-valores de la prueba t de sig-

Significancia de parámetros, modelo GARCH(1, 1)		
Parámetro	Estimación puntual	P-valor
α_0	2.220×10^{-07}	0.106
α_1	0.0581	7.15×10^{-12}
β_1	0.9396	$< 2 \times 10^{-16}$

Tabla 5.6: Significancia estadística de los parámetros del modelo GARCH(1, 1) ajustado a la serie del Tipo de Cambio FIX.

nificancia estadística de estos parámetros. Dicha Tabla nos permite concluir que los parámetros α_1 y β_1 de nuestro modelo son estadísticamente significativos y que el modelo GARCH que subyace a nuestros datos es estacionario en covarianza ($\alpha_1 + \beta_1 = 0.9977 < 1$).

No obstante, notemos que el parámetro α_0 del GARCH(1, 1) ajustado al FIX no es estadísticamente significativo, por lo que éste puede considerarse estadísticamente igual a cero. Dado que, según lo estudiado en el Capítulo 2 nuestra fórmula para pronosticar volatilidad con base en un modelo GARCH(1,1) está dada por

$$\hat{\sigma}_{t+h} = \hat{\alpha}_0 \sum_{i=0}^{h-1} (\hat{\alpha}_1 + \hat{\beta}_1)^i + (\hat{\alpha}_1 + \hat{\beta}_1)^{h-1} (\hat{\alpha}_1 X_t^2 + \hat{\beta}_1 \hat{\sigma}_t^2), \quad (5.2)$$

este hecho nos dice que dicho modelo podría tener problemas para pronosticar la volatilidad del FIX varios días adelante, pues dicho pronóstico pierde un factor importante al desvanecerse el parámetro α_0 .

Sin embargo, el GARCH(1,1) podría, de cualquier forma, modelar apropiadamente la dinámica de nuestras observaciones, pues la fórmula de estimación de volatilidad con base en el GARCH(1,1) es menos dependiente del parámetro α_0 :

$$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_{t-1}^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2. \quad (5.3)$$

De este modo, tenemos dos procesos potenciales para modelar la dinámica de la serie del FIX: un ARCH(2) y un GARCH(1, 1). Dado que ambos procesos son estacionarios en covarianza, debemos utilizar otro criterio para elegir al mejor de ellos para modelar el FIX. El criterio que utilizaremos será, nuevamente, el AICC. Al comparar esta medida entre ambos candidatos podemos observar que el modelo GARCH(1, 1) es señalado como el mejor de ellos.

No obstante, antes de declarar al GARCH(1, 1) como nuestro modelo final para el FIX, debemos asegurar que dicho proceso modela apropiadamente la

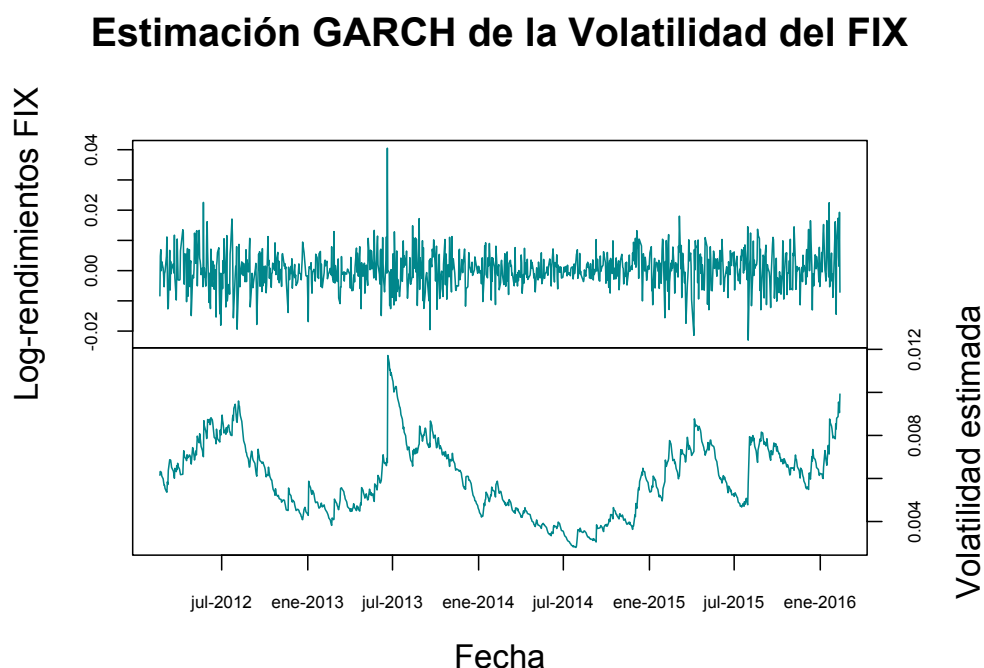


Figura 5.8: Series de log-rendimientos y de volatilidad estimada del Tipo de Cambio FIX. La volatilidad del FIX fue estimada vía el modelo GARCH(1,1) ajustado a la serie del FIX.

dinámica de nuestra serie. Es decir, debemos verificar la calidad del ajuste del GARCH(1, 1) a la serie del FIX.

Evaluación de la calidad de ajuste del GARCH(1,1) a la dinámica del FIX

La evaluación de la calidad de ajuste del GARCH(1, 1) al FIX puede llevarse a cabo mediante un análisis de residuales: obteniendo una versión muestral de las innovaciones (residuos estandarizados) del proceso GARCH ajustado al FIX y verificando que éstas se comporten como un ruido blanco estricto.

Para obtener la versión muestral de los residuales del GARCH(1,1) ajustado al FIX estimamos la volatilidad de dicha serie de tiempo vía el modelo ajustado. Esto puede lograrse al sustituir los valores estimados de los parámetros de dicho modelo en la fórmula (5.3) y evaluándola recursivamente para obtener los valores $\hat{\sigma}_t^2$, $t \in \{1, \dots, 1,000\}$. Claramente, necesitamos un par de

valores iniciales $\hat{\sigma}_0^2$ y \hat{X}_0^2 . En este trabajo, fijaremos dichos valores, respectivamente, como la varianza y media muestrales del FIX. Así, la volatilidad estimada puede ser calculada aplicando la función $f(x) = \sqrt{x}$ a los valores $\hat{\sigma}_t^2$.

En la Figura 5.8, podemos observar la serie de log-rendimientos del FIX y la volatilidad del FIX estimada vía el GARCH(1,1). Dicha Figura nos permite observar el comportamiento de la volatilidad estimada mediante el GARCH: ésta reacciona de forma dinámica a las observaciones del FIX; aumenta en los periodos con *clusters* de valores grandes (periodos de estrés) y disminuye en aquellos de relativa calma (donde no se observan cúmulos de observaciones extremas).

Podemos entonces utilizar los valores de la volatilidad estimada para obtener la versión muestral de la innovaciones del FIX, a fin de evaluar la calidad de ajuste del GARCH(1,1) a la serie del FIX.

Para evaluar el comportamiento como ruido blanco estricto en las innovaciones del FIX podemos hacer uso de pruebas tanto gráficas como numéricas. Comenzando por las gráficas, en la Figura 5.9 podemos observar una serie de gráficos de dichas innovaciones: las gráficas como serie de tiempo de las innovaciones crudas, al cuadrado y en valor absoluto. Dicha Figura nos muestra que la serie de innovaciones ya no presenta los *clusters* de valores extremos que aparecían en la serie original del FIX, lo que proporciona evidencia inicial de que la dinámica del FIX fue correctamente estimada por el GARCH.

Continuando con la evidencia gráfica, en la Figura 5.10 podemos observar las gráficas de las funciones de autocorrelación y autocorrelación parcial de las innovaciones del FIX crudas, al cuadrado y en valor absoluto. Dichas gráficas proporcionan evidencia adicional de que la hipótesis de un comportamiento ruido blanco estricto en las innovaciones del FIX es plausible: todas las autocorrelaciones y autocorrelaciones parciales consideradas son estadísticamente iguales a cero.

Por otro lado, pasando a las pruebas numéricas, podemos aplicar la prueba de Ljung-Box a las innovaciones del FIX. Basaremos nuestra aplicación de la prueba en distintos retrasos y la aplicaremos a los residuos estandarizados crudos, en valor absoluto y al cuadrado. En la Tabla 5.7 podemos encontrar los p-valores de las pruebas de Ljung-Box aplicadas: en ningún caso rechazamos la hipótesis de independencia. Claramente, esto refuerza de forma importante la evidencia de que las innovaciones del FIX presentan un comportamiento ruido blanco estricto.

El análisis anterior nos brinda certeza de que, aún a pesar de que el parámetro

Variabilidad de las Innovaciones del FIX

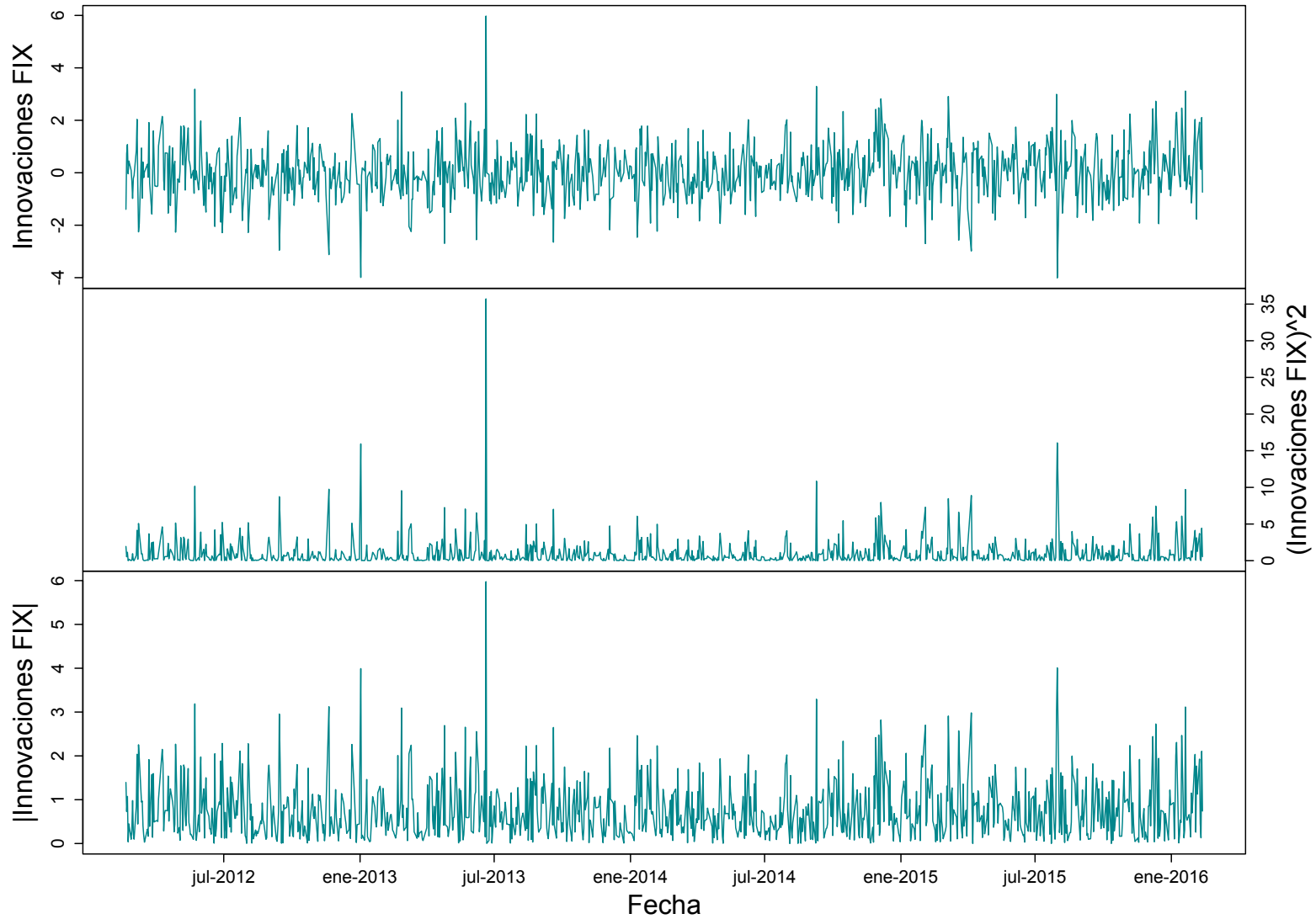


Figura 5.9: Gráficas como serie de tiempo de las Innovaciones del GARCH(1,1) ajustado al Tipo de Cambio FIX: innovaciones crudas (arriba), al cuadrado (en medio) y en valor absoluto (abajo).

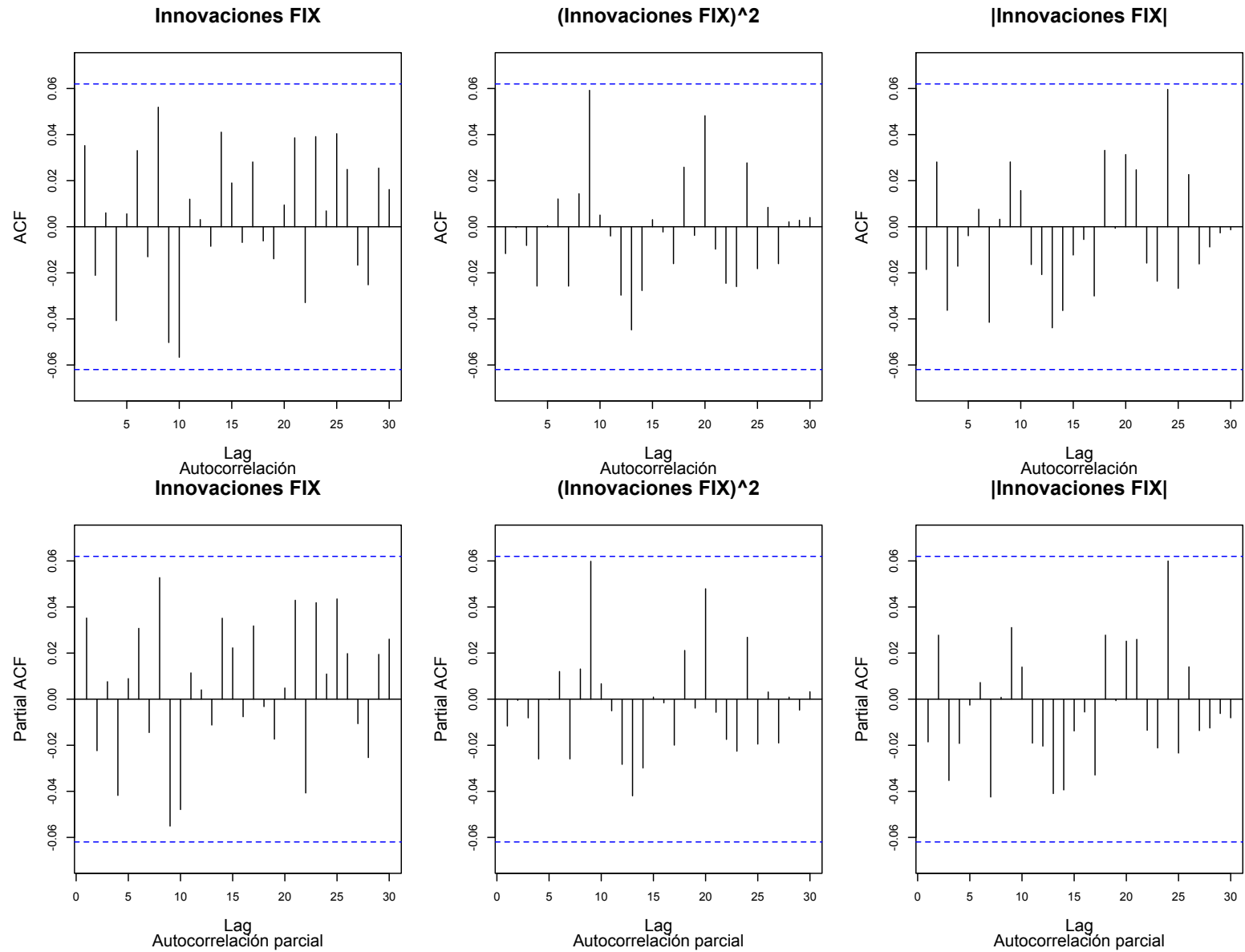


Figura 5.10: Autocorrelación y autocorrelación parcial de los residuos estandarizados del modelo GARCH(1,1) ajustado al Tipo de Cambio FIX. Se muestran las gráficas del ACF (primer renglón) y PACF (segundo renglón) de los datos crudos, al cuadrado, y en valor absoluto.

Prueba de Ljung-Box en las innovaciones del FIX: p-valores			
Datos	$h = \ln(n)$	$h = 20$	$h = 30$
FIX original	0.6967	0.6721	0.7181
FIX al cuadrado	0.9752	0.8967	0.9871
FIX en valor absoluto	0.7131	0.8834	0.9272

Tabla 5.7: P-valores de las pruebas de Ljung-Box aplicadas a las innovaciones del FIX crudas, al cuadrado y en valor absoluto.

α_0 no resultó estadísticamente significativo, el modelo GARCH(1,1) fue capaz de modelar apropiadamente la dinámica del FIX: existe evidencia consistente de pruebas gráficas y numéricas de que las innovaciones del GARCH(1,1) ajustado al FIX se comportan como un ruido blanco estricto. Por lo tanto, este será nuestro modelo final para la dinámica del FIX.

Pronóstico de la volatilidad del FIX

En el Capítulo 2, estudiamos dos metodologías para pronosticar la volatilidad de una serie financiera: el suavizamiento exponencial (EWMA) y el pronóstico basado en el ajuste de un modelo GARCH. La fórmula de pronóstico de volatilidad a un paso con base en un modelo GARCH(1,1) ajustado resultó:

$$\hat{\alpha}_0 + \hat{\alpha}_1 X_t^2 + \hat{\beta}_1 \hat{\sigma}_t^2$$

Si bien el GARCH(1,1) ajustado al FIX logró estimar adecuadamente la volatilidad de dicha serie, el hecho de que su parámetro α_0 no sea estadísticamente significativo nos impide basar nuestro pronóstico de la volatilidad del FIX en dicho modelo (pues dicha fórmula fue desarrollada bajo la hipótesis de todos los parámetros del modelo GARCH(1,1) considerado son positivos). Por lo tanto, en este caso haremos uso del EWMA para pronosticar la volatilidad de nuestra serie de tiempo.

La fórmula del EWMA para pronosticar volatilidad al tiempo $t+1$ está dada por:

$$\hat{\sigma}_{t+1}^2 = \lambda X_t^2 + (1 - \lambda) \hat{\sigma}_t^2.$$

Ahora bien, dadas las consideraciones anteriores nos queda una pregunta importante por responder: ¿qué coeficiente λ deberíamos utilizar para realizar nuestro pronóstico de volatilidad? Para proponer un λ , podemos aprovechar

el hecho de que el modelo GARCH(1,1) ajustado al FIX sí modeló apropiadamente la volatilidad de dicha serie.³⁴

Para ello, notemos la fuerte similitud entre la fórmula de estimación de volatilidad del GARCH(1,1) y la del pronóstico EWMA: en ambas fórmulas se ponderan los factores $\hat{\sigma}_t^2$ y X_t^2 ; la única diferencia entre éstas es el factor α_0 , el cual es despreciable en nuestro caso. Adicionalmente, notemos que $\alpha_1 + \beta_1 = 0.9396 + 0.0581 = 0.9977$, el cual es un valor muy cercano a 1.

De este modo, una propuesta para λ sería fijar el valor de dicho coeficiente como: $\lambda = \alpha_1 = 0.0581$. Así, la técnica EWMA de pronóstico de volatilidad recupera casi de forma exacta la ecuación de volatilidad basada en el GARCH(1,1) para el FIX, pues $1 - \lambda = 0.9419$, el cual es un valor bastante cercano al de β_1 .

Siguiendo la metodología descrita, nuestro pronóstico de volatilidad para el FIX al tiempo $t + 1$ toma la forma:

$$\begin{aligned}\hat{\sigma}_{t+1} &= (\lambda X_t^2 + (1 - \lambda)\hat{\sigma}_t^2)^{1/2} \\ &= ((0.0581)X_t^2 + (0.9419)\hat{\sigma}_t^2)^{1/2} \\ &= ((0.0581)(-0.007128)^2 + (0.9419)(0.009923)^2)^{1/2} \\ &= 0.009782.\end{aligned}$$

Una desventaja de basar un pronóstico de volatilidad en el EWMA y no en el GARCH es que sólo pueden obtenerse estimaciones de volatilidad 1 paso adelante. Sin embargo, para alcanzar nuestros objetivos finales nosotros solo requeriremos, en cada ocasión, pronósticos de volatilidad a un día, por lo que esto no es una preocupación.

Con el análisis anterior hemos terminado de modelar la dinámica del FIX. Recapitulando: fuimos capaces de ajustar exitosamente un modelo de volatilidad estocástica a dicha serie, lo cual se verificó al probar que las innovaciones del FIX bajo nuestro modelo se comportan como un ruido blanco estricto. Además, derivamos una técnica de estimación y pronóstico de volatilidad para dicha serie de tiempo.

Procederemos entonces con la siguiente etapa de nuestra metodología (analizar a profundidad la naturaleza de la distribución de las innovaciones del

³⁴ Esta clase de propuestas ya se han utilizado con anterioridad. Para referencias sobre otros casos de estudio en donde un enfoque similar a este se ha utilizado véase [29] o [1].

Innovaciones del FIX: Medidas descriptivas					
Mínimo	Máximo	Mediana	Media	Varianza	Ex. Curtosis
-4.0090	5.9760	0.0101	0.0367	1.0006	1.8505

Tabla 5.8: Medidas de Tendencia Central y Valores Importantes de las innovaciones del Tipo de Cambio FIX, entre febrero de 2012 y febrero de 2016.

FIX), a fin de acercarnos a nuestros objetivos finales: estimar medidas de riesgo condicionales para el FIX.

5.1.3. Modelando las innovaciones del FIX

Una vez determinado el modelo para la dinámica del FIX, el siguiente paso de la metodología estudiada en el Capítulo 4 es investigar la naturaleza de las innovaciones del GARCH(1,1) ajustado a dicha serie. En consecuencia, nuestro objetivo en esta Sección será modelar de forma apropiada la distribución de tales innovaciones.

Al modelar la distribución de las innovaciones del FIX, centraremos nuestros esfuerzos especialmente en el área de las colas. Esto se debe a que nuestras estimaciones de medidas de riesgo para la serie del FIX (Var_α^t y ES_α^t) dependerán fuertemente de la cola del modelo ajustado a las innovaciones, por lo que un ajuste óptimo de nuestro modelo en dicha región de la distribución es de vital importancia para nuestros objetivos finales.

En consecuencia, en esta Sección haremos uso de las técnicas y herramientas estudiadas en el Capítulo 3. Específicamente, utilizaremos el método POT para ajustar un modelo Pareto generalizado a la cola derecha de las innovaciones del FIX, a fin de analizarla de forma especializada. Después, contrastaremos la calidad del ajuste obtenido a través dicha metodología con la que obtendríamos al ajustar por máxima verosimilitud un modelo Gaussiano o t -Student a dichos datos. Para concluir, compararemos las estimaciones que obtengamos del Var_α y ES_α que obtengamos a partir de los tres modelos para las innovaciones. De este modo, al final de esta Sección, estaremos en posibilidades de obtener estimaciones de medidas de riesgo condicionales para la serie del FIX.

No obstante, a fin de conocer los datos que modelaremos, primeramente exploraremos de forma más profunda las innovaciones del FIX, a través de un análisis de estadística descriptiva. En la Tabla 5.8 podemos observar una serie de medidas descriptivas de la innovaciones del FIX. Los valores contenidos en

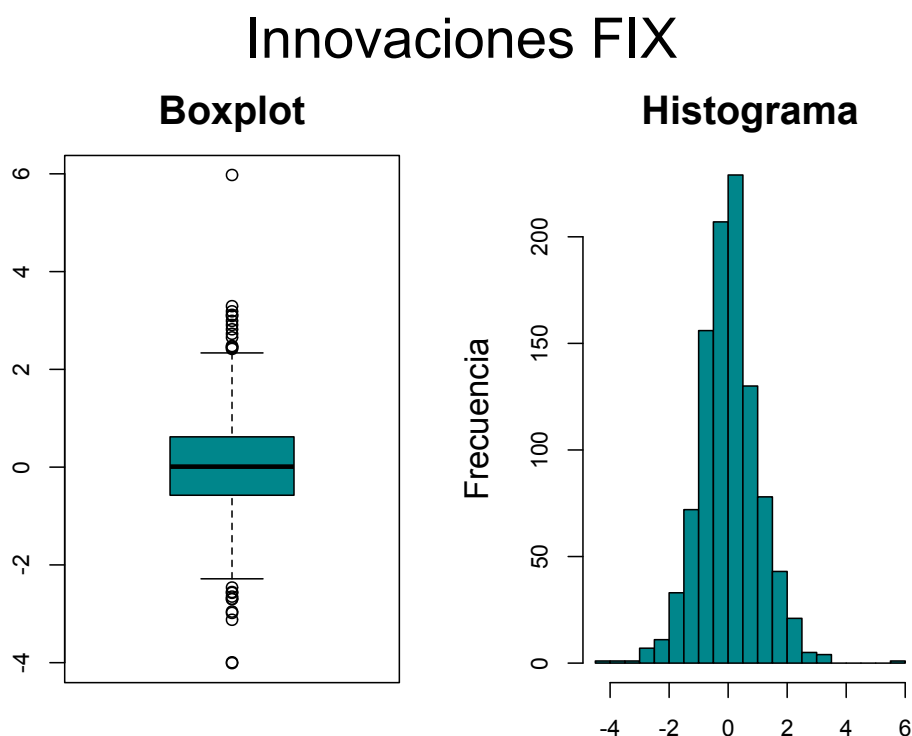


Figura 5.11: Histograma y *boxplot* de los residuos estandarizados del Tipo de Cambio FIX.

la Tabla 5.8 nos dicen que la distribución de las innovaciones podría tener una cola derecha más larga que la izquierda (la innovación más grande es mayor, en valor absoluto, que la innovación más pequeña) y estar centrada en cero (su media y mediana están muy cercanas a dicho valor). Adicionalmente, podemos comprobar que la varianza de dichos datos es muy cercana a 1, lo cual está en línea con nuestra hipótesis de RBE. Finalmente, el valor de la curtosis de exceso nos señala que las colas de estas innovaciones podrían ser más pesadas que las de una distribución normal, lo que nos da la primera pieza de evidencia de que tal distribución podría no constituir un modelo óptimo para las innovaciones del FIX.

Ahora bien, para visualizar la forma de la distribución de las innovaciones del FIX, en la Figura 5.11 podemos observar el histograma y *boxplot* de estos residuales. Dicha Figura nos confirma de forma gráfica que las innovaciones del FIX están centradas en cero y que su distribución podría tener una cola derecha más larga que la izquierda. Adicionalmente, el *boxplot* nos revela la presencia de una cantidad considerable de datos de grandes en las innovaciones del FIX; lo cual puede apreciarse al notar la cantidad de datos graficados

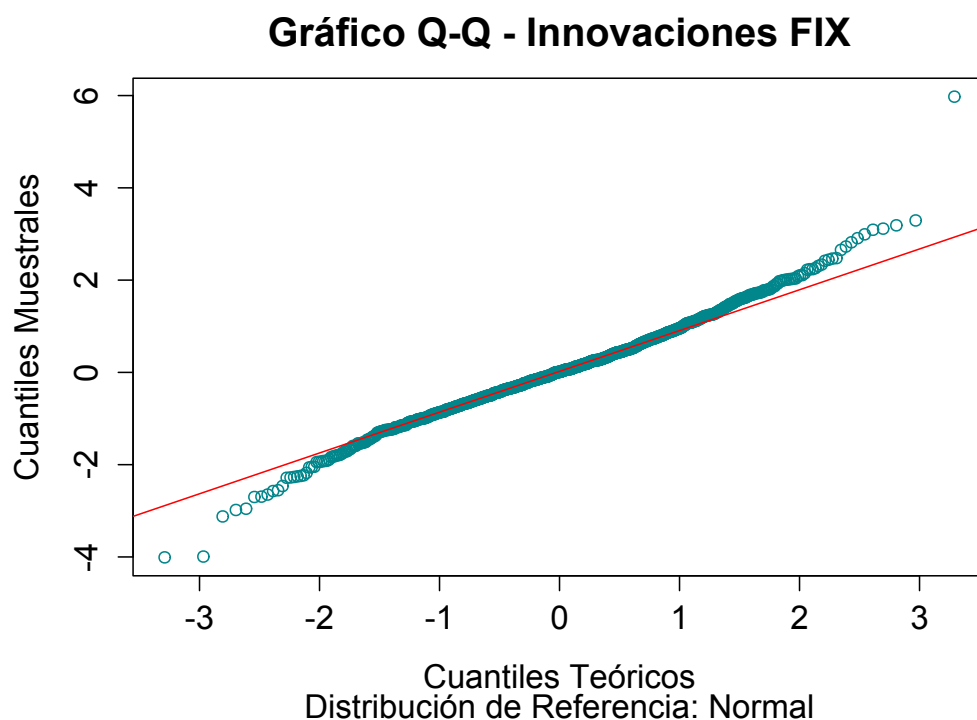


Figura 5.12: Gráfico QQ de los residuos estandarizados del Tipo de Cambio FIX, contra una distribución de referencia normal. Los parámetros de la distribución normal fueron estimados vía máxima verosimilitud.

por encima de los brazos del *boxplot*. No obstante, dada la mayor cantidad de datos por encima del brazo superior, es probable que la cola derecha de las innovaciones sea más pesada que la izquierda.

Para investigar de forma más profunda lo pesado de las colas de las innovaciones del FIX, en la Figura 5.12 podemos observar un QQ-Plot de dichas observaciones contra una distribución de referencia Gaussiana. La Figura 5.12 nos permite ver que las colas de las innovaciones del FIX son, en efecto, más pesadas que las de una distribución normal. No obstante, dichas colas no parecen ser mucho más pesadas que las de la distribución normal: la forma de “S” volteada no es tan pronunciada como en el QQ-Plot del FIX original.

Para concluir nuestro análisis descriptivo y ganar más información sobre lo pesado de la cola (derecha) de las innovaciones del FIX, en la Figura 5.13 podemos observar la gráfica de la función muestral media de exceso de dichos datos. La Figura 5.13, nos permite ver que existen umbrales a partir de los cuales la gráfica de la función muestral media de exceso presenta un

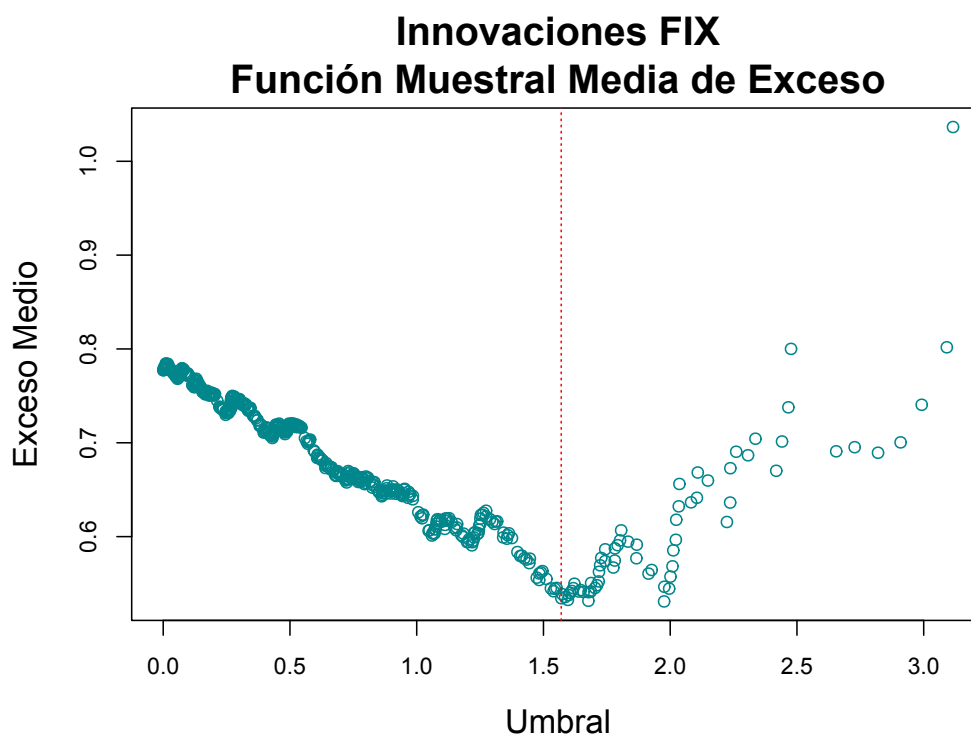


Figura 5.13: Gráfica de la función muestral media de exceso de los residuos estandarizados del Tipo de Cambio FIX; la línea roja vertical muestra el umbral $u = 1.57$ a partir de cual el gráfico presenta un comportamiento lineal con pendiente positiva.

comportamiento aproximadamente lineal con pendiente positiva; por ejemplo $u = 1.57$ o $u = 1.97$. No obstante, antes del umbral $u = 1.57$ la gráfica de la función muestral media presenta un claro comportamiento lineal con pendiente negativa, lo que podría indicar, en combinación con el comportamiento para $u > 1.57$, que una tendencia con pendiente aproximadamente igual a cero también podría resultar plausible para describir el comportamiento general de dicha gráfica.

Recapitulando, nuestro análisis descriptivo nos mostró diversas características de las innovaciones del FIX: primeramente, su distribución está centrada en cero y tiene una varianza igual a 1, lo cual está en línea con nuestras hipótesis sobre las innovaciones de un modelo GARCH. Además, la cola derecha de la distribución de las innovaciones parece ser más larga que la izquierda. No obstante, lo pesado de las colas de tal distribución aún no es claro. Hasta el momento, solo sabemos (gracias a la curtosis de exceso y el QQ-Plot) que dichas colas son más pesadas que las de una distribución normal. Por lo

Ajuste DPG a las Innovaciones del FIX

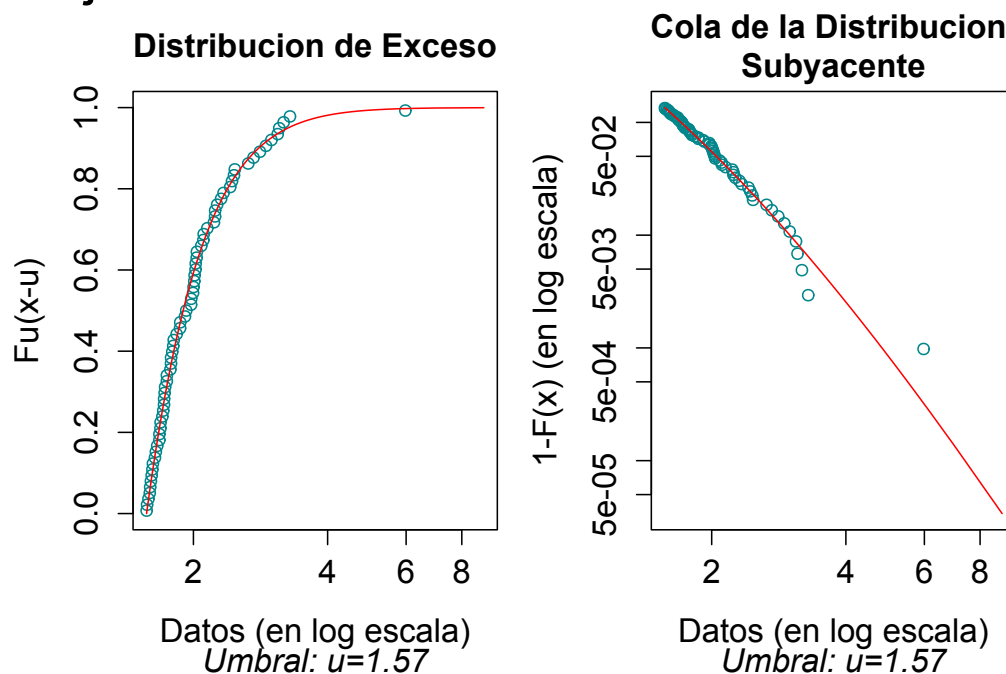


Figura 5.14: Ajuste de una Distribución Pareto Generalizada a la cola derecha de las innovaciones del FIX, para el umbral $u = 1.57$. Se muestran las curvas de la distribución de exceso (izquierda) y la cola (derecha) del modelo DPG para las innovaciones del FIX con las versiones empíricas de dichas curvas sobreimpuestas de forma respectiva. El valor estimado, vía máxima verosimilitud, del parámetro de forma ξ tomó el valor 0.1515.

tanto, requeriremos de mayor información para determinar si el $DAM(H_\xi)$ al que pertenecen las innovaciones del FIX es aquel con $\xi > 0$ o bien aquel con $\xi = 0$.

En el Capítulo 3 estudiamos que el primer paso para aplicar el método POT a un conjunto de datos es la elección de un umbral a partir del cual consideremos que comienza la cola de su distribución. Además, estudiamos que la gráfica de la función muestral media de exceso es una buena herramienta para seleccionar dicho umbral. La Figura 5.13 nos mostró que para las innovaciones del FIX dicha gráfica presenta un comportamiento aproximadamente lineal con pendiente positiva a partir del umbral $u = 1.57$. Por lo tanto, nuestro ajuste Pareto generalizado inicial a las innovaciones del FIX será realizado con base en dicho umbral.

Bajo la consideración anterior, nuestro ajuste DPG inicial se realizará con base en 69 excesos sobre el umbral. Con base en dichos datos, obtenemos que los estimadores de máxima verosimilitud de los parámetros de forma y escala son, respectivamente, $\hat{\xi} = 0.1515$ y $\hat{\beta} = 0.4465$. El valor de $\hat{\xi}$ sugiere que las innovaciones del FIX tienen una cola derecha pesada.

En la Figura 5.14 podemos observar, sobreimpuestas de forma respectiva, las curvas de la distribución de exceso y la cola de nuestro modelo DPG inicial y las versiones empíricas de dichas curvas. Así, la Figura 5.14 nos permite evaluar de forma gráfica la calidad del ajuste del modelo DPG inicial a la cola derecha de las innovaciones del FIX. De este modo, al analizarla, la Figura 5.14 nos permite concluir que el ajuste de la DPG a la cola de las innovaciones del FIX es muy satisfactorio.

No obstante, como mencionamos en el Capítulo 3, al aplicar el método POT es recomendable realizar análisis con base en más de un umbral. Por lo tanto, a continuación evaluaremos la calidad de ajuste de un modelo DPG a las innovaciones del FIX considerando un umbral $u = 1.97$. Elegimos este umbral pues a partir de él la gráfica de la función muestral media de exceso de las innovaciones del FIX (ver Figura 5.13) también presenta un comportamiento lineal con pendiente positiva.

Con este nuevo umbral nuestro ajuste DPG se realiza con base en 34 excesos. Los estimadores de los parámetros de la DPG con base en dichos excesos toman los valores $\hat{\xi} = 0.3467$ y $\hat{\beta} = 0.3497$. Nuevamente, la estimación puntual de ξ sugiere un modelo de colas pesadas para las innovaciones del FIX.

En la Figura 5.15 se puede observar la calidad de ajuste de la DPG a los excesos de las innovaciones del FIX al umbral $u = 1.97$. Dicha Figura nos permite apreciar que en este caso también se logra una buena calidad de ajuste del modelo DPG a la cola de las innovaciones del FIX; a pesar de que dicho ajuste se realizó con base en menos excesos que aquel para el umbral $u = 1.57$.

Los dos modelos anteriores tuvieron una muy buena calidad de ajuste a la cola de las innovaciones del FIX y ambos sugirieron un modelo de colas pesadas para tales datos; las estimaciones puntuales del parámetro de forma ξ resultaron en ambos casos mayores que cero. No obstante, dado que el parámetro crucial para la estimación de altos cuantiles y déficits esperados utilizando el modelo DPG es justamente el parámetro de forma, a continuación investigaremos qué tan estable es dicho parámetro al variar el umbral sobre el cual se basa el ajuste DPG a las innovaciones del FIX, así como el

Ajuste DPG a las Innovaciones del FIX

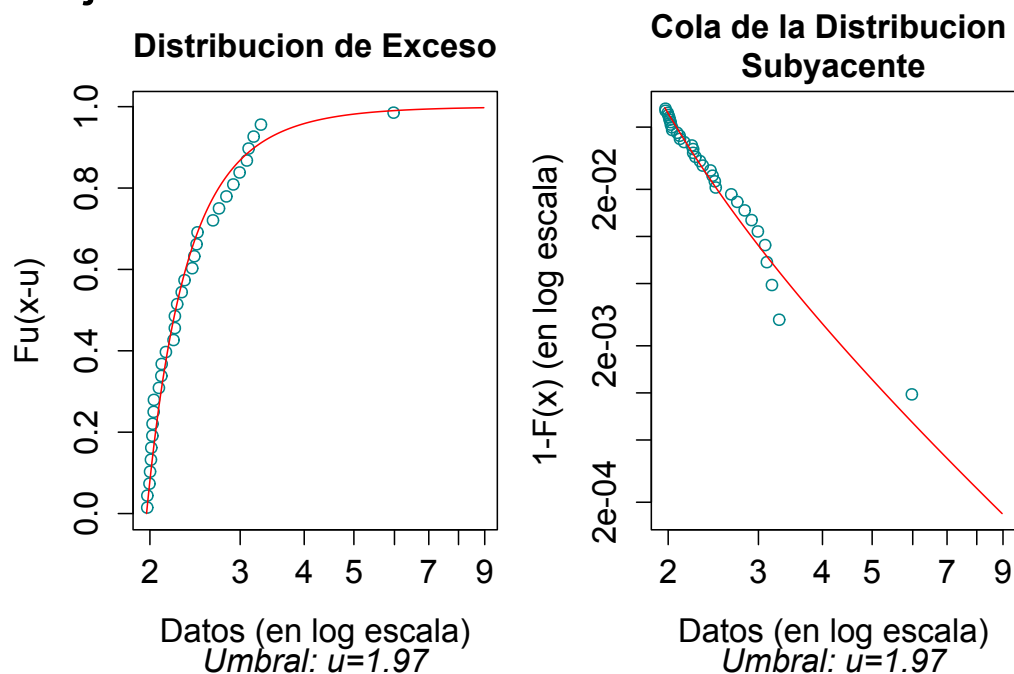


Figura 5.15: Ajuste de una Distribución Pareto Generalizada a la cola derecha de las innovaciones del FIX, para el umbral $u = 1.97$. Se muestran las curvas de la distribución de exceso (izquierda) y la cola (derecha) del modelo DPG para las innovaciones del FIX con las versiones empíricas de dichas curvas sobreimpuestas de forma respectiva. El valor estimado, vía máxima verosimilitud, del parámetro de forma ξ tomó el valor 0.3467.

hecho de si dicho parámetro puede ser considerado estadísticamente mayor que cero.

Así, en la Figura 5.16 se ajustan modelos DPG con distintos umbrales para obtener estimadores de ξ , así como intervalos de confianza asintóticos para los estimadores del parámetro. En el eje x inferior se marca el número de excesos al umbral; en el eje x superior se marca el umbral. El parámetro de forma ξ y sus intervalos de confianza se grafican en el eje y . En total se ajustan 30 modelos.

La Figura 5.16 nos permite ver que los valores estimados de ξ incrementan conforme aumenta el umbral sobre el que basamos el ajuste de la DPG a las innovaciones del FIX. De hecho, se puede apreciar que el parámetro de forma sufre un cambio de signo (de negativo a positivo) a partir de cierto umbral.

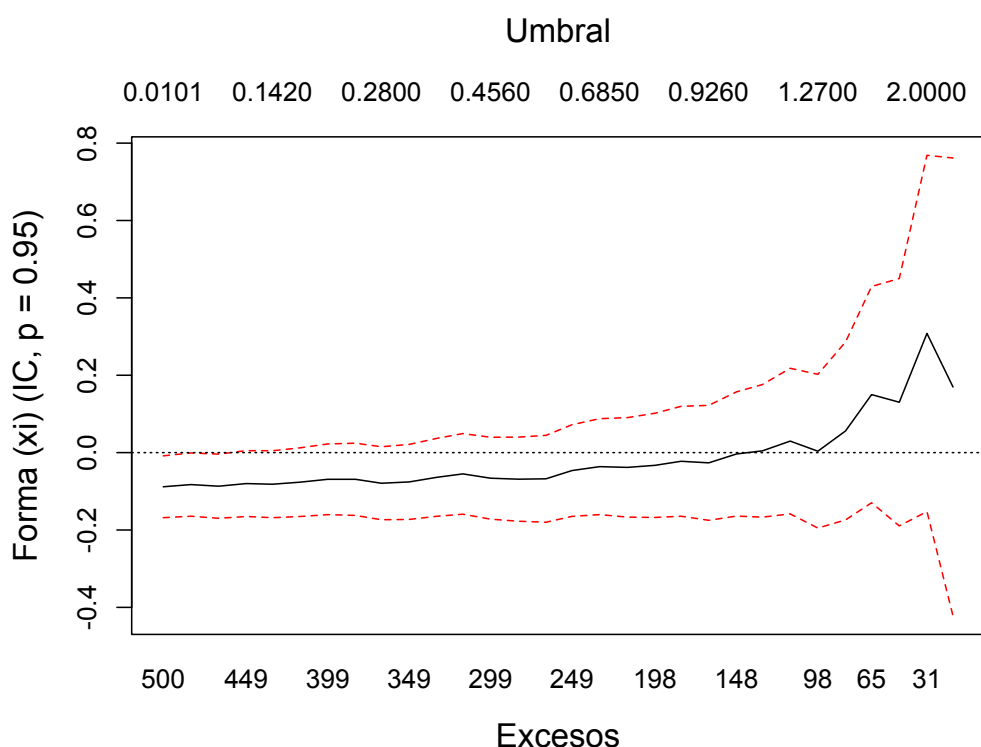


Figura 5.16: Parámetro del forma ξ de la Distribución Pareto Generalizada ajustada a la cola derecha de las innovaciones del FIX al variar el umbral en el que se basa el ajuste. En total, se ajustan 30 modelos distintos.

Adicionalmente, podemos apreciar que el valor cero siempre se encuentra contenido dentro de los intervalos de confianza al 95% del parámetro ξ .

Los hechos anteriores nos indican que con una confianza de 95%, podemos considerar que la distribución de las innovaciones del FIX pertenece al $DAM(H_\xi)$, donde $\xi = 0$. No obstante, esto no quiere decir que las colas de las innovaciones del FIX sean ligeras; de hecho, hemos comprobado que éstas son más pesadas que las de una distribución normal. Por el contrario, recordemos que en Capítulo 3 discutimos que el $DAM(H_0)$ contiene una gran gama de distribuciones y que algunas de ellas presentan colas cuyo comportamiento se asemeja de forma importante al de distribuciones de colas pesadas; ejemplos de estas distribuciones son la log-normal o las distribuciones hiperbólicas generalizadas. Por lo tanto, este hecho nos indica que la distribución de las innovaciones del FIX tiene la característica de presentar un comportamiento de colas pesadas, sin que el comportamiento de éstas pueda necesariamente ser explicado en términos de funciones de variación regular.

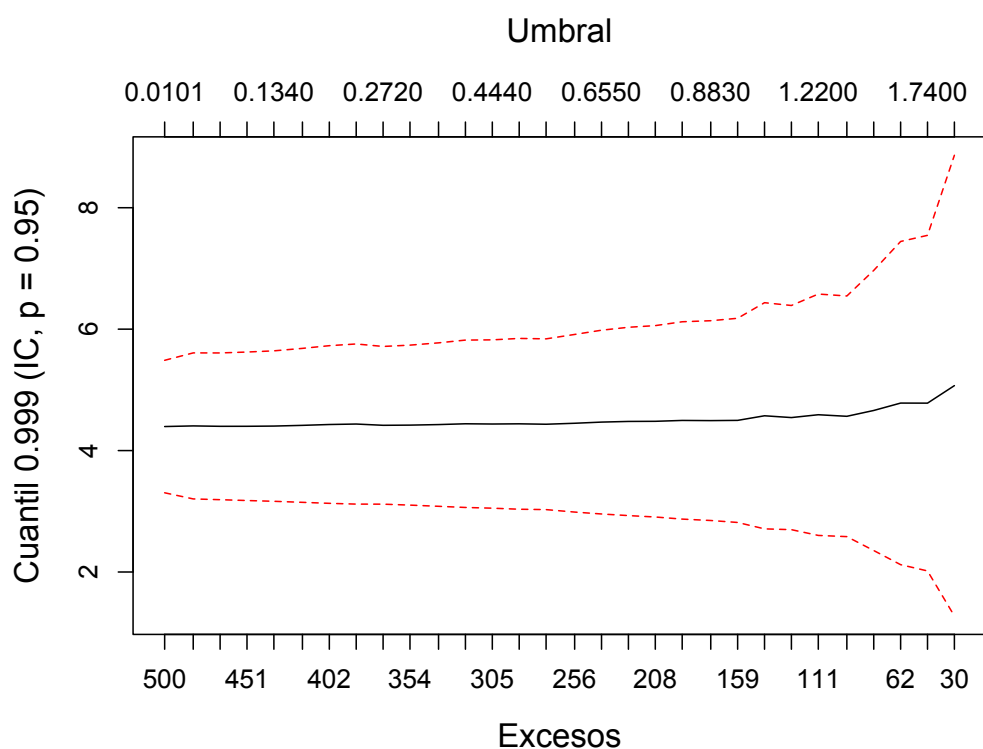


Figura 5.17: Variación del cuantil 0.999 al variar el umbral en el que se basa el ajuste de la Distribución Pareto Generalizada a la cola derecha de las innovaciones del FIX. En total, se ajustan 30 modelos distintos.

Ahora bien, al utilizar la Figura 5.16 para elegir un umbral “óptimo” nos enfrentamos a un intercambio entre sesgo y varianza. Puesto que el ajuste del modelo DPG se basa en un resultado límite para umbrales altos, si escogemos un umbral demasiado bajo nuestros estimadores podrían estar sesgados porque dicho resultado puede no ser válido. Por otro lado, si elegimos un umbral demasiado alto, basaríamos nuestro ajuste en muy pocos datos y nuestras estimaciones estarían sujetas a altos errores estándar. Así, una elección sensible de u a partir de la gráfica de la Figura 5.16 podría encontrarse al inicio de la última cuarta parte de dicha gráfica.

Como guía adicional para seleccionar un umbral para basar el ajuste de un modelo DPG a la cola de una distribución, podemos analizar qué tan sensibles son las estimaciones del VaR ante cambios en tal parámetro. En la Figura 5.17 se grafican los estimadores del cuantil 0.999 para distintos umbrales; dicha Figura nos permite observar que los estimadores de los cuantiles dependen de la elección del umbral. Si el modelo es bueno, este cuantil debería ser excedido por aproximadamente una innovación de cada mil. Por lo tanto, en

Innovaciones del FIX: Estimación del VaR_{0.999}

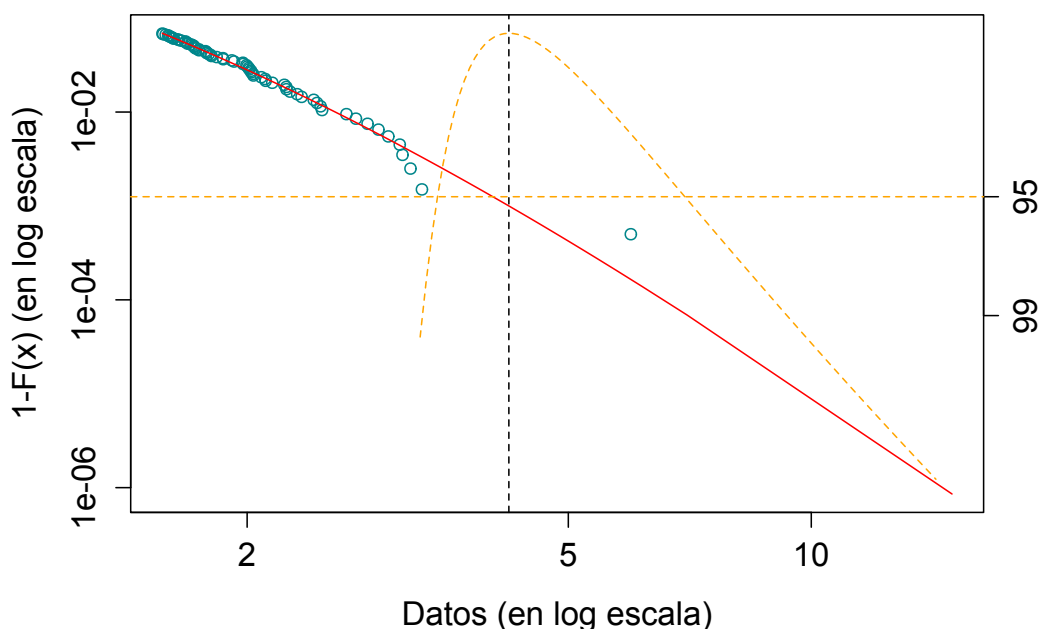


Figura 5.18: Estimación del $VaR_{0.999}$ para las innovaciones del FIX a través de un modelo DPG. El umbral en el que se basó el ajuste fue $u = 1.57$ y el valor estimado del parámetro de forma de la DPG fue de $\xi = 0.1515$.

la serie de las innovaciones del FIX es posible que solo hayamos visto una innovación de dicha magnitud, de manera que este es un problema difícil de estimación de cuantiles.

Al analizar la Figura 5.17 podemos observar que las estimaciones del $VaR_{0.999}$ se comportan de forma bastante estable al variar el umbral, si bien estas aumentan ligeramente en la última parte de la gráfica debido a que el parámetro ξ también es mayor para dichos umbrales. Sin embargo, dichas diferencias no parecen ser significativas para umbrales menores a 2.

De este modo, y tomando en cuenta que el comportamiento de la distribución del FIX se asemeja al de una distribución con colas pesadas, podríamos basar nuestro análisis en el modelo con $u = 1.57$. Así, obtenemos que el estimador del $VaR_{0.999}$ para las innovaciones del FIX es 4.22. Esto se muestra en la Figura 5.18. La línea punteada vertical se intersecta con el estimador de la cola en el punto $(4.22, 0.001)$, lo que nos permite leer el estimador del VaR en el eje de las x . La curva naranja en forma de parábola es una herramienta que

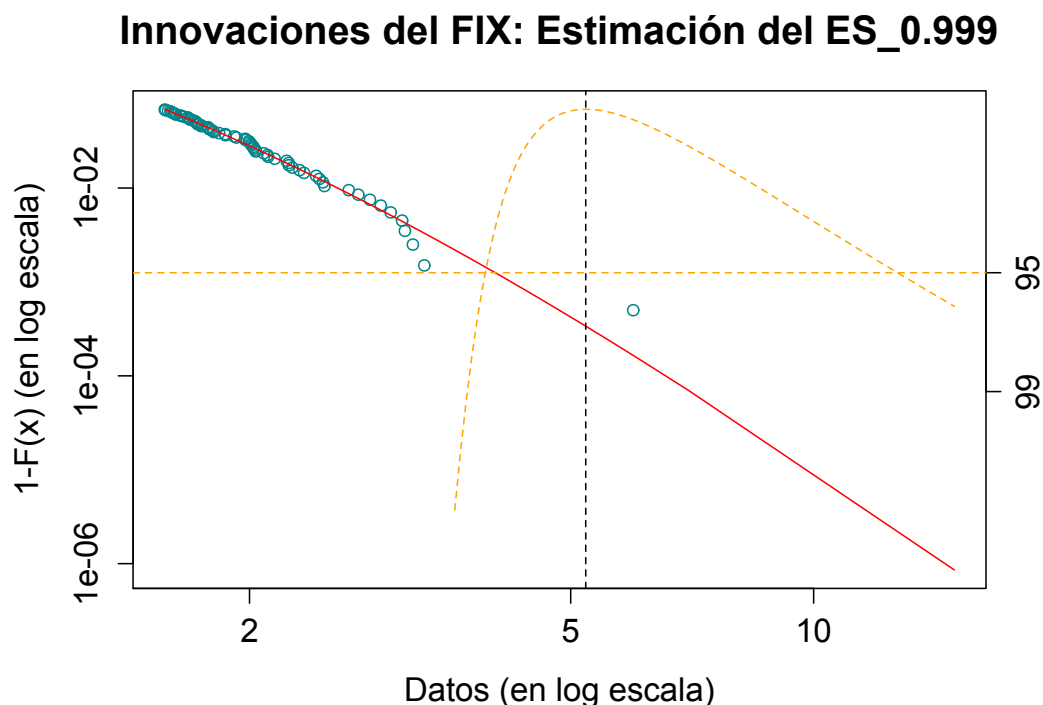


Figura 5.19: Estimación del $ES_{0.999}$ para las innovaciones del FIX a través de un modelo DPG. El umbral en el que se basó el ajuste fue $u = 1.57$ y el valor estimado del parámetro de forma de la DPG fue de $\xi = 0.1515$.

nos permite calcular intervalos de confianza para el VaR (mediante el método de verosimilitud de perfiles). El eje y del lado derecho de la gráfica es una escala de confianza (no una escala de cuantiles). La línea naranja horizontal corresponde a una confianza del 95%; las coordenadas en x de los dos puntos donde la curva en forma de parábola se intersecta con la línea horizontal son los extremos del intervalo de confianza del 95% para el VaR . En nuestro caso dichas coordenadas son: $(3.4974, 6.9222)$. Notemos que podemos obtener un intervalo de confianza del 99% más amplio al bajar la línea horizontal al valor 99 del eje de confianza.

Una ventaja de la visualización del $VaR_{0.999}$ y su intervalo de confianza al 95% que nos ofrece la Figura 5.18 es que nos permite observar de forma clara que dicho intervalo es asimétrico. Este hecho refuerza que las colas de las innovaciones del FIX no son ligeras: acotar los intervalos de confianza de sus cuantiles por abajo es más fácil que hacerlo por arriba. Esto nos da también una idea de la dificultad que conlleva estimar altos cuantiles para datos con comportamiento de colas pesadas: los intervalos asimétricos reflejan

Innovaciones FIX: parámetros		
Modelo	Parámetros	
Normal	$\mu = 0.0367$	$\sigma = 0.9998$
DPG	$\xi = 0.1515$	$\beta = 0.4465$
<i>t</i> -Student	$\nu = 6.9818$	

Tabla 5.9: Estimaciones puntuales de los parámetros de los modelos ajustados a las innovaciones del modelo GARCH(1, 1) ajustado a la serie del Tipo de Cambio FIX.

que se debe tener mayor cautela al tratar de predecir el comportamiento de las observaciones extremas en este tipo de distribuciones.

Finalmente, podemos tener una guía sencilla de la calidad de nuestra estimación del $VaR_{0.999}$ al fijarnos en lo siguiente. Dada la longitud de la serie de las innovaciones del FIX, esperaríamos haber observado solo una pérdida que exceda al $VaR_{0.999}$. En la Figura 5.18 podemos observar que nuestra estimación del $VaR_{0.999}$ para la serie de las innovaciones del FIX solo fue excedida una vez, lo que nos dice que dicha estimación está en línea con lo observado en nuestros datos.

Calculemos ahora estimadores del *Expected Shortfall* para las innovaciones del FIX. Estimando dicha medida mediante nuestro modelo DPG ajustado con base en un umbral $u = 1.57$, tenemos que el estimador puntual de $ES_{0.999}$ es 5.2193. Dicha estimación, así como su intervalo de confianza al 95 % pueden observarse en la Figura 5.19.

La visualización de obtenida en Figura 5.19 nos permite notar que el intervalo de confianza para el $ES_{0.999}$ también es asimétrico; dicho intervalo es incluso más amplio que aquel del $VaR_{0.999}$. Los extremos del intervalo de confianza del $ES_{0.999}$ resultaron: (3.9342, 12.6594).

Con las Figuras 5.18 y 5.19 hemos ejemplificado la estimación del VaR_{α} y ES_{α} a partir de un modelo DPG, con lo que podemos concluir nuestro análisis POT de las innovaciones del FIX. Dicho análisis, en general, entregó muy buenos resultados.

No obstante, a continuación compararemos los resultados obtenidos a partir de nuestro análisis POT con los que obtendríamos al utilizar otra clase de enfoque: ajustar, mediante máxima verosimilitud, una distribución conocida a las innovaciones del FIX; específicamente, las distribuciones Normal y *t*-Student.

En consecuencia, en la Tabla 5.9 podemos observar los valores estimados

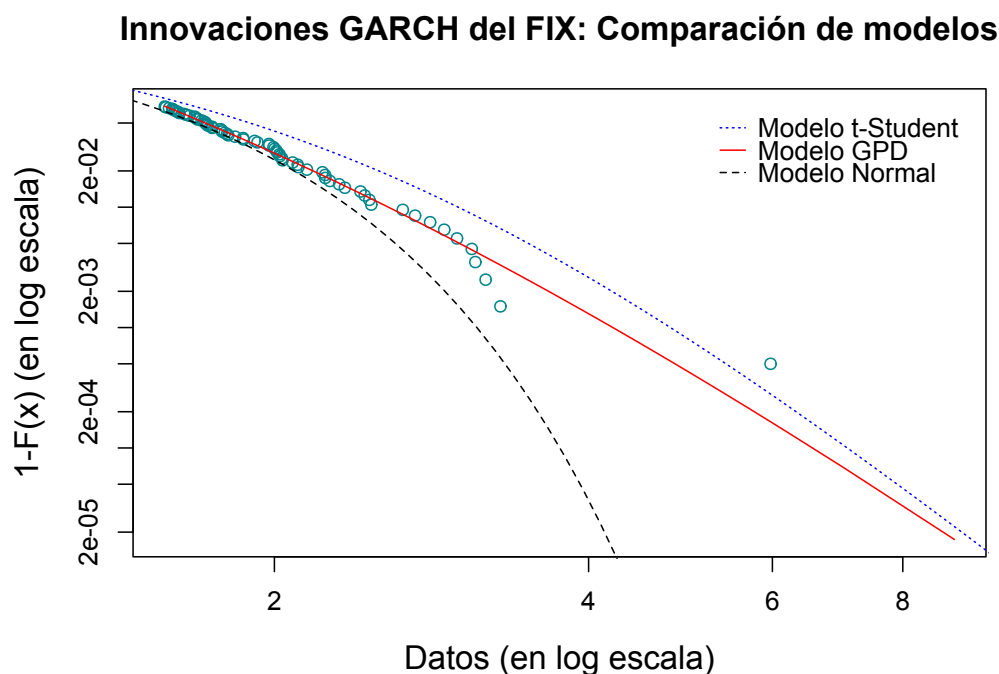


Figura 5.20: Comparación gráfica de la calidad de ajuste de diversos modelos a la cola derecha de la distribución de las innovaciones del FIX. Los modelos ajustados son: normal (línea de guiones negra), t -Student (línea azul punteada) y DPG (línea roja sólida).

de los parámetros de las distribuciones Normal, t -Student y nuestro modelo DPG con base en el umbral $u = 1.57$. Al analizar los parámetros de cada modelo podemos confirmar diversas características de las innovaciones del FIX. Primeramente, el parámetro ν del modelo t -Student no está cercano al valor crítico 2, por lo que podemos inferir que la cola de estos datos no es tan pesada como para que la varianza de éstos se vuelva infinita. Adicionalmente, los parámetros del modelo normal nos reafirman que las innovaciones del FIX están, efectivamente, centradas en cero y que el valor de su varianza es muy cercano a 1.

Ahora bien, en la Figura 5.20 podemos ver un comparativo gráfico de la calidad de ajuste de los modelos Normal, t -Student y DPG a la cola (derecha) de las innovaciones del FIX. Dicha Figura nos permite apreciar que:

1. El ajuste de la distribución normal es muy pobre para altos cuantiles de la distribución de las innovaciones del FIX: para cuantiles superiores a

Innovaciones FIX: Medidas de Riesgo								
α	0.95		0.99		0.995		0.999	
Modelo	VaR_α	ES_α	VaR_α	ES_α	VaR_α	ES_α	VaR_α	ES_α
Normal	1.681	2.099	2.363	2.702	2.612	2.929	3.126	3.404
t -Student	1.895	2.597	3.000	3.774	3.503	4.327	4.792	5.773
DPG	1.717	2.270	2.572	3.277	3.009	3.792	4.220	5.219

Tabla 5.10: Estimaciones puntuales del VaR_α y ES_α obtenidas a partir de los modelos ajustados a las innovaciones de FIX.

- 0.95 se aprecia que este modelo subestima de forma importante la cola derecha de las innovaciones.
- El modelo t -Student ofrece cierta mejoría en comparación con el modelo normal; la distancia entre la cola de este modelo y la de la distribución empírica de las innovaciones del FIX es menor que aquella con el modelo Normal. Sin embargo, y en contraste con el caso anterior, el modelo t -Student sobrestima claramente los cuantiles superiores a 0.95 de las innovaciones del FIX.
 - El modelo DPG es el que parece entregar los mejores resultados en la comparación. Dicho modelo presenta un ajuste bastante satisfactorio a lo largo de toda la cola de la distribución empírica de las innovaciones del FIX: no parece haber evidencia de que este modelo tienda a subestimar o sobrestimar cuantiles a lo largo de toda la cola de la distribución de las innovaciones.

De este modo, el análisis de la Figura 5.20 señala de que el método POT de la TVE parece constituir una mejor alternativa para modelar la cola de las innovaciones del FIX que el ajuste por máxima verosimilitud de modelos Normal o t -Student. Adicionalmente, dicho análisis nos muestra que el modelo GARCH(1,1) con innovaciones DPG tiene, en general, un ajuste bastante satisfactorio a la serie de log-rendimiento del FIX.

Procederemos entonces a analizar medidas de riesgo para las innovaciones del FIX. En la Tabla 5.10 podemos observar un comparativo de las estimaciones del VaR *Expected Shortfall* obtenidas a partir de los tres modelos para las innovaciones. Específicamente, para cada modelo se muestran el VaR_α y el ES_α para $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$. Dichas medidas de riesgo fueron calculadas mediante las fórmulas derivadas en los Capítulos 1 y 3.

Al analizar la Tabla 5.10 podemos apreciar que las medidas de riesgo estimadas a partir de los modelos Normal, t -Student y DPG guardan una relación

de monotonía: en todos los casos, las estimaciones obtenidas a través del modelo Normal tienen los valores más pequeños, les siguen aquellas del modelo DPG y, finalmente, las estimaciones del modelo t -Student tienen los valores más grandes. Esto refleja el hecho de que estos modelos estiman colas con diferentes pesos para las innovaciones del FIX, siendo la Gaussiana la más ligera y la t -Student la más pesada.

Adicionalmente, notemos que las estimaciones DPG tanto del Var_α como del ES_α son más parecidas a las del modelo normal para umbrales relativamente bajos ($\alpha \leq 0.99$) mientras que para umbrales altos éstas se asemejan más a aquellas entregadas por el modelo t -Student. Este hecho nos dice nuevamente que la cola derecha de la distribución de las innovaciones del FIX tiene un comportamiento similar al de distribuciones con colas pesadas para altos cuantiles, a pesar de que se tenga una confianza del 95% de que ésta se encuentre en el $DAM(H_0)$. No obstante, más adelante, investigaremos qué modelo es el que estimó de forma más precisa la verdadera naturaleza de las colas de las innovaciones del FIX.

Para avanzar hacia nuestro objetivo final, procederemos a estimar medidas de riesgo condicionales, Var_α^t y ES_α^t , para la serie del FIX. En este punto contamos con todos los elementos necesarios para ello: estimaciones de la volatilidad del FIX un paso adelante y de las medidas de riesgo de los modelos para las innovaciones de dicha serie.

De este modo, podemos utilizar las fórmulas en (5.1) para estimar el Var_α^t y ES_α^t para el FIX a partir del modelo GARCH(1,1) ajustado a dicha serie y los tres modelos considerados anteriormente para sus innovaciones. Así, en la Tabla 5.11 podemos observar las estimaciones de dichas medidas de riesgo, para $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$ y $t = 12/02/2016$. Notemos que las medidas de riesgo contenidas en la Tabla 5.11 son un escalamiento de aquellas contenidas en la Tabla 5.10, por lo que entre ellas se guarda la misma relación de monotonía que entre las medidas de riesgo para las innovaciones.

Con lo anterior hemos concluido nuestra estimación inicial de medidas de riesgo condicionales para la serie del Tipo de Cambio FIX. Adicionalmente, los análisis realizados nos han brindado cierta intuición sobre qué modelo podría entregar las estimaciones con un mejor desempeño en la práctica: el modelo GARCH-DPG. No obstante, para concluir la aplicación de la metodología estudiada en el Capítulo 4 en la serie del FIX, en la siguiente Sección validaremos tal intuición: evaluaremos de forma histórica las estimaciones condicionales de medidas de riesgo entregadas por los modelos utilizados en esta Sección para modelar la serie del FIX.

Tipo de Cambio FIX: Medidas de Riesgo condicionales								
α	0.95		0.99		0.995		0.999	
Modelo	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t	VaR_α^t	ES_α^t
Normal	0.0164	0.0205	0.0231	0.0264	0.0256	0.0286	0.0306	0.0333
t -Student	0.0185	0.0254	0.0293	0.0369	0.0343	0.0423	0.0469	0.0565
DPG	0.0165	0.0226	0.0264	0.0325	0.0306	0.0367	0.0405	0.0466

Tabla 5.11: Estimaciones puntuales del VaR_α^t y ES_α^t obtenidas a partir del modelo GARCH(1,1) ajustado a la serie del Tipo de Cambio FIX al variar el modelo ajustado a la cola de la distribución de las innovaciones.

5.1.4. *Backtesting*

En esta última Sección aplicaremos una metodología de *backtesting* a nuestras estimaciones tanto del VaR_α^t como del ES_α^t para la serie del FIX. El objetivo central de estos análisis será evaluar el comportamiento histórico de las estimaciones de dichas medidas de riesgo y así comparar el desempeño de los tres modelos estudiados en la Sección anterior: GARCH-GPG, GARCH-Normal y GARCH- t -Student.

Para realizar los *backtests* de esta Sección seguiremos de cerca las metodologías estudiadas en los Capítulos 2 y 4.³⁵ No obstante, nuestro análisis inicial de la serie del FIX mostró que pueden existir ciertas dificultades prácticas al ajustar un modelo GARCH a una serie de tiempo financiera. Por lo tanto, debido a que un *backtest* consiste en el ajuste de dicha clase de modelos de forma automatizada para una posterior estimación de medidas de riesgo, la metodología utilizada en el ajuste de tales modelos a lo largo de un *backtest* debe considerar la potencial aparición de ese tipo problemas; especialmente si el *backtests* a realizar es largo.

Existen al menos dos escenarios típicos que podrían afectar la aplicación directa de las metodologías de *backtest* estudiadas en Capítulos anteriores:

1. El parámetro α_0 del modelo ajustado a la serie no es estadísticamente significativo. Lo cual se traduce en que la prueba t de significancia estadística aplicada a la estimación de dicho parámetro entrega un p -valor mayor a nuestro nivel de significancia.
2. El proceso GARCH ajustado a la serie no es estacionario en covarianza. Lo cual se refleja en un factor $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j$ mayor que 1.

³⁵ Recordemos que la única diferencia entre dichas metodologías es la forma en que se estiman el VaR_α y el ES_α de la distribución de las innovaciones.

Estos dos escenarios afectan nuestra estimación del VaR_α^t o el ES_α^t , pues impiden una correcta estimación de la volatilidad futura de una serie de tiempo con base en un modelo GARCH. En ambos casos, el problema reside en que la fórmula de pronóstico de volatilidad con base en el ajuste de un GARCH fue derivada bajo las hipótesis de que el parámetro α_0 de dicho modelo es positivo y de que el proceso resultante es estacionario covarianza. Por lo tanto, al no satisfacerse cualquiera de tales hipótesis, la fórmula dada en (5.2) no resulta válida.

Así, a fin de realizar *backtests* que tomen en consideración estos dos escenarios, debemos realizar alguna modificación a las metodologías de los Capítulos anteriores. En consecuencia, en los análisis siguientes se propone utilizar la metodología ilustrada en la Sección 5.1.2 de este trabajo, cuando en nuestro análisis inicial de la serie del FIX se presentó el primero de los escenarios mencionados anteriormente.

De esta manera, la propuesta puede aplicarse cuando el modelo seleccionado para modelar la dinámica de la serie de tiempo estudiada es un GARCH(1,1) y consiste en modificar la forma en que se estima la volatilidad de la serie un paso adelante, en los días en que alguno de estos dos escenarios se presente. Específicamente, utilizamos un suavizamiento exponencial de volatilidad (EWMA) y el parámetro α_1 del GARCH(1,1) ajustado a la serie para pronosticar la volatilidad de ésta un paso adelante.

En consecuencia, dado que en los análisis de las secciones anteriores determinamos que un proceso GARCH(1,1) captura de forma parsimoniosa la dinámica del FIX, la metodología de *backtest* del VaR_α^t y ES_α^t que utilizaremos a continuación será igual a la estudiada en los Capítulos 2 y 4, con la salvedad de que el cuadrado de la volatilidad un paso adelante de la serie de tiempo estudiada será estimado como:

$$\hat{\sigma}_{t+1}^2 \begin{cases} \hat{\alpha}_1 X_t^2 + (1 - \hat{\alpha}_1) \hat{\sigma}_t^2, & \text{si } \hat{\alpha}_1 + \hat{\beta}_1 > 1 \text{ ó } \mathcal{P}(\alpha_0) > 0.05, \\ \hat{\alpha}_0 + \hat{\alpha}_1 X_t^2 + \hat{\beta}_1 \hat{\sigma}_t^2, & \text{en otro caso,} \end{cases} \quad (5.4)$$

donde $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1$ son los valores estimados de los parámetros del modelo GARCH(1,1) ajustado a la serie estudiada en el día t y $\mathcal{P}(\alpha_0)$ es el p-valor de la prueba t de significancia estadística del parámetro α_0 de dicho modelo.

Adicionalmente, para guardar consistencia con dichos análisis iniciales, los *backtests* del VaR_α^t y ES_α^t para el FIX se realizarán con base en una ventana temporal móvil de $n = 1,000$ días hábiles bancarios. Esto quiere decir que el conjunto de días sobre el que realizaremos nuestros *backtests* abarcará del 29 de diciembre de 1999 al 12 de febrero de 2016; la información de los días entre

el 2 de enero de 1996 y el 28 de diciembre de 1999 se utilizará para realizar el ajuste del primer modelo GARCH(1,1). De esta forma, los *backtests* del FIX llevarán a cabo sobre un conjunto de 4,060 datos y cada estimación del VaR_α^t y ES_α^t se realizara con aproximadamente 4 años de información diaria. Cabe mencionar que la longitud de este periodo se extendió tanto como fue posible debido a que se realizaremos evaluaciones sobre estimaciones de altos cuantiles y déficits esperados.³⁶

De esta manera, para realizar el *backtest* del VaR_α^t en la serie del FIX primeramente debemos calcular las versiones muestrales de las indicadores de violación (o exceso) al VaR_α^t :

$$I_{\{X_{t+1} > VaR_\alpha^t\}} = I_{\{Z_{t+1} > q_\alpha(Z)\}} \sim Ber(1 - \alpha). \quad (5.5)$$

Esto se logra al sustituir las estimaciones correspondientes del VaR_α^t o $q_\alpha(Z)$ en dichas fórmulas. En nuestro caso, haremos esto para cada t en nuestra ventana de estudio y $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$. Adicionalmente, para lograr el contraste que deseamos, realizaremos este procedimiento para cada uno de los modelos propuestos para las innovaciones del FIX en secciones anteriores: Normal, t -Student y DPG.

De este modo, una vez obtenida, en cada caso, la versión muestral de las indicadores de violación del VaR_α^t , el siguiente paso será realizar un *test* binomial a cada una de las muestras obtenidas.³⁷ Dicha prueba evaluará, para cada modelo seleccionado para las innovaciones, si la suma de las indicadores de violación al VaR_α^t se comporta como una variable aleatoria $Binomial(4, 034, 1 - \alpha)$. En caso afirmativo, ganaremos certeza de que el VaR_α^t ha sido correctamente estimado, pues querrá decir las indicadores de cada muestra se comportan como variables aleatorias $Ber(\alpha)$; la cual es la hipótesis sobre la que fueron construidas.

Por otro lado, para realizar el *backtest* del ES_α^t debemos obtener las versiones muestrales de los residuales de exceso definidos en los Capítulos 2 y 4, a saber:

$$R_{t+1} = \frac{X_{t+1} - ES_\alpha^t}{\sigma_{t+1}} = Z_{t+1} - ES_\alpha(Z), \quad (5.6)$$

los cuales, dado el evento $\{Z_{t+1} > VaR_\alpha\}$, deberían tener media cero.

³⁶ La cantidad de datos es relevante para realizar *backtests* a altos niveles de confianza pues, por ejemplo, en el caso del $VaR_{0.995}^t$ esperamos tener, en promedio, $0.005 \times 4,060 \approx 20$ violaciones a dicho cuantil. Esto se traduce en tener, aproximadamente, solo 20 datos para realizar las pruebas correspondientes sobre las indicadores de violación al $VaR_{0.995}^t$.

³⁷ Para detalles sobre el *test* binomial, véase el Apéndice B.

Dichas versiones muestrales, al igual que las indicadores de exceso al VaR_α^t , son obtenidas al sustituir las estimaciones correspondientes de ES_α^t o $ES_\alpha(Z)$ en la expresión anterior. Este procedimiento se llevará a cabo para cada t en nuestra ventana de estudio, $\alpha \in \{0.95, 0.99, 0.995\}$ y para cada uno de modelos elegidos para las innovaciones del GARCH(1, 1) ajustado al Tipo de Cambio FIX.

Después, para considerar de forma práctica solo aquellos residuos asociados al evento $\{Z_{t+1} > VaR_\alpha\}$, se seleccionarán sólo aquellos de los días en que la observación X_{t+1} del FIX resulte mayor que el correspondiente VaR_α^t . Posteriormente, a dichos residuales seleccionados se les aplicará una prueba *bootstrap* para verificar la hipótesis de que la media de dichas sub-muestras sea, en efecto, estadísticamente igual a cero.³⁸

Ahora bien, para los modelos GARCH-Normal y GARCH- t -Student lo anterior puede ser logrado de forma sencilla, siguiendo de forma automatizada los pasos de ajuste de modelo ilustrados en las secciones previas de este análisis. Sin embargo, para el modelo GARCH-DPG tenemos un problema.

En efecto, hasta este momento hemos utilizado la metodología estudiada en el Capítulo 3 para ajustar un modelo DPG a un conjunto de datos: analizar la gráfica de la función muestral media de exceso de los datos, con el objetivo de encontrar un valor del umbral, u , a partir del cual comenzar el ajuste del modelo DPG a la cola de la distribución de tales observaciones. No obstante, en un proceso automatizado, como el requerido para aplicar un método de *backtests*, resultaría imposible detenernos en cada ocasión a analizar la gráfica de la función muestral media de exceso para determinar candidatos para el valor inicial de u . Por lo tanto, en este punto necesitamos implementar una metodología que nos permita encontrar valores razonables de u para nuestros *backtests*.

Estudio de simulación para la elección de un umbral

Para seleccionar un umbral razonable a utilizar en nuestros *backtests*, realizaremos un *estudio de simulación*. Analizando una vez más la Figura 5.20, podemos ver que la distribución t -Student provee al menos una aproximación a la distribución observada de las innovaciones del FIX. La idea de nuestro estudio de simulación consiste en aprovechar este hecho.

³⁸ Para detalles sobre la prueba *bootstrap* de media cero aplicada en el *backtest* del ES_α^t , véase el Apéndice B.

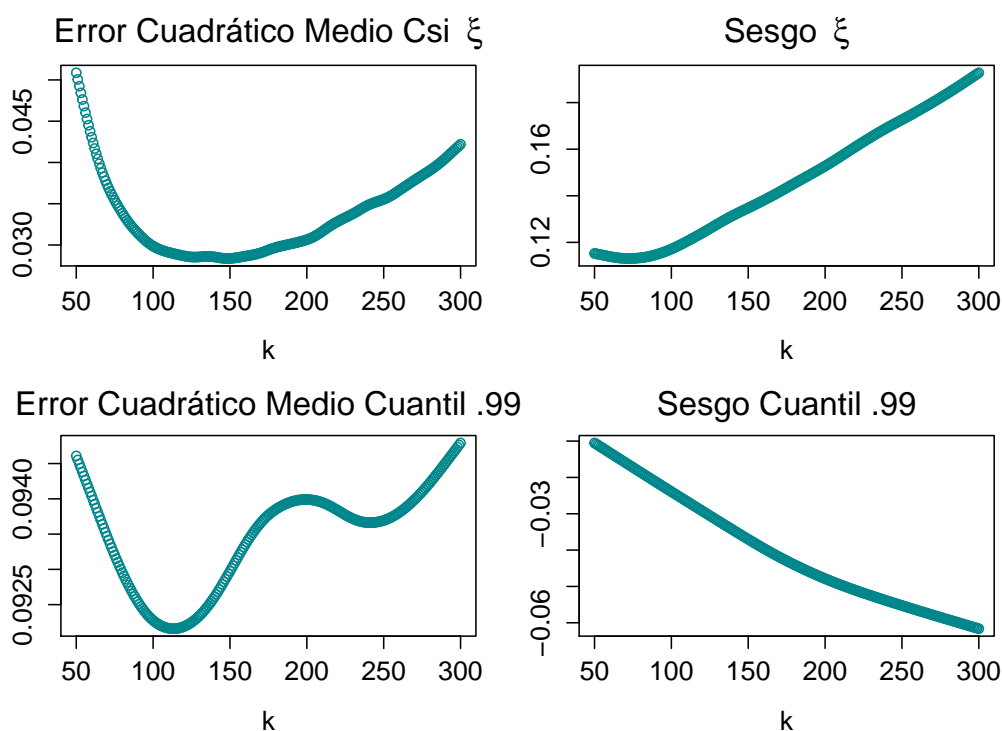


Figura 5.21: Resultados del estudio de simulación. Se muestran las gráficas del sesgo y error cuadrático medio estimados del parámetro ξ (arriba) y cuantil 0.99 (abajo) de una DPG, al variar el número de observaciones en la cola de la distribución, k .

Específicamente, simularemos muestras de tamaño 1,000 (para ser consistentes con la ventana temporal móvil de nuestros *backtests*) de la distribución t -Student e investigaremos qué valor de u reduce el sesgo y el error cuadrático medio de las estimaciones, vía máxima verosimilitud, del parámetro de forma ξ y del cuantil 0.99 de la DPG.³⁹ Esto es posible gracias a la relación entre el parámetro ξ de la DPG y el parámetro ν de la distribución t -Student, a saber: $\nu = 1/\xi$

Así, para cada estimador estimamos el error cuadrático medio y el sesgo utilizando estimaciones Monte Carlo basadas en 1,000 muestras independientes.

³⁹ Dos características deseables de un estimador es que sea insesgado y con error cuadrático medio mínimo, por lo que minimizar estas dos cantidades nos da certeza de que nuestra metodología brindará buenas estimaciones de parámetros o cantidades estadísticas (medidas de riesgo).

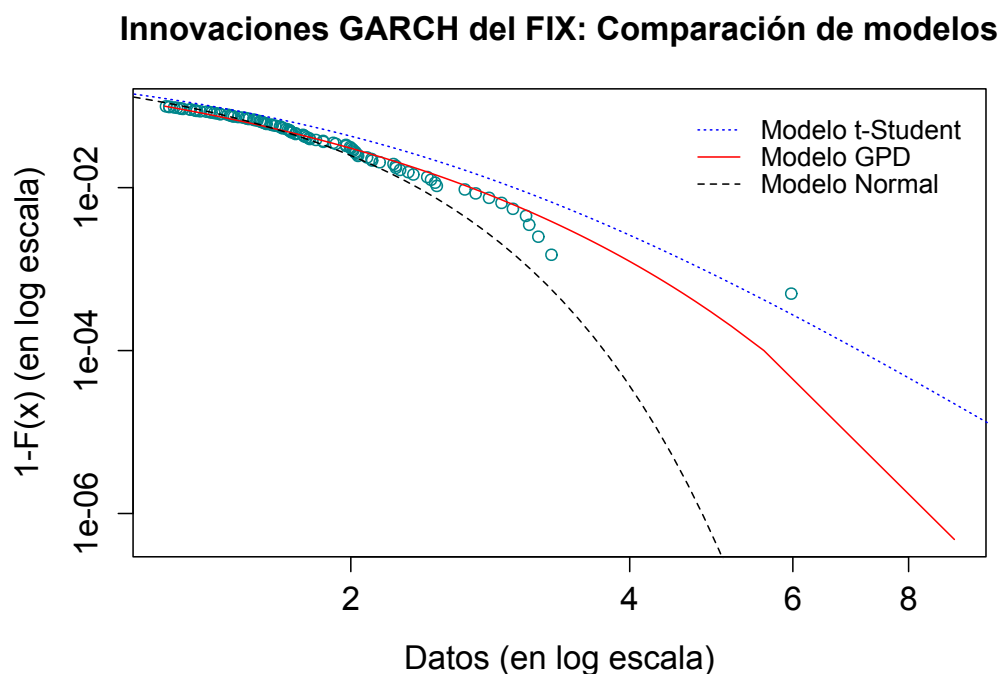


Figura 5.22: Resultados del estudio de simulación. Se muestra una comparación de modelos para la cola derecha de la distribución de la innovaciones del GARCH(1, 1) ajustado al FIX. Los modelos ajustados son: normal (línea de guiones negra), t -Student (línea azul punteada) y DPG (línea roja sólida). La DPG fue ajustada utilizando las 100 innovaciones más grandes como excesos al umbral.

De este modo, por ejemplo, el $ECM(\hat{\xi})$ es estimado como:

$$\widehat{ECM}(\hat{\xi}) = \sum_{j=1}^{1000} (\hat{\xi}^{(j)} - \xi)^2,$$

donde $\hat{\xi}^{(j)}$ representa la estimación de ξ obtenida de la j -ésima muestra.

Ahora bien, es claro que una forma alternativa de determinar el valor de u (hasta este momento hemos tomado u con ayuda de la gráfica de función media de exceso) consiste en determinar el número de observaciones que deseemos considerar como pertenecientes al área de la cola. Es decir, la cantidad de datos de nuestra muestra que pensemos que, debido a su magnitud, se encuentran en el área de la cola de la distribución y no en la parte central.

Bajo este enfoque, para una muestra cualquiera podemos determinar u como

el valor de la $(k + 1)$ -ésima estadística de orden de dicha muestra;⁴⁰ de forma que, en efecto, en cada ocasión tengamos k observaciones por encima de u . De este modo, por cuestiones de implementación numérica, en este estudio determinaremos el valor de u de dicha forma; la cual también fue sugerida en el Capítulo 4 de este trabajo.

En nuestro estudio de simulación utilizamos muestras de una distribución t -Student con $\nu = 4$ grados de libertad.⁴¹ Los resultados del estudio se pueden observar en la Figura 5.21. Dicha Figura nos permite ver que un valor de $k = 100$ parece minimizar el sesgo y error cuadrático medio de los estimadores de ξ y del cuantil 0.99 de la DPG.

A fin de ganar certeza sobre los resultados entregados por nuestro estudio de simulación, utilizaremos dicho valor para realizar el ajuste de una DPG a las innovaciones del FIX calculadas en la Sección 5.1.2. De esta manera, podremos comparar la calidad de ajuste obtenida con la metodología de nuestro estudio de simulación contra aquella lograda cuando al analizar la gráfica de la función muestral media de exceso.

En la Figura 5.22 se muestra un comparativo gráfico de la calidad de ajuste de una DPG ajustada con la metodología derivada de nuestro estudio de simulación, contra el ajuste, vía máxima verosimilitud, de un modelo normal y un modelo t -Student. En dicha Figura, se puede apreciar que la calidad del ajuste de la DPG utilizando la metodología derivada de nuestro estudio de simulación es muy satisfactoria; de hecho, dicho ajuste tiene una calidad comparable con la lograda anteriormente, cuando se analizó detenidamente la gráfica de la función media de exceso (véase la Figura 5.20).

Así, en vista de los resultados favorables de nuestro estudio de simulación, utilizaremos la metodología sugerida por dicho estudio para realizar los *back-test* del FIX: fijaremos el valor del umbral a partir de cual realizaremos el ajuste de una DPG a las innovaciones del FIX, en cada ocasión, como aquel de la (101) -ésima estadística de orden de cada muestra de innovaciones de los modelos GARCH(1,1) ajustados al FIX. De este modo, en cada ocasión consideraremos que la cola de la distribución las innovaciones del FIX está constituida por el 10% de sus observaciones más grandes.

⁴⁰ Aquí, estamos pensando que para una muestra aleatoria Z_1, Z_2, \dots, Z_n sus estadísticas de orden están dadas por $Z_{1,n} \geq Z_{2,n} \geq \dots \geq Z_{n,n}$.

⁴¹ Es importante mencionar que, aunque en este trabajo estamos presentando resultados para una distribución t -Student con $\nu = 4$, más simulaciones fueron realizadas en el desarrollo del estudio presentado. Dichas simulaciones sugieren que el mismo resultado se sostiene para otros valores de ν .

Tipo de Cambio FIX: <i>Backtest</i> del VaR_{α}^t				
α	0.95		0.99	
Modelo	Viol.Hist(<i>203</i>)	P-valor	Viol.Hist(<i>41</i>)	P-valor
GARCH-Normal	217	0.3132	77	3.25×10^{-07}
GARCH- <i>t</i> -Student	169	0.0143	34	0.3432
GARCH-DPG	205	0.8854	37	0.6359
α	0.995		0.999	
Modelo	Viol.Hist(<i>20</i>)	P-valor	Viol.Hist(<i>4</i>)	P-valor
GARCH-Normal	45	1.77×10^{-06}	19	6.30×10^{-08}
GARCH- <i>t</i> -Student	15	0.2665	4	1
GARCH-DPG	19	0.9111	3	0.8046

Tabla 5.12: Resultados del *backtest* de las estimaciones del VaR_{α}^t . Se muestran, por modelo ajustado a las innovaciones del GARCH, los p-valores de los *tests* binomiales aplicados a las muestras de indicadores de violación al VaR_{α}^t , las violaciones históricas a las estimaciones del VaR_{α}^t y, entre paréntesis, las violaciones esperadas al VaR_{α}^t por nivel de confianza.

Una vez resuelto el problema de elección de umbral para las innovaciones del FIX, podemos continuar y realizar nuestras evaluaciones históricas del VaR_{α}^t y ES_{α}^t .

Resultados del *backtest* del VaR_{α}^t

En la Tabla 5.12 podemos ver los resultados de los *tests* binomiales aplicados a las indicadores de violación al VaR_{α}^t de la serie del FIX. Por lo tanto, los valores contenidos en dicha Tabla nos permiten determinar si el comportamiento de las estimaciones del VaR_{α}^t obtenidas a partir de los diferentes modelos estudiados fue o no satisfactorio. Así, al analizar la Tabla 5.12 podemos observar que:

1. Modelo GARCH-Normal:

- Las estimaciones obtenidas a partir de modelos GARCH con innovaciones Gaussianas tuvieron un desempeño adecuado para niveles de confianza relativamente bajos, $\alpha = 0.95$. Esto se deduce del siguiente hecho: las violaciones a las estimaciones del $VaR_{0.95}^t$ obtenidas a partir de este modelo fueron aproximadamente iguales a las esperadas.
- Para niveles de confianza superiores a 0.95, las estimaciones del

VaR_{α}^t obtenidas a partir de este modelo subestimaron considerablemente el verdadero valor de dicha medida de riesgo; lo cual se refleja en que para $\alpha > 0.95$ el número de violaciones históricas a dichas estimaciones fue considerablemente mayor al número de violaciones esperadas.

- Por lo tanto, la hipótesis nula del *test* binomial aplicado a las muestras de las indicadores de violación al VaR_{α}^t fue rechazada para cada $\alpha > 0.95$.
- En consecuencia, podemos concluir que el modelo GARCH-Normal subestimó de forma importante lo pesado de la cola derecha de la distribución condicional de pérdida del FIX; lo cual vuelve dicho modelo una herramienta muy pobre para medir el riesgo acarreado por este tipo de cambio.
- De este modo, podemos concluir que, en la práctica, el uso de un modelo GARCH-Normal para medir el riesgo de mercado del FIX llevaría a una institución a esperar un número de pérdidas considerablemente menor al que podría materializarse para una serie de la naturaleza del FIX.

2. Modelo GARCH-*t*-Student:

- Las estimaciones del VaR_{α}^t obtenidas a partir de modelos GARCH con innovaciones *t*-Student presentaron un desempeño pobre para niveles de confianza por debajo de 0.99.
- Específicamente, para $\alpha = 0.95$ dichas estimaciones sobrestimaron de forma importante los verdaderos valores del VaR_{α}^t de esta serie; lo cual se refleja en que la cantidad de violaciones al $VaR_{0.95}^t$ estimado con el modelo GARCH-*t*-Student fue considerablemente menor a la esperada.
- No obstante, para niveles de confianza $\alpha = 0.99$ y superiores, las estimaciones obtenidas a partir de este modelo resultaron consistentes con la información histórica del FIX; no parece existir evidencia de que este modelo tienda a subestimar o sobrestimar el verdadero valor del VaR_{α}^t para dichos niveles de confianza.
- Por lo tanto, podemos concluir que el modelo GARCH-*t*-Student brinda solo una solución parcial para medir el riesgo de mercado acarreado por el FIX; dicho modelo no logró estimar de forma apropiada toda la cola de la distribución condicional de los log-rendimientos de dicha serie de tiempo.

- En consecuencia, el modelo GARCH- t -Student tendría una utilidad limitada en la práctica. Sólo sería recomendable utilizarlo al trabajar con niveles de confianza superiores a 0.95. Para α 's menores, el uso de dicho modelo podría llevar a una institución financiera a reservar más capital del necesario para administrar el riesgo de mercado del FIX.

3. Modelo GARCH-DPG:

- Las estimaciones del VaR_{α}^t obtenidas a partir de modelos GARCH con innovaciones DPG tuvieron un desempeño en línea con la información histórica del Tipo de Cambio FIX, para todos los niveles de confianza considerados.
- El hecho anterior se refleja en que la hipótesis nula de los *tests* binomiales aplicados a las indicadores de violación al VaR_{α}^t no fue rechazada para ninguno de los niveles de confianza considerados. En otras palabras, la cantidad de violaciones históricas a las estimaciones del VaR_{α}^t obtenidas a través del modelo GARCH-DPG fue aproximadamente igual a las esperadas para la serie del FIX, para todos los niveles de confianza considerados.
- En consecuencia, podemos concluir que no parece existir evidencia de que este modelo tienda a subestimar o sobrestimar lo pesado de toda la cola derecha de la distribución condicional de los rendimientos del FIX.
- Por lo tanto, podemos concluir que este modelo ha logrado estimar satisfactoriamente los cuantiles condicionales de la serie del FIX, para todos los niveles de confianza relevantes en la administración de riesgo de mercado.
- En la práctica, este modelo permitiría a una institución financiera reservar justamente el capital necesario para administrar las pérdidas ocasionadas por movimientos de este tipo de cambio; lo cual lo constituye como una importante alternativa medir el riesgo de mercado del FIX.

Para ilustrar el comportamiento de las estimaciones condicionales del VaR , en la Figura 5.23 podemos ver cómo se comportan las estimaciones dicha medida de riesgo para la serie del FIX; para un nivel de confianza $\alpha = 0.99$ y desde el inicio de la crisis financiera del 2008 hasta el final de la ventana temporal de nuestro estudio. Específicamente, se muestra el periodo comprendido entre el 1 de julio de 2008 y el 12 de febrero de 2016.

De este modo, la Figura 5.23 nos permite observar las estimaciones del $VaR_{0.99}^t$ obtenidas a partir de los modelos GARCH-Normal y GARCH-DPG, así como los excesos a estas medidas de riesgo durante dicho periodo. Así, en dicha Figura podemos apreciar la adaptabilidad y dinamismo de las medidas de riesgo condicionales: reaccionan rápidamente para ajustarse a los niveles que va presentando la serie. Este comportamiento es el que vuelve ideales este tipo de estimaciones para medir el riesgo de mercado: permite a las instituciones financieras reservar más capital en periodos de mayor estrés en los mercados y reducirlo en tiempos de relativa calma.

Por otro lado, puesto que la Figura 5.23 abarca una ventana temporal de 1895 días, esperaríamos ver alrededor de 19 violaciones al $VaR_{0.99}^t$; en la realidad, las estimaciones obtenidas a partir del modelo con innovaciones Gaussianas fueron excedidas 40 veces, mientras que las del modelo GARCH-DPG presentaron 18 violaciones. Esto confirma nuevamente que el modelo GARCH-DPG entrega mejores estimaciones de medidas de riesgo condicionales que los modelos GARCH-Normal para la serie del FIX. Adicionalmente, nos permite apreciar las ventajas de incluir un toque de la Teoría de Valores Extremos en la administración de riesgos condicional: las estimaciones obtenidas a partir del modelo híbrido GARCH-DPG se comportan de forma excepcional aún en periodos que contienen tiempos de crisis (donde los rendimientos y pérdidas extremos ocurren).

Resultados del *backtest* del ES_{α}^t

En la Tabla 5.13 se pueden observar los resultados del *backtest* del ES_{α}^t del Tipo de Cambio FIX. Específicamente, se muestran los p-valores de las prueba *bootstrap* aplicada a las muestras de residuales de exceso, los días en que el correspondiente VaR_{α}^t fue excedido. Al analizar dicha Tabla podemos observar que:

1. Modelo GARCH-Normal:
 - Para todos los niveles de confianza considerados se rechazó la hipótesis de que las muestras de residuales del *backtest* tuvieran una media estadísticamente igual a cero.
 - El hecho anterior, en conjunto con los resultados del *backtest* del VaR , nos permiten concluir que el modelo GARCH-Normal subestimó de forma significativa el verdadero valor de ES_{α}^t .
 - En consecuencia, los resultados de este *backtest* confirman que

Tipo de Cambio FIX: Backtest del VaR

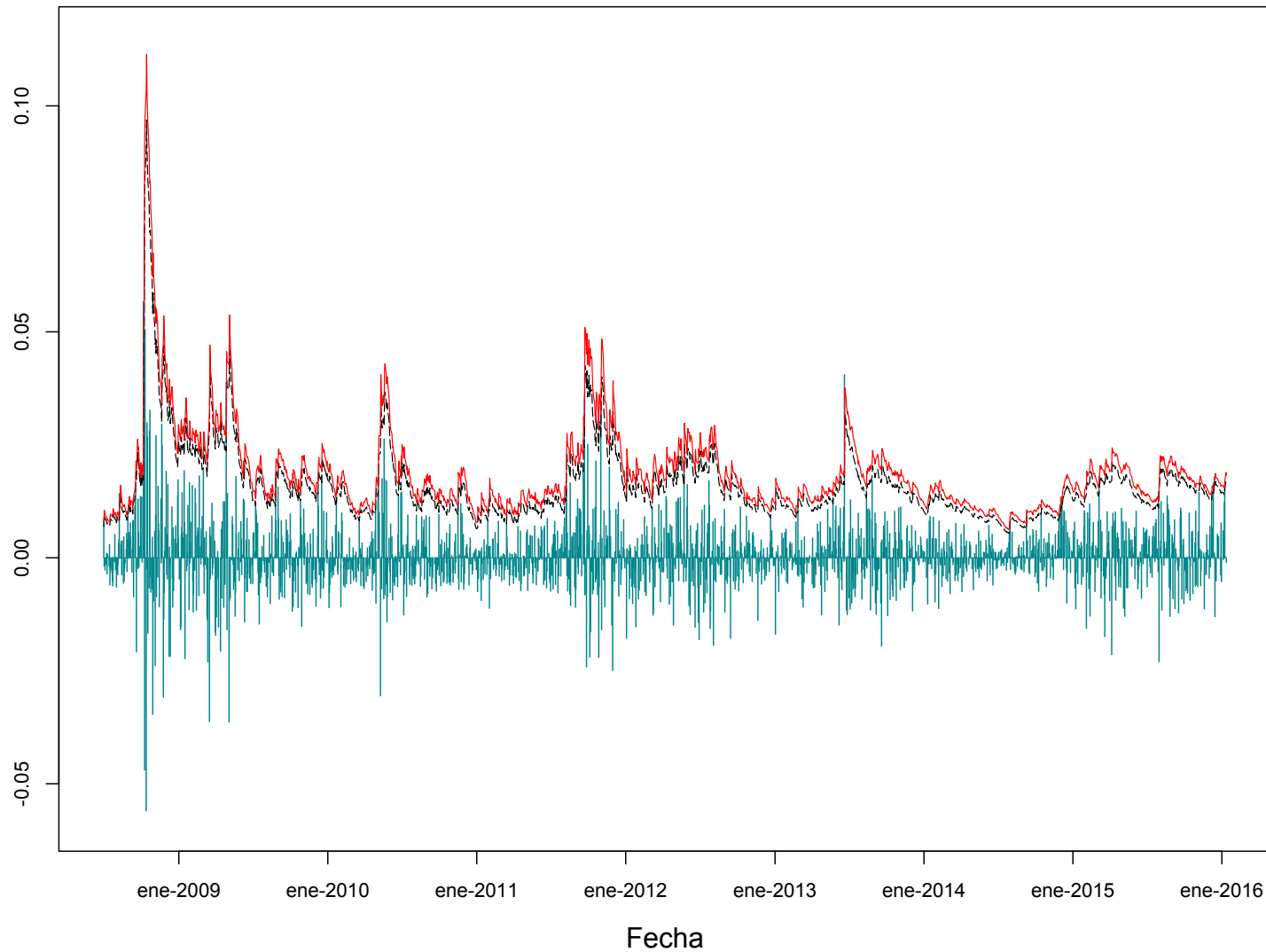


Figura 5.23: *Backtests* del $VaR_{0.99}^t$ desde la crisis financiera del 2008 hasta febrero de 2016. Se muestran las estimaciones obtenidas a partir de un modelo GARCH-Normal (línea de guiones negra) y GARCH-DPG (línea sólida roja). Las estimaciones obtenidas del modelo GARCH-Normal se encuentran por debajo de las del modelo GARCH-DPG y tienden a ser excedidas en más ocasiones debido a que no consideran la naturaleza leptocúrtica de los residuales del FIX.

Tipo de Cambio FIX: <i>Backtest</i> del ES_{α}^t			
α	0.95	0.99	0.995
Modelo	P-valor	P-valor	P-valor
GARCH-Normal	0.0048	0.0095	0.0429
GARCH- <i>t</i> -Student	0.5948	0.9959	0.6775
GARCH-DPG	0.5941	0.7734	0.8051

Tabla 5.13: Resultados del *backtest* de las estimaciones del ES_{α}^t . Se muestran los p-valores de las pruebas *bootstrap* aplicadas a las muestras de residuales de diferencia entre X_{t+1} y ES_{α}^t en los días que el VaR_{α}^t fue excedido.

la verdadera distribución condicional de las innovaciones del FIX tiene una cola derecha considerablemente más pesada que la de un modelo GARCH con innovaciones Gaussianas.

- De esta forma, para el caso específico $\alpha = 0.95$, el *backtest* del $ES_{0.95}^t$ nos revela que los días en que el $VaR_{0.95}^t$ sea excedido, las pérdidas en las que incurrirá el FIX serán considerablemente mayores, en promedio, que aquellas predichas por el modelo GARCH-Normal; a pesar de que este modelo entregó resultados razonables en la estimación del $VaR_{0.95}^t$,
- Por lo tanto, con base en los hechos anteriores podemos y los resultados del *backtest* del VaR_{α}^t , podemos concluir que el modelo GARCH-Normal ha resultado ser una herramienta muy pobre para estimar el riesgo de mercado acarreado por el Tipo de Cambio FIX.

2. Modelo GARCH-*t*-Student:

- Para ningún nivel de confianza se rechazó la hipótesis de que los residuos analizados en el *backtest* tengan media estadísticamente cero.
- No obstante, para $\alpha = 0.95$ se cuenta con el precedente de que este modelo sobrestima de forma importante el $VaR_{0.95}^t$, por lo que a pesar de haber obtenido un buen resultado en este *backtest*, el modelo GARCH-*t*-Student continua teniendo un desempeño pobre para un nivel de confianza $\alpha = 0.95$.
- El hecho anterior se desprende de que el desempeño del ES_{α}^t está fuertemente ligado al desempeño del VaR_{α}^t ; en este caso, al estar sobrestimado el $VaR_{0.95}^t$, con el $ES_{0.95}^t$ estaríamos calculando

pérdidas esperadas cuando excedemos un cantidad con una confianza mayor a 0.95.

- En la práctica, dicho fenómeno podría causar que una compañía reserve más capital para mitigar el riesgo de mercado acarreado por FIX del realmente necesario, generándole costos excesivos debido a una pobre medición del riesgo de mercado.

3. Modelo GARCH-DPG:

- Para todos los niveles de confianza considerados se obtuvieron resultados satisfactorios; en ningún caso se rechazó la hipótesis de que los residuos del *backtest* del ES_{α}^t tengan una media estadísticamente igual a cero.
- El hecho anterior, en conjunto con los resultados satisfactorios en el *backtest* del VaR_{α}^t de este modelo, brinda certeza de que el ES_{α}^t ha sido correctamente estimado por el modelo GARCH-DPG, para todos los niveles de confianza considerados.
- Por lo tanto, los resultados de este *backtest* brindan certeza adicional sobre el hecho de que el modelo GARCH-DPG ha logrado estimar de forma apropiada lo pesado de la cola derecha de la distribución condicional de los rendimientos del FIX.
- En consecuencia, podemos concluir que el modelo GARCH-DPG parece ideal para cuantificar el riesgo de mercado de esta serie de tiempo. Su uso en la práctica permitiría a una compañía, que trabaje con cualquier nivel relevante de tolerancia al riesgo, reservar el capital adecuado para administrar las pérdidas que pudiera acarrear este tipo de cambio.

Así, para ilustrar el comportamiento de los residuales utilizados en la evaluación del ES_{α}^t , en la Figura 5.24 podemos observar la gráfica de dichos residuos. Específicamente, la Figura 5.24 nos permite observar los residuos del *backtest* del $ES_{0.95}^t$ obtenidos a partir de los modelos GARCH-Normal y GARCH-DPG, desde la crisis financiera de 2008 y hasta febrero de 2016; nuestra ventana temporal en este caso comprendió nuevamente del 1 de julio de 2008 al 16 de febrero de 2016.

La Figura 5.24 nos apreciar de forma gráfica que, en efecto, la hipótesis de media cero es más razonable para los residuos del modelo GARCH-DPG que para aquellos del modelo GARCH-Normal: los residuales positivos del modelo GARCH-Normal fueron, en cada ocasión, de mayor magnitud que los del modelo GARCH-DPG y muestran evidencia en contra de la hipótesis de

Tipo de Cambio FIX: Backtest del ES

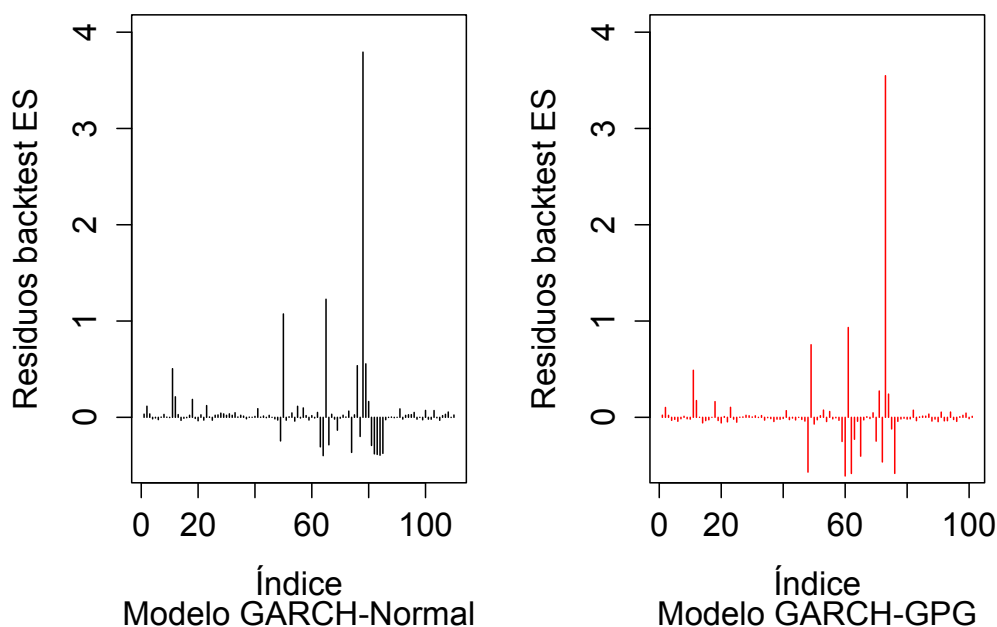


Figura 5.24: *Backtests* del $ES_{0.95}^t$ desde la crisis financiera del 2008 hasta febrero de 2016. Se muestran los residuales del *backtest* del $ES_{0.95}^t$ de los modelos GARCH-Normal (lado izquierdo, líneas negras) y GARCH-DPG (lado derecho, líneas rojas).

media cero. Adicionalmente, existe una mayor cantidad de residuos con signo negativo para el modelo GARCH-DPG que para el GARCH-Normal; en este caso, los residuos del modelo GARCH-DPG tienen una mayor magnitud que los del modelo GARCH-Normal.

Como dato adicional, el promedio de los residuos del modelo GARCH-DPG en el periodo seleccionado fue 0.0205, mientras que el de los residuos del modelo GARCH-Normal ascendió a 0.0523. Esto nos indica que, durante dicho periodo, los residuales del modelo GARCH-Normal fueron, en promedio, 255% más grandes que los del modelo GARCH-DPG.

De este modo, podemos observar de nueva cuenta la potencia que la Teoría de Valores Extremos es capaz de otorgar a nuestras estimaciones condicionales de medidas de riesgo: el modelo híbrido GARCH-DPG fue también capaz de entregar estimaciones excepcionales de ES_{α}^t , apegadas a lo observado en el mercado, aun en periodos que contienen tiempos de crisis.

Con lo anterior, hemos concluido el análisis del ES_α^t para la serie del FIX. La conclusión de dicho análisis es que, en efecto, la metodología estudiada en el Capítulo 4 de este trabajo también entrega mejores estimaciones de ES_α^t para el FIX que los modelos GARCH-Normal o GARCH- t -Student.

5.2. Índice de Precios y Cotizaciones

El Índice de Precios y Cotizaciones (IPC) es el principal indicador del mercado mexicano de valores; expresa el rendimiento del mercado accionario en función de las variaciones de precios de una muestra balanceada, ponderada y representativa del conjunto de emisoras cotizadas en la bolsa, basado en las mejores prácticas internacionales. Adicionalmente, el IPC, con base octubre de 1978, tiene como principal objetivo constituirse como un indicador representativo del Mercado Mexicano para servir como referencia y subyacente de productos financieros.

La serie del IPC considerada para este trabajo abarca cotizaciones del 02 de enero de 1996 al 12 de enero de 2016 y corresponderá a los log-rendimientos de las cotizaciones históricas de dicho índice, tal que las pérdidas tengan asociado un signo positivo y las ganancias uno negativo. Esto es, si la cotización al tiempo t del Índice de Precios y Cotizaciones es de S_t unidades, entonces el rendimiento logarítmico negativo al tiempo t estará dado por

$$x_t = -\ln\left(\frac{S_t}{S_{t-1}}\right).$$

La gráfica del conjunto de información a analizar en esta Sección puede observarse en la Figura 5.25.

Ahora bien, nuestro enfoque al modelar la serie del IPC será el mismo que el utilizado al analizar el FIX. Esto es: modelaremos pérdidas diarias futuras de la serie del IPC a través del estudio la función de distribución $F_{X_{t+1}|\mathcal{F}_t}$.

De este modo, la ventana temporal sobre la que se realizarán nuestros análisis iniciales de la serie del IPC tendrá una longitud de $n = 1,000$ datos y corresponderá al periodo comprendido entre el 7 de septiembre de 2011 y el 31 de agosto de 2015. No obstante, en la parte final de esta Sección haremos uso de la totalidad de nuestra serie, a fin de realizar *backtests* de nuestras estimaciones de medidas condicionales de riesgo para el IPC.

Índice de Precios y Cotizaciones

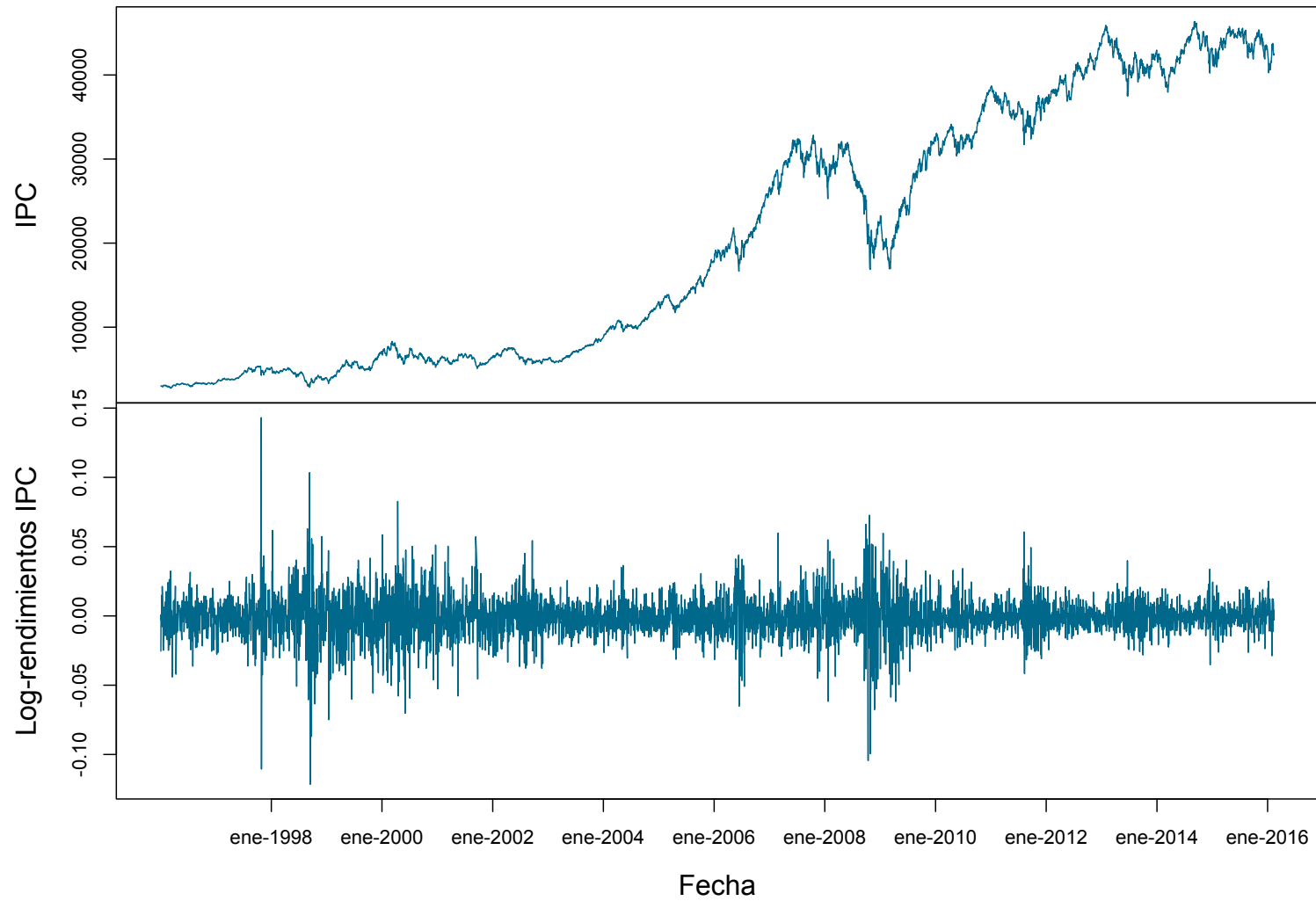


Figura 5.25: Índice de Precios y Cotizaciones (arriba) y pérdidas y ganancias de dicho índice (abajo); las series van de enero 1996 a febrero de 2016.

Log-Rendimientos IPC: Medidas descriptivas					
Mínimo	Máximo	Mediana	Media	Varianza	Ex. Curtosis
-0.0352	0.0494	-6.4×10^{-5}	-2.4×10^{-4}	8.3×10^{-5}	1.7851

Tabla 5.14: Medidas de Tendencia Central y Valores Importantes de los Log-Rendimientos del Índice de Precios y Cotizaciones entre septiembre de 2011 y octubre de 2015.

5.2.1. Análisis descriptivo

Iniciaremos el estudio de nuestro conjunto inicial de información del IPC a través de un análisis descriptivo. Así, en la Tabla 5.14 podemos observar una serie de medidas importantes de estadística descriptiva: los valores máximo y mínimo, la media, la mediana, la varianza y la curtosis de exceso.

Los valores contenidos en la Tabla 5.14 nos permiten notar que la media y la mediana de la serie del IPC se encuentran muy cercanas a cero, aunque presentan ambas un signo negativo. Este par de hechos reflejan, respectivamente, que durante nuestro periodo de estudio la ganancia promedio del IPC fue positiva y que en más de la mitad de las ocasiones terminó el día con ganancias. En otras palabras, esta serie presenta una tendencia a la alza, rindiendo más ganancias que pérdidas. No obstante, Tabla 5.14 también nos dice que la pérdida más grande de la serie es mayor, en valor absoluto, que la ganancia más grande. Finalmente, el valor de la curtosis de exceso refleja que las colas de esta serie podrían ser más pesadas que las de una distribución normal.

Una vez realizado este primer acercamiento a la serie del IPC, podemos proceder a explorarla con un poco más de profundidad. Por lo tanto, a continuación se presentan una serie de gráficos descriptivos de la serie; los cuales nos ayudarán a conocer mejor la naturaleza de la misma. Estos análisis descriptivos nos permitirán ganar intuición sobre los resultados que obtendremos más adelante, en estudios más especializados.

Histograma y *boxplot*

El primer par de gráficos exploratorios que utilizaremos con la serie del IPC serán el Histograma y el *boxplot* (diagrama de caja y brazos). En la Figura 5.26 podemos observar este par de gráficos para la serie del IPC.

Los dos gráficos de Figura 5.26 nos permite confirmar de forma gráfica que la

Índice de Precios y Cotizaciones

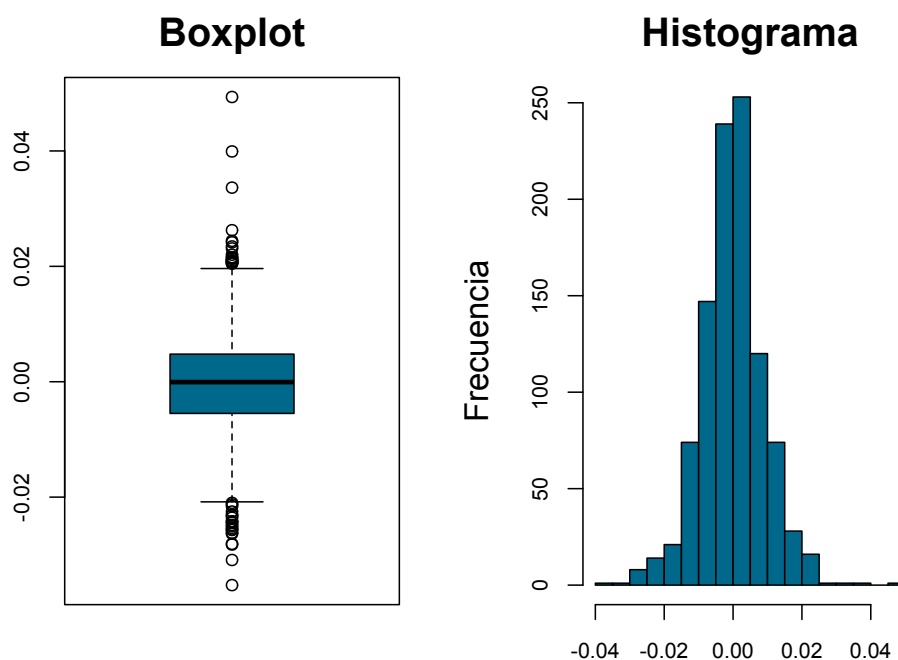


Figura 5.26: Histograma (izquierda) y *boxplot* (derecha) de los log-rendimientos del Índice de Precios y Cotizaciones de 2011 a 2015.

serie del IPC se encuentra centrada en cero. Adicionalmente, dicha Figura nos permite observar que la serie del IPC parece ser aproximadamente simétrica respecto a dicho valor.

Ahora bien, al analizar el *boxplot* en la Figura 5.26, podemos ver que los valores más extremos de la serie del IPC fueron negativos. Adicionalmente, al existir una cantidad importante de observaciones por encima y por debajo de los brazos del *boxplot*, dicha gráfica nos brinda la primera pieza de evidencia de que las colas del IPC podrían no ser ligeras. No obstante, la mayor parte de los datos de grandes magnitudes parecen tener un signo negativo (asociado a ganancias): existen 20 pérdidas por arriba del valor 0.02, mientras que hay 24 ganancias que, en valor absoluto, exceden dicho valor.

Así, dado nuestro particular interés en la naturaleza de las colas del IPC, a continuación exploraremos de forma más profunda qué tan pesadas son dichas colas.

Gráfico Q-Q - Índice de Precios y Cotizaciones

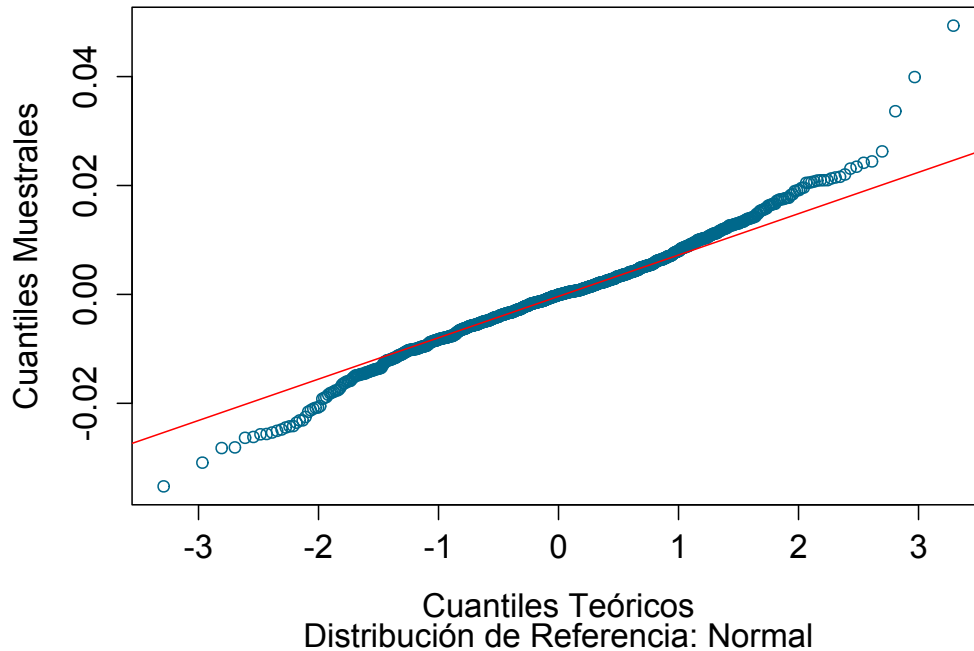


Figura 5.27: Gráfico QQ de los log-rendimientos del Índice de Precios y Cotizaciones contra una distribución de referencia Normal. Los parámetros de la distribución Normal se obtuvieron al ajustar dicho modelo a los log-rendimientos del IPC vía Máxima Verosimilitud.

QQ-Plot

Dada la evidencia inicial que apunta a que las colas del IPC podrían no ser ligeras (valor de la curtosis de exceso y *boxplot* del IPC), investigaremos vía un QQ-Plot qué tan pesadas son éstas en comparación con las de una distribución bien conocida: la distribución Normal.

De este modo, en la Figura 5.27 podemos observar un QQ-Plot de la serie del IPC contra una distribución de referencia Gaussiana. Dicha Figura nos permite apreciar que debido a las observaciones extremas presentes en la serie del IPC, la cola derecha de dicha serie parece ser más pesada que la izquierda. No obstante, fuera de dichos valores extremos, las colas de la serie del IPC no parecen ser mucho más pesadas que las de una distribución normal: la forma de “S” volteada no es muy pronunciada para esta serie.

El hecho anterior sugiere que la serie del IPC podría pertenecer al $DAM(H_\xi)$

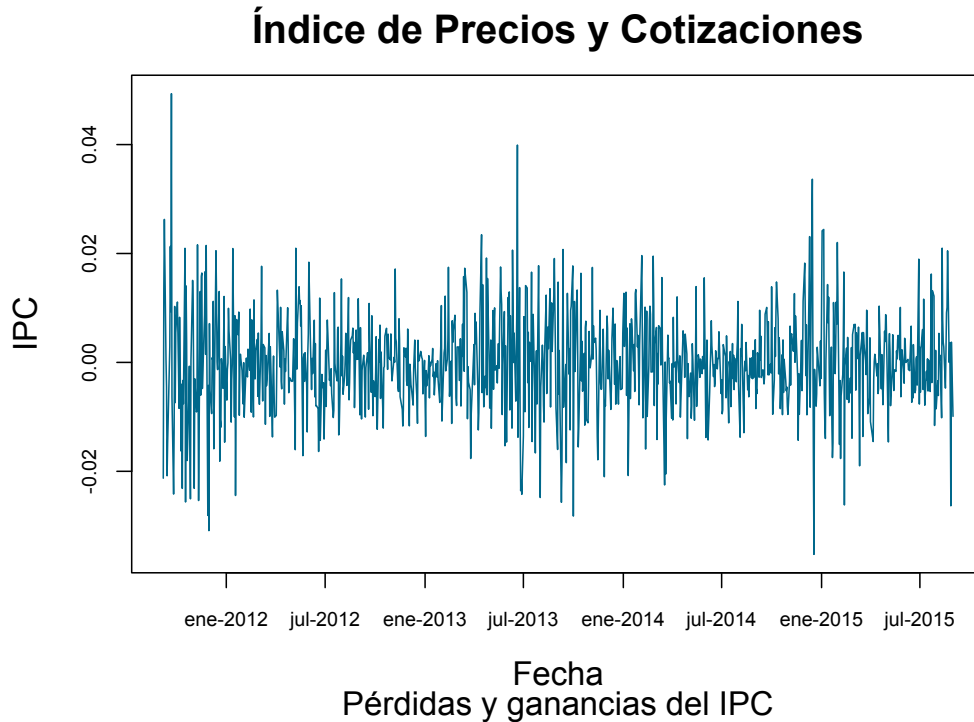


Figura 5.28: Pérdidas y ganancias del Índice de Precios y Cotizaciones de septiembre de 2011 a agosto de 2015.

con $\xi = 0$; a pesar de tener colas más pesadas que las de una distribución normal, éstas no parecen tan pesadas como las de una distribución de la clase Fréchet.

Ahora bien, para concluir nuestro análisis exploratorio de la serie del IPC tomaremos en cuenta un hecho obviado hasta este momento: los datos del IPC constituyen una serie de tiempo; pues constituyen un conjunto de observaciones indexadas por un conjunto de fechas.

Gráfica del IPC

En la Figura 5.28 podemos ver las observaciones del IPC graficadas como una serie de tiempo. Dicha Figura nos permite ver que las observaciones del IPC parecen presentar periodos de estrés: la gráfica del IPC presenta periodos donde grandes valores (quizá de signo distinto) se sucedieron entre sí hasta llegar a un periodo de relativa calma (con observaciones de menor magnitud).

El hecho anterior nos dice que la serie del IPC podría no ser homocedástica y que las observaciones que la componen podrían tampoco ser independientes entre sí. No obstante, por ahora no investigaremos de forma más profunda este par de hechos; dejaremos su análisis y tratamiento para secciones posteriores, donde los exploraremos de forma especializada.

De este modo, podemos concluir nuestro análisis inicial de la serie del IPC. Dicho análisis nos permitió tener un primer acercamiento con dichos datos, permitiéndonos conocer algunas de sus principales características.

Recapitulando: es probable que las observaciones del IPC no sean independientes y que su variabilidad tienda a cambiar a lo largo del tiempo; por otro lado, los valores más extremos del IPC estuvieron asociados a pérdidas, a pesar de que esta serie tiene una tendencia importante a rendir ganancias; finalmente, a pesar de que las colas de esta serie son más pesadas que las de una distribución normal, ésta podría pertenecer $DAM(H_0)$.

Así, una vez reunido este conocimiento inicial de la serie del IPC, podemos proceder a analizar qué tan apropiada es la metodología estudiada en el Capítulo 4 para medir el riesgo acarreado por dicho índice. Por lo tanto, a continuación procederemos a estudiar si el ajuste de un modelo de volatilidad estocástica como los estudiados en el Capítulo 2 podría resultar plausible para la serie del IPC.

5.2.2. Modelando la dinámica del IPC

Dado que la metodología estudiada en el Capítulo 4 considera modelar la dinámica de una serie de tiempo financiera mediante un modelo ARCH o GARCH, a continuación exploraremos qué tan razonable podría resultar el ajuste de dicha clase de modelos a la serie del IPC. Para ello, exploraremos dicha serie de forma más especializada que en la Sección anterior, a fin de verificar en ella la presencia de una serie de hechos estilizados de series de tiempo financieras que pueden ser modeladas de forma satisfactoria por procesos ARCH y GARCH.

Dicha serie de hechos estilizados consta de 6 características esenciales, las cuales enumeramos al analizar la serie del FIX, en la Sección 5.1.2. Así, a continuación exploraremos si dichos hechos estilizados se encuentran presentes en la serie del IPC.

Hechos estilizados de series con volatilidad estocástica

Comenzaremos esta exploración notando que la información recopilada en nuestro análisis exploratorio inicial nos dio evidencia de algunos de los hechos estilizados que buscamos.

Primeramente, el QQ-Plot de la serie del IPC y el valor de su curtosis de exceso, nos dieron la primera pieza de evidencia de que esta serie podría ser leptocúrtica. Así, podemos concluir que el hecho estilizado número 5 se encuentra presente en la serie del FIX.

Por otro lado, nuestro análisis de la gráfica del IPC como serie de tiempo señaló que es probable que las observaciones de dicha serie no sean independientes y que su variabilidad tienda a cambiar a lo largo del tiempo. No obstante, a fin de investigar este par de hechos de forma más profunda, a continuación utilizaremos una serie de pruebas tanto gráficas como numéricas para determinar si dichas observaciones se sostienen en la serie del IPC.

Comenzaremos por evaluar la independencia de las observaciones del IPC. Así, notemos que si éstas fueran independientes, los valores de su función de autocorrelación deberían ser todos iguales a cero (para $h > 0$). Asimismo, si les aplicáramos una transformación, los datos resultantes deberían seguir siendo independientes y sus correspondientes funciones de autocorrelación deberían tener valores iguales a cero para $h > 0$.

De este modo, nuestra primera prueba será gráfica y consistirá en evaluar los hechos anteriores para un par de transformaciones: el cuadrado y el valor absoluto. Así, en la Figura 5.29 podemos observar la gráfica de la función de autocorrelación muestral de las observaciones del IPC crudas, al cuadrado y en valor absoluto.

La Figura 5.29 nos permite observar que los valores de la función de autocorrelación de las observaciones crudas del IPC son todos estadísticamente iguales a cero. Sin embargo, el comportamiento de las gráficas de autocorrelación muestral de las observaciones del IPC al cuadrado y en valor absoluto es completamente distinto: presentan valores estadísticamente mayores que cero para todos los retrasos considerados. Este hecho presenta una fuerte evidencia en contra de la hipótesis de independencia de las observaciones del IPC.

No obstante, para investigar de forma más profunda la independencia de la serie del IPC, procederemos a aplicar una prueba de independencia numérica en dicha serie: la prueba de Ljung-Box.

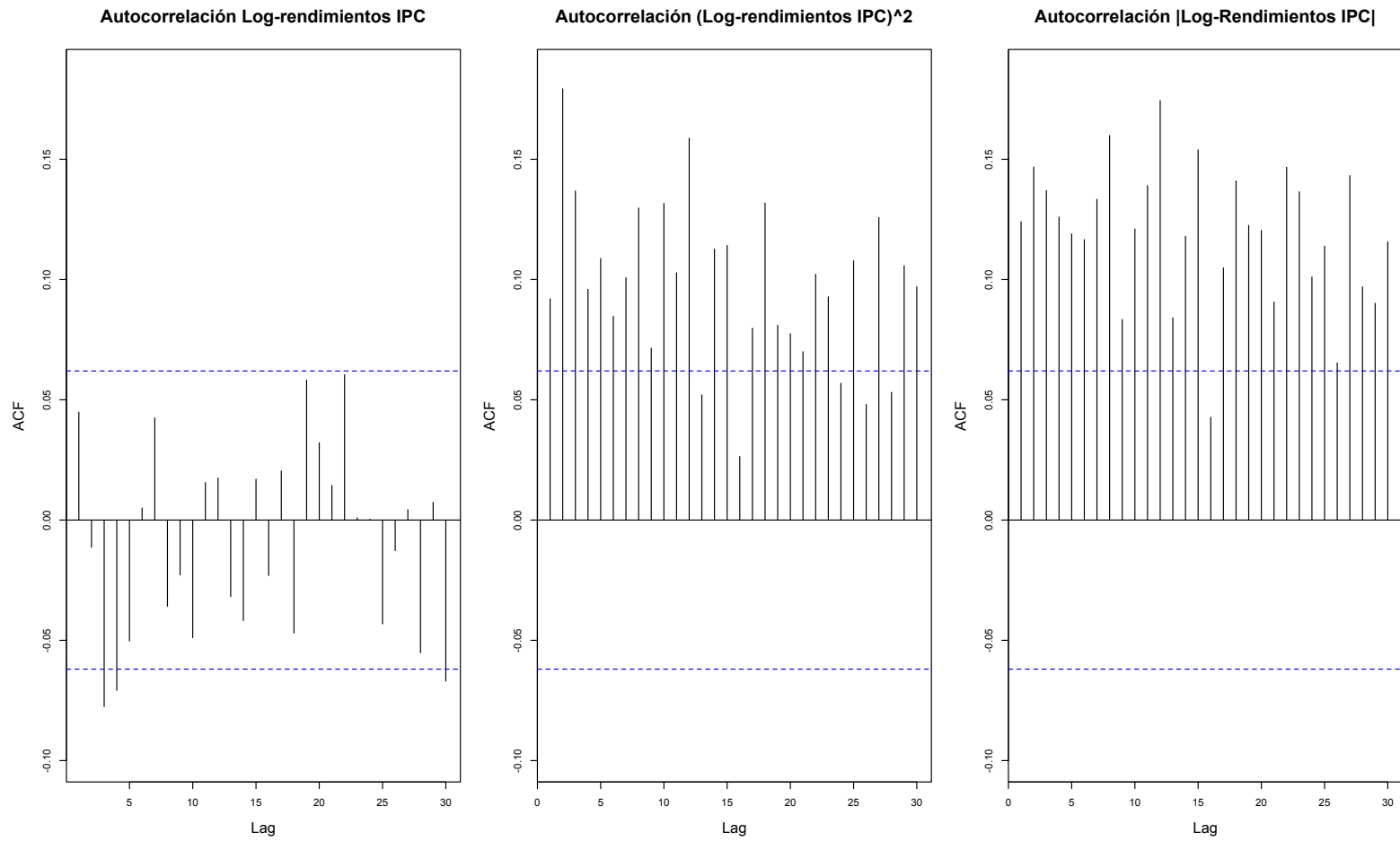


Figura 5.29: Autocorrelación del IPC. Se muestran las gráficas de la función de autocorrelación muestral para los datos crudos (izquierda), sus cuadrados (centro) y valores absolutos (derecha).

Prueba de Ljung-Box en el IPC: p-valores			
Datos	$h = \ln(n)$	$h = 20$	$h = 30$
IPC original	0.0276	0.0446	0.0307
IPC al cuadrado	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$
IPC en valor absoluto	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

Tabla 5.15: P-valores de las pruebas de Ljung-Box aplicadas a las observaciones del Índice de Precios y Cotizaciones crudas, al cuadrado y en valor absoluto.

Nuestra aplicación de la prueba de Ljung-Box al IPC se basa en distintos retrasos. Además, para ganar mayor certeza sobre la independencia de los datos originales, dicha prueba se aplica también a las observaciones crudas, al cuadrado y en valor absoluto del IPC.

En la Tabla 5.15 podemos observar los p-valores resultantes de la aplicación de la prueba de Ljung-Box al IPC. Así, los valores contenidos en dicha Tabla nos indican que la hipótesis de independencia en el IPC fue rechazada, para todos los retrasos y tipos de observaciones considerados.

Al tener evidencia consistente por parte de pruebas gráficas y numéricas, podemos concluir que las observaciones del IPC no son independientes. No obstante, la Figura 5.29 nos mostró que las observaciones crudas del IPC presentan una correlación serial muy leve. Por lo tanto, podemos concluir que el hecho estilizado número 1 se encuentra presente en la serie del IPC.

Ahora bien, nuestra elección de transformaciones para realizar las pruebas de independencia de la serie del IPC claramente no fue arbitraria. La Figura 5.29 también nos muestra que las series de los cuadrados y valores absolutos del IPC muestran una correlación serial muy fuerte. Así, podemos concluir que el hecho estilizado número 2 también se encuentra presente en la serie del IPC.

Pasamos ahora a investigar la presencia del hecho estilizado número 4 en el IPC. Para ello, analizaremos la gráfica de un par de series que nos ayudarán a apreciar de mejor forma la variabilidad del IPC a lo largo del tiempo: las series del cuadrado y valor absoluto del IPC.

De este modo, en la Figura 5.30 podemos observar las gráficas como series de tiempo de las observaciones del IPC crudas, al cuadrado y en valor absoluto. Así, al analizar dichas gráficas podemos ver de forma clara que la serie del IPC sí presenta periodos de estrés (*clusters* de volatilidad).

Específicamente, al analizar las gráficas de la Figura 5.30 es evidente la exis-

Variabilidad del IPC

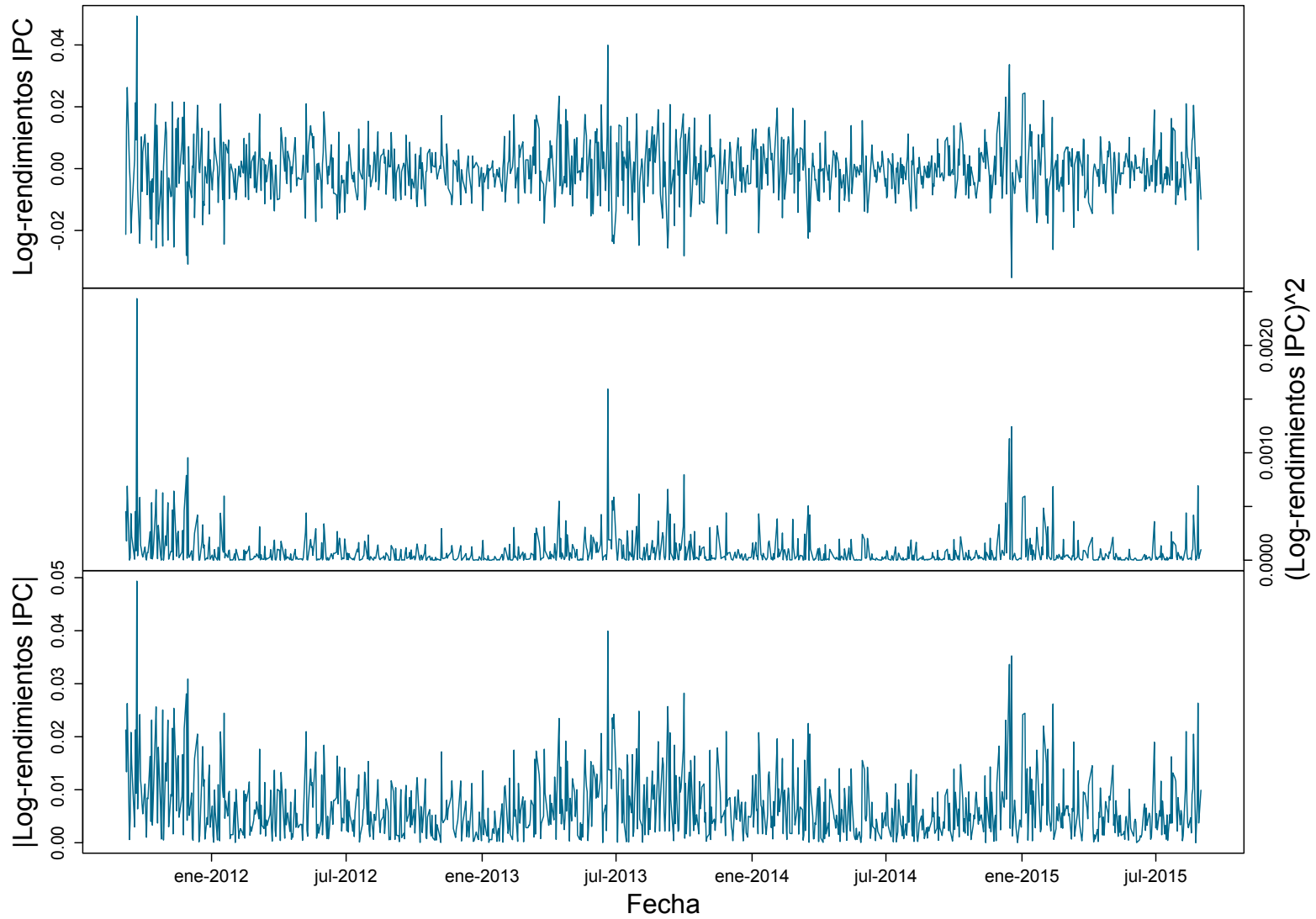


Figura 5.30: Gráficas del Índice de Precios y Cotizaciones. Observaciones crudas (arriba), al cuadrado (en medio) y en valor absoluto (abajo).

tencia de periodos llenos de observaciones de grandes magnitudes (quizá con signo distinto), que se suceden entre sí hasta que la serie llega a un nueva etapa de relativa calma. Ejemplos de esta clase de *clusters* se encuentran del inicio de la serie hasta enero de 2012, de septiembre de 2014 a marzo de 2015 o de julio de 2015 hasta el final de nuestra ventana temporal inicial.

Por lo tanto, dada la evidencia reunida en la Figura, podemos concluir que los hechos estilizados 4 y 6 están presentes en la serie del IPC.

Finalmente, recordemos que los valores de la función de autocorrelación de la serie del IPC son estadísticamente nulos y que ésta se encuentra centrada en el cero. La combinación de este par de hechos deriva en que nuestra mejor predicción de rendimientos futuros con base en rendimientos pasados es que éstos serán iguales a cero. Por lo tanto, el hecho estilizado número 3 se encuentra presente en la serie del IPC.

Con el análisis anterior hemos terminado de documentar la presencia de los hechos estilizados de series con volatilidad estocástica en el IPC, concluyendo que todos ellos se encuentran presentes en dicha serie. Por lo tanto, podemos concluir que un modelo de volatilidad estocástica, como los procesos ARCH y GARCH estudiados en el Capítulo 2, podría modelar de forma adecuada la dinámica del IPC. No obstante, aún debemos determinar qué proceso ARCH o GARCH podría modelar de mejor forma al IPC.

Ajustando un modelo de volatilidad estocástica al IPC

Dada la evidencia que sugiere que un proceso de las familias ARCH o GARCH podría modelar de forma apropiada la dinámica del IPC, en esta Sección analizaremos qué proceso de dichas familias es el más adecuado para dicha serie de tiempo. Adicionalmente, investigaremos si el modelo que seleccionemos para la dinámica del IPC logra tener una buena calidad de ajuste.

Para determinar el orden del proceso ARCH o GARCH que podría subyacer a nuestros datos utilizaremos de nueva cuenta la relación entre procesos ARCH y GARCH al cuadrado con procesos AR y ARMA respectivamente. Por lo tanto, en la Figura 5.31 podemos observar las gráficas de las funciones de autocorrelación y autocorrelación parcial muestrales de las observaciones del IPC cuadrado.

Al analizar la Figura 5.31 podemos deducir que:

1. La gráfica de autocorrelación parcial muestral del IPC tiene sus valores más grandes para los retrasos $h = 1, 2, 3$. Ello sugiere que el ajuste

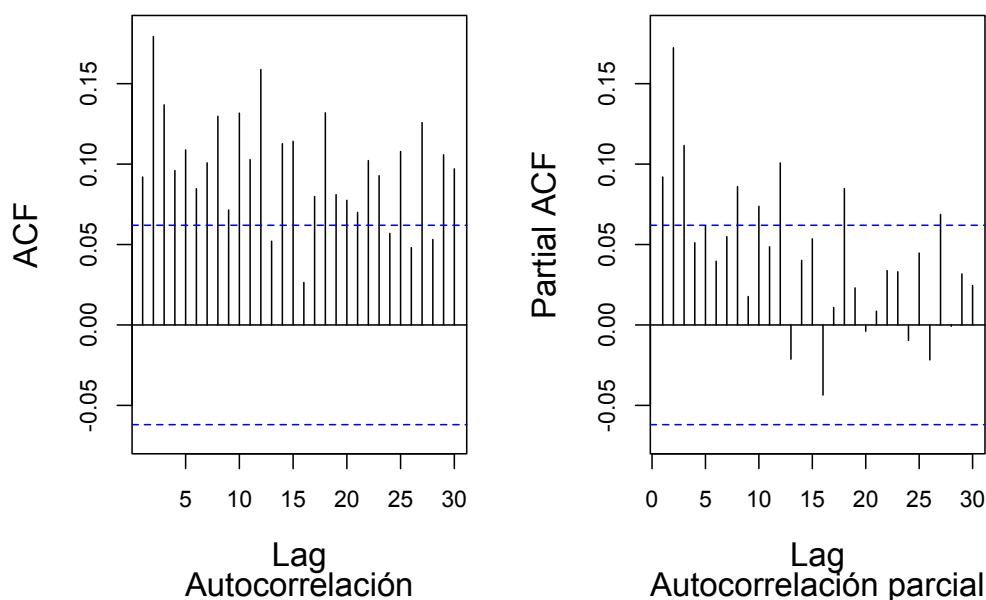
(Log-rendimientos IPC)²

Figura 5.31: Autocorrelación (izquierda) y autocorrelación parcial (derecha) de los log-rendimientos al cuadrado del Índice de Precios y Cotizaciones.

de un modelo $\text{ARCH}(p)$ con $p = 1, 2, 3$ podría resultar razonable para la serie del IPC. Ahora bien, si bien es cierto que la gráfica de autocorrelación parcial muestral tiene valores estadísticamente distintos de cero para otros retrasos, éstos son menores que los primeros tres y se encuentran en retrasos muy grandes. Por lo tanto, tomarlos en consideración podría llevarnos a generar modelos demasiado complicados que podrían sobreajustar la serie del IPC. De este modo, solo consideraremos el ajuste de modelos $\text{ARCH}(p)$ con $p = 1, 2, 3$ para dicha serie.

- Por otro lado, la gráfica de la función de autocorrelación muestral tiene muchos valores estadísticamente distintos de cero. Ello sugiere que el ajuste de modelos $\text{GARCH}(p, q)$ con $q = 1, \dots, 30$ podría resultar plausible para la serie del IPC. No obstante, realizar el ajuste de modelos GARCH de órdenes tan altos resultaría poco práctico y correríamos el riesgo de sobreajustar nuestro modelo para la dinámica del IPC. Por lo tanto, en esta ocasión tomaremos un enfoque más parsimonioso y ajustaremos a la serie del IPC modelos $\text{GARCH}(p, q)$ con $q = 1, 2, 3, 4, 5$.

Log-Rendimientos IPC: AICC modelos ARCH(p)			
p	1	2	3
AICC	-6.557992	-6.580914	-6.563010

Tabla 5.16: AICC de los modelos ARCH(p) ajustados a la serie del Índice de Precios y Cotizaciones.

Significancia de parámetros, modelo ARCH(2)		
Parámetro	Estimación puntual	P-valor
α_0	5.163×10^{-05}	$< 2 \times 10^{-16}$
α_1	0.2277	1.50×10^{-05}
α_2	0.1749	1.27×10^{-05}

Tabla 5.17: Significancia estadística de los parámetros del modelo ARCH(2) ajustado a la serie del Índice de Precios y Cotizaciones.

Con el análisis anterior, contamos ya con diversos candidatos de procesos ARCH y GARCH para modelar la dinámica del IPC. Por lo tanto, a continuación seleccionaremos el mejor de ellos, a fin de tener un modelo final para dicha dinámica. El criterio que utilizaremos para seleccionar los mejores modelos ARCH y GARCH será el criterio de Akaike.

Comenzando por los procesos ARCH, en la Tabla 5.16 podemos encontrar el AICC de los modelos ARCH(p), con $p = 1, 2, 3$, ajustados a la serie del IPC. Así, los valores contenidos en dicha Tabla señalan que el mejor modelo ARCH para la dinámica del IPC es el proceso ARCH(2).

No obstante, a fin de considerar el proceso ARCH(2) como el modelo ARCH final para la dinámica del IPC, debemos verificar que los parámetros de dicho modelo sean estadísticamente significativos y que el proceso resultante sea estacionario en covarianza.

Por lo tanto, en la Tabla 5.17 podemos encontrar los valores estimados de los parámetros del modelo ARCH(2) ajustado al IPC, así como los p-valores de la prueba de t de significancia estadística de dichos parámetros. Así, los valores contenidos en dicha Tabla nos permiten concluir que, en efecto, los parámetros del modelo ARCH(2) ajustado al IPC son todos estadísticamente significativos así como que dicho modelo es estacionario en covarianza (pues $\alpha_1 + \alpha_2 = 0.4026 < 1$). Por lo tanto, tomaremos el proceso ARCH(2) como modelo ARCH final para el IPC.

Continuamos ahora con la elección del mejor modelo GARCH para la dinámica del IPC. Por lo tanto, en la Tabla 5.18 podemos encontrar el AICC de

Log-Rendimientos IPC: AICC modelos GARCH(p, q)					
p/q	1	2	3	4	5
1	-6.653637	-6.628635	-6.601171	-6.567152	-6.533591
2	-6.629757	-6.606172	-6.573828	-6.537735	-6.499221
3	-6.601642	-6.572282	-6.539423	-6.499294	-6.457126

Tabla 5.18: AICC de los modelos GARCH(p, q) ajustados a la serie del Índice de Precios y Cotizaciones.

Significancia de parámetros, modelo GARCH(1, 1)		
Parámetro	Estimación puntual	P-valor
α_0	1.333×10^{-6}	0.0118
α_1	0.0671	4.46×10^{-7}
β_1	0.9164	$< 2 \times 10^{-16}$

Tabla 5.19: Significancia estadística de los parámetros del modelo GARCH(1, 1) ajustado a la serie del Índice de Precios y Cotizaciones.

los procesos GARCH(p, q), $q = 1, 2, 3, 4, 5$, ajustados al IPC. Así, los valores contenidos en dicha Tabla señalan que el proceso GARCH(1,1) es el mejor modelo GARCH para la serie del IPC.

Sin embargo, antes de poder declarar al proceso GARCH(1,1) como nuestro modelo GARCH final para el IPC, debemos verificar la significancia de sus parámetros, así como que dicho proceso sea estacionario en covarianza. Por lo tanto, en la Tabla 5.19 podemos encontrar los valores estimados y p-valores de la prueba t de significancia estadística de los parámetros de dicho modelo.

De este modo, los valores contenidos en la Tabla 5.19 nos permiten concluir que los parámetros del GARCH(1,1) ajustado al IPC son todos estadísticamente significativos y que dicho modelo es estacionario en covarianza (pues $\alpha_1 + \beta_1 = 0.9835 < 1$). Por lo tanto, tomaremos al proceso GARCH(1,1) como modelo GARCH final para la dinámica de la serie del IPC.

Ahora bien, con los análisis anteriores contamos con dos candidatos potenciales para modelar la dinámica del IPC: un ARCH(2) y un GARCH(1,1). Debido a que ambos modelos cuentan con parámetros estadísticamente significativos y son estacionarios en covarianza, debemos utilizar algún otro criterio para seleccionar el mejor de ellos para la serie del IPC.

En nuestro caso, dicho criterio será el AICC. Al comparar dicha medida entre los dos procesos seleccionados para modelar al IPC, podemos concluir que el GARCH(1,1) es el modelo que resulta favorecido por el criterio de Akaike.

Estimación GARCH de la Volatilidad del IPC

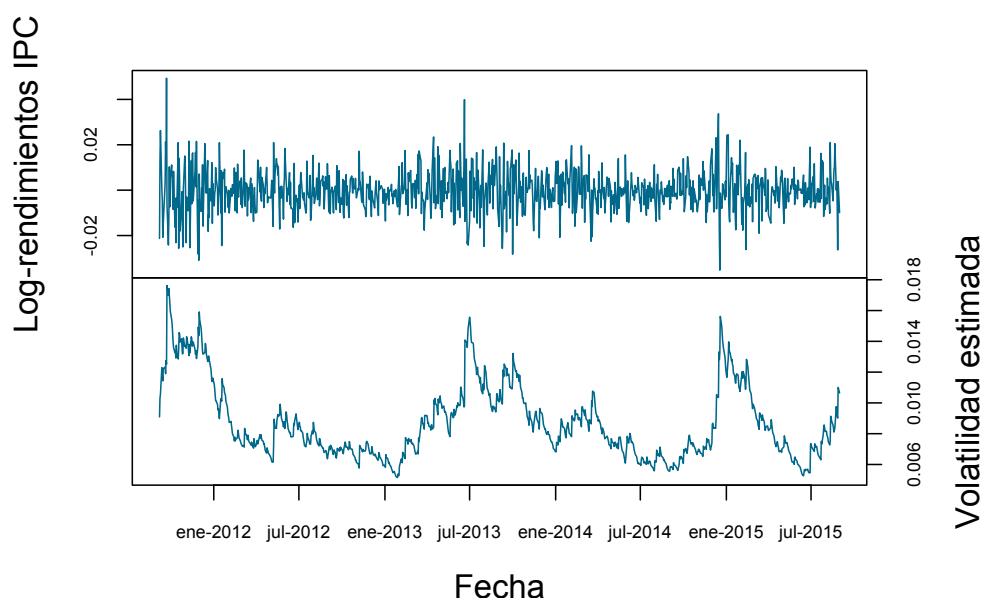


Figura 5.32: Series de log-rendimientos y de volatilidad estimada del Índice de Precios y Cotizaciones. La volatilidad del IPC fue estimada vía el modelo GARCH(1,1) ajustado a dicha serie.

Por lo tanto, dicho proceso será nuestra propuesta de modelo final para la dinámica del IPC. No obstante, antes de proceder a realizar un mayor análisis del IPC con base en el modelo GARCH(1,1), debemos asegurarnos que dicho proceso modela de forma adecuada dicha dinámica.

Evaluación de la calidad de ajuste del GARCH(1,1) a la dinámica del IPC

Para evaluar la calidad de ajuste del GARCH(1,1) al IPC podemos realizar un análisis de los residuales estandarizados de dicho modelo. Esto pues, como estudiamos en el Capítulo 2, si el modelo GARCH ajustado a una serie ha logrado capturar todos los elementos de su dinámica, dichos residuos estandarizados deberían comportarse como un ruido blanco estricto; hecho que puede ser evaluado a través de diversas pruebas, tanto gráficas como numéricas.

De este modo, el primer paso para evaluar la calidad de ajuste del GARCH(1,1) al IPC es obtener las versiones muestrales de los residuos estandarizados de dicho modelo. Para ello, debemos estimar la volatilidad del IPC a través del modelo GARCH(1,1) ajustado. Dicha estimación puede obtenerse al sustituir los valores estimados de los parámetros del GARCH(1,1) ajustado al IPC en la fórmula dada en (5.3). En nuestro caso, los valores iniciales X_0^2 y σ_0^2 serán tomados como la media y varianza muestrales, respectivamente.

De esta forma, en la Figura 5.32 podemos observar la gráfica de la volatilidad del IPC estimada vía el modelo GARCH(1,1) ajustado a dicha serie. Así, la Figura 5.32 nos permite observar el dinamismo de la volatilidad estimada con base en un modelo GARCH: mayor en los periodos donde se concentran grandes valores de la serie y moderada en aquellos donde las observaciones son menores.

Ahora bien, una vez obtenidos los valores estimados de la volatilidad del IPC, podemos obtener las versiones muestrales de los residuos del GARCH(1,1) ajustado a dicha serie y evaluar la hipótesis de que éstos se comporten como un ruido blanco estricto. Realizaremos dicha evaluación a través de diversas pruebas gráficas y numéricas.

Comenzaremos con la evidencia gráfica. En la Figura 5.33 podemos observar las gráficas como serie de tiempo de las observaciones del IPC crudas, al cuadrado y en valor absoluto. Dichas gráficas nos permiten apreciar que los *clusters* de valores grandes ya no se encuentran presentes en las innovaciones del IPC, lo que da evidencia de que la varianza de dichas observaciones podría ser constante a lo largo del tiempo.

Adicionalmente, al haberse eliminado los cúmulos de valores grandes, se tiene evidencia de que la dependencia que existía entre las observaciones de grandes magnitudes del IPC se ha logrado eliminar de las innovaciones de dicha serie; ello vuelve más plausible la hipótesis de que los residuos estandarizados del GARCH(1,1) ajustado al IPC se comporten como un ruido blanco estricto.

Por otro lado, en la Figura 5.34 podemos observar las gráficas de las funciones de autocorrelación y autocorrelación parcial muestrales de las innovaciones del IPC crudas, al cuadrado y en valor absoluto. Así, al observar dicha Figura podemos comprobar que los valores de dichas funciones son todos estadísticamente iguales a cero, para las tres series mencionadas y todos los retrasos considerados.

Así, la Figura 5.34 constituye una evidencia importante de que las innovaciones del IPC sí podrían soportar la hipótesis de comportarse como un ruido blanco estricto.

Variabilidad de las Innovaciones del IPC

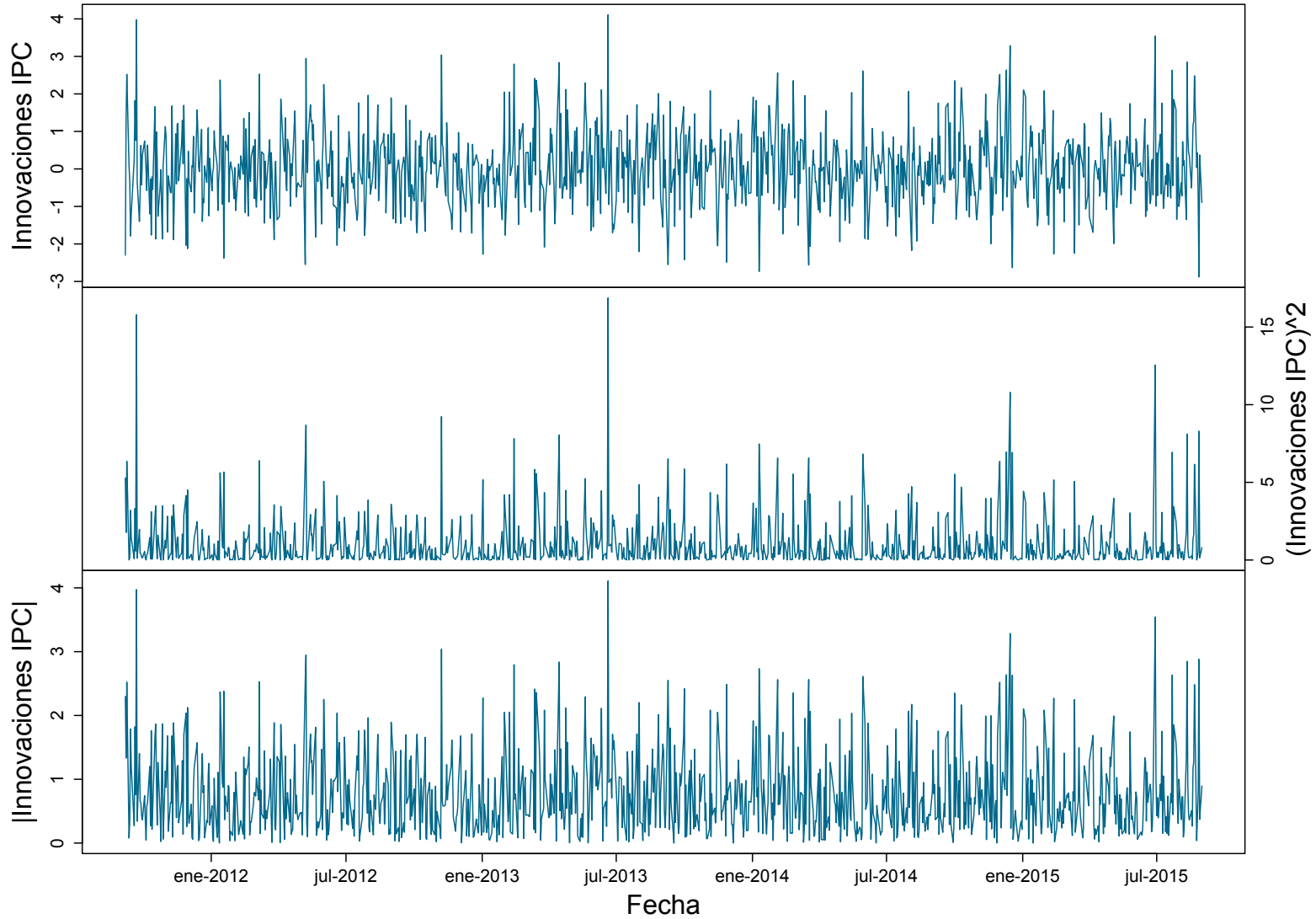


Figura 5.33: Gráficas como serie de tiempo de las Innovaciones del GARCH(1,1) ajustado al Índice de Precios y Cotizaciones: innovaciones crudas (arriba), al cuadrado (en medio) y en valor absoluto (abajo).

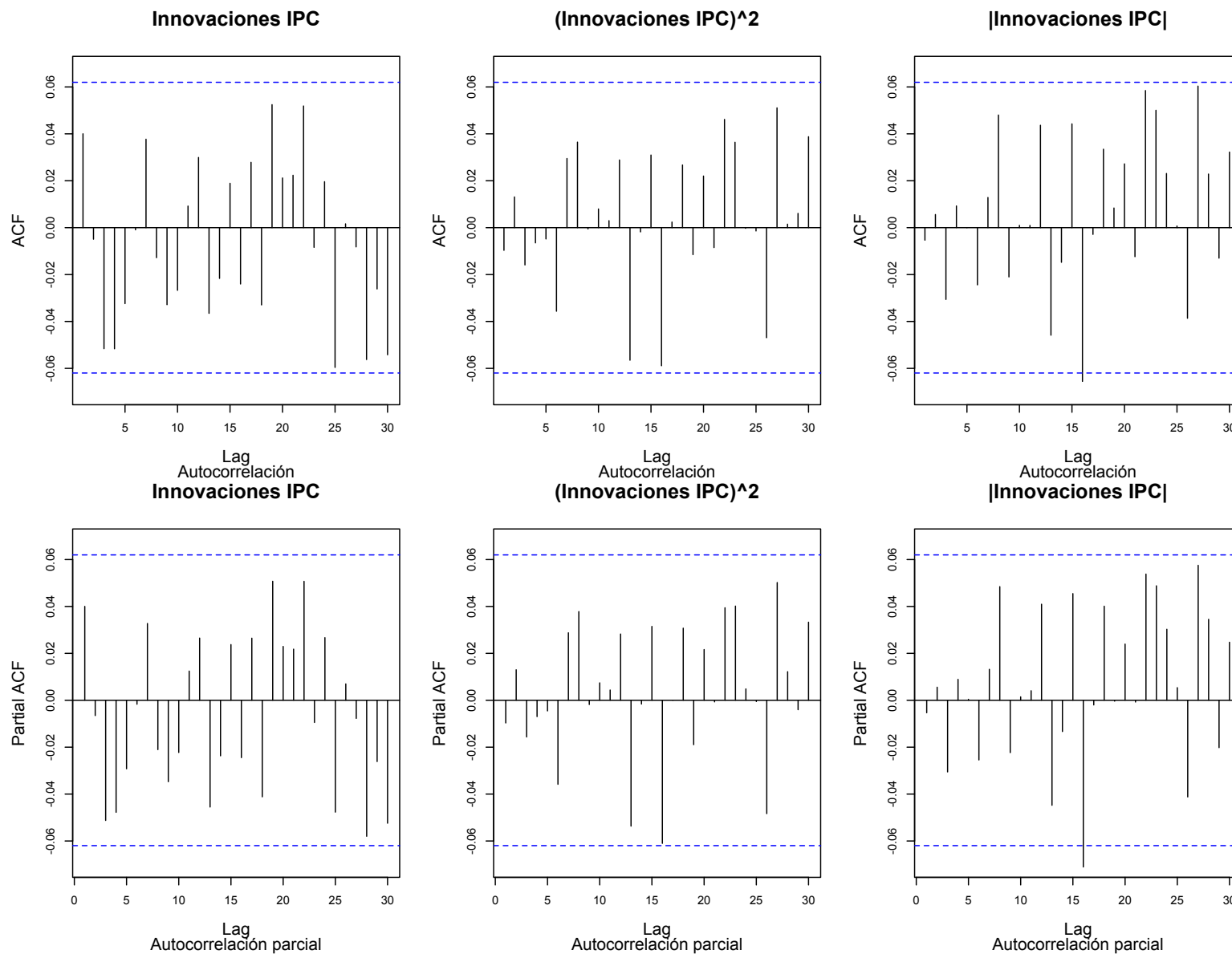


Figura 5.34: Autocorrelación y autocorrelación parcial de los residuos estandarizados del modelo GARCH(1,1) ajustado al Índice de Precios y Cotizaciones. Se muestran las gráficas del ACF (primer renglón) y PACF (segundo renglón) de los datos crudos, al cuadrado, y en valor absoluto.

Prueba de Ljung-Box en las innovaciones del IPC: p-valores			
Datos	$h = \ln(n)$	$h = 20$	$h = 30$
FIX original	0.2196	0.4325	0.2474
FIX al cuadrado	0.9075	0.8241	0.7588
FIX en valor absoluto	0.9674	0.6353	0.4096

Tabla 5.20: P-valores de las pruebas de Ljung-Box aplicadas a las innovaciones del IPC crudas, al cuadrado y en valor absoluto.

Pasando a pruebas numéricas, en la Tabla 5.20 podemos observar los p-valores de la prueba de Ljung-Box aplicada a las innovaciones del IPC. La aplicación de dicha prueba a las innovaciones del IPC se realizó sobre las observaciones crudas, al cuadrado y en valor absoluto; además, la prueba se basó en tres retrasos distintos.

De esta manera, al analizar los valores contenidos en la Tabla 5.20 podemos apreciar que en ningún caso se rechaza la hipótesis de que las innovaciones del IPC sean independientes entre sí. Por lo tanto, dicha Tabla provee mayor evidencia de que dichas innovaciones se comportan como un ruido blanco estricto.

Por lo tanto, al contar con evidencia consistente de pruebas tanto gráficas como numéricas que señala que las innovaciones del IPC se comportan como un ruido blanco estricto, podemos concluir que dichas observaciones sí soportan tal hipótesis; lo cual quiere decir que el modelo GARCH(1,1) ajustado al IPC ha logrado modelar de forma satisfactoria la dinámica de dicha serie de tiempo.

De este modo, tomaremos el proceso GARCH(1,1) como modelo final para la dinámica del IPC y lo utilizaremos como la base de los análisis siguientes.

Pronóstico de la volatilidad del IPC

Para avanzar hacia nuestros objetivos finales, el primer paso a seguir es realizar un pronóstico de la volatilidad del IPC. En esta ocasión, podemos basar dicho pronóstico en el modelo GARCH(1,1) ajustado al IPC, puesto que todos los parámetros de dicho modelo resultaron estadísticamente significativos.

Por lo tanto, el pronóstico de volatilidad del IPC se obtiene al sustituir los valores estimados de los parámetros del GARCH(1,1) ajustado a dicha serie en la fórmula (5.2). Dicho pronóstico cuenta con un par de características

Innovaciones del IPC: Medidas descriptivas					
Mínimo	Máximo	Mediana	Media	Varianza	Ex. Curtosis
-2.7790	4.1040	0.0298	0.0233	1.0103	0.6269

Tabla 5.21: Medidas de Tendencia Central y Valores Importantes de las innovaciones del Índice de Precios y Cotizaciones, entre septiembre de 2011 y agosto de 2015.

atractivas desde un punto de vista práctico. Primeramente, permite realizar pronósticos de volatilidad para más de un paso adelante; es decir, pronósticos al tiempo $t + h$, con $h \geq 1$. Adicionalmente, dicho pronóstico tiene una propiedad de convergencia; específicamente, conforme $h \rightarrow \infty$, el pronóstico dado en (5.2) converge a la raíz cuadrada de la varianza no condicional del proceso GARCH(1,1) en el que se está basando la estimación: $\alpha_0/(1-\alpha_1-\beta_1)$.

Ahora bien, en virtud de que estamos primordialmente interesados en pronósticos de volatilidad un paso adelante, notemos que en este caso la estimación de $\hat{\sigma}_{t+1}^2$ está dada por:

$$\begin{aligned}
 \hat{\sigma}_{t+1}^2 &= \alpha_0 + \alpha_1 X_t^2 + \beta_1 \hat{\sigma}_t^2 \\
 &= 1.333 \times 10^{-06} + (0.0671)X_t^2 + (0.9164)\hat{\sigma}_t^2 \\
 &= 1.333 \times 10^{-06} + (0.0671)(-0.009909)^2 + (0.9164)(0.010649)^2 \\
 &= 0.000111.
 \end{aligned} \tag{5.7}$$

De esta forma, al calcular la raíz cuadrada de la expresión dada en (5.7) obtendríamos una estimación de la volatilidad del IPC un paso adelante.

Por otro lado, en la Figura 5.35 podemos observar la gráfica de la volatilidad estimada y pronosticada del IPC; en particular, se muestra el pronóstico a un año de la volatilidad del IPC. Así, la Figura 5.35 nos permite observar el comportamiento de un pronóstico de volatilidad realizado con base en un modelo GARCH: una curva suave que converge a la raíz cuadrada de la varianza no condicional de dicho proceso.

Así, una vez realizado este pronóstico de volatilidad, podemos pasar a analizar de forma detallada la naturaleza de la distribución de las innovaciones del IPC. Dados nuestros objetivos finales, dicho análisis se centrará en el comportamiento de las colas de tal distribución.

Índice de Precios y Cotizaciones: Predicción de Volatilidad

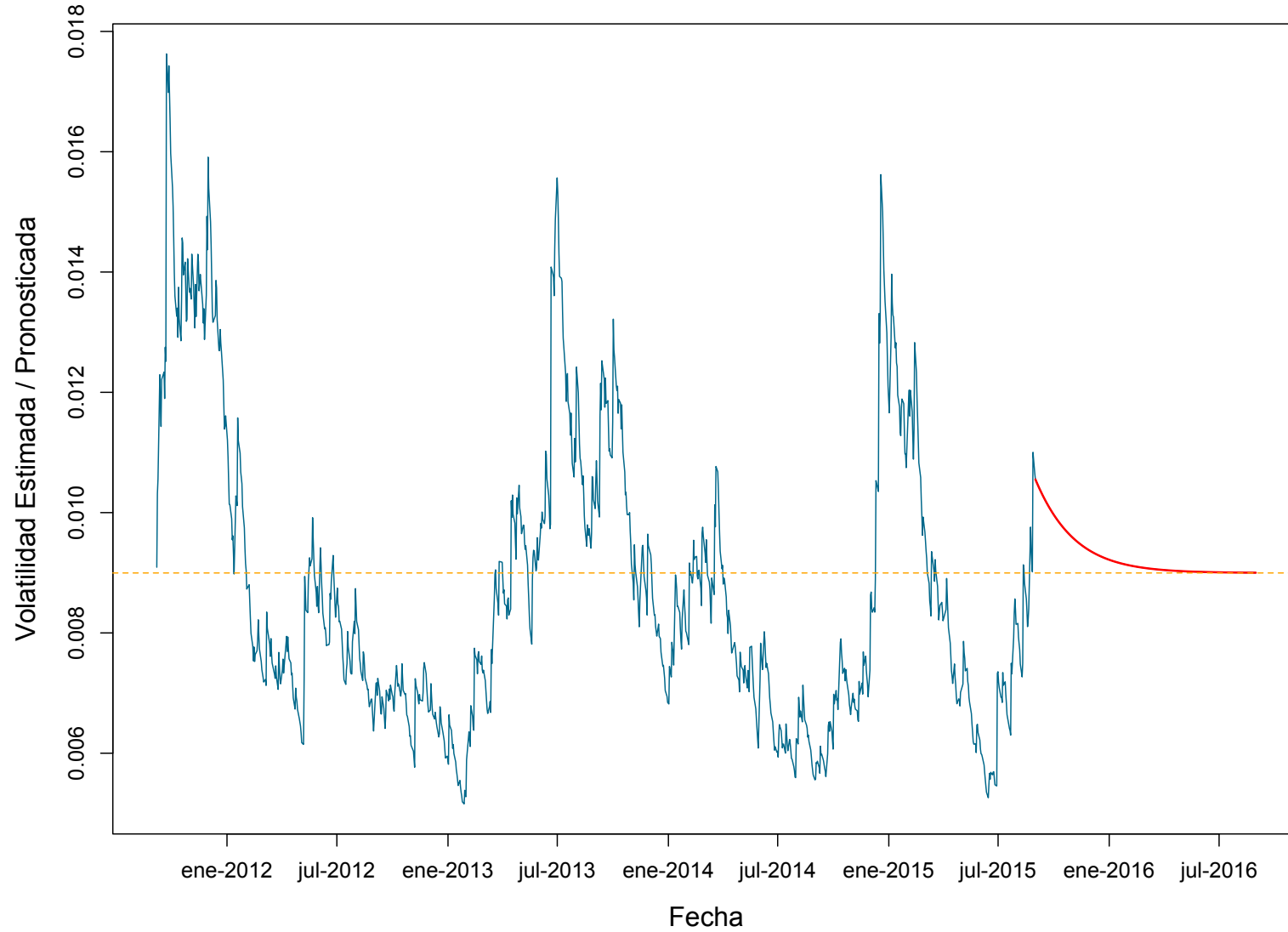


Figura 5.35: Estimación (azul) y pronóstico (rojo) de la volatilidad de la serie de log-rendimientos del Índice de Precios y Cotizaciones; también se muestra (en naranja) la estimación de la volatilidad no condicional de la serie.

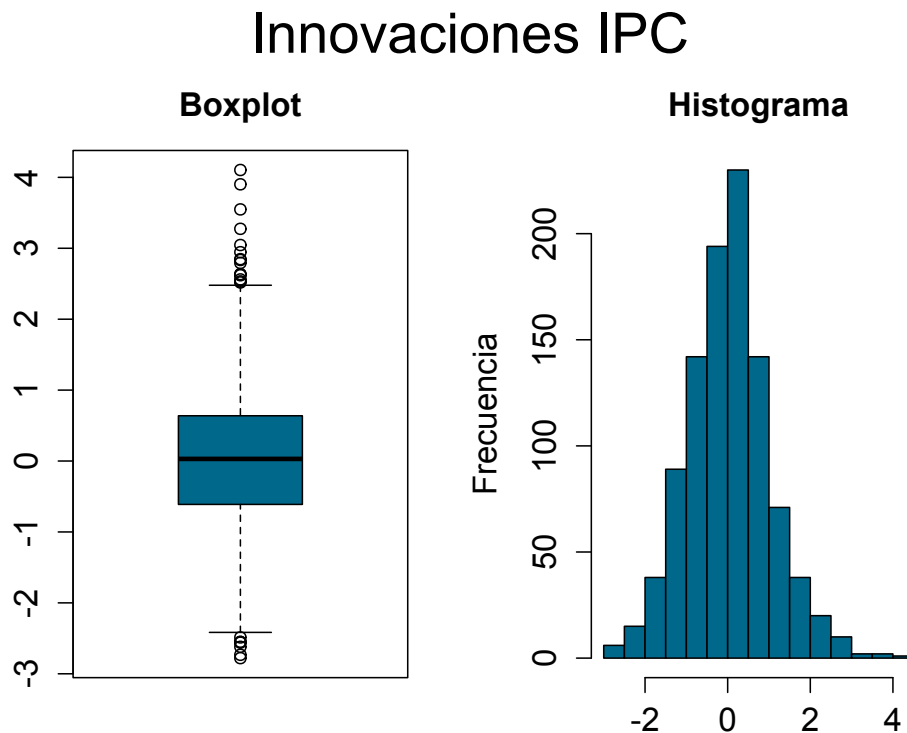


Figura 5.36: Histograma y *boxplot* de los residuos estandarizados del Índice de Precios y Cotizaciones.

5.2.3. Modelando las innovaciones del IPC

El siguiente paso de nuestro estudio consistirá en estimar tanto el VaR_α como el ES_α de las innovaciones del IPC. Dado que dichas estimaciones se realizan en el área de la cola de una distribución, en esta Sección utilizaremos una técnica de la Teoría de Valores Extremos para modelar dicha región en la distribución de las innovaciones del IPC: el método POT, el cual nos permitirá ajustar una Distribución Pareto Generalizada a la cola de las innovaciones del IPC.

Adicionalmente, compararemos la calidad de ajuste de la DPG al área de la cola de las innovaciones del IPC contra aquel que lograríamos al modelar dicha distribución mediante un par de modelos paramétricos muy conocidos: las distribuciones Normal y t -Student. No obstante, antes de iniciar dichos estudios realizaremos un análisis descriptivo de las innovaciones del IPC, a fin de conocer más a fondo su comportamiento; lo cual nos permitirá ganar intuición sobre los resultados de análisis posteriores.

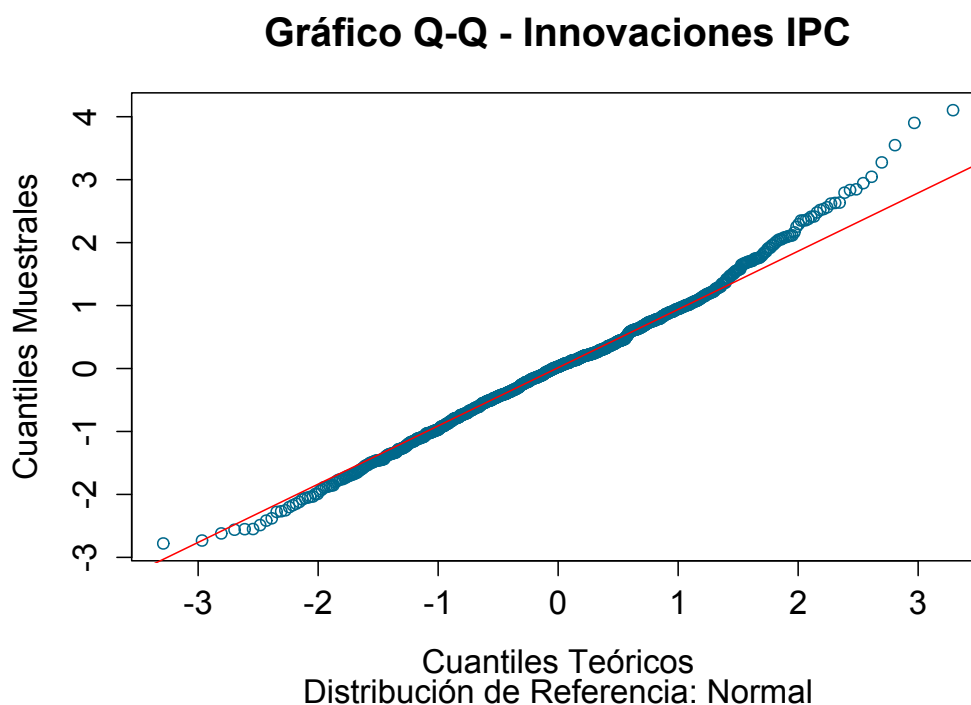


Figura 5.37: Gráfico QQ de los residuos estandarizados del Índice de Precios y Cotizaciones, contra una distribución de referencia normal. Los parámetros de la distribución normal fueron estimados vía máxima verosimilitud.

De este modo, en la Tabla 5.21 podemos observar una serie de medidas de estadística descriptiva para las innovaciones del IPC. Los valores contenidos en dicha Tabla nos permiten notar que: el valor máximo de esta serie es mayor, en valor absoluto, que el valor mínimo de la misma; la media y varianza de las innovaciones se encuentran muy cercanas a los valores 0 y 1, respectivamente, lo cual está en línea con nuestra hipótesis sobre estos datos; finalmente, el valor de la curtosis de exceso nos dice que las colas de la distribución de las innovaciones podrían ser ligeramente más pesadas que las de una distribución normal.

Una vez concluido este primer análisis sobre las características distribucionales de las innovaciones del IPC, pasamos a examinar una serie gráficos descriptivos. Comenzaremos por el histograma y el *boxplot*. Así, en la Figura 5.36 podemos observar dichos gráficos para la innovaciones del IPC.

Los gráficos en la Figura 5.36 nos permiten observar que las innovaciones del IPC están centradas en el valor cero y que la cola derecha de su distribución

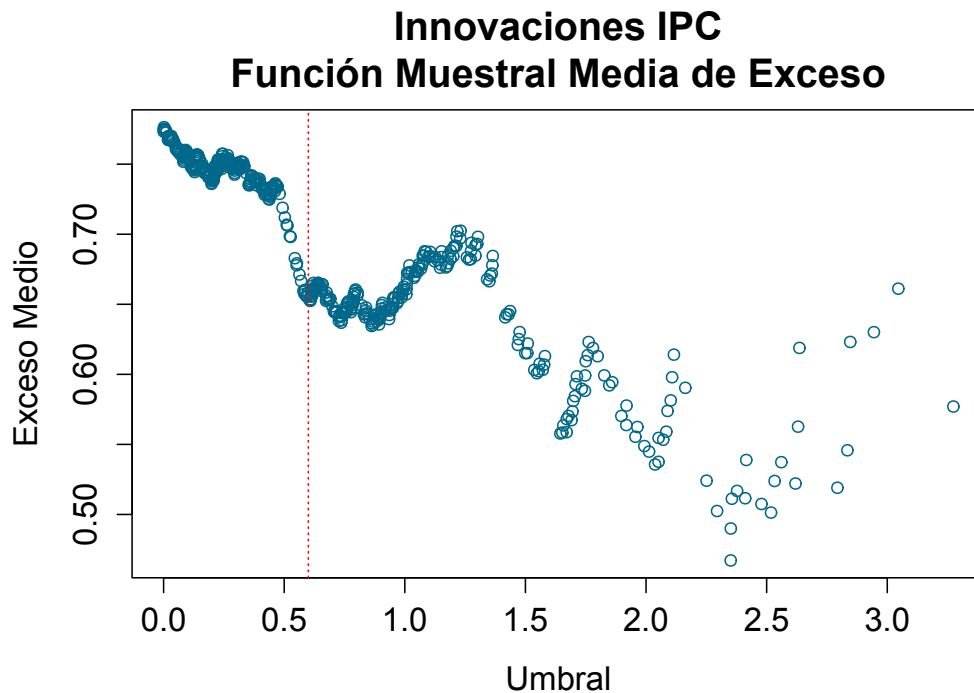


Figura 5.38: Gráfica de la función muestral media de exceso de los residuos estandarizados del Índice de Precios y Cotizaciones; la línea roja vertical muestra el umbral $u = 0.6$ a partir de cual el gráfico presenta un comportamiento aproximadamente lineal con pendiente cero.

parece ser considerablemente más larga que la izquierda. Adicionalmente, al analizar la cantidad de datos por encima y por debajo de los brazos del *boxplot*, podemos notar que la cola derecha de dicha distribución podría ser pesada, mientras que la izquierda parece ser de naturaleza ligera.

Para investigar de forma más profunda el comportamiento de las colas de las innovaciones del IPC, construimos un gráfico QQ de dichas observaciones contra una distribución de referencia Normal. Así, en la Figura 5.37 podemos observar dicho gráfico.

Al analizar la Figura 5.37 podemos notar que, en efecto, la cola derecha de las innovaciones del IPC es más pesada que la de una distribución Gaussiana. Asimismo, podemos apreciar que la cola izquierda de dichas observaciones es tan pesada como la de una distribución Normal.

Dado que nuestros análisis se realizarán sobre la cola derecha de las innovaciones del IPC, será dicha cola la que estudiaremos de forma exhaustiva.

Ajuste DPG a las Innovaciones del IPC

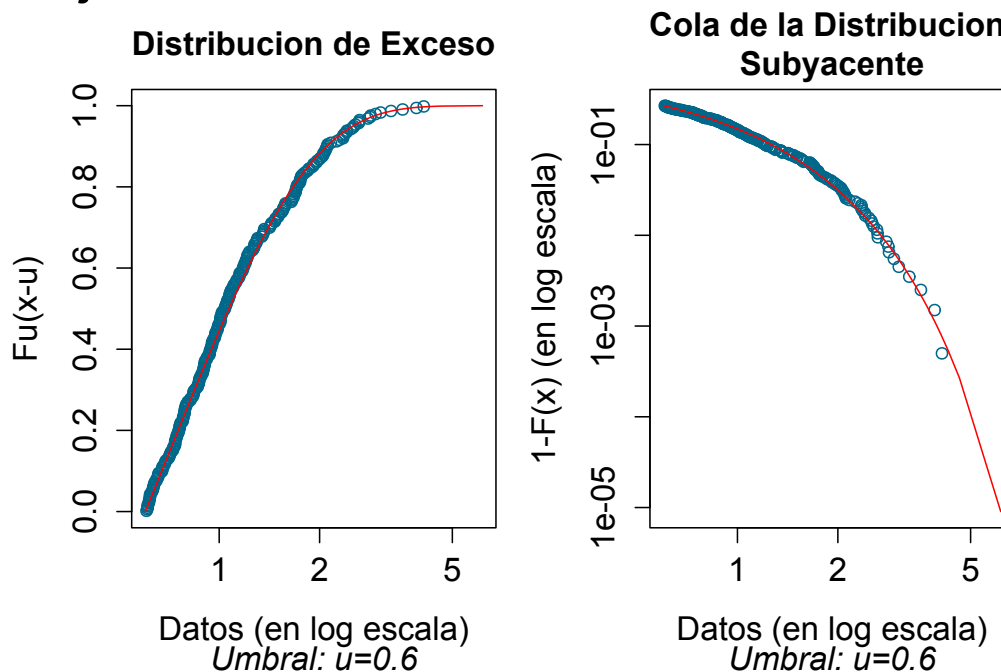


Figura 5.39: Ajuste de una Distribución Pareto Generalizada a la cola derecha de las innovaciones del IPC, para el umbral $u = 0.6$. Se muestran las curvas de la distribución de exceso (izquierda) y la cola (derecha) del modelo DPG para las innovaciones del IPC con las versiones empíricas de dichas curvas sobrepuestas de forma respectiva. El valor estimado, vía máxima verosimilitud, del parámetro de forma ξ tomó el valor -0.0495 .

No obstante, cabe mencionar que aunque ésta es más pesada que la de una distribución Normal, la diferencia no es tan marcada como para considerarla dentro del Dominio de Atracción del Máximo de la distribución Fréchet; hecho que exploraremos de forma cuidadosa más adelante.

Concluiremos nuestro análisis de gráficos descriptivos para las innovaciones del IPC con la gráfica de la función muestral media de exceso. Así, en la Figura 5.38 podemos observar dicho gráfico, construido a partir de las innovaciones con signo positivo.

Al analizar la Figura 5.38 podemos notar que la gráfica de la función muestral media de exceso parece tener una tendencia negativa. No obstante, los últimos puntos en dicha gráfica presentan gran variabilidad, debido a que los modelos en los que se basa su estimación contemplan muy pocos datos.

Ajuste DPG a las Innovaciones del IPC

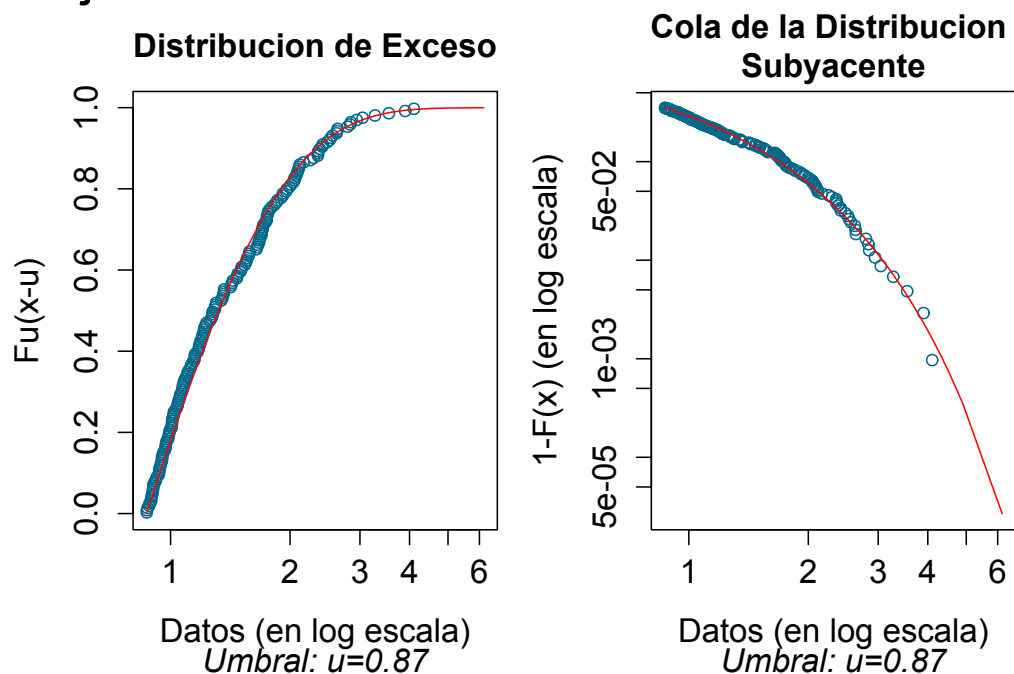


Figura 5.40: Ajuste de una Distribución Pareto Generalizada a la cola derecha de las innovaciones del IPC, para el umbral $u = 0.87$. Se muestran las curvas de la distribución de exceso (izquierda) y la cola (derecha) del modelo DPG para las innovaciones del IPC con las versiones empíricas de dichas curvas sobreimpuestas de forma respectiva. El valor estimado, vía máxima verosimilitud, del parámetro de forma ξ tomó el valor -0.0370 .

De este modo, si fijamos nuestra atención únicamente en la Sección de la gráfica comprendida entre los umbrales 0.6 y 2, podemos notar que ésta mantiene una forma aproximadamente horizontal en dicho intervalo. Por lo tanto, un comportamiento casi lineal con media cero podría describir la gráfica de la función muestral media de exceso de las innovaciones del IPC; lo cual sugiere que la distribución de tales observaciones podría pertenecer al Dominio de Atracción del Máximo de la distribución Gumbel.

Bajo las consideraciones anteriores, podríamos realizar nuestro ajuste Pareto generalizado inicial con base en un umbral $u = 0.6$, que es el valor a partir del cual comienza el comportamiento previamente descrito de la función muestral media de exceso de las innovaciones del IPC.

De esta manera, en nuestro ajuste DPG inicial se tomaron en consideración

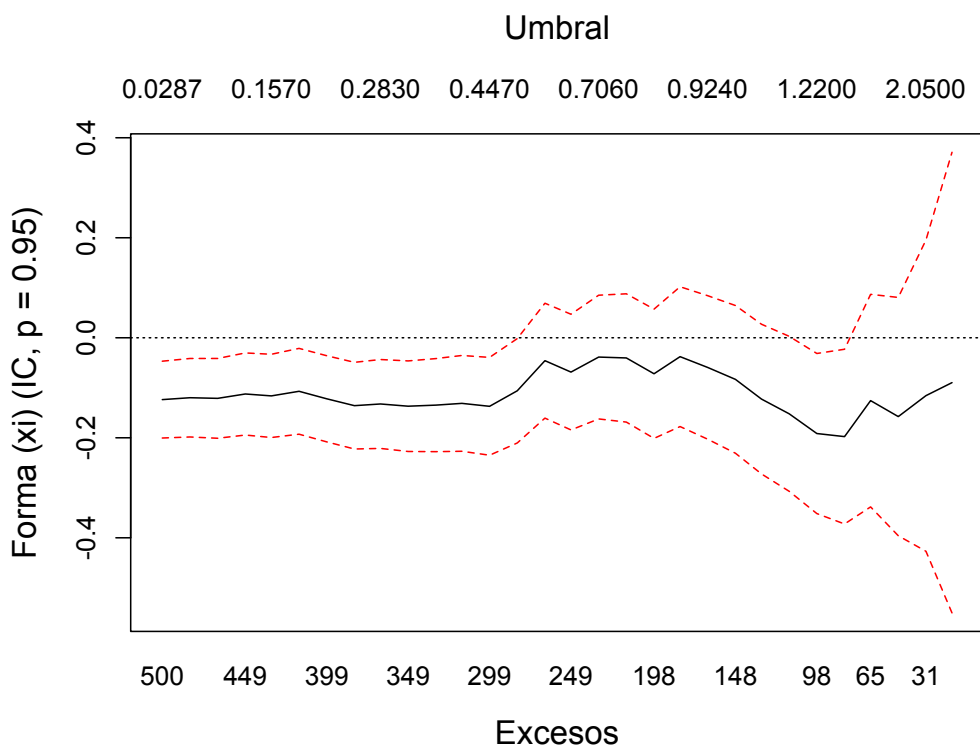


Figura 5.41: Parámetro del forma ξ de la Distribución Pareto Generalizada ajustada a la cola derecha de las innovaciones del IPC al variar el umbral en el que se basa el ajuste. En total, se ajustan 30 modelos distintos.

268 excesos al umbral $u = 0.6$. Los valores estimados de los parámetros de forma y escala de la DPG tomaron los valores $\hat{\xi} = -0.0495$ y $\hat{\beta} = 0.6887$, respectivamente. Así, notemos que el valor estimado del parámetro ξ se encuentra, en efecto, cercano al valor cero; lo cual robustece la hipótesis de la distribución de las innovaciones del IPC podría pertenecer al $DAM(H_0)$

Para evaluar la calidad de ajuste de dicho modelo DPG a las innovaciones del IPC, en la Figura 5.39 podemos observar las gráficas de las curvas de la cola y la distribución exceso de dicho modelo y, sobreimpuestas de forma respectiva, las versiones empíricas de dichas curvas. De esta manera, la Figura 5.39 nos permite apreciar que la calidad de ajuste lograda es muy satisfactoria.

No obstante, como hemos mencionado anteriormente, al realizar el ajuste de una DPG siempre es recomendable realizar análisis para distintos umbrales, a fin de obtener el mejor ajuste posible. Por lo tanto, a continuación realizaremos un segundo ajuste, esta vez con base en el umbral $u = 0.87$.

Dicho umbral fue seleccionado debido a que a partir de él gráfica de la fun-

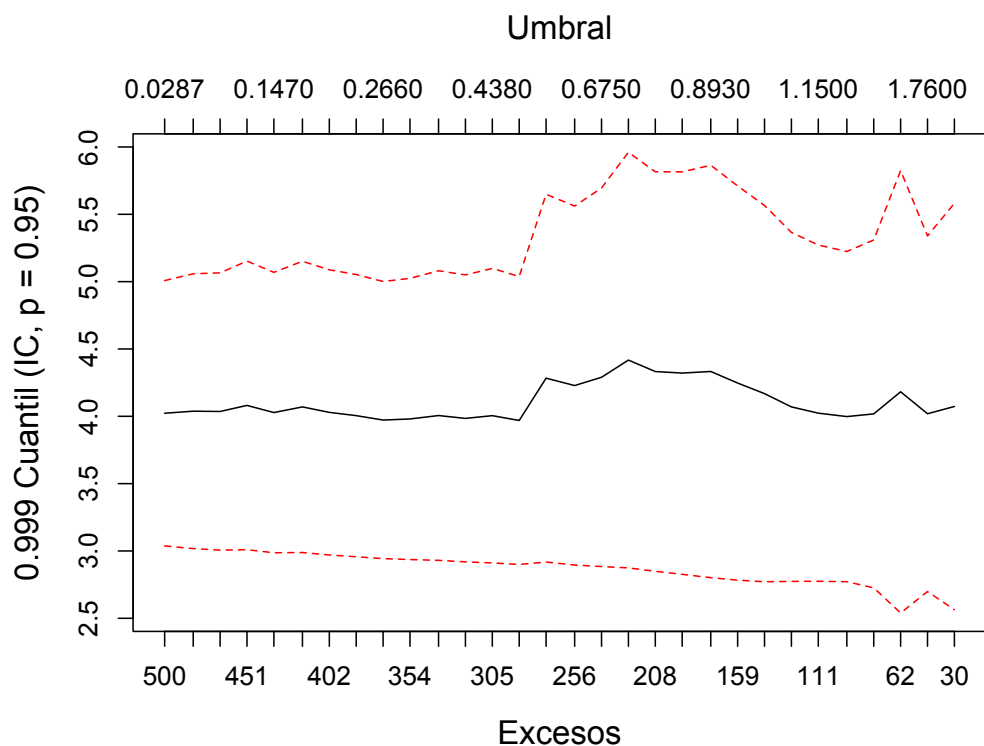


Figura 5.42: Estabilidad del cuantil 0.999 al variar el umbral en el que se basa el ajuste de la Distribución Pareto Generalizada a la cola derecha de las innovaciones del IPC. En total, se ajustan 30 modelos distintos.

ción muestral media de exceso muestra un comportamiento casi lineal con pendiente cero. Así, nuestro segundo ajuste DPG a la cola de las innovaciones del IPC se realizó sobre 182 excesos al umbral. En esta ocasión, los valores estimados de los parámetros de la DPG resultaron $\hat{\xi} = -0.037$ y $\hat{\beta} = 0.661$. Por lo tanto, de nueva cuenta el valor estimado de ξ se encuentra muy cercano al valor cero.

De esta manera, la Figura 5.40 muestra la calidad de ajuste del modelo DPG con base en el umbral $u = 0.87$. De esta forma, dicha Figura nos permite ver la calidad de ajuste obtenida vuelve a ser muy satisfactoria, a pesar de que este modelo está basado en menos excesos al umbral.

Ahora bien, dado que el parámetro más importante para la estimación de medidas de riesgo con base en una DPG es el de forma, a continuación investigaremos la estabilidad de tal cantidad ante cambios en el umbral. Asimismo, investigaremos si dicho parámetro puede ser considerado estadísticamente igual a cero para las innovaciones del IPC.

Innovaciones del IPC: Estimación del VaR_{0.999}

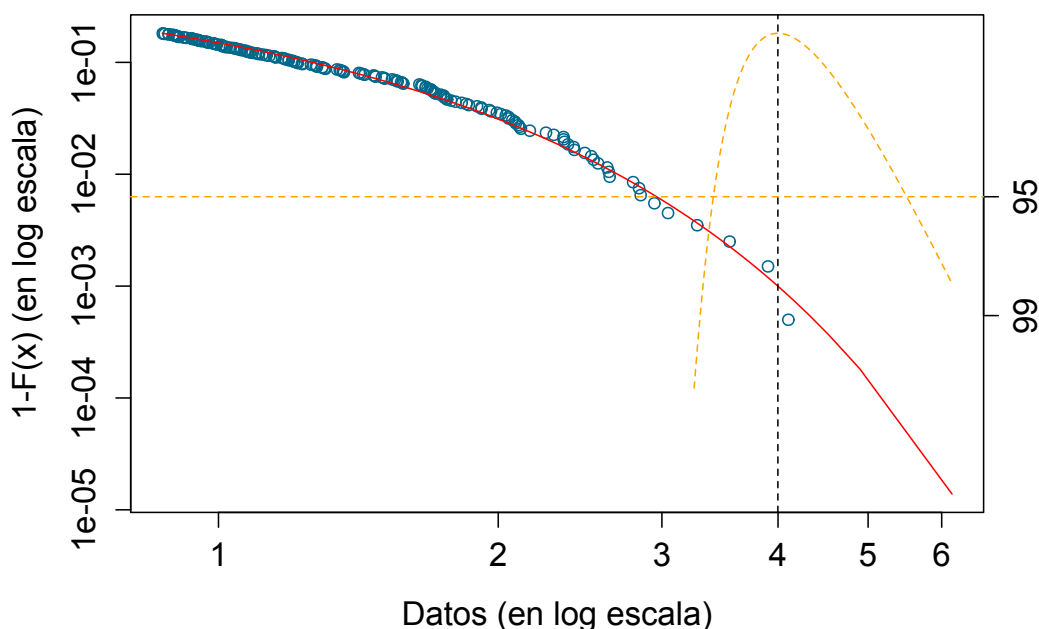


Figura 5.43: Estimación del $VaR_{0.999}$ para las innovaciones del IPC a través de un modelo DPG. El umbral en el que se basó el ajuste fue $u = 0.87$ y el valor estimado del parámetro de forma de la DPG fue de $\xi = -0.0370$.

Por lo tanto, en la Figura 5.41 podemos observar una gráfica de los niveles del parámetro ξ de modelos DPG ajustados a las innovaciones del IPC con base en distintos umbrales. Adicionalmente, dicha en dicha gráfica también se ilustran intervalos de confianza asintóticos para las estimaciones de dicho parámetro. En total, en la Figura 5.41 se contemplan 30 modelos DPG distintos.

De este modo, al analizar la Figura 5.41 podemos ver que antes del umbral $u = 0.45$, las estimaciones y extremos de los intervalos de confianza de ξ son negativos. No obstante, dado que el ajuste de la DPG a la cola de una distribución está basado en un resultado límite para umbrales altos, es altamente probable que dichas estimaciones estén sesgadas; pues están basadas en umbrales muy bajos para considerar válida la aproximación de la DPG a la cola de la distribución de estudio. Por otro lado, los ajustes basados en menos de 100 innovaciones ($u > 1.3$), presentan una alta variabilidad, debido a que están basados en muy poca información.

Bajo las dos consideraciones anteriores, la región más confiable para elegir un umbral y , y por tanto, un valor de ξ en el cual basar el ajuste DPG a la cola de las innovaciones del IPC, podría encontrarse entre las dos zonas anteriormente mencionadas. Así, al observar dicha región de la Figura 5.41, podemos notar que los extremos superiores de los intervalos de confianza de ξ al 95 % contienen el valor cero. Por lo tanto, podemos considerar que, con un 95 % de confianza, la distribución de las innovaciones del IPC pertenece al $DAM(H_0)$.

Ahora bien, a fin de investigar qué tan sensibles pueden llegar a ser las estimaciones de medidas de riesgo basadas en un modelo DPG para las innovaciones del IPC ante cambios en el umbral, en la Figura 5.42 podemos observar una gráfica de las estimaciones del $VaR_{0.999}$ obtenidas a través de modelos DPG ajustados a dicho conjunto de datos, con base en diversos umbrales. En total, en la Figura 5.42 se ajustan 30 modelos distintos.

Al analizar la Figura 5.42 podemos notar que las estimaciones del $VaR_{0.999}$ obtenidas a través modelos DPG se comportan de forma similar al parámetro ξ : presentan valores más pequeños para $u < 0.5$ y luego crecen conforme u aumenta; presentando mayor variabilidad en umbrales mayores a 1.3. No obstante, entre los umbrales $u = 0.5$ y $u = 1.3$, dichas estimaciones parecen mantenerse relativamente estables alrededor del valor 4.25.

Procederemos entonces a ilustrar el cálculo del $VaR_{0.999}$ para las innovaciones del IPC. Para ello, recordemos que los dos modelos DPG propuestos anteriormente para modelar la cola de dichos datos están basados en umbrales dentro de la zona confiable para realizar el ajuste de una DPG a la cola de las innovaciones del IPC; asimismo ambos lograron una buena calidad de ajuste. No obstante, el segundo de ellos está basado en un umbral mayor, por lo que la aproximación DPG a la cola de las innovaciones del IPC es teóricamente más válida; y por tanto, disminuye la probabilidad de que nuestras estimaciones de medidas de riesgo estén sesgadas.

De esta forma, en la Figura 5.43 podemos observar una representación gráfica del $VaR_{0.999}$; dicha estimación fue obtenida a partir del modelo DPG basado en el umbral $u = 0.87$. Así, la línea punteada vertical se intersecta con el estimador de la cola en el punto $(4.0163, 0.001)$, lo cual nos permite leer el estimador del $VaR_{0.999}$ en la coordenada x de dicho punto. Adicionalmente, la curva naranja en forma de parábola nos indica que los extremos del intervalo de confianza al 95 % para el $VaR_{0.999}$ (calculados mediante el método de verosimilitud de perfiles) son: $(3.427019, 5.518214)$.

Notemos que a pesar de que el intervalo de confianza para el $VaR_{0.999}$ de

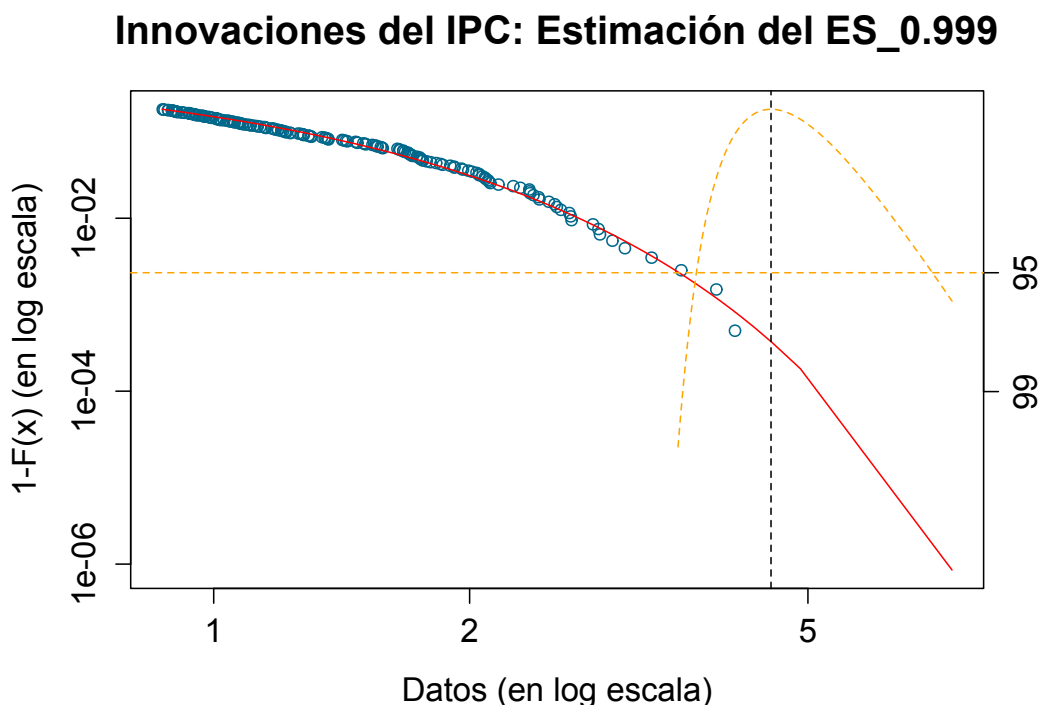


Figura 5.44: Estimación del $ES_{0.999}$ para las innovaciones del IPC a través de un modelo DPG. El umbral en el que se basó el ajuste fue $u = 0.87$ y el valor estimado del parámetro de forma de la DPG fue de $\xi = -0.0370$.

las innovaciones del IPC es asimétrico, dicha asimetría no es tan marcada como en el caso del FIX (véase Figura 5.18), lo que refleja que las colas de las innovaciones del IPC son más ligeras que aquellas de las innovaciones del FIX.

Por otro lado, podemos evaluar la calidad de nuestra estimación del $VaR_{0.999}$ de forma sencilla considerando lo siguiente: para una serie con una longitud de 1,000 observaciones esperaríamos haber observado alrededor de 1 dato por encima de dicha medida de riesgo. Por lo tanto, si nuestra estimación del $VaR_{0.999}$ es razonable, esperaríamos que cumpla con dicho criterio. En la realidad este es el caso: solo una innovación del IPC resultó mayor a nuestra estimación DPG del $VaR_{0.999}$; hecho que puede ser verificado en la Figura 5.43.

Pasamos ahora a ilustrar el cálculo del $ES_{0.999}$ para las innovaciones del IPC. Basando la estimación de dicha medida de riesgo en el modelo DPG ajustado a las innovaciones del IPC con base en el umbral $u = 0.87$ tenemos

Innovaciones IPC: parámetros		
Modelo	Parámetros	
Normal	$\mu = 0.0244$	$\sigma = 1.005$
DPG	$\xi = -0.0353$	$\beta = 0.6619$
<i>t</i> -Student	$\nu = 9.9583$	

Tabla 5.22: Estimaciones puntuales de los parámetros de los modelos ajustados a las innovaciones del modelo GARCH(1, 1) ajustado a la serie del Índice de Precios y Cotizaciones.

que $ES_{0.999} = 4.5482$. La Figura 5.44 nos permite visualizar dicha estimación, así como su intervalo de confianza al 95 %; los extremos de dicho intervalo resultaron: (3.7338, 7.0353).

De esta manera, la Figura 5.44 nos permite ver que el intervalo de confianza para el $ES_{0.999}$ también es asimétrico. No obstante, en esta ocasión la diferencia en la magnitud de dicha asimetría con respecto a la observada en la estimación del $ES_{0.999}$ para las innovaciones del FIX es considerablemente más marcada; lo cual es un reflejo de que las colas de las innovaciones del FIX resultaron considerablemente más pesadas que las de las innovaciones del IPC. Este hecho nos permite apreciar de forma práctica la diversidad en la pesadez de las colas de las distribuciones que pertenecen a las clase Gumbel; en ella pueden coexistir distribuciones cuyas colas se comportan de forma similar a las de distribuciones de colas pesadas y distribuciones cuyas colas se asemejan más bien a las de distribuciones como la Normal.

Con el análisis anterior hemos concluido nuestra aplicación del Método POT a las innovaciones del IPC. Dicha aplicación nos permitió concluir que existe una confianza del 95 % de que la distribución de las innovaciones del IPC pertenezca al Dominio de Atracción del Máximo de la distribución Gumbel; asimismo, nos permitió ajustar un modelo DPG a la cola derecha de las innovaciones del IPC y obtener así estimaciones tanto del *VaR* como del *Expected Shortfall*. No obstante, dados nuestros intereses finales, a continuación compararemos dichos resultados con aquellos que obtendríamos al modelar la distribución de las innovaciones del IPC a través de un modelo Normal o *t*-Student.

De esta forma, comenzaremos por ajustar dichos modelos a las innovaciones del IPC. Esto puede lograrse estimando los parámetros de dichas distribuciones a partir de las innovaciones del IPC. En nuestro caso, estimamos dichos parámetros vía máxima verosimilitud.

Así, en la Tabla 5.22 podemos observar los valores de los estimadores máxi-

Innovaciones GARCH del IPC: Comparación de modelos

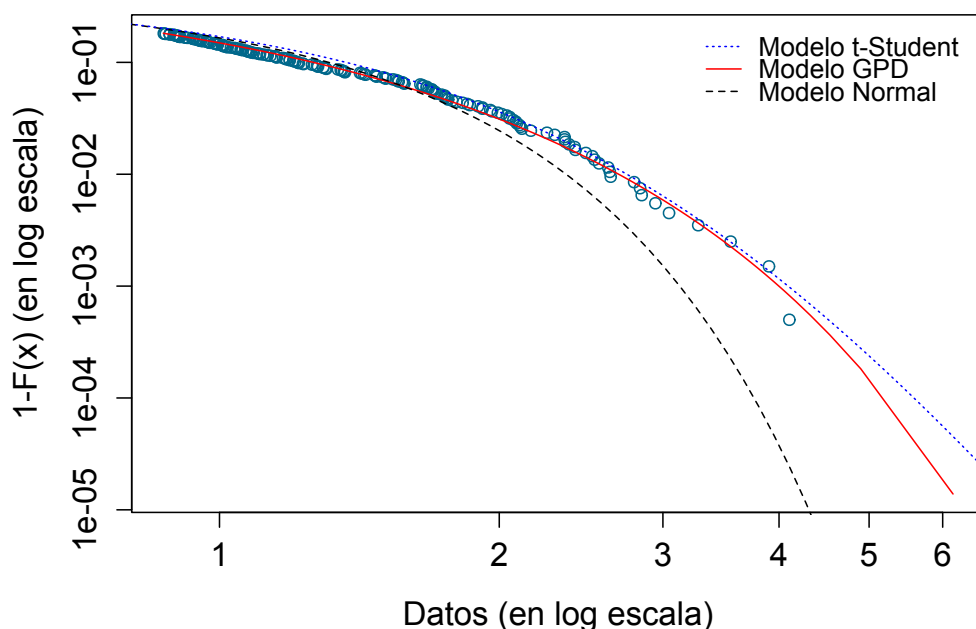


Figura 5.45: Comparación gráfica de la calidad de ajuste de diversos modelos a la cola derecha de la distribución de las innovaciones del IPC. Los modelos ajustados son: normal (línea de guiones negra), t -Student (línea azul punteada) y DPG (línea roja sólida).

mo verosímiles de los parámetros de las distribuciones Normal y t -Student; adicionalmente, se muestran los parámetros del modelo DPG ajustado a las innovaciones del IPC con base en el umbral $u = 0.87$. Los valores contenidos en la Tabla 5.22 nos reafirman que las innovaciones del IPC están centradas en cero y tienen una varianza aproximadamente igual a 1. Adicionalmente, el parámetro ν de la distribución t -Student nos refleja que la varianza de estos datos es finita; pues está alejada del valor crítico 2.

Procederemos ahora a comparar la calidad de ajuste de a las de los modelos especificados en la Tabla 5.22. Para ello, en la Figura 5.45 se muestra la calidad de ajuste que cada uno de estos modelos logró tener a la cola derecha de las innovaciones del IPC. Específicamente, se muestran la versión empírica de la cola de las innovaciones del IPC y las curvas de la cola de cada modelo ajustado sobreimpuestas.

Al analizar la Figura 5.45 podemos apreciar que:

1. El modelo normal subestima de forma importante la cola derecha de las innovaciones del IPC; lo cual nos dice que dicho modelo resulta una alternativa muy pobre para describir el comportamiento de las innovaciones más grandes del IPC.
2. El modelo t -Student parece tener una muy buena calidad de ajuste a lo largo de toda la cola de las innovaciones del IPC, por lo que representa una opción considerablemente mejor que el modelo Normal para describir el comportamiento dicha región de la distribución de las innovaciones. Adicionalmente, podemos ver que este es el modelo que presenta la colas más pesada de entre los tres considerados.
3. El modelo DPG también logró tener una calidad de ajuste excelente a la cola de las innovaciones del IPC, representando una muy buena alternativa para describir el comportamiento de las innovaciones de mayor magnitud. De hecho, las colas de los modelos DPG y t -Student presentan un comportamiento muy similar en el área donde existe información; es en la región extrapolada de la cola que dichos modelos difieren: el modelo DPG estima una cola un poco más ligera que el modelo t -Student.

De esta forma, el análisis anterior indica que los modelos DPG y t -Student representan buenas alternativas para describir el comportamiento de las innovaciones del IPC, mientras que el modelo Gaussiano presenta una tendencia importante a subestimar las observaciones más grandes de dicho conjunto de información. Esto sugiere que los modelos GARCH-DPG y GARCH- t -Student podrían modelar de forma satisfactoria la serie del IPC.

De esta forma, y a fin de avanzar hacia nuestros objetivos finales, procederemos a estimar tanto el VaR_α como el ES_α a partir de los tres modelos considerados para las innovaciones del IPC; lo cual puede ser logrado utilizando las fórmulas derivadas en los Capítulos 1 y 3, respectivamente.

En la Tabla 5.23 podemos observar las estimaciones de dichas medidas de riesgo, para los tres modelos para las innovaciones del IPC y $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$.

Al analizar los valores contenidos en la Tabla 5.23 podemos apreciar que existe una clara relación de monotonía entre las estimaciones de medidas de riesgo entregadas por los modelos Normal, DPG y t -Student. Específicamente, las estimaciones del modelo normal son menores que las del modelo DPG, las cuales a su vez son menores que aquellas del modelo t -Student.

De esta manera, podemos ver que tanto el VaR_α como el ES_α confirman

Innovaciones IPC: Medidas de Riesgo								
α	0.95		0.99		0.995		0.999	
Modelo	VaR_α	ES_α	VaR_α	ES_α	VaR_α	ES_α	VaR_α	ES_α
Normal	1.678	2.087	2.363	2.690	2.613	2.916	3.130	3.391
t -Student	1.813	2.410	2.766	3.367	3.172	3.788	4.149	4.821
DPG	1.706	2.317	2.695	3.272	3.104	3.667	4.016	4.548

Tabla 5.23: Estimaciones puntuales del VaR_α y ES_α obtenidas a partir de los modelos ajustados a las innovaciones de IPC.

la relación entre las colas de los modelos para las innovaciones que señaló nuestro análisis de la Figura 5.45.

Con la consideración anterior avanzaremos a la parte final de nuestro análisis inicial de la serie del IPC: la estimación de medidas de riesgo condicionales. Para ello, hacemos uso de las fórmulas dadas en (5.1); sustituyendo en ella los valores del pronóstico a un paso de la volatilidad del IPC (obtenido de calcular la raíz cuadrada de la expresión dada en (5.7)) y las medidas de riesgo de la Tabla 5.23.

De esta forma, en la Tabla 5.24 podemos observar los valores de nuestras estimaciones tanto del VaR_α^t como del ES_α^t , para los tres modelos para las innovaciones considerados y $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$. Puesto que las cantidades en la Tabla 5.24 son un escalamiento de aquellas reportadas en la Tabla 5.23, éstas conservan la relación de monotonía descrita en el análisis de la Tabla 5.23. Por lo tanto, podemos concluir que en nuestro análisis inicial el modelo GARCH- t -Student fue el que estimó la cola más pesada para la serie del IPC, seguido de cerca por el modelo GARCH-DPG; el modelo GARCH-Normal estimó la cola menos pesada.

Adicionalmente, podemos ver que para niveles de confianza relativamente bajos ($\alpha = 0.95$), las estimaciones de medidas de riesgo de los modelos GARCH-DPG y GARCH-Normal son más parecidas entre sí, mientras que las del modelo GARCH- t -Student son un tanto más grandes. No obstante, para niveles de confianza superiores ($\alpha > 0.95$), esta relación cambia. Específicamente, las estimaciones del VaR_α^t y ES_α^t entregadas por los modelos GARCH-DPG y GARCH- t -Student se asemejan más entre sí y aquellas entregadas por el modelo GARCH-Normal se quedan muy por debajo de ellas.

El análisis anterior nos permitió conocer la relación entre las colas que estima cada uno de los modelos considerados para la serie del IPC. Sin embargo, para saber cuál de ellos es el que mide de mejor forma el riesgo acarreado por el IPC, debemos someterlos a una evaluación. En nuestro caso, dicha

Índice de Precios y Cotizaciones: Medidas de Riesgo condicionales								
α	0.95		0.99		0.995		0.999	
Modelo	VaR_{α}^t	ES_{α}^t	VaR_{α}^t	ES_{α}^t	VaR_{α}^t	ES_{α}^t	VaR_{α}^t	ES_{α}^t
Normal	0.0181	0.0225	0.0255	0.0290	0.0282	0.0314	0.0337	0.0366
t -Student	0.0195	0.0258	0.0296	0.0359	0.0339	0.0403	0.0441	0.0511
DPG	0.0188	0.0254	0.0297	0.0346	0.0335	0.0379	0.0407	0.0440

Tabla 5.24: Estimaciones puntuales del VaR_{α}^t y ES_{α}^t obtenidas a partir del modelo GARCH(1, 1) ajustado a la serie del Índice de Precios y Cotizaciones al variar el modelo ajustado a la cola de la distribución de las innovaciones.

evaluación será histórica; utilizaremos *backtests* para realizar dicho contraste.

5.2.4. *Backtesting*

En la Sección anterior analizamos a profundidad la serie de log-rendimientos del IPC e ilustramos cómo utilizar la metodología estudiada en el Capítulo 4 para estimar medidas de riesgo condicionales a 1 periodo. Dichos análisis señalaron que los modelos GARCH-DPG y GARCH- t -Student pueden modelar de forma apropiada la serie del IPC. Por lo tanto, es probable que dichos modelos entreguen buenas estimaciones del VaR_{α}^t y ES_{α}^t .

En esta Sección investigaremos si dicha hipótesis es verdadera. Para ello, realizaremos *backtests* de las dos medidas de riesgo mencionadas. Así, los resultados de dichas evaluaciones nos permitirán conocer qué modelo logra modelar de mejor manera la serie del IPC y, en consecuencia, saber cuál de ellos es el mejor para medir el riesgo acarreado por dicha serie.

Iniciaremos entonces con una descripción de las características de los *backtests* que aplicaremos a la serie del IPC. Tomando en consideración los resultados de nuestro análisis inicial de dicha serie, en las evaluaciones siguientes utilizaremos un proceso GARCH(1,1) para modelar la dinámica del IPC; pues, como estudiamos en la Sección 5.2.2, dicho proceso logró una calidad de ajuste excelente a la serie del IPC.

Adicionalmente, para guardar consistencia con nuestros estudios iniciales del IPC, en las evaluaciones siguientes fijaremos la de nuestra ventana móvil en $n = 1,000$ observaciones. De esta forma, cada estimación del VaR_{α}^t y ES_{α}^t se realizará con aproximadamente 4 años de información diaria.

Ahora bien, en *backtests* del IPC seguiremos la misma metodología que con el Tipo de Cambio FIX. En otras palabras, para realizar el *backtest* del

VaR_α^t sustituiremos las correspondientes estimaciones del VaR_α^t o $q_\alpha(Z)$ en la fórmula dada en (5.5) y obtendremos las versiones muestrales de las indicadoras de exceso al VaR_α^t . Una vez contemos con dicha versión muestral de las indicadoras de exceso, le aplicaremos un *test* binomial y obtenemos el p-valor que nos ayudará a evaluar el desempeño de las estimaciones del VaR_α^t . Realizaremos esto para cada modelo considerado para las innovaciones, $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$ y a lo largo de toda la serie del IPC reservada para el *backtesting*.

Por otro lado, para realizar la evaluación del ES_α^t identificaremos los días en que el VaR_α^t es excedido y en dichos días calcularemos los residuales de exceso definidos en la ecuación (5.5). Así, una vez contemos con la versión muestral de dichos residuales, investigaremos si éstos se comportan como una muestra aleatoria de una distribución con media cero; lo cual puede lograrse mediante la aplicación de una prueba *bootstrap* de media cero como la descrita en el Apéndice B. Dicha evaluación se realizará para cada modelo de las innovaciones considerado y $\alpha \in \{0.95, 0.99, 0.995\}$; esto pues para niveles de confianza más altos nuestras muestras de residuales se ven reducidas a alrededor de 3 observaciones para los modelos GARCH-DPG y GARCH-*t*-Student.

Finalmente, cabe mencionar que en los *backtests* del IPC también pronosticaremos la volatilidad diaria IPC mediante la fórmula dada en (5.4) y fijaremos en $k = 100$ la cantidad de datos que consideraremos que conforman la cola de las innovaciones del IPC. Las razones para seguir este par de reglas con la serie del IPC son:

1. El *backtest* del IPC será de una longitud similar al del FIX, por lo que alguno de los dos casos que impiden el pronóstico de volatilidad con base en un GARCH podría aparecer en algunos días del *backtest* del IPC.⁴² Por lo tanto, nuestra metodología debe tomarlos en consideración y darles el tratamiento más apropiado posible.
2. La Figura 5.45 mostró que la distribución *t*-Student es una muy buena aproximación a la verdadera distribución de las innovaciones del IPC, por lo que la metodología de nuestro estudio de simulación para elección de un umbral también puede aplicarse al caso el IPC.

No obstante, a fin de ganar certeza sobre la aplicabilidad en la serie del IPC de

⁴² Específicamente, los casos que podrían surgir en algún día del *backtest* del IPC son: el parámetro α_0 del modelo GARCH(1,1) ajustado a la serie no es estadísticamente significativo o; dicho modelo no resulta estacionario en covarianza (es decir, $\alpha_1 + \beta_1 > 1$). Para mayores detalles sobre dichos casos problemáticos, véase la Sección 5.1.4.

la metodología sugerida por nuestro estudio de simulación, podemos realizar una evaluación gráfica. Así, en la Figura 5.46 podemos apreciar la calidad de ajuste de un modelo DPG a las innovaciones del IPC de nuestro análisis inicial siguiendo tal metodología; la Figura 5.46 también compara dicha calidad de ajuste con la obtenida al ajustar modelos Normal y t -Student a dichos datos, vía máxima verosimilitud.

De esta forma, la Figura 5.46 muestra que la calidad de ajuste del modelo DPG obtenida siguiendo la metodología de nuestro estudio de simulación es excelente; dicha calidad de ajuste es comparable con aquella obtenida al analizar la gráfica de la función muestral media de exceso (véase Figura 5.45). Adicionalmente, al comparar dicha calidad de ajuste con aquella de un modelo Gaussiano o t -Student, podemos observar que de nueva cuenta los modelos DPG y t -Student son los que describen de mejor forma el comportamiento de las innovaciones del IPC, mientras que el modelo Normal claramente subestima los cuantiles más grandes de dichos datos. Sin embargo, es claro que la metodología sugerida por nuestro estudio de simulación ha logrado entregar buenos resultados al aplicarse a las innovaciones del IPC, por lo que procederemos a utilizarla en la aplicación de los *backtests* siguientes.

Así, una vez definida nuestra metodología de *backtesting* para el IPC, podemos pasar a su aplicación. La ventana temporal sobre la que se realizarán los *backtests* del IPC abarcará del 29 de diciembre de 1999 al 12 de febrero de 2016; el periodo comprendido entre el 2 de enero de 1996 y el 28 de diciembre de 1999 se utilizará para realizar el ajuste del primer modelo GARCH(1,1) del *backtest*. De esta manera, la serie sobre la que aplicaremos los *backtests* del VaR_α^t y ES_α^t consistirá de 4,056 observaciones diarias de los log-rendimientos del IPC; longitud necesaria para lograr evaluar medidas de riesgo a altos niveles de confianza, como $\alpha = 0.995$ o $\alpha = 0.999$.

Resultados del backtest del VaR_α^t

En la Tabla 5.25 podemos observar las violaciones históricas y esperadas para cada estimación del VaR_α^t y los p-valores de los *tests* binomiales aplicados a las versiones muestrales de las indicadores de exceso a dicha medida de riesgo; la información se muestra clasificada por modelo para las innovaciones y para $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$.

Analizando los valores de la Tabla 5.25 podemos evaluar el comportamiento de las estimaciones del VaR_α^t . Dicha Tabla nos permite observar que:

1. Modelo GARCH-Normal:

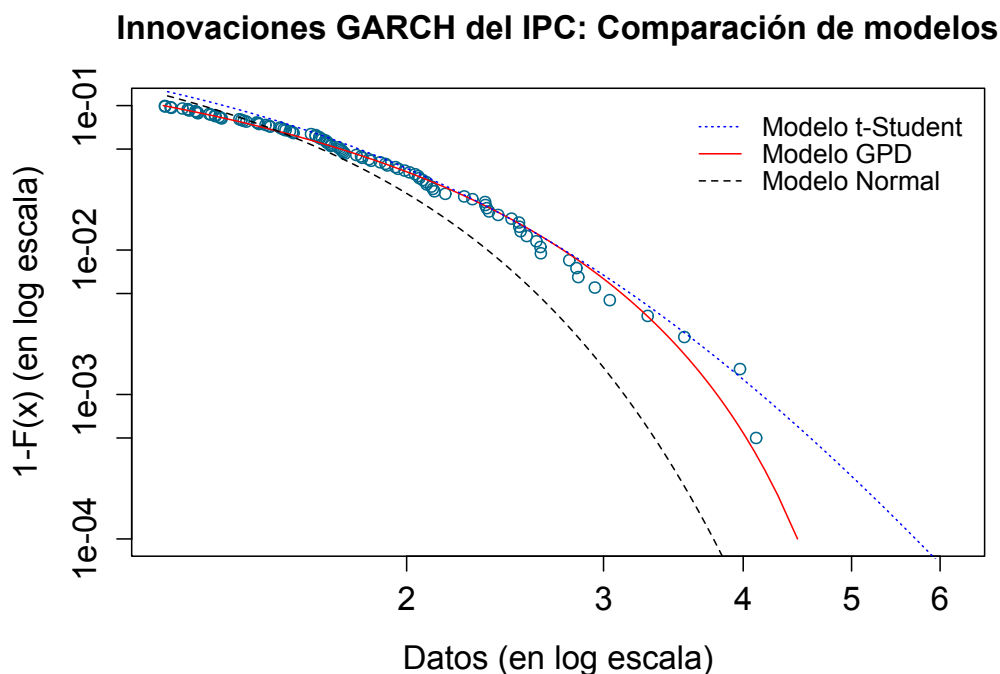


Figura 5.46: Ajuste de una DPG a la cola de las innovaciones GARCH(1,1) del IPC. El ajuste se realizó considerando que las 100 observaciones más grandes constituyen la cola de la distribución de las innovaciones del IPC.

Índice de Precios y Cotizaciones: *Backtest* del VaR_{α}^t

α	0.95		0.99	
Modelo	Viol.Hist(203)	P-valor	Viol.Hist(41)	P-valor
GARCH-Normal	201	0.9426	72	5.94×10^{-06}
GARCH- <i>t</i> -Student	154	0.0003	27	0.0326
GARCH-DPG	187	0.2642	34	0.3431
α	0.995		0.999	
Modelo	Viol.Hist(20)	P-valor	Viol.Hist(4)	P-valor
GARCH-Normal	43	9.50×10^{-06}	21	2.36×10^{-09}
GARCH- <i>t</i> -Student	14	0.1813	2	0.4535
GARCH-DPG	18	0.7377	3	0.8046

Tabla 5.25: Resultados del *backtest* de las estimaciones del VaR_{α}^t . Se muestran los p-valores de los *tests* binomiales aplicados a las muestras de indicadores de violación, las violaciones históricas a las estimaciones del VaR_{α}^t por modelo ajustado a las innovaciones del GARCH y, entre paréntesis, las violaciones esperadas al VaR_{α}^t por nivel de confianza.

- Este modelo estimó de forma apropiada el VaR_{α}^t para niveles de confianza relativamente bajos, como $\alpha = 0.95$.
- Sin embargo, para niveles de confianza superiores, este modelo entregó estimaciones muy pobres del VaR_{α}^t . Lo cual se refleja en que para todos los casos donde se consideró $\alpha > 0.95$, la hipótesis nula de los *tests* binomiales aplicados a las muestras de indicadores de exceso al $VaR_{0.95}^t$ de este modelo fue rechazada.
- Adicionalmente, el gran número de violaciones históricas al VaR_{α}^t , cuando $\alpha > 0.95$, indican que este modelo presentó una clara tendencia a subestimar de forma significativa el valor real de VaR_{α}^t .
- Por lo tanto, podemos concluir que la cola derecha de la distribución condicional de las pérdidas del IPC es considerablemente más pesada que la del modelo GARCH-Normal; lo que vuelve dicho modelo una opción muy pobre para medir el riesgo acarreado por el IPC.
- En consecuencia, utilizar un modelo GARCH-Normal para cuantificar el riesgo de mercado del IPC podría llevar a una institución financiera, que trabaje con prácticamente cualquier nivel de confianza relevante en administración de riesgos, a esperar una cantidad considerablemente menor de pérdidas a las que podrían observarse históricamente en una serie de la naturaleza del IPC.

2. Modelo GARCH-*t*-Student:

- Este modelo entregó estimaciones pobres del VaR_{α}^t cuando $\alpha = 0.95$ y $\alpha = 0.99$; para dichos niveles de confianza se rechazó la hipótesis nula de los *tests* binomiales aplicados a las indicadores de exceso al VaR_{α}^t del modelo GARCH-*t*-Student.
- El punto anterior se debe a que se tuvieron muy pocos excesos a los $VaR_{0.95}^t$ y $VaR_{0.99}^t$ estimados por el modelo GARCH-*t*-Student, lo que significa que este modelo sobrestimó el verdadero valor de dichas medidas de riesgo. En consecuencia, podemos concluir que el modelo GARCH-*t*-Student tuvo un desempeño muy pobre al estimar cuantiles relativamente bajos de la serie del IPC.
- Por otro lado, para niveles de confianza elevados ($\alpha > 0.99$), este modelo entregó estimaciones del VaR_{α}^t consistentes con lo observado históricamente en la serie del IPC; para $\alpha = 0.995$ y $\alpha = 0.999$ no se rechazó la hipótesis nula de los *tests* binomiales aplicados a las indicadores de exceso de dichas medidas de riesgo.

- Por lo tanto, puede concluirse que para niveles de confianza altos el modelo GARCH- t -Student entregó estimaciones condicionales razonables de cuantiles para la serie del Índice de Precios y Cotizaciones.
- Sin embargo, el pobre desempeño de este modelo para niveles de confianza relativamente bajos podría llevar a una institución financiera a esperar más pérdidas originadas por movimientos IPC de las que históricamente se podrían observar.

3. Modelo GARCH-DPG:

- Con este modelo, en ningún caso se rechazó la hipótesis nula de los *tests* binomiales aplicados a las indicadoras de exceso al VaR_{α}^t .
- En cada caso, las violaciones a los VaR_{α}^t obtenidas a través de este modelo fueron aproximadamente iguales a las esperadas para una serie de la longitud del IPC.
- En consecuencia, podemos concluir que al utilizar este modelo no se subestimó ni sobrestimó de forma sistemática ninguna región de la cola derecha de la distribución condicional de la serie de log-rendimientos del IPC.
- Por lo tanto, podemos concluir que el modelo híbrido GARCH-DPG permite medir de forma sensible y precisa el riesgo acarreado por la serie del IPC, para cualquier nivel de confianza relevante en la administración de riesgo de mercado.
- De esta manera, en caso de ser utilizado en la práctica, este modelo permitiría a una institución financiera prever de forma adecuada las pérdidas poco probables ocasionadas por movimientos del IPC, por lo que podrían reservar el capital necesario para administrarlas.

Procederemos entonces a ilustrar el comportamiento de nuestras estimaciones condicionales del VaR_{α}^t del IPC. Para ello, en la Figura 5.47 podemos observar gráficamente el comportamiento de dichas cantidades, cuando se trabaja con un nivel de confianza $\alpha = 0.99$ y éstas son estimadas a partir de los modelos GARCH-Normal y GARCH-DPG.

De forma específica, la Figura 5.47 muestra las estimaciones del $VaR_{.99}^t$ para la serie del IPC desde los inicios de la crisis financiera del 2008 hasta el final de la ventana temporal de nuestro estudio; en otras palabras, se muestra el comportamiento del $VaR_{.99}^t$ del 1 de julio de 2008 al 12 de febrero de 2016.

Al analizar la Figura 5.47, podemos notar que las estimaciones del $VaR_{.99}^t$ obtenidas a partir del modelo GARCH-DPG son, en general, más grandes que las obtenidas a partir del modelo GARCH-Normal. Lo cual refleja que la cola derecha del modelo GARCH-DPG es más pesada que la del modelo GARCH-Normal.

Adicionalmente, dado que el periodo considerado en la Figura 5.47 consta de 1,916 días bancarios, esperaríamos tener alrededor de 19 violaciones al $VaR_{.99}^t$. En la práctica, las estimaciones obtenidas a través del modelo GARCH-Normal fueron excedidas 38 veces, mientras que las del modelo GARCH-DPG tuvieron 17 violaciones.

Así, los hechos anteriores nos permiten apreciar de nueva cuenta las bondades que los modelos de la Teoría de Valores Extremos pueden traer a la administración dinámica de riesgo de mercado: el modelo GARCH-DPG fue capaz de entregar buenas estimaciones condicionales de cuantiles, aún en periodos que contienen tiempos de crisis. La relevancia de esto reside en que dada la alta volatilidad en el periodo en cuestión, obtener resultados tan satisfactorios no es una tarea sencilla. Por lo tanto, el modelo GARCH-DPG parece ser una herramienta considerablemente mejor para estimar dicha medida de riesgo en el IPC que el modelo GARCH-Normal; el cual, según nuestros análisis, tiende a subestimar el riesgo que puede acarrear este índice.

Finalmente, la Figura 5.47 nos permite observar de nueva cuenta el comportamiento dinámico de las estimaciones condicionales del VaR . Dichas estimaciones tienen la capacidad de adaptarse rápidamente a las condiciones del mercado, permitiendo a las instituciones financieras reservar más capital en periodos de alta volatilidad y reducirlo en épocas de relativa calma.

Resultados del *backtest* del ES_{α}^t

Pasamos ahora al análisis de los resultados de la evaluación del ES_{α}^t . De esta forma, en la Tabla 5.26 pueden observarse los p-valores de la prueba *bootstrap* de media cero aplicada a los residuales de exceso al VaR_{α}^t . Dichos p-valores se encuentran clasificados por modelo para las innovaciones y nivel de confianza.

De esta manera, al analizar los valores contenidos en la Tabla 5.26 podemos observar que:

1. GARCH-Normal

Índice de Precios y Cotizaciones: Backtest del VaR

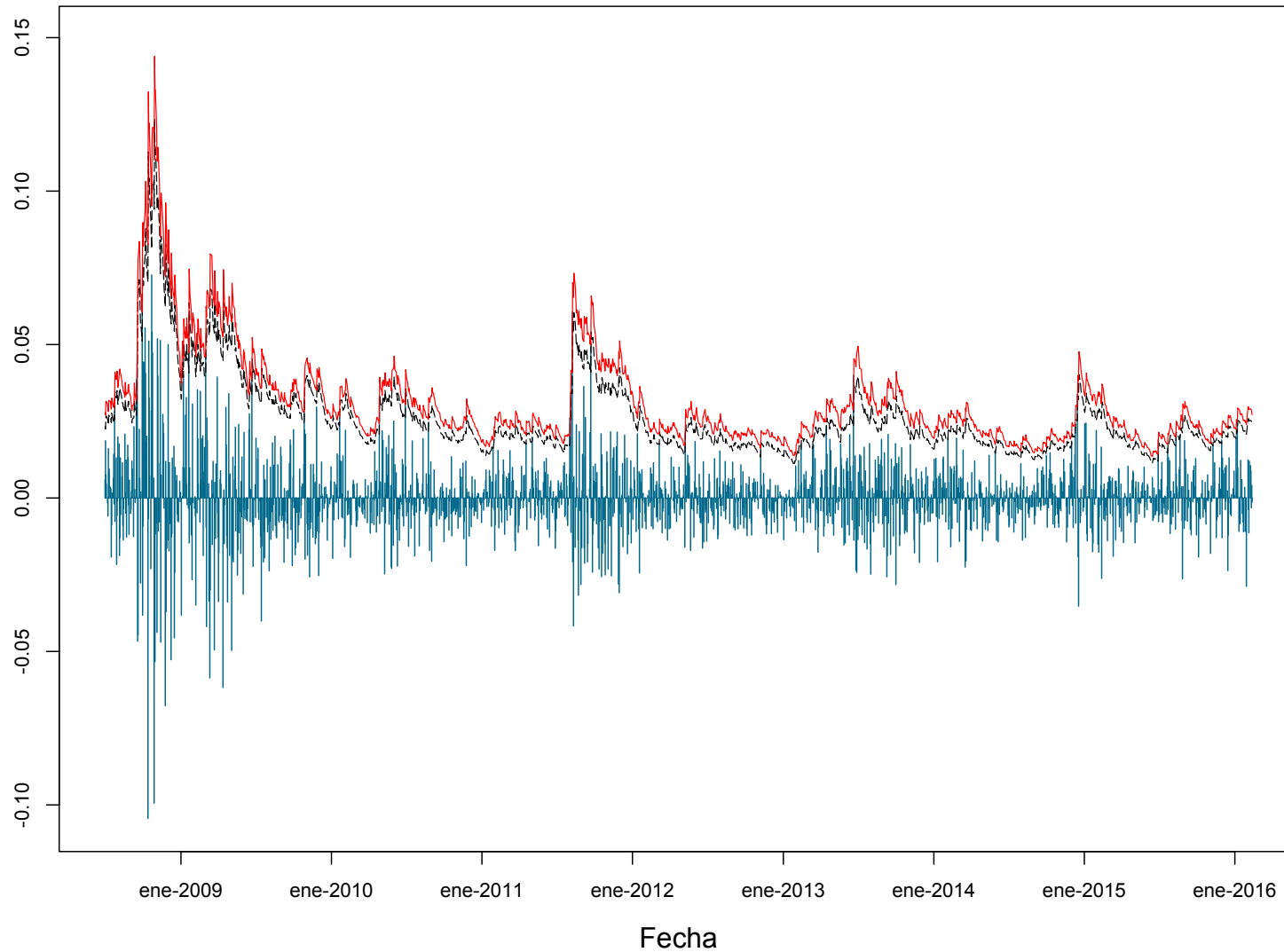


Figura 5.47: *Backtests* del $VaR_{0.99}^t$ desde la crisis financiera del 2008 hasta febrero de 2016. Se muestran las estimaciones obtenidas a partir de un modelo GARCH-Normal (línea de guiones negra) y GARCH-DPG (línea sólida roja). Las estimaciones obtenidas del modelo GARCH-Normal se encuentran por debajo de las del modelo GARCH-DPG y tienden a ser excedidas en más ocasiones debido a que la cola derecha de la distribución Gaussiana es más ligera que aquella de los residuales del IPC.

Índice de Precios y Cotizaciones: <i>Backtest</i> del ES_{α}^t			
α	0.95	0.99	0.995
Modelo	P-valor	P-valor	P-valor
GARCH-Normal	0	0.0075	0.0154
GARCH- <i>t</i> -Student	0.9197	0.9536	0.7554
GARCH-DPG	0.7949	0.7980	0.7332

Tabla 5.26: Resultados del *backtest* de las estimaciones del ES_{α}^t . Se muestran los p-valores de las pruebas *bootstrap* aplicadas a las muestras de residuales de diferencia entre X_{t+1} y ES_{α}^t en los días que el Var_{α}^t fue excedido.

- Para este modelo la hipótesis de que los residuales de exceso tengan media cero fue rechazada, para todos los niveles de confianza considerados.
- A pesar de que este modelo entregó resultados razonables en la estimación del $Var_{0.95}^t$, rechazar la hipótesis nula del *backtest* del $ES_{0.95}^t$ nos indica que este modelo subestima de forma importante lo pesado de la cola de la distribución condicional de pérdida del IPC.
- El rechazo de la hipótesis nula de la prueba *bootstrap* para las muestras asociadas a niveles de confianza $\alpha > 0.95$ confirma que este modelo subestima el riesgo de la serie de log-rendimientos del IPC para cualquier nivel de confianza relevante en la administración de riesgos.
- En consecuencia, podemos concluir que medir el riesgo del IPC con este modelo podría llevar a una compañía a reservar menos capital del necesario para enfrentar pérdidas poco probables (aunque previsibles) ocasionadas por movimientos de dicho índice; lo cual vuelve al modelo GARCH-Normal una alternativa muy pobre para cuantificar el riesgo de mercado acarreado por el IPC.

2. GARCH-*t*-Student

- No se rechazó la hipótesis nula de la prueba *bootstrap* para ninguno de los niveles de confianza considerados.
- Sin embargo, el *backtest* del Var_{α}^t nos reveló que este modelo tiende a sobrestimar el valor de tal medida de riesgo para niveles de confianza relativamente bajos ($\alpha \in \{0.95, 0.99\}$), por lo que el ES_{α}^t podría estar, de cualquier forma, erróneamente estimado

para tales niveles de confianza.

- En contraste con el modelo GARCH-Normal, este modelo podría llevar a una compañía a reservar más capital del necesario si dicha empresa calibrara sus modelos de administración de riesgo de mercado para operar con un nivel de confianza menor o igual a 0.99.
- Para niveles de confianza elevados ($\alpha > 0.99$), este modelo entregó estimaciones razonables tanto de ES_α^t como de VaR_α^t . En consecuencia, esta podría ser una alternativa razonable para cuantificar el riesgo del IPC para tales niveles de confianza.

3. GARCH-DPG

- Para este modelo no se rechazó la hipótesis nula de las pruebas *bootstrap* del *backtest* del ES_α^t , para ninguno de los niveles de confianza considerados.
- Estos resultados, en conjunto con los obtenidos en el *backtest* del VaR_α^t , nos permiten concluir que el modelo GARCH-DPG entregó buenas estimaciones de nuestras dos medidas de riesgo, para todos los niveles de confianza considerados.
- Adicionalmente, podemos concluir que este modelo no subestimó ninguna región de la cola derecha de la distribución condicional del IPC. Este fue el único modelo capaz de lograr resultados tan satisfactorios.
- Por lo tanto, y en contraste con los otros dos modelos considerados, utilizar el modelo GARCH-DPG podría permitir a una compañía reservar justamente el capital necesario para administrar las pérdidas ocasionadas por movimientos del IPC, independientemente del nivel de confianza con el que ésta trabaje. Lo cual vuelve ideal al modelo GARCH-DPG para cuantificar el riesgo de mercado de esta serie de tiempo.

Para ilustrar el comportamiento de los residuales del *backtest* del ES_α^t . En la Figura 5.48 podemos observar los residuales obtenidos de los modelos GARCH-Normal y GARCH-DPG, al trabajar con un nivel de confianza $\alpha = 0.95$, desde la crisis financiera de 2008 y hasta el final de la ventana temporal de nuestro estudio. Específicamente, se muestran los residuales obtenidos en el periodo comprendido entre el 1 de julio de 2008 y el 12 de febrero de 2016.

Dado que la longitud del periodo considerado abarcó 1,916 días bancarios,

IPC: Backtest del ES

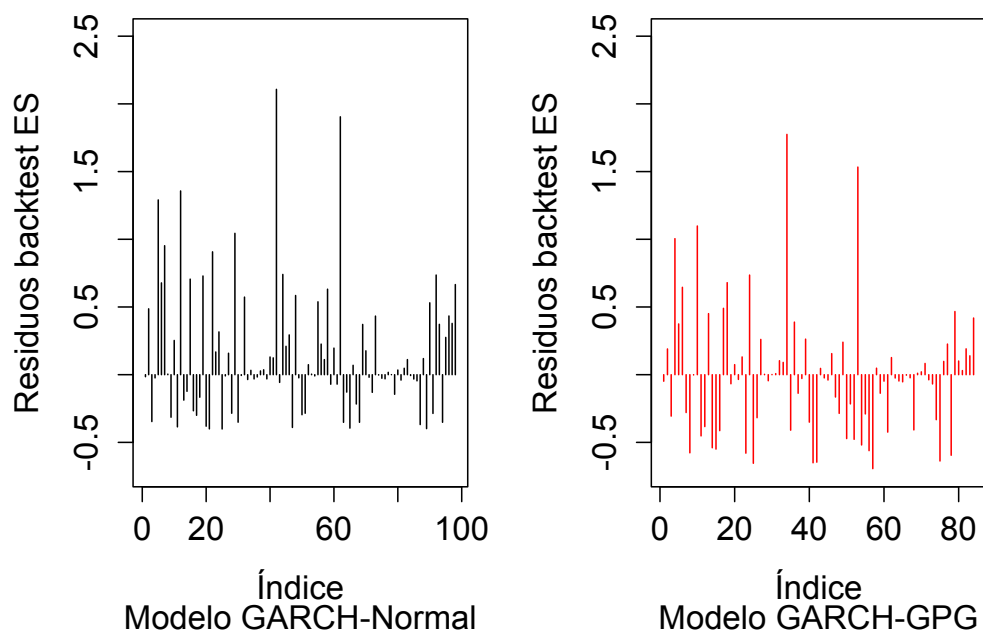


Figura 5.48: *Backtests* del $ES_{0.95}^t$ desde la crisis financiera del 2008 hasta febrero de 2016. Se muestran los residuales del *backtest* del $ES_{0.95}^t$ de los modelos GARCH-Normal (lado izquierdo, líneas negras) y GARCH-DPG (lado derecho, líneas rojas).

esperaríamos observar alrededor de 96 violaciones al $Var_{0.95}^t$. No obstante, la Figura 5.48 muestra 201 residuales para el modelo GARCH-Normal y 187 para el modelo GARCH-DPG, lo cual se debe a que dichos residuales fueron calculados solo en los días en que la estimación $Var_{0.95}^t$ de cada modelo fue excedida por el rendimiento X_{t+1} del IPC.

Así, la Figura 5.48 nos permite apreciar que la hipótesis de media cero es más plausible para los residuales obtenidos a través del modelo GARCH-DPG que para aquellos del modelo GARCH-Normal: los residuales positivos del modelo GARCH-Normal presentan en general mayores magnitudes que los del modelo DPG; adicionalmente, existe una mayor cantidad de residuales negativos para el modelo GARCH-DPG que para el modelo GARCH-Normal; finalmente, se aprecia que los residuales negativos del modelo GARCH-DPG tienen en general una mayor magnitud que aquellos del modelo GARCH-Normal.

Como evidencia adicional del hecho anterior, en el periodo contemplado en la Figura 5.48 los residuales del modelo GARCH-DPG promediaron un valor de -0.0233 , mientras que la media de los residuales del modelo GARCH-Normal ascendió a 0.1436 ; lo que significa que la media de los residuos del modelo GARCH normal fue, en valor absoluto, 617% más grande que la los residuos del modelo GARCH-DPG en dicho periodo. Por lo tanto, es claro que las estimaciones de $ES_{0.95}^t$ entregadas por el modelo GARCH-DPG lograron medir de mejor forma la pesadez de las colas y el riesgo acarreado por la serie del IPC.

Cabe mencionar que este éxito puede atribuirse a la potencia para modelar colas que otorgan las técnicas de la Teoría de Valores Extremos; en particular, el Método POT para ajustar una DPG a la cola de una distribución. La ventaja del modelo híbrido GARCH-DPG recae en que éste es capaz de combinar dicha potencia para modelar eventos extremos con el dinamismo de los modelos GARCH para capturar y describir el fenómeno de la volatilidad.

Con las observaciones anteriores hemos terminado el análisis de nuestra segunda medida de riesgo condicional, el ES_{α}^t . Siendo la conclusión de dicho análisis la que esperábamos: el modelo híbrido GARCH-DPG es una potente herramienta para estimar medidas de riesgo condicionales. Como evidencia, dicho modelo fue capaz de entregar, a lo largo de muchos años (e incluso en medio de crisis financieras) y para altos niveles de confianza, estimaciones excepcionales de medidas condicionales de riesgo en el principal indicador del Mercado Mexicano de Valores.

Conclusiones

Recordemos que al inicio de este trabajo se plantearon dos objetivos: i) evaluar si el modelo híbrido GARCH-DPG, que combina la potencia del Método POT para modelar extremos con el dinamismo de los modelos GARCH para modelar volatilidad, es capaz de entregar estimaciones razonables de medidas de riesgo condicionales a altos niveles de confianza para series de origen mexicano; y ii) comparar los resultados de dicho enfoque con aquellos entregados por la metodología que consiste en modelar las innovaciones de un GARCH ajustado a una serie financiera mediante una distribución Normal o *t*-Student.

Para realizar tal evaluación de forma práctica nos basamos en dos casos de estudio con datos mexicanos reales: el Tipo de Cambio FIX y el Índice de Precios y Cotizaciones.

Al analizar ambas series de tiempo fuimos capaces de comprobar que, en efecto, el modelo híbrido GARCH-DPG logró estimar ambas medidas de riesgo de forma excepcional, a prácticamente cualquier nivel de confianza relevante en la administración de riesgos. Más aún, pudimos comprobar que dichas estimaciones sí tuvieron un mejor desempeño que el de aquellas obtenidas a través de modelos GARCH-Normal o GARCH-*t*-Student.

En ambas series observamos un comportamiento similar por parte de los modelos GARCH-Normal y GARCH-*t*-Student. El primero logró estimar el Var_{α}^t de forma adecuada solo para niveles de confianza relativamente bajos ($\alpha = 0.95$). Para niveles de confianza más grandes, las estimaciones del Var_{α}^t obtenidas a través del modelo GARCH-Normal resultaron muy pobres: subestimaron fuertemente el verdadero valor de los cuantiles de ambas series de tiempo.

En contraste, el modelo GARCH-*t*-Student presentó una clara tendencia a sobrestimar el nivel del Var_{α}^t para niveles de confianza bajos ($\alpha \leq 0.99$). No obstante, para niveles de confianza más altos, este modelo entregó estimacio-

nes de cuantiles consistentes con la historia de ambas series de tiempo.

Por otro lado, en el caso del ES_α^t , el modelo GARCH-Normal entregó estimaciones muy pobres para todos los niveles de confianza considerados, lo cual refleja que la cola de la distribución condicional de las series de tiempo financieras de origen mexicano es considerablemente más pesada que la de modelos tipo GARCH con innovaciones Gaussianas. En consecuencia, resulta inútil tratar de cuantificar el riesgo de mercado de este tipo de series mediante procesos con innovaciones normales (normalidad condicional): la probabilidad que asignan a eventos extremos resulta estar muy por debajo de lo que se observa en la realidad.

Por su lado, los modelos GARCH- t -Student entregaron buenas estimaciones del ES_α^t para altos niveles de confianza, ($\alpha > 0.99$). No obstante, debido a que estos modelos sobrestimaron de forma importante el verdadero valor del VaR_α^t para niveles de confianza bajos, las estimaciones del ES_α^t para dichos niveles de confianza resultaron pobres, independientemente de lo observado en los *backtests* del ES_α^t : recordemos que la definición de ES_α^t es “la pérdida esperada cuando el VaR_α^t es excedido”; por lo tanto, un error de estimación en el VaR_α^t afecta de forma inmediata cualquier estimación del ES_α^t .

Finalmente, el modelo GARCH-DPG fue el único capaz de entregar estimaciones consistentes con lo observado históricamente en ambas series de tiempo, para nuestras dos medidas de riesgo y para todos los niveles de confianza considerados. Esto significa que este modelo fue el único capaz de estimar apropiadamente lo pesado de la cola derecha de la distribución condicional de pérdidas de ambas series de tiempo.

Lo anterior es particularmente relevante en el caso del *Expected Shortfall*: nuestro trabajo presenta la primera pieza de evidencia de que el modelo GARCH-DPG es capaz de estimar dicha medida de riesgo a altos niveles de confianza de forma satisfactoria para series financieras de origen mexicano. Más aún, el análisis elaborado también reveló que los modelos GARCH-Normal (para cualquier nivel de confianza) y GARCH- t -Student (para niveles de confianza debajo del 99 %) tienen un desempeño pobre al estimar dicha medida de riesgo en series mexicanas.

Adicionalmente, nuestros dos casos de estudio mostraron que el modelo híbrido GARCH-DPG es capaz de entregar estimaciones razonables del VaR_α^t y el ES_α^t tanto en tiempos de crisis, donde pérdidas de grandes magnitudes aparecen en *clusters* a lo largo de un periodo considerable de tiempo, como en épocas de relativa calma en los mercados.

En efecto, nuestros *backtests* fueron de una longitud suficiente para probar el

modelo GARCH-DPG en ambos escenarios; específicamente, contuvieron en su totalidad la crisis del 2008 y la alta volatilidad en el Tipo de Cambio FIX presentada a finales de 2015 y principios de 2016 (la cual estuvo fuertemente relacionada con la caída a nivel mundial de los precios del petróleo en dicho periodo y a la incertidumbre sobre la decisión de política monetaria de la Reserva Federal de los Estados Unidos de América).

El éxito del modelo GARCH-DPG tanto en tiempos de alta volatilidad como en aquellos de relativa calma puede atribuirse a la flexibilidad de la Distribución Pareto Generalizada para modelar la cola de una distribución: a través del signo y magnitud de su parámetro de forma se pueden modelar de forma apropiada colas pesadas, ligeras e incluso cortas.

La característica anterior permite al modelo GARCH-DPG ajustar de forma adecuada y oportuna su estimación de la pesadez de la cola de las innovaciones; lo cual termina por tener el mismo efecto en la cola de la distribución condicional de pérdida. En otras palabras, la DPG otorga al modelo la flexibilidad suficiente para adaptarse a periodos donde se presentan tanto pérdidas grandes (asociadas a innovaciones de mayores magnitudes) como moderadas (asociadas a innovaciones más bien módicas).

Al respecto, nuestros dos casos de estudio mostraron que, en fechas recientes, las innovaciones de los GARCH ajustados a series financieras mexicanas exhiben un comportamiento interesante en términos de la pesadez de sus colas: presentan características de una distribución de colas pesadas pero sus distribuciones pueden llegar a pertenecer al Dominio de Atracción del Máximo de la distribución Gumbel (con un 95% de confianza). Es decir, dichas innovaciones pueden presentar un comportamiento similar al de distribuciones como la lognormal. No obstante, es claro que dicho comportamiento puede llegar a variar a lo largo del tiempo (en épocas de crisis, es probable que la pesadez de sus colas aumente y su comportamiento obedezca más bien al de una distribución Pareto) e investigar cómo se da dicha variabilidad podría ser interesante.

Por otro lado, nuestro trabajo también confirmó que el método de selección de un umbral óptimo para efectos de *backtests* sugerido por el estudio de simulación propuesto en [3] es aplicable a series financieras de origen mexicano. Esto nos revela un hecho empírico interesante de las series financieras mexicanas: al filtrar su volatilidad a través de un GARCH, es posible considerar que la cola de las innovaciones de dicho modelo está constituida por, aproximadamente, del 10% más grande de la muestra de dichas innovaciones.

Más aún, nuestros *backtests* también muestran que utilizar un EWMA para

complementar un GARCH que modela la volatilidad de una serie de tiempo mexicana en horizontes de tiempo largos puede entregar buenos resultados.

Ahora bien, los resultados de nuestro trabajo muestran que modelo GARCH-DPG constituye, en general, una excelente herramienta para cuantificar el riesgo de mercado en series mexicanas de rendimientos diarios. Además, podemos concluir que dicho modelo puede considerarse mejor que aquellos que modelan las innovaciones de un GARCH a través de una distribución Gaussiana o *t*-Student.

El hecho anterior tiene implicaciones importantes para las instituciones financieras con exposición a series mexicanas: muestra una metodología que podría volver considerablemente más precisas sus estimaciones del riesgo de mercado; es decir, una metodología que podría volver su capital regulatorio más cercano a su capital económico.

No obstante, los resultados de esta tesis son solo un primer paso para verificar que tal objetivo es alcanzable a través de dicha metodología. En efecto, a futuro deben realizarse mayores análisis de la efectividad de este modelo para describir rendimientos de series mexicanas; en particular, podría estudiarse si los resultados de nuestro ejercicio se sostienen para rendimientos de frecuencias más bajas, como 5 o 10 días. Esto se deriva de que hoy en día el Comité de Basilea exige a los bancos calcular dichos requerimientos con base en sus *VaR* a 10 días, a un 99% de confianza.

En consecuencia, análisis como el propuesto permitirían saber si los bancos mexicanos tendrían suficientes incentivos para apegarse a las mejores prácticas en materia de administración del riesgo de mercado: podrían calcularse de forma directa los requerimientos de capital que las instituciones deben satisfacer e investigar su comportamiento al calcularse bajo una metodología sensible al riesgo como la utilizada en nuestro trabajo.

No obstante, al realizarse el ejercicio anterior no debe perderse de vista al *Expected Shortfall*, pues en el futuro las cargas de capital de las instituciones podrían llegar a depender de tal medida de riesgo, según lo establecido en el documento *Minimum capital requirements for market risk* publicado por el Comité de Basilea en enero del 2016.⁴³

Por último, se considera que estudios futuros deben explorar el comportamiento de las series mexicanas en un contexto multivariado. Esto es, se debe tomar en consideración la dependencia entre los rendimientos de activos mexicanos y llevar esta aplicación al caso de portafolios de inversión.

⁴³ Dicho documento puede consultarse en <http://www.bis.org/bcbs/publ/d352.pdf>.

No obstante, se piensa que dichos análisis no deberían basarse únicamente en medidas de asociación tradicionales, como el coeficiente de correlación de Pearson, pues éstas suelen estar basadas en desviaciones de la media que dan el mismo peso a observaciones extremas y al resto de las mismas. Más aún, la mayoría de dichas medidas de dependencia consideran únicamente dependencia lineal entre variables aleatorias.

En su lugar, el análisis de dependencia de esta clase de datos podría basarse el uso de cópulas. En particular, se podría investigar qué cópula permite una mejor estimación de medidas condicionales de riesgo para un portafolio compuesto por las series utilizadas en esta tesis (u otras series mexicanas), cuando sus distribuciones marginales son modeladas, en cada periodo, mediante un modelo GARCH-DPG (u otro más general que modele adecuadamente dichas distribuciones).

Un análisis como el anterior también debería ejecutarse en series mexicanas de rendimientos de diferentes frecuencias, como 1, 5 o 10 días hábiles bancarios. El último caso podría investigar el comportamiento de los requerimientos de capital por riesgo de mercado de una institución financiera en un contexto más próximo a la realidad.

Apéndice A

Teoría estadística complementaria

A.1. Teorema del Límite Central

El Teorema del Límite Central es uno de los resultados más notables en la teoría de la probabilidad. En pocas palabras, el teorema establece que la suma de un gran número de variables aleatorias independientes tiene una distribución que es aproximadamente normal.

Por lo tanto, este teorema no solo nos provee un método simple para calcular probabilidades aproximadas para sumas de variables aleatorias independientes, sino que nos ayuda a explicar el notable hecho de que frecuencias empíricas de muchas poblaciones naturales exhiban curvas con forma de campana (es decir, normales). En su versión más simple, el Teorema del Límite Central es como sigue.

Teorema A.1. Teorema del Límite Central. *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas, cada una con una media μ y varianza σ^2 . Entonces, la distribución de*

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

tiende a la normal estándar conforme $n \rightarrow \infty$. Esto es, para $-\infty < a < \infty$,

$$P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx, \quad \text{conforme } n \rightarrow \infty.$$

A.2. Teorema de Glivenko-Cantelli

Dentro de la teoría de la probabilidad, el Teorema de Glivenko-Cantelli determina el comportamiento asintótico de la *función de distribución empírica* conforme crece el número de variables aleatorias e idénticamente distribuidas. A continuación, se presenta el enunciado de este importante resultado.

Teorema A.2. Glivenko-Cantelli. *Sea $\{X_n\}_{n=1}^\infty$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de distribución F . Supongamos que F es desconocida y que deseamos estimarla con base en la muestra. El estimador será la función de distribución empírica, definida mediante*

$$\hat{F}_n(x) := \sum_{j=1}^n I_{\{X_j \leq x\}}.$$

Una aplicación de la Ley Fuerte de los Grandes Números nos dice que

$$\hat{F}_n(x) \rightarrow E(I_{\{X_j \leq x\}}) = F(x),$$

para todo $x \in \mathbb{R}$. Esto último es equivalente a

$$\Delta_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{c.s.} 0, \quad n \rightarrow \infty.$$

A.3. P-Valor

El p-valor se utiliza en el contexto de pruebas de hipótesis con el fin de cuantificar la noción de *significancia estadística* de cierta evidencia.⁴⁴ Las pruebas de hipótesis son un argumento de reducción al absurdo adaptado a la estadística. En esencia, una afirmación es probada válida al demostrar lo improbable de las implicaciones que se derivan de suponer lo contrario a dicha afirmación.

Por tanto, la única hipótesis que necesita ser especificada en estas pruebas, y que expresa el complemento lógico de la afirmación a probar, se conoce como la hipótesis nula. El rechazo de la hipótesis nula implica que la hipótesis correcta se encuentra en el complemento lógico de dicha hipótesis. Sin embargo, a menos que exista una única alternativa a la hipótesis nula, el rechazo de la hipótesis nula no nos dice cuál de las alternativas es la correcta.

Por ejemplo, si la hipótesis nula de una prueba dada establece que una población sigue una distribución $Normal(0, 1)$, el rechazo de esta hipótesis nula puede implicar: i) la media no es cero, ii) la varianza no es 1 o, iii) la distribución no es normal, dependiendo del tipo de prueba realizada. No obstante, incluso si se conoce que la distribución de la población es normal y que su varianza es 1, el rechazo de la hipótesis nula de media cero no nos dice el valor de media que deberíamos tomar para tal población.

En estadística, una hipótesis se refiere a la distribución de probabilidad que uno supone que gobierna los datos observados. Si X es una variable aleatoria que representa los datos observados y H es la hipótesis estadística bajo consideración, entonces la noción de significancia estadística podría ser calculada de forma natural como la probabilidad condicional $P(X | H)$, que da la probabilidad de la observación dado que la hipótesis supuesta es verdadera. No obstante, si X es una variable aleatoria continua y en la práctica observamos un valor x , entonces $P(X = x | H) = 0$. Por lo tanto, esta definición natural es inadecuada y necesita ser modificada con el fin de considerar las variables aleatorias continuas.

El p-valor se define como la probabilidad, bajo la suposición de que la hipótesis H es verdadera, de obtener un resultado igual o más extremo que lo que

⁴⁴ En el presente Apéndice solo se presenta una síntesis del concepto de p-valor, la cual supone un conocimiento básico de la noción de prueba de hipótesis y los conceptos fundamentales involucrados en su estudio. Al lector no familiarizado con tales nociones lo referimos a [18], donde podrá encontrar una discusión más detallada del concepto de p-valor.

se observó en realidad. Dependiendo de cómo se mire, “más extremo que lo que se observó en realidad” puede significar $\{X \geq x\}$ (prueba o evento de cola derecha) o $\{X \leq x\}$ (prueba o evento de cola izquierda) o el “más pequeño” de $\{X \leq x\}$ y $\{X \geq x\}$ (prueba o evento de dos colas). Por lo tanto, el p-valor puede estar dado por

- $P(X \geq x | H)$ para pruebas de cola derecha.
- $P(X \leq x | H)$ para pruebas de cola izquierda.
- $2 \min\{P(X \leq x | H), P(X \geq x | H)\}$ para pruebas de dos colas.

Mientras más pequeño sea el p-valor mayor será la significancia porque le dice al investigador que la hipótesis en cuestión no puede explicar adecuadamente la observación. La hipótesis H se rechaza si cualquiera de estas probabilidades es menor que o igual a un umbral pequeño, fijo y predefinido arbitrariamente, α , el cual es conocido como nivel de significancia. A diferencia del p-valor, el nivel de α no se deriva de los datos de observación y no depende de la hipótesis probada; más bien, el valor de α corresponde al consenso de la comunidad científica en la que trabaje el investigador que está realizando la prueba.

Puesto que el valor de x que define la prueba de cola izquierda o cola derecha es una variable aleatoria, el p-valor es en sí mismo una variable aleatoria *Uniforme* $[0, 1]$, suponiendo que x es continua. Por lo tanto, el nivel fijo y predefinido α puede interpretarse como la tasa de rechazar falsamente la hipótesis nula (o error de tipo I), pues $P(\text{Rechazar } H | H) = P(p \leq \alpha | H) = \alpha$.

Por lo general, en lugar de las observaciones crudas, X , se utiliza en cambio una estadística de prueba. Una estadística de prueba es una función escalar de todas las observaciones, tal como la media o el coeficiente de correlación, que resume las características de los datos en un solo número y que corresponde a una investigación específica. La estadística de prueba sigue por sí misma una distribución, la cual está determinada por la función que se utiliza para definir dicha estadística de prueba y la distribución de los datos utilizados para calcularla.

Por lo tanto, el cálculo de un p-valor requiere una hipótesis nula, una prueba estadística (junto con una decisión sobre si el investigador desea realizar una prueba de una cola o una prueba de dos colas), y datos. A pesar de que el cálculo de la estadística de prueba en los datos dados puede ser fácil, el cálculo de la distribución muestral bajo la hipótesis nula, y el posterior cálculo de su función de distribución es a menudo difícil. Hoy en día, este cálculo se

realiza utilizando el software estadístico, comúnmente a través de métodos numéricos (en lugar de fórmulas exactas), pero a principios y mediados del siglo XX, esto era hecho a través de tablas de valores; se solía interpolar o extrapolar p-valores a partir de dichos valores discretos.

A.4. T-Valor

Suele llamársele t-valor a los valores obtenidos al calcular una estadística t (en inglés *t-statistic*, cuyo nombre se derivó de *hypothesis test statistic*). Dicha estadística está definida como un cociente cuyo numerador corresponde a la diferencia entre el valor estimado de un parámetro y su valor poblacional hipotetizado, mientras que el denominador es el error estándar de la estadística.

De este modo, si deseamos estimar un parámetro β a partir de la estadística $\hat{\beta}$, entonces la estadística t sería calculada como

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{s.e.(\hat{\beta})},$$

donde β_0 es un valor no aleatorio conocido y $s.e.(\hat{\beta})$ denota el error estándar de la estadística $\hat{\beta}$.

El valor de elegido para β_0 nos ayuda a realizar distintos contrastes. Si elegimos $\beta_0 = 0$ podemos probar la significancia estadística de un parámetro, mientras que si elegimos $\beta_0 \neq 0$ podemos probar la hipótesis de que el parámetro β tenga un valor de β_0 para nuestra población o modelo estadístico.

La distribución del estadístico $t_{\hat{\beta}}$ dependerá del modelo e hipótesis a partir de los cuales se construya dicha estadística. Por ejemplo, en el modelo clásico de regresión lineal (es decir, aquel cuyas hipótesis incluyen residuos homocedásticos y normalmente distribuidos), cuando $\hat{\beta}$ es un estimador por mínimos cuadrados y bajo la hipótesis nula $H_0 : \beta = \beta_0$, la estadística $t_{\hat{\beta}}$ tiene una distribución t -Student con $(n - k)$ grados de libertad, donde n es el tamaño de la muestra y k el número de regresores utilizados.

Finalmente, cabe mencionar que el valor $P(> |t|)$ usualmente reportado por paquetes estadísticos representa la probabilidad de caer en la región de rechazo, la cual variará dependiendo de la distribución de $t_{\hat{\beta}}$ y el contraste planteado al construirla. Por ejemplo, si $t_{\hat{\beta}}$ sigue una distribución t -Student con k grados de libertad y una prueba de dos colas fue planteada al construir la estadística, la región de rechazo estará dada por $2(1 - P(X < t_{\hat{\beta}}))$, donde X es una variable aleatoria con distribución t -Student con k grados de libertad. En consecuencia, si dicha cantidad, $P(> |t|)$, resulta ser menor que nuestro nivel de significancia, entonces rechazamos la hipótesis nula planteada al construir la estadística $t_{\hat{\beta}}$.

Apéndice B

Desarrollo de pruebas de hipótesis

B.1. Pruebas Portmanteau

Las aplicaciones financieras usualmente requieren que seamos capaces de probar conjuntamente si varias autocorrelaciones son estadísticamente iguales a cero. Por tanto, para realizar el contraste $H_0 : \rho(1) = \dots = \rho(m) = 0$ contra $H_1 : \rho_i \neq 0$ para alguna $i \in \{1, \dots, m\}$, Box y Pierce propusieron la estadística Portmanteau:

$$Q^*(m) = T \sum_{h=1}^m \hat{\rho}(h)^2,$$

donde m es el número de *retrasos* en los que se basa la prueba y T representa el tamaño de la muestra.

Bajo la suposición de que la estadística se calcula con base en una muestra i.i.d. y bajo con ciertas condiciones para los momentos de su distribución, $Q^*(m)$ es una variable aleatoria asintóticamente ji-cuadrada con m grados de libertad.

Ljung y Box modificaron después la estadística $Q^*(m)$ para incrementar su potencia en muestras finitas. Su propuesta fue:

$$Q(m) = T(T+2) \sum_{h=1}^m \frac{\hat{\rho}(h)^2}{T-h}.$$

La regla de decisión es rechazar H_0 si $Q(m) > \chi_\alpha^2$, donde χ_α^2 es el cuantil $(1 - \alpha)$ de una distribución ji-cuadrada con m grados de libertad; la mayoría de los paquetes estadísticos proveen el p-valor de $Q(m)$. Por lo tanto, en aplicaciones se rechaza H_0 si el p-valor entregado por el paquete estadístico es menor o igual a α , el nivel de significancia con el que se esté trabajando.

En la práctica, la elección de m puede afectar el desempeño de la estadística $Q(m)$. Por lo tanto, es recomendable utilizar varios valores de m al trabajar con ella. Como guía, puede tenerse en cuenta que diversos estudios de simulación sugieren que la elección $m \approx \ln(T)$ provee un mejor desempeño en la potencia de la prueba.⁴⁵ Otro valor altamente utilizado es $m = 20$.

En nuestro caso, cuando se emplearon pruebas Portmanteau, se utilizó la estadística $Q(m)$ propuesta por Ljung y Box. Más aún, se tomó un nivel de significancia $\alpha = 0.05$ y se utilizaron tres valores distintos de m ; a saber, $\ln(T)$, 20 y 30.

⁴⁵ Véase la página 27 de [30].

B.2. Criterio de información de Akaike

El criterio de información de Akaike (AIC) es una medida de la calidad relativa de ajuste de modelos estadísticos a un conjunto dado de datos. Específicamente, dada una colección de modelos para los datos, el AIC estima la calidad de ajuste de cada modelo, en relación con cada uno de los otros modelos. Por tanto, el AIC proporciona una herramienta para la selección del *mejor* (según la medida del criterio) modelo para un conjunto de datos.

En otras palabras, el AIC ofrece una estimación relativa de la pérdida de información cuando se utiliza un determinado modelo para representar el proceso que genera los datos. En consecuencia, el AIC nos auxilia al lidiar con el *trade-off* entre la bondad de ajuste del modelo y la complejidad del mismo.

El AIC no evalúa un modelo en el sentido de probar una hipótesis nula; es decir, el AIC no es capaz de brindarnos información sobre la calidad del modelo en un sentido absoluto: esta medida no nos brindará alerta alguna si todos los modelos candidatos ajustan mal a los datos.

Definición B.1. Criterio de información de Akaike (AIC). *Supongamos que tenemos un modelo estadístico para un conjunto de datos. Sea L el valor máximo de la función de verosimilitud para el modelo y sea k el número de parámetros estimados en el modelo. Entonces, el valor AIC del modelo es:*

$$AIC = 2k - 2 \ln(n). \quad (\text{B.1})$$

Dado un conjunto de modelos candidatos para los datos, el modelo preferido es el que tiene el valor AIC mínimo. Por lo tanto AIC recompensa la bondad de ajuste de cada modelo (medida a través de la evaluación de la función de verosimilitud), pero también incluye una penalización, que es una función creciente del número de parámetros estimados. De esta forma, dicha penalización desincentiva el sobreajuste del modelo a utilizar.

Ahora bien, como mencionamos antes, la cantidad dada en (B.1) es una estimación de la información perdida al utilizar cierto modelo para representar el proceso que genera un conjunto de datos. No obstante, debe considerarse que dicha estimación es válida de forma asintótica, por lo que si el número de datos en nuestra muestra es pequeño, debe realizarse algún tipo de corrección a la cantidad (B.1) a fin de utilizarla para seleccionar modelos. Una corrección de este problema se expone a continuación.

Definición B.2. Criterio de información de Akaike corregido para muestras finitas. La corrección del AIC para muestras finitas depende de la clase de modelo estadístico ajustar. La siguiente corrección ha sido recomendada para el ajuste de series de tiempo por diversos autores, tales como [12]. A esta cantidad suele denotársele por AICC y, en nuestro caso, está dada por:

$$AICC = AIC + \frac{2k(k+1)}{n-k-1}, \quad (\text{B.2})$$

donde n es el tamaño de nuestra muestra y k denota el número de parámetros del modelo al que se está calculando el AICC.

Al utilizar esta medida el criterio de selección es el mismo que con el AIC: se prefiere el modelo que tenga el AICC mínimo.

B.3. Test binomial

Un *test* binomial es una prueba de hipótesis que compara las frecuencias observadas de las categorías de una variable dicotómica con las frecuencias esperadas en una distribución binomial con un parámetro de probabilidad especificado (p). El objetivo de la prueba es investigar si dicha variable dicotómica soporta la hipótesis de provenir de la distribución binomial de referencia.

El desarrollo de esta prueba se basa en la relación entre las distribuciones *Bernoulli* y *Binomial*. Específicamente, el contraste realizado considera que la suma de n variables aleatorias *Bernoulli*(p) sigue una distribución *Binomial*(n, p). Más aún, la relación anterior implica que si promediamos variables aleatorias *Bernoulli*(p), dicho promedio debería ser muy parecido al parámetro de dicha distribución: p . Con estas ideas en mente, a continuación se expone el desarrollo estadístico del *test* binomial.

Sea X_1, \dots, X_n una muestra aleatoria de la distribución *Bernoulli*(p). Consideremos, primeramente, realizar el contraste $H_0 : p \leq p_0$ contra $H_1 : p > p_0$, donde $0 < p_0 < 1$ es un valor específico.

El estimador máximo verosímil de p , basado en una muestra de tamaño n , está dado por $\hat{p}_n = \sum_{i=1}^n X_i/n$. Por tanto, el Teorema del Límite Central puede ser aplicado a \hat{p}_n , puesto que dicho estimador es simplemente una media muestral. En consecuencia, sabemos que, para cualquier $0 < 1 < p$, la cantidad $(p_n - p)/\sigma_n$ converge a una variable aleatoria normal estándar.

En la expresión anterior, puede tomarse $\sigma_n = \sqrt{p(1-p)/n}$, valor que claramente depende del parámetro desconocido p . Un estimador razonable de σ_n está dado por $S_n = \sqrt{\hat{p}_n(1-\hat{p}_n)/n}$.⁴⁶ De esta forma, para cualquier $0 < p < 1$, se tiene que:

$$\frac{\hat{p}_n - p}{\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}} \rightarrow N(0, 1).$$

La estadística de Wald, Z_n , es obtenida al reemplazar p por p_0 en la expresión anterior. Dicha estadística rechaza H_0 si $Z_n > z_\alpha$, donde z_α representa el cuantil α de una distribución *Normal*($0, 1$).

Ahora bien, podemos modificar la prueba anterior para realizar el contraste $H_0 : p = p_0$ vs $H_1 : p \neq p_0$. Para nuestros fines, este es el caso de mayor

⁴⁶ Puede demostrarse que el ratio σ_n/S_n converge a 1 en probabilidad. Para mayores detalles sobre esta convergencia, puede consultarse [18].

interés, pues esta es la versión del *test* binomial que se aplicó a lo largo de este trabajo.

Para realizar la evaluación anterior notemos que, por el Teorema del Límite Central, para cualquier $0 < p < 1$ tenemos:

$$\frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow N(0, 1).$$

Por lo tanto, si la hipótesis nula de nuestro nuevo contraste se satisface, tendríamos que la estadística

$$Z'_n = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1), \quad (\text{aproximadamente}).$$

De esta manera, la prueba basada en la estadística Z'_n rechaza la hipótesis nula si $|Z'_n| > z_{\alpha/2}$.

B.4. Prueba *bootstrap*

El *bootstrap* (o *bootstrapping*) es un método de remuestreo propuesto por Bradley Efron en 1979. El método es comúnmente utilizado para aproximar la distribución de un estadístico, construir intervalos de confianza o realizar contrastes de hipótesis sobre parámetros de interés.

En nuestro caso, utilizaremos el *bootstrapping* para desarrollar una prueba que nos permita contrastar la hipótesis de que la media de una población sea exactamente igual a un valor dado. En otras palabras, para una muestra x_1, x_2, \dots, x_n , deseamos responder lo siguiente: ¿Es la media de la muestra estadísticamente igual a μ_0 ?

El procedimiento *bootstrap* que implementaremos para realizar el contraste anterior consta de los pasos enseguida descritos.

Primeramente, declaramos las hipótesis nula y alternativa de la prueba:

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.$$

Después, elegimos la estadística de prueba, T , y fijamos nuestro nivel de significancia. Dado que estamos interesados en probar si la media de una población es igual a un valor específico, elegimos el promedio simple como estadística de prueba. Adicionalmente, fijamos nuestro nivel de significancia en $\alpha = 0.05$.

Bajo las consideraciones anteriores, calculamos el valor observado de la estadística de prueba para nuestros datos. Es decir, calculamos el valor

$$\hat{\mu} = T(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Necesitamos ahora el p-valor de nuestra prueba. Para obtenerlo, podríamos basarnos en el hecho de que, en virtud del Teorema del Límite Central, la distribución de nuestra estadística de prueba es aproximadamente normal. Sin embargo, dicho enfoque se sigue de un resultado asintótico, el cual podría no resultar válido si n es relativamente pequeña.

Por lo tanto, en lugar de utilizar el Teorema del Límite Central realizamos simulaciones en las que sabemos que la hipótesis nula se satisface. Para ello, definimos $y_i = x_i - \hat{\mu} + \mu_0$, con $1 \leq i \leq n$. De esta forma, tenemos que,

$$\frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu} + \mu_0) = \hat{\mu} - \hat{\mu} + \mu_0 = \mu_0,$$

por lo que la hipótesis nula se sostiene para las y_i 's.

Procedemos entonces a remuestrear de nuestro nuevo conjunto de datos $\{y_i\}_{i=1}^n$. Esto es, extraemos aleatoriamente y con reemplazo n (pues este es el tamaño de la muestra original) observaciones de la muestra y_1, y_2, \dots, y_n .

Posteriormente, calculamos el promedio de la muestra obtenida; es decir, calculamos nuestra estadística de prueba a partir de los datos remuestreados de las y_i 's. El valor obtenido será cercano a μ_0 , pues esa es la media de la población. Repetiremos este procedimiento un gran número de veces y almacenaremos los valores obtenidos de la estadística de prueba.

Ahora bien, la idea central de nuestro contraste vía *bootstrap* consiste en comparar los valores de la estadística de prueba, calculados mediante el procedimiento anterior, contra el valor obtenido a partir de la muestra original, es decir, $\hat{\mu}$. Esto nos permitirá saber si $\hat{\mu}$ es un valor consistente, o no, con la hipótesis nula.

Para obtener un p-valor, recordemos que dicha cantidad refleja la probabilidad de que la estadística de prueba elegida tome un valor más extremo que aquel obtenido al calcularla con nuestra muestra. Así, sea $\delta = |\hat{\mu} - \mu_0|$, entonces, el p-valor de esta prueba está dado por $P(T < \mu_0 - \delta) + P(T > \mu_0 + \delta)$.

Nuestro procedimiento *bootstrap* permite estimar la probabilidad anterior como

$$\frac{1}{m} \left(\sum_{i=1}^m I_{\{T_i > \mu_0 + \delta\}} + \sum_{i=1}^m I_{\{T_i < \mu_0 - \delta\}} \right), \quad (\text{B.3})$$

donde T_i denota el valor de la estadística de prueba obtenido a partir del i -ésimo remuestreo de los datos $\{y_i\}_{i=1}^n$ y m es el número de remuestreos realizados (normalmente, $m \gg n$).

Finalmente, para tomar una decisión comparamos el valor obtenido en (B.3) contra nuestro nivel de significancia, α . Si el valor obtenido en (B.3) resulta menor que α , entonces rechazamos la hipótesis nula; en caso contrario no la rechazamos y podríamos concluir que nuestros datos soportan la hipótesis de que su media sea estadísticamente igual a μ_0 .

Apéndice C

Códigos en R

Como se mencionó al inicio del Capítulo 5, todas las estimaciones, gráficas y pruebas realizadas en este trabajo fueron desarrolladas a través del *software* estadístico R.

Por lo tanto, a fin de que el lector pueda ganar una idea de cómo utilizar R para realizar un análisis como el expuesto en este trabajo, en el presente Apéndice se exponen las rutinas creadas para su elaboración. Dichas rutinas se exponen en el orden en que fueron utilizadas para crear los resultados expuestos en el Capítulo 5.

C.1. Análisis exploratorios

La función *analisis_exp* permite obtener los gráficos y medidas descriptivas empleados en nuestros análisis exploratorios. La función despliega un menú al usuario para que éste elija que elemento del análisis exploratorio desea obtener.

Los parámetros de la función son dos: x : el vector de observaciones (numéricas) al que deseamos realizar el análisis exploratorio y; *Umbral*: el umbral a partir del cual se elaborará la gráfica de la función de vida media residual.

Por defecto, la función *analisis_exp* trabaja solo con la cola derecha de los datos a explorar, no obstante, ésta puede ser fácilmente modificada (línea 23) para analizar la cola izquierda o ambas colas de las observaciones.

Las opciones que tiene el usuario al correr la función *analisis_exp* son: el histograma, el diagrama de caja y brazos, el gráfico QQ contra una distribución

de referencia Normal, la gráfica de la función muestral media de exceso y una matriz que contiene diversas medidas de tendencia central. El código de dicha función es el siguiente:

```

1 #Análisis exploratorio
2
3 analisis_explor<-function(x, umbral) {
4   library(evir) #Libreria para hacer la gráfica de la función
5     muestral media de exceso
6   library(e1071) #Libreria que alberga la función kurtosis
7   x<-as.numeric(x) #Trabajando con observaciones numéricas
8   #Enumerando las opciones que da la función
9   choices<-c("Histograma", "Boxplot", "QQ-Plot (Dist. Normal)",
10     "Función muestral media de exceso",
11     "Medidas de tendencia central")
12   tmenu<-paste("Opción:", choices) #Para desplegar el menú
13   pick <- 1 #Variable que marca la selección del usuario
14   lastcurve <- NULL
15   while (pick > 0) {
16     pick<-menu(tmenu, title="\nSelecciona una opción (o 0 para
17       salir):")
18     if(pick == 1) {
19       hist(x, ylab = "Frecuencia", xlab = "Datos", ...) }
20     if(pick == 2)
21       boxplot(x, ...)
22     if(pick == 3){
23       qqnorm(x)
24       qqline(x, col="red") }
25     if(pick == 4) {
26       meplot(x[x>0], umbral)
27       abline(v=umbral, col="red", lty="dashed") }
28     if(pick == 5){
29       print(matrix(c(min(x), max(x), median(x), mean(x),
30         var(x), kurtosis(x)),
31         nrow=1, ncol=6,
32         dimnames = list(c("Medidas"), c("Mínimo",
33           "Máximo", "Mediana", "Media", "Varianza",
34           "Ex. Curtosis")))) ) } }
35   invisible(lastcurve) }

```

C.2. Modelando la dinámica de la volatilidad de una serie de tiempo

Las funciones creadas para modelar la dinámica de la volatilidad de una serie de tiempo se resumen en 2 rutinas.

La primera rutina tiene por nombre *explora_garch* y nos ayuda a evaluar la plausibilidad de ajustar un proceso GARCH a una serie de tiempo. Los parámetros de dicha rutina son dos: *Datos*, el conjunto de observaciones numéricas de la serie de tiempo a analizar y, *Fechas*, el conjunto de fechas asociado a los datos.

De este modo, al correr la función *explora_garch* se despliega un menú al usuario, las opciones que éste tiene para elegir son: gráfica de la variabilidad de los datos; gráfica de ACF de los datos crudos, al cuadrado y en valor absoluto; p-valores de la prueba de Ljung-Box; gráfica del ACF y PACF de los datos al cuadrado. El código de dicha rutina es el siguiente:

```

1 #Evaluando la plausibilidad de ajustar un modelo GARCH a un
  conjunto de datos
2
3 explora_garch<-function(datos,fechas) {
4   #Librería para hacer las gráficas de las ACF y PACF
    muestrales
5   library(timeSeries)
6   datos<-as.numeric(datos)
7   #Asignando la fecha de cada dato
8   attr(datos,"times")<-fechas
9   op<-par() #Valores gráficos estándar
10  #Enumerando las opciones de la función
11  choices<-c("Gráfica de la variabilidad de los datos",
12            "ACF (datos crudos, al cuadrado y en
13              valor absoluto)",
14            "Prueba de Ljung-Box (datos crudos, al cuadrado
15              y en valor absoluto)",
16            "ACF y PACF de datos al cuadrado")
17  tmenu<-paste("Opción:", choices)
18  pick<-1
19  lastcurve <- NULL
20  while(pick > 0) {
21    pick<-menu(tmenu, title = "\nSelecciona una opción (o 0
22      para salir):")
23    if(pick == 1) {
24      par(op)
25      datos_plot<-timeSeries(datos,fechas)

```

```

25 A<-matrix(c(datos_plot , datos_plot ^2, abs(datos_plot)) ,
26           ncol=3, nrow=length(datos_plot))
27 datos_plot_comp<-timeSeries(A, fechas)
28 colnames(datos_plot_comp) <- c("Datos", "(Datos)^2",
29                               "|Datos|")
30 plot(datos_plot_comp, main="Variabilidad de los datos",
31       plot.type=c("multiple"), col="slategrey",
32       xlab="Fecha", at=pretty(fechas, 8), yax.flip=T,
33       format="%a-%Y") }
34 if(pick == 2){
35   par(mfrow=c(3,1), mar=c(4.08, 4.1, 3.3, 1.68))
36   acf(datos, main="Autocorrelación Datos", cex.main=0.7)
37   acf((datos)^2, main="Autocorrelación (Datos)^2", cex=1,
38       cex.main=0.7)
39   acf(abs(datos), main="Autocorrelación |Datos|",
40       cex.main=0.7) }
41 if(pick == 3) {
42   print("P-valores de la Prueba de Ljung-Box:")
43   print(matrix(c(Box.test(datos,
44                   lag=round(log(length(datos)), 0),
45                   type="Ljung")$p.value,
46                   Box.test(datos^2,
47                   lag=round(log(length(datos)), 0),
48                   type="Ljung")$p.value,
49                   Box.test(abs(datos),
50                   lag=round(log(length(datos)), 0),
51                   type="Ljung")$p.value),
52               nrow=3, ncol=1,
53               dimnames=list(c("Datos crudos",
54                               "Datos al cuadrado",
55                               "Datos en valor absoluto"),
56                             c("P-valor"))))) }
57 if(pick == 4) {
58   par(mfrow=c(1,2), oma = c(0, 0, 2, 0),
59       mar=c(5.1, 4.1, 3.1, 2.1))
60   acf((datos)^2, main="", sub="Autocorrelación")
61   pacf((datos)^2, main="", sub="Autocorrelación parcial")
62   mtext("(Datos)^2", outer = T, cex=1.7) } }
63 invisible(lastcurve) }

```

La segunda rutina de esta Sección se llama *ajusta_garch*. Dicha función tiene dos propósitos: ajustar un modelo GARCH del orden indicado por el usuario y evaluar la calidad de ajuste dicho modelo.

La función *ajusta_garch* cuenta con tres parámetros. Los dos primeros coinciden con los de la función *explora_garch* y el tercero es un vector de dos entradas llamado *orden*. A través del tercer parámetro, el usuario puede ingresar el orden del proceso GARCH que desea ajustar a los datos; la primera entrada

del vector corresponde al orden p y la segunda al q del modelo GARCH.

Las opciones que tiene el usuario al correr la función `ajusta_garch` son: graficar la volatilidad estimada (vía el GARCH ajustado) de los datos; graficar la variabilidad de las innovaciones; graficar el ACF y PACF de las innovaciones (crudas, al cuadrado y en valor absoluto); obtener los p-valores de la prueba de Ljung-Box aplicada a las innovaciones; asignar el modelo ajustado a un objeto. El código de dicha función es el siguiente:

```

1 #Evaluando la calidad de ajuste de un modelo GARCH dado
2
3 ajusta_garch<-function(datos, fechas, orden_garch=c(p,q)) {
4 #Librería para ajustar modelos GARCH a datos
5 library(fGarch)
6 orden_garch<-as.numeric(orden_garch)
7 datos<-as.numeric(datos)
8 attr(datos,"times")<-fechas
9 fit<-garchFit(formula=substitute(~garch(p,q),
10                                     list(p=orden_garch[1],
11                                           q=orden_garch[2])),
12               data=datos,cond.dist=c("QMLE"),
13               trace=F,include.mean=F) #Ajuste del GARCH
14 #Obtenemos las innovaciones del modelo
15 res<-residuals(fit,standardize=TRUE)
16 attr(res,"times")<-fechas
17 op<-par() #Valores gráficos estándar
18 #Enumerando las opciones de la función
19 choices<-c("Gráfica de volatilidad estimada de los datos",
20            "Gráfica de la variabilidad de las innovaciones",
21            "ACF y PACF de las innovaciones (crudas, al
22             cuadrado y en valor absoluto)",
23            "Prueba de Ljung-Box (innovaciones crudas, al
24             cuadrado y en valor absoluto)",
25            "Asignar el modelo a un objeto")
26 tmenu<-paste("Opción:",choices)
27 pick<-1
28 lastcurve<-NULL
29 while(pick > 0) {
30   pick<-menu(tmenu, title = "\nSelecciona una opción (o 0
31                               para salir):")
32   if(pick == 1) {
33     #Estimando la volatilidad de los datos
34     vol<-volatility(fit)
35     D<-matrix(c(datos,vol),ncol=2,nrow=length(datos))
36     vol_completo_plot<-timeSeries(D,fechas_final)
37     colnames(vol_completo_plot) <- c("Datos", "Volatilidad
38                                     estimada")

```

```

39   par(mfrow=c(1,1),oma=c(0,0,0,0),mar=c(5.1,4.1,3.1,2.1),
40       cex.axis=0.52) #Gráfica de la Serie de Tiempo
41   plot(vol_completo_plot,
42       main="Estimación GARCH de la Volatilidad de los
43       datos",
44       plot.type=c("multiple"),
45       col="slategrey", xlab="Fecha",
46       at=pretty(fechas_final,8),
47       yax.flip=T,format="%a-%Y" ) }
48   if(pick == 2) {
49     res_plot<-timeSeries(res, fechas)
50     C<-matrix(c(res_plot, res_plot^2, abs(res_plot)),
51             ncol=3,nrow=length(res_plot))
52     res_plot_comp<-timeSeries(C, fechas)
53     colnames(res_plot_comp) <- c("Innovaciones",
54                                 "(Innovaciones)^2",
55                                 "|Innovaciones|")
56     plot(res_plot_comp, main="Variabilidad de las
57           innovaciones",
58          plot.type=c("multiple"), col="slategrey",
59          xlab="Fecha", at=pretty(fechas,8), yax.flip=T,
60          format="%a-%Y" ) }
61   if(pick == 3) {
62     par(mfrow=c(3,2),mar=c(4.08,4.1,3.3,1.68))
63     acf(res, main="Innovaciones",sub="Autocorrelación")
64     pacf(res, main="Innovaciones",
65          sub="Autocorrelación parcial")
66     acf(res^2, main="(Innovaciones)^2",
67          sub="Autocorrelación")
68     pacf(res^2, main="(Innovaciones)^2",
69          sub="Autocorrelación parcial")
70     acf(abs(res), main="|Innovaciones|",
71          sub="Autocorrelación")
72     pacf(abs(res), main="|Innovaciones|",
73          sub="Autocorrelación parcial") }
74   if(pick == 4) {
75     print("P-valores de la Prueba de Ljung-Box:")
76     print(matrix(c(Box.test(res,
77                     lag=round(log(length(res)),0),
78                     type="Ljung")$p.value,
79                     Box.test(res^2,
80                     lag=round(log(length(res)),0),
81                     type="Ljung")$p.value,
82                     Box.test(abs(res),
83                     lag=round(log(length(res)),0),
84                     type="Ljung")$p.value),
85                nrow=3,ncol=1,
86                dimnames = list(c("Innovaciones crudas",
87                                "Innovaciones al

```

```
88         "cuadrado",
89         "Innovaciones en valor
90         absoluto"),
91         c("P-valor")))) }
92     if(pick == 5)
93         return(fit) }
94 invisible(lastcurve) }
```

C.3. Modelando las innovaciones de un proceso GARCH

Una vez ajustado un proceso GARCH de forma satisfactoria, el siguiente paso de nuestra aplicación es modelar las innovaciones de dicho modelo.

Para ello, primeramente exploramos de forma general dichos residuos, lo cual puede ser logrado al utilizar la función *analisis_explor* con las innovaciones como parámetro. No obstante, debemos utilizar otra función a fin de explorar las innovaciones de un GARCH con un enfoque de valores extremos. Dicha función se llama *explora_evt*.

La función *explora_evt* tiene dos parámetros: *res*, los residuos estandarizados del modelo GARCH ajustado y; *alfa*, variable que indica el cuantil a graficar en el análisis exploratorio.

Al correr la función *explora_evt* se despliega un menú al usuario, sus opciones son: graficar la función muestral media de exceso; graficar la variación del parámetro ξ al variar el umbral y; graficar la variación, ante cambios en el umbral, del cuantil *alfa* de un modelo DPG ajustado a las innovaciones. El código de la función *explora_evt* es el siguiente:

```

1  #Explorando las innovaciones del GARCH con un enfoque de
    valores extremos
2
3  explora_evt<-function(res , alfa) {
4    #Librería para trabajar con valores extremos
5    library(evir)
6    res<-as.numeric(res)
7    #Enumerando las opciones de la función
8    choices<-c("Gráfica de la función muestral media de
9              exceso",
10             "Gráfica del parámetro de forma de la DPG al
11             variar el umbral",
12             "Gráfica del cuantil alfa al variar el umbral")
13    tmenu<-paste("Opción:", choices)
14    pick<-1
15    lastcurve<-NULL
16    while(pick > 0) {
17      pick<-menu(tmenu, title = "\nSelecciona una opción (o 0
18                        para salir):")
19      if(pick == 1)
20        meplot(res [res >0])
21      if(pick == 2)
22        shape(res [res >0])

```

```

23     if(pick == 3)
24         quant(res[res > 0], p=alfa) }
25     invisible(lastcurve) }

```

Una vez exploradas las innovaciones, el siguiente paso de nuestra aplicación consistió en ajustar un modelo DPG a dichos residuos y comparar la calidad del ajuste obtenido contra aquella de las distribuciones Normal y *t*-Student.

La función *compara_modelos* nos ayuda a realizar las tareas anteriores. Dicha función cuenta con dos parámetros: *res*, las innovaciones del GARCH ajustado y; *umbral*, ya sea expresado como número (en la variable *umbral*) o como en función de la cantidad de excesos a éste (en la variable *extremos*).

Las opciones desplegadas por la función *compara_modelos* son dos: ajuste de DPG con base en el umbral dado y comparación DPG vs Normal vs *t*-Student.

Al seleccionar la primera opción se despliega un menú adicional, que ofrece 4 gráficas para evaluar la calidad de ajuste de la DPG a los datos. Estas opciones están dadas por defecto por la función *plot.gpd* de la librería *evir*. En este trabajo utilizamos solo las primeras dos opciones de dicho menú.

Por otro lado, la segunda opción de la función *compara_modelos* nos permite elaborar una gráfica que compara la calidad de ajuste de la DPG a la cola de las innovaciones analizadas contra aquella de los modelo Normal y *t*-Student ajustados por máxima verosimilitud. El código de la función *compara_modelos* es el siguiente:

```

1  #Modelando las innovaciones del GARCH
2
3  compara_modelos<-function(res , umbral=NA, extremos) {
4      library(evir)
5      res <- as.numeric(res)
6      #Verificando si el umbral se indica numéricamente
7      if(is.na(umbral))
8          gpd<-gpd(res , nextremes = extremos) #Ajustando la DPG
9      #Verificando si el umbral se determina en función de la
10     #cantidad de observaciones que lo exceden
11     else
12         gpd<-gpd(res , threshold = umbral) #Ajustando la DPG
13     #Enumerando las opciones de la función
14     choices <- c("Ajuste de DPG a con base en el umbral dado",
15                "Comparación DPG vs Normal vs t-Student")
16     tmenu <- paste("Opción:", choices)
17     pick <- 1
18     lastcurve <- NULL

```

```

19 while (pick > 0) {
20   pick <- menu(tmenu, title = "\nSelecciona una opción (o 0
21     para salir):")
22   if (pick == 1) {
23     out1<-gpd(res[res >0], threshold=umbral,
24       nextremes=extremos)
25     plot.gpd(out1) }
26   if (pick == 2) {
27     #Estimación MV de parámetros del modelo normal
28     par.norm<-fitdistr(res, "normal")$estimate
29     #Estimación MV de parámetros del modelo t-Student
30     par.t<-fitdistr(res, "t")$estimate
31     par(mfrow=c(1,1))
32     tailplot(gpd, main="Innovaciones GARCH:
33       comparación de modelos")
34     curve(1-pnorm(x, par.norm[ 'mean' ], par.norm[ 'sd' ]),
35       from=0.6089704, to=30, lty="dashed",
36       col="red", add=TRUE)
37     curve(1-pt(x, par.t[ "df" ]), from=0.6089704,
38       to=30, lty="dotted", col="blue", add=TRUE)
39     legend("topright",
40       c("Modelo GPD", "Modelo Normal", "Modelo t-Student"),
41       col=c("black", "red", "blue"),
42       lty=c("solid", "dashed", "dotted"), bty="n", cex=0.8) } }
43 invisible(lastcurve) }

```

C.4. Estimación de medidas de riesgo

Una vez modeladas la dinámica de la volatilidad de la serie de tiempo estudiada y las innovaciones del GARCH ajustado, el siguiente paso de nuestra aplicación es estimar medidas de riesgo.

Nuestra aplicación requiere que estimemos medidas de riesgo en dos etapas: primeramente, para las innovaciones del GARCH ajustado y, después, para la serie de tiempo estudiada. Por lo tanto, para estimar dichas medidas de riesgo se presentan dos rutinas, una por cada etapa mencionada.

En primer lugar, presentaremos la rutina creada para las innovaciones del GARCH ajustado. Dicha rutina consta de tres funciones. Las dos primeras, *ES.norm* y *ES.t*, nos ayudan a estimar el *Expected Shortfall* de las distribuciones Normal y *t*-Student; los valores requeridos por dichas funciones son los parámetros del modelo con el que trabajan y el nivel de confianza al que se desea calcular el ES_α .

Para calcular el *VaR* y *Expected Shortfall* de una DPG ajustada nos ayudamos de la función *riskmeasures* de la librería *evir* de R. Por otro lado, los cuantiles de las distribuciones Normal y *t*-Student pueden ser obtenidos mediante las funciones *qnorm* y *qt*, que vienen cargadas por defecto en R.

Las funciones mencionadas previamente nos permiten calcular nuestras dos medidas de riesgo, VaR_α y ES_α , a partir de los tres modelos para las innovaciones considerados en este trabajo. No obstante, a fin de sintetizar dichas estimaciones en un solo procedimiento, se creó la función *riesgo_innov*, la cual es la última función de nuestra rutina para las innovaciones.

La función *riesgo_innov* cuenta con 4 parámetros: *res*, las innovaciones del un modelo GARCH; *medida*, cadena de caracteres que indica qué medida de riesgo se desea estimar (*VaR* o *ES*); *alfa*, nivel de confianza al que se desean estimar las medidas de riesgo (puede ser un vector numérico conteniendo valores entre 0 y 1); *umbral*, el umbral en el que se basa el ajuste de la DPG a partir de la que se estimarán las medidas de riesgo para las innovaciones (expresado numéricamente o en función de la cantidad de datos que lo exceden).

La función *riesgo_innov* devuelve una matriz. Las filas de dicha matriz corresponden a los modelos a partir de los cuales se estiman el *VaR* y el *Expected Shortfall*. Sus columnas corresponden a los niveles de confianza a los que se solicita estimar dichas medidas de riesgo.⁴⁷

⁴⁷ Las funciones utilizadas para estimar medidas de riesgo para las innovaciones están

El código de las tres funciones de la rutina para estimar medidas de riesgo para las innovaciones de un GARCH es el siguiente:

```

1 #Medidas de riesgo para las innovaciones
2
3 #ES modelo Normal
4 ES.norm<-function(alpha, mu, sigma) {
5   n<-length(alpha)
6   ES<-c()
7   for(i in 1:n) {
8     ES[i]<-mu+(sigma*(dnorm(qnorm(alpha[i], mu, sigma),
9                           mu, sigma)/(1-alpha[i]))) }
10  return(ES) }
11
12 #ES modelo t-Student
13 ES.t<-function(alpha, df) {
14   n<-length(alpha)
15   ES<-c()
16   for(i in 1:n) {
17     ES[i]<-(dt(qt(alpha[i], df), df)/(1-alpha[i])) *
18             ((df+(qt(alpha[i], df))^2)/(df-1))) }
19  return(ES) }
20
21 #Medidas de riesgo para las innovaciones – Síntesis
22
23 riesgo_innov<-function(res, medida, alfa, umbral=NA, extremos) {
24   res<-as.numeric(res)
25   par.norm<-fitdistr(res, "normal")$estimate
26   par.t<-fitdistr(res, "t")$estimate
27   if(is.na(umbral))
28     gpd<-gpd(res, nextremes = extremos)
29   else
30     gpd<-gpd(res, threshold = umbral)
31   if(medida == "VaR") {
32     return(matrix(c(qnorm(alfa, par.norm["mean"],
33                       par.norm["sd"]),
34                     qt(alfa, par.t["df"]),
35                     riskmeasures(gpd, alfa)[,2]),
36                 nrow=3, col=length(alfa),
37                 dimnames=list(c("VaR, modelo Normal",
38                                 "VaR, modelo t-Student",
39                                 "VaR, modelo DPG"),
40                               as.character(alfa)), byrow=TRUE)) }
41   if(medida == "ES") {
42     return(matrix(c(ES.norm(alfa, par.norm["mean"],
43                       par.norm["sd"]),
44

```

basadas en las fórmulas derivadas en los Capítulos 1 y 3.


```

44     ES.t(alfa, par.t["df"]),
45     riskmeasures(gpd, alfa)[,3]),
46     nrow=3, ncol=length(alfa),
47     dimnames = list(c("ES, modelo Normal",
48                       "ES, modelo t-Student",
49                       "ES, modelo DPG"),
50     as.character(alfa)), byrow=TRUE)) } }

```

Ahora bien, la rutina desarrollada para estimar medidas de riesgo condicionales para la serie analizada consta de siete funciones. Las primeras seis funciones estiman el VaR_α^t y ES_α^t : se cuenta con tres funciones por medida de riesgo debido a los diferentes modelos considerados para las innovaciones.

Dichas primeras seis funciones tienen los mismos parámetros: *alfa*, nivel de confianza al que se desean estimar las medidas de riesgo (puede ser un vector numérico conteniendo valores entre 0 y 1); *fit*, objeto que contiene el ajuste GARCH realizado a la serie estudiada e; *ind*, variable dicotómica que indica cómo se estima la volatilidad: el valor 1 indica que se deben tener en consideración los casos problemáticos discutidos en el Capítulo 5, el valor 0 hace caso omiso de dichos problemas y siempre pronostica la volatilidad del día siguiente con base en el GARCH ajustado.

De esta forma, las seis primeras funciones devuelven la estimación puntual de la medida de riesgo condicional sugerida por el nombre de la función (e.g. *ES_garch_norm* devuelve la estimación puntual de ES_α^t considerando innovaciones Gaussianas).⁴⁸

De forma análoga al caso de las innovaciones, a fin de sintetizar los resultados de las primeras seis funciones de nuestra segunda rutina en un solo procedimiento, se creó la función *medidas_riesgo_condicionales*.

La función *medidas_riesgo_condicionales* tiene cuatro parámetros, tres de éstos corresponden a los requeridos por las primeras seis funciones de nuestra segunda rutina. El cuarto parámetro, *medida*, es una cadena de caracteres que indica qué medida de riesgo se desea estimar (*VaR* o *ES*).

De este modo, la función *medidas_riesgo_condicionales* devuelve una matriz. Las filas de dicha matriz corresponden a los modelos a partir de los cuales se estiman el VaR_α^t y el ES_α^t . Sus columnas corresponden a los niveles de

⁴⁸ Estas seis funciones se auxilian de las funciones utilizadas para estimar medidas de riesgo para las innovaciones de un GARCH. El objetivo de pasarles como parámetro el ajuste completo de dicho modelo es concentrar tanto los parámetros estimados del GARCH como las versiones muestrales de las innovaciones en un solo parámetro de la función.

confianza a los que se solicita estimar dichas medidas de riesgo.⁴⁹

El código de las siete funciones que comprenden la rutina para estimar medidas de riesgo condicionales es el siguiente:

```

1 #Medidas de riesgo condicionales
2
3 #VaR condicional innovaciones t
4 VaR_garch_t<-function(alfa , fit , ind) {
5   if(ind == 1) {
6     if((fit@fit$par[ 'alpha1 ']+fit@fit$par[ 'beta1 '])>1 ||
7       round(fit@fit$matcoef[2,4],4) >0.05) {
8       vol<-sqrt((fit@fit$par[ 'alpha1 ']*
9                 (fit@data[ length(fit@data)]^2))+
10                (volatility(fit)[ length(volatility(fit))]^2)*
11                (1-fit@fit$par[ 'alpha1 ']))) }
12     else
13       vol<-predict(fit ,1)$standardDeviation }
14   if(ind == 0)
15     vol<-predict(fit ,1)$standardDeviation
16
17   res<-residuals(fit , standardize = TRUE)
18   par.t<-fitdistr(res , "t")$estimate
19   nu<-par.t[ 'df' ]
20   qz<-qt(alfa , nu)
21   VaR_alfa<-qz*vol
22   return(VaR_alfa) }
23
24 #VaR condicional innovaciones Normal
25 VaR_garch_norm<-function(alfa , fit , ind) {
26   if(ind == 1) {
27     if((fit@fit$par[ 'alpha1 ']+fit@fit$par[ 'beta1 '])>1 ||
28       round(fit@fit$matcoef[2,4],4) >0.05) {
29       vol<-sqrt((fit@fit$par[ 'alpha1 ']*
30                 (fit@data[ length(fit@data)]^2))+
31                (volatility(fit)[ length(volatility(fit))]^2)*
32                (1-fit@fit$par[ 'alpha1 ']))) }
33     else
34       vol<-predict(fit ,1)$standardDeviation }
35   if(ind == 0)
36     vol<-predict(fit ,1)$standardDeviation
37
38   res<-residuals(fit , standardize = TRUE)
39   par.norm<-fitdistr(res , "normal")$estimate
40   med<-par.norm[ 'mean' ]
41   sigma<-par.norm[ 'sd' ]

```

⁴⁹ Las funciones utilizadas para estimar medidas de riesgo condicionales están basadas en las fórmulas derivadas en los Capítulos 2 y 4.

```

42   qz<-qnorm(alfa ,med, sigma)
43   VaR_alfa<-qz*vol
44   return(VaR_alfa) }
45
46 #VaR condicional innovaciones DPG
47 VaR_garch_evt<-function(alfa , fit , ind) {
48   if(ind == 1) {
49     if(( fit@fit$par[ 'alpha1 ']+ fit@fit$par[ 'beta1 '])>1 ||
50        round( fit@fit$matcoef[2 ,4] ,4) >0.05) {
51       vol<-sqrt(( fit@fit$par[ 'alpha1 ']*
52                 (fit@data[ length( fit@data)]^2))+
53                 (volatility( fit)[ length( volatility( fit))]^2)*
54                 (1- fit@fit$par[ 'alpha1 ']))) }
55     else
56       vol<-predict( fit ,1)$standardDeviation }
57   if(ind == 0)
58     vol<-predict( fit ,1)$standardDeviation
59
60   res<-residuals( fit , standardize = TRUE)
61   gpd<-gpd( res ,nextremes=100)
62   qz<-riskmeasures( gpd , alfa) [2]
63   VaR_alfa<-qz*vol
64   return(VaR_alfa) }
65
66 #ES condicional innovaciones t
67 ES_garch_t<-function( alfa , fit , ind) {
68   if(ind == 1) {
69     if(( fit@fit$par[ 'alpha1 ']+ fit@fit$par[ 'beta1 '])>1 ||
70        round( fit@fit$matcoef[2 ,4] ,4) >0.05) {
71       vol<-sqrt(( fit@fit$par[ 'alpha1 ']*
72                 (fit@data[ length( fit@data)]^2))+
73                 (volatility( fit)[ length( volatility( fit))]^2)*
74                 (1- fit@fit$par[ 'alpha1 ']))) }
75     else
76       vol<-predict( fit ,1)$standardDeviation }
77   if(ind == 0)
78     vol<-predict( fit ,1)$standardDeviation
79
80   res<-residuals( fit , standardize = TRUE)
81   par.t<-fitdistr( res , "t")$estimate
82   nu<-par.t[ 'df' ]
83   ESz<-ES.t( alfa , par.t )
84   ES_alfa<-ESz*vol
85   return(ES_alfa) }
86
87 #ES condicional innovaciones Normal
88 ES_garch_norm<-function( alfa , fit , ind) {
89   if(ind == 1) {
90     if(( fit@fit$par[ 'alpha1 ']+ fit@fit$par[ 'beta1 '])>1 ||

```



```
140         as.character(alfa)),
141         byrow=TRUE)) }
142     if(medida == "ES") {
143         return(matrix(c(ES_garch_norm(alfa, fit, ind),
144             ES_garch_t(alfa, fit, ind),
145             ES_garch_evt(alfa, fit, ind)),
146             nrow=3, ncol=length(alfa),
147             dimnames = list(c("ES, GARCH-Normal",
148                 "ES, GARCH-t-Student",
149                 "ES, GARCH-DPG"),
150             as.character(alfa)),
151             byrow=TRUE)) } }
```

C.5. *Backtests* del *VaR* y *Expected Shortfall* condicionales

La última parte de nuestra aplicación consiste en evaluar las estimaciones realizadas de medidas de riesgo condicionales, vía *backtests*. Para llevar a cabo dicho procedimiento se crearon dos rutinas en R, una por medida de riesgo a evaluar.

Expondremos primeramente la rutina que nos ayuda a realizar el *backtest* del *VaR*. Dicha rutina consta de una sola función llamada *VaR_backtest_garch*, la cual tiene cuatro parámetros: *datos*, el conjunto de observaciones numéricas de la serie de tiempo a la que se aplicará el *backtest*; *fechas*, el conjunto de fechas asociado a los datos; *n* la longitud de la ventana temporal en la que se desea basar el *backtest* (en nuestra aplicación se utilizó $n = 1,000$) e; *ind*, variable dicotómica que indica cómo se estima la volatilidad a lo largo del *backtest*: el valor 1 indica que se deben tener en consideración los casos problemáticos discutidos en el Capítulo 5, el valor 0 hace caso omiso de dichos problemas y siempre pronostica la volatilidad del día siguiente con base en el GARCH ajustado.

La función *VaR_backtest_garch* se auxilia de las funciones del apartado para estimar de forma automatizada el VaR_{α}^t , $\{0.95, 0.99, 0.995, 0.999\}$, a partir de los modelos GARCH-Normal, GARCH-*t*-Student y GARCH-DPG. Así, en función de las violaciones a dichos cuantiles, la rutina genera cuatro vectores de ceros y unos por modelo: las versiones muestrales de las indicadoras de violación al VaR_{α}^t .

Posteriormente, una vez se cuenta dichas muestras de indicadoras de violación al *VaR*, la rutina se auxilia de la función *binom.test* de R para aplicarles un *test* binomial como el descrito en el apartado .

Después, la rutina genera cinco matrices en las que se resume la información de las pruebas realizadas: cuatro conteniendo la cantidad de violaciones históricas y los p-valores y en los *tests* binomiales y una con información sobre el número de violaciones esperadas a los cuantiles estimados.

Para terminar, la función genera una lista que contiene la 5 matrices anteriores; dicha lista es el objeto que devuelve la función *VaR_backtest_garch*.

El código de la función de la rutina de este apartado es el siguiente:

```
1 #Backtest del VaR.
2
```

```

3 VaR_backtest_garch<-function(datos , fechas ,n, ind) {
4   #Indicadoras de exceso al VaR para GARCH con innovaciones
   Gaussianas (versiones muestrales de las indicadoras de
   violación al VaR)
5   ind.norm.95<-c()
6   ind.norm.99<-c()
7   ind.norm.995<-c()
8   ind.norm.999<-c()
9   #Indicadoras para GARCH con innovaciones t
10  ind.t.95<-c()
11  ind.t.99<-c()
12  ind.t.995<-c()
13  ind.t.999<-c()
14  #Indicadoras GARCH con innovaciones GPD
15  ind.evt.95<-c()
16  ind.evt.99<-c()
17  ind.evt.995<-c()
18  ind.evt.999<-c()
19  #Evitando el uso innecesario de la función length
20  m2<-(length(datos)-n)
21  #Empieza el ciclo
22  for(i in 1:(m2+1)) {
23    #Seleccionando los datos a utilizar en el i-simo ajuste
24    datosfit<-datos[i:((n-1)+i)]
25    #Seleccionando las fechas del i-ésimo ajuste
26    attr(datosfit, "times")<-fechas[i:((n-1)+i)]
27    #Realizando el i-ésimo ajuste GARCH(1,1)
28    fit.qmle<-garchFit(formula=~garch(1,1),
29                      data=datosfit, cond.dist=c("QMLE"),
30                      trace=F, include.mean=F,
31                      hessian = "ropt")
32    #Detectando violaciones al VaR
33    #Innovaciones Gaussianas
34    if((VaR_garch_norm(0.95, fit.qmle, ind) < datos[(n+i)]) &
35        ((n+i) <= length(datos)))
36      ind.norm.95[i]<-1
37    else
38      ind.norm.95[i]<-0
39    if((VaR_garch_norm(0.99, fit.qmle, ind) < datos[(n+i)]) &
40        ((n+i) <= length(datos)))
41      ind.norm.99[i]<-1
42    else
43      ind.norm.99[i]<-0
44    if((VaR_garch_norm(0.995, fit.qmle, ind) < datos[(n+i)]) &
45        ((n+i) <= length(datos)))
46      ind.norm.995[i]<-1
47    else
48      ind.norm.995[i]<-0
49    if((VaR_garch_norm(0.999, fit.qmle, ind) < datos[(n+i)]) &

```

```

50         ((n+i) <= length(datos)))
51     ind.norm.999[i]<-1
52     else
53         ind.norm.999[i]<-0
54     #Innovaciones t-Student
55     if((VaR_garch_t(0.95, fit.qmle, ind) < datos[(n+i)]) &
56         ((n+i) <= length(datos)))
57         ind.t.95[i]<-1
58     else
59         ind.t.95[i]<-0
60     if((VaR_garch_t(0.99, fit.qmle, ind) < datos[(n+i)]) &
61         ((n+i) <= length(datos)))
62         ind.t.99[i]<-1
63     else
64         ind.t.99[i]<-0
65     if((VaR_garch_t(0.995, fit.qmle, ind) < datos[(n+i)]) &
66         ((n+i) <= length(datos)))
67         ind.t.995[i]<-1
68     else
69         ind.t.995[i]<-0
70     if((VaR_garch_t(0.999, fit.qmle, ind) < datos[(n+i)]) &
71         ((n+i) <= length(datos)))
72         ind.t.999[i]<-1
73     else
74         ind.t.999[i]<-0
75     #Innovaciones DPG
76     if((VaR_garch_evt(0.95, fit.qmle, ind) < datos[(n+i)]) &
77         ((n+i) <= length(datos)))
78         ind.evt.95[i]<-1
79     else
80         ind.evt.95[i]<-0
81     if((VaR_garch_evt(0.99, fit.qmle, ind) < datos[(n+i)]) &
82         ((n+i) <= length(datos)))
83         ind.evt.99[i]<-1
84     else
85         ind.evt.99[i]<-0
86     if((VaR_garch_evt(0.995, fit.qmle, ind) < datos[(n+i)]) &
87         ((n+i) <= length(datos)))
88         ind.evt.995[i]<-1
89     else
90         ind.evt.995[i]<-0
91     if((VaR_garch_evt(0.999, fit.qmle, ind) < datos[(n+i)]) &
92         ((n+i) <= length(datos)))
93         ind.evt.999[i]<-1
94     else
95         ind.evt.999[i]<-0 }
96 n<-m2
97 #P-Valores del Test Binomial
98 #Innovaciones Gaussianas

```



```

99  pvalue.norm.95<-binom.test(sum(ind.norm.95),n,
100                               p=0.05)$p.value
101  pvalue.norm.99<-binom.test(sum(ind.norm.99),n,
102                               p=0.01)$p.value
103  pvalue.norm.995<-binom.test(sum(ind.norm.995),n,
104                               p=0.005)$p.value
105  pvalue.norm.999<-binom.test(sum(ind.norm.999),n,
106                               p=0.001)$p.value
107  #Innovaciones t-Student
108  pvalue.t.95<-binom.test(sum(ind.t.95),n,
109                               p=0.05)$p.value
110  pvalue.t.99<-binom.test(sum(ind.t.99),n,
111                               p=0.01)$p.value
112  pvalue.t.995<-binom.test(sum(ind.t.995),n,
113                               p=0.005)$p.value
114  pvalue.t.999<-binom.test(sum(ind.t.999),n,
115                               p=0.001)$p.value
116  #Innovaciones GPD
117  pvalue.evt.95<-binom.test(sum(ind.evt.95),n,
118                               p=0.05)$p.value
119  pvalue.evt.99<-binom.test(sum(ind.evt.99),n,
120                               p=0.01)$p.value
121  pvalue.evt.995<-binom.test(sum(ind.evt.995),n,
122                               p=0.005)$p.value
123  pvalue.evt.999<-binom.test(sum(ind.evt.999),n,
124                               p=0.001)$p.value
125  #Armando la matriz del VaR 0.95
126  viol.hist.95<-c(sum(ind.norm.95),sum(ind.t.95),sum(ind.evt
127  .95))
128  pvalue.95<-c(pvalue.norm.95,pvalue.t.95,pvalue.evt.95)
129  VAR.95<-matrix(c(viol.hist.95,pvalue.95),nrow=3,
130                ncol=2,byrow=FALSE,
131                dimnames = list(c("Modelo GARCH-Normal",
132                "Modelo GARCH-t-Student",
133                "Modelo GARCH-DPG"),
134                c("Violaciones Histricas al
135                VaR 0.95",
136                "P-Value Test Binomial")))
137  #Armando la matriz del VaR 0.99
138  viol.hist.99<-c(sum(ind.norm.99),sum(ind.t.99),sum(ind.evt
139  .99))
140  pvalue.99<-c(pvalue.norm.99,pvalue.t.99,pvalue.evt.99)
141  VAR.99<-matrix(c(viol.hist.99,pvalue.99),nrow=3,
142                ncol=2,byrow=FALSE,
143                dimnames = list(c("Modelo GARCH-Normal",
144                "Modelo GARCH-t-Student",
145                "Modelo GARCH-DPG"),
146                c("Violaciones Históricas al
147                VaR 0.99",

```

```

145                                     "P-Value Test Binomial"))
146 #Armando la matriz del VaR 0.995
147 viol.hist.995<-c(sum(ind.norm.995),sum(ind.t.995),sum(ind.
    evt.995))
148 pvalue.995<-c(pvalue.norm.995,pvalue.t.995,pvalue.evt.995)
149 VAR.995<-matrix(c(viol.hist.995,pvalue.995),nrow=3,
150                 ncol=2,byrow=FALSE,
151                 dimnames = list(c("Modelo GARCH-Normal",
152                                   "Modelo GARCH-t-Student",
153                                   "Modelo GARCH-DPG"),
154                                   c("Violaciones Históricas
155                                       al VaR 0.995",
156                                       "P-Value Test Binomial")))
157 #Armando la matriz del VaR 0.999
158 viol.hist.999<-c(sum(ind.norm.999),sum(ind.t.999),sum(ind.
    evt.999))
159 pvalue.999<-c(pvalue.norm.999,pvalue.t.999,pvalue.evt.999)
160 VAR.999<-matrix(c(viol.hist.999,pvalue.999),nrow=3,
161                 ncol=2,byrow=FALSE,
162                 dimnames = list(c("Modelo GARCH-Normal",
163                                   "Modelo GARCH-t-Student",
164                                   "Modelo GARCH-DPG"),
165                                   c("Violaciones Históricas
166                                       al VaR 0.999",
167                                       "P-Value Test Binomial")))
168 #Violaciones Esperadas
169 nombres<-c("0.95","0.99","0.995","0.999")
170 viol.esp1<-c(n*(1-0.95),n*(1-0.99),n*(1-0.995),n*(1-0.999))
171 viol.esp<-matrix(viol.esp1,nrow=1,ncol=4,byrow=T,
172                 dimnames = list(c("Violaciones esperadas"),
173                                   c("0.95","0.99","0.995",
174                                       "0.999")))
175 #Formamos la lista que regresar esta funcin
176 Backtest_VaR<-list(VaR.95=VAR.95,VaR.99=VAR.99,
177                   VaR.995=VAR.995,VaR.999=VAR.999,
178                   Viol.Esp=viol.esp)
179 return(Backtest_VaR) }

```

Para concluir, expondremos la rutina que nos ayuda a evaluar las estimaciones condicionales del *Expected Shortfall*. Dicha rutina consta de dos funciones.

La primera corresponde a la implementación en R de la prueba *bootstrap* descrita en el apartado B.4. Dicha función tiene por nombre *bootstrap1* y nos ayuda a evaluar si la media de una muestra es estadísticamente igual a un valor dado. En nuestra aplicación, dicho valor fue cero.

La función *bootstrap1* tiene tres parámetros: *datos*, la muestra de números a los que se aplicará la prueba; *m0*, el valor que se desea investigar si corres-

ponde a la media de la muestra y ; m , el número de remuestreos en los que se basará la prueba (en nuestra aplicación se utilizó $m = 100,000$). La función *bootstrap1* devuelve el p-valor de la prueba realizada.

La segunda función de la rutina se llama *ES_backtest_garch* y tiene los mismos parámetros que la función *VaR_backtest_garch*. Dicha función es la encargada de evaluar las estimaciones condicionales del *Expected Shortfall*.

Para ello, la función se auxilia de las funciones del apartado C.5 y de la función *bootstrap1*: con ayuda de las primeras, calcula de forma automatizada el ES_{α}^t , $\alpha \in \{0.95, 0.99, 0.995, 0.999\}$, los días que el correspondiente VaR_{α}^t es excedido; luego, calcula el correspondiente residual de exceso. Dicho procedimiento se lleva a cabo a lo largo de toda la serie estudiada, para los modelos GARCH-Normal, GARCH-*t*-Student y GARCH-DPG.

Una vez calculados todos los residuales de exceso, la función *ES_backtest_garch* utiliza la función *bootstrap1* para probar si la media de dichos residuos es estadísticamente igual a cero; esta prueba se realiza para cada nivel de confianza y modelo para las innovaciones considerados.

La función *ES_backtest_garch* devuelve una lista. Dicha lista contiene 5 matrices: las primeras cuatro, una por cada nivel de confianza considerado, contienen los p-valores de la prueba *bootstrap* media cero aplicada a los residuales de exceso; la quinta, contiene información sobre la cantidad esperada de violaciones al VaR_{α}^t para cada nivel de confianza considerado.

El código de las dos funciones de la rutina de esta Sección es el siguiente:

```

1 #Backtest del Expected Shortfall
2
3 #Función para evaluar, vía bootstrap, si la media de una
4   población es estadísticamente igual a un valor dado
5 bootstrap1<-function(datos,m0,B) {
6   #Valor observado de la estadística de prueba
7   med<-mean(datos)
8   #Definiendo los datos que satisfacen la hipótesis nula
9   nuevosdatos<-datos-mean(datos)+m0
10  bstrap<-c()
11  #Remuestrando de los nuevos datos para calcular la
12     estadística de prueba
13  for(i in 1:B) {
14    nuevamuestra<-sample(nuevosdatos,length(datos),replace=T)
15    bstrap<-c(bstrap,mean(nuevamuestra))
16  }
17  dif<-abs(m0-med)
18  a<-m0+dif
19  b<-m0-dif

```

```

17 #Estiando el p-valor vía bootstrapa
18 pvalor<-(sum(bstrap>a)+sum(bstrap<b))/B
19 return(pvalor) }
20
21 #Función para evaluar las estimaciones del Expected Shortfall
    condicional
22 ES_backtest_garch<-function(datos, fechas, n, ind) {
23 #Vectores para almacenar residuales de exceso
24 #Innovaciones Gaussianas
25 rts.norm.95<-c()
26 n.95<-1
27 rts.norm.99<-c()
28 n.99<-1
29 rts.norm.995<-c()
30 n.995<-1
31 rts.norm.999<-c()
32 n.999<-1
33 #Innovaciones t
34 rts.t.95<-c()
35 t.95<-1
36 rts.t.99<-c()
37 t.99<-1
38 rts.t.995<-c()
39 t.995<-1
40 rts.t.999<-c()
41 t.999<-1
42 #Innovaciones GPD
43 rts.evt.95<-c()
44 evt.95<-1
45 rts.evt.99<-c()
46 evt.99<-1
47 rts.evt.995<-c()
48 evt.995<-1
49 rts.evt.999<-c()
50 evt.999<-1
51 #Variable para la Volatilidad de cada modelo
52 vol<-0
53 #Evitando el uso innecesario de la función length
54 m2<-(length(datos)-n)
55 #Comienza el ciclo
56 for(i in 1:(m2+1)) {
57   datosfit<-datos[i:((n-1)+i)]
58   attr(datosfit, "times")<-fechas[i:((n-1)+i)]
59   fit.qmle<-garchFit(formula=~garch(1,1),
60                     data=datosfit, cond.dist=c("QMLE"),
61                     trace=F, incluye.mean=F,
62                     hessian = "ropt")
63   #Indicando cómo se estima la volatilidad
64   if(ind == 0)

```

```

65     vol<-predict(fit.qmle,1)$standardDeviation
66   if(ind == 1) {
67     if(fit.qmle@fit$par['alpha1']+fit.qmle@fit$par['beta1']
68       >1 || round(fit.qmle@fit$matcoef[2,4],4) > 0.05)
69       vol<-sqrt((fit.qmle@fit$par['alpha1']*
70                 (fit.qmle@data[length(fit.qmle@data)]^2))+
71                 ((volatility(fit.qmle)[length(volatility(
fit.qmle))]) *
72                 (1-fit.qmle@fit$par['alpha1'])))
73   else
74     vol<-predict(fit.qmle,1)$standardDeviation }
75   #Calculando residuales de exceso los días que se viola el
VaR
76   #Innovaciones Gaussianas
77   if((VaR_garch_norm(0.95, fit.qmle, ind) < datos[(n+i)]) &
78     ((n+i) <= length(datos))) {
79     rts.norm.95[n.95]<-(datos[(n+i)] -
80                       ES_garch_norm(0.95, fit.qmle, ind))/
81                       vol
82     n.95<-n.95+1 }
83   if((VaR_garch_norm(0.99, fit.qmle, ind) < datos[(n+i)]) &
84     ((n+i) <= length(datos))) {
85     rts.norm.99[n.99]<-(datos[(n+i)] -
86                       ES_garch_norm(0.99, fit.qmle, ind))/
87                       vol
88     n.99<-n.99+1 }
89   if((VaR_garch_norm(0.995, fit.qmle, ind) < datos[(n+i)]) &
90     ((n+i) <= length(datos))) {
91     rts.norm.995[n.995]<-(datos[(n+i)] -
92                       ES_garch_norm(0.995, fit.qmle, ind))/
93                       vol
94     n.995<-n.995+1 }
95   if((VaR_garch_norm(0.999, fit.qmle, ind) < datos[(n+i)]) &
96     ((n+i) <= length(datos))) {
97     rts.norm.999[n.999]<-(datos[(n+i)] -
98                       ES_garch_norm(0.999, fit.qmle, ind))/
99                       vol
100    n.999<-n.999+1 }
101   #Innovaciones t-Student
102   if((VaR_garch_t(0.95, fit.qmle, ind) < datos[(n+i)]) &
103     ((n+i) <= length(datos))) {
104     rts.t.95[t.95]<-(datos[(n+i)] -
105                       ES_garch_t(0.95, fit.qmle, ind))/vol
106     t.95<-t.95+1 }
107   if((VaR_garch_t(0.99, fit.qmle, ind) < datos[(n+i)]) &
108     ((n+i) <= length(datos))) {
109     rts.t.99[t.99]<-(datos[(n+i)] -
110                       ES_garch_t(0.99, fit.qmle, ind))/vol
111     t.99<-t.99+1 }

```

```

112   if((VaR_garch_t(0.995, fit.qmle, ind) < datos[(n+i)]) &
113       ((n+i) <= length(datos))) {
114     rts.t.995[t.995]<-(datos[(n+i)] -
115                       ES_garch_t(0.995, fit.qmle, ind))/vol
116     t.995<-t.995+1 }
117   if((VaR_garch_t(0.999, fit.qmle, ind) < datos[(n+i)]) &
118       ((n+i) <= length(datos))) {
119     rts.t.999[t.999]<-(datos[(n+i)] -
120                       ES_garch_t(0.999, fit.qmle, ind))/vol
121     t.999<-t.999+1 }
122   #Innovaciones DPG
123   if((VaR_garch_evt(0.95, fit.qmle, ind) < datos[(n+i)]) &
124       ((n+i) <= length(datos))) {
125     rts.evt.95[evt.95]<-(datos[(n+i)] -
126                          ES_garch_evt(0.95, fit.qmle, ind))/
127     vol
128     evt.95<-evt.95+1 }
129   if((VaR_garch_evt(0.99, fit.qmle, ind) < datos[(n+i)]) &
130       ((n+i) <= length(datos))) {
131     rts.evt.99[evt.99]<-(datos[(n+i)] -
132                          ES_garch_evt(0.99, fit.qmle, ind))/
133     vol
134     evt.99<-evt.99+1 }
135   if((VaR_garch_evt(0.995, fit.qmle, ind) < datos[(n+i)]) &
136       ((n+i) <= length(datos))) {
137     rts.evt.995[evt.995]<-(datos[(n+i)] -
138                          ES_garch_evt(0.995, fit.qmle, ind))/
139     vol
140     evt.995<-evt.995+1 }
141   if((VaR_garch_evt(0.999, fit.qmle, ind) < datos[(n+i)]) &
142       ((n+i) <= length(datos))) {
143     rts.evt.999[evt.999]<-(datos[(n+i)] -
144                          ES_garch_evt(0.999, fit.qmle, ind))/
145     vol
146     evt.999<-evt.999+1 } }
147   #Calculando p-Valores de la prueba bootstrap para evaluar
148   las estimaciones condicionales de ES
149   #Innovaciones Gaussianas
149   pvalue.norm.95<-bootstrap1(rts.norm.95,0,10000)
150   pvalue.norm.99<-bootstrap1(rts.norm.99,0,10000)
151   pvalue.norm.995<-bootstrap1(rts.norm.995,0,10000)
152   pvalue.norm.999<-bootstrap1(rts.norm.999,0,10000)
153   #Innovaciones t-Student
154   pvalue.t.95<-bootstrap1(rts.t.95,0,10000)
155   pvalue.t.99<-bootstrap1(rts.t.99,0,10000)
156   pvalue.t.995<-bootstrap1(rts.t.995,0,10000)
157   pvalue.t.999<-bootstrap1(rts.t.999,0,10000)
158   #Innovaciones DPG
159   pvalue.evt.95<-bootstrap1(rts.evt.95,0,10000)

```

```

160 pvalue.evt.99<-bootstrap1(rts.evt.99,0,10000)
161 pvalue.evt.995<-bootstrap1(rts.evt.995,0,10000)
162 pvalue.evt.999<-bootstrap1(rts.evt.999,0,10000)
163 #Armando la matriz del ES 0.95
164 pvalue.95<-c(pvalue.norm.95,pvalue.t.95,pvalue.evt.95)
165 ES.95<-matrix(pvalue.95,nrow=3,ncol=1,byrow=FALSE,
166             dimnames = list(c("Modelo GARCH-Normal",
167                             "Modelo GARCH-t-Student",
168                             "Modelo GARCH-DPG"),
169                             c("ES_0.95: p-Value
170                               Bootstrap Test")))
171 #Armando la matriz del ES 0.99
172 pvalue.99<-c(pvalue.norm.99,pvalue.t.99,pvalue.evt.99)
173 ES.99<-matrix(pvalue.99,nrow=3,ncol=1,byrow=FALSE,
174             dimnames = list(c("Modelo GARCH-Normal",
175                             "Modelo GARCH-t-Student",
176                             "Modelo GARCH-DPG"),
177                             c("ES_0.99: p-Value
178                               Bootstrap Test")))
179 #Armando la matriz del ES 0.995
180 pvalue.995<-c(pvalue.norm.995,pvalue.t.995,pvalue.evt.995)
181 ES.995<-matrix(pvalue.995,nrow=3,ncol=1,byrow=FALSE,
182             dimnames = list(c("Modelo GARCH-Normal",
183                             "Modelo GARCH-t-Student",
184                             "Modelo GARCH-DPG"),
185                             c("ES_0.995: p-Value
186                               Bootstrap Test")))
187 #Armando la matriz del ES 0.999
188 pvalue.999<-c(pvalue.norm.999,pvalue.t.999,pvalue.evt.999)
189 ES.999<-matrix(pvalue.999,nrow=3,ncol=1,byrow=FALSE,
190             dimnames = list(c("Modelo GARCH-Normal",
191                             "Modelo GARCH-t-Student",
192                             "Modelo GARCH-DPG"),
193                             c("ES_0.999: p-Value
194                               Bootstrap Test")))
195 #Violaciones Esperadas al VaR
196 nombres<-c("0.95","0.99","0.995","0.999")
197 viol.esp1<-c(n*(1-0.95),n*(1-0.99),n*(1-0.995),n*(1-0.999))
198 viol.esp<-matrix(viol.esp1,nrow=1,ncol=4,byrow=T,
199             dimnames = list(c("Violaciones esperadas"),
200                             c("0.95","0.99","0.995",
201                               "0.999")))
202 #Formamos la lista que regresará esta función
203 Backtest_ES<-list(Sfall.95=ES.95,Sfall.99=ES.99,
204                 Sfall.995=ES.995,Sfall.999=ES.999,
205                 Viol.Esp=viol.esp)
206 return(Backtest_ES) }

```

Bibliografía

- [1] Jonathan Martínez Martínez. *Modelos de volatilidad estocástica y su aplicación a la administración de riesgos financieros*. Universidad Nacional Autónoma de México, 2014.
- [2] María Antonieta Campa Rojas. *Teoría de Valores Extremos con aplicaciones a medidas de riesgo*. Universidad Nacional Autónoma de México, 2001.
- [3] Alexander J McNeil and Rüdiger Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300, 2000.
- [4] Lancine Kourouma, Denis Dupre, Gilles Sanfilippo, and Ollivier Tarasmasco. Extreme value at risk and expected shortfall during financial crisis. *International Journal of Forecasting*, 2011.
- [5] Viviana Paulina Fernández. Extreme value theory: value at risk and returns dependence around the world. 2003.
- [6] Ramazan Gencay and Faruk Selcuk. Extreme value theory and value-at-risk: Relative performance in emerging markets. *International Journal of Forecasting*, 20(2):287–303, 2004.
- [7] Alejandro Iván Aguirre Salado, Humberto Vaquera Huerta, Martha Elva Ramírez Guzmán, José René Valdez Lazalde, Carlos Arturo Aguirre Salado, et al. Value-at-risk-estimation in the mexican stock exchange using conditional heteroscedasticity models and theory of extreme values. *Economía Mexicana NUEVA ÉPOCA*, 22(1):177–205, 2013.
- [8] Edna López Ávila. *Valor en Riesgo del Peso Mexicano y otras divisas latinoamericanas ante eventos de crisis utilizando la Teoría de Valores Extremos*. Universidad Nacional Autónoma de México, 2013.

- [9] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton university press, 2005.
- [10] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.
- [11] Philippe Artzner, David Heath, F Delbaen, and JM Eber. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.
- [12] Peter J Brockwell and Richard A Davis. *Time series: Theory and Methods*. Springer-Verlag New York, 1991.
- [13] Henrik Madsen. *Time series analysis*. CRC Press, 2007.
- [14] Peter J Brockwell and Richard A Davis. *Introduction to Time Series and Forecasting*. Springer-Verlag New York, 2002.
- [15] Andreas Brandt. The stochastic equation $y_{n+1} = a_n y_n + b_n$ with stationary coefficients. *Advances in Applied Probability*, pages 211–220, 1986.
- [16] Thomas Mikosch. Modeling dependence and tails of financial time series. *Extreme Values in Finance, Telecommunications, and the Environment*, pages 185–286, 2003.
- [17] Philippe Bougerol and Nico Picard. Stationarity of garch processes and of some non-negative time series. *Journal of econometrics*, 52(1):115–127, 1992.
- [18] George Casella and Roger L Berger. *Statistical inference*. Duxbury Pacific Grove, CA, 2002.
- [19] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events for insurance and finance*, volume 33. Springer Science & Business Media, 1997.
- [20] Malcolm R Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Science & Business Media, 1983.
- [21] Janos Galambos. *Extreme value theory for applications*. Springer, 1994.
- [22] Jon Danielsson and Casper G de Vries. Tail index and quantile estimation with very high frequency data. *Journal of empirical Finance*, 4(2):241–257, 1997.
- [23] J. P. Morgan and Risk Metrics Reuters. Technical document. *Morgan Guaranty Trust Company of New York*, 91, 1996.

- [24] Ulrich A Müller, Michel M Dacorogna, Rakhal D Davé, Richard B Olsen, Olivier V Pictet, and Jacob E von Weizsäcker. Volatilities of different time resolutions: analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2):213–239, 1997.
- [25] Neil Shephard. Statistical aspects of arch and stochastic volatility. *Monographs on Statistics and Applied Probability*, 65:1–68, 1996.
- [26] Christian Gouriéroux. *ARCH models and financial applications*. Springer Science & Business Media, 1997.
- [27] Richard L Smith. Estimating tails of probability distributions. *The annals of Statistics*, pages 1174–1207, 1987.
- [28] Alexander J McNeil and Thomas Saladin. *The peaks over thresholds method for estimating high quantiles of loss distributions*. 1997.
- [29] Carol Alexander. *Market Risk Analysis, Practical Financial Econometrics (Volume II)*. John Wiley & Sons Ltd, 2008.
- [30] Ruey S Tsay. *Analysis of financial time series*, volume 543. John Wiley & Sons, Inc., 2005.