

CUANTIFICACION DE RIESGOS EN PRESENCIA DE EVENTOS EXTREMOS

Andrés Mora Valencia

Trabajo de investigación 013/013

Master en Banca y Finanzas Cuantitativas

Tutores: Dr. Gonzalo García Donato
Dr. Antonio Díaz

Universidad Complutense de Madrid

Universidad del País Vasco

Universidad de Valencia

Universidad de Castilla-La Mancha

Cuantificación de riesgo en presencia de eventos extremos

Autor: Andrés Mora
Tutores: Gonzalo García-Donato
Antonio Díaz

Junio, 25 de 2013

Índice general

Resumen	vi
1. Introducción	1
1.1. Motivación	1
1.2. Revisión de la Literatura	3
2. Conceptos básicos	9
2.1. Value-at-Risk	14
2.2. Modelos GARCH	16
2.3. Distribuciones	24
3. Métodos de estimación	38
3.1. Método de estimación para la distribución normal y t-Student	39
3.2. Método de estimación para la distribución de Pareto generalizada (GPD)	39
3.3. Métodos de estimación para la distribución α -estable	40
3.4. Método de estimación para la distribución g-h	41
4. Backtesting y Aplicación	43
4.1. Backtesting	43
4.2. Aplicación	51

4.3. Propuesta para la mejora de eficiencia computacional	55
5. Conclusiones	60
A. Regresión robusta	62
B. Modelos de vectores autorregresivos (VAR)	65

Índice de cuadros

1.1. Resumen de la revisión de la literatura	8
2.1. Estadística descriptiva de las rentabilidades del DJIA	13
2.2. Estimación Modelo GARCH(1,1)	21
2.3. Estimación Modelo ARMA(1,1)GARCH(1,1)	23
4.1. Resultados backtesting - DJIA	48
4.2. Contrastes de cobertura incondicional y de independencia para DJIA	49
4.3. Resultados de Rachev et al. (2010)	49
4.4. Estadística Descriptiva de la Aplicación	52
4.5. Resultados del Backtesting - Aplicación	53
4.6. Contrastes de cobertura incondicional y de independencia para MSCI	54
4.7. Contrastes de cobertura incondicional y de independencia para USD/EUR	54
4.8. Ajuste de las predicciones	57

Índice de figuras

1.1. Representación esquemática de la metodología que seguiremos en este trabajo.	7
2.1. Precios y rentabilidades del DJIA	12
2.2. Correlograma de las rentabilidades y valor absoluto de las rentabilidades del DJIA	12
2.3. Densidad de las rentabilidades del DJIA y ajuste a la normal .	13
2.4. Representación gráfica del VaR	15
2.5. Serie temporal de rentabilidades y de volatilidad	22
2.6. Correlograma y QQ-Plot de residuos estandarizados	22
2.7. Correlograma y QQ-Plot de residuos estandarizados	23
2.8. Tres casos especiales de la α -estable	25
2.9. Forma de la densidad a diferentes valores del parámetro α . .	26
2.10. Función de distribución y de densidad de GPD	28
2.11. Representación gráfica del método POT	30
2.12. Gráfico de Hill	32
2.13. Ejemplo de Horror Hill Plot	33
2.14. Densidad de g-h	35
4.1. Backtesting para las rentabilidades del DJIA	45
4.2. Evolución del parámetro - t-Student	46

4.3. Evolución de los parámetros - g-h	47
4.4. Evolución del índice de estabilidad	47
4.5. Gráfico de precios y rentabilidades de la Aplicación	51
4.6. Orden del VAR	56
4.7. Comparación de las predicciones y error de pronóstico	57
4.8. Comparación del VaR usando vectores autoregresivos	58

Resumen

En este trabajo se analiza la cuantificación del valor en riesgo (VaR) mediante varias distribuciones empleadas en gestión de riesgos en periodos de fuertes turbulencias en los mercados financieros. Los datos analizados, correspondientes a rendimientos bursátiles, se filtran mediante un proceso (ARMA)GARCH. De esta manera, se modelizan los hechos estilizados de varianza cambiante y aglomeración de la volatilidad. A las distribuciones habituales en la literatura (normal y t-Student), se incorporan la distribución de Pareto generalizada (GPD) empleada en la teoría del valor extremo (EVT), la distribución α -estable que brinda un mejor ajuste a ciertos rendimientos financieros que los modelos clásicos gaussianos, y la distribución g-h que tiene en cuenta asimetría y curtosis. Para el periodo y muestra analizada se recomienda el uso de la distribución GPD y α -estable.

Agradecimientos

En primer lugar doy gracias a Dios, sin Él no soy nada. Quiero agradecer también a mis tutores Gonzalo y Antonio, por su orientación, dedicación y paciencia. También agradezco a los profesores y coordinadores del Máster por su contribución y diligencia al resolver mis dudas académicas. Agradezco a mi madre por su infinito apoyo desde la distancia. Finalmente, pero no menos importante, agradezco al Gobierno del País Vasco y la Universidad del País Vasco (UPV/EHU) por la ayuda económica a la movilidad, sin el cual no hubiese sido posible realizar el Máster. Este trabajo está dedicado a la memoria de mi padre Guillermo, a mi madre Hilda y a mi hermano Diego Fernando.

Capítulo 1

Introducción

1.1. Motivación

A raíz de las recientes crisis, hipotecaria y de deuda soberana, los bancos, entidades financieras y el Comité de Basilea han estado buscando formas más efectivas de medir el riesgo de sus inversiones. La medida de riesgo estándar de la industria financiera es el Value-at-Risk (VaR). Para cuyo cálculo los primeros modelos paramétricos del VaR suponían rendimientos distribuidos normal. No obstante, de un tiempo aquí se conoce que los rendimientos presentan colas más pesadas que las de una distribución normal (Mandelbrot, 1963), de esta manera, se comienzan a contemplar distribuciones cuyas colas decaen de manera potencial, es decir, decaen a cero más lentamente que las de una normal, permitiendo así capturar de mejor manera los eventos de cola. Por lo que en la literatura financiera más reciente se ha reemplazado la distribución normal por la distribución t-Student. Aún así esta distribución también es simétrica, y aunque la media de los rendimientos bursátiles (diarios) tienden a ser cero, se observa cierta asimetría en la distribución de pérdidas y ganancias, teniendo mayor probabilidad las pérdidas.

El objetivo de este trabajo es medir el desempeño del VaR con diferentes distribuciones mediante la comparación del número de excepciones resultante del modelo del VaR y las excepciones esperadas, procedimiento conocido como backtesting, y así elegir qué modelo es más adecuado para calcular el VaR. Las distribuciones escogidas en este trabajo son la t-Student, α -estable,

distribución de Pareto generalizada (GPD, por sus siglas en inglés) y la distribución g-h. Estas distribuciones presentan la característica de capturar eventos de cola, comparadas con un caso base que es la distribución normal, como se observa en la revisión de la literatura. Estimar los parámetros de un modelo presenta un desafío. Cuando se cuenta con una forma analítica de la función de verosimilitud, se procede a maximizar esta función y el resultado de esta optimización son los parámetros estimados. Generalmente, se emplean métodos numéricos, pero algunas veces la función de verosimilitud puede presentar varios máximos locales o ser muy plana. Adicionalmente, algunas distribuciones no cuentan con una fórmula cerrada para su densidad, lo cual hace que el método de maximizar la función de verosimilitud no sea aplicable o requiera de una extracción numérica. Este es el caso de la distribución α -estable y g-h. Razón por la cual en este trabajo también se investigan otros métodos de estimación y se presentan en el capítulo tres. Una de las novedades de este trabajo es la incorporación al estudio comparativo de la distribución α -estable ya que existen pocos estudios donde se incluye esta distribución bajo un marco de backtesting. La razón puede ser, como ya se ha comentado, que la estimación de los parámetros no es fácil comparada con otras distribuciones, puesto que no existe una forma explícita de su función de densidad. Adicionalmente, en nuestro conocimiento, este es el primer trabajo donde se evalúa el desempeño de la distribución g-h mediante backtesting.

En este trabajo se recomienda el uso de la distribución α -estable en conjunto con la teoría del valor extremo, puesto que sus resultados son más confiables en especial en épocas de turbulencia financiera. Aunque no existe una fórmula cerrada de la función de densidad de α -estable, en este documento se aplica el método de Koutrouvelis (1980) para la estimación de sus parámetros, puesto que el método de máxima verosimilitud resulta ser computacionalmente costoso. Otro aporte de este trabajo consiste en reducir el tiempo computacional que toma este método. La idea consiste en dividir el backtesting en dos periodos. Para el primer periodo se estiman los dos parámetros de la distribución mediante máxima verosimilitud, y a partir de estas estimaciones, en el segundo periodo del backtesting se emplea una técnica de predicción de estos dos parámetros, y estos valores se usan para calcular el VaR. Los resultados muestran que la predicción es muy aproximada a la estimación por el método de máxima verosimilitud. Adicionalmente, este trabajo presenta una propuesta para estimar los parámetros de la distri-

bución g-h. Esta propuesta consiste en emplear una regresión robusta para estimar el parámetro h en el método sugerido por Hoaglin (1985).

Entonces la pregunta de investigación de este trabajo es ¿Existe un modelo que permita calcular VaR de manera adecuada, y en especial, capturar eventos de cola? Existe un consenso entre la academia y la industria financiera en que las rentabilidades de activos financieros presentan colas pesadas. Por ello, los gestores de riesgo deben tener en cuenta los eventos extremos en sus modelos. Adicionalmente, se deben tener en cuenta fenómenos como la asimetría, aglomeración de volatilidad y, en algunos casos, autocorrelación de dichas rentabilidades para tener una medida confiable del VaR. Entonces, este trabajo pretende ofrecer a la industria financiera, en especial al área de gestión de riesgos, un análisis de modelos que permitan calcular VaR de manera confiable en épocas de turbulencia financiera.

El resto del documento se divide de la siguiente forma: el capítulo 2 introduce los conceptos básicos. El capítulo 3 trata de los métodos de estimación, en el cuarto capítulo se realiza la aplicación empírica y se validan los métodos mediante un backtesting, y finalmente el capítulo 5 concluye.

1.2. Revisión de la Literatura

Existen en la literatura financiera varios trabajos relacionados con este tema, y generalmente se observa la comparación entre la distribución GPD (empleada en el método POT¹), la t-Student y normal. En uno de los trabajos pioneros, McNeil y Frey (2000) comparan VaR diario empleando las distribuciones normal, t-Student y GPD aplicados a variaciones logarítmicas diarias de los índices bursátiles SP 500; DAX; la acción de BMW; tipo de cambio dólar/libra y de precios del oro. Estas variaciones se filtran mediante un modelo AR(1)-GARCH(1,1). Los autores encuentran que el enfoque GPD presentan un buen desempeño al calcular VaR a diferentes niveles de probabilidad. Kiesel, Perraudin y Taylor (2003) emplean el método propuesto por McNeil y Frey (2000) en datos de rendimientos de bonos en mercados emergentes, y encuentran que los resultados bajo EVT (Extreme Value Theory, por sus siglas en inglés) son similares a los cuantiles de la distribución empírica. Otro estudio que concluye el buen desempeño de EVT, no sólo en periodos

¹Peaks over Threshold, por sus siglas en inglés

de relativa calma sino también en periodos volátiles, es el de Byström (2004). El autor compara las distribuciones usadas en EVT (GEV² y GPD) con la normal y t-Student. También emplea modelos AR-GARCH y utiliza datos de los índices Dow Jones Industrial Average (DJIA) y el AFF³ sueco. Otros estudios que también encuentran buen desempeño de EVT son: Lauridsen (2000), Ourir y Snoussi (2012), De Jesús, Ortiz y Cabello (2013).

Además, varios autores demuestran que variantes o combinaciones de las distribuciones de EVT también funcionan bien. Chavez-Demoulin y McGill (2012) emplean procesos de Hawkes para capturar aglomeración de volatilidad junto con la distribución GPD para las pérdidas en exceso, y se aplica a datos intradía de la acción NVIDIA Corporation. Araújo Santos, Fraga Alves y Hammoudeh (2013) proponen usar los métodos Picos sobre el Umbral basados en Duración (DPOT) y Picos sobre el Umbral Aleatorio (PORT). Estos métodos se comparan con las distribuciones normal y t-asimétrica, y son aplicados a datos del índice Dow Jones Industrial Average. Wei, Chen y Lin (2013) combinan un modelo de volatilidad multifractal con la distribución GPD y compara su desempeño con diferentes modelos tipos GARCH. Los resultados del backtesting, al calcular VaR a datos intradía del índice Shanghai Stock Exchange Component, muestran que el método propuesto es más preciso que los obtenidos por los modelos tipo-GARCH.

Angelidis, Benos y Degiannakis (2004) emplean las distribuciones normal, t-Student y la distribución de errores generalizada (GED) y muestran que las distribuciones leptocúrticas presentan un buen comportamiento. Se emplean rentabilidades diarias de los índices SP 500; NIKKEI 225; DAX 30; CAC 40 y FTSE 100. Otro estudio similar, incluyendo también la distribución GPD, es desarrollado por Bekiros y Georgoutsos (2005). Los autores muestran que las técnica POT (basada en la GPD) se desempeña mejor que las otras distribuciones (normal, t-Student y GED), cuando se calcula VaR al 99.5%. Para niveles “bajos” de probabilidad, los otros métodos tienen un resultado satisfactorio. Los datos empleados corresponden a rentabilidades diarias de los índices Dow Jones Industrial Average y Cyprus Stock Exchange.

Con respecto a la distribución α -estable, Khindanova, Rachev y Schwartz (2001) muestran que el VaR calculado con la distribución estable presenta mejor desempeño que la distribución normal. Marinelli, d’Addona y Rachev

²Generalized Extreme Value, por sus siglas en inglés

³Abreviación del índice general sueco Affärsvärlden

(2007), bajo backtesting, comparan el desempeño del VaR con distribuciones α -estable y GPD, y encuentran que la distribución estable se desempeña mejor que GPD cuando se calcula VaR al 95 %; sin embargo, los resultados son similares cuando se calcula VaR al 99 %. Los autores escogen datos del SP 500; NASDAQ; Amazon y Microsoft. Güner, Rachev, Edelman y Fabozzi (2010) calculan VaR con esta distribución de manera condicional, es decir, en una primera etapa filtran las rentabilidades de fondos de cobertura mediante un modelo ARMA(1,1)-GARCH(1,1) con innovaciones t-Student y en la segunda etapa ajustan la distribución α -estable a los residuos estandarizados de la primera etapa. Rachev, Racheva-Iotova y Stoyanov (2010) emplean rentabilidades diarias del DJIA y mediante backtesting comparan la distribución normal, t-Student, GPD y α -estable. Los autores encuentran que esta última distribución es la que mejor se desempeña.

En cuanto a otras distribuciones, ciertos trabajos utilizan la t-asimétrica, como por ejemplo el trabajo de Kuester, Mittnik y Paoletta (2006). Los autores comparan el desempeño de esta distribución frente a la distribución normal, t-Student y GPD a datos del índice NASDAQ Composite Index. Los resultados muestran que los modelos basados tanto en GPD como en t-asimétrica son los que mejores se desempeñan, puesto que estas distribuciones tienen en cuenta asimetría y colas pesadas. Junto con la t-asimétrica está también la distribución Pearson-IV (Stavroyiannis, Makris, Nikolaidis y Zarangas, 2012). En ese estudio se emplean rentabilidades diarias de los índices DJIA; NASDAQ Composite; FTSE100; CAC40; DAX y SP 500. Los autores muestran que la distribución Pearson tipo-IV presenta mejores resultados que la t-asimétrica cuando se calcula VaR a altos niveles de probabilidad. Estos resultados van en la línea de los obtenidos por Grigoletto y Lisi (2009, 2011). Otras distribuciones son las mixturas Gaussianas y de t-Student. Basado en el estudio de Zhang y Cheng (2005), Haas (2009) prueba este tipo de distribuciones y encuentra que la de mejor desempeño es la mixtura de dos distribuciones t-Student. Se calcula VaR a rentabilidades diarias de los principales índices bursátiles europeos.

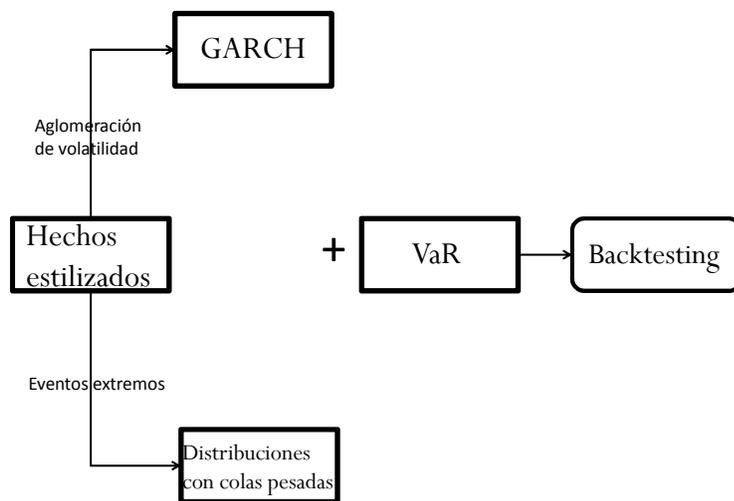
En otro tipo de estudios empíricos, Berkowitz y O'Brien (2002) encuentran que el modelo VaR-ARMA(1,1)-GARCH(1,1)-normal es satisfactorio para pronosticar riesgos extremos para los principales bancos comerciales en Estados Unidos, y que las medidas de VaR usadas por los bancos son muy conservadoras. El VaR es estimado al 99 %. En línea con este resultado está el estudio de Pérignon, Deng y Wang (2008), basado en datos de bancos

comerciales en Canadá.

Por otro lado, Wong (2010) propone sumar el tamaño de pérdidas mediante el uso de técnicas de punto de silla; en lugar de contar el número de excepciones. Los autores analizan datos diarios del SP 500 y encuentran que distribuciones asimétricas y de colas pesadas fallan en capturar el riesgo de cola durante el crash de 1987.

El Cuadro 1.1 presenta un resumen de la revisión de la literatura. En la mayoría de estos trabajos, cuando se emplea el método POT, se observa el uso de la técnica de dos pasos de McNeil y Frey (2000) que está basado en Diebold, Shuermann y Stroughair (1998). Esta técnica consiste en filtrar las rentabilidades mediante un modelo AR(1)-GARCH(1,1) en un primer paso y luego calcular los parámetros de la GPD sobre los residuos estandarizados de este modelo. Jalal y Rockinger (2008) muestran que esta técnica tiene muy buenas propiedades predictivas para el cálculo del VaR. Por tal razón, en el presente documento se emplea esta misma técnica, pero con un modelo ARMA(1,1)-GARCH(1,1). La Figura 1.1 resume la metodología de este proyecto de investigación.

Figura 1.1: Representación esquemática de la metodología que seguiremos en este trabajo.



Cuadro 1.1: Resumen de la revisión de la literatura

Autor (año)	Distribuciones empleadas	Datos	Mejor modelo
McNeil y Frey (2000)	normal, t-Student, GPD	SP 500; DAX 30; BMW; dólar/libra y oro	GPD
Byström (2004)	normal, t-Student, GEV, GPD	DJIA y AFF	GEV y GPD
Angelidis et al. (2004)	normal, t-Student, GED	SP 500; NIKKEI 225; DAX 30; CAC 40 y FTSE 100	t-Student y GED
Bekiros y Georgoutsos (2005)	normal, t-Student, GED, GPD	DJIA y Cyprus Stock Exchange	GPD
Khindanova et al. (2001)	normal, α -estable	divisas e índices	α -estable
Marinelli et al. (2007)	GPD, α -estable	SP 500; NASDAQ; Amazon y Microsoft	α -estable (VaR-95 %)
Güner et al. (2010)	t-Student, α -estable	Hedge Funds	α -estable
Rachev et al. (2010)	normal, t-Student, GPD, α -estable	DJIA	α -estable
Kuester et al. (2006)	normal, t-Student, t-asimétrica, GPD	NASDAQ	GPD y t-asimétrica
Stravroyiannis et al. (2012)	t-asimétrica y Pearson IV	DJIA; NASDAQ; FTSE 100; CAC 40; DAX 30 y SP 500	Pearson IV
Haas (2009)	mixturas Gaussianas, mixturas t-Student	Principales índices europeos	mixturas de t-Student

Capítulo 2

Conceptos básicos

Según McNeil et al. (2005, p. 1), riesgo es cualquier evento o acción que puede afectar de manera adversa la habilidad que tiene una organización de lograr sus objetivos y ejecutar sus estrategias, o la verosimilitud cuantificable de pérdida o de rendimientos menores a lo esperado.

El riesgo financiero se clasifica, generalmente, en tres clases: riesgo de crédito, riesgo de mercado y riesgo operacional (Crohuy, Galai y Mark, 2005, p. 14). Riesgo de crédito se define como el riesgo de que una contraparte no cumpla con una obligación en el momento del vencimiento o en un momento anterior. El riesgo de mercado es el riesgo de que el valor de una inversión decrezca debido a movimientos en los factores de riesgo (por ej. tipos de cambio, tipos de interés, precios de commodities, fluctuación en los precios, entre otros). Riesgo operacional se define como el riesgo causado por deficiencias en los sistemas de información o en controles internos que podrían resultar en pérdidas inesperadas. Este tipo de riesgo también se puede definir como el riesgo de errores humanos o falla de un componente en el hardware, software o sistemas de comunicaciones que son cruciales para la liquidación de transacciones¹. El primer Acuerdo de Basilea (Basilea I o BIS 88) establece el capital mínimo que un banco requiere para solventar exposiciones de crédito. Mientras que la Enmienda a Basilea I, conocida como BIS 98, requiere que las instituciones financieras calculen requerimientos de capital incluyendo riesgo de mercado.

¹Estas definiciones se pueden encontrar en el Glosario de “A glossary of terms used in payments and settlement systems”; March 2003, BIS.

Los precios de activos financieros son variables aleatorias, puesto que no es posible determinar exactamente cuáles serán sus valores en el futuro. Una primera aproximación a la modelización estadística de los activos es suponer que las variaciones de los precios de estos activos en un periodo corto de tiempo siguen un modelo lognormal, y por ende, los rendimientos de los activos (cambios relativos en el precio) comúnmente se miden como la diferencia en el logaritmo de los precios que, supuestamente, siguen una distribución normal (Alexander, 2001, p.4). Sin embargo, en la práctica, los factores de riesgo de mercado tienden a presentar distribuciones con colas más pesadas que las de una normal (Hull, 2006, p. 117). Adicionalmente, se supone que los rendimientos de los activos financieros son generados por un proceso estocástico estacionario, es decir, el valor esperado, varianza y auto-covarianza es la misma en cada momento de tiempo.

La volatilidad es una medida de dispersión en una densidad de probabilidad. La medida más comúnmente usada es la desviación estándar (o desviación típica) σ de una variable aleatoria, que es la raíz cuadrada de la varianza (Alexander, 2001, p.4). La volatilidad del proceso estocástico que gobierna los rendimientos se denomina proceso de volatilidad, mientras que la realización de un proceso de volatilidad se denomina volatilidad realizada, que puede medirse usando datos históricos de precios. Si el proceso de rendimientos presenta una volatilidad cambiante con el tiempo gobernada por un modelo GARCH (modelo de heteroscedasticidad condicional auto-regresiva generalizada), entonces la volatilidad realizada es la volatilidad estimada mediante GARCH sobre un periodo de tiempo (Alexander, 2001, p.11). Este modelo se describirá más adelante en el apartado 2.2.

En cuanto al comportamiento de las rentabilidades diarias de las acciones, varios autores han estudiado las características empíricas (conocidos también como hechos estilizados) de las rentabilidades de acciones (Granger y Ding 1995; Rydén, Teräsvirta y Asbrink, 1998; Cont, 2001, McNeil, Frey y Embrechts, 2005, p. 117, entre otros). Entre las más importantes están:

- (CE-1) Ausencia de autocorrelación lineal
- (CE-2) La distribución incondicional (y en algunos casos, condicional) de los rendimientos presenta colas pesadas
- (CE-3) Asimetría en la distribución de pérdidas y ganancias

(CE-4) Aglomeración de volatilidad

(CE-5) Decaimiento lento de autocorrelación en el valor absoluto de los rendimientos².

Debido a estas características, usualmente, se filtran las rentabilidades de acciones mediante un modelo GARCH, y se emplea la medida de Value-at-Risk (VaR) para cuantificar riesgo. Puesto que el modelo GARCH es una herramienta flexible que permite capturar las características empíricas previamente mencionadas (Teräsvirta y Zhao, 2011, p. 79).

En los siguientes apartados de este capítulo se describe el VaR, el modelo GARCH y las distribuciones a usar. Antes de entrar en detalle, se realiza un ejemplo con datos del Dow Jones Industrial Average (DJIA) para observar las características empíricas mencionadas anteriormente.

Ejemplo 1. (Características empíricas)

Los datos³ a analizar en éste y los siguientes ejemplos de este trabajo corresponden a precios del índice Dow Jones Industrial Average (DJIA) desde el 7 de agosto de 2003 hasta el 31 de diciembre de 2009 para un total de 1671 datos diarios. Los rendimientos logarítmicos se calculan como $r_t = 100 \log(P_t/P_{t-1})$, donde P_t el precio en el momento t (es decir, se analizan 1670 datos diarios de rentabilidad). El objetivo de emplear estos datos, en los ejemplos de este documento, es comparar los resultados obtenidos en Rachev et al. (2010) y los de este trabajo de investigación.

En el lado derecho de la Figura 2.1, se observa la característica empírica (CE-4) Aglomeración de volatilidad, donde periodos de baja (alta) volatilidad vienen seguidos de baja (alta) volatilidad.

La característica empírica (CE-1) ausencia de autocorrelación no es muy clara para esta serie de rentabilidades, como se observa en el lado izquierdo de la Figura 2.2. Según el estadístico Ljung-Box, se rechaza la hipótesis nula de cero autocorrelación a cualquier nivel razonable de significación, empleando

²El valor absoluto de los rendimientos es una aproximación a una medida de dispersión de las rentabilidades. También se suele usar el cuadrado de las rentabilidades, como aproximación a la varianza de las mismas.

³Fuente: Datastream, Thomson Financial.

Figura 2.1: Precios y rentabilidades del DJIA

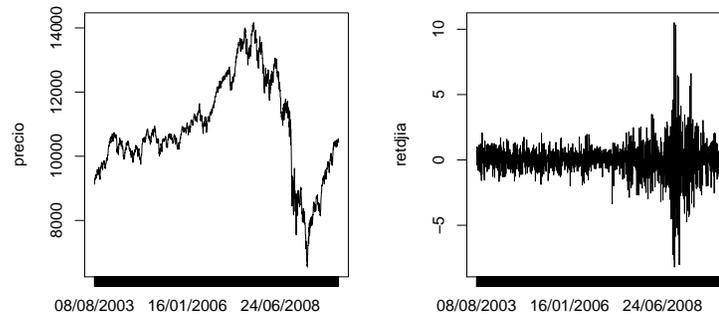


Figura 2.2: Correlograma de las rentabilidades y valor absoluto de las rentabilidades del DJIA

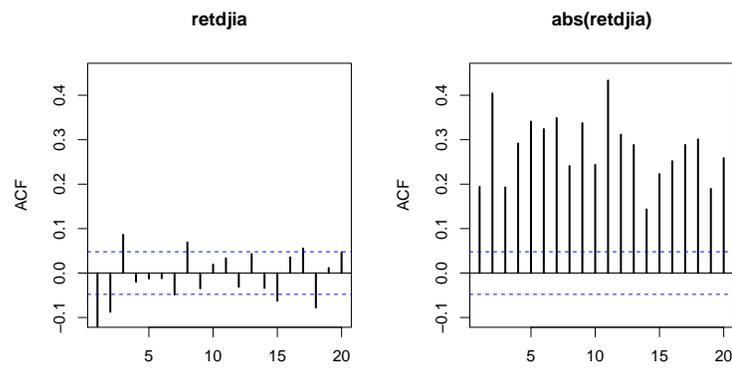
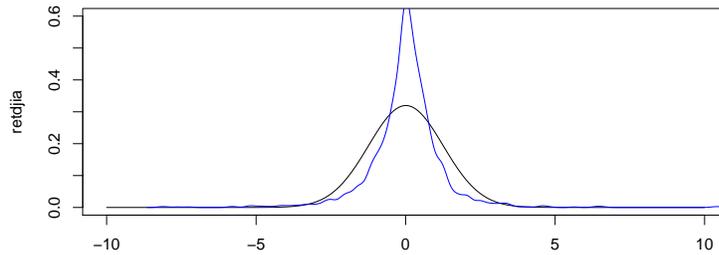


Figura 2.3: Densidad de las rentabilidades del DJIA y ajuste a la normal



varios valores de retardo: 20, $\log(T)$ (Tsay, 2010, p. 33), $T/4$; donde T es la longitud de la serie.

Mientras que la característica empírica (CE-5) se observa en el decaimiento lento de la autocorrelación en el valor absoluto de los rendimientos del DJIA.

La Figura 2.3 presenta la densidad de los datos (línea azul) y el ajuste a la densidad normal (línea negra). No se aprecia claramente la característica empírica de asimetría en la distribución de las rentabilidades (CE-3); sin embargo, se observa que los datos presentan colas más pesadas que las de la distribución normal (CE-2). Esto se corrobora apreciando el valor de exceso de curtosis, 12.37 y el coeficiente de asimetría igual a 0.03 en el Cuadro 2.1.

Cuadro 2.1: Estadística descriptiva de las rentabilidades del DJIA

Estadístico	DJIA
Media	0.0080
Mediana	0.0298
Desviación estándar	1.2476
Varianza	1.5564
Coficiente de asimetría	0.0351
Exceso de curtosis	12.3739
Mínimo	-8.2005
Máximo	10.5083

Dadas las características de las rentabilidades diarias del DJIA, se puede emplear un modelo ARMA-GARCH con innovaciones cuya distribución presente colas pesadas para caracterizar dichas rentabilidades. Esto se verá más adelante en los Ejemplos 2 y 3.

2.1. Value-at-Risk

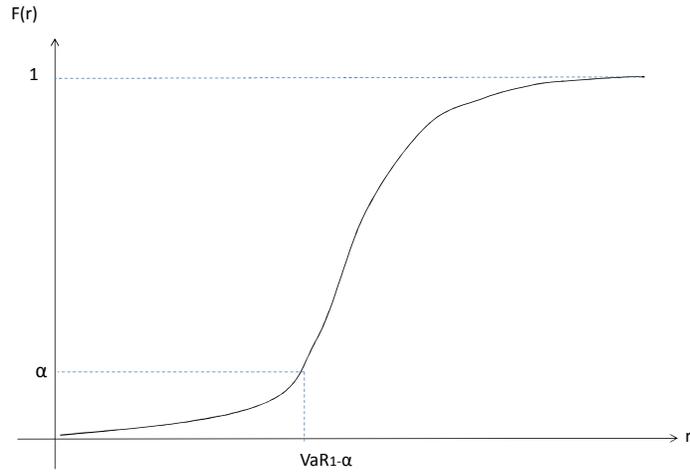
Existe una amplia literatura dedicada al VaR, como por ejemplo Jorion (2006), Alexander (2008b), Hubbert (2012), Hull (2012), entre otros. Las medidas actuales de riesgo se basan en medidas estadísticas que describen la distribución condicional de pérdidas, y la más usada en la industria financiera es la medida Value-at-Risk (VaR). El VaR se puede definir como la máxima pérdida que se puede esperar de un título o portafolio en un periodo de tiempo (un día ó 10 días para propósitos regulatorios), dado un nivel específico, $100(1 - \alpha) \%$, de probabilidad o también denominado nivel de confianza (Crouhy, Galai y Mark, 2005, p. 154). Para riesgo de mercado, el Comité de Basilea sugiere estimar el VaR al 99 % de nivel de confianza. El VaR trata de responder a la pregunta de cuánto es la pérdida que será excedida solamente en el $\alpha 100 \%$ de los casos en el siguiente periodo de negociación. Suponiendo que la variable aleatoria que representa la rentabilidad, R , tiene asociada una función de distribución F , el VaR es:

$$VaR_{1-\alpha}(R) = \inf \{r \in \mathbb{R} : F(r) \geq \alpha\}. \quad (2.1)$$

En otras palabras, $VaR_{1-\alpha}(R) = F^{-1}(\alpha)$, es decir, el VaR es un cuantil de la función de distribución de rentabilidades. Esto se muestra en la Figura 2.4, en el gráfico de la función de distribución acumulada (FDA). Por tanto, el VaR es una medida estadística que define un nivel particular de pérdida en términos de sus probabilidades de ocurrencias o nivel de confianza (Crouhy, Galai y Mark, 2005, p. 8). En adelante asumiremos la hipótesis usual de que las rentabilidades son de la forma $R_{t+1} = \mu_{t+1} + \sigma_{t+1}Z_{t+1}$.⁴ En este caso, el VaR atiende a una expresión que es sencillo de derivar. Efectivamente, si α el nivel de probabilidad al cual se desea calcular el VaR, entonces usaremos la siguiente expresión:

⁴Véase por ejemplo McNeil, Frey y Embrechts, 2005, p. 159-161.

Figura 2.4: Representación gráfica del VaR



$$\begin{aligned}
 Pr(R_{t+1} \leq VaR_{1-\alpha}) &= \alpha, \\
 Pr(\mu_{t+1} + \sigma_{t+1}Z_{t+1} \leq VaR_{1-\alpha}) &= \alpha, \\
 Pr\left(Z_{t+1} \leq \frac{VaR_{1-\alpha} - \mu_{t+1}}{\sigma_{t+1}}\right) &= \alpha, \\
 \frac{VaR_{1-\alpha} - \mu_{t+1}}{\sigma_{t+1}} &= q_{\alpha}(Z_{t+1}), \\
 VaR_{1-\alpha} &= \mu_{t+1} + \sigma_{t+1}q_{\alpha}(Z_{t+1}),
 \end{aligned} \tag{2.2}$$

donde Z_{t+1} es el proceso de innovaciones con función de distribución G centrada en cero y varianza unitaria, y $q_{\alpha}(Z_{t+1})$ es el cuantil a nivel α de Z_{t+1} . Para modelizar la varianza usaremos el modelo GARCH que se describe a continuación.

2.2. Modelos GARCH

Suponiendo que la serie temporal de rentabilidades, r_1, r_2, \dots, r_{t+1} , sigue una función de densidad conjunta dada por $f(r_1, r_2, \dots, r_{t+1}|\theta)$, se deduce que la varianza condicional es $Var(r_{t+1}|\theta, r_1, r_2, \dots, r_t)$. Entonces los modelos de volatilidad cambiante con el tiempo describen un proceso para la volatilidad condicional, es decir, $Var(r_{t+1}|\theta, r_1, \dots, r_t)$. Una distribución condicional es una distribución que gobierna un rendimiento en un instante de tiempo. La volatilidad condicional en el momento t es la raíz cuadrada de la varianza de la distribución condicional en el momento t , y se denota como σ_t^2 (Alexander, 2001, p.12). Los modelos GARCH son los más populares para modelizar la varianza cambiante con el tiempo o heteroscedasticidad condicional. Los modelos GARCH fueron introducidos por Engle (1982) y Bollerslev (1986)⁵ con el fin de capturar la característica de volatilidad cambiante y aglomeración de volatilidad de las rentabilidades financieras. El modelo GARCH deriva su nombre debido a que es un modelo de varianza cambiante (heteroscedasticidad condicional), es un modelo auto-regresivo y es la versión generalizada de los modelos ARCH propuesto por Engle (1982).

En este documento usaremos el modelo ARMA(1,1)-GARCH(1,1), y a manera de introducción se presenta el modelo GARCH(1,1). Este modelo, generalmente, es el mejor modelo en cuanto a capacidad predictiva, entre toda la familia de modelos tipo GARCH (Lunde y Hansen, 2005).

$$\begin{aligned} R_t &= \epsilon_t, \\ \epsilon_t &= Z_t \sigma_t, \quad Z_t \sim G(0, 1), \\ \sigma_{t+1}^2 &= \omega + \alpha \epsilon_t^2 + \beta \sigma_t^2. \end{aligned} \tag{2.3}$$

En la ecuación anterior, hemos denotado $Var(r_{t+1}|\theta, r_1, \dots, r_t)$ mediante σ_{t+1}^2 , y es la notación que usaremos de aquí en adelante. En (2.3) ϵ_t denota el shock del mercado o el rendimiento inesperado ($R_t = \epsilon_t$), mientras que Z_t es la innovación que se asume se distribuye según una densidad conocida G (es decir, no depende de parámetros extra) centrada en cero y varianza unitaria. Además, ω es la constante del modelo, α representa la sensibilidad de la va-

⁵Por estos trabajos, Engle y Bollerslev recibieron el Premio Nobel de Economía en 2003.

varianza a shocks del mercado,⁶ y β representa la persistencia en la varianza. Entonces $\alpha + \beta$ determina la tasa de convergencia de la varianza condicional al nivel medio de largo plazo⁷ (Alexander, 2008a, p. 137). Usualmente, para series de rentabilidades financieras, los valores estimados para α están alrededor de 0.10, y para β por encima de 0.90. En los estudios iniciales se usaba la distribución normal estándar como G , pero como ya argumentamos en el capítulo anterior, esta aproximación no es buena porque sus colas decaen exponencialmente y, por consiguiente, no permite capturar de manera adecuada los eventos extremos. En este trabajo utilizaremos como G modelos de colas pesadas, como se verá en 2.3.

La varianza de largo plazo está dada por (al tomar esperanza incondicional de (2.3)):

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}. \quad (2.4)$$

Para asegurar que la varianza de largo plazo exista y sea positiva, de (2.4) se infiere que $\omega > 0$ y $\alpha + \beta < 1$. El GARCH(1,1) es de covarianza-estacionaria si y sólo si $\alpha + \beta \leq 1$ (Proposición 4.21 de McNeil, Frey y Embrechts, 2005, p. 146). Otra restricción para el modelo es $\alpha, \beta \geq 0$. Un caso específico del modelo GARCH es el modelo de alisado exponencial (EWMA, por sus siglas en inglés), donde $\omega = 0$ y $\alpha + \beta = 1$. Con estos valores, la varianza de largo plazo no está definida para el modelo de alisado exponencial, lo cual lo convierte en una desventaja frente al modelo GARCH.

Al despejar ω de (2.4) y reemplazarla en (2.3) se obtiene:

$$\sigma_{t+1}^2 = \sigma^2 + \alpha(\epsilon_t^2 - \sigma^2) + \beta(\sigma_t^2 - \sigma^2). \quad (2.5)$$

Esto es, la varianza en $t + 1$ es un promedio ponderado⁸ de la varianza de largo plazo, el rendimiento inesperado al cuadrado y la varianza en t . En otras palabras, la varianza de mañana es la varianza de largo plazo con algo añadido (menos) si el rendimiento inesperado de hoy está por encima

⁶Cuando α es relativamente alto, por encima de 0.1, entonces la varianza es muy sensible a eventos del mercado.

⁷Cuando $\alpha + \beta$ es relativamente alto, por encima de 0.99, entonces la predicción de la varianza será relativamente plana.

⁸Donde los pesos son α y β .

(debajo) de su varianza de largo plazo, y algo añadido (menos) si la varianza de hoy está por encima (debajo) de su varianza de largo plazo (Christoffersen, 2003, p. 23).

Otra característica del modelo GARCH es que permite modelizar rentabilidades con colas pesadas. Efectivamente, si denotamos K_Z y K_R la curtosis de Z_t y de R_t del modelo (2.3), respectivamente, entonces

$$K_R = K_Z \frac{1 - (\alpha + \beta)^2}{1 - (\alpha + \beta)^2 - (K_Z - 1)\alpha^2}. \quad (2.6)$$

La curtosis de R_t es mayor que la de Z_t si $K_Z > 1$. Note que si el término del denominador, $(K_Z - 1)\alpha^2 > 0$, entonces el numerador sería mayor al denominador, y esta fracción multiplica a K_Z . Para el caso $Z \sim N(0, 1)$, es decir, $G = N(0, 1)$, se tiene:

$$K_R = 3 \frac{1 - (\alpha + \beta)^2}{1 - (\alpha + \beta)^2 - 2\alpha^2}, \quad (2.7)$$

y $K_R > 3$, puesto que $K_Z = 3$ y $\alpha^2 > 0$. Sin embargo, esta curtosis puede ser pequeña comparada con la de las rentabilidades financieras (Alexander, 2008a, p. 157).

Los parámetros del modelo GARCH se estiman mediante máxima verosimilitud⁹ (ML, por sus siglas en inglés). La idea consiste en maximizar la función de (log)verosimilitud mediante algoritmos numéricos, un método muy usado es BHHH¹⁰ (McNeil, Frey y Embrechts, 2005, p. 152). Sin embargo, algunas funciones de verosimilitud pueden presentar varios máximos locales. Entonces, es recomendable variar las condiciones iniciales en el algoritmo numérico y fijar un buen criterio de convergencia. Otro problema, es que la función de verosimilitud puede ser muy plana y esto complicaría aún más la estimación de los parámetros mediante máxima verosimilitud. En otras ocasiones, algunos parámetros pueden ocasionar el problema de división entre dos ceros, causando un resultado de indefinición (Danielsson, 2011, p. 43).

⁹Véase por ejemplo Casella y Berger, 2002, Sección 7.2.2; para más detalles.

¹⁰Representa las iniciales de sus autores: Berndt, Hall, Hall and Hausmann.

Suponiendo $Z_t \sim N(0, 1)$ la función de logverosimilitud,¹¹ para estimar los parámetros del modelo GARCH, está dada por:

$$\text{loglik} = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log \sigma_t^2 - \frac{1}{2} \sum_{t=1}^T \frac{\epsilon_t^2}{\sigma_t^2}. \quad (2.8)$$

Una vez se han estimado los parámetros del modelo GARCH y verificar que el modelo está correctamente especificado (ausencia de autocorrelación en los residuos estandarizados al cuadrado y significancia de los coeficientes estimados), se procede a predecir. Entonces se puede utilizar un modelo ARMA para obtener la predicción de R_{t+1} , esto es útil para el presente trabajo de investigación, y en la literatura se suele denominar modelo ARMA con “errores” GARCH, nosotros lo llamaremos ARMA-GARCH.

Modelo ARMA-GARCH

El modelo ARMA-ARCH fue inicialmente introducido por Weiss (1984), para una revisión de los modelos ARMA-GARCH véase por ejemplo Li, Ling y McAleer (2002). En un modelo ARMA-GARCH, la media μ_t sigue un proceso ARMA y la varianza sigue un proceso GARCH (como el visto). En concreto

$$\mu_t = \mu + \phi_1 \mu_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t \quad (2.9)$$

Donde μ es la constante del modelo, ϕ_1 es el parámetro de un modelo AR(1), θ_1 es el parámetro de un modelo MA(1) y $\epsilon_t \sim G(0, \sigma^2)$.

Sea R_t el proceso de rentabilidades logarítmicas dado por $R_t = \mu_t + \sigma_t Z_t$, y se asume que R_t es un proceso estacionario (de covarianza-estacionaria). Sea Z_t las innovaciones independientes e idénticamente distribuidas (i.i.d.) que siguen una distribución G , con media cero y varianza unitaria. El modelo ARMA(1,1)-GARCH(1,1) es de la forma:

¹¹Algunas veces se elimina el primer término del lado derecho de la ecuación (2.8) por ser una constante.

$$\begin{aligned}
\mu_t &= \mu + \phi_1 \mu_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t, \\
Z_t &= \epsilon_t / \sigma_t, \quad Z_t \sim G(0, 1), \\
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2.
\end{aligned} \tag{2.10}$$

Para modelizar la media condicional μ_{t+1} se emplea un ARMA(1,1), y para la volatilidad condicional σ_{t+1} un GARCH(1,1). Estos parámetros son estimados mediante máxima verosimilitud. Ejemplos y más detalles de modelos ARMA y GARCH en aplicaciones a administración del riesgo se pueden encontrar en McNeil et al. (2005; capítulo 4 y sección 7.2.6), y en Jondeau et al. (2007; capítulos 4 y 8).

Para calcular VaR, entonces se combina el modelo presentado en (2.10) junto con la expresión (2.2):

$$\begin{aligned}
\mu_{t+1} &= \mu + \phi_1 \mu_t + \theta_1 \epsilon_t + \epsilon_{t+1}, \\
Z_{t+1} &= \epsilon_{t+1} / \sigma_{t+1}, \quad Z_{t+1} \sim G(0, 1), \\
\sigma_{t+1}^2 &= \omega + \alpha \epsilon_t^2 + \beta \sigma_t^2, \\
VaR_{t+1} &= \mu_{t+1} + \sigma_{t+1} q_\gamma(Z_{t+1}).
\end{aligned} \tag{2.11}$$

Donde $q_\gamma(Z_{t+1})$ es el cuantil de Z_{t+1} a nivel γ .

Ejemplo 2. (GARCH-normal)

Se aplica un modelo GARCH(1,1) y ARMA(1,1)-GARCH(1,1) a los datos analizados en el Ejemplo 1. El Cuadro 2.2 muestra los resultados de la estimación de los parámetros del modelo GARCH(1,1).

Las estimaciones resultan ser significativas a cualquier nivel razonable de significación. El valor de $\alpha + \beta$ es de 0.989, lo que indica alta persistencia en volatilidad.

Se observa en la Figura 2.5, un periodo de alta volatilidad, en especial 2008, correspondiente al efecto de la crisis subprime. En la Figura 2.6 se presenta el correlograma de los residuos estandarizados al cuadrado (lado

Cuadro 2.2: Estimación Modelo GARCH(1,1)

Modelo GARCH(1,1)		
Parámetro	Estimado	P-valor
ω	0.0098	0.0002
α	0.0670	0.0000
β	0.9220	0.0000

izquierdo) y el QQ-Plot¹² (Quantile-Quantile plot) de los residuos estandarizados.

Del correlograma se observa que el modelo está bien especificado, mientras que el QQPlot muestra que el modelo GARCH (con innovaciones distribuidas normal) no alcanza a capturar las colas pesadas de los datos y es necesario una distribución que permita recoger esta característica.¹³

Finalmente, el valor de logverosimilitud es de -2180.397. Este valor sirve para comparar con otros modelos, y generalmente se escoge el de mayor logverosimilitud (si los modelos son comparables en complejidad).

El Cuadro 2.3 muestra los resultados al estimar el modelo ARMA(1,1)-GARCH(1,1). Se observa que las estimaciones de los parámetros de la parte ARMA resultan ser no significativas, y las estimaciones de los parámetros de la parte GARCH no varían sustancialmente con respecto al anterior modelo.

Los resultados del correlograma y del QQPlot son similares al caso anterior, como se observa en la Figura 2.7. El valor de logverosimilitud es de -2173.88, mayor que el calculado para el modelo GARCH(1,1) (-2180.397).

Un criterio de información usualmente empleado para comparar dos mo-

¹²Esta es una herramienta estadística valiosa para determinar si una serie de datos de tamaño n proviene de una población con una distribución paramétrica específica. Si los cuantiles de una distribución paramétrica están linealmente relacionados con los cuantiles de los datos de una muestra en el QQ-plot, se obtiene un ajuste de dicha distribución paramétrica a los datos a la muestra. Para este caso se utiliza distribución normal como distribución paramétrica.

¹³El p-valor del contraste Ljung-Box es de 0.1489 y 0.6310 cuando se emplean $\log(T)$ y $T/4$ en los retardos. Esto quiere decir que no existe evidencia estadística para rechazar la hipótesis nula de cero autocorrelación en los residuos estandarizados. Mientras que el contraste de Jarque-Bera rechaza la hipótesis nula de normalidad de los residuos estandarizados.

Figura 2.5: Serie temporal de rentabilidades y de volatilidad

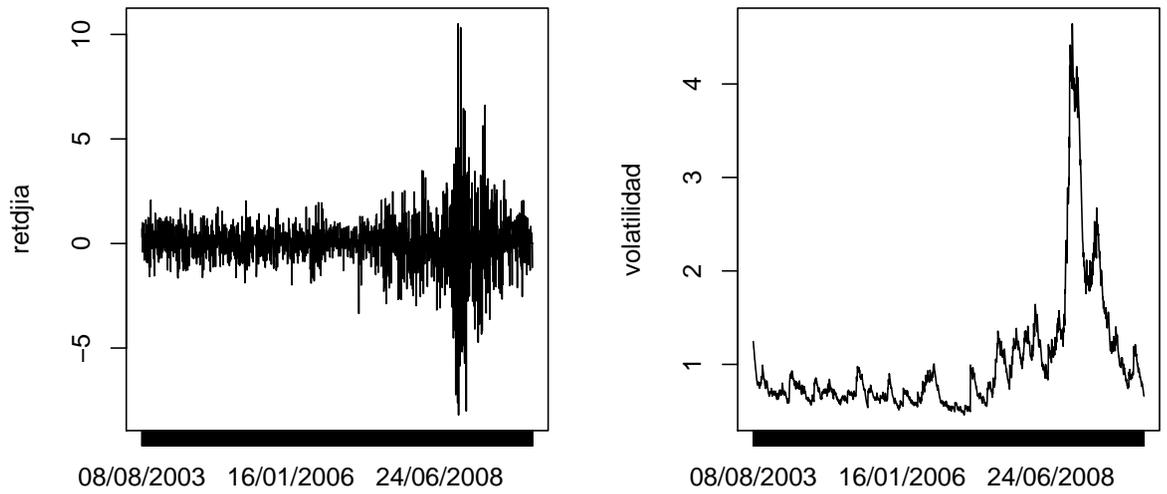
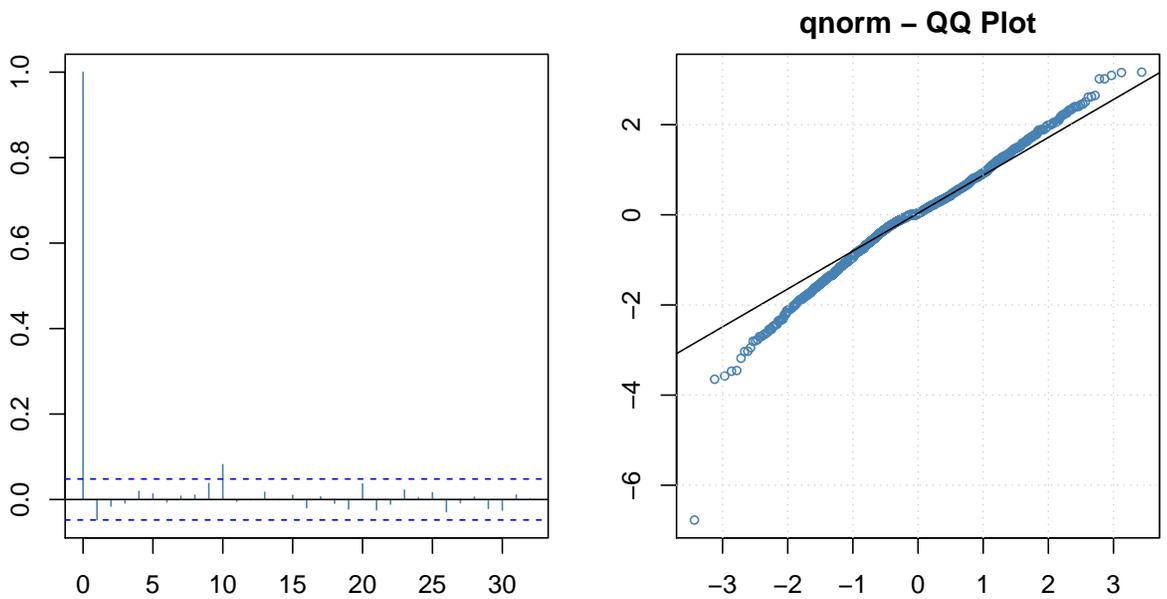


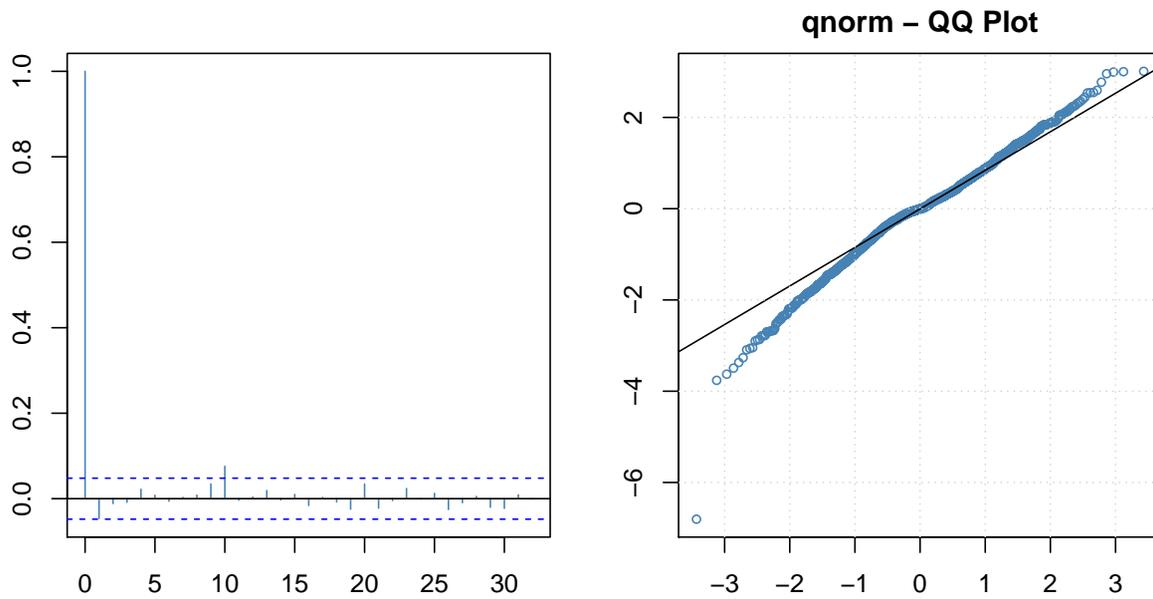
Figura 2.6: Correlograma y QQ-Plot de residuos estandarizados



Cuadro 2.3: Estimación Modelo ARMA(1,1)GARCH(1,1)

Modelo ARMA(1,1)-GARCH(1,1)		
Parámetro	Estimado	P-valor
μ	0.0307	0.1257
ϕ_1	0.2654	0.4782
θ_1	-0.3321	0.3621
ω	0.0099	0.0002
α	0.0682	0.0000
β	0.9208	0.0000

Figura 2.7: Correlograma y QQ-Plot de residuos estandarizados



delos es el AIC (Akaike Information Criterion), el cual penaliza el valor de logverosimilitud con el número de parámetros incluidos en el modelo.

$$AIC = 2p - 2\loglik. \quad (2.12)$$

Donde p es el número de parámetros del modelo y \loglik es el valor de logverosimilitud.

El AIC para el modelo GARCH(1,1) es de 4366.794, mientras que para el de ARMA(1,1)-GARCH(1,1) es de 4359.760, y se escoge el modelo con menor valor de AIC, en este caso es el modelo ARMA(1,1)-GARCH(1,1).

2.3. Distribuciones

En este trabajo se emplean las siguientes distribuciones para las innovaciones (lo que en la ecuación 2.11 llamamos G) del modelo GARCH: (i) la distribución normal, (ii) la distribución t-Student, (iii) la distribución α -estable, (iv) la distribución de Pareto generalizada¹⁴ (GPD, por sus siglas en inglés), y (v) la distribución g-h.

(i) Distribución normal

La función de densidad de una normal viene dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}. \quad (2.13)$$

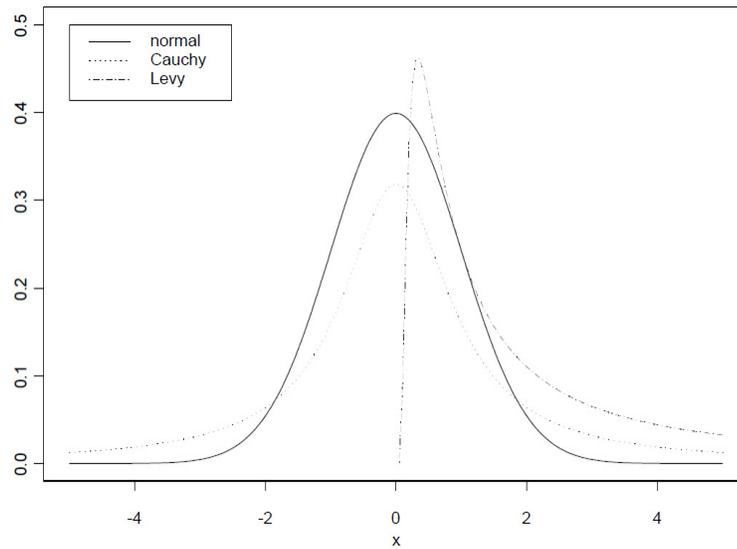
Donde μ y σ^2 son los parámetros de media y varianza de una normal.

(ii) Distribución t-Student

La densidad de una t-Student está dada por:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}. \quad (2.14)$$

¹⁴Esta es la función empleada en el método picos sobre el umbral (POT, por sus siglas en inglés) basado en la teoría del valor extremo (EVT, por sus siglas en inglés).

Figura 2.8: Tres casos especiales de la α -estable

Fuente: Tomado de Nolan (2013)

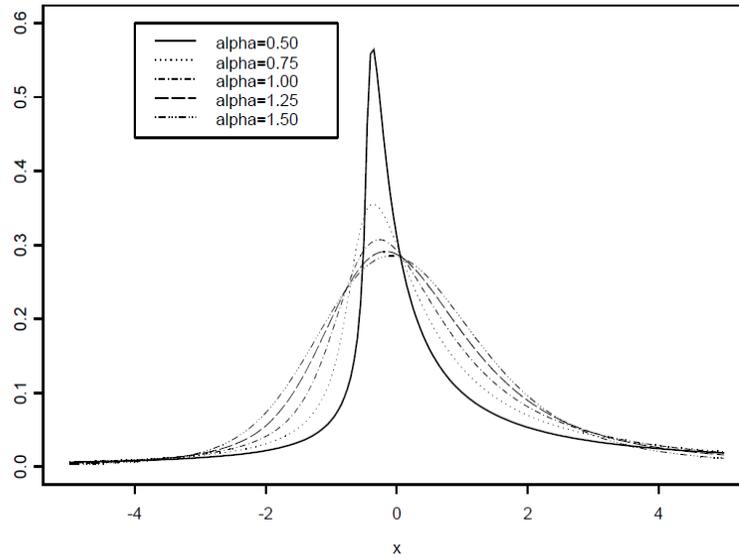
Donde Γ representa la función gamma y ν los grados de libertad. En la práctica, este valor está entre 6 y 10 (Danielsson, 2011, p. 41). Recuerde que la distribución t-Student se aproxima a una normal cuando $\nu \rightarrow \infty$.

(iii) Distribución α -estable

En general la distribución α -estable se describe mediante su función característica. Sólo en tres casos su función de densidad tiene fórmula cerrada que representan la distribución normal, Cauchy y de Lévy. La Figura 2.8 muestra las densidades para cada uno de estos casos.

La función característica está dada por:

$$E[e^{i\theta X}] = \begin{cases} \exp \left\{ -\sigma^\alpha |\theta|^\alpha \left(1 - i\beta \operatorname{sgn}(\theta) \tan \frac{\pi\alpha}{2} \right) + i\mu\theta \right\}, & \text{si } \alpha \neq 1, \\ \exp \left\{ -\sigma |\theta| \left(1 + i\beta \frac{2}{\pi} \operatorname{sgn}(\theta) \log |\theta| + i\mu\theta \right) \right\}, & \text{si } \alpha = 1. \end{cases}$$

Figura 2.9: Forma de la densidad a diferentes valores del parámetro α 

Fuente: Tomado de Nolan (2013)

Donde sgn es la función signo:

$$\text{sgn}(\theta) = \begin{cases} 1, & \text{si } \theta > 0, \\ 0, & \text{si } \theta = 0, \\ -1, & \text{si } \theta < 0. \end{cases}$$

El parámetro α es el índice de estabilidad, β es el parámetro de asimetría, μ el de localización y σ el de escala. Los rangos permisibles para los parámetros son: $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\sigma \geq 0$ y μ puede tomar cualquier valor real. La Figura 2.9 presenta la forma de la densidad α -estable a diferentes valores de α y con $\beta = 0,5$.

Cuando $\alpha = 2$ y $\beta = 0$, la distribución α -estable toma la forma de la distribución normal, y la cola de la distribución tiene varianza finita. Entre menor sea el valor del índice de estabilidad ($0 < \alpha \leq 2$) más pesada será la cola de la distribución. Si $1 < \alpha < 2$ se tiene el caso de media finita, pero

varianza infinita. Mientras que si $\alpha < 1$ no se tiene media finita. La ventaja de usar este tipo de distribuciones es que la suma de dos variables independientes α -estables es una distribución α -estable. Esta propiedad es deseable para los rendimientos de activos financieros tanto para la administración de riesgos como para la administración de carteras. Sin embargo, una desventaja es que puede presentar varianza infinita y que no hay una fórmula cerrada para la función de densidad, lo que complica un poco la estimación de sus parámetros. Generalmente se emplean tres métodos de estimación: el método de cuantiles, método generalizado de momentos (GMM) y mediante máxima verosimilitud.

La técnica basada en cuantiles fue inicialmente sugerido por Fama y Roll (1971), y posteriormente McCulloch (1986) desarrolló una técnica que emplea cinco cuantiles de una muestra para estimar α y β sin sesgo asintótico; sin embargo, se restringe a que $\alpha > 0.6$.

Otro método consiste en igualar los momentos obtenidos por la función característica empírica junto con la teórica. Carrasco y Florens (2000) introducen el método GMM para el caso continuo de condiciones de momentos.

Para utilizar el método de máxima verosimilitud, se emplean métodos numéricos, puesto que no se conoce la función de densidad en forma explícita. Para aproximar la función de densidad se usan métodos como la transformada rápida de Fourier (FFT, por sus siglas en inglés) o métodos de integración directa. Para el primer caso véase por ejemplo Mittnik et al. (1999), y para el segundo Nolan (2001).

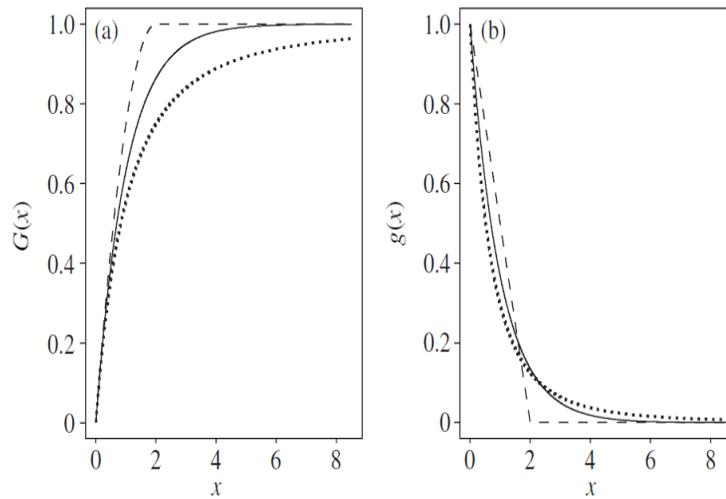
(iv) Distribución de Pareto Generalizada

La distribución de Pareto generalizada (GPD, por sus siglas en inglés) es empleada en el método POT que se describirá más adelante. La función de distribución está dada por:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}}, & \text{si } \xi \neq 0, \\ 1 - \exp\left\{-\frac{x}{\beta}\right\}, & \text{si } \xi = 0. \end{cases}$$

Donde ξ representa el parámetro de forma, y β el de escala. Si $\xi \geq 0$,

Figura 2.10: Función de distribución y de densidad de GPD



Fuente: Tomado de McNeil, Frey y Embrechts (2005, p. 276)

entonces $\beta > 0$ y $x \geq 0$. Mientras que si $\xi < 0$, entonces $0 \leq x \leq -\beta/\xi$. Esta distribución contiene tres casos dependiendo del valor del parámetro de forma. Cuando $\xi > 0$ toma la forma de la distribución Pareto, cuando $\xi = 0$ es la distribución exponencial, y cuando $\xi < 0$ es la distribución Pareto tipo II. El caso de interés en finanzas es cuando $\xi > 0$ que permite modelizar rentabilidades con colas pesadas.

En el panel (a) de la Figura 2.10 se presentan tres casos de la función de distribución Pareto generalizada. En línea sólida se representa el caso de la exponencial ($\xi = 0$), en línea punteada la Pareto ($\xi = 0,5$) y el otro caso la Pareto tipo II ($\xi = -0,5$). En el panel (b) se representan sus respectivas funciones de densidad. Para todos los casos, el parámetro de escala $\beta = 1$. Se observa como la cola de la distribución Pareto decae más lentamente que los otros casos.

Gracias al teorema de Pickands, Balkema y de Haan (veáse por ejemplo Embrechts et al., 1997), la distribución de las pérdidas en exceso (pérdidas por encima de un umbral, u), se pueden aproximar a la distribución de Pareto

generalizada cuando el nivel de umbral es lo suficientemente alto. El teorema dice que para una función de distribución F (que está en el dominio de atracción de la distribución de valor extremo generalizada), cuando el umbral u se incrementa progresivamente, la distribución de excesos F_u converge a la distribución de Pareto generalizada.

Cálculo del cuantil en GPD

Para calcular el cuantil a un nivel α se usa (véase McNeil et al., 2005, p. 283):

$$q(Z) = u + \frac{\beta}{\xi} \left\{ \left(\frac{1 - \alpha}{N_u/n} \right)^{-\xi} - 1 \right\}. \quad (2.18)$$

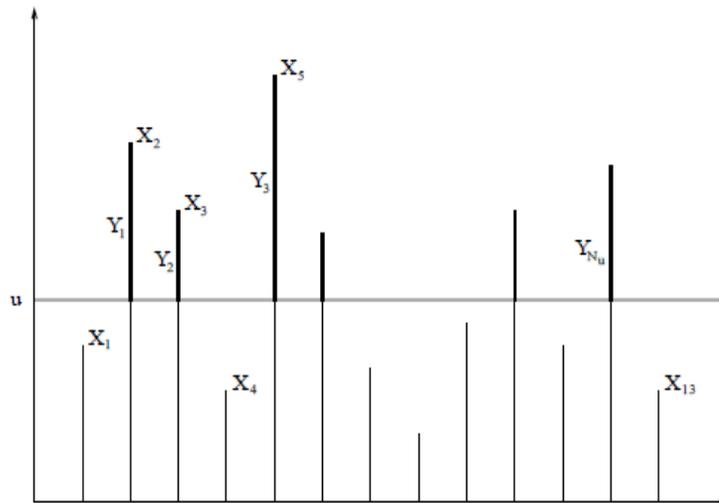
Donde u es el umbral escogido, β y ξ son los parámetros de escala y de forma de la GPD, N_u es la cantidad de datos que exceden el umbral y n el total de datos de la muestra. Por lo tanto, N_u/n es un estimador empírico de la cola de los datos que han superado un umbral u .

El Método Picos sobre el Umbral (POT)

La teoría del valor extremo (EVT) es una técnica estadística basada en modelos distribucionales obtenidos a partir de la teoría asintótica y es útil para analizar las colas de las distribuciones. Existen dos clases de modelos para tratar valores extremos: máximos por bloque (block maxima) y picos sobre el umbral (POT, por sus siglas en inglés). El método POT es útil para grandes pérdidas (o ganancias) las cuales exceden un umbral alto u . Este método es la técnica más usada para analizar la cola de una función de distribución, debido al uso eficiente de los datos en valores extremos. Adicionalmente, desde el punto de vista de un gestor de riesgos, se está interesado en las pérdidas que exceden un umbral u . Por tal razón, la metodología POT es la que se empleará en este trabajo. Este enfoque ha sido estudiado por Smith (1989), Davison y Smith (1990) y Leadbetter (1991) quienes muestran el uso práctico de POT en la teoría del valor extremo.

En la Figura 2.11, X_1, \dots, X_{13} pueden representar datos de pérdida que provienen de una función de distribución (desconocida) F . Dado un nivel de

Figura 2.11: Representación gráfica del método POT



Fuente: Tomado de Embrechts, Klüppelberg y Mikosch (1997), p. 353.

umbral u , un número aleatorio Nu de pérdidas excederá el umbral. Dichos excesos se representan en el gráfico como Y_1, \dots, Y_{Nu} , y se denominan las pérdidas en exceso.

Sin embargo, una desventaja que presenta este método es la selección del umbral. La selección del umbral conlleva a un compromiso entre sesgo y varianza en la estimación de los parámetros de la GPD mediante máxima verosimilitud. Un umbral muy alto resultaría en alta varianza en la estimación de los parámetros; mientras que un umbral muy bajo resultaría en alto sesgo, en especial en la estimación de ξ . Existen varias alternativas para tratar de solucionar este problema (veáse por ejemplo Beirlant et al., 2004); sin embargo, no existe un método óptimo que permita la correcta selección del umbral. Un método alternativo para estimar ξ es mediante el gráfico de Hill que es un método visual y está basado en el estimador del mismo nombre. Este estimador se calcula como:

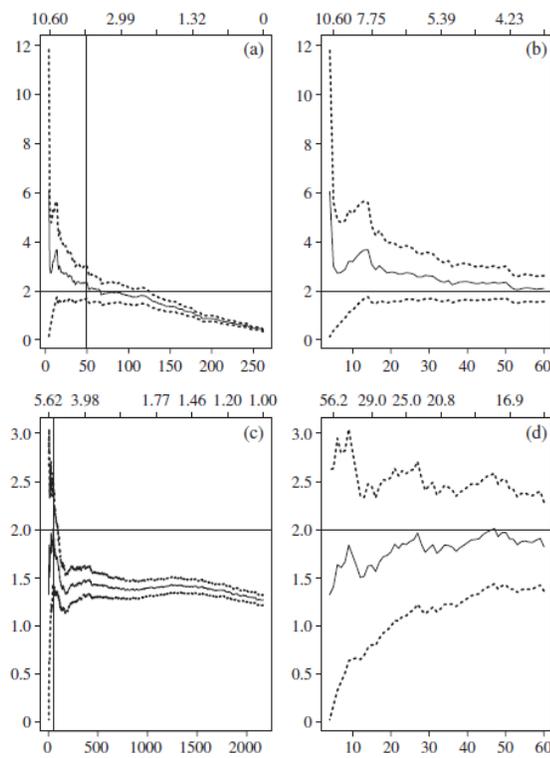
$$\hat{\xi}_{k,n} = \frac{1}{k} \sum_{j=1}^k \log X_{j,n} - \log X_{k,n}, \quad k = 2, \dots, n. \quad (2.19)$$

Donde X puede representar los datos de pérdida y n el total de los datos. Entonces el gráfico de Hill es de la forma $\{k, \hat{\xi}_{k,n} : k = 2, \dots, n\}$. La idea consiste en calcular el estimador de Hill para diferentes valores de k y escoger el estimador de ξ donde se observe una región plana en el gráfico de Hill (Reiss y Thomas 2007, p. 136).

En el panel (a) y en el panel (b) de la Figura 2.12 se representan pérdidas semanales en porcentaje de AT& T, mientras que en el panel (c) y en el panel (d) las pérdidas por indemnizaciones de incendio a compañías de seguros en Dinamarca. En el eje de las abscisas se representa k y en el de las ordenadas $1/\hat{\xi}$. En la práctica se sugiere escoger k entre 10 y 50. En este caso en el panel (b) y en el panel (d) se observa una parte plana en el gráfico de Hill.

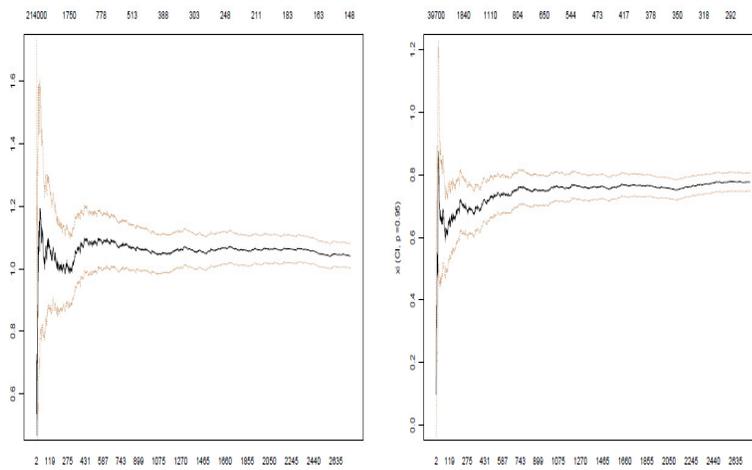
Sin embargo, en algunas ocasiones este método puede llevar a estimaciones erróneas. En la Figura 2.13 se han simulado datos provenientes de una distribución g-h con parámetros $g = 0.1$ y $h = 1$ para el lado izquierdo. Mientras que para el lado derecho $g = 2$ y $h = 0.2$. Los verdaderos valores de ξ serían 1 y 0.2 respectivamente. En el eje de las ordenadas se representa $\hat{\xi}$.

Figura 2.12: Gráfico de Hill



Fuente: Tomado de McNeil, Frey y Embrechts (2005, p. 288).

Figura 2.13: Ejemplo de “Horror Hill Plot”



Fuente: Tomado de Degen, Embrechts y Lambrigger (2007).

Se observa como el gráfico de Hill es plano alrededor de 1 en el lado izquierdo, mientras que para el lado derecho se observa una parte plana alrededor de 0.7, cuando el verdadero valor es 0.2. Otro método visual empleado para estimar el cuantil en la teoría del valor extremo es el gráfico de excesos medios (mean excess plot). Sin embargo, existen problemas al usar esta herramienta. Una de ellas es que algunas distribuciones (no cercanas a la GPD) pueden engañar el diagnóstico del gráfico de excesos medios (Ghosh y Resnick, 2010).

En este trabajo se empleará un umbral correspondiente al 10 % de los datos de la cola de la distribución como en McNeil y Frey (2000). Para soportar la selección del umbral en este trabajo, Chavez-Demoulin (1999), basado en estudios de simulación recomienda escoger un umbral dado el percentil al 90 % de las datos de pérdidas.

(v) Distribución g-h

La distribución g-h fue introducida por Tukey en 1977 y ha sido utilizada en finanzas en aplicaciones de rentabilidades de acciones e índices, como también para modelar tipos de interés, y opciones en tipos de interés (Veáse por ejemplo Dutta y Perry, 2006, y las referencias allí contenidas).

Sea $Z \sim N(0, 1)$ una variable aleatoria normal estándar. Una variable aleatoria X se dice que tiene una distribución g-h con parámetros g y h (con $g \neq 0$ y $h \in \mathbb{R}$), si X satisface:

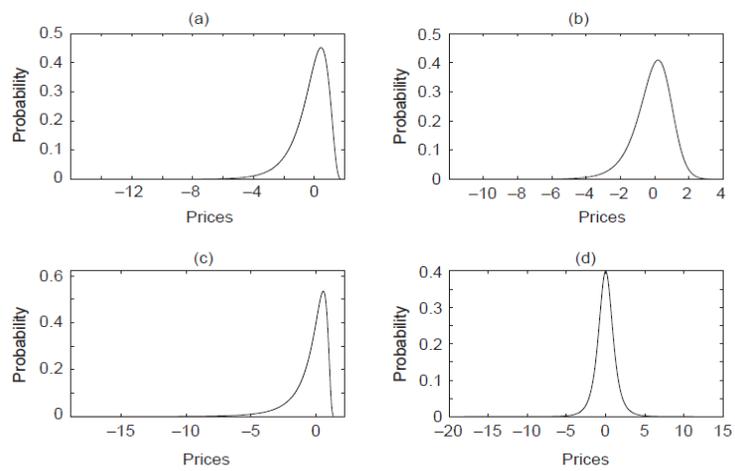
$$X = \frac{\exp(gZ) - 1}{g} \exp(hZ^2/2), \quad (2.20)$$

y se denota $X \sim \text{g-h}$.

El parámetro g controla la cantidad y dirección de asimetría, mientras que el parámetro h controla la cantidad de elongación (curtosis). La distribución es más sesgada a la derecha a mayores valores de g , y entre mayor es el valor de h , mayor es la elongación. La Figura 2.14 presenta diversas formas de la densidad de la distribución g-h.

Si además introducimos un parámetro de localización, a , y uno de escala, b , la distribución g-h tiene cuatro parámetros.

Figura 2.14: Densidad de g-h



(a) $g = -0.5, h = 0$; (b) $g = -0.2446, h = 0.053356$; (c) $g = -0.78714, h = 0.016356$; (d) $g = 0, h = 0.16355$.

Fuente: Tomado de Jiménez y Arunachalam (2011, p. 103).

$$Y = a + b \frac{\exp(gZ) - 1}{g} \exp(hZ^2/2) = a + bX. \quad (2.21)$$

Cuando $h = 0$, la ecuación (2.21) es $Y = \frac{\exp(gZ)-1}{g}$ y la distribución g-h se convierte en la distribución g que corresponde a una distribución lognormal escalada. Cuando $g = 0$, la ecuación (2.21) se interpreta como $Y = a + bZ \exp(hZ^2/2)$ y la distribución es referida como la distribución h. Cuando $g = h = 0$, Y se distribuye normal.

Propiedades de la distribución g-h

Martínez (1981), Martínez Iglewicz (1984), Hoaglin (1985), entre otros, han estudiado las propiedades de la distribución g-h. Una propiedad importante es que muchas distribuciones pueden ser generadas a partir de la g-h para valores específicos de los parámetros. Las siguientes propiedades se basan en Dutta y Babel (2002):

1. X es una transformación estrictamente creciente de Z . Esto es, la transformación de una normal estándar a una g-h es uno a uno.
2. Si $a = 0$, entonces $Y_{-g,h}(Z) = -Y_{g,h}(-Z)$. Cambiar el signo de g cambia la dirección pero no el valor de asimetría de la distribución g-h.
3. Para $h = 0$ y $g \rightarrow 0$, la distribución g-h converge a una distribución normal.
4. La distribución g es una distribución lognormal escalada.
5. El parámetro de localización a de la distribución g-h es la mediana del conjunto de datos, y el p-ésimo percentil del valor de g está dado por $g_p = -(\frac{1}{Z_p}) \log(\frac{X_{1-p} - X_{0,5}}{X_{0,5} - X_p})$. Como $Z_{0,5} = 0$, y al reemplazarlo en (2.21), se obtiene $a = X_{0,5}$. Al escoger diversos valores de p se obtienen diferentes estimados del parámetro g . Para solucionar este problema, algunos autores han sugerido usar la mediana de los diferentes valores de g_p . Jiménez y Martínez (2006) proponen la estimación del parámetro g de una manera más sencilla cuando h tiende a cero.

6. Para un cierto valor de g , el valor de h está dado por $\log \frac{g(X_p - X_{1-p})}{\exp(gZ_p) - \exp(-gZ_p)} = \log b + h(Z_p^2/2)$. De esta manera, se puede obtener el valor de h como el coeficiente de la pendiente de la regresión de $\log \frac{g(X_p - X_{1-p})}{\exp(gZ_p) - \exp(-gZ_p)}$ contra $(Z_p^2/2)$. El valor de b se obtiene tomando la función exponencial del intercepto de la regresión.

Un resultado importante del artículo de Degen et al. (2007), es que si los datos de pérdida son bien modelados por una distribución g-h (con valores $g, h > 0$), la estimación de cuantiles altos mediante el método POT, por lo general convergerá muy lentamente, y por consiguiente su estimación mediante EVT será imprecisa. Por tal razón hemos decidido incluir esta distribución en nuestro análisis. A continuación se presenta como estimar cuantiles de una distribución g-h cuando se han estimado los parámetros de la distribución. Este apartado está basado en la sección 2.1 de Degen et al. (2007).

Estimación del cuantil mediante la aproximación g-h

Puesto que $k(x) = \frac{\exp(gx) - 1}{g} \exp(hx^2/2)$ es estrictamente creciente (para $h > 0$), la función de distribución acumulada de una variable aleatoria g-h X , se puede escribir como:

$$F(x) = \Phi(k^{-1}(x)), \quad (2.22)$$

donde Φ denota la función de distribución acumulada de una normal estándar. Entonces el cuantil que nos interesa se estima de la siguiente forma:

$$q_\alpha = k(\Phi^{-1}(\alpha)), \quad 0 < \alpha < 1. \quad (2.23)$$

Capítulo 3

Métodos de estimación

Como se comentó en la sección de Motivación, cuando se cuenta con una forma analítica de la función de verosimilitud, se procede a maximizar esta función y el resultado de esta optimización son los parámetros estimados. Generalmente, se emplean métodos numéricos, pero algunas veces la función de verosimilitud puede presentar varios máximos locales o ser muy plana. Adicionalmente, algunas distribuciones no cuentan con una fórmula cerrada para su densidad, lo cual hace que el método de maximizar la función de verosimilitud no sea aplicable o requiera de una extracción numérica. Este es el caso de la distribución α -estable y g-h. El objetivo de este capítulo es presentar el método de máxima verosimilitud para las distribuciones que presentan una forma explícita de la función de verosimilitud (normal, t-Student, GPD) y otros métodos de estimación para los que no se cuenta con una fórmula cerrada de su función de verosimilitud (g-h y α -estable).

Los datos analizados correspondientes a rentabilidades de activos financieros, como en algunos casos (véase por ejemplo, Berkowitz y O'Brien, 2002; Güner et al., 2010), se filtran mediante un modelo ARMA(1,1)-GARCH(1,1). Para estimar los parámetros del proceso ARMA(1,1)-GARCH(1,1) se emplea el método de máxima verosimilitud cuando las innovaciones se distribuyen normal y t-Student. Para los otros casos (α -estable, GPD y g-h), se emplea un método de dos pasos. En un primer paso se filtran las rentabilidades mediante un ARMA(1,1)-GARCH(1,1) usando el método de cuasi-máxima verosimilitud (QMLE, por sus siglas en inglés), y en un segundo paso se ajustan las distribuciones analizadas a los residuos estandarizados resultantes de

la primera etapa.

3.1. Método de estimación para la distribución normal y t-Student

La función de verosimilitud para estos casos está dada por:

$$L(\Theta; \mathbf{R}) = \prod_{t=1}^n \frac{1}{\sigma_t} g\left(\frac{r_t - \mu_t}{\sigma_t}\right). \quad (3.1)$$

Donde σ_t sigue un modelo GARCH(1,1) y μ_t un modelo ARMA(1,1), Θ representa el conjunto de parámetros a ser estimados y g la función de densidad de las innovaciones, en este caso normal o t-Student. Para obtener las estimaciones se resuelven las condiciones de primer orden de la función de verosimilitud (o de la logverosimilitud). Usualmente se emplean algoritmos numéricos que consisten en modificaciones del método Newton-Raphson. En este trabajo se emplea la librería `fGarch` del paquete R que usa el método BFGS (Broyden–Fletcher–Goldfarb–Shanno) con Quasi-Newton.

3.2. Método de estimación para la distribución de Pareto generalizada (GPD)

Para estimar los parámetros de la GPD se emplea MLE. La función de logverosimilitud está dada por:

$$\log L(Z; \xi, \beta) = -N_u \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \log \left(1 + \xi \frac{Z_j}{\beta}\right). \quad (3.2)$$

Donde Z son los residuos estandarizados del modelo (2.3), N_u es la cantidad de datos que exceden el umbral escogido u , β y ξ son los parámetros de escala y de forma de la GPD. Como se argumentó anteriormente, el umbral corresponde al 10 % de la cola de la distribución de los datos.

3.3. Métodos de estimación para la distribución α -estable

Para estimar los parámetros de la distribución α -estable se emplean dos métodos. Un método basado en MLE (que llamaremos α -estable-ML), donde se estima la función de densidad a partir de la función característica mediante integración directa. Este método es computacionalmente costoso y sólo se emplea en el Ejemplo 3. Por tal razón, el otro método empleado es el propuesto por Koutrouvelis (1980), que se basa en regresión (y la llamaremos α -estable-RegK) sobre la función característica muestral.¹ Por la ley de los grandes números, la función característica muestral es un estimador consistente de la función característica (Čížek, Härdle y Weron, 2011, p. 31). De la ecuación característica de la α -estable se puede obtener:

$$\log(-\log|\phi(t)|^2) = \log(2\sigma^\alpha) + \alpha \log(t). \quad (3.3)$$

Como esta ecuación depende sólo de α y de σ se propone una regresión para estimar estos parámetros. La regresión es de la forma $y_k = m + \alpha\omega_k + \epsilon_k$, donde $y = \log(-\log|\phi_n(t)|^2)$, $m = \log(2\sigma^\alpha)$, $\omega = \log(t)$ y ϵ_k es el término de error. Koutrouvelis (1980) propone usar $t_k = \frac{\pi k}{25}$ con $k = 1, \dots, K$ con valores de K entre 9 y 134 para diferentes valores de α y tamaños de muestra. Una vez estimados α y σ se estiman μ y β . La parte real e imaginaria de la función característica están dadas por (para $\alpha \neq 0$):

$$\operatorname{Re}\phi(t) = \exp(-|\sigma t|^\alpha) \cos\left[\mu t + |\sigma t|^\alpha \beta \operatorname{sgn}(t) \tan\frac{\pi\alpha}{2}\right], \quad (3.4)$$

¹ $\hat{\phi}(t) = \frac{1}{n} \sum_{j=1}^n \exp(itx_j)$.

$$\operatorname{Im}\phi(t) = \exp(-|\sigma t|^\alpha) \sin \left[\mu t + |\sigma t|^\alpha \beta \operatorname{sgn}(t) \tan \frac{\pi\alpha}{2} \right]. \quad (3.5)$$

De estas expresiones se obtiene:

$$\arctan \frac{\operatorname{Im}\phi(t)}{\operatorname{Re}\phi(t)} = \mu t + \beta \sigma^\alpha \tan \frac{\pi\alpha}{2} \operatorname{sgn}(t) |t|^\alpha. \quad (3.6)$$

Entonces se calcula otra regresión para obtener los estimados de μ y β . Las regresiones se repiten con $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$ y $\hat{\mu}$ como parámetros iniciales. Se itera tantas veces hasta que se satisfaga un criterio de convergencia. Como estimado inicial de σ , Koutrouvelis (1980) sugiere usar la fórmula de Fama y Roll (1971), y para μ el 25 % de la media truncada. En este documento se emplea la función de Matlab, `stblfit`, seleccionando el método ‘ecf’.

3.4. Método de estimación para la distribución g-h

En las aplicaciones de la distribución g-h se emplea la metodología propuesta por Hoaglin (1985), en especial la sección 11D para estimar g . Esta metodología se basa en cuantiles y en comparación con otras metodologías, como el método de momentos y máxima verosimilitud, el método basado en cuantiles puede ser más preciso en el ajuste de las colas de la distribución, como lo afirman varios autores. Este método no es apropiado para todas las distribuciones; sin embargo, sí es adecuado para g-h, que es una transformación de una normal estándar. Para el método de los momentos se podrían plantear cuatro ecuaciones para resolver cuatro incógnitas, que son la cantidad de parámetros a estimar. Sin embargo, el cuarto momento exige que $0 \leq h < 1/4$, y por ejemplo, Dutta y Perry (2006) encuentran un rango para el parámetro h en $[0,1, 0,35]$. Esto se convierte en una desventaja para el método de los momentos. Aunque no existe una expresión analítica para obtener la función de verosimilitud de una distribución g-h, Rayner y MacGillivray (2002) estudian procedimientos de estimación de los parámetros para una distribución g-h generalizada mediante estimación numérica de la

función de máxima verosimilitud. Mientras que Haynes y Mengersen (2005) exploran computación bayesiana para estimar tales parámetros. En este trabajo se presenta una propuesta de estimar el parámetro h como la pendiente de una regresión robusta de $\log \frac{g(X_p - X_{1-p})}{\exp(gZ_p) - \exp(-gZ_p)}$ vs. $(Z_p^2/2)$ (veáse la Propiedad 6), en lugar de usar una regresión lineal clásica. Veáse el Apéndice A para una introducción de regresión robusta.

Habiendo elegido los métodos de estimación de los parámetros de los modelos, se procede a calcular el VaR a tres muestras de datos, y medir su desempeño mediante backtesting, empleando los métodos de estimación descritos en este capítulo, para las cinco distribuciones analizadas en el documento: normal, t-Student, GPD, α -estable y g-h.

Capítulo 4

Backtesting y Aplicación

Este capítulo introduce la idea del backtesting y se aplica a los datos de DJIA, analizados en los ejemplos de capítulos anteriores. Adicionalmente, en el apartado 4.2 se aplica el backtesting a dos series temporales diferentes a DJIA, correspondientes a un índice bursátil y otro a tipo de cambio. Finalmente, el apartado 4.3 propone una mejora a la eficiencia computacional en la estimación de los parámetros de la distribución α -estable mediante el método MLE. Estas estimaciones se utilizan en la cuantificación del VaR.

4.1. Backtesting

El backtesting es la técnica más comúnmente empleada en la industria financiera para validar la confiabilidad de los modelos VaR. Para el Comité de Basilea es una herramienta que permite verificar la pertinencia de modelos internos de las instituciones financieras para cálculos de requerimientos de capital.

La idea del backtesting es probar lo bien que se han comportado las estimaciones del VaR en el pasado. Esto permite comparar objetivamente distintos procedimientos para el cálculo del VaR. Se calcula el número de veces que la pérdida real haya excedido el VaR estimado. Los días en que

la pérdida haya superado el VaR estimado se denominan violaciones o excepciones. Si se calcula VaR al 99 % se espera que el número de excepciones esté alrededor de 1 % de los casos. Si el porcentaje de excepciones es mucho más alto (bajo) que el 1 %, entonces el modelo del VaR sub(sobre)-estima el riesgo.

Se elige un tamaño de ventana, que es el tamaño de muestra a ser usado para estimar el VaR. Luego se emplea la metodología de ventanas móviles para ir estimando el VaR, es decir, se actualiza la ventana con rentabilidad siguiente al último dato de la ventana inicial, y se elimina el primer dato. El proceso se realiza sucesivamente hasta finalizar el periodo del backtesting.

El backtesting se basa en el supuesto que el número de excesos son generadas por un proceso independiente e idénticamente distribuido (i.i.d.) Bernoulli (Alexander, 2008b, p. 334). En este caso, la variable aleatoria que recoge el número de excesos (X) toma el valor de 1 si la pérdida excedió el estimado del VaR, en caso contrario toma el valor de cero. Por lo general, se define la función indicadora de la siguiente manera:

$$I_{1-\alpha,t+1} = \begin{cases} 1, & \text{si } r_{t+1} < VaR_{1-\alpha,t+1}, \\ 0, & \text{en otro caso.} \end{cases} \quad (4.1)$$

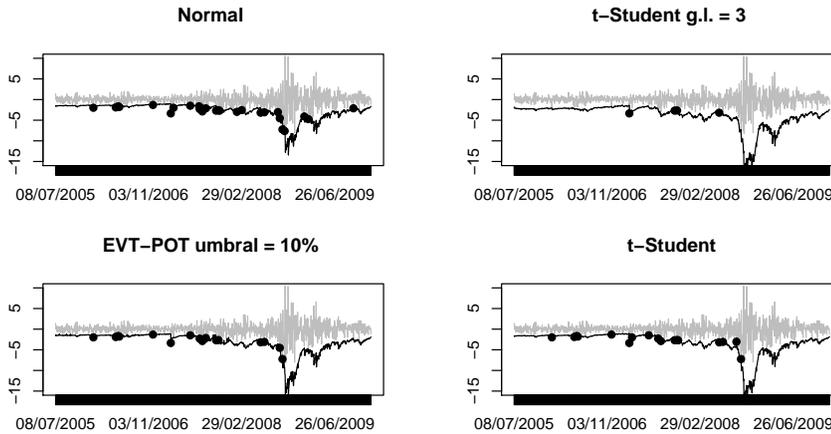
Teniendo en cuenta el valor esperado y varianza de una variable aleatoria binomial, se tiene:

$$\begin{aligned} E(X) &= T\alpha, \\ var(X) &= T\alpha(1 - \alpha). \end{aligned} \quad (4.2)$$

Donde T es el número de días de backtesting. Si T es muy grande, entonces la variable aleatoria X se distribuye como una normal de manera aproximada. Por lo tanto, se puede construir un intervalo de confianza al 95 % para el número esperado de violaciones.

$$IC_X^{95\%} = \left(T\alpha - 1,96\sqrt{T\alpha(1 - \alpha)}, T\alpha + 1,96\sqrt{T\alpha(1 - \alpha)} \right). \quad (4.3)$$

Figura 4.1: Backtesting para las rentabilidades del DJIA



Por ejemplo, si el número de días de backtesting es de 1170, y se desea calcular VaR al 99 %, el número esperado de violaciones sería 11.7, y el error estándar es de 3.40. Entonces, el intervalo de confianza al 95 % es aproximadamente (5.03, 18.37). Este intervalo será utilizado en el siguiente ejemplo.

Ejemplo 3. (Backtesting)

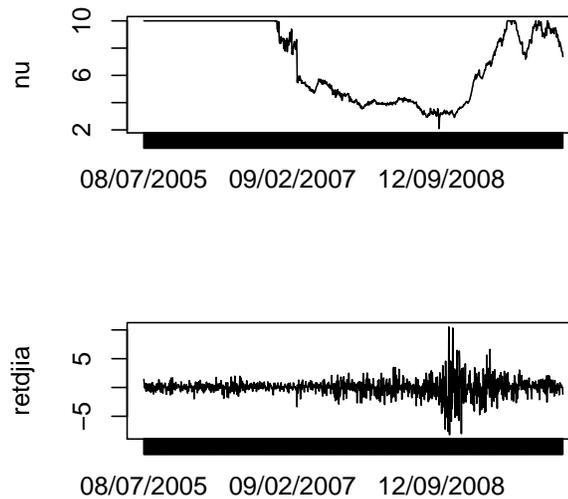
Continuando con los datos de DJIA, se escoge un tamaño de ventana igual a 500. Entonces, el periodo de backtesting es igual a 1170 días. Para filtrar las rentabilidades se utiliza un modelo ARMA(1,1)-GARCH(1,1). La Figura 4.1 presenta el backtesting para cuatro modelos (normal, t-Student, t-Student con grados de libertad fijo, y el método POT).

Para el método Picos sobre el Umbral (POT), se usa el procedimiento de dos pasos propuesto por McNeil y Frey (2000):

(1) Se ajusta el modelo ARMA-GARCH mediante el enfoque de cuasi-máxima verosimilitud, con el cual no se asume una distribución G específica para las innovaciones, y se estiman μ_{t+1} y σ_{t+1} usando el modelo ajustado.

(2) En este paso se usan los residuos estandarizados del modelo, los cuales se asumen ser i.i.d., y se usa el método POT, con un umbral del 10% de la cola de la distribución de los datos para estimar el cuantil al nivel α .

Figura 4.2: Evolución del parámetro - t-Student



La Figura 4.2 compara el valor del parámetro de la distribución t-Student con la evolución de las rentabilidades del DJIA. Recuérdese que a menos grados de libertad, más pesadez estimada en las colas. Se observa como los grados de libertad alcanzan sus mínimos valores en periodos de alta volatilidad y altas pérdidas, efecto de la crisis subprime. Mientras que la Figura 4.3 muestra la evolución de los parámetros g-h. Se observa que el parámetro que controla la curtosis (h) es positivo y alcanza sus máximos valores en periodos de alta volatilidad.

La Figura 4.4 muestra el índice de estabilidad estimado mediante MLE y el backtesting cuando las distribuciones se distribuyen α -estable. Los valores estimados de α están entre 1.72 y 2. Mientras que el método de regresión estima α entre 1.49 y 2. Este último método es el que se empleará en 4.2, al estimar VaR mediante la distribución α -estable.

El Cuadro 4.1 resume el resultado del backtesting. Donde α -estable-ML indica que los parámetros se han estimado mediante máxima verosimilitud (utilizando integración directa para obtener la función de densidad a partir de la función característica), y α -estable-RegK indica que los parámetros han sido estimados mediante el método de Koutrouvelis (1980).

Figura 4.3: Evolución de los parámetros - g-h

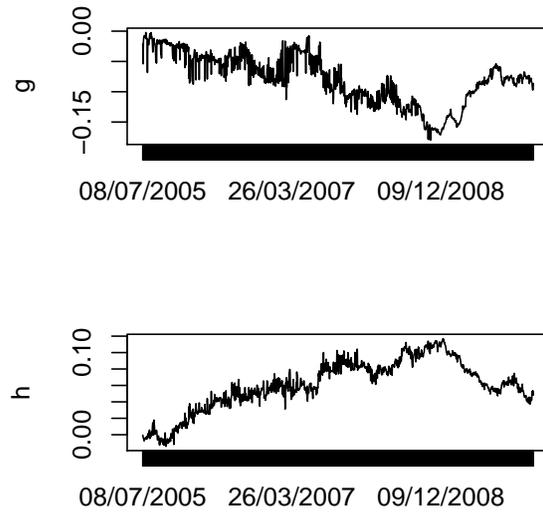
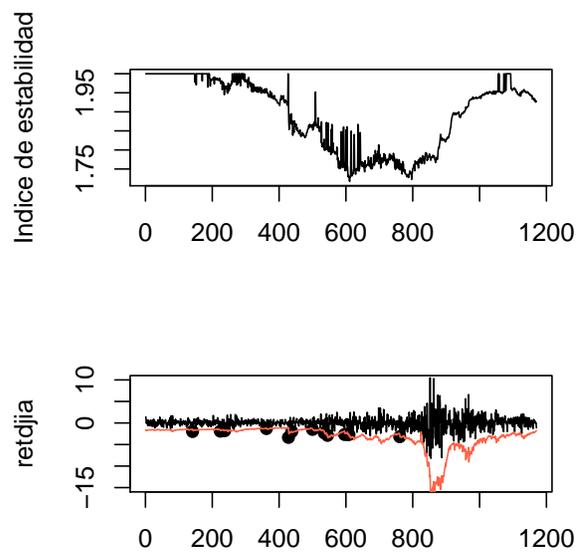


Figura 4.4: Evolución del índice de estabilidad



Cuadro 4.1: Resultados backtesting - DJIA

Intervalo de confianza para excepciones [5,18]	
Modelo	Excepciones
ARMA-GARCH-normal	27
ARMA-GARCH-t ($\nu = 3$)	4
ARMA-GARCH-t	17
ARMA-GARCH-POT (u=10%)	16
ARMA-GARCH- α -estable-ML	14
ARMA-GARCH- α -estable-RegK	10
ARMA-GARCH-g-h	8

Para corroborar el análisis anterior, usualmente se calcula el ratio de excepciones (VR, por sus siglas en inglés), que mide la proporción entre excepciones obtenidas por el modelo VaR y las excepciones esperadas. Entonces, se considera que el mejor modelo es el que cuenta con un VR igual o muy cercano a 1. Si VR es mucho mayor (menor) que 1 el modelo sub(sobre)-estima riesgo. El contraste estadístico para validar el VR es el de cobertura incondicional (UC, por sus siglas en inglés). La hipótesis nula es que el modelo del VaR es correcto, según el ratio de excepciones. Otro contraste es el de independencia. Un modelo del VaR puede ser correcto en el sentido del contraste de cobertura, pero las excepciones pueden ocurrir en días consecutivos. Si esto sucede podría ocasionar una quiebra con mayor probabilidad que si las excepciones ocurren de manera independiente. La hipótesis nula de este contraste es independencia en la ocurrencia de las excepciones. El Cuadro 4.2 muestra los valores de los estadísticos y sus respectivos p-valores, donde LR_{uc} es el estadístico del contraste de cobertura incondicional y LR_{ind} es el estadístico del contraste de independencia. Si consideramos un nivel de significación del 5%, se rechaza la hipótesis nula cuando el p-valor sea menor a 0.05. Para más detalles del cálculo de los contrastes, véase por ejemplo Christoffersen (2003, sección 8.2) y Danielsson (2011, cap. 8).

Cuadro 4.2: Contrastes de cobertura incondicional y de independencia para DJIA

Modelo	VR	LR_{uc}	p-valor	LR_{ind}	p-valor
Normal	2.3076	14.7603	0.0001	1.2768	0.2585
t ($\nu = 3$)	0.3419	6.8647	0.0088	0.0274	0.8684
t-Student	1.4529	2.1275	0.1446	0.5017	0.4787
POT (u=10%)	1.3675	1.4320	0.2314	0.4440	0.5052
α -estable-ML	1.1966	0.4296	0.5121	0.3394	0.5602
α -estable-RegK	0.8547	0.2624	0.6085	0.1726	0.6778
g-h	0.6837	1.3294	0.2489	0.1102	0.7398

Como era de esperarse, las excepciones de los modelos que no satisfacen el intervalo de confianza del Cuadro 4.1, no satisfacen el contraste de cobertura incondicional. Ningún modelo rechaza la hipótesis nula del contraste de independencia. Esto quiere decir que el modelo ARMA(1,1)-GARCH(1,1) se ajusta bien a los cambios en volatilidad.

La idea de utilizar este ejemplo es comparar los resultados obtenidos en este documento con los de Rachev et al. (2010). Estos autores obtienen los resultados que se muestran en el Cuadro 4.3, para los datos del Ejemplo 3.

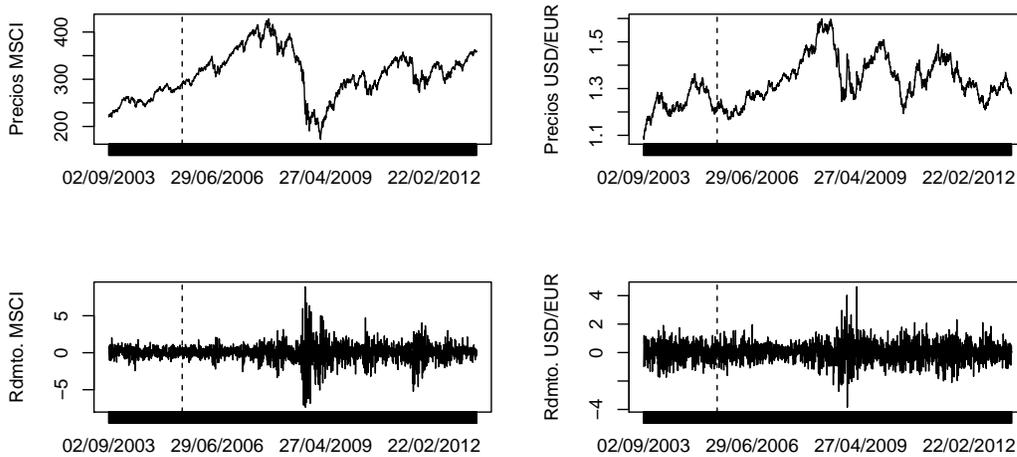
Cuadro 4.3: Resultados de Rachev et al. (2010)

Modelo	Excepciones
GARCH-normal	20
GARCH-t	3
GARCH-EVT (POT)	1
GARCH-estable	11

El artículo de Rachev et al. (2010) no especifica que modelo GARCH ha sido empleado, aunque indican que se han basado en un modelo ARMA-GARCH. Las excepciones esperadas para el caso de la distribución normal son similares; sin embargo, existen diferencias en los otros casos. Por ejemplo, en el caso de la distribución t-Student, los autores emplean un parámetro fijo de cinco grados de libertad. En nuestro análisis obtenemos un valor similar a estas excepciones cuando se fija el parámetro en tres grados de libertad. Sin embargo, cuando se permite ajustar el parámetro de la t-Student para cada ventana, obtenemos 17 excepciones, que están en el intervalo de confianza al 95 % para el número esperado de excepciones en este caso. Para el método POT, Rachev et al. (2010) emplean un umbral del 1.02 % necesitando una ventana móvil de 3000 datos y obtienen un pobre desempeño de este método. Mientras que en nuestro caso, se emplea un umbral del 10 % y el tamaño de la ventana móvil igual al de las demás distribuciones (500). Nuestros resultados muestran que el número de excepciones es 16 para POT, y también se encuentra en el intervalo de confianza del 95 %. Finalmente, el estimado de excepciones mediante la distribución estable es el que más se acerca al esperado y este resultado va en la línea del estudio de Rachev et al. (2010). Otra diferencia es que a pesar de contar con el mismo rango de datos, el periodo de backtesting para Rachev et al. (2010) es de 1130 días, mientras que en nuestro estudio es de 1170 días.

En el apartado del método POT, ya habíamos comentado la dificultad en la selección del umbral para este método, y las consecuencias en la estimación de los parámetros de la distribución de Pareto generalizada. Aunque existen métodos visuales para la selección del umbral, como el gráfico de Hill y el gráfico de excesos medios, estos métodos presentan problemas. No existe un método óptimo para la selección del umbral, y éste es un problema abierto de investigación. En el artículo de Rachev et al. (2010), la selección del umbral del 1.02 % de la cola de los datos, conlleva a una pobre estimación del VaR mediante POT. Sin embargo, en este trabajo (como en otros estudios mencionados en la revisión de la literatura) se recomienda usar como umbral el 10 % de la cola de los datos. Para los datos de DJIA, observamos una adecuada estimación del VaR de acuerdo a los resultados del backtesting con dicho umbral.

Figura 4.5: Gráfico de precios y rentabilidades de la Aplicación



4.2. Aplicación

En este apartado se aplican los métodos descritos en los capítulos anteriores y que se aplicaron en el Ejemplo 3 de la sección anterior. La idea es probar el desempeño de las distribuciones analizadas a dos series temporales, teniendo en cuenta las recientes crisis financieras, la crisis subprime y de riesgo soberano. Los datos corresponden al índice bursátil MSCI ACWI¹ (MSCI) y el tipo de cambio USD/EUR.² El periodo de precios de estas series comprende desde el 1 de septiembre de 2003 hasta 1 de abril de 2013, para un total de 2501 datos diarios de precio. Los rendimientos logarítmicos se calculan como $r_t = 100 \log(P_t/P_{t-1})$, con esto se estiman 2500 datos de rendimientos diarios. El tamaño de la ventana empleado es de 500 días, y el periodo del backtesting comprende desde el 2 de agosto de 2003 hasta el 1 de abril de 2013. Por lo tanto, los días de backtesting son 2000, y al calcular VaR al 99%, se espera un promedio de 20 excepciones al VaR calculado dependiendo de la distribución asumida.

¹Es un índice ponderado por capitalización del mercado y mide el desempeño bursátil de mercados desarrollados y emergentes.

²Fuente: Datastream, Thomson Financial.

La Figura 4.5 muestra en la parte superior la serie de precios y en la inferior la serie de rendimientos. La línea vertical punteada representa el inicio del periodo del backtesting.

El Cuadro 4.4 muestra que el promedio del rendimiento logarítmico para las dos series está alrededor de cero. El coeficiente de asimetría es positivo para los rendimientos el USD/EUR, esto indica que para el periodo analizado se tiene más probabilidad de obtener rendimientos logarítmicos negativos. El caso contrario sucede con los rendimientos del MSCI. Las series presentan valores de curtosis mayores que tres, que es el valor de curtosis de la distribución normal. Sin embargo, el exceso de curtosis es mucho mayor en MSCI que en el tipo de cambio USD/EUR. Esto muestra que hay más eventos extremos en la serie de MSCI. Adicionalmente, la desviación estándar y el rango son mayores en MSCI que en el tipo de cambio, lo que indica que los datos de rentabilidad de MSCI están más dispersos. Se pasará directamente a los resultados del backtesting donde los rendimientos se filtran por un proceso ARMA(1,1)-GARCH(1,1) como en el Ejemplo 3.

Cuadro 4.4: Estadística Descriptiva de la Aplicación

Estadístico	MSCI	USD/EUR
Media	0.0194	0.0063
Mediana	0.0825	0.0072
Desviación estándar	1.0960	0.6458
Varianza	1.2014	0.4171
Coeficiente de asimetría	-0.4593	0.1174
Exceso de curtosis	9.1004	2.9421
Mínimo	-7.3713	-3.8440
Máximo	8.9030	4.6171

Cuadro 4.5: Resultados del Backtesting - Aplicación

Intervalo de confianza para excepciones [11,29]		
Modelo	MSCI	USD/EUR
ARMA-GARCH-normal	44	25
ARMA-GARCH-t	33	17
ARMA-GARCH-POT (u=10%)	28	22
ARMA-GARCH- α -estable-RegK	29	26
ARMA-GARCH-g-h	18	10

Resultados del Backtesting

El Cuadro 4.5 presenta los resultados del backtesting. Mediante la ecuación (4.3) se calcula el intervalo de confianza al 95% para las excepciones entre 11 y 29.

Se observa que la distribución normal y t-Student ofrecen un buen resultado para los rendimientos del tipo de cambio USD/EUR. Este resultado es razonable, puesto que esta serie no presenta muchos eventos extremos. Sin embargo, estas distribuciones fallan en capturar los eventos de cola para la serie de rendimientos del MSCI. El método basado en la teoría del valor extremo y la distribución α -estable ofrecen resultados satisfactorios para ambas series. Mientras que la distribución g-h presenta un buen desempeño para los datos de MSCI pero falla en el caso del tipo de cambio.

El Cuadro 4.6 presenta los contrastes de cobertura incondicional y de independencia para MSCI, mientras que el Cuadro 4.7 para los datos del USD/EUR. Al igual que con los resultados del Cuadro 4.2, los modelos que no cumplen con el intervalo de confianza de excepciones, no satisfacen el contraste de cobertura incondicional, y ningún modelo rechaza la hipótesis nula del contraste de independencia.

Cuadro 4.6: Contrastes de cobertura incondicional y de independencia para MS-CI

Modelo	VR	LR_{uc}	p-valor	LR_{ind}	p-valor
Normal	2.2	21.6763	0.0000	1.9807	0.1593
t-Student	1.65	7.1367	0.0075	1.1078	0.2925
POT (u=10 %)	1.4	2.8748	0.0900	0.7955	0.3724
α -estable-RegK	1.45	3.5917	0.0581	0.8538	0.3555
g-h	0.9	0.2090	0.6475	0.3271	0.5673

Cuadro 4.7: Contrastes de cobertura incondicional y de independencia para USD/EUR

Modelo	VR	LR_{uc}	p-valor	LR_{ind}	p-valor
Normal	1.25	1.1698	0.2794	0.6332	0.4261
t-Student	0.85	0.4788	0.4889	0.2916	0.5892
POT (u=10 %)	1.1	0.1956	0.6582	0.4896	0.4841
α -estable-RegK	1.3	1.6611	0.1974	0.6853	0.4078
g-h	0.5	6.1875	0.0129	0.1005	0.7512

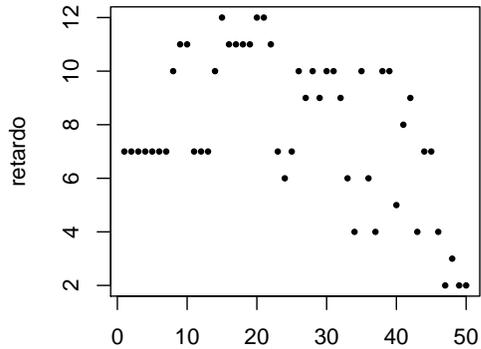
4.3. Propuesta para la mejora de eficiencia computacional en estimación de parámetros mediante MLE

Este apartado plantea una posible mejora a la eficiencia computacional mediante el método de máxima verosimilitud. La idea consiste en utilizar una serie de parámetros previamente estimados mediante este método, y luego emplear una técnica de series temporales para predecir los valores futuros de estos parámetros. Para la distribución α -estable los parámetros de interés son el índice de estabilidad, α ; y el parámetro del coeficiente de asimetría, β . Estos parámetros controlan la forma de la distribución. Específicamente, se emplea la técnica de vectores autorregresivos (VAR, por sus siglas en inglés). Generalmente se aplica VAR (véase Apéndice B) a variables estacionarias, y según Brooks (2008, p. 292-3) no se recomienda hacer diferenciación de las variables, pues al hacerlo se estaría desperdiciando información de largo plazo. Por lo tanto, se aplicará VAR a las series como están, sin diferenciar. Para elegir el orden del modelo VAR se emplea el algoritmo de Tsay (VARorder) que está basado en el criterio AIC. Adicionalmente, Brooks anota que para realizar contrastes individuales o conjuntos para examinar la significación estadística de los coeficientes, las variables deberían ser estacionarias. En general, los modelos de regresión con variables no estacionarias ofrecerán resultados espúreos, pero no cuando los residuos son estacionarios, es decir, cuando las variables están cointegradas. Sin embargo, un resultado importante es que aplicar VAR a variables no estacionarias se obtienen residuos estacionarios. A continuación se presenta esta propuesta aplicado al Ejemplo del DJIA.

Ejemplo 4. (Predicción de los parámetros de una α -estable)

Para este caso se cuentan con 1670 rendimientos diarios, y en el Ejemplo 3 se estimaron 1170 pares de parámetros α y β , puesto que para la primera

Figura 4.6: Orden del VAR



estimación de parámetros se emplearon los primeros 500 días que corresponden a los días de backtesting. El ejercicio consiste en predecir los últimos 500 valores (de los 1170) de α y β , y compararlas con las últimas 500 estimaciones del Ejemplo 3. Entonces se emplea una ventana móvil de 670 valores estimados para predecir α y β para los siguientes 10 días. Para la siguiente ventana, se emplean estos 10 valores pronosticados que se agregan al final de la ventana anterior, eliminando los primeros 10 valores de la ventana previa. Se repite este procedimiento hasta completar las 500 predicciones requeridas. Adicionalmente, se restringe los valores de α y β a los valores permisibles ($0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$). La Figura 4.6 presenta el retardo seleccionado para el modelo VAR en cada una de las 50 iteraciones.

Para la estimación, se empleó la librería `vars` de R. Aquí se considera que los estimados obtenidos en el Ejemplo 3 son los valores reales. Entonces los valores de MSE, RMSE y RRMSE para este ejemplo se muestran en el Cuadro 4.8. Donde $MSE = 1/n \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$, $RMSE = \sqrt{MSE}$ y $RRMSE = RMSE/E(Y_i)$.

La Figura 4.7 muestra el valor estimado de los parámetros mediante MLE y los valores predichos mediante el modelo VAR. También se presenta el error de pronóstico.

Cuadro 4.8: Ajuste de las predicciones

	α	β
MSE	0.0003	0.0465
RMSE	0.0180	0.2157
RRMSE	0.0097	-0.2702

Figura 4.7: Comparación de las predicciones y error de pronóstico

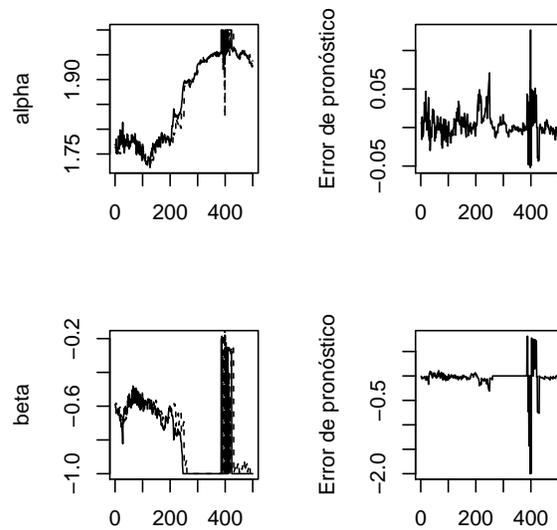
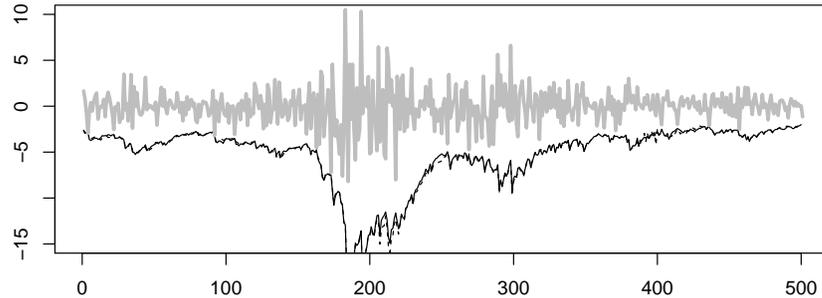


Figura 4.8: Comparación del VaR usando vectores autoregresivos



De esta manera, se utilizan los valores de los parámetros estimados mediante el VAR (modelo de vectores autoregresivos) para cuantificar Value-at-Risk como se realizó anteriormente. En lugar, de estimar los parámetros de la distribución α -estable, se emplean los valores predichos mediante el modelo de vectores autoregresivos. La Figura 4.8 presenta la comparación del VaR calculado con la distribución α -estable mediante MLE como se realizó en el Ejemplo 3 (línea sólida) con respecto al VaR calculado con los parámetros predichos mediante el modelo de vectores autoregresivos (línea punteada). Note que las rentabilidades y el VaR calculado (línea sólida) de la Figura 4.8 corresponden a los últimos 500 datos de la Figura 4.4.

Como se observa en la Figura 4.8, no se aprecia diferencias entre los dos cálculos del VaR. Por lo tanto, este apartado tiene dos implicaciones importantes en la gestión del riesgo. Una es que, gracias a la técnica de vectores autoregresivos, se puede mejorar el tiempo computacional cuando se emplea la técnica MLE al estimar los parámetros de la distribución α -estable. Puesto que estos parámetros son el insumo para calcular el cuantil de los residuos estandarizados del modelo $ARMA(1,1)GARCH(1,1)$, se obtiene un método confiable de cálculo de VaR mediante la distribución α -estable usando la expresión (2.11).

La otra implicación de este apartado es que el gestor de riesgo puede emplear la técnica de vectores autoregresivos para predecir los valores futuros del parámetro α o índice de estabilidad. Como se mencionó anteriormente,

este parámetro indica que tan pesada es la cola de la distribución, y normalmente en activos financieros, este índice está entre 1.6 y 2.0. Recuerde que entre menor sea el valor del índice de estabilidad más pesada es la cola de la distribución. Si los valores predichos del parámetro α de las rentabilidades de un activo financiero tienden a disminuir, ésto servirá de alerta temprana al gestor de riesgos para que tome la posición adecuada con respecto a dicho activo financiero y no incurra en pérdidas.

Como futura investigación, se propone emplear el modelo de vectores autoregresivos para predecir parámetros de otras distribuciones como la g-h y GPD, analizadas en este trabajo de investigación, y estudiar su relación con la cuantificación del VaR.

Capítulo 5

Conclusiones

La pregunta de investigación de si existe un modelo que permita calcular VaR de manera adecuada, y en especial, capturar eventos de cola es aún estudiada en la literatura financiera. En la revisión de la literatura se observa el uso de diferentes distribuciones que recogen asimetría y colas pesadas de las rentabilidades. En este trabajo se emplearon las distribuciones t-Student, GPD, α -estable y g-h comparada con la distribución normal.

A partir de nuestro análisis empírico, podemos concluir que, para la muestra analizada, en los datos donde se presentan eventos extremos, la distribución GPD, α -estable y g-h funcionan bien, lo cual es plenamente consistente con la postura mayoritaria de la literatura. Por lo tanto, también se recomienda el uso de esta última distribución, puesto que es una manera de incorporar asimetría y curtosis en los cálculos del Value-at-Risk (VaR), y aunque no se cuenta con una forma explícita de su función de densidad, la estimación de sus parámetros es relativamente sencilla comparada con otros métodos, como por ejemplo, para la distribución α -estable. Sin embargo, en el caso de pocos eventos de cola, la distribución g-h muestra un comportamiento muy conservador al estimar riesgo. Mientras que la distribución GPD y α -estable funcionan bien para estos casos. Al calcular VaR mediante el método POT se emplean sólo los datos que exceden un umbral, mientras que la distribución α -estable emplea todos los datos. Otro inconveniente del método POT, es la selección del umbral, debido al tradeoff entre sesgo y varianza presente en la

estimación de los parámetros de la GPD. Sin embargo, se observa que emplear un umbral correspondiente al 10 % de los datos de la cola, se obtienen resultados satisfactorios en el cálculo del VaR. La estimación de parámetros de la distribución α -estable también presenta un desafío para la investigación financiera. De la revisión de la literatura se observa que el método de máxima verosimilitud ofrece estimaciones más confiables que otros métodos; pero este método resulta ser computacionalmente costoso. Este trabajo presenta una propuesta para mitigar este problema. La propuesta consiste en predecir valores de los parámetros a partir de estimaciones previas mediante el uso del modelo de vectores autorregresivos (VAR). Los resultados muestran que los valores predichos son muy cercanos a las estimaciones mediante máxima verosimilitud. De esta manera estos valores predichos pueden ser usados para calcular VaR.

Futura investigación estará encaminada a probar el desempeño de estas distribuciones en la medida de riesgo Expected Shortfall (ES) cuando exista un método apropiado de backtesting para esta medida de riesgo, debido a que ES no cumple la propiedad de “elicibility” (Véase por ejemplo Gneiting, 2011).

Apéndice A

Regresión robusta

Un modelo de regresión lineal estimado mediante mínimos cuadrados no se puede comportar bien cuando los errores no se distribuyen normal, especialmente cuando su distribución presenta colas pesadas. Esto se debe a la presencia de observaciones influyentes. Una solución es usar el método de la regresión robusta, cuyos estimados no son tan sensibles en la presencia de datos influyentes (Fox, 2002).

El método más común en regresión robusta es la estimación- M , desarrollada por Huber en 1964. Sea el modelo de regresión lineal:

$$\begin{aligned} y_k &= \beta_1 + \beta_2 x_{k2} + \beta_3 x_{k3} + \cdots + \beta_j x_{kj} + \epsilon_k, \\ y_k &= \mathbf{x}'_k \hat{\beta} + e_k. \end{aligned} \tag{A.1}$$

Donde \mathbf{x}'_k y $\hat{\beta}$ es la notación matricial para las variables explicativas y los estimados de los parámetros respectivamente. Entonces el estimador- M es el que minimiza la función:

$$\sum_{k=1}^n \rho(e_k) = \sum_{k=1}^n \rho(y_k - \mathbf{x}'_k \hat{\beta}). \tag{A.2}$$

Donde ρ brinda la contribución de cada residuo a la función objetivo. ρ

debería tener las siguientes propiedades:

- $\rho(e) \geq 0$.
- $\rho(0) = 0$.
- $\rho(e) = \rho(-e)$.
- $\rho(e_k) \geq \rho(e_{k'})$ para $|e_k| > |e_{k'}|$.

Sea ϕ la derivada de ρ . Al diferenciar la función objetivo con respecto a los estimados de los coeficientes de la regresión $\hat{\beta}$, y al igualar las derivadas parciales a cero, se obtienen las j ecuaciones para estimar los coeficientes:

$$\sum_{k=1}^n \phi(y_k - \mathbf{x}'_k \hat{\beta}) \mathbf{x}'_k = \mathbf{0}. \quad (\text{A.3})$$

Sea la función de pesos $\omega(e) = \phi(e)/e$ y llame $\omega_k = \omega(e_k)$. Entonces, las ecuaciones a ser estimadas se pueden reescribir como:

$$\sum_{k=1}^n \omega(y_k - \mathbf{x}'_k \hat{\beta}) \mathbf{x}'_k = \mathbf{0}. \quad (\text{A.4})$$

Y este se vuelve un problema de mínimos cuadrados ponderados, al minimizar $\sum_{k=1}^n \omega_k^2 e_k^2$. Sin embargo, los pesos dependen de los residuos, éstos dependen de los coeficientes estimados, y éstos a su vez de los pesos. Por lo tanto, se requiere de una solución iterativa, técnica conocida como mínimos cuadrados iterativamente reponderados (IRLS, por sus siglas en inglés). Finalmente, se resuelve iterativamente hasta que los coeficientes convergen.

$$\hat{\beta}^{(t)} = [\mathbf{X}' \mathbf{W}^{(t-1)} \mathbf{X}]^{-1} \mathbf{X}' \mathbf{W}^{(t-1)} \mathbf{y}. \quad (\text{A.5})$$

Donde $\mathbf{W}^{(t-1)} = \left\{ \omega_k^{(t-1)} \right\}$ es la matriz (diagonal) de pesos para la iteración $t-1$. La función objetivo del estimador- M , en específico, el estimador de Huber está dada por:

$$\rho_H(e) = \begin{cases} \frac{1}{2}e^2 & \text{para } |e| \leq l, \\ l|e| - \frac{1}{2}l^2 & \text{para } |e| > l. \end{cases}$$

Mientras que la función objetivo para el estimador de mínimos cuadrados es e^2 . l es una constante de afinamiento. Entre más pequeña es esta constante hay más resistencia a los valores atípicos, pero produce menor eficiencia cuando los errores se distribuyen normal. En particular, para el estimador de Huber, $l = 1,345\sigma$, donde σ es la desviación estándar de los errores.

La estimación mediante mínimos cuadrados ordinarios se puede ver como un caso específico de la estimación mediante mínimos cuadrados ponderados, donde la función de pesos es igual a 1. Esta función en la estimación robusta (Huber) es de la forma:

$$\omega_H(e) = \begin{cases} 1 & \text{para } |e| \leq l, \\ k/|e| & \text{para } |e| > l. \end{cases}$$

Apéndice B

Modelos de vectores autorregresivos (VAR)

Este tipo de modelos es útil cuando se requiere caracterizar interacciones simultáneas entre dos o más variables. Sean y_1 y y_2 dos variables, la forma de VAR(1) está dada por:

$$\begin{aligned}y_{1t} &= \beta_{10} + \beta_{11}y_{1,t-1} + \beta_{12}y_{2,t-1} + u_{1t}, \\y_{2t} &= \beta_{20} + \beta_{21}y_{1,t-1} + \beta_{22}y_{2,t-1} + u_{2t}.\end{aligned}\tag{B.1}$$

Las aplicaciones más comunes del VAR en el campo de la economía están orientadas a pruebas de causalidad de Granger, analizar función de respuesta al impulso y análisis de la descomposición de la varianza de errores. Para más detalles de las dos últimas aplicaciones véase por ejemplo Hamilton (1994, p. 167-169). En nuestro caso, se emplea VAR para predecir dos variables. Si las dos variables están cointegradas se recomienda usar el modelo vectorial de ajuste al equilibrio de largo plazo (VECM), si no lo están se sugiere usar VAR. Se dice que dos variables están cointegradas, si son variables integradas cuya combinación lineal es estacionaria. Sin embargo, un resultado importante es que aplicar VAR a variables no estacionarias se obtienen residuos estacionarios. Esto implica que se puede omitir el paso de la cointegración y realizar predicción mediante VAR sobre variables no estacionarias. Por tal razón, se

emplea VAR en nuestro trabajo. Puesto que el objetivo de este documento no es describir el modelo VAR, para más detalles véase por ejemplo Tsay (2010, Cap. 8) y Hamilton (1994, Cap. 11).

Bibliografía

Alexander, C. 2001. *Market Models: A Guide to Financial Data Analysis*. Wiley.

Alexander, C. 2008a. *Market Risk Analysis, Volume II: Practical Financial Econometrics*. Wiley

Alexander, C. 2008b. *Market Risk Analysis, Volume IV: Value at Risk Models*. Wiley.

Angelidis, T., Benos, A., Degiannakis, S. 2004. The use of GARCH models in VaR estimation. *Statistical Methodology* 1(1–2), 105-128.

Araújo Santos, P., et al. 2013. High quantiles estimation with Quasi-PORT and DPOT: An application to value-at-risk for financial variables. *North American Journal of Economics and Finance*.
<http://dx.doi.org/10.1016/j.najef.2013.02.017>.

Beirlant, J. Goegebeur, Y., Segers, J., Teugels, J. 2004. *Statistics of Extremes: Theory and Applications*. John Wiley Sons.

Berkowitz, J., O'Brien, J. 2002. How accurate are Value-at-Risk models at commercial banks? *Journal of Finance* 57, 1093–1111.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307-327.

Carrasco, M., Florens J. P. 2002. Efficient GMM Estimation Using the Empirical Characteristic Function. Working Paper, Institut d'Economie Industrielle, Toulouse.

Casella, G., Berger, R. L. 2001. *Statistical Inference, Second Edition*. Cengage Learning.

Čížek, P., Härdle, W. K., Weron, R. 2011. *Statistical Tools for Finance and Insurance*. Springer, Berlin.

Chavez-Demoulin, V. 1999. Two problems in environmental statistics: Capture-recapture analysis and smooth extremal models, Ph.D. thesis. Department of Mathematics, Swiss Federal Institute of Technology, Lausanne.

Chavez-Demoulin, McGill, J.A. 2012. High-frequency financial data modeling using Hawkes processes. *Journal of Banking Finance* 36 (12) 3415-3426.

Christoffersen, P. 2003. *Elements of Financial Risk Management*, San Diego, CA: Academic Press.

Crouhy, M., Galai, D., Mark, R. 2005. *The Essentials of Risk Management*. McGraw-Hill.

Danielsson, J. 2011. *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab*. Wiley-Blackwell.

Davison, A.C., Smith R.L. 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society Series B* 52, 393-442.

De Jesús, R., Ortiz, E., Cabello, A. 2013. Long run peso/dollar exchange rates and extreme value behavior: Value at Risk modeling. *The North American Journal of Economics and Finance* 24, 139-152.

Degen, M., Embrechts, P., Lambrigger, D. 2007. The quantitative modeling of operational risk: between g-and-h and EVT. *ASTIN Bulletin* 37(2) 265-291.

Diebold, F. X., Schuermann, T., Stroughair, J. D. 1998. Pitfalls and Opportunities in the use of extreme value theory in risk management. Working Paper, 98-10, Wharton School, University of Pennsylvania.

Dutta, K., Babbel D. F. 2002. On Measuring Skewness and Kurtosis in Short Rate Distributions: The Case of the US Dollar London Inter Bank Offer Rates. Wharton Financial Institutions Center Working Paper.

Dutta, K., Perry, J. 2006. A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital. Federal Reserve Bank of Boston, Working Paper No. 06-13.

Embrechts, P., Klüppelberg, C., Mikosch, T. 1997. *Modelling Extremal*

Events for Insurance and Finance, Springer, Berlin.

Engle, R.F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.

Fama, E. F., Roll, R. 1971. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association* 66, 331-338.

Fox, J. 2002. *An R and S-PLUS Companion to Applied Regression*. Sage Publications.

Garcia, R., Renault, E., Veredas, D. 2011. Estimation of stable distributions with indirect inference. *Journal of Econometrics*, 161 (3) 325-337.

Ghosh, S., Resnick, S. 2010. A discussion on mean excess plots. *Stochastic Processes and their Applications* 120, 1492-1517.

Gilli, M., Kellezi, E. 2006. An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics* 27, 207-228.

Gneiting, T. 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106, 746-762.

Gourieroux, C., Monfort, A., Renault, E. 1993. Indirect inference *Journal of Applied Econometrics*. 8 (S1) S85-S118.

Granger, C. W. J., Ding, Z. 1995. Some properties of absolute returns. An alternative measure of risk. *Annales d'économie et de statistique*, 40, 67-91.

Grigoletto, M., Lisi, F. 2009. Looking for skewness in financial time series. *The Econometrics Journal* 12, 310-323.

Grigoletto, M., Lisi, F. 2011. Practical implications of higher moments in risk management. *Statistical Methods and Applications* 20, 487-506.

Güner, B., Rachev, S., Edelman, D., Fabozzi, F. J. 2010. Bayesian Inference for Hedge Funds with Stable Distribution of Returns. *En Rethinking Risk Measurement and Reporting: Volume II* (Ed. K. Böcker), 95-136.

Haas, M. 2009. Value-at-Risk via mixture distributions reconsidered. *Applied Mathematics and Computation*, 215 (6) 2103-2119.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton University Press: New Jersey.

Haynes, M. y Mengersen, K. 2005. Bayesian estimation of g-and-k distri-

butions using MCMC. *Comput. Stat.* 20(1) 7–30.

Hoaglin, D. C. 1985. Summarizing Shape Numerically: The g-and-h Distributions. Chapter 11 en *Exploring Data Tables Trends, and Shapes*. Eds. Hoaglin, Mosteller, and Tukey. New York, NY: John Wiley.

Hubbert, S. 2012. *Essential Mathematics for Market Risk Management*, Second Edition, The Wiley Finance Series.

Hull, J. 2006. *Risk Management and Financial Institutions*. Prentice Hall.

Hull, J. 2012. *Risk Management for Financial Institutions*, Third Edition, Wiley Finance.

Jiménez, J. A. y Martínez J. 2006. Una estimación del parámetro de la distribución g de Tukey. *Revista Colombiana de Estadística* 29(1), 1 – 16.

Jondeau, E., Poon, S.-H., Rockinger M. 2007. *Financial Modeling Under Non-Gaussian Distributions*, Springer Verlag.

Jorion, P. 2006. *Value at Risk: The New Benchmark for Managing Financial Risk*, Third Edition, McGraw-Hill.

Khindanova, I., Rachev, S., Schwartz, E. 2001. Stable modeling of value at risk. *Mathematical and Computer Modelling* 34(9-11), 1223-1259.

Kiesel, R., Perraudin, W., Taylor, A. 2003. An Extreme Analysis of VaRs for Emerging Market Benchmark Bonds. En: *Credit Risk: Measurement, Evaluation and Management*, eds: G.Bol, et al., Physica-Verlag.

Kuester, K., Mittnik, S., Paolella, M. S. 2006. Value-at-risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics* 4, 53–89.

Lauridsen, S. 2000. Estimation of Value at Risk by Extreme Value Methods. *Extremes* 3 (2), 107-144.

Leadbetter, M. R. 1991. On a basis for ‘peaks over thresholds’ modeling. *Statistics Probability Letters* 4, 357-362.

Li, W. K., Ling, S., McAleer, M. 2002. Recent Theoretical Results for Time Series Models with GARCH Errors. *Journal of Economic Surveys* 16 (3) 245–269.

Lunde, A, Hansen, P. R. 2005. A forecast comparison of volatility models: Does anything beat a garch(1,1)?. *Journal of Applied Econometrics* 20 (7)

873–889.

Mandelbrot, B. 1963. The variation of certain speculative prices. *Journal of Business* 36, 394-419.

Marinelli, C., D'Addona, S., Rachev, S. 2007. A comparison of some univariate models for Value-at-Risk and expected shortfall. *International Journal of Theoretical and Applied Finance*, 10 (6).

Martínez, J. 1981. Some Applications of Robust Scale Estimators, Doctoral Thesis, Temple University, Department of Statistics. Philadelphia.

Martínez, J. & Iglewicz, B. 198). Some properties of the Tukey g and h family of distributions, *Communications in Statistics: Theory and Methods* 13 (3) 353-369.

McCulloch, J. H. 1986. Simple consistent estimators of stable distribution parameters, *Communications in Statistics - Simulations* 15, 1109-1136.

McNeil, A.J., Frey, R. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7, 271–300.

McNeil, A. J., Frey, R., Embrechts, P. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.

Mittnik, S., Rachev, S. T., Doganoglu, T., Chenyao, D. 1999. Maximum likelihood estimation of stable Paretian models, *Mathematical and Computer Modelling* 29: 275-293.

Nolan, J. P. 2001. Maximum likelihood estimation and diagnostics for stable distributions, in O. E. Barndorff-Nielsen, T. Mikosch, S. Resnick (eds.), *Lévy Processes*, Birkhäuser, Boston.

Nolan, J. P. 2013. *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser. Boston. In progress.

Ourir , A., Snoussi, W. (2012). Markets liquidity risk under extremal dependence: Analysis with VaRs methods *Economic Modelling* 29 (5) 1830-1836.

Perignon , C., Deng, Z. Y., Wang, Z. J. 2008. Do banks overstate their Value-at-Risk? *Journal of Banking Finance* 32, 783-794.

Rachev, S., Racheva-Iotova, B., Stoyanov, S. 2010. Capturing fat tails. *Risk Magazine*, May 2010, 72-77.

Rayner, G. y MacGillivray, H. 2002. Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics and Computing* 12, 57-75.

Reiss, R.-D., Thomas, M. 2007. *Statistical Analysis of Extreme Values*. Third Edition. Basel: Birkhäuser.

Rydén, T., Teräsvirta, T., Asbrink, S. 1998. Stylized facts of daily return series and the hidden markov model. *Journal of Applied Econometrics* 13, 217-244.

Smith, R. L. 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science* 4, 367-393.

Stavroyiannis, S. Makris, I. Nikolaidis, V. Zarangas, L. 2012. Econometric modeling and value-at-risk using the Pearson type-IV distribution. *International Review of Financial Analysis* 22, 10-17.

Teräsvirta, T., Zhao, Z. 2011. Stylized facts of return series, robust estimates and three popular models of volatility. *Applied Financial Economics* 21, 67-94.

Tsay, R. 2010. *Analysis of Financial Time Series*, Third Edition. Wiley.

Wei, Y., Cheng, W., Lin Y. 2013. Measuring daily Value-at-Risk of SSEC index: A new approach based on multifractal analysis and extreme value theory. *Physica A: Statistical Mechanics and its Applications* 392 (9), 2163-2174.

Weiss, A. 1984. ARMA Models with ARCH Errors. *Journal of Time Series Analysis*, 5, 129-143.

Wong, W. K. 2010. Backtesting value-at-risk based on tail losses. *Journal of Empirical Finance* 17, 526-538.

Zhang, M.-H., Cheng, Q.-S. 2005. An Approach to VaR for capital markets with Gaussian mixture. *Applied Mathematics and Computation* 168, 1079-1085.