

# APLICACIONES DE LA TEORÍA DE VALORES EXTREMOS A LA GESTIÓN DEL RIESGO

**Tesis Doctoral**

Autor: José Olmo

Director: Jesús Gonzalo



Departamento de Economía

UNIVERSIDAD CARLOS III DE MADRID

Getafe, Mayo 2005



Esta tesis está dedicada a mi padre, José.



# Índice general

<b>1. Introducción</b>	<b>15</b>
<b>2. Una Definición formal de los Valores Extremos</b>	<b>29</b>
2.1. Introducción . . . . .	30
2.2. Resultados de la Teoría de Valores Extremos . . . . .	32
2.3. Selección del umbral para definir los Valores Extremos . . . . .	36
2.4. Contraste de Hipótesis . . . . .	43
2.4.1. Aproximación Bootstrap . . . . .	46
2.4.2. Comportamiento en muestras finitas: Potencia empírica . . . . .	48
2.5. Aplicación: Estimación del VaR en índices Financieros . . . . .	51
2.5.1. Características de los datos . . . . .	53
2.6. Conclusiones . . . . .	54
2.7. Apéndice A: Demostraciones . . . . .	56
2.8. Apéndice B: Tablas . . . . .	58
<b>3. Dependencia en los Extremos</b>	<b>61</b>
3.1. Introducción . . . . .	63

3.2.	Definición del extremal index . . . . .	67
3.3.	Estimación del extremal index . . . . .	71
3.3.1.	Propiedades estadísticas de los diferentes estimadores . . . . .	73
3.3.2.	Inferencia para el Extremal Index . . . . .	75
3.3.3.	Algunos comentarios sobre la selección del <i>tamaño</i> del bloque . . . . .	78
3.4.	Simulaciones: Algunos ejemplos . . . . .	80
3.5.	Clustering en Series Financieras: El Caso del índice DaX . . . . .	82
3.6.	Conclusión . . . . .	84
3.7.	Apéndice: Figuras . . . . .	86
<b>4.</b>	<b>Transmisión del Riesgo en Mercados Financieros</b>	<b>99</b>
4.1.	Introducción . . . . .	101
4.2.	El modelo . . . . .	106
4.2.1.	La estructura de dependencia: La función cópula . . . . .	107
4.3.	Contagio: tipos y definiciones . . . . .	114
4.3.1.	Estimación de la Cópula: Contraste del Efecto Contagio . . . . .	119
4.4.	Aplicación: Vuelo hacia la calidad frente a Contagio . . . . .	120
4.5.	Conclusiones . . . . .	124
4.6.	Apéndice: Tablas y Figuras . . . . .	126
<b>5.</b>	<b>Futuras Lineas de Investigación</b>	<b>137</b>

*”La fantasía, abandonada de la razón,  
solo produce monstruos imposibles.  
Unida a ella, en cambio, es la madre  
del arte y fuente de sus deseos.”*

Francisco de Goya

*”Sólo la fantasía permanece siempre joven;  
lo que no ha ocurrido jamás no envejece nunca.*

J. C. Friedrich von Schiller





## AGRADECIMIENTOS

Quiero agradecer en primer lugar la labor de mi director de tesis, Jesús Gonzalo. Me ha enseñado a ser riguroso con mi trabajo y su apoyo y sugerencias han sido ingredientes imprescindibles en esta tesis.

También quiero agradecer su nombre, y a la Universidad Carlos III de Madrid sus apellidos. Por otra parte esta tesis no hubiera sido posible sin el soporte de los Departamentos de Economía y en especial de Estadística y Econometría. Sin sus medios y sus profesores no me hubiera sido posible realizar este trabajo. Me gustaría destacar el esfuerzo constante del programa de doctorado por mantener un altísimo nivel de calidad y que me ha dado la posibilidad de conocer a profesores de la talla de Robert Engle, Clive Granger, Winfried Stute, David Nualart ó Murad Taqqu. Los comentarios de algunos de ellos han sido de gran ayuda.

Quiero agradecer a la Fundación Universidad Carlos III sus becas para estancias breves en el extranjero que me dieron la posibilidad de visitar las Universidades de Edimburgo (verano de 2000) y de Oxford (verano de 2001).

Esta investigación ha sido financiada por el Ministerio de Ciencia y Tecnología y después por el Ministerio de Educación y Ciencia, proyecto número SEC2001-0890, del que he sido becario de formación del personal investigador (FPI) durante los tres últimos *años*. Mi gratitud para estas becas que me han permitido la realización de dos estancias breves en el Departamento de Estadística e Investigación Operativa de la Universidad de Carolina del Norte en Chapel Hill.

No querría olvidarme de este departamento. Su hospitalidad y ambiente de trabajo han sido fundamentales para el resultado final de mi investigación. De entre sus profesores me gustaría destacar a Richard Smith y a Ross Leadbetter, ejemplos de dedicación a la carrera universitaria. Del Departamento de Economía de dicha universidad me gustaría mencionar a Eric Renault.

De todos estos sitios lo que más he ganado han sido amigos. De la Universidad Carlos III

me gustaría destacar primero a Ricardo Laborda y a Pablo Villaplana, verdaderos amigos. A Sara López, Oscar Martínez, Ramiro Losada, Juan Carlos Escanciano, Andrés Ubierna, Mónica Benito, Carmen Broto, María Angeles Pulido y otras/os que también han estado en algún momento.

De Carolina del Norte a Vassilis Maroulas, Arka Gosh, Brian Lopes, Gustavo Didier y muy en especial a Francisco Chamú y su esposa Jennifer.

De Oxford a Asghar Zaidi y a Rosa Fernández y de Edimburgo a Andy Snell.

Me gustaría dedicar las últimas líneas de agradecimiento a mi familia. A mi hermano, a mi madre, porque sin ella no habría lagrangianos, ni martingalas, ni segundos momentos. A Ana, porque es lo más dulce que he conocido y porque se lo merece. Y finalmente a mi padre, porque aunque no ha podido verlo, él siempre supo que lo conseguiría. Y si él estaba orgulloso de mi, yo lo estoy mucho más de él.

## RESUMEN

La intención de esta tesis es conocer más sobre la gestión del riesgo por medio de una metodología muy diferente de las técnicas estadísticas normalmente utilizadas: varianza y correlación. La alternativa utilizada es la teoría de valores extremos, que se presenta como el medio natural para cuantificar el riesgo en econometría financiera.

La tesis se concentra en el riesgo. Hay diferentes interpretaciones de este concepto que dan lugar a diversas metodologías para cuantificar su magnitud e impacto en diferentes características de la econometría financiera. En la introducción de la tesis se discute la distinción entre incertidumbre y riesgo desde diferentes puntos de vista, teoría de la decisión y gestión del riesgo. Se sigue con una definición formal del riesgo motivada por teoría de la decisión pero consistente con la metodología usada en la gestión del riesgo. El riesgo se puede cuantificar por medio de técnicas estadísticas. Se caracteriza por las colas de la distribución de los datos, en particular por la verosimilitud de cualquier suceso que conlleve una característica negativa. En econometría financiera esta definición de riesgo se denomina normalmente “downside risk” y se asocia con la cola izquierda de la distribución de los rendimientos.

El objetivo del segundo capítulo es dar medidas adecuadas para cuantificar el riesgo en series financieras. Para conseguir esto, se aplican herramientas derivadas de la teoría de valores extremos.

Todas estas medidas del riesgo recientemente consideradas en la literatura basadas en valores extremos se caracterizan en la práctica por métodos de selección ad-hoc de los valores extremos (5 %, 1 %, etc.) La principal contribución en el segundo capítulo es proponer una definición formal para estos valores. Los valores extremos de una muestra aleatoria simple de tamaño  $n$  de una distribución  $F$  se definen como las observaciones que exceden cierto umbral y siguen una distribución Generalizada de Pareto ( $GPD$ ) donde el “tail index” de  $F$  juega un papel principal. El umbral es el estadístico de orden que minimiza un estadístico del tipo de Kolmogorov-Smirnov entre la distribución empírica de las corre-

spondientes observaciones mayores y la correspondiente *GPD*. Para formalizar la definición usamos un bootstrap semiparamétrico para contrastar la correspondiente aproximación por la distribución Generalizada de Pareto. Finalmente, usamos nuestra metodología para cuantificar el riesgo estimando el tail index (es decir, el ratio de decaimiento de la cola negativa), y el Valor en Riesgo (VaR) de algunos índices financieros de los principales mercados de acciones.

Una vez que el riesgo se define y es formalmente cuantificado el siguiente objetivo de la tesis es analizar los mecanismos de transmisión del riesgo en diferentes marcos. El capítulo III se dedica a la transmisión del riesgo en series temporales. El riesgo se mide por la ocurrencia de observaciones de gran magnitud y el canal de transmisión es la dependencia temporal que se encuentra en los valores extremos y que pueden originar el agrupamiento de estas observaciones. En este contexto existe un parámetro, el “extremal index” que gobierna la dependencia temporal en las observaciones más altas, y tal que su recíproco mide el nivel de agrupamiento (clustering) en los extremos. La contribución de la tesis en este capítulo comienza por redefinir este parámetro. La definición provee un sencillo e inmediato método de estimación para el extremal index con interesantes propiedades estadísticas como son la consistencia y la distribución asintótica gaussiana. La existencia de clustering en las observaciones más grandes es una consecuencia de la transmisión del riesgo derivado de la ocurrencia de sucesos extremos. Una contribución muy importante en esta parte es la posibilidad de contrastar la transmisión del riesgo en series financieras mediante el contraste del clustering en los valores extremos. Esta teoría contrasta con teorías fundadas en modelos para la volatilidad que modelizan la dependencia condicional en los segundos momentos.

El siguiente capítulo trata sobre la transmisión del riesgo entre mercados financieros. El interés en esta sección radica en distinguir interdependencia entre mercados, que surge de los lazos normales entre diferentes economías, de los efectos de contagio, originados por unas conexiones que se hacen más fuertes en periodos de crisis. Para hacer esto, las nociones de interdependencia y contagio se revisan. La contribución en este punto se basa en nuevas definiciones para estos conceptos basados en propiedades de las funciones cópula y

en monotonicidad en las colas, que se usarán para analizar el contagio direccional (causalidad entre extremos). Esto es posible gracias a una innovadora función cópula que se deriva de la teoría de valores extremos multivariante. Esta cópula nos permite modelizar diferentes patrones de dependencia entre las variables de acuerdo al estado de los mercados, por ejemplo en mercados a la baja o en mercados al alza. Este modelo es suficientemente flexible para describir asimetrías entre las variables de tal manera que el contagio direccional se puede contrastar. El modelo se aplica para medir el fenómeno de vuelo hacia la calidad (flight to quality), es decir, flujos de capital que salen de los mercados de acciones hacia los mercados de bonos cuando los primeros afrontan periodos de crisis.

Finalmente el capítulo V esboza las líneas de investigación futuras que implican diferentes aspectos del análisis del riesgo.



# Capítulo 1

## Introducción

### Resumen

Este capítulo presenta los problemas y retos que motivaron esta investigación. La sección comienza con las nociones de incertidumbre y riesgo que se derivan de la teoría de la decisión y desarrolla el correspondiente tratamiento estadístico. En la actual literatura la varianza se utiliza para cuantificar ambos conceptos. En esta parte en cambio, las diferencias entre incertidumbre y riesgo se discuten y las metodologías estadísticas apropiadas para cada problema se introducen. La teoría de valores extremos se motiva como el medio natural para cuantificar el riesgo. Los defectos de esta técnica y las alternativas introducidas en la tesis se esbozan en esta parte. La sección concluye con el marco multivariante y el uso de funciones cópula para modelizar la dependencia motivadas por los problemas de medidas lineales como la correlación.

En la última década la gestión del riesgo se ha convertido en una disciplina importante en finanzas. Se estudia en diferentes campos dentro de finanzas: econometría financiera, matemáticas financieras ó en ingeniería financiera. El principal objetivo de la gestión del riesgo es analizar las causas y consecuencias de sucesos negativos para los intereses del inversor. Esta basta definición depende de la definición de suceso negativo, y en el perfil del inversor. Además, uno de los campos de investigación más prometedores en finanzas en los últimos *años* es el desarrollo de instrumentos financieros y estrategias de inversión que permiten a uno cubrirse contra estos sucesos negativos. Algunos ejemplos de lo último son el uso de opciones financieras, la posibilidad de posicionarnos en corto en una cartera de activos o invertir en derivados con diferentes fechas de vencimiento.

El concepto de riesgo, por lo tanto, no está claramente definido. Al menos no hay una deficinición única y universal que permita a académicos e inversores privados progresar en la misma dirección para resolver los mismos problemas. Ambos grupos sin embargo se han concentrado en estudiar la incertidumbre más que el riesgo guiados por la creencia común de que ambos términos son intercambiables. Knight (1921) definió incertidumbre de ocurrencia de un suceso por la imposibilidad de asignar una probabilidad al suceso. De esta forma el autor definió situaciones de incertidumbre por la ausencia de mercados de seguros. En contraste, riesgo se presenta cuando podemos asignar una medida de probabilidad al suceso, y por lo tanto los mercados de seguros se desarrollan (existe intercambio).

Otra distinción importantte entre estos conceptos viene dada por el significado negativo del suceso que conlleva riesgo. La incertidumbre no conlleva necesariamente un resultado negativo. Por el contrario, el riesgo implica una probabilidad estrictamente positiva de un resultado negativo en el universo de posibles resultados. Ambos conceptos sin embargo se caracterizan por la presencia de aleatoriedad y por lo tanto pueden ser atribuidos a una variable aleatoria. De esta forma, las anteriores definiciones de riesgo e incertidumbre se pueden traducir a la teoría de la probabilidad.

En este contexto la incertidumbre es intrínseca a la definición de variable aleatoria y es normalmente descrita por la varianza. El riesgo sin embargo conlleva algo más no capturado por la varianza. Consideremos por ejemplo la predicción del peso de una persona adulta a



los 10 *años* de edad. No hay duda de que la incertidumbre estadística se mide mediante la varianza, aunque en absoluto es informativa sobre el riesgo. Riesgo en esta situación viene de valores muy bajos o excepcionalmente altos. En este ejemplo las observaciones extremas pueden derivar en diferentes enfermedades: anorexia, obesidad.

Este ejemplo motiva la definición de riesgo como algo que ocurre en las colas de la distribución de la variable aleatoria y que conlleva el conocimiento de su función de probabilidad. Es interesante sin embargo la visión del riesgo para estadísticos y econométricos. El riesgo se reduce a medir la varianza de la variable aleatoria que describe el comportamiento del suceso aleatorio. Esto es verdad sólo si la función de probabilidad es conocida y la única incógnita es la varianza. Consideremos el ejemplo de una distribución normal  $N(\mu, \sigma^2)$  donde  $\mu$  es desconocida pero  $\sigma$  es conocida. El conocimiento de  $\sigma$  no es suficiente para conocer ni la incertidumbre ni el riesgo. La función de probabilidad no es conocida, sino un conjunto de funciones de probabilidad. Este ejemplo refleja la ambigüedad de conocer sólo la varianza. En el ejemplo particular de analizar series financieras se suele asumir que el valor esperado de los rendimientos es 0, y entonces tiene sentido pensar en la volatilidad (varianza) como medida del riesgo. No obstante este ejemplo derivado de la econometría financiera necesita de otros supuestos. El precio de los instrumentos financieros (bonos, acciones) se asume que sigue una distribución log-normal, y en consecuencia la distribución de los rendimientos es normal. Otra situación donde la varianza es suficiente para describir el riesgo de los rendimientos financieros es cuando la función de preferencias del inversor (función de utilidad) es cuadrática.

Hay toda una serie de técnicas econométricas para estimar el riesgo bajo estos supuestos. El objetivo es la estimación y modelización del proceso de la volatilidad. La metodología estándar es la estimación de la distribución histórica donde la volatilidad se considera constante, y todas las observaciones tienen el mismo peso en la estimación de la varianza. En cambio, si la serie tiene cierta dinámica técnicas de suavizado son estimadores más adecuados para la volatilidad, en estas técnicas las observaciones más recientes  $\{x_t\}$  tienen más

protagonismo que las observaciones anteriores.

$$\sigma_t^2 = (1 - \lambda)x_{t-1}^2 + \lambda\sigma_{t-1}^2, \quad (1.1)$$

con  $0 < \lambda < 1$ .

La misma filosofía siguen los modelos GARCH introducidos en Engle and Bollerslev (1986). El modelo general GARCH(p,q) toma esta forma

$$\sigma_t^2 = \omega + \alpha_1 x_{t-1}^2 + \dots + \alpha_p x_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2, \quad t = 1, \dots, T \quad (1.2)$$

con  $\omega, \alpha_i, \beta_j > 0$  y  $\sum_i \alpha_i + \sum_j \beta_j < 1$  para obtener estacionariedad.

Hay distintas modificaciones de este modelo que reflejan los diferentes hechos estilizados que se observan en las series financieras. Ejemplos de estos modelos son los modelos EGARCH que tienen en cuenta el efecto de apalancamiento (leverage effect), IGARCH donde  $\sum_i \alpha_i + \sum_j \beta_j = 1$  que describen varianza infinita (riesgo infinito?), etc.

Otras formas más sofisticadas de medir el riesgo en este marco son la volatilidad implícita y la volatilidad realizada, que se han desarrollado en la última década. La volatilidad implícita se deriva de la valoración de opciones por el mercado y por lo tanto de la fórmula de Black-Scholes, Black and Scholes (1973). Se asume que los precios siguen un movimiento browniano geométrico. Este supuesto es suficiente, por ejemplo, para encontrar el precio de no arbitraje para la opción europea (European plain-vanilla option),

$$C_t(S_t) = S_t \Phi(d_1) - K e^{-r(T-t)} K \Phi(d_2), \quad (1.3)$$

con  $C_t$  el precio de la opción,  $\Phi(\cdot)$  la función de distribución normal estándar,  $r$  el tipo de interés libre de riesgo,  $K$  el strike,  $T$  la fecha de vencimiento, y  $d_1, d_2$  constantes que satisfacen

$$d_1 = \frac{\log \frac{S_t}{K} + (r + \frac{\sigma^2}{2})(T-t)}{\sigma \sqrt{T-t}}, \quad d_2 = d_1 - \sigma \sqrt{T-t}.$$

El conocimiento de los precios de las opciones para diferentes fechas de vencimiento es

observable del mercado. Por lo tanto, de la ecuación (1.3) se puede obtener la varianza. Esta varianza, que se denomina volatilidad implícita, se deriva por lo tanto de las expectativas de los agentes sobre el futuro.

Otra medida de volatilidad que se basa en ecuaciones diferenciales estocásticas es la volatilidad realizada. Este concepto aparece por primera vez en los artículos de Andersen et al (2001), y Barndorff-Nielsen y Shephard (2002). La expresión para la volatilidad viene de la teoría de procesos de valoración sin arbitraje y en tiempo continuo, y de la teoría de variación cuadrática. Denotemos  $[x]$  para la variación cuadrática de  $x$  definida como

$$[x]_t = \underset{q \rightarrow \infty}{plim} \sum_{i=0}^{m-1} \{x_{t_{i+1}^q} - x_{t_i^q}\}^2, \quad (1.4)$$

para cualquier secuencia de particiones  $t_0^q = 0 < t_1^q < \dots < t_m^q = t$  con  $\sup_i (t_{i+1}^q - t_i^q) \rightarrow 0$  para  $q \rightarrow \infty$ . La notación *plim* indica el límite en probabilidad de la suma. Si el logaritmo de los precios sigue una ecuación diferencial estocástica de este tipo

$$x_t = \mu_t dt + \sigma_t dw_t, \quad (1.5)$$

con  $\sigma_t^2$  la volatilidad instantánea, la variación cuadrática toma la misma expresión que la volatilidad integrada definida como

$$\sigma_t^{2*} = \int_0^t \sigma_u^2 du. \quad (1.6)$$

Este método explota la disponibilidad de datos de alta frecuencia intra-diarios. La volatilidad realizada, denotada  $\{x\}_n$  es el estimador natural de la variación cuadrática, y se define como la suma del cuadrado de los rendimientos de  $M$  observaciones intra-diarias durante cada día. Toma esta expresión

$$\{x\}_n = \sum_{j=1}^M \left\{ x_{(n-1)+\frac{j}{M}} - x_{(n-1)+\frac{j-1}{M}} \right\}^2, \quad (1.7)$$

que es consistente cuando  $M \rightarrow \infty$ . La teoría de variación cuadrática revela que bajo

ciertas condiciones la volatilidad realizada no es sólo un estimador ex-post insesgado de la volatilidad de los rendimientos diarios, donde el índice del día es  $n$ , sino también es asintóticamente libre de error de medida. Empíricamente, si se trata la volatilidad como observada en vez de latente se facilita la modelización mediante métodos simples y directos de variables observables. Por otro lado la volatilidad considerada como una variable latente, no observada, puede ser modelizada por modelos de volatilidad estocástica. Estos modelos son de esta forma

$$\sigma_t = \rho \sigma_{t-1} + \epsilon_t, \quad (1.8)$$

con  $0 < \rho < 1$  el parámetro autorregresivo y  $\epsilon_t$  la innovación.

Todas las metodologías mencionadas para cuantificar el riesgo fracasan si la distribución de los rendimientos está lejos de la gaussianidad. Este hecho está ganando popularidad entre académicos e inversores privados que han planteado la necesidad de una modelización más realista de la distribución de los rendimientos y del análisis del riesgo. El objetivo entonces cambia, en vez de calcular una medida de dispersión para los datos se mide la probabilidad en las colas de la distribución. El riesgo subyacente en las secuencias financieras se denomina *downside risk* ya que se asocia a resultados negativos de la variable aleatoria que describe la serie. Estas observaciones se representan en la cola izquierda de la distribución de los rendimientos. Merece la pena resaltar la existencia de “*upside risk*” debido a la existencia de instrumentos de cobertura que se utilizan para compensar valores en la cola izquierda de la distribución (posibles pérdidas), por lo tanto valores positivos puedan resultar en pérdidas en la cartera del inversor.

El interés de los gestores del riesgo se encuentra en estimar la distribución de los datos, en particular la distribución en las colas. Los resultados en Kolmogorov (1933) y en Gnedenko (1943) derivados de la distribución del máximo muestral, son la base para una nueva y excitante area en Estadística que envuelve el análisis de los valores extremos de secuencias aleatorias y del análisis de las colas de la distribución de los datos. Este área se denomina Teoría de Valores Extremos (EVT en sus siglas en inglés) y es la base teórica para la mayoría de las técnicas desarrolladas en la tesis.

Supongamos una muestra aleatoria simple de una distribución  $F$  desconocida, y sea  $G$  la distribución límite del máximo muestral  $M_n$ . La teoría de valores extremos clásica muestra que bajo ciertas condiciones de regularidad en la cola de  $F$  y para ciertas condiciones de normalización  $a_n, b_n$ ,

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \rightarrow G(x), \quad (1.9)$$

donde  $G$  debe ser de uno de los siguientes tipos (ver de Haan (1976)),

Tipo I: (Gumbel)  $G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty.$

Tipo II: (Fréchet)  $G(x) = \begin{cases} 0 & x \leq 0, \\ e^{-x^{-\frac{1}{\xi}}} & x > 0, \xi > 0. \end{cases}$

Tipo III: (Weibull)  $G(x) = \begin{cases} 1 & x \geq 0, \\ e^{-(-x)^{-\frac{1}{\xi}}} & x < 0, \xi < 0. \end{cases}$

El parámetro  $\xi$  es el tail index de  $F$  y caracteriza el comportamiento de la cola de la función de distribución. Los tres tipos se pueden agrupar en lo que se llama la distribución generalizada de valores extremos, propuesta por primera vez por von Mises (1936),

$$G(x) = e^{-(1+\xi\frac{x-\nu}{\sigma})^{-\frac{1}{\xi}}}, \quad (1.10)$$

donde  $\nu$  es un parámetro de localización,  $\sigma$  un parámetro de escala y  $\xi \neq 0$ . Esta expresión se reduce a  $G(x) = e^{-e^{-\frac{x-\nu}{\sigma}}}$  cuando  $\xi = 0$ .

En consecuencia la distribución del máximo muestral estandarizado  $F^n(a_n x + b_n)$  converge a  $e^{-(1+\xi\frac{x-\nu}{\sigma})^{-\frac{1}{\xi}}}$  para todo  $x$ . Por medio de cálculo simple encontramos que

$$\lim_{n \rightarrow \infty} n(1 - F(a_n x + b_n)) = \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}. \quad (1.11)$$

Este resultado se explota para derivar la convergencia débil de las observaciones más altas determinadas por una secuencia umbral  $u_n = a_n \nu + b_n$ , con  $\nu$  que satisface  $-\log G(\nu) = 1$ .

Este es el principal resultado del teorema de Pickands, Pickands (1975),

$$\lim_{u_n \rightarrow x_F} \sup_{0 \leq y < \infty} |F_{u_n}(y) - GPD_{\xi, \sigma(u_n)}(y)| = 0, \quad (1.12)$$

con

$$F_{u_n}(y) = \frac{F(u_n + y) - F(u_n)}{1 - F(u_n)}, \quad (1.13)$$

$y = a_n(x - \nu)$ ,  $\sigma(u_n) = \sigma a_n$  y  $x > \nu$ .  $F_{u_n}(y)$  es la función de distribución condicionada dada  $u_n$ , y

$$GPD_{\xi, \sigma(u_n)}(y) = \begin{cases} 1 - (1 + \xi \frac{y}{\sigma(u_n)})^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma(u_n)}} & \text{si } \xi = 0 \end{cases}, \quad (1.14)$$

es la distribución Generalizada de Pareto.

El teorema de Pickands parece prometer una estimación precisa de los cuantiles extremos y de las probabilidades de la cola de los rendimientos financieros cuando la distribución de los datos  $F$  es desconocida. De esta forma *EVT* irrumpió en la econometría financiera como una técnica muy potente que solucionaba los problemas derivados de la ausencia de información en las colas. Esta metodología ha sido implementada antes en otras ciencias como hidrología ó meteorología con cierto éxito. Después de un periodo de euforia reflejado por un gran número de artículos teóricos y aplicados utilizando esta técnica a mediados de la década de los noventa surgió cierto desencanto y desilusión entre los académicos debido a las carencias en el desarrollo de la parte estadística de la teoría de valores extremos, y en los inversores privados debido a su complejidad comparada con la simulación histórica o con los métodos derivados de los supuestos de gaussianidad en los datos. En esta línea se pronuncia un artículo de Diebold, Schuerman y Stroughair (1998) que enumera algunos de los defectos y retos que quedan por delante en la teoría de valores extremos. La recomendación final de estos autores es cierta precaución a la hora de usar esta metodología en econometría financiera y en la gestión del riesgo, y un mejor entendimiento de las situaciones donde estas potentes herramientas probabilísticas y estadísticas son realmente adecuadas y para las que están *diseñadas*.

La mayoría de los temas que se discuten en Diebold et al. están relacionados con los aspectos estadísticos de la teoría, en particular con la estimación de la cola de la distribución y con la aproximación por la distribución *GPD*. Estos autores asumen que la cola de  $F$  tiene un decaimiento exponencial y por lo tanto pertenece al máximo dominio de atracción de una distribución Fréchet, esto es, la distribución del máximo muestral de  $F$  converge débilmente a ese tipo de distribución *EVT*. El estimador del tail index considerado es el estimador de Hill, Hill (1975),

$$\hat{\xi}_n^{Hi}(u_n) = \frac{1}{k} \sum_{i=n-k+1}^n \log \frac{x^{(i)}}{x^{(n-k)}}, \quad (1.15)$$

con  $u_n = x_{(n-k)}$ , la secuencia umbral, y  $x_{(n-k+1)} \leq \dots \leq x_{(n)}$  denotando los estadísticos de orden crecientes.

Este estimador tiene interesantes propiedades teóricas. Es consistente y asintóticamente normal, si se asume que los datos son independientes e idénticamente distribuidos (*iid*) y que  $k$ , el número de las observaciones más altas definidas por el umbral  $u_n = x_{(n-k)}$ , crece a un ratio adecuado,  $k \rightarrow \infty$  con  $k/n \rightarrow 0$ . Se puede ver que la selección del umbral para el estimador de Hill afecta al sesgo y a su varianza. En particular, hay un importante intercambio entre error de sesgo y varianza cuando  $k$  varía para muestras dadas. Diebold et al. *señalan* el problema de la selección del umbral como el primer defecto de esta teoría, *EVT*, en cuanto a la estimación de las colas. Ellos abogan por un procedimiento formal y riguroso para determinar el umbral que sustituyan las reglas ad-hoc basadas en métodos gráficos. Las reglas formales para decidir el umbral podrían ser complementadas por técnicas de remuestreo y simulación Monte-Carlo.

El segundo defecto importante de *EVT* es la ausencia de estudios serios sobre las propiedades finito-dimensionales del estimador del tail index bajo varias elecciones de valor umbral. Ellos mantienen que se necesita un estudio Monte-Carlo intensivo para diversos procesos generadores de los datos. Otros problemas relacionados no mencionados en Diebold et al. se refieren al uso de la aproximación asintótica normal para la distribución del estimador del 'tail index' para *tamaños* de muestra *pequeños*. Los autores remarcan el pobre tratamiento de los aspectos estadísticos de *EVT* en contraste con los resultados proba-

bilísticos. Algunos ejemplos importantes son la ausencia de intervalos de confianza para cuantiles de  $F$ , para estimaciones en la cola de la distribución y la ausencia de contrastes de hipótesis fiables para *tamaños* de muestra *pequeños*.

El segundo capítulo de esta tesis reúne estos problemas y desarrolla una metodología que soluciona la mayoría de estos problemas atribuidos a *EVT*. El capítulo presenta una definición para los valores extremos de una secuencia aleatoria. Esta definición es dada por las observaciones que superan un valor umbral determinado por la teoría asintótica derivada en el teorema de Pickands. En consecuencia la selección del umbral es crucial para la definición de los valores extremos. Se implementa un procedimiento iterativo formal para la selección del umbral de acuerdo a la aproximación *GPD* dada en el teorema de Pickands. La selección del umbral depende de los datos (es un estadístico de orden) y de la longitud de la secuencia. Por lo tanto para formalizar la definición de los valores extremos introducimos un innovador contraste de hipótesis bootstrap. Este bootstrap semi-paramétrico se beneficia de la información de la serie histórica para la mayoría de los datos y del ajuste *GPD* en la cola. La distribución muestral bootstrap obtenida hace inmediato el estudio de las propiedades muestrales de los diferentes estimadores del tail index considerados y la inferencia estadística (intervalos de confianza para cuantiles y medidas en la cola de la distribución). Los principales beneficiarios de estos métodos de estimación tan precisos son las medidas de riesgo definidas para las series financieras. En particular VaR y el ratio de decaimiento de la cola que son estimados en este capítulo para series financieras de los mayores mercados bursátiles mundiales, y se validan mediante los intervalos de confianza bootstrap desarrollados en el capítulo.

Los supuestos mantenidos en Diebold et al. son el decaimiento polinomial en las colas y las observaciones *iid*. En consecuencia, situaciones fuera de estos supuestos traerán más problemas en los aspectos estadísticos de *EVT*. En particular el supuesto *iid* para datos financieros econométricos no es realista. La disponibilidad de datos intra-diarios de alta frecuencia mejora y motiva el uso de técnicas *EVT*, pero por otro lado conllevan dependencia temporal. La literatura financiera explica la dependencia mediante la correlación en los segundos momentos. En otras palabras, mientras los rendimientos son incorrelados, la



volatilidad condicional depende de la información pasada. Junto con la dependencia en la volatilidad hay otros hechos estilizados que caracterizan los rendimientos financieros. Estos son el agrupamiento (clustering) de las observaciones más grandes en ambas colas, la magnitud de estas observaciones que está lejos de ser explicada por las distribuciones normales (colas pesadas?), la asimetría que se observa en la distribución de los rendimientos, y el efecto apalancamiento (leverage effect), esto es, después de periodos de alta volatilidad se producen periodos de rendimientos negativos. Un primer vistazo a este tipo de secuencias rápidamente descartaría el uso de modelos gaussianos, sin embargo los inversores y los investigadores siguen *empeñados* en proponer modelos para describir la dinámica en estos datos modelizando la volatilidad.

El tercer capítulo reúne los retos estadísticos de *EVT* relativos a la dependencia en los extremos y los problemas encontrados en la econometría financiera para modelizar los hechos estilizados mencionados anteriormente.

Las contribuciones en este capítulo pueden ser divididas en dos grupos. Primero, se propone un nuevo estimador para el parámetro que describe la dependencia en los extremos, denotado extremal index ( $\theta$ ), y segundo considera el clustering en las secuencias financieras como una característica de las observaciones más grandes en vez de recurrir a la dependencia en los segundos momentos de la distribución condicional al pasado de los datos. Las propiedades estadísticas del estimador para  $\theta$  son muy interesantes. En particular, es asintóticamente normal haciendo la inferencia muy sencilla. Por lo tanto, el clustering en los extremos puede ser contrastado. Otros contrastes de hipótesis derivados de la distribución asintótica del estimador se desarrollan con aplicación inmediata a contrastes para el resto de los hechos estilizados hallados en las series financieras. Las conclusiones están de acuerdo con las teorías que consideran la presencia de colas pesadas en series financieras más que con distribuciones condicionalmente gaussianas con volatilidad heteroscedástica.

Un problema mayor en *EVT* es el caso multivariante. No hay una extensión natural de esta teoría a varias variables. Más aún, los problemas típicos encontrados para mayores dimensiones, sirva como ejemplo la maldición de la dimensionalidad, aumentan en el contexto de valores extremos debido a la ausencia de datos. Por lo tanto, aunque las aplicaciones

en el mundo real son numerosas, en particular en econometría financiera y en gestión del riesgo para modelizar dependencia entre mercados financieros en periodos de crisis, estas son oscurecidas por la ausencia de datos y de técnicas rigurosas para dimensiones mayores que dos.

El cuarto capítulo se centra en estos lazos entre mercados resaltando el fenómeno de contagio, esto es, la transmisión de crisis de economías enfermas a economías sanas. Hasta ahora esto se ha medido mediante medidas lineales como la correlación. Esta metodología no está *diseñada* para medir causalidad entre las variables y en consecuencia fracasa en describir la fuente y el nivel de dependencia entre las variables en periodos de crisis. Embrechts, McMeil y Straumann (1999) revisan los problemas de la correlación de Pearson y presentan diversas alternativas a esta medida estándar de dependencia. El mensaje viene a ser que la correlación de Pearson es una medida adecuada para la dependencia sólo para la distribución normal multivariante. En este caso las distribuciones marginales y la matriz de correlaciones son suficiente para describir completamente la dependencia existente en los datos.

Basándonos en la correlación dos funciones de distribución multivariantes con la misma matriz de covarianzas no pueden ser diferenciadas, y más importante para nuestro caso, la correlación no nos dice nada acerca del grado de dependencia en las colas de las distribuciones de los datos. Otros problemas de la correlación son enumerados en la siguiente lista.

1. Correlación es simplemente una medida escalar de dependencia. No está *diseñada* para describir completamente la estructura de dependencia en los datos.
2. Correlación depende de las distribuciones marginales. Todos los valores entre  $-1$  y  $1$  no son necesariamente obtenibles.
3. Dos variables con una dependencia positiva perfecta no tienen necesariamente una correlación de  $1$ . A su vez una dependencia perfecta negativa no alcanza necesariamente un valor de  $-1$ .

4. Una correlación de cero no indica independencia entre las variables.
5. Correlación no es invariante bajo transformaciones de los riesgos.
6. Correlación está sólo definida cuando las varianzas de las correspondientes variables son finitas.

Embrechts et al. recomiendan las funciones cópula, un método nuevo y prometedor en economía, para modelizar dependencia. Este método se basa en la misma idea planteada en la discusión sobre los conceptos de varianza y riesgo. Mientras que la varianza es una medida lineal que describe la dispersión en los datos, el riesgo necesita de la función de distribución completa. En el contexto multivariante el papel de la varianza es asumido por la correlación, y la noción de riesgo es reemplazada por el contagio y el nivel de dependencia en las colas. Por lo tanto, del mismo modo que el riesgo necesitaba de  $F$ , el contagio necesita de la distribución multivariante completa. La función cópula extrae la estructura de dependencia de la función de distribución conjunta. Sklar (1959) muestra que toda función de distribución conjunta se puede escribir como

$$H(x_1, \dots, x_m) = C(F_1(x_1), \dots, F_m(x_m)), \quad (1.16)$$

con  $H$  la función de distribución multivariante,  $F_i$ ,  $i = 1, \dots, m$  las marginales,  $C$  la función cópula, y  $m$  el número de variables aleatorias.

Como todo en la vida esta poderosa herramienta estadística no está libre de problemas. Hay un conjunto enorme de posibles cópulas, pero no hay un método formal para discriminar entre ellas. Contrastes de bondad de ajuste en el caso multivariante no son sencillos, y dependen fuertemente del conocimiento de las distribuciones marginales. Además, la mayoría de las técnicas para contrastar bondad de ajuste en las cópulas tienen problemas computacionales que hacen la implementación en práctica inviable. Por lo tanto la elección de la cópula suele reemplazarse en la práctica por métodos ad-hoc. Las otras deficiencias importantes de la mayoría de las cópulas es la simetría. Las funciones cópula son normalmente simétricas y por lo tanto no están *diseñadas* para recoger diferentes contribuciones

en la dependencia de las correspondientes variables aleatorias. Finalmente, la dinámica en la dependencia multivariante no han sido exploradas en detalle. Las cópulas condicionales introducidas en Patton (2001) son un primer intento de modelizar la dependencia que varía en el tiempo, aunque estas cópulas se reducen a imponer cierta dinámica a los parámetros que rigen la dependencia en las cópulas estándar. Por otro lado los supuestos condicionales no están bien especificados en general.

En el capítulo IV la dependencia multivariante entre diferentes mercados financieros se divide en dos grupos: dependencia racional e irracional. La dependencia racional se debe a fundamentos económicos, y se describe mediante modelos de regresión univariantes y filtros para la volatilidad. La dependencia irracional se refleja en los lazos en el vector de innovaciones y es modelizada por las funciones cópula. El énfasis en este capítulo está en los lazos entre los mercados financieros de bonos y los mercados de acciones, y en contrastar el fenómeno de contagio frente al de vuelo hacia la calidad (*flight to quality*), que consiste en salidas de capital de los mercados de acciones a los de bonos en épocas de crisis de los primeros. La cópula introducida es una nueva variante de la cópula Gumbel y es suficientemente flexible para describir los efectos asimétricos entre las variables. Esta cópula es *diseñada* para ser capaz de reflejar estos efectos asimétricos, y por lo tanto de describir causalidad en los valores extremos. La elección de la cópula Gumbel está motivada por la teoría de valores extremos multivariante y las propiedades de dependencia entre el vector del máximo. Los conceptos de contagio e interdependencia son revisados y se adaptan para ser definidos como propiedades de la cola de las distribuciones.

Finalmente, el último capítulo esboza las líneas de investigación futuras a seguir por el autor y las principales conclusiones encontradas en la tesis.

## Capítulo 2

# El Riesgo está en las colas: Una Definición formal de los Valores Extremos

El objetivo de este capítulo es definir y estimar los extremos de cualquier muestra aleatoria simple de tamaño  $n$ , de una función de distribución  $F$ . Esto se hace mediante la selección de una secuencia umbral y de un contraste de bondad de ajuste. Los valores extremos son las observaciones que exceden tal umbral y que a su vez siguen un tipo de distribución Generalizada de Pareto ( $GPD$ ) que incluye el tail index de  $F$ . El umbral es el estadístico de orden que minimiza una generalización de la distancia del supremo entre la función de distribución empírica de las correspondientes observaciones más grandes y la correspondiente  $GPD$ . Esta generalización incluye el estadístico de Kolmogorov-Smirnov como un caso particular. Una vez que el conjunto de candidatos a valores extremos se identifica usamos un bootstrap semi-paramétrico para contrastar la correspondiente aproximación  $GPD$  (segunda parte de la definición). Las simulaciones Monte-Carlo muestran muy buenos resultados en muestra del citado contraste. Finalmente, usamos nuestra metodología para estimar el tail index y el Valor en Riesgo de algunos de los índices financieros de los mercados bursátiles más importantes del mundo.

## 2.1. Introducción

La gestión del riesgo es una de las innovaciones más importantes del siglo XX en Economía. Durante la última década los mercados financieros se han dado cuenta de la importancia de controlar y vigilar el riesgo. La pregunta que a uno le gustaría contestar es: “si las cosas van mal, cómo de mal pueden llegar a ir ?” La varianza utilizada como una medida de riesgo es incapaz de responder a esta cuestión.

Necesitamos definir medidas alternativas que tienen en cuenta posibles valores fuera del rango de la información disponible. La teoría de valores extremos (*EVT*) proporciona las herramientas para modelizar la distribución asintótica del máximo de una secuencia de variables aleatorias  $\{X_n\}$ , y en este sentido esta teoría puede ser muy útil para obtener una primera impresión sobre cómo de mal las cosas pueden llegar a ir. Un análisis más profundo en *EVT* nos permite saber no sólo el orden de convergencia del máximo sino también la distribución límite de las observaciones más grandes de la muestra aleatoria simple. Estas observaciones son el principal ingrediente de medidas de riesgo más informativas que la varianza. Estas medidas han sido introducidas en los últimos tiempos, entre estas se encuentra el Valor en Riesgo (*VaR*) y el Expected Shortfall (*ES*). Estas medidas son funciones de cuantiles extremos de la distribución de los datos. El intento para modelizar las colas de estas distribuciones es engorroso y metodologías estándar como la simulación histórica ó la distribución gaussiana no ofrecen aproximaciones fiables para cuantiles muy altos.

Por otro lado, la metodología que se deriva de *EVT* cubre estos problemas, y produce un marco paramétrico para derivar el *VaR* o cualquier función de este cuantil extremo. Está claro que la primera tarea es identificar qué valores son realmente extremos. En la práctica esto se hace mediante métodos gráficos como el QQ-plot, Sample Mean Excess Plot o por cualquier otro método ad-hoc que imponga un umbral arbitrario (5 %, 10 %, ...), ver Embrechts, Klüppelberg y Mikosch (1997). Estos métodos no proponen un método formal y calculable, más aún, sólo dan estimaciones muy bastas del conjunto de valores extremos. En este capítulo proponemos un modo formal de identificar y estimar los valores extremos de cualquier muestra aleatoria de tamaño  $n$  y viniendo de una función de distribución  $F$ .

Estos valores van a ser definidos como las observaciones que exceden una secuencia umbral  $\{u_n\}$  y que siguen un tipo determinado de distribución Generalizada de Pareto (GPD). La selección del umbral es una parte crucial en esta definición de extremos y en la estimación de los parámetros de la *GPD*. La secuencia de valores extremos depende de la longitud de la secuencia mediante la selección del umbral  $\{u_n\}$ . Por lo tanto necesitamos introducir un contraste apropiado para validar estadísticamente si la función de distribución del conjunto de candidatos a extremos dados por el umbral realmente satisface la convergencia débil a la *GPD* o no, con parámetros dirigidos por el proceso generador de los datos,  $F$ . Para conseguir esto, proponemos un contraste bootstrap semi-paramétrico y estudiamos su comportamiento asintótico y en muestras finitas.

El propósito final de nuestra metodología es obtener una aproximación fiable de  $F$  prestando especial atención a las colas de la distribución. Nuestra estimación de la cola ofrece una aproximaciones acertadas de los cuantiles extremos de  $F$ , y de estos es inmediato calcular las medidas de riesgo introducidas en la literatura financiera.

El capítulo se estructura de esta forma. En la sección 2 presentamos algunos resultados generales sobre la teoría de valores extremos centrándonos en la convergencia débil de las observaciones mayores de la secuencia aleatoria. La sección 3 introduce diferentes enfoques para seleccionar la secuencia umbral y da un breve repaso de los diferentes métodos de estimación para los parámetros de la *GPD*. Resultados de simulación muestran el comportamiento de nuestro enfoque en términos de estimación del tail index. La definición completa de la secuencia de valores extremos se da en la sección 4 por medio de contrastes de hipótesis bootstrap. Simulaciones Monte-Carlo proporcionan el comportamiento en muestras finitas de el test propuesto en este capítulo. La sección 5 presenta una aplicación empírica donde el riesgo de las secuencias financieras de los índices financieros de los mercados bursátiles más importantes del mundo es analizado por medio del tail index y del *VaR*. Finalmente, la sección 6 concluye con varios comentarios y extensiones. Las pruebas están reunidas en el apéndice.

## 2.2. Resultados de la Teoría de Valores Extremos

El propósito de esta sección es introducir brevemente los resultados más importantes de la teoría de valores extremos y que serán necesarios para desarrollar la teoría usada en el capítulo. El punto de partida es el estudio de la convergencia débil para el máximo muestral de una secuencia de variables aleatorias  $\{X_n\}$  con función de distribución  $F$ . Nuestra intención es usar la distribución límite de este estadístico para derivar la convergencia débil de las observaciones más grandes de una secuencia aleatoria bajo el menor número posible de supuestos sobre la distribución  $F$ .

Sea  $M_n = \max\{X_1, \dots, X_n\}$  el máximo muestral de la secuencia y sea  $F$  la distribución común de los datos  $\{X_n\}$ . Nuestro primer objetivo es introducir las condiciones bajo las cuales  $M_n$  converge débilmente a una función de distribución no degenerada.

**Resultado 2.2.1.** *Sea  $\{X_n\}$  una secuencia de observaciones independientes e idénticamente distribuidas (iid). Sea  $0 \leq \tau \leq \infty$  y supongamos que  $\{u_n\}$  es una secuencia de números reales tal que*

$$n(1 - F(u_n)) \rightarrow \tau \quad \text{cuando } n \rightarrow \infty. \quad (2.1)$$

Entonces

$$P\{M_n \leq u_n\} \rightarrow e^{-\tau} \quad \text{cuando } n \rightarrow \infty. \quad (2.2)$$

La implicación en la otra dirección también es cierta, si la condición (2.2) se cumple para algún  $\tau$ ,  $0 \leq \tau \leq \infty$ , entonces también se cumple la condición (2.1).

La prueba de este resultado sale de forma inmediata de

$$P\{M_n \leq u_n\} = F^n(u_n) = \left(1 - \frac{n(1 - F(u_n))}{n}\right)^n. \quad (2.3)$$

Sin embargo este resultado no garantiza la existencia de una función de distribución no degenerada para  $M_n$ . Definamos el límite por la derecha del rango de la distribución  $F$



como  $x_F = \sup\{x|F(x) < 1\} \leq +\infty$ . Está claro que  $M_n \rightarrow x_F$  con probabilidad 1 cuando  $n \rightarrow \infty$ . Supongamos ahora que  $F$  tiene un salto en  $x_F$  con  $x_F < \infty$  (i.e.  $F(x_{F-}) < 1$ ) y consideremos una secuencia  $\{u_n\}$  que satisface (2.2) con  $0 \leq \tau \leq \infty$ . Entonces, o  $u_n < x_F$  para infinitos valores de  $n$  y  $n(1 - F(u_n)) \rightarrow \infty$ , o  $u_n > x_F$  y  $n(1 - F(u_n)) = 0$ . Por lo tanto necesitamos también cierta condición de regularidad en la cola de  $F$  para evitar la existencia de saltos.

**Resultado 2.2.2.** *Sea  $F$  una distribución con límite por la derecha  $x_F$  tal que*

$$\lim_{x \uparrow x_F} \frac{1 - F(x)}{1 - F(x^-)} = 1, \quad (2.4)$$

y sea  $\{u_n\}$  una secuencia con  $u_n < x_F$  y  $n(1 - F(u_n)) \rightarrow \tau$ . Entonces  $0 < \tau < \infty$ .

La selección de la secuencia  $\{u_n\}$  determina el valor de  $\tau$ . Supongamos que  $v_n > u_n$  y (2.2) se cumplen, entonces  $n(1 - F(v_n)) \rightarrow \tau'$  con  $\tau' < \tau$ . Podemos escribir la expresión (2.2) como  $P\{M_n \leq u_n(x)\} \rightarrow e^{-\tau(x)}$ , con  $u_n$  dependiendo de  $x$ . Más aún, existen ciertas secuencias de normalización  $a_n, b_n$  que varían de acuerdo a  $F$  tal que

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \rightarrow G(x) \quad \text{cuando } n \rightarrow \infty, \quad (2.5)$$

con  $u_n(x) = a_n x + b_n$  y  $G(x) = e^{-\tau(x)}$  una función de distribución. Esta función ha sido completamente caracterizada por Gnedenko (1943) ó de Haan (1976) via el análisis de dominios de atracción para el máximo, y se puede resumir como sigue:

**Resultado 2.2.3.** *La función de distribución  $G(x)$  derivada en la expresión (2.5) solo puede ser de una de las tres siguientes formas,*

Tipo I: (Gumbel)  $G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty,$

Tipo II: (Frèchet)  $G(x) = \begin{cases} 0 & x \leq 0, \\ e^{-x^{-\frac{1}{\xi}}} & x > 0, \xi > 0 \end{cases}$

$$\text{Tipo III: (Weibull)} \quad G(x) = \begin{cases} 1 & x \geq 0, \\ e^{-(-x)^{-\frac{1}{\xi}}} & x < 0, \xi < 0 \end{cases}.$$

El parámetro  $\xi$  es el tail index de  $F$  y caracteriza el comportamiento de la función de distribución. Los tres tipos se pueden agrupar en la denominada Distribución Generalizada de Pareto, propuesta por primera vez en von Mises (1936),

$$G(x) = e^{-(1+\xi\frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}}, \tag{2.6}$$

donde  $\mu$  es un parámetro de localización,  $\sigma$  un parámetro de escala y  $\xi \neq 0$ . Esta expresión se puede escribir como  $G(x) = e^{-e^{-\frac{x-\mu}{\sigma}}}$  cuando  $\xi = 0$ .

Claramente  $\tau(x) = (1 + \xi\frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}$  en expresión (2.5), y entonces  $n(1 - F(u_n(x))) \rightarrow (1 + \xi\frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}$  para todo  $x$ , donde  $a_n, b_n$  son constantes apropiadas. Este es el resultado que explotamos para derivar la convergencia débil de las observaciones más grandes determinadas por una secuencia umbral  $u_{on} = a_n\mu + b_n$ , con  $\mu$  que verifica  $-\log G(\mu) = 1$ . De esta forma

$$\frac{1 - F(u_n(x))}{1 - F(u_{on})} \rightarrow (1 + \xi\frac{x - \mu}{\sigma})^{-\frac{1}{\xi}}, \quad \text{cuando } n \rightarrow \infty. \tag{2.7}$$

Esta expresión se puede reescribir como

$$\frac{F(u_n(x)) - F(u_{on})}{1 - F(u_{on})} \rightarrow 1 - (1 + \xi\frac{x - \mu}{\sigma})^{-\frac{1}{\xi}}, \tag{2.8}$$

para todo  $x > \mu$  punto de continuidad. La secuencia que define el umbral satisface que  $u_n(x) = u_{on} + a_n(x - \mu)$ , y podemos definir

$$F_{u_{on}}(a_n(x - \mu)) = \frac{F(u_{on} + a_n(x - \mu)) - F(u_{on})}{1 - F(u_{on})}, \tag{2.9}$$

como la función de distribución condicional dada por  $u_{on}$  con  $x > \mu$ . Esto nos lleva al siguiente resultado:

**Resultado 2.2.4.** Sea  $y = a_n(x - \mu)$ , entonces

$$\lim_{u_{on} \rightarrow x_F} \sup_{[0 \leq y < \infty]} |F_{u_{on}}(y) - GPD_{\xi, \sigma(u_{on})}(y)| = 0, \quad (2.10)$$

con

$$GPD_{\xi, \sigma(u_{on})}(y) = \begin{cases} 1 - (1 + \xi \frac{y}{\sigma(u_{on})})^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma(u_{on})}} & \text{si } \xi = 0 \end{cases}, \quad (2.11)$$

la distribución Generalizada de Pareto y  $\sigma(u_{on}) = \sigma a_n$ .

Este resultado es conocido como el teorema de Pickands (1975). Pickands propuso una secuencia  $u_{on}$  en el intervalo  $[b_n, b_{n+1}]$  con  $b_n$  una secuencia apropiada en (2.5). Esta aproximación para la distribución de las observaciones más grandes vistas como los valores que exceden la secuencia umbral se puede mejorar cuando la cola de  $F$  decae a un ratio polinomial.

Supongamos  $1 - F(x) = x^{-\frac{1}{\xi}} L(x)$  con  $L(tx)/L(x) \rightarrow 1$  cuando  $x \rightarrow x_F$  y  $\xi > 0$ ; entonces la función de distribución  $F$  satisface

$$\lim_{x \uparrow x_F} \frac{1 - F(tx)}{1 - F(x)} = t^{-\frac{1}{\xi}}, \quad t > 0. \quad (2.12)$$

Este tipo de distribuciones se dicen que varían regularmente a un ratio  $\frac{1}{\xi}$  y el dominio de atracción del máximo muestral es la distribución de Fréchet (ver Resnick (1987) ó de Haan (1976)). La función  $L(x)$  se dice que varía lentamente y se introduce para incluir las desviaciones de  $F$  de la familia de distribuciones de Pareto. Cuando la “distancia” de estas distribuciones de las distribuciones Pareto son *pequeñas*,  $F_{u_{on}}(y)$  se aproxima mejor por la función de distribución Pareto. Consideremos una secuencia  $u_n(x) = u_{on}x$  donde  $u_{on} = u_n(1)$  es el umbral que satisface  $1 - F(u_{on}) = u_{on}^{-\frac{1}{\xi}} L(u_{on})$ . La función de distribución condicional de los excesos definida por  $u_{on}$  como  $F_{u_{on}}(u_n(x)) = \frac{F(u_n(x)) - F(u_{on})}{1 - F(u_{on})}$  satisface

$$F_{u_{on}}(u_n(x)) \rightarrow 1 - \left(\frac{u_n(x)}{u_{on}}\right)^{-\frac{1}{\xi}}, \quad \text{cuando } n \rightarrow \infty, \quad (2.13)$$

para  $u_n(x) \geq u_{on}$  ó equivalentemente para  $x \geq 1$ . Esta convergencia se cumple para todo punto de continuidad de  $F$  y por lo tanto para este caso podemos reescribir el resultado anterior como

$$\lim_{u_{on} \rightarrow x_F} \sup_{[u_{on} \leq y < \infty]} |F_{u_{on}}(y) - PD_\xi(y)| = 0, \quad (2.14)$$

con  $y = u_n(x)$  y  $PD_\xi(y) = 1 - \left(\frac{y}{u_{on}}\right)^{-\frac{1}{\xi}}$ .

Finalmente, la elección del umbral tiene también un efecto en el error cometido por las aproximaciones demostradas en el teorema de Pickands. Este error surge de la relación asintótica  $n(1 - F(u_n)) \rightarrow \tau$  y de la aproximación de  $F^n(u_n)$  por la distribución exponencial. Esta última aproximación es de orden  $o(n^{-1})$  ya que

$$0 \leq e^{-x} - \left(1 - \frac{x}{n}\right)^n \leq 0,3 \frac{1}{n-1},$$

para  $0 \leq x \leq n$  (ver, e.g., Leadbetter, Lindgren y Rootzén (1983)). No obstante si  $F$  es continua uno siempre puede obtener la igualdad en la expresión (2.2) tomando  $u_n = F^{-1}\left(e^{-\frac{x}{n}}\right)$  y haciendo que los errores de aproximación desaparezcan. Sin embargo secuencias del tipo  $u_n(x) = a_n x + b_n$ , con  $a_n, b_n$  ciertas constantes son más apropiadas para estudiar la convergencia débil de  $M_n$ . En estos casos la igualdad ó cualquier cota uniforme para todo  $x$  no son normalmente alcanzables en la expresión (2.5).

### **2.3. Selección del umbral para definir los Valores Extremos**

La última sección se ha centrado en encontrar las leyes de probabilidad asintóticas que rigen las observaciones más grandes de una secuencia aleatoria de una función de distribución  $F$ . Este conjunto de observaciones se define por medio de una secuencia umbral y del tail index que caracteriza la correspondiente distribución Generalizada de Pareto o Pareto. La elección de esta secuencia es problemática y engorrosa ya que  $u_{on} \rightarrow x_F$  cuando  $n \rightarrow \infty$ , pero a cierto ratio. Este orden de convergencia depende de  $F$  representado

por las secuencias  $a_n$  y  $b_n$  cuando  $u_n(x)$  es de la forma  $u_n(x) = a_n x + b_n$ . Entonces la secuencia umbral  $u_{on}$  se puede definir por las secuencias  $a_n$ ,  $b_n$  y el valor de  $x$  que satisface la condición  $-\log G(x) = 1$ , o equivalentemente  $n(1 - F(u_{on})) \rightarrow 1$ . Usaremos a partir de ahora la notación  $u_n$  en vez de  $u_{on}$  para denotar la secuencia umbral que satisface estas condiciones. Esta secuencia se deriva de forma inmediata mediante cálculos analíticos cuando  $F$  es conocida. Consideremos como ejemplo el caso  $F(x) = 1 - e^{-x}$ . Por continuidad de  $F$  podemos elegir  $u_n(x) = F^{-1}(1 - \frac{\tau(x)}{n})$  con  $\tau(x) > 0$ , y tal que  $u_n(x) = -\log \tau(x) + \log n$ . La expresión (2.2) se escribe como

$$P\{M_n \leq -\log \tau(x) + \log n\} \rightarrow e^{-\tau(x)},$$

y entonces  $P\{M_n - \log n \leq x\} \rightarrow e^{-e^{-x}}$ , con  $\tau(x) = e^{-x}$  para todo  $x > 0$ . Las secuencias de normalización son  $a_n = 1$ ,  $b_n = \log n$ , y la secuencia umbral queda  $u_n = \log n$ , ya que  $-\log G(0) = 1$ . Más ejemplos de derivación analítica de estas secuencias se pueden encontrar en Leadbetter, Lindgren y Rootzén (1983).

En general  $F$  no es conocido, y en este marco no podemos obtener las secuencias de forma directa ni por comparación de diferentes secuencias umbral. Este método de comparación se lleva a cabo analizando las propiedades del estimador del tail index de  $F$ , ya que la mayoría de estos estimadores para  $\xi$  vienen unidos a una selección del umbral. Por lo tanto sus sesgos y varianzas están influenciados por el efecto de la selección del umbral  $u_n$ . Hay una abundante literatura en estimación del tail index (capítulo VI de Embrechts, Klüppelberg y Mikosch (1997) da un excelente repaso). Entre estos estimadores, los más populares son el estimador de Hill (1975) y el estimador de Pickands (1975). El primero de éstos toma esta forma

$$\hat{\xi}_n^{Hi}(u_n) = \frac{1}{k} \sum_{i=n-k+1}^n \log \frac{x_{(i)}}{x_{(n-k)}}, \quad (2.15)$$

con  $u_n = x_{(n-k)}$ ,  $x_{(n-k+1)} \leq \dots \leq x_{(n)}$  que denota los estadísticos de orden de forma creciente y  $k$  un valor entero positivo en el intervalo  $[1, n]$ . El estimador de Pickands para

el tail index toma esta expresión

$$\hat{\xi}_n^{Pi}(u_n) = \frac{1}{\log(2)} \log\left(\frac{x_{(n-k+1)} - x_{(n-2k+1)}}{x_{(n-2k+1)} - x_{(n-4k+1)}}\right), \quad (2.16)$$

y

$$\hat{\sigma}_n^{Pi}(u_n) = \frac{x_{(n-2k+1)} - x_{(n-4k+1)}}{\int_0^{\log 2} e^{\hat{\xi}_n^{Pi}(x_{(n-4k+1)})t} dt}, \quad (2.17)$$

para la varianza con  $u_n = x_{(n-4k+1)}$  y  $k = 1, \dots, n/4$ . Hay ciertas características de ambos estimadores que merece la pena mencionar. Estos estimadores dependen en gran medida de la selección del umbral  $u_n$ , y ambos se pueden derivar bajo el supuesto de que la función de distribución condicionada  $F_{u_n}$  es exactamente Pareto con parámetro  $\xi$  o Generalizada de Pareto con parámetros  $\xi$  y  $\sigma(u_n)$ . Más aún, si  $F_{u_n} = PD_\xi$ , el estimador de Hill es el estimador de máxima verosimilitud (Mv) de  $\xi$  y posee todas las propiedades asintóticas de este tipo de estimadores: consistencia y distribución asintótica normal. Este enfoque sólo es válido para distribuciones que varían regularmente, *i.e.*  $\xi > 0$ , de otro modo las propiedades asintóticas de este estimador varían de acuerdo a la forma de  $F$  (ver Davis y Resnick (1984)).

El estimador de Pickands para el tail index se obtiene bajo el supuesto  $F_{u_n} = GPD_{\xi, \sigma(u_n)}$  y tomando la función inversa de la distribución paramétrica  $GPD$ . Este estimador es consistente y converge también a una distribución normal; pero en cambio es muy sensible a la elección de la secuencia  $u_n$ . Alternativamente, bajo este último supuesto paramétrico en  $F_{u_n}$  se puede obtener el estimador de máxima verosimilitud para los parámetros  $\xi$  y  $\sigma(u_n)$  de la  $GPD$ . En este caso no existe una expresión cerrada para los estimadores de máxima verosimilitud de estos parámetros, y nos tenemos que basar en procedimientos numéricos (ver Press (1992)). El estimador de máxima verosimilitud para el tail index es consistente y asintóticamente normal para  $\xi > -\frac{1}{2}$ , tal y como se comenta en Smith (1985).

La selección del umbral se lleva a cabo por medio del estudio del error cuadrático medio de estos  $\xi$  estimadores, cuando  $u_n$  varía. Sin embargo se requiere que la función de

distribución  $F$  tenga cierta forma funcional explícita. Bajo el supuesto

$$1 - F(x) = Cx^{-\frac{1}{\xi}}[1 + Dx^{-\beta} + o(x^{-\beta})], \quad (2.18)$$

donde  $\xi > 0, C > 0, \beta > 0$  y  $D$  es un número real, Hall (1982) propuso diferentes estimadores para el tail index basados en una selección óptima de estadísticos de orden intermedios como candidatos para la secuencia umbral. No obstante uno de los primeros trabajos en selección del umbral es Pickands (1975), donde  $F$  no es necesariamente como en (2.18). La estimación del tail index y la selección del umbral se hacen en un solo paso. Pickands propuso como candidato para la secuencia umbral al estadístico de orden de una muestra  $\{x_n\}$  que minimiza la distancia  $d^\infty$  entre las funciones de distribución  $F_{u_n, n}$  y  $GPD_{\hat{\xi}_n^{Pi}(u_n), \hat{\sigma}_n^{Pi}(u_n)}$ . La función de distribución empírica de los valores que exceden  $u_n$ ,  $F_{u_n, n}(x)$  con  $x > u_n$  se define como

$$F_{u_n, n}(x) = \frac{\sum_{i=1}^n \mathbf{1}_{\{u_n < x_i \leq x\}}}{\sum_{j=1}^n \mathbf{1}_{\{x_j > u_n\}}}, \quad (2.19)$$

o equivalentemente, vía la transformación  $y = a_n(x - u_n) > 0$ , por

$$F_{u_n, n}(y) = \frac{\sum_{i=1}^n \mathbf{1}_{\{0 < y_i \leq y\}}}{\sum_{j=1}^n \mathbf{1}_{\{y_j > 0\}}}. \quad (2.20)$$

La distancia  $d^\infty$  se puede escribir como función de un parámetro  $\theta$ ,

$$d^\infty(F_{\theta, n}, GPD_{\hat{\xi}_n^{Pi}(\theta), \hat{\sigma}_n^{Pi}(\theta)}) = \sup_{0 \leq y < \infty} |F_{\theta, n}(y) - GPD_{\hat{\xi}_n^{Pi}(\theta), \hat{\sigma}_n^{Pi}(\theta)}(y)|. \quad (2.21)$$

El umbral óptimo es entonces

$$u_n^{Pi} = \arg \min_{\theta} d^\infty(F_{\theta, n}, GPD_{\hat{\xi}_n^{Pi}(\theta), \hat{\sigma}_n^{Pi}(\theta)}), \quad (2.22)$$

con  $\theta$  tomando valores a lo largo de la muestra ordenada  $x_{(3n/4)} \leq \dots \leq x_{(n)}$ . Más concretamente,  $u_n^{Pi} = x_{(n-k)}$  con  $k \rightarrow \infty$ ,  $n \rightarrow \infty$  y  $k = o(n)$  para beneficiarse de una *tamaño* de muestra que aumenta (consistencia).

De forma alternativa, proponemos una versión de la distancia  $d^\infty$  donde el número de observaciones de la cola se pondera de forma diferente. Este nuevo enfoque tiene en cuenta los posibles problemas de estimación que se derivan de la falta de observaciones cuando  $\theta$  se aproxima a  $x_F$ .

**Definición 2.3.1.** Sea  $F_{\theta,n}$  la versión empírica de  $F_\theta$  y  $GPD_{\hat{\xi}_n^{Ml}(\theta), \hat{\sigma}_n^{Ml}(\theta)}$  la función de distribución de las observaciones más grandes con parámetros estimados por máxima verosimilitud ( $Mv$ ). Definamos la distancia de Pickands ponderada  $d^{Wp}$  como

$$d^{Wp}(F_{\theta,n}, GPD_{\hat{\xi}_n^{Mv}(\theta), \hat{\sigma}_n^{Mv}(\theta)}) = k^\varepsilon \sup_{0 \leq y < \infty} |F_{\theta,n}(y) - GPD_{\hat{\xi}_n^{Mv}(\theta), \hat{\sigma}_n^{Mv}(\theta)}(y)|, \quad (2.23)$$

con  $0 \leq \varepsilon \leq 1/2$  y  $k = \sum_{j=1}^n 1_{\{x_j > \theta\}}$ .

El parámetro  $\varepsilon$  determina el peso asignado por la distancia  $d^{Wp}$  a las observaciones de la cola definidas por el correspondiente parámetro  $\theta$ . Merece la pena notar que esta distancia es la que usa Pickands cuando  $\varepsilon = 0$ , y el estadístico de Kolmogorov-Smirnov ( $KS$ ), Kolmogorov (1933) cuando  $\varepsilon = 1/2$ . La correspondiente selección del umbral es el estadístico de orden que minimiza la distancia,

$$u_n = \arg \min_{\theta} d^{Wp}(F_{\theta,n}, GPD_{\hat{\xi}_n^{Mv}(\theta), \hat{\sigma}_n^{Mv}(\theta)}), \quad (2.24)$$

con  $\theta$  tomando valores a lo largo de la muestra ordenada  $x_{(1)} \leq \dots \leq x_{(n)}$ . El parámetro  $\varepsilon$  puede ser útil para estudiar el efecto de diferentes escenarios de ponderación en la selección del umbral; sin embargo esto se aleja del objetivo de este capítulo donde nos centraremos únicamente en el valor  $\varepsilon = 1/2$  (estadístico  $KS$ ).

Un análisis preliminar indica que los posibles candidatos a la secuencia umbral que están lejos de  $x_F$  producen estimaciones sesgadas del tail index. Por otro lado  $u_n$  próximo a  $x_F$



dará lugar a estimaciones ineficientes de  $\xi$ .

Goldie y Smith (1987) y Smith (1987) derivan las funciones de distribución asintóticas de ambos estimadores  $Mv$  y el estimador de Hill del tail index para una clase de distribuciones que satisfacen  $1 - F(x) = x^{-\frac{1}{\xi}}L(x)$ , donde  $L(x)$  son funciones de diferentes tipos que varían lentamente en  $x_F$ . Estos autores también discuten en detalle el sesgo y la varianza asintótica para estos estimadores y encuentran que funciones de distribución  $F$  lejos de la función de distribución de Pareto dan lugar a estimaciones sesgadas e ineficientes del tail index para ambos estimadores. Como resultado, una selección adecuada de la secuencia umbral resulta ser de importancia crucial para determinar el error cuadrático medio (MSE).

Hall (1982) deriva una expresión analítica para el MSE del estimador de Hill cuando  $F$  satisface (2.18). Todos estos resultados se consiguen para determinadas clases de funciones de distribución. En contraste, bajo los supuestos presentados en la pasada sección no es posible derivar analíticamente la expresión para el error cuadrático medio para el estimador del tail index. Por lo tanto proponemos intervalos de confianza bootstrap para medir el sesgo y la incertidumbre de los diferentes estimadores del tail index que hemos considerado.

El bootstrap naïve no paramétrico es consistente ya que la función de distribución empírica  $F_n$  es un estimador consistente de  $F$  y  $\sqrt{k}(\hat{\xi}_n^{(i)}(u_n) - \xi)$ ,  $i = Hi, Mv$  converge débilmente a una distribución normal, con  $k$  el número de valores que exceden  $u_n$ . Entonces, la aproximación bootstrap  $J_n(x, F_n)$  a la verdadera distribución muestral  $J_n(x, F)$  de este estadístico se puede utilizar para construir regiones de confianza a un nivel de confianza  $1 - \alpha$  de la siguiente forma

$$\xi \in [\hat{\xi}_n^{Mv}(u_n) - \frac{1}{\sqrt{k}}J_n^{-1}(1 - \frac{\alpha}{2}, F_n), \hat{\xi}_n^{Mv}(u_n) + \frac{1}{\sqrt{k}}J_n^{-1}(\frac{\alpha}{2}, F_n)], \quad (2.25)$$

donde  $J_n^{-1}(1 - \alpha, F_n)$  es el cuantil bootstrap  $1 - \alpha$ . Para implementar (2.25) la aproximación bootstrap se estima mediante

$$\hat{J}_n(x, F_n) = \frac{1}{B} \sum_{j=1}^B 1_{\{\sqrt{k}(\hat{\xi}_{j,n}^{*Mv}(u_{j,n}^*) - \hat{\xi}_n^{Mv}(u_n)) \leq x\}}, \quad (2.26)$$

con  $B$  el número de iteraciones bootstrap,  $\hat{\xi}_{j,n}^{*Mv}(u_{j,n}^*)$  el estimador de máxima verosimilitud para la muestra bootstrap  $j$ , y  $u_{j,n}^*$  la selección correspondiente de la secuencia umbral.

El comportamiento en muestras finitas de los diferentes estimadores se analiza en la Tabla 2.8.1. El umbral  $u_n$  se elige por ambos métodos, Pickands y Pickands ponderado con  $\varepsilon = 1/2$ . Para remarcar la importancia de la selección del umbral para estimar el tail index se incluye también en el análisis un umbral ad-hoc ( $u_n = x_{(\frac{95}{100}n)}$ ).

El experimento de simulación de la Tabla 2.8.1 se realiza para diferentes distribuciones  $t$ -student, donde el tail index  $\xi$  se aproxima de forma adecuada por el inverso de los grados de libertad de la distribución (ver capítulo III de Embrechts, Klüppelberg y Mikosch (1997)).

Antes de discutir los resultados de la Tabla 2.8.1 es importante notar que aunque  $F$  es conocida, reemplazamos esta función por  $F_n$  para calcular la aproximación bootstrap  $J_n(x, F_n)$ . La razón para hacer esto es que el procedimiento bootstrap funciona bien incluso cuando  $F$  no es conocida y sólo tenemos una realización de la secuencia aleatoria  $\{X_n\}$ .

Hay dos resultados claros de esta tabla, Tabla 2.8.1: Primero, los intervalos de confianza para nuestro estimador contienen el verdadero tail index, algo que no ocurre para el método de Pickands; y segundo, los intervalos de confianza estimados de la selección ad-hoc del umbral son más amplios que los derivados de nuestro método cuando  $\xi$  es significativamente más grande que cero.

En la Tabla 2.8.2 se analiza más detalladamente las ventajas del método  $u_n$  cuando los datos vienen de una distribución de colas pesadas. En este caso la distribución  $GPD_{\xi, \sigma(u_n)}$  se reemplaza por la distribución  $PD_\xi$  en (2.3.1) y (2.24).

De la Tabla 2.8.2 concluimos que cuando tratamos con distribuciones de colas pesadas ( $\xi > 0$ ), nuestro método es más eficiente con  $PD$  que con la distribución  $GPD$ . Estos resultados de simulación están en la misma línea que los resultados teóricos derivados en Smith (1987).

## 2.4. Contraste de Hipótesis

Diferentes métodos para elegir el umbral definen diferentes posibles conjuntos de valores extremos de una secuencia concreta  $\{X_n\}$ . En este capítulo las observaciones que exceden cierto umbral se consideran valores extremos sólo si adicionalmente se distribuyen como una  $GPD_{\xi, \sigma(u_n)}$ , con  $\xi$  el tail index de  $F$ . Para verificar esta condición proponemos un contraste de bondad de ajuste para la siguiente hipótesis:

$H_{n,0}$ : la muestra  $\{(x_1 - u_n)_+, \dots, (x_n - u_n)_+\}$  se distribuye como  $GPD_{\xi, \sigma(u_n)}$

frente a la alternativa general de la forma

$H_{n,1}$ : la muestra  $\{(x_1 - u_n)_+, \dots, (x_n - u_n)_+\}$  no se distribuye como  $GPD_{\xi, \sigma(u_n)}$

con  $u_n \in \mathbb{R}$ ,  $\xi$  el tail index de  $F$  y  $(x)_+ = \max(x, 0)$ .

Un contraste de bondad de ajuste natural es el estadístico de  $KS$  (para otros procedimientos de bondad de ajuste consultar Anderson y Darling (1952)),

$$R_k(y; \xi, \sigma(u_n)) = \sqrt{k} \sup_{0 \leq y < \infty} |P_k(y) - GPD_{\xi, \sigma(u_n)}(y)|, \quad (2.27)$$

con  $k = \sum_{j=1}^n 1_{\{x_j > u_n\}}$  y  $P_k$  la función de distribución empírica de las observaciones mayores que  $u_n$ . Cuando los parámetros son conocidos, la distribución asintótica de este test estadístico es tabulada y los valores críticos se pueden derivar. Si los parámetros no son conocidos pero se pueden estimar de forma consistente, la distribución bootstrap es una aproximación adecuada para la verdadera distribución muestral de  $R_k(y; \xi, \sigma(u_n))$ . En este caso se puede probar (ver Romano (1988)) que los valores críticos bootstrap son estimaciones consistentes de los verdaderos valores.

Nuestro interés, sin embargo, no está en la definición de los valores extremos de una secuencia determinada  $\{X_n\}$ ; sino en la definición de los valores extremos de cualquier secuencia de longitud  $n$  con función de distribución  $F$ . En este caso se necesita un contraste

de hipótesis diferente para determinar si el umbral seleccionado es un buen candidato para definir los valores extremos de  $F$  dado el *tamaño* muestral  $n$ . Más formalmente, el problema del contraste de hipótesis que consideramos es

$$H_0 : F_{u_n} = GPD_{\xi, \sigma(u_n)}$$

frente a una alternativa general del tipo

$$H_1 : F_{u_n} \neq GPD_{\xi, \sigma(u_n)},$$

con  $\xi$  el tail index de  $F$ .

Ahora se puede definir de forma formal el conjunto de valores extremos de cualquier secuencia con función de distribución  $F$ .

**Definición 2.4.1.** *Sea  $\{X_n\}$  cualquier secuencia de una función de distribución  $F$ . Los valores extremos de cualquier secuencia de longitud  $n$  de esta distribución vienen dados por las observaciones que exceden el umbral  $u_n$  y que satisfacen  $F_{u_n} = GPD_{\xi, \sigma(u_n)}$ .*

El estadístico del contraste en este caso es una versión de la familia de estadísticos de Kolmogorov-Smirnov,

$$T_n(y_n; \xi, \sigma(u_n)) = \sqrt{n} \sup_{0 \leq y < \infty} |F_{u_n, n}(y) - GPD_{\xi, \sigma(u_n)}(y)|, \quad (2.28)$$

con  $y_i = (x_i - u_n)_+$ ,  $i = 1, \dots, n$ . Este estadístico depende de  $u_n$ ,  $\xi$  y  $\sigma(u_n)$ .

Para derivar la distribución asintótica de (2.28) y dar validez a la aproximación bootstrap se requieren los siguientes resultados. Sea

$$U_\lambda(t) = \frac{P\{\lambda < T \leq t\}}{P\{T > \lambda\}} \quad (2.29)$$

la función de distribución condicionada con parámetro  $\lambda$  en  $[0, 1]$  de una variable aleatoria  $T$  uniforme en  $[0, 1]$ . Su correspondiente empírica

$$U_{\lambda,n}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1_{\{\lambda < t_i \leq t\}}}{\frac{1}{n} \sum_{j=1}^n 1_{\{t_j > \lambda\}}}, \quad (2.30)$$

con  $t_1, \dots, t_n$  y  $t \in [0, 1]$ , define un proceso empírico  $B_n(t) = \sqrt{n}(U_{\lambda,n}(t) - U_\lambda(t))$  similar al proceso empírico uniforme  $\sqrt{n}(U_n(t) - U(t))$ . Es bien sabido que este último converge débilmente a la distribución de un proceso gaussiano de media cero,  $Z_U(\cdot)$  (ver capítulo V de Pollard (1984)). Por un razonamiento análogo se deriva inmediatamente la ley de probabilidad de este proceso  $S_n(y) = \sqrt{n}(F_{u_n,n}(y) - F_{u_n}(y))$  donde la secuencia  $u_n$  toma el papel del parámetro  $\lambda$ .

**Teorema 2.4.1.** *Consideremos una función de distribución continua y estrictamente creciente  $F$  y un umbral  $u_n$ , con  $u_n < x_F$ . El proceso empírico  $S_n(y)$  converge débilmente a la distribución de un proceso gaussiano de media cero  $Z_{F_{u_n}}(\cdot)$  con función de covarianzas*

$$\text{Cov}(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{(F(\min(y_1, y_2)) - F(u_n)) - (F(y_1) - F(u_n))(F(y_2) - F(u_n))}{(1 - F(u_n))^2}, \quad (2.31)$$

con  $y_1, y_2 \in \mathbb{R}$ . Más aún, bajo la hipótesis nula  $H_0$ , este proceso empírico toma esta expresión  $\sqrt{n}(F_{u_n,n}(y) - \text{GPD}_{\xi,\sigma(u_n)}(y))$  y la función de covarianzas es

$$\text{Cov}(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{\text{GPD}_{\xi,\sigma(u)}(\min(y_1, y_2))}{1 - F(u_n)} - \text{GPD}_{\xi,\sigma(u_n)}(y_1)\text{GPD}_{\xi,\sigma(u_n)}(y_2). \quad (2.32)$$

Por el teorema de continuidad funcional (“continuous mapping theorem”), la distribución límite, denotada  $L(x, F)$ , del estadístico  $T_n$  es la distribución del supremo de un proceso gaussiano de media cero con función de covarianzas (2.32). La demostración se encuentra en el apéndice.

Para contrastar  $H_0$  y por tanto para rechazar la hipótesis nula, deberíamos usar el

siguiente criterio

$$\{T_n(y_n; \xi, \sigma(u_n)) > L_n^{-1}(1 - \alpha, F)\}, \quad (2.33)$$

donde  $L_n^{-1}(1 - \alpha, F)$  es el cuantil  $1 - \alpha$  de la distribución muestral  $L_n(x, F)$  del estadístico  $T_n$ . Esta distribución  $L_n$  no es conocida y en la práctica tiene que ser aproximada por la distribución asintótica  $L(x, F)$ . Esta distribución límite toma una forma complicada y que depende del conocimiento de  $F$ , de los parámetros de la GPD, así como del umbral  $u_n$ . La dependencia en los parámetros ruidosos nos obliga a buscar un método alternativo para aproximar la distribución  $L_n(x, F)$ .

### 2.4.1. Aproximación Bootstrap

Sea  $L_n(x, Q_n)$  la distribución bootstrap que aproxima  $L_n(x, F)$ , y  $L_n^{-1}(1 - \alpha, Q_n)$  el cuantil bootstrap que aproxima el cuantil correspondiente de la distribución verdadera muestral  $L_n^{-1}(1 - \alpha, F)$ . Para que el bootstrap sea consistente  $Q_n$  tiene que verificar ciertas condiciones.

**Lema 2.4.1.** *Sea  $Q_n$  un estimador de  $F$  basado en  $\{x_1, \dots, x_n\}$  que satisface  $\sup_{x \in \mathbb{R}} |Q_n(x) - F(x)| \xrightarrow{p} 0$  cuando  $F \in H_0$ , y sea  $L(x, F)$ , la distribución límite del estadístico  $T_n$ , continua y estrictamente creciente. Entonces*

$$P\{T_n > L_n^{-1}(1 - \alpha, Q_n)\} \rightarrow \alpha, \quad \text{cuando } n \rightarrow \infty. \quad (2.34)$$

El bootstrap naïve no paramétrico calculado de la distribución  $Q_n = F_n$  no es capaz de producir una estimación consistente de una función de distribución bajo la hipótesis nula  $H_0$  si  $F$  no pertenece a la nula. Por otro lado, el bootstrap paramétrico derivado de la distribución  $GPD_{\xi, \sigma(u_n)}$  (ver (2.27)) no recoge la estructura de  $F$  para las observaciones menores que el umbral  $u_n$ .

Entonces para cumplir las condiciones del lema (2.4.1) correspondientes a  $Q_n$  y por lo

tanto resolver los dos problemas mencionados anteriormente, se desarrolla una metodología bootstrap semi-paramétrica. Definamos

$$Q_n(x) = \begin{cases} F_n(x), & x \leq u_n \\ GPD_{\xi, \sigma(u_n)}(x - u_n) + F_n(u_n)(1 - GPD_{\xi, \sigma(u_n)}(x - u_n)), & x > u_n. \end{cases} \quad (2.35)$$

Esta función de distribución se deriva del teorema de la probabilidad condicionada. Entonces

$$P\{X \leq x\} = P\{X \leq u_n\}P\{X \leq x \mid X \leq u_n\} + P\{X > u_n\}P\{X \leq x \mid X > u_n\}, \quad (2.36)$$

donde  $P\{X \leq u_n\}$  se aproxima consistentemente por  $F_n(u_n)$ , y bajo la nula  $P\{X \leq x \mid X > u_n\} = GPD_{\xi, \sigma(u_n)}(y)$  con  $y = x - u_n$ .

Usemos  $\{x_n^*\}$  para denotar una muestra bootstrap obtenida de  $Q_n$  y consideremos la muestra transformada bootstrap  $y_i^* = x_i^* - u_n$  con  $i = 1, \dots, n$ . El valor del estadístico es  $t_n(y_1^*, \dots, y_n^*; \xi, \sigma(u_n))$  y por facilidad de notación se denota como  $t_n^*(y_n; \xi, \sigma(u_n))$ . La aproximación bootstrap  $L_n(x, Q_n)$  se estima entonces mediante la distribución empírica de los  $B$  (número de muestras bootstrap) valores de  $T_n$ ,

$$\hat{L}_n(x, Q_n) = \frac{1}{B} \sum_{j=1}^B 1_{\{t_{n,j}^*(y_n; \xi, \sigma(u_n)) \leq x\}}. \quad (2.37)$$

El cuantil  $1 - \alpha$  de  $\hat{L}_n(x, Q_n)$  es el estadístico de orden  $t_{n, \lceil (1-\alpha)B \rceil}^*(y_n; \xi, \sigma(u_n))$  de la secuencia  $\{t_{n,j}^*(y_n; \xi, \sigma(u_n))\}$  de  $B$  elementos, donde  $\lceil x \rceil$  es la parte superior entera de  $x$ .

El criterio de rechazo (2.33) se sustituye ahora por

$$\{T_n(y_n; \xi, \sigma(u_n)) > t_{n, \lceil (1-\alpha)B \rceil}^*(y_n; \xi, \sigma(u_n))\}, \quad (2.38)$$

y entonces para una muestra  $\{x_n\}$  la hipótesis nula se rechaza si  $t_n(y_1, \dots, y_n; \xi, \sigma(u_n))$  está en la región de rechazo. Esto significa que la función de distribución condicionada definida por  $u_n$  no es una  $GPD_{\xi, \sigma(u_n)}$ , y de acuerdo a nuestra definición estos candidatos a

extremos no son realmente extremos.

Recordemos que hasta ahora se ha asumido que los parámetros se conocen. No obstante esta condición raramente se satisface en la práctica. Para hacer nuestro contraste operacional reemplazamos estos parámetros por sus estimadores de máxima verosimilitud, y en vez de  $Q_n$  definimos su distribución análoga empírica  $\hat{Q}_n$ :

$$\hat{Q}_n(x) = \begin{cases} F_n(x), & x \leq u_n \\ GPD_{\hat{\xi}_n^{Ml}(u_n), \hat{\sigma}_n^{Ml}(u_n)}(x - u_n) + F_n(u_n)(1 - GPD_{\hat{\xi}_n^{Ml}(u_n), \hat{\sigma}_n^{Ml}(u_n)}(x - u_n)), & x > u_n \end{cases} \quad (2.39)$$

Notar que la nueva distribución bootstrap  $L_n(x, \hat{Q}_n)$  se reduce a  $L_n(x, Q_n)$  para  $x \leq u_n$ , y para  $x > u_n$  la anterior  $\sqrt{k}$ -converge a la última donde  $k$  es el número de observaciones de la cola definidas por  $u_n$ . Más aún, si  $F$  pertenece a la hipótesis nula definida por  $u_n$ , las condiciones del lema (2.4.1) todavía se verifican y la región de rechazo (2.38) pasa a ser

$$\{\hat{T}_n(y_n; \hat{\xi}_n^{Mv}(u_n), \hat{\sigma}_n^{Mv}(u_n)) > t_{n, ([1-\alpha]B)}^*(y_n; \hat{\xi}_n^{*,Mv}(u_n), \hat{\sigma}_n^{*,Mv}(u_n))\}, \quad (2.40)$$

donde  $\hat{T}_n$  y  $\hat{\xi}_n^{Mv}(u_n), \hat{\sigma}_n^{Mv}(u_n)$  se calculan de la muestra original  $\{x_n\}$  y  $\hat{\xi}_n^{*,Mv}(u_n), \hat{\sigma}_n^{*,Mv}(u_n)$  se estiman de las correspondientes secuencias bootstrap.

### 2.4.2. Comportamiento en muestras finitas: Potencia empírica

La potencia de nuestro contraste,

$$P\{\hat{T}_n > L_n^{-1}(1 - \alpha, \hat{Q}_n)\}, \quad (2.41)$$

depende de tres parámetros clave: la selección del umbral, la función de distribución  $F$  y la longitud  $n$  de la secuencia. Para calcular la potencia de este contraste es importante darse cuenta de que las estimaciones de máxima verosimilitud  $\hat{\xi}_n^{Mv}(u_n), \hat{\sigma}_n^{Mv}(u_n)$  que entran en la expresión de  $\hat{T}_n$  son las mismas utilizadas para definir la hipótesis nula y por lo tanto  $\hat{Q}_n$ .

Este contraste se basa en construir  $\hat{Q}_n$ , tal que su función de distribución condicionada



definida por  $u_n$  sigue una distribución  $GPD_{\hat{\xi}_n^{Mv}(u_n), \hat{\sigma}_n^{Mv}(u_n)}$ . De esta forma las observaciones que vienen de la hipótesis nula son generadas por  $\hat{Q}_n$  y no por  $F$ . El tamaño empírico del contraste se calcula de la distribución anterior. Para entender mejor cómo calcular la potencia vía bootstrap, ver Beran (1986) y Romano (1988).

Los siguientes algoritmos se dedican a describir el experimento de simulación. El Algoritmo 2.4.1 genera muestras bootstrap  $\{x_n^*\}$  de la función de distribución  $\hat{Q}_n$  y calcula la aproximación empírica bootstrap de  $L_n(x, F)$ . El valor del umbral  $u_n$  y las estimaciones de máxima verosimilitud se obtienen de una muestra particular  $\{x_n\}$  generada por  $F$ , y se utilizan para construir  $\hat{Q}_n$ .

**Algoritmo 2.4.1. (Procedimiento Bootstrap):**

1.  $l = 1$ .
2. Generar  $x_{1,l}^*, \dots, x_{n,l}^*$  extraídas de  $\hat{Q}_n$ .
3. Calcular  $\hat{\xi}_n^{*Mv}(u_n)$  y  $\hat{\sigma}_n^{*Mv}(u_n)$  de la muestra bootstrap.
4.  $t_{n,l}^*(y; \hat{\xi}_n^{*Mv}(u_n), \hat{\sigma}_n^{*Mv}(u_n)) = \sqrt{n} \sup_{0 \leq y < \infty} |F_{u_n, n}(y) - GPD_{\hat{\xi}_n^{*Mv}(u_n), \hat{\sigma}_n^{*Mv}(u_n)}(y)|$   
con  $y = x - u_n$ .
5.  $l++$ . Ir al paso 2 mientras  $l \leq B$ .
6.  $\hat{L}_n(x, \hat{Q}_n) = \frac{1}{B} \sum_{j=1}^B 1_{\{t_{n,j}^*(y; \hat{\xi}_n^{*Mv}(u_n), \hat{\sigma}_n^{*Mv}(u_n)) \leq x\}}$

En la práctica, el  $p$ -valor reemplaza el criterio de rechazo dado en expresión (2.40). El  $p$ -valor empírico es

$$p = \frac{1}{B} \sum_{j=1}^B 1_{\{t_{n,j}^* > \hat{t}_n\}}, \quad (2.42)$$

con  $\hat{t}_n$  obtenido de la muestra  $\{x_n\}$ .

La probabilidad (2.41) no se puede derivar directamente y tenemos que utilizar experimentos de simulación de Monte-Carlo para calcularlo. El siguiente algoritmo describe cómo implementar el procedimiento.

**Algoritmo 2.4.2. (Potencia Empírica):**

1.  $j = 1$ .
2. Sea  $\{x_{1,j}, \dots, x_{n,j}\}$  una muestra de  $F$  y obtenemos  $u_n$ ,  $\hat{\xi}_n^{Mv}(u_n)$  y  $\hat{\sigma}_n^{Mv}(u_n)$ .
3. Construimos  $\hat{Q}_n$  and  $\hat{L}_n(x, \hat{Q}_n)$  como en Algoritmo 2.4.1.
4. Generar  $\{x'_1, \dots, x'_n\}$  de  $F_1$ .
5. Calcular  $\hat{t}_n(x'_n; \hat{\xi}_n^{Mv}(u_n), \hat{\sigma}_n^{Mv}(u_n))$  si  $F_1 \neq F$ . Sino  $\hat{t}_n(x'_n; \hat{\xi}_n^{*Mv}(u_n), \hat{\sigma}_n^{*Mv}(u_n))$  con  $\hat{\xi}_n^{*Mv}(u_n), \hat{\sigma}_n^{*Mv}(u_n)$  derivada de  $\{x'_n\}$ .
6. Calcular el  $p$ -valor  $p$  como en (2.42).
7. 
$$\delta_j = \begin{cases} 1 & \text{si } p < \alpha \\ 0 & \text{en otro caso.} \end{cases}$$
8.  $j++$ . Repetir mientras  $j \leq m$ .
9.  $\hat{\alpha} = \frac{1}{m} \sum_{j=1}^m \delta_j$ .

Cuando  $n \rightarrow \infty$ , la estimación  $\hat{\alpha}$  se acerca al *tamaño* del contraste si el umbral  $u_n$  realmente define los extremos de  $F$  para un *tamaño*  $n$  de muestra dado. Por otro lado si la función de distribución condicionada definida por el umbral no es  $GPD_{\xi, \sigma(u_n)}$ , o la secuencia de los datos no proviene de  $F$  la estimación  $\hat{\alpha}$  tiende a uno.

La siguiente tabla, Tabla 2.8.3 describe los resultados de simulación de la potencia empírica para una familia de distribuciones  $t$ -student con el umbral  $u_n$  obtenido mediante nuestro método de Pickands ponderado.

La Tabla 2.8.3 indica dos claros resultados. Primero, el hecho de que la diagonal esté muy cerca del *tamaño* nominal revela que nuestro procedimiento funciona muy bien a la hora de capturar los extremos de secuencias de *tamaño*  $n$  generadas por  $F_0$ . Segundo, los candidatos a valores extremos que vienen de una distribución  $F_1 \neq F_0$  se rechazan como valores extremos de  $F_0$ . Una consecuencia de esta tabla es que nuestro contraste se puede

considerar un contraste de bondad de ajuste vía las colas de la distribución. En principio nuestro contraste es más sensible que el estándar de Kolmogorov-Smirnov para detectar desviaciones en las colas de la distribución (ver Mason y Schuenemeyer (1983)).

Otra alternativa para seleccionar el umbral es fijar un estadístico de orden. En este caso el conjunto de valores extremos se define por un número fijo de observaciones dado el tamaño de muestra  $n$ .

El mensaje de la Tabla 2.8.4 es claro: Estas selecciones ad-hoc del conjunto de valores extremos pueden ser válidas para secuencias particulares de  $F$ ; pero en general se rechazan como cotas inferiores que definen los extremos de cualquier secuencia generada por  $F$  con la misma longitud  $n$ .

## 2.5. Aplicación: Estimación del VaR en índices Financieros

Una aplicación importante de la aproximación semi-paramétrica  $\hat{Q}_n$  de  $F$  es la estimación de cuantiles en la región de las colas, donde normalmente hay falta de observaciones porque reflejan sucesos extremos. Esta cuestión se está volviendo de importancia vital en una amplia gama de campos de investigación, *ej.* finanzas, climatología ó hidrología.

El objetivo de esta sección es conocer en mayor profundidad la gestión del riesgo de mercados financieros para índices bursátiles de las mayores bolsas de todo el mundo. La gestión del riesgo financiero está relacionada a la probabilidad de ocurrencia de sucesos extremos, esto es, rendimientos muy grandes (positivos ó negativos). Nos centramos en este apartado en una medida particular de este riesgo de mercado: el Valor en Riesgo (VaR), es decir, la cantidad de dinero necesaria para hacer frente a las pérdidas que puedan ocurrir con una probabilidad  $p$  sobre un cierto periodo. No es nuestra intención entrar en los detalles de la metodología *VaR*, sólo perseguimos presentar algunos resultados sobre la estimación del tail index (el comportamiento en las colas) y un cálculo sencillo del *VaR* bajo supuestos *iid* para datos financieros. Está claro que estos supuestos no son realistas y deberíamos ir

más allá teniendo en cuenta modelos de heteroscedasticidad condicionada en la volatilidad; pero esto lo abordaremos en futuros trabajos.

Los usuarios financieros generalmente calculan VaRs de dos formas diferentes: (i) de una forma completamente paramétrica, donde se supone cierta forma explícita para la distribución de los datos (normal, t-student, etc.) y (ii) de una forma completamente no paramétrica, donde el principal componente es la distribución empírica  $F_n$ . Nuestro enfoque se puede considerar en el medio de ambos, porque usamos una aproximación semi-paramétrica  $\hat{Q}_n$ .

La distribución inversa de  $\hat{Q}_n$  nos da una estimación consistente del VaR para la función de distribución  $F$ . En este caso,

$$\widehat{\text{VaR}}_p = \begin{cases} \text{inf}\{x|F_n(x) \geq 1-p\}, & 1-p \leq F_n(u_n) \\ u_n + \frac{\hat{\sigma}_n^{Mv}(u_n)}{\hat{\xi}_n^{Mv}(u_n)} \left( \left( \frac{p}{1-F_n(u_n)} \right)^{-\hat{\xi}_n^{Mv}(u_n)} - 1 \right), & 1-p > F_n(u_n) \end{cases}. \quad (2.43)$$

Cuando la función de distribución varía regularmente en  $x_F$ , ( $\xi > 0$ ), la cola de  $\hat{Q}_n$  se modeliza como una distribución de Pareto y la inversa de  $F$  se estima consistentemente mediante

$$\widehat{\text{VaR}}_p = \begin{cases} \text{inf}\{x|F_n(x) \geq 1-p\}, & 1-p \leq F_n(u_n) \\ u_n \left( \frac{1-F_n(u_n)}{p} \right)^{\hat{\xi}_n^{Mv}(u_n)}, & 1-p > F_n(u_n) \end{cases}. \quad (2.44)$$

La incertidumbre en las estimaciones se puede medir mediante intervalos de confianza bootstrap, ya que la función de distribución exacta de  $V_n = \sqrt{n}(\widehat{\text{VaR}}_p - \text{VaR}_p)$  no es conocida y su distribución asintótica depende de parámetros ruidosos. Sea  $J_n(x, \hat{Q}_n)$  la aproximación bootstrap de la distribución exacta de  $V_n$ . Un intervalo de confianza para  $\text{VaR}_p$ , a un nivel de significación  $\alpha$ , es por lo tanto dado por la siguiente expresión

$$I.C_\alpha(\text{VaR}_p) = [\widehat{\text{VaR}}_p - \frac{1}{\sqrt{n}} J_n^{-1}(1 - \frac{\alpha}{2}, \hat{Q}_n), \widehat{\text{VaR}}_p + \frac{1}{\sqrt{n}} J_n^{-1}(\frac{\alpha}{2}, \hat{Q}_n)], \quad (2.45)$$

donde  $J_n^{-1}(1 - \alpha, \hat{Q}_n)$  es el cuantil bootstrap  $1 - \alpha$ .

### 2.5.1. Características de los datos

Los datos que utilizamos para ilustrar cómo la metodología propuesta en este trabajo se puede aplicar a series financieras consiste de cinco series de índices bursátiles de los mayores mercados financieros mundiales durante el periodo 19/12/1994 – 20/04/2001: Frankfurt (DaX), Londres (Ftse-100), Madrid (Ibex), Tokio (Nikkei) y Nueva York (Dow-Jones). Estos datos han sido recogidos de la página web *www.freelunch.com*. Las observaciones consideradas para el análisis son los rendimientos logarítmicos medidos en tanto por ciento y denotados como  $r_t$ :

$$r_t = 100 (\log P_t - \log P_{t-1}),$$

con  $P_t$  los precios originales en tiempo  $t$ . Para facilidad de cálculo las observaciones negativas (pérdidas) se representan en la cola positiva.

Una primera mirada al estadístico estándar para la curtosis muestra que la mayoría de estas series son leptocúrticas. Por ejemplo el índice Dax tiene un coeficiente corregido de curtosis de 5,70; Ftse: 1,34; Ibex: 3,88; Nikkei: 2,77, y el Dow-Jones tiene un coeficiente de 3,25. Tradicionalmente esta medida se ha considerado como un indicador de las colas pesadas. No obstante el coeficiente de curtosis no da información adecuada sobre la fuente de la pesadez en la distribución. El tail index, sin embargo, proporciona este tipo de información ya que se centra en el ratio de decaimiento de una cola en particular. Por ejemplo,  $\xi > 0$  corresponde a las distribuciones donde la cola tiene un decaimiento polinomial (un análisis más detallado se puede encontrar en Shiryaev (2001)).

La Tabla 2.8.5 presenta los intervalos de confianza bootstrap no paramétricos para el tail index (ver (2.25)) obtenidos mediante las diferentes técnicas investigadas a lo largo de este trabajo.

De esta tabla parece claro que el tail index  $\xi$  es mayor que cero, indicando la existencia de una cola pesada en la cola derecha (correspondiente a las pérdidas). La única excepción es el índice *Ftse*, donde hay dudas razonables sobre su ratio de decaimiento. Por esta razón en la siguiente tabla el *VaR* se calcula bajo ambas metodologías, *GPD* y *PD*.

En la Tabla 2.8.6 presentamos estimaciones puntuales e intervalos de confianza para el  $VaR$  bajo cuatro técnicas diferentes. Las dos primeras corresponden a los métodos desarrollados en este trabajo, y las dos últimas corresponden a las metodologías empíricas estándar que usaremos aquí como benchmark.

De la Tabla 2.8.6 se pueden extraer tres conclusiones: (i) Comparando nuestros dos enfoques y teniendo en cuenta los resultados de la tabla previa, el método  $PD$  actúa mejor que el  $GPD$  desde un punto de vista de eficiencia, dado que las estimaciones puntuales son muy similares. Este es el resultado bajo la presencia de colas pesadas; (ii) el enfoque basado en la distribución empírica es menos eficiente comparado con el método  $PD$ . La razón principal para esto es la falta de observaciones provenientes de la cola, algo que nuestra metodología  $PD$  soluciona mediante una parametrización adecuada de la cola; y (iii) el enfoque basado en gaussianidad, como era de esperar, es muy conservativo en el sentido de que requiere menores cantidades de capital ( $VaR$  más *pequeño*).

## 2.6. Conclusiones

El riesgo y la incertidumbre no son la misma cosa (ver Granger (2002)) y por lo tanto se necesitan diferentes medidas para caracterizarlos. Hoy en día es comunmente aceptado que la varianza es capaz de capturar la incertidumbre pero no el riesgo. Para medir el riesgo, en otras palabras, para responder a la pregunta de *si las cosas van mal, cómo de mal pueden llegar a ir?* necesitamos primero definir las observaciones extremas que determinan el riesgo que existe en la secuencia de datos. Este es el principal objetivo de este trabajo, donde siguiendo la metodología de Pickands (1975) no sólo definimos formalmente el conjunto de observaciones extremas de una secuencia particular, sino que también por medio de un contraste de hipótesis definimos los valores extremos de cualquier secuencia de la misma longitud y con la misma distribución. La identificación de estas observaciones extremas nos permite estimar muy acertadamente medidas de riesgo como el Valor en Riesgo, así como hacer inferencia en diferentes parámetros de la cola que pueden ser de interés.

La transmisión del riesgo en series temporales, que implica dependencia en las observaciones más grandes se desarrolla en el siguiente capítulo.

## 2.7. Apéndice A: Demostraciones

**Demostración del teorema 2.4.1:** Sea  $\{U_n\}$  una secuencia de variables aleatorias uniformes en  $[0, 1]$  independientes e idénticamente distribuidas (*iid*) y sea  $\lambda$  un parámetro en  $0 < \lambda < 1$ . Definamos el proceso empírico  $B_n(t) = \sqrt{n}(U_{\lambda,n}(t) - U_\lambda(t))$  con  $U_{\lambda,n}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1_{\{\lambda < t_i \leq t\}}}{\frac{1}{n} \sum_{j=1}^n 1_{\{t_j > \lambda\}}}$ . Este proceso sigue una distribución binomial  $Bin(n, U_\lambda(t))$ . Por el teorema central del límite empírico (CLT),  $B_n(t)$  converge débilmente a una distribución  $N(0, U_\lambda(t)(1 - U_\lambda(t)))$ , por lo tanto las distribuciones finito dimensionales son normales para cualquier  $t \in [0, 1]$  fijado. Además el proceso es 'tight' debido a la continuidad uniforme de la función de distribución  $U$  y de  $U_\lambda(t)$ . Esto implica que  $B_n(t)$  converge débilmente a un proceso gaussiano de media cero  $Z_{U_\lambda}(t)$ . Sólo queda encontrar la función de covarianzas asintótica,

$$Cov(B_n(s), B_n(t)) = Cov[\sqrt{n}(U_{\lambda,n}(s) - U_\lambda(s)), \sqrt{n}(U_{\lambda,n}(t) - U_\lambda(t))],$$

con  $0 < s, t < 1$ . Como  $U_\lambda(t)$  es constante dado  $t \in (0, 1)$  la función de covarianzas queda así

$$Cov(B_n(s), B_n(t)) = \frac{n}{(1 - U_n(\lambda))^2} Cov\left(\frac{1}{n} \sum_{i=1}^n 1_{\{\lambda < t_i \leq s\}}, \frac{1}{n} \sum_{i=1}^n 1_{\{\lambda < t_i \leq t\}}\right).$$

Las observaciones  $\{t_1, \dots, t_n\}$  son *iid*, y por lo tanto  $Cov(1_{\{\lambda < t_i \leq s\}}, 1_{\{\lambda < t_j \leq t\}}) = 0$  con  $i \neq j$ .

La función de covarianzas es en este caso

$$\begin{aligned} Cov(B_n(s), B_n(t)) &= \frac{1}{(1 - U_n(\lambda))^2} Cov(1_{\{\lambda < t_i \leq s\}}, 1_{\{\lambda < t_i \leq t\}}) = \\ &= \frac{1}{(1 - U_n(\lambda))^2} [E(1_{\{\lambda < t_i \leq \min(s,t)\}}) - E(1_{\{\lambda < t_i \leq s\}})E(1_{\{\lambda < t_i \leq t\}})] = \\ &= \frac{(U(\min(s, t)) - U(\lambda)) - (U(s) - U(\lambda))(U(t) - U(\lambda))}{(1 - U_n(\lambda))^2}, \end{aligned} \tag{2.46}$$



con  $0 < s, t < 1$ . Por lo tanto  $B_n(t)$  converge débilmente a la distribución de un proceso gaussiano de media cero  $Z_{U_\lambda}(t)$  con función de covarianzas dada por

$$Cov(Z_{U_\lambda}(s), Z_{U_\lambda}(t)) = \frac{(\min(s, t) - \lambda) - (s - \lambda)(t - \lambda)}{(1 - \lambda)^2}.$$

Sea  $F$  continua y estrictamente creciente y definamos  $u_n = F^{-1}(\lambda)$ . Construyamos  $x_1, \dots, x_n$  iid de  $F$  vía  $x_i = F^{-1}(t_i)$  y utilizemos  $F_n(x)$  para denotar la función de distribución empírica basada en  $x_1, \dots, x_n$ . Por el teorema de la continuidad funcional (“continuous mapping theorem”)  $\sum_{i=1}^n 1_{\{u_n < x_i \leq x\}} = \sum_{i=1}^n 1_{\{F(u_n) < F(x_i) \leq F(x)\}}$  y por lo tanto  $F_{u_n, n}(x) = U_{\lambda, n}(t)$ . Entonces  $B_n(t)$  se puede escribir como  $\sqrt{n}(F_{u_n, n}(y) - F_{u_n}(y))$  con  $y = x - u_n$  (ver (2.19) y (2.20)) y la función de covarianzas es

$$Cov(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{(F(\min(y_1, y_2)) - F(u_n)) - (F(y_1) - F(u_n))(F(y_2) - F(u_n))}{(1 - F(u_n))^2},$$

con  $y_1 = F^{-1}(s)$  y  $y_2 = F^{-1}(t)$ .

Bajo la hipótesis nula, el proceso empírico  $S_n(y)$  se puede escribir como  $\sqrt{n}(F_{u_n, n}(y) - GPD_{\xi, \sigma(u_n)}(y))$  y la función de covarianzas del proceso límite se puede escribir como

$$Cov(Z_{F_{u_n}}(y_1), Z_{F_{u_n}}(y_2)) = \frac{GPD_{\xi, \sigma}(\min(y_1, y_2))}{1 - F(u_n)} - GPD_{\xi, \sigma}(y_1)GPD_{\xi, \sigma}(y_2).$$

**Demostración del lema 2.4.1:** Sea  $0 < \alpha < 1$  el nivel de significación del contraste y consideremos  $L(x, F)$  continua y estrictamente creciente. Por definición

$$P\{T_n > L^{-1}(1 - \alpha, F)\} = \alpha,$$

con  $L^{-1}(1 - \alpha, F)$  el cuantil asintótico  $1 - \alpha$ .

Consideremos  $L_n(x, Q_n)$  la aproximación bootstrap de  $L_n(x; F)$  y  $L_n^{-1}(1 - \alpha, Q_n)$  su cuantil  $1 - \alpha$ . Por lo tanto si  $\sup_{x \in \mathbb{R}} |Q_n(x) - F(x)| \xrightarrow{p} 0$  entonces  $L_n^{-1}(1 - \alpha, Q_n) \rightarrow L^{-1}(1 -$

$\alpha, F)$  con probabilidad uno y por el teorema de Slutsky

$$P\{T_n > L_n^{-1}(1 - \alpha, Q_n)\} \rightarrow P\{T_n > L^{-1}(1 - \alpha, F)\} = \alpha.$$

## 2.8. Apéndice B: Tablas

	$t_1(\xi \sim 1)$	$t_5(\xi \sim 0,2)$	$t_{10}(\xi \sim 0,1)$	$t_{30}(\xi \sim 0)$
$\hat{\xi}_n^{Mv}(u_n)$	[0,70, 1,69]	[-0,17, 0,24]	[-0,28, 0,39]	[-0,43, 0,68]
$\hat{\xi}_n^{Pi}(u_n^{Pi})$	[0,29, 1,06]	[-0,39, 0,08]	[-0,63, -0,06]	[-0,64, -0,17]
$\hat{\xi}_n^{Mv}(x_{(\frac{95}{100}n)})$	[0,34, 1,75]	[0,19, 0,91]	[-0,26, 0,33]	[-0,28, 0,57]

**Tabla 2.8.1.** Intervalos de confianza bootstrap a un nivel de significación  $\alpha = 0,05$  para diferentes estimadores del tail index:  $\hat{\xi}_n^{Mv}(u_n)$  con  $u_n$  estimado por  $d^{Wp}$  y por  $x_{(\frac{95}{100}n)}$ ; y  $\hat{\xi}_n^{Pi}(u_n^{Pi})$  con  $u_n$  estimado por  $d^\infty$ .  $B = 1000$  muestras bootstrap de tamaño  $n = 1000$  se extraen de una secuencia generada de  $t_\nu$ , con  $\nu = 1, 5, 10$  y  $30$ .

	$t_1(\xi \sim 1)$	$t_5(\xi \sim 0,2)$	$t_{10}(\xi \sim 0,1)$	$t_{30}(\xi \sim 0)$
$\hat{\xi}_n^{Mv}(u_n)$	[0,70, 1,69]	[-0,17, 0,24]	[-0,28, 0,39]	[-0,43, 0,68]
$\hat{\xi}_n^{Hi}(u_n)$	[0,82, 1,23]	[0,08, 0,37]	[-0,42, 0,23]	[0,04, 0,20]

**Tabla 2.8.2.** Intervalos de confianza bootstrap a un nivel de significación  $\alpha = 0,05$  para diferentes estimadores del tail index cuando  $u_n$  se obtiene de la distribución  $GPD_{\xi, \sigma(u_n)}$  y de la  $PD_\xi$  respectivamente. Notar que  $\hat{\xi}_n^{Mv}(u_n)$  es  $\hat{\xi}_n^{Hi}(u_n)$  para el caso  $PD_\xi$ .  $B = 1000$  muestras bootstrap de tamaño  $n = 1000$  se extraen de una secuencia generada de  $t_\nu$ , con  $\nu = 1, 5, 10$  y  $30$ .

$n = 1000$	$F_1$			
$F_0$	$t_{30}(\xi \sim 0)$	$t_{10}(\xi \sim 0,1)$	$t_5(\xi \sim 0,2)$	$t_1(\xi \sim 1)$
$t_{30}$	0,06	0,63	0,91	0,97
$t_{10}$	0,59	0,08	0,79	0,98
$t_5$	0,95	0,72	0,06	0,99
$t_1$	0,94	0,94	0,94	0,05

**Tabla 2.8.3.** *Potencia empírica de  $T_n$  para una familia de funciones de distribución  $t$ -student con  $u_n$  calculado por medio de  $d^{Wp}$ .  $F_0$  denota el proceso generador de los datos y  $F_1$  la función de distribución bajo la hipótesis alternativa.  $B = 1000$  replicaciones bootstrap se obtienen,  $m = 500$  simulaciones Monte-Carlo. El nivel de significación es  $\alpha = 0,05$ .*

$F_0$	$x_{(700)}$	$x_{(800)}$	$x_{(900)}$	$x_{(950)}$
$t_{30}$	0,49	0,48	0,43	0,44
$t_{10}$	0,48	0,48	0,46	0,46
$t_5$	0,54	0,50	0,48	0,47
$t_1$	0,64	0,58	0,52	0,48

**Tabla 2.8.4.** *Potencia empírica de  $T_n$  para una familia de funciones de distribución  $t$ -student con diferentes elecciones ad-hoc de la secuencia umbral  $u_n$  para un tamaño de muestra  $n = 1000$ .  $B = 1000$  replicaciones bootstrap se obtienen,  $m = 500$  simulaciones Monte-Carlo. El nivel de significación es  $\alpha = 0,05$ .*

	$\hat{\xi}_n^{Mv}(u_n)$	$\hat{\xi}_n^{Hi}(u_n)$	$\hat{\xi}_n^{Pi}(u_n^{Pi})$	$\hat{\xi}_n^{Mv}(x_{(\frac{95}{100}n)})$
<i>Dax</i>	[-0,02; 0,24; 0,84]	[0,30; 0,31; 0,36]	[-0,50; -0,37; -0,20]	[-0,13; 0,22; 0,65]
<i>Ftse</i>	[-0,57; -0,26; 0,04]	[0,07; 0,11; 0,12]	[-0,44; -0,28; -0,08]	[-0,54; -0,29; 0,13]
<i>Ibex</i>	[-0,12; 0,28; 0,87]	[0,32; 0,37; 0,38]	[-0,43; -0,21; -0,04]	[-0,04; 0,46; 0,90]
<i>Nikkei</i>	[-0,13; 0,11; 0,55]	[0,33; 0,34; 0,39]	[-0,34; -0,19; -0,03]	[-0,25; 0,07; 0,50]
<i>Dow-Jones</i>	[-0,11; 0,63; 1,52]	[0,33; 0,41; 0,44]	[-0,24; -0,22; -0,03]	[0,05; 0,76; 1,72]

**Tabla 2.8.5.** Intervalos de confianza bootstrap ( $\alpha = 0,05$ ) y estimaciones puntuales del tail index  $\xi$  para rendimientos de acciones durante el periodo 19/12/1994 – 20/04/2001.

Muestras bootstrap  $B = 1000$ .

VaR	GPD	PD	$F_n$	Gaussiana
<i>Dax</i>	[3,57; 4,16; 7,83]	[3,48; 4,25; 4,93]	[2,96; 4,33; 5,04]	[3,52; 3,62; 3,71]
<i>Ftse</i>	[2,81; 3,04; 3,40]	[2,83; 3,05; 3,31]	[2,83; 3,08; 3,32]	[2,65; 2,78; 2,85]
<i>Ibex</i>	[3,25; 3,92; 4,69]	[2,94; 3,91; 4,62]	[3,02; 4,50; 5,80]	[3,08; 3,19; 3,32]
<i>Nikkei</i>	[3,69; 4,24; 8,30]	[3,33; 4,31; 5,00]	[4,09; 4,73; 5,95]	[3,75; 3,79; 3,83]
<i>Dow-Jones</i>	[1,47; 2,09; 2,60]	[1,56; 2,09; 2,49]	[1,36; 1,90; 2,15]	[1,55; 1,73; 1,97]

**Tabla 2.8.6.** Intervalos de confianza ( $\alpha = 0,05$ ) y estimaciones puntuales del VaR para los diferentes índices financieros calculados con diferentes metodologías: nuestro enfoque GPD y PD, un enfoque no paramétrico  $F_n$ , y un enfoque paramétrico basado en supuestos gaussianos. El VaR indica el porcentaje de pérdidas que puede ocurrir con un  $p = 0,01$  de probabilidad y durante un periodo de 1 día. Los datos cubren el periodo 19/12/1994 – 20/04/2001. Muestras bootstrap  $B = 1000$ .

## Capítulo 3

# Dependencia en los Extremos: Un Canal para la transmisión del Riesgo

El “extremal index” es el parámetro clave para extender la teoría de valores extremos para variables aleatorias *iid* a procesos estacionarios, reflejando el nivel de dependencia existente en las observaciones más altas definidas por una secuencia umbral  $\{u_n\}$ . Este capítulo introduce un estimador para este parámetro como el ratio del número de elementos de dos procesos puntuales definidos por una partición de la muestra en diferentes bloques, y por la secuencia de los máximos definidos sobre cada bloque y que exceden los umbrales correspondientes  $\{v_n\}$  y  $\{u_n\}$  con  $v_n > u_n$ . El estimador es asintóticamente insesgado bajo condiciones muy generales sobre  $\{u_n\}$ , consistente (la varianza converge a 0 cuando  $n \rightarrow \infty$ ) y el teorema central del límite se puede aplicar. Por lo tanto admite un contraste de hipótesis para el extremal index y por lo tanto para contrastar la existencia de agrupamiento (de ahora en adelante “clustering”) en los valores extremos. Otras ventajas de este método son que permite cierta libertad para elegir  $\{u_n\}$ , y no es muy sensible a la elección de la partición. El análisis del clustering de las observaciones más grandes en el mercado financiero de Frankfurt (índice DaX) arroja cierta luz acerca de los patrones de dependencia en las series financieras. La transmisión del riesgo se puede deber a la dependencia encontrada en las observaciones más grandes más que a una estructura en la

dinámica del proceso de volatilidad.

### 3.1. Introducción

Supongamos una muestra aleatoria de una función de distribución desconocida  $F$ , y sea  $G$  la distribución límite del máximo muestral  $M_n$ . La teoría clásica de valores extremos muestra que bajo ciertas condiciones de regularidad sobre la cola de  $F$  y para ciertas constantes  $a_n > 0$ ,  $b_n$ ,

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \rightarrow G(x), \quad (3.1)$$

donde  $G$  debe ser de uno de los siguientes tipos (ver de Haan (1976)),

Tipo I: (Gumbel)  $G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty.$

Tipo II: (Fréchet)  $G(x) = \begin{cases} 0 & x \leq 0, \\ e^{-x^{-\frac{1}{\xi}}} & x > 0, \xi > 0. \end{cases}$

Tipo III: (Weibull)  $G(x) = \begin{cases} 1 & x \geq 0, \\ e^{-(-x)^{-\frac{1}{\xi}}} & x < 0, \xi < 0. \end{cases}$

Este importante resultado se puede extender para estudiar el máximo de una amplia clase de procesos dependientes. Nos centramos en este capítulo en secuencias estacionarias donde la dependencia está restringida por diferentes condiciones distribucionales *mixing*. Distinguimos dos tipos de dependencia: dependencia a largo plazo y dependencia a corto plazo. Para limitar el primer tipo de dependencia suponemos una variación de la condición distribucional *mixing*  $D(u_n)$  de Leadbetter et al. (1983). La condición *mixing* de Leadbetter se dice que se verifica para una secuencia umbral  $\{u_n\}$  si para cualquier conjunto de enteros  $1 \leq i_1 < \dots < i_p < j_1 < \dots < j_{p'} \leq n$  para los cuales  $j_1 - i_p \geq l$ , se tiene

$$D(u_n) : |F_{i_1, \dots, i_p, j_1, \dots, j_{p'}}(u_n) - F_{i_1, \dots, i_p}(u_n)F_{j_1, \dots, j_{p'}}(u_n)| \leq \alpha_{n,l},$$

donde  $\alpha_{n,l_n} \rightarrow 0$  cuando  $n \rightarrow \infty$  para algún  $l_n = o(n)$ , y  $F_{i_1, \dots, i_p}(u_n)$  denota  $P\{X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n\}$ . Sea  $D'(u_n)$  la condición alternativa *mixing* que usaremos a lo largo del capítulo. Esta condición es como sigue,

$$D'(u_n) : |P \{X_{i_1} > u_n \circ \dots \circ X_{i_p} > u_n \circ X_{j_1} > u_n \circ \dots \circ X_{j_{p'}} > u_n\} - P \{X_{i_1} > u_n \circ \dots \circ X_{i_p} > u_n\} P \{X_{j_1} > u_n \circ \dots \circ X_{j_{p'}} > u_n\}| \leq \alpha_{n,l}. \quad (3.2)$$

Notemos que estas condiciones sólo implican condiciones del tipo  $\{X_i > u_n\}$  en contraste con condiciones mixing más restrictivas como por ejemplo la condición “strong mixing” introducida en Rosenblatt (1956). Estas condiciones mixing por sí solas son suficientes para extender el resultado central dado en (3.1) a secuencias estacionarias para constantes apropiadas no necesariamente las obtenidas del contexto *iid*. En particular estas constantes  $a_n > 0$ ,  $b_n$  y la distribución extrema  $G$  son las mismas del caso *iid* bajo una condición  $D''(u_n)$  restringiendo la dependencia entre observaciones contiguas, Leadbetter (1983), que evita la presencia de clustering,

$$D''(u_n) : \limsup_{n \rightarrow \infty} n \sum_{j=2}^{[n/k_n]} P\{X_1 > u_n, X_j > u_n\} \rightarrow 0 \quad \text{cuando } k_n \rightarrow \infty, \quad (3.3)$$

con  $k_n$  una secuencia que define una partición de la muestra. De otro modo, para una secuencia estacionaria  $\{X_n\}$  que verifique sólo  $D'(u_n)$  con  $u_n = a_n x + b_n$ , típicamente tenemos

$$P\{a_n^{-1}(M_n - b_n) \leq x\} \rightarrow G^\theta(x), \quad (3.4)$$

donde  $\theta$  es el parámetro clave para extender la teoría de valores extremos para variables aleatorias *iid* a secuencias estacionarias. Este concepto, originado en los artículos de Loynes (1965), O'Brien (1974) y desarrollado en detalle en Leadbetter (1983), refleja el efecto del clustering de las observaciones que exceden  $u_n$  en la distribución límite del máximo.

Hay diferentes interpretaciones del extremal index  $\theta$ , concerniendo diferentes características del clustering de las observaciones más grandes. Loynes (1965) bajo condiciones mixing ligeramente diferentes encontró que

$$P\{M_n \leq u_n\} = F^{n\theta}(u_n). \quad (3.5)$$



O'Brien (1987) mostró que

$$P\{M_{2,r_n} \leq u_n | X_1 > u_n\} \rightarrow \theta, \quad (3.6)$$

donde  $M_{2,r_n}$  es el máximo de  $\{X_2, \dots, X_{r_n}\}$ , y  $r_n = o(n)$  satisface ciertas condiciones de crecimiento. Notemos que de esta definición del extremal index es inmediato ver que  $0 \leq \theta \leq 1$ . Alternativamente Leadbetter (1983) mostró que el inverso del extremal index es el número límite medio de los valores que exceden  $u_n$  en un intervalo de longitud  $r_n$ , *i.e.*

$$E \left[ \sum_{j=1}^{r_n} I(X_j > u_n) \middle| \sum_{j=1}^{r_n} I(X_j > u_n) \geq 1 \right] \rightarrow \theta^{-1}, \quad (3.7)$$

con  $I(X > 0)$  la función indicadora. Por estacionariedad esto se llama el *tamaño* límite medio del cluster del proceso. Finalmente, Hsing (1993) y Ferro y Segers (2003) se aprovechan de la distribución límite del máximo,

$$P\{M_n \leq u_n\} \rightarrow e^{-\theta\tau}, \quad (3.8)$$

con  $0 < \tau < \infty$ , de dos formas diferentes. Hsing aproxima la distribución de  $n(1 - F(M_n))$  por una distribución exponencial con media  $\theta^{-1}$ , y Ferro y Segers modelizan el proceso de los tiempos entre valores que exceden  $u_n$  mediante la misma distribución límite exponencial.

La expresión (3.8) es una transformación de (3.4) donde  $\tau$  es el exponente de una distribución de valores extremos y  $u_n = a_n x + b_n$ . De la misma forma la probabilidad límite (3.1) se puede escribir como  $P\{M_n \leq u_n\} \rightarrow e^{-\tau}$ . Tomando logaritmos en esta expresión es inmediato derivar que  $n(1 - F(u_n)) \rightarrow \tau$  para  $u_n$  suficientemente grande. Entonces para secuencias *iid*,  $B_n^{(u_n)} = \sum_{j=1}^n I(X_j > u_n)$  converge en distribución a una variable aleatoria de Poisson con media  $\tau$ .

Sin embargo para secuencias estacionarias dependientes donde  $D''(u_n)$  no se satisface  $B_n^{(u_n)}$  no converge a una variable aleatoria de Poisson (los valores que superan  $u_n$  no son mutuamente independientes), no obstante podemos definir un proceso puntual como resul-

tado de estrechar el proceso  $B_n^{(u_n)}$ . Este estrechamiento (de ahora en adelante “thinning”) define el proceso  $N_{k_n}^{(u_n)}$  compuesto por el máximo sobre  $k_n$  bloques de longitud  $r_n$  y excediendo  $u_n$ , y que converge a un proceso de Poisson  $N$  con media  $\theta\tau$ , ver Leadbetter (1983) ó Leadbetter et al. (1983). Este capítulo presenta una interpretación alternativa del extremal index como el resultado de estrechar dos veces el proceso  $B_n^{(u_n)}$ . El segundo estrechamiento de  $B_n^{(u_n)}$ , y por lo tanto estrechamiento de  $N_{k_n}^{(u_n)}$ , define otro proceso puntual  $N_{k_n}^{(v_n)}$  que converge en distribución a un proceso Poisson con intensidad  $\theta^2\tau$ . La secuencia  $\{v_n\}$  satisface  $n(1 - F(v_n)) \rightarrow \theta\tau$  y se define por  $E[\sum_{j=1}^{r_n} I(X_j > v_n) | \sum_{j=1}^{r_n} I(X_j > u_n) \geq 1] \rightarrow 1$ . Bajo ciertas condiciones débiles sobre la secuencia umbral este método ofrece un estimador consistente del extremal index que mejora la mayoría de los estimadores populares y tal que no es muy sensible a la elección del *tamaño* del bloque  $r_n$  ni a la elección del umbral  $\{u_n\}$  en contraste con el resto de candidatos que estiman  $\theta$ .

El capítulo se estructura como sigue. La sección 2 introduce una definición del extremal index como el ratio de dos procesos puntuales que se derivan de la distribución asintótica del máximo muestral. Un estimador natural para este parámetro basado en estas técnicas se introduce en el apartado 3. Este apartado también revisa algunos de los estimadores más populares que se encuentran en la literatura y sus propiedades estadísticas, en particular sesgo y varianza. Las propiedades correspondientes de nuestro estimador se estudian también con especial énfasis en el análisis del error cuadrático medio de los diferentes métodos. La selección del bloque óptimo se considera también esta sección que concluye con un contraste de hipótesis para el extremal index que es suficiente para contrastar la hipótesis de clustering en los extremos. Un experimento de simulación para diferentes ejemplos presentados en la literatura se lleva a cabo en el apartado 4 remarcando el experimento Monte-Carlo para el error cuadrático medio. La Sección 5 presenta una aplicación a los rendimientos del índice DaX para conocer en más profundidad el clustering en los extremos de la secuencia de rendimientos y en la volatilidad del proceso. Finalmente las conclusiones se encuentran en la sección 6.

## 3.2. Definición del extremal index

Supongamos que tenemos  $n$  observaciones de una secuencia estacionaria  $\{X_i, i \geq 1\}$  con función de distribución marginal  $F$  y que satisface  $[1 - F(x)]/[1 - F(x^-)] \rightarrow 1$  cuando  $x \rightarrow \infty$ . Esta condición es suficiente para definir una secuencia  $\{u_n\}$  para cada  $0 < \tau < \infty$  tal que

$$n(1 - F(u_n)) \rightarrow \tau. \quad (3.9)$$

Consideremos de ahora en adelante que  $\{X_n\}$  satisface  $D'(u_n)$ , como se ha definido en (3.2), para cada  $\tau > 0$ . Intuitivamente esta condición da una medida del grado de dependencia existente en el proceso y permite la construcción de bloques *casi* independientes por medio de la definición de secuencias  $\{k_n\}$ ,  $\{r_n\}$  con  $k_n \rightarrow \infty$ ,  $k_n = o(n)$  y  $k_n l_n = o(n)$ , mientras  $r_n$  es la parte entera de  $n/k_n$ . La interpretación de estas secuencias es:  $k_n$  es el número de bloques de la secuencia de longitud  $n$ , y  $r_n$  el tamaño de cada bloque.

Bajo estos supuestos, si  $P\{M_n \leq u_n\}$  converge para algún  $\tau > 0$  entonces

$$P\{M_n \leq u_n\} \rightarrow e^{-\theta\tau}, \quad (3.10)$$

para todo  $\tau > 0$ , con  $0 \leq \theta \leq 1$  (ver teorema 3,7,1. de Leadbetter et al. (1983) para una demostración detallada). El parámetro  $\theta$  se llama extremal index de la secuencia  $\{X_n\}$  y es el parámetro clave para extender la teoría de valores extremos de variables aleatorias *iid* a procesos estacionarios.

Consideremos  $\{k_n\}$ ,  $\{r_n\}$  que definen una partición apropiada de la secuencia  $\{X_n\}$ , entonces una condición suficiente para la existencia del extremal index es

$$k_n(1 - F_{1,\dots,r_n}(u_n)) \rightarrow \theta\tau. \quad (3.11)$$

Este resultado es inmediato por la aproximación de  $P\{M_n \leq u_n\}$  por  $P^{k_n}\{M_{r_n} \leq u_n\}$  para elecciones adecuadas de  $k_n$  y  $r_n$ , y (3.10) y la expansión polinomial de la función exponencial. La otra implicación de este resultado también es cierta, *i.e.* una secuencia

estacionaria con extremal index  $\theta$  verifica (3.11) para cada  $\tau > 0$ . La demostración se obtiene tomando logaritmos en la expresión  $P^{k_n}\{M_{r_n} \leq u_n\}$  que aproxima  $e^{-\theta\tau}$ .

Consideremos el número de valores que exceden  $u_n$  dentro de un bloque de tamaño  $r_n$ . Este suceso define una secuencia de variables aleatorias  $B_{r_n}^{(u_n)} = \sum_{j=1}^{r_n} I(X_j > u_n)$  for  $r_n \rightarrow \infty$ , y  $r_n = o(n)$  cuyo valor esperado, debido a la estacionariedad del proceso, converge al tamaño medio del cluster de los valores que exceden  $u_n$  en la secuencia  $\{X_n\}$ , esto es, al inverso del extremal index,

$$E [B_{r_n}^{(u_n)} | B_{r_n}^{(u_n)} \geq 1] \rightarrow \theta^{-1}.$$

Esto se prueba inmediatamente ya que  $E [B_{r_n}^{(u_n)} | B_{r_n}^{(u_n)} \geq 1] = \sum_{j=1}^{\infty} j P \{B_{r_n}^{(u_n)} = j | B_{r_n}^{(u_n)} \geq 1\} = \frac{r_n P\{X_j > u_n\}}{P\{\bigcup_{j=1}^{r_n} (X_j > u_n)\}}$ , y por tanto  $E [B_{r_n}^{(u_n)} | B_{r_n}^{(u_n)} \geq 1] = \frac{r_n(1-F(u_n))}{1-F_{1,\dots,r_n}(u_n)} \rightarrow \theta^{-1}$ , si (3.11) se verifica.

El mismo argumento se puede aplicar al proceso  $B_{r_n}^{(v_n)}$  con  $v_n \geq u_n$  que verifica

$$E [B_{r_n}^{(v_n)} | B_{r_n}^{(v_n)} \geq 1] \rightarrow 1. \quad (3.12)$$

Es interesante notar que la secuencia  $\{v_n\}$  satisface la condición  $D'(v_n)$  ya que  $v_n \geq u_n$  y

$$n(1 - F(v_n)) \rightarrow \theta\tau \quad \text{cuando } n \rightarrow \infty. \quad (3.13)$$

Además, por la estructura de dependencia (ver (3.9) y (3.10)) tenemos  $P\{M_n \leq v_n\} \rightarrow e^{-\theta^2\tau}$ . Entonces es inmediato ahora ver que (3.11) se cumple para la secuencia  $\{v_n\}$  por

$$k_n(1 - F_{1,\dots,r_n}(v_n)) \rightarrow \theta^2\tau. \quad (3.14)$$

El suceso  $\{X_i > u_n\}$  y las secuencias  $\{k_n\}$ ,  $\{r_n\}$  dividen la secuencia  $\{X_n\}$ , con extremal index  $\theta$ , en aproximadamente grupos independientes de valores que exceden  $u_n$  donde  $M_{(j-1)r_n+1, jr_n}$  es el máximo sobre los bloques para  $j = 1, \dots, k_n$ . Está claro que la secuencia  $\{M_{(j-1)r_n+1, jr_n}\}$  es aproximadamente independiente cuando  $n$  crece si  $D'(u_n)$  se verifica para  $\{X_n\}$ .

Consideremos los puntos  $j$  como puntos en el tiempo y definamos para cada  $n$ , y  $k_n$ ,

un proceso  $\eta_{k_n}(j/k_n) = M_{(j-1)r_n+1, jr_n}$ . La escala del tiempo se normaliza  $t = j/k_n$  en el intervalo unidad  $(0, 1]$ . Entonces los valores que exceden  $u_n$  en el proceso  $\eta_{k_n}(t)$  definen un proceso puntual  $N_{k_n}^{(u_n)}$  en el intervalo unidad (ver Kallenberg (1976) para la teoría de procesos puntuales). Más aún, el proceso puntual  $N_{k_n}^{(u_n)}$  converge en distribución a un proceso de Poisson  $N$  en  $(0, 1]$  con parámetro de intensidad  $\theta\tau$ . Para probar este resultado sólo es necesario mostrar que  $E[N_{k_n}^{(u_n)}(a, b)] \rightarrow E[N(a, b)]$  para  $0 < a < b \leq 1$  y  $P\{N_{k_n}^{(u_n)}(A) = 0\} \rightarrow P\{N(A) = 0\}$  para cada unión disjunta  $A$  de conjuntos  $(a_i, b_i) \subset (0, 1]$ . La demostración es análoga a la correspondiente encontrada en el teorema 4.1. en Leadbetter (1983).

Es interesante ver que el mismo argumento se puede aplicar para construir estrechamientos de  $N_{k_n}^{(u_n)}$  mediante una secuencia  $\{v_n\}$  que satisfaga (3.12). Esta secuencia define el proceso puntual  $N_{k_n}^{(v_n)}$  en el intervalo unidad que converge a un proceso de Poisson con intensidad  $\theta^2\tau$ . La demostración es idéntica al caso  $N_{k_n}^{(u_n)}$  ya que (3.14) y  $D'(v_n)$  se verifican con  $v_n \geq u_n$ .

Estos resultados nos dan el marco para definir el extremal index como el ratio del valor esperado límite de los procesos puntuales  $N_{k_n}^{(u_n)}$  y  $N_{k_n}^{(v_n)}$ ,

$$\theta = \lim_{n \rightarrow \infty} \frac{E[N_{k_n}^{(v_n)}]}{E[N_{k_n}^{(u_n)}]}. \quad (3.15)$$

El extremal index se puede interpretar también como la probabilidad condicionada definida por  $u_n$ . De los resultados dados en (3.9) and (3.13) se deriva que

$$\theta = 1 - \lim_{n \rightarrow \infty} F_{u_n}(v_n), \quad (3.16)$$

con  $F_{u_n}(v_n) = \frac{F(v_n) - F(u_n)}{1 - F(u_n)}$ . Está claro que cuando la dependencia en los extremos (valores que exceden  $u_n$ ) de la secuencia estacionaria decrece,  $v_n$  se aproxima a  $u_n$  y  $\theta$  se acerca a uno como ocurre para el caso *iid* o para dependencia débil ( $D'(u_n)$  y  $D''(u_n)$  se cumplen).

Estas definiciones del extremal index son también válidas para secuencias umbral donde (3.9) no se cumple pero la condición mixing en (3.2) todavía se cumple. Consideremos

$\tilde{u}_n$  tal que  $n(1 - F(\tilde{u}_n)) = \tau_n$ , con  $\tau_n \rightarrow \infty$ , y  $\tau_n = o(n)$ . Esta condición implica que  $P\{M_n \leq \tilde{u}_n\} \rightarrow 0$ .

Una condición necesaria sobre  $\tilde{u}_n$  para que el extremal index se defina de la misma forma que en (3.15) es que el ratio  $\frac{-\log P\{M_n \leq \tilde{u}_n\}}{n(1-F(\tilde{u}_n))}$  converja a una constante en  $(0, 1)$ . Si la secuencia  $\{X_n\}$  tiene extremal index  $\theta$  las condiciones (3.9) y (3.11) se verifican para ciertas secuencias  $u_n$ . Entonces, una condición suficiente para  $\tilde{u}_n$  es que

$$\frac{(1 - F(\tilde{u}_n))(1 - F_{1,\dots,r_n}(u_n))}{(1 - F(u_n))(1 - F_{1,\dots,r_n}(\tilde{u}_n))} \rightarrow 1. \quad (3.17)$$

Esta condición conlleva que  $k_n(1 - F_{1,\dots,r_n}(\tilde{u}_n)) = \tau'_n$  con  $\tau'_n \rightarrow \infty$  y  $\tau'_n/\tau_n \rightarrow \theta$ . Los mismos resultados que para  $u_n$  y  $\tau$  constante se obtienen ahora para  $\tilde{u}_n$  y  $\tau_n$ . Por lo tanto, la secuencia  $B_{r_n}(\tilde{u}_n)$  satisface que

$$E [B_{r_n}^{(\tilde{u}_n)} | B_{r_n}^{(\tilde{u}_n)} \geq 1] \rightarrow \theta^{-1},$$

y existe una secuencia  $\tilde{v}_n$  tal que  $n(1 - F(\tilde{v}_n)) = \tau'_n$ . Bajo la condición (3.17) para  $\{\tilde{v}_n\}$  en vez de  $\{\tilde{u}_n\}$  se obtiene que  $k_n(1 - F_{1,\dots,r_n}(\tilde{v}_n)) = \tau''_n$ , con  $\tau''_n \rightarrow \infty$  y  $\tau''_n/\tau'_n \rightarrow \theta$  y el extremal index se puede definir como en (3.15) para el correspondiente  $\tilde{u}_n$  y  $\tilde{v}_n$  dado que  $D'(\tilde{u}_n)$  se cumple.

Cuando tratemos con la estimación nos referiremos al número de elementos de los procesos  $N_{k_n}^{(\tilde{u}_n)}$  y  $N_{k_n}^{(\tilde{v}_n)}$  como  $Z_{\tilde{u}_n}^*$  y  $Z_{\tilde{v}_n}^*$  respectivamente, y  $Z_{\tilde{u}_n}$  y  $Z_{\tilde{v}_n}$  se utilizarán para denotar el número de valores que exceden  $\tilde{u}_n$  y  $\tilde{v}_n$  en la secuencia  $\{X_n\}$ . Notación análoga se usará para los correspondientes valores que exceden  $u_n$  y  $v_n$ . Notemos que las variables  $Z_{\tilde{u}_n}^*$  y  $Z_{u_n}^*$  se pueden interpretar como el número de bloques de la partición definida por  $\{k_n\}$ ,  $\{r_n\}$  donde hay al menos una observación que excede  $\tilde{u}_n$  y  $u_n$  respectivamente.

### 3.3. Estimación del extremal index

El extremal index representa el clustering de las observaciones más grandes determinadas por una secuencia  $\{u_n\}$  suficientemente grande como para satisfacer condiciones del tipo de (3.9). La dependencia temporal en estas observaciones afecta a la distribución del máximo de la secuencia estacionaria, esto es,  $P\{M_n \leq u_n\}$  es  $F^{n\theta}(u_n)$  en vez de  $F^n(u_n)$  para  $n$  y  $u_n$  suficientemente grande.

Este resultado nos lleva al primer estimador para el extremal index para secuencias apropiadas  $k_n, r_n$  cumpliendo que  $P^{k_n}\{M_{r_n} \leq u_n\}$  aproxima  $P\{M_n \leq u_n\}$ . Entonces, tomando logaritmos en ambas expresiones,  $\theta = \frac{\log P\{M_{r_n} \leq u_n\}}{r_n \log F(u_n)}$ . Un estimador natural para el extremal index en este caso es

$$\hat{\theta}_n^{(1)} = \frac{\log(1 - Z_{u_n}^*/k_n)}{r_n \log(1 - Z_{u_n}/n)}, \quad (3.18)$$

con la notación introducida en la última sección. El ratio  $Z_{u_n}/n$  es un estimador de  $1 - F(u_n)$ , y  $Z_{u_n}^*/k_n$  un estimador de  $1 - F_{1, \dots, r_n}(u_n)$ .

Por otro lado el concepto de extremal index introducido en Leadbetter (1983),  $\theta^{-1}$  el límite del *tamaño* medio del cluster de los valores que exceden el umbral da este estimador

$$\hat{\theta}_n^{(2)} = \frac{Z_{u_n}^*}{Z_{u_n}}. \quad (3.19)$$

Este método se denomina el método de los bloques y se puede considerar una versión simplificada de  $\hat{\theta}_n^{(1)}$ . Otro método popular es el método de las rachas ('runs estimator'), que se puede ver como el estimador del extremal index para las definiciones introducidas en O'Brien (1987) ó en Hsing (1993),

$$\bar{\theta}_n = \frac{W_{u_n}}{Z_{u_n}}, \quad (3.20)$$

donde  $W_{u_n} = \sum_{i=1}^{n-r_n} I(X_i > u_n)(1 - I(X_{i+1} > u_n)) \cdots (1 - I(X_{i+r_n} > u_n))$ .

Nuestra definición del extremal index proporciona un atractivo estimador de  $\theta$  dado por

el ratio de  $Z_{v_n}^*$  y  $Z_{u_n}^*$  o alternativamente por  $Z_{\tilde{v}_n}^*$  y  $Z_{\tilde{u}_n}^*$ . Para  $u_n$  y  $v_n$  secuencias verificando (3.9) y (3.13) nuestro estimador  $\tilde{\theta}_n$  viene dado por

$$\tilde{\theta}_n = \frac{Z_{v_n}^*}{Z_{u_n}^*}, \quad (3.21)$$

representando los correspondientes estrechamientos de la secuencia original definidos por la secuencia  $k_n$  y los umbrales  $u_n$  y  $v_n$ . El estimador, sin embargo, no está completamente especificado ya que estas secuencias no están determinadas. Por (3.9) un candidato apropiado para la secuencia umbral viene dado por los estadísticos extremos de orden (ver sección 2.5. en Leadbetter et al. (1983)). Por lo tanto un estimador adecuado para  $v_n$  viene dado por el estadístico de orden de la secuencia estacionaria  $\{X_n\}$  que verifica el correspondiente empírico (3.12), *i.e.*

$$v_n = \max_{1 \leq i \leq n} \left\{ x_i, i = 1, \dots, n \mid \frac{1}{Z_{u_n}^*} \sum_{j=1}^{k_n} B_{r_n, j}^{(x_i)} = 1 \right\}, \quad (3.22)$$

con  $B_{r_n, j}^{(u_n)} = \sum_{k=(j-1)r_n+1}^{jr_n} I(X_k > u_n)$ . Esta expresión se reduce a  $v_n = x_{(n-Z_{u_n}^*)}$ , estadístico extremo de orden, con  $x_{(1)} \leq \dots \leq x_{(n)}$  la secuencia de estadísticos de orden. Por (3.17) las expresiones correspondientes se aplican a  $\tilde{u}_n$  y a  $\tilde{v}_n$  siendo éstos estadísticos de orden intermedios.

Si el umbral  $u_n$  se estima mediante un estadístico de orden extremo el proceso puntual  $N_{k_n}^{(u_n)}$  converge a un proceso de Poisson, y su varianza por tanto converge a una constante. Este es un serio inconveniente para la consistencia de la mayoría de los estimadores de  $\theta$  que es solucionado en nuestro marco usando  $\tilde{u}_n$  (estadístico de orden intermedio).



### 3.3.1. Propiedades estadísticas de los diferentes estimadores

Consideremos primero el caso de  $\tilde{\theta}_n$  como el cociente de las variables aleatorias  $Z_{v_n}^*$  y  $Z_{u_n}^*$  donde  $v_n$  y  $u_n$  verifican (3.13) y (3.9) respectivamente, esto es,

$$\tilde{\theta}_n = \frac{Z_{v_n}^*}{Z_{u_n}^*}.$$

Por la expansión de Taylor de segundo orden de  $E[Z_{v_n}^*/Z_{u_n}^*]$  sobre los respectivos valores esperados (método delta) tenemos que

$$E[\tilde{\theta}_n] = \frac{E[Z_{v_n}^*]}{E[Z_{u_n}^*]} \left( 1 + \frac{V[Z_{u_n}^*]}{E[Z_{u_n}^*]^2} - \frac{Cov[Z_{v_n}^*, Z_{u_n}^*]}{E[Z_{u_n}^*]E[Z_{v_n}^*]} \right) + O\left(\frac{1}{\tau^2}\right). \quad (3.23)$$

Las diferentes contribuciones a  $Z_{u_n}^*$  no son independientes dos a dos. En particular,  $E[Z_{u_n}^{*2}] = k_n P\{M_1 > u_n\} + \sum_{i=1}^{k_n} \sum_{j \neq i}^{k_n} P\{M_i > u_n, M_j > u_n\}$ , donde  $M_i$  se usa para denotar el máximo de  $\{X_{(i-1)r_n+1}, \dots, X_{ir_n}\}$ . Por estacionariedad la varianza se puede expresar como  $V[Z_{u_n}^*] = E[Z_{u_n}^*] + k_n^2 P\{M_1 > u_n, M_2 > u_n\} - E^2[Z_{u_n}^*] - k_n P\{M_1 > u_n, M_2 > u_n\}$ . Bajo  $D'(u_n)$  la diferencia entre  $k_n^2 P\{M_1 > u_n, M_2 > u_n\}$  y  $E^2[Z_{u_n}^*]$  converge a 0 cuando  $n$  crece, y  $k_n P\{M_1 > u_n, M_2 > u_n\}$  se aproxima adecuadamente por  $E[Z_{u_n}^*]P\{M_1 > u_n\}$  que por tanto también converge a 0. La covarianza también toma una expresión similar  $Cov[Z_{u_n}^*, Z_{v_n}^*] = E[Z_{v_n}^*] + k_n^2 P\{M_1 > u_n, M_2 > v_n\} - E[Z_{u_n}^*]E[Z_{v_n}^*] - k_n P\{M_1 > u_n, M_2 > v_n\}$  que se reduce a esta expresión  $Cov[Z_{u_n}^*, Z_{v_n}^*] = E[Z_{v_n}^*]$ .

Por lo tanto la expresión (3.23) para  $n$  suficientemente grande es como sigue

$$E[\tilde{\theta}_n] = \frac{E[Z_{v_n}^*]}{E[Z_{u_n}^*]} \left( 1 + \frac{E[Z_{u_n}^*]}{E[Z_{u_n}^*]^2} - \frac{E[Z_{v_n}^*]}{E[Z_{u_n}^*]E[Z_{v_n}^*]} \right) + O\left(\frac{1}{\tau^2}\right),$$

y es inmediato ver que el valor esperado de nuestro estimador toma esta expresión,

$$E[\tilde{\theta}_n] = \theta + O\left(\frac{1}{\tau^2}\right). \quad (3.24)$$

Para el análisis de la varianza es útil derivar los momentos condicionales. Consideremos  $Z_{u_n}^* = z_{u_n}^*$  conocido, y notemos que las secuencias  $u_n$  y  $v_n$  están relacionadas por esta expresión,

$$1 - F_{1,\dots,r_n}(v_n) = (1 - F_{1,\dots,r_n}(u_n)) \left( 1 - \frac{F_{1,\dots,r_n}(v_n) - F_{1,\dots,r_n}(u_n)}{1 - F_{1,\dots,r_n}(u_n)} \right). \quad (3.25)$$

Entonces por (3.11),  $E[\tilde{\theta}_n | Z_{u_n}^* = z_{u_n}^*] = 1 - \frac{F_{1,\dots,r_n}(v_n) - F_{1,\dots,r_n}(u_n)}{1 - F_{1,\dots,r_n}(u_n)}$ , y la varianza condicional toma esta forma,

$$V[\tilde{\theta}_n | Z_{u_n}^* = z_{u_n}^*] = \frac{1}{z_{u_n}^*} \left( 1 - \frac{F_{1,\dots,r_n}(v_n) - F_{1,\dots,r_n}(u_n)}{1 - F_{1,\dots,r_n}(u_n)} \right). \quad (3.26)$$

Por la ley de la esperanzas iteradas la varianza incondicional se puede expresar en dos términos diferentes,  $V[\tilde{\theta}_n] = V[E[\tilde{\theta}_n | Z_{u_n}^*]] + E[V[\tilde{\theta}_n | Z_{u_n}^*]]$ . Está claro que el primer término es 0, y por la expansión de Taylor de  $E[1/Z_{u_n}^*]$  sobre  $E[Z_{u_n}^*]$  obtenemos que

$$E[V[\tilde{\theta}_n | Z_{u_n}^* = z_{u_n}^*]] = \left( 1 - \frac{F_{1,\dots,r_n}(v_n) - F_{1,\dots,r_n}(u_n)}{1 - F_{1,\dots,r_n}(u_n)} \right) \left( \frac{1}{E[Z_{u_n}^*]} + \frac{V[Z_{u_n}^*]}{E^3[Z_{u_n}^*]} \right). \quad (3.27)$$

En consecuencia,

$$V[\tilde{\theta}_n] = \left( 1 - \frac{F_{1,\dots,r_n}(v_n) - F_{1,\dots,r_n}(u_n)}{1 - F_{1,\dots,r_n}(u_n)} \right) \left( \frac{1}{\theta\tau} + O\left(\frac{1}{\tau^2}\right) \right) = O\left(\frac{1}{\tau}\right). \quad (3.28)$$

Por lo tanto el error cuadrático medio (MSE) de nuestro estimador es de orden  $O(\frac{1}{\tau})$  con  $\tau$  constante. Este resultado implica que este estimador no es consistente para  $u_n$  y  $v_n$  definidos por estadísticos de orden extremos. La consistencia, sin embargo, se conseguirá cuando estas secuencias se reemplazan por  $\tilde{u}_n$  y  $\tilde{v}_n$  estadísticos de orden intermedios como se muestra en la sección siguiente.

Nuestro estimador se puede interpretar como un refinamiento del método estándar de los bloques  $\hat{\theta}_n^{(2)}$  escribiendo  $\tilde{\theta}_n = \frac{Z_{v_n}^*/Z_{u_n}}{\hat{\theta}_n^{(2)}}$ . Las propiedades asintóticas del último estimador  $\hat{\theta}_n^{(2)}$  se derivan en Hsing (1991) ó en Smith y Weissman (1994). Por medio del método delta estos autores encuentran que  $E[\hat{\theta}_n^{(2)}] = \theta + O(\frac{1}{\tau})$ , y la varianza es  $V[\hat{\theta}_n^{(2)}] = O(\frac{1}{\tau})$ . Por lo

tanto el sesgo de este estimador es mayor que el sesgo de  $\tilde{\theta}_n$ , pero el error cuadrático medio (MSE) de ambos estimadores es  $O(1/\tau)$ .

Para el método de los logaritmos,

$$E[\hat{\theta}_n^{(1)}] = \frac{E[Z_{u_n}^*]}{E[Z_{u_n}]} \left( 1 + \frac{E[Z_{u_n}^*]}{2k_n} + \frac{E^2[Z_{u_n}^*]}{6k_n^2} \right) = \theta + O\left(\frac{\tau}{k_n}\right), \text{ y } V[\hat{\theta}_n^{(1)}] = O\left(\frac{1}{\tau}\right).$$

Este estimador es asintóticamente insesgado pero no es consistente tampoco para  $\tau$  constante.

### 3.3.2. Inferencia para el Extremal Index

Consideremos ahora las secuencias  $\tilde{u}_n$  y  $\tilde{v}_n$  definidas por las condiciones  $\tau'_n = k_n(1 - F_{1,\dots,r_n}(\tilde{u}_n))$ ,  $\tau''_n = k_n(1 - F_{1,\dots,r_n}(\tilde{v}_n))$ , con  $\tau'_n \rightarrow \infty$ ,  $\tau''_n \rightarrow \infty$  y  $\tau''_n/\tau'_n \rightarrow \theta$ . En este caso los primeros dos momentos de las variables aleatorias  $Z_{\tilde{u}_n}^*$  y  $Z_{\tilde{v}_n}^*$  divergen a infinito. Por estacionariedad la varianza viene dada por esta expresión,

$$V[Z_{\tilde{u}_n}^*] = E[Z_{\tilde{u}_n}^*] + (k_n^2 P\{M_1 > \tilde{u}_n, M_2 > \tilde{u}_n\} - E^2[Z_{\tilde{u}_n}^*]) - k_n P\{M_1 > \tilde{u}_n, M_2 > \tilde{u}_n\}.$$

Notemos que en este caso, bajo  $D'(u_n)$  para  $n$  suficientemente grande, la varianza es

$$V[Z_{\tilde{u}_n}^*] = E[Z_{\tilde{u}_n}^*] - E[Z_{\tilde{u}_n}^*]P\{M_1 > \tilde{u}_n\}. \quad (3.29)$$

La covarianza por lo tanto toma esta expresión,

$$Cov[Z_{\tilde{u}_n}^*, Z_{\tilde{v}_n}^*] = E[Z_{\tilde{v}_n}^*] - E[Z_{\tilde{v}_n}^*]P\{M_1 > \tilde{u}_n\}.$$

Por lo tanto la expresión (3.23) es como sigue

$$E[\tilde{\theta}_n] = \frac{\tau''_n}{\tau'_n} \left( 1 + \frac{\tau'_n P\{M_1 \leq \tilde{u}_n\}}{(\tau'_n)^2} - \frac{\tau''_n P\{M_1 \leq \tilde{u}_n\}}{\tau'_n \tau''_n} \right) + o\left(\frac{1}{\tau_n}\right), \quad (3.30)$$

que se reduce a  $E[\tilde{\theta}_n] = \theta + o\left(\frac{1}{\tau_n}\right)$ , por la definición de  $\tau'_n$  and  $\tau''_n$ .

Este estimador de  $\theta$  ahora es asintóticamente insesgado, y para  $\tau_n < k_n$ ,  $\tau_n^2 > k_n$ ,  $\tilde{\theta}_n$  es superior a  $\hat{\theta}_n^{(1)}$  en este sentido. Para  $\tau_n > k_n$  este resultado es trivial.

Para encontrar la varianza incondicional en este caso, calculamos primero los momentos condicionales.

$$V[\tilde{\theta}_n | Z_{\tilde{u}_n}^* = z_{\tilde{u}_n}^*] = \frac{1}{z_{\tilde{u}_n}^{*2}} V[Z_{\tilde{v}_n}^* | Z_{\tilde{u}_n}^* = z_{\tilde{u}_n}^*]. \quad (3.31)$$

Aplicando (3.29) a la variable aleatoria  $Z_{\tilde{v}_n}^* | Z_{\tilde{u}_n}^*$ ,

$$V[Z_{\tilde{v}_n}^* | Z_{\tilde{u}_n}^* = z_{\tilde{u}_n}^*] = E[Z_{\tilde{v}_n}^* | Z_{\tilde{u}_n}^* = z_{\tilde{u}_n}^*] P\{M_1 \leq \tilde{v}_n | M_1 > \tilde{u}_n\},$$

que equivale a

$$V[Z_{\tilde{v}_n}^* | Z_{\tilde{u}_n}^* = z_{\tilde{u}_n}^*] = z_{\tilde{u}_n}^* P\{M_1 > \tilde{v}_n | M_1 > \tilde{u}_n\} P\{M_1 \leq \tilde{v}_n | M_1 > \tilde{u}_n\}. \quad (3.32)$$

Entonces, de la misma forma que en (3.27),

$$V[\tilde{\theta}_n] = \left(1 - \frac{F_{1,\dots,r_n}(\tilde{v}_n) - F_{1,\dots,r_n}(\tilde{u}_n)}{1 - F_{1,\dots,r_n}(\tilde{u}_n)}\right) \left(\frac{F_{1,\dots,r_n}(\tilde{v}_n) - F_{1,\dots,r_n}(\tilde{u}_n)}{1 - F_{1,\dots,r_n}(\tilde{u}_n)}\right) \left(\frac{1}{E[Z_{\tilde{u}_n}^*]} + \frac{V[Z_{\tilde{u}_n}^*]}{E^3[Z_{\tilde{u}_n}^*]}\right),$$

que por tanto queda

$$V[\tilde{\theta}_n] = \left(1 - \frac{F_{1,\dots,r_n}(\tilde{v}_n) - F_{1,\dots,r_n}(\tilde{u}_n)}{1 - F_{1,\dots,r_n}(\tilde{u}_n)}\right) \left(\frac{F_{1,\dots,r_n}(\tilde{v}_n) - F_{1,\dots,r_n}(\tilde{u}_n)}{1 - F_{1,\dots,r_n}(\tilde{u}_n)}\right) \frac{1}{\tau'_n} + o\left(\frac{1}{\tau_n}\right). \quad (3.33)$$

Bajo  $D'(\tilde{u}_n)$  la distribución de  $Z_{\tilde{v}_n}^* | Z_{\tilde{u}_n}^*$  se aproxima fielmente ( $\sim$ ) por una distribución binomial de parámetros  $bin\left(Z_{\tilde{u}_n}^*, 1 - \frac{F_{1,\dots,r_n}(\tilde{v}_n) - F_{1,\dots,r_n}(\tilde{u}_n)}{1 - F_{1,\dots,r_n}(\tilde{u}_n)}\right)$ , y  $Z_{\tilde{u}_n}^*$  por una  $bin(k_n, 1 - F_{1,\dots,r_n}(\tilde{u}_n))$ .

Entonces, la distribución de  $\tilde{\theta}_n$  se puede aproximar por una distribución normal con parámetros dados en (3.30) y (3.33).

OPor otro lado la relación entre las colas introducida en (3.25) se mantiene para  $\tilde{u}_n$  y  $\tilde{v}_n$ , y por definición  $\tau''_n/\tau'_n \rightarrow \theta$ , dando este resultado  $1 - \frac{F_{1,\dots,r_n}(\tilde{v}_n) - F_{1,\dots,r_n}(\tilde{u}_n)}{1 - F_{1,\dots,r_n}(\tilde{u}_n)} \rightarrow \theta$ . La

distribución de  $\tilde{\theta}_n$  se aproxima por

$$\tilde{\theta}_n \stackrel{w}{\sim} N\left(\theta, \frac{\theta(1-\theta)}{\tau'_n}\right). \quad (3.34)$$

Por la estructura de dependencia  $\frac{\tau'_n}{\tau_n} \rightarrow \theta$  cuando  $n$  tiende a infinito, y entonces  $\tilde{\theta}_n \stackrel{w}{\sim} N\left(\theta, \frac{1-\theta}{\tau_n}\right)$  resulta una aproximación válida para la distribución de  $\tilde{\theta}_n$ . Más formalmente, podemos obtener un test estadístico que es asintóticamente libre de parámetros,

$$T_n = \frac{\tilde{\theta}_n - \theta}{\sqrt{1-\theta}} \sqrt{\tau_n} \stackrel{w}{\rightarrow} N(0, 1). \quad (3.35)$$

Los intervalos de confianza asintóticos para  $\theta$  se calculan fácilmente de la expresión anterior.

$$\theta \in \left[ \tilde{\theta}_n \pm z_{1-\alpha/2} \sqrt{\frac{1-\tilde{\theta}_n}{\tau_n}} \right], \quad (3.36)$$

con  $z_{1-\alpha/2}$  el cuantil de la distribución normal estándar. Este intervalo es una aproximación de los verdaderos intervalos de confianza para muestras finitas. La región exacta de confianza para *tamaños* de muestra *pequeños* se puede aproximar mejor mediante técnicas de remuestreo. El intervalo de confianza toma esta expresión

$$\theta \in \left[ \tilde{\theta}_n - \sqrt{\frac{1-\tilde{\theta}_n}{\tau_n}} J_n^{-1}\left(1 - \frac{\alpha}{2}, F\right), \tilde{\theta}_n - \sqrt{\frac{1-\tilde{\theta}_n}{\tau_n}} J_n^{-1}\left(\frac{\alpha}{2}, F\right) \right], \quad (3.37)$$

donde  $J_n^{-1}(1-\alpha, F)$  es el cuantil  $1-\alpha$  de la distribución muestral  $J_n(F)$  del estadístico  $T_n$ . En la práctica este cuantil se aproxima por el estadístico de orden  $T_{n,((1-\alpha)B)}$  de la muestra  $T_{n,1}, \dots, T_{n,B}$  con  $B$  el número de iteraciones. La notación  $F$  en la distribución  $J_n(F)$  se refiere a la simulación Monte-Carlo, esto es, el proceso generador de los datos es conocido. De otro modo  $J_n(F)$  debe ser aproximado por  $J_b(F)$  con  $b < n$ ,  $b/n \rightarrow 0$  (subsampling), ó por  $J_n(F^*)$  con  $F^*$  representando métodos bootstrap por bloques. El bootstrap naïve no funciona en este contexto debido a la dependencia temporal existente en los datos, no obstante, como se ve en las simulaciones, los intervalos derivados de la distribución asintótica

normal ofrece aproximaciones adecuadas de las regiones de confianza exactas para *tamaños* de muestra moderados.

Nuestro interés sin embargo recae sobre intervalos de confianza del tipo  $H_0 : \theta = \theta_0$  vs  $H_1 : \theta < \theta_0$ . Este contraste de hipótesis de una cola (one-sided) permite dar validez a la condición mixing  $D''(u_n)$  introducida en (3.3) bajo el supuesto de la hipótesis nula  $\theta_0 = 1$ . En otras palabras, si  $\theta = 1$  no hay clusters de valores extremos (valores que exceden cierto umbral) en la secuencia estacionaria. La hipótesis nula equivale a ver si  $\theta_0$  está contenido en el intervalo  $\left(-\infty, \tilde{\theta}_n + z_{1-\alpha} \sqrt{\frac{1-\tilde{\theta}_n}{\tau_n}}\right]$  ó alternativamente en  $\left(-\infty, \tilde{\theta}_n - \sqrt{\frac{1-\tilde{\theta}_n}{\tau_n}} J_n^{-1}(\alpha, F)\right]$ , la aproximación bootstrap.

Consideremos el ejemplo debido a Chernick (1981) para  $\{X_n\}$  una secuencia estacionaria autorregresiva de primer orden generada por el modelo  $X_i = \frac{1}{r}X_{i-1} + \varepsilon_i$ , donde  $r \geq 2$  es un número entero,  $\varepsilon_i$  son variables aleatorias uniformes discretas en  $\{0, 1/r, \dots, (r-1)/r\}$ , siendo además independientes de  $X_{i-1}$ , y  $X_i$  y teniendo una distribución uniforme en  $[0, 1]$ . El extremal index viene dado por  $\theta = \frac{r-1}{r}$ . El gráfico dado en 3.7.1 describe la curva de las estimaciones de  $\theta$  para diferentes particiones de la muestra y diferentes *tamaños* de muestra. El panel superior es para  $n = 200$  y el inferior considera  $n = 1000$ .

Se pueden extraer dos conclusiones de estos gráficos. Primero, está claro que la condición  $D''(\hat{u}_n)$  se rechaza con  $\alpha = 0,05$ , y segundo, los intervalos de confianza para  $\theta$  son más *pequeños* cuando  $n$  incrementa. Esto se debe a la elección de un estadístico de orden intermedio,  $\hat{u}_n = x_{(n-k)}$  con  $k = \sqrt{2n}$ , como umbral.

### 3.3.3. Algunos comentarios sobre la selección del *tamaño* del bloque

La partición de la secuencia  $\{X_n\}$  en  $k_n$  bloques de *tamaño*  $r_n$  tiene dos características principales: Primero, esta partición define un proceso puntual  $N_{k_n}^{(u_n)}$  que converge a un proceso donde las observaciones son independientes, y segundo, la distribución de esta secuencia converge a un proceso de Poisson cuando  $k_n$  tiende a infinito. La mayoría de los estimadores para el extremal index encontrados en la literatura están estrechamente rela-

cionados a la partición correspondiente. Esta dependencia por ejemplo se vuelve explícita para el método de los logaritmos donde  $k_n$  aparece en la expresión  $\hat{\theta}_n^{(1)}$ , así como en su valor esperado.

Si las observaciones son independientes ó débilmente dependientes,  $N_{k_n}^{(u_n)}$  con  $k_n = n$  define por si mismo un proceso puntual *iid* y  $\theta = 1$ . De otra forma el extremal index es menor que 1 y la partición  $k_n < n$  juega un papel central en la estimación del extremal index.

Dado  $n$ , una selección adecuada de la secuencia  $k_n$  junto con un umbral  $u_n$  definen una secuencia dada por el máximo sobre los correspondientes bloques de observaciones. En la práctica contrastar la condición  $D'(u_n)$  se reemplaza por contrastar la independencia en esta serie dada por los máximos. Los contrastes de hipótesis para esta última condición requieren *tamaños* de muestra muy grandes y la mayoría de ellos se basan en supuestos gaussianos. Una alternativa naïve a estos métodos es descartar las primeras particiones definidas por valores de  $k_n$  como  $k_n = n, n - 1, \dots$ , que tienen una probabilidad muy alta de conllevar dependencia en la serie  $N_{k_n}^{(u_n)}$ , y analizar el comportamiento de los estimadores estudiando el valor de las correspondientes estimaciones a lo largo de las diferentes particiones.

La influencia de la elección de  $k_n$  en las estimaciones del extremal index también depende de la selección del umbral  $u_n$ . Por ejemplo en el método de los bloques las estimaciones de  $\theta$  están dirigidas por la partición correspondiente para valores bajos de  $u_n$ . En este caso  $Z_{u_n}^*$  y  $k_n$  toman los mismos valores resultando en una secuencia de estimaciones que se aproxima a 0 cuando  $k_n$  decrece.

Una situación similar ocurre para el método de los logaritmos con  $u_n$  un umbral bajo. El numerador en este caso colapsa a  $-\infty$  y el estimador no está definido para la partición correspondiente. Estos efectos no están presentes en  $\tilde{\theta}_n$  porque un ratio mayor  $Z_{u_n}^*/k_n$  dado por un umbral que toma valores bajos es compensada por un valor grande de  $Z_{v_n}^*/k_n$ , esto es, la secuencia  $v_n$  varía de acuerdo a  $u_n$ .

### 3.4. Simulaciones: Algunos ejemplos

Consideremos ahora algunos ejemplos de la literatura que muestran dependencia a corto plazo ( $0 < \theta < 1$ ) reflejada en una distribución del máximo que satisface la condición (3.5).

El siguiente ejemplo es el modelo doble estocástico estudiado por Smith y Weissman (1994). Sea  $\{\xi_i, i \geq 1\}$  observaciones *iid* con función de distribución  $F$ , y supongamos que  $Y_1 = \xi_1$ , y para  $i > 1$ ,  $Y_i = Y_{i-1}$  con probabilidad  $\psi$ , y  $Y_i = \xi_i$  con probabilidad  $1 - \psi$ . La secuencia doble estocástica  $\{X_i\}$  se define por  $X_i = Y_i$  con probabilidad  $\eta$ , y  $X_i = 0$  con probabilidad  $1 - \eta$ , con estos eventos diferentes mutuamente independientes. El extremal index es

$$\theta = \frac{1 - \psi}{1 - \psi + \psi\eta}.$$

Los siguientes gráficos, figura (3.7.2), representan las sendas de los diferentes estimadores para el extremal index. La secuencia umbral se estima por un estadístico de orden:  $\hat{u}_n = x_{(n-k)}$ . En este trabajo hemos implementado dos tipos de estadísticos de orden para muestras de tamaño  $n = 200$  y  $n = 1000$  observaciones. Un estadístico de orden extremo ( $k = 20$  fijado), y un estadístico de orden intermedio ( $k = \sqrt{2n}$ ). Presentamos sólo las estimaciones de  $\theta$  para el estadístico de orden intermedio ya que la otra secuencia umbral ofrece resultados similares para estos tamaños de muestra en este ejemplo.

Las curvas describen las medias muestrales de las diferentes estimaciones del extremal index y para diferentes particiones de la muestra para  $m = 100$  secuencias simuladas generadas del modelo introducido en Smith y Weissman con  $\psi = 0,9$  y  $\eta = 0,7$ . Supongamos también que  $F$ , es una distribución de Fréchet  $F(x) = \exp(-x^{-\alpha})$  con  $\alpha = 1$  y  $x \in (0, \infty)$ .

Los intervalos de confianza para  $\tilde{\theta}_n$  derivados en la última sección no están representados gráficamente. En cambio, hemos representado la desviación estándar simulada de los diferentes estimadores para  $m = 100$  para poder presentar una comparación en los mismos términos entre los tres competidores. La desviación estándar para las diferentes particiones



se estima vía simulación Monte-Carlo por  $\hat{\sigma}_{k_n}$  con

$$\hat{\sigma}_{k_n}^2 = \frac{1}{m-1} \sum_{i=1}^m (\theta_{i,est} - \bar{\theta}_{est})^2, \quad (3.38)$$

y  $\bar{\theta}_{est}$  la media muestral de las diferentes estimaciones.

Aparentemente el método de los bloques es el mejor método. Después de las primeras particiones de la muestra, la curva de las estimaciones de  $\theta$  se mantiene estable muy próxima al objetivo en los tres métodos. No obstante el estimador del método de los bloques tiene menor varianza. Además, si nos centramos en la figura 3.7.3 está claro que los diferentes estimadores de  $\theta$  analizados en este ejemplo son consistentes y el método de los bloques es más eficiente. El error cuadrático medio es estimado de las secuencias simuladas generadas para la figura 3.7.2 y toma esta expresión,

$$MSE(\theta_{est}) = \frac{1}{m} \sum_{i=1}^m (\theta_{i,est} - \theta)^2.$$

Estos resultados están en la misma línea que las conclusiones encontradas en Smith y Weissman (1994). Sin embargo el impresionante comportamiento del método de los bloques puede ser debido al valor tan bajo del verdadero parámetro ( $\theta = 0,137$ ) y a la selección de una secuencia umbral baja. Bajo estas circunstancias la curva de estimaciones de  $\theta$  por el método de los bloques decae cuando  $k_n$  decrece ( $r_n$  crece) y se aproxima al verdadero parámetro. Para conseguir cierta perspectiva entorno a esto, estudiamos también el proceso doble estocástico donde el extremal index es significativamente mayor. Supongamos  $\psi = 0,5$  y  $\eta = 0,5$ , *i.e.*  $\theta = 0,66$ . Los gráficos presentados en las figuras 3.7.4 y 3.7.5 son los análogos de las figuras 3.7.2 y 3.7.3.

El método de los bloques en este ejemplo no funciona. El número de bloques con algún valor superando  $\hat{u}_n$  ( $Z_{\hat{u}_n}^*$ ) es similar a  $k_n$  para cada partición. Por lo tanto el estimador decrece cuando  $k_n$  decrece ya que  $Z_{\hat{u}_n}$  se mantiene constante. Por otro lado el método de los logaritmos mejora cuando  $n$  crece y el error cuadrático medio de  $\tilde{\theta}_n$  y  $\hat{\theta}_n^{(1)}$  son despreciables para  $n = 1000$ .

Finalmente los intervalos de confianza exactos y asintóticos para  $\theta$  se muestran en la figura 3.7.6 para dar validez a las estimaciones dadas por  $\tilde{\theta}_n$ .

### 3.5. Clustering en Series Financieras: El Caso del índice DaX

Los rendimientos financieros se caracterizan por una serie de hechos estilizados: el efecto apalancamiento, “leverage effect”, (después de periodos de alta volatilidad la probabilidad de que haya pérdidas es mayor que en periodos de calma), colas pesadas, clustering de las observaciones más grandes y cierta asimetría hacia la cola de las pérdidas. El artículo pionero de Engle y Bollerslev (1986) propuso los populares modelos GARCH, Generalized Auto-Regressive Conditional Heteroscedastic, donde la volatilidad condicional es heteroscedástica y sigue un proceso autorregresivo. La función de este modelo era explicar estas características de los datos. En general, el GARCH(1,1) es suficiente para modelizar la mayoría de los rendimientos financieros. Este modelo toma esta expresión,

$$X_i = \varepsilon_i \sigma_i, \quad \sigma_i^2 = \omega + \alpha X_{i-1}^2 + \beta \sigma_{i-1}^2,$$

con  $\omega, \alpha, \beta > 0$ , y  $\alpha + \beta < 1$ , que se pueden interpretar como un modelo ARMA(1,1) para los cuadrados,

$$X_i^2 = \omega + (\alpha + \beta) X_{i-1}^2 + \nu_i - \beta \nu_{i-1},$$

con  $\nu_i = \sigma_i^2 (\varepsilon_i^2 - 1)$ .

De acuerdo a este modelo, la dependencia encontrada en los rendimientos financieros se debe a los segundos momentos. La literatura relacionada con este aspecto es enorme; hasta el punto de que existen diferentes modelos GARCH para explicar características concretas de las series financieras.

Proponemos analizar algunos de estos hechos estilizados, en particular el clustering de

las observaciones más grandes, por medio del extremal index. Un valor de  $\theta$  significativamente menor que 1 muestra cierta dependencia a corto plazo reflejada en el clustering de las observaciones más grandes. Esto se puede interpretar como un patrón en la ocurrencia de los valores extremos, esto es, dado que se ha producido una pérdida muy grande de un activo financiero podemos esperar un periodo de grandes pérdidas (valores que exceden un umbral). La longitud media de este periodo es el recíproco del extremal index.

Los datos que usamos para ilustrar esta metodología consisten en el análisis del mercado bursátil de Frankfurt ( índice DaX) durante el periodo 19/12/1994 – 20/04/2001. Estos datos se han recogido de la página web *www.freelunch.com*. Las observaciones consideradas para el análisis son los rendimientos logarítmicos medidos en porcentaje y denotados como  $r_t$ ,

$$r_t = 100 (\log P_t - \log P_{t-1}),$$

con  $P_t$  los precios originales en el momento  $t$ . En la figura 3.7.7 se representan los rendimientos del índice DaX,  $r_t$ , y la correspondiente secuencia de los rendimientos al cuadrado,  $r_t^2$ .

El análisis del extremal index para ambas colas (figura 3.7.8) muestra cierto clustering en la ocurrencia de los valores extremos tanto positivos como negativos. Los intervalos de confianza derivados de  $\tilde{\theta}_n$  no contienen  $\theta = 1$  para  $\alpha = 0,05$ .

Estos gráficos también muestran un mayor nivel de clustering para los rendimientos negativos más grandes que para los valores positivos. Este hecho se puede contrastar estadísticamente por medio de intervalos de confianza para la diferencia de los extremal indexes correspondientes a la cola positiva y negativa. Este intervalo de confianza toma esta expresión

$$\theta_{pos} - \theta_{neg} \in \left[ \tilde{\theta}_{n,pos} - \tilde{\theta}_{n,neg} \pm z_{1-\alpha/2} \sqrt{\frac{1 - \tilde{\theta}_{n,pos}}{\tau_{n,pos}} + \frac{1 - \tilde{\theta}_{n,neg}}{\tau_{n,neg}}} \right]. \quad (3.39)$$

Es importante mencionar que  $\tilde{\theta}_{n,pos}$  y  $\tilde{\theta}_{n,neg}$  se consideran independientes. Esto puede llevar a obtener intervalos de confianza más *pequeños* dado  $\alpha$  en relación a considerar estimadores dependientes con correlación positiva. Para algunas particiones de la muestra es

estadísticamente significativo que el clustering de los rendimientos extremos positivos es menor que para los rendimientos extremos negativos (figura 3.7.9).

El análisis del clustering de los valores más grandes para la secuencia de la volatilidad (cuadrados de los rendimientos) merece comentarios de interés. El intervalo de confianza introducido en (3.39) se puede aplicar a contrastar la diferencia entre el extremal index de la secuencia de la volatilidad  $\theta_{vol}$  y  $\theta_{pos}$  ó  $\theta_{neg}$  (figura 3.7.10). Los resultados derivados de ambos contrastes,  $\theta_{pos} - \theta_{vol}$  y  $\theta_{neg} - \theta_{vol}$ , señalan que los valores extremos de la secuencia de la volatilidad son dirigidos por los valores extremos negativos. Por lo tanto estas observaciones son más grandes en valor absoluto que los rendimientos positivos más grandes. Este hecho explica la asimetría negativa existente en la secuencia de los rendimientos.

Finalmente merece la pena mencionar el hecho estilizado relacionado con las colas pesadas. Por la condición de Berman (Berman, 1964), si  $\{r_t\}$  es una secuencia normal estándar y  $Cov(r_t, r_{t-j}) \log j \rightarrow 0$  as  $j \rightarrow \infty$ , el extremal index de la secuencia es  $\theta = 1$ . En la práctica, la función de autocorrelación de los rendimientos de una serie financiera es normalmente cercana a cero, también lo cumple en este caso, entonces la segunda parte de la condición de Berman se cumple. Por lo tanto, si  $\theta < 1$  la secuencia de los rendimientos del índice DaX no se distribuye como una normal sino con colas pesadas. Esto sugiere que la existencia de clustering en los valores extremos de una secuencia financiera implica que la distribución de las observaciones es de colas pesadas. Entonces no es suficiente con los segundos momentos de  $\{r_t\}$  para conocer la estructura de dependencia de la serie. Más aún, la dependencia en los extremos juega un importante papel y esta dependencia surge de la existencia de colas pesadas.

### 3.6. Conclusión

El objetivo de este capítulo ha sido proponer un estimador del extremal index definido como el ratio del número de valores que exceden las dos secuencias umbral. Este estimador posee dos propiedades que lo hacen muy interesante: Primero, no es necesario elegir una

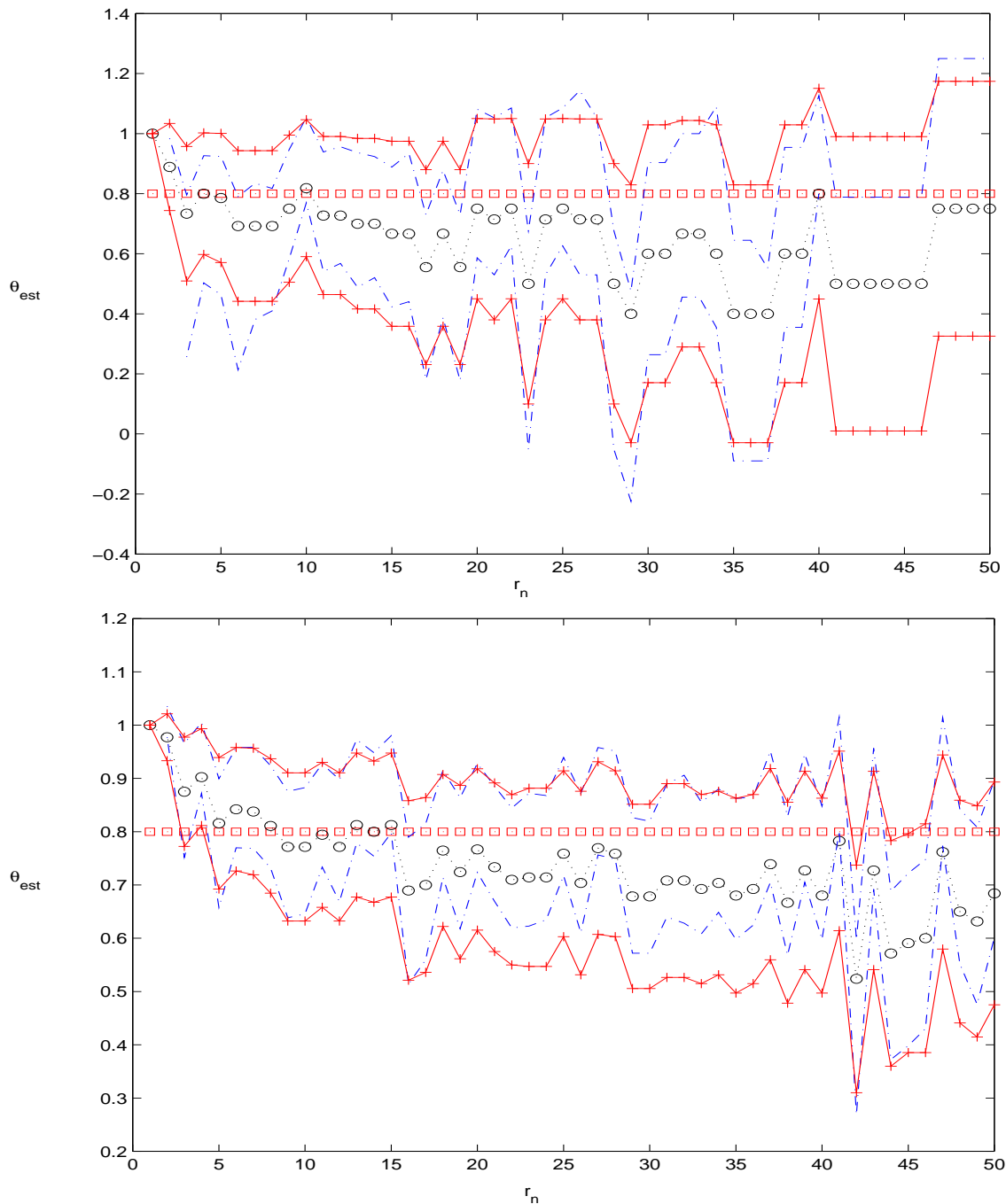
secuencia  $\{u_n\}$  que verifique la condición de Poisson en el límite, y segundo no es muy sensible a la selección del *tamaño* del bloque.

Considerando las propiedades asintóticas podemos concluir que nuestro estimador tiene el mismo orden de convergencia que los métodos estándar (las respectivas varianzas son del mismo orden). Sin embargo, bajo condiciones muy generales nuestro estimador es asintóticamente insesgado mejorando los otros dos métodos que no están libres de un término residual. Nuestro estimador también funciona mejor que estos otros métodos de dos maneras: no depende tanto de la correspondiente partición de la secuencia, y relaja la selección de la secuencia umbral.

Además, la ausencia de dependencia de la condición de Poisson permite proponer un contraste de hipótesis para el extremal index. Este contraste es útil en varias formas: formalmente valida las estimaciones del extremal index, introduce un procedimiento innovador para contrastar la existencia de clustering en la ocurrencia de los valores extremos y puede ser útil para determinar la asimetría y kurtosis de la distribución de los datos mediante contrastes para la diferencia de los extremal indexes entre ambas colas.

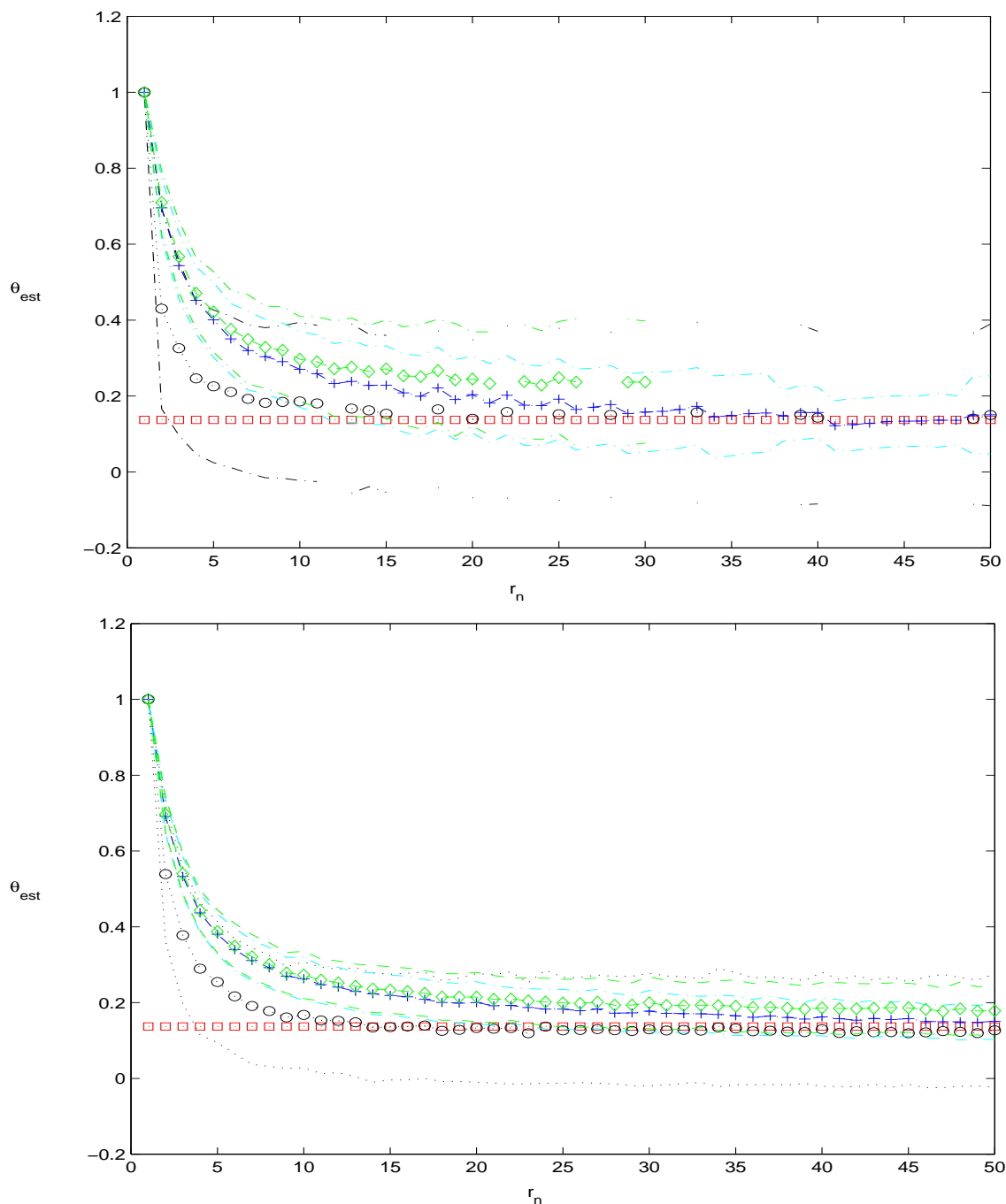
Finalmente, la aplicación de estas metodologías a series financieras ( índice DaX) confirma la existencia de dependencia a corto plazo en las observaciones extremas; esto es, cierto agrupamiento de los valores extremos de los rendimientos positivos y negativos. El clustering es mayor para la cola negativa. Por la condición de Berman, la distribución de las observaciones es de colas pesadas ya que  $\theta$  es significativamente menor que 1. Estos resultados concuerdan con los hechos estilizados encontrados en la mayoría de las series financieras.

## 3.7. Apéndice: Figuras



**Figura 3.7.1.** Valores estimados de  $\theta$  para el modelo de Chernick con  $r = 5$ .  $\theta = 0,8$  se representa por una línea  $\square$ . Las particiones  $r_n$  consideradas están en el rango  $[1, 50]$ .  $\tilde{\theta}_n$  se representa por  $(\dots)$  y  $o$ ; la línea discontinua describe los intervalos de confianza bootstrap con  $B = 1000$  y  $(+-)$  se emplea para los intervalos asintóticos. El nivel de significación es  $\alpha = 0,05$ . Los tamaños de muestra son  $n = 200$  y  $n = 1000$  respectivamente. El umbral es

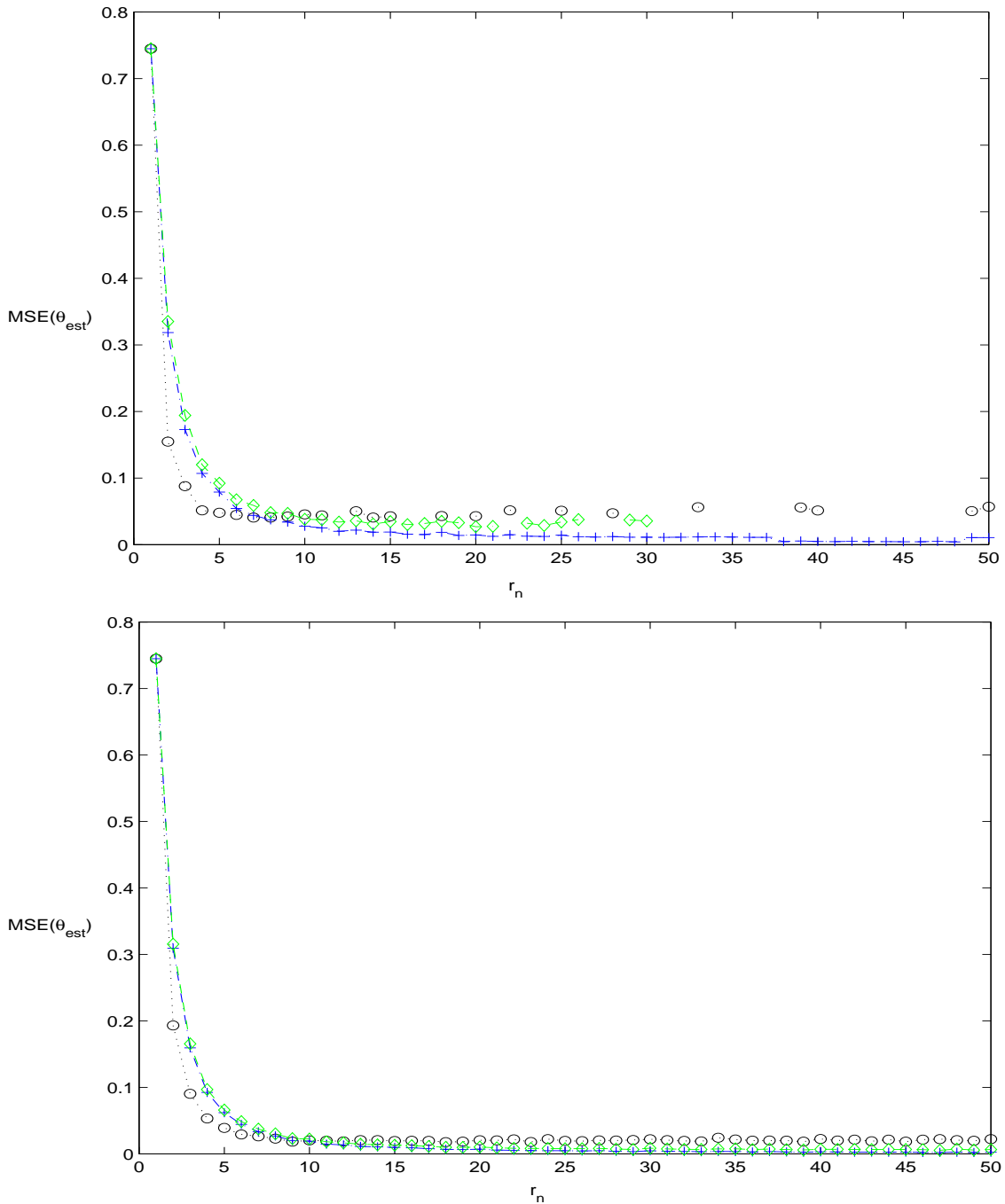
$$\hat{u}_n = x_{(n-k)} \text{ con } k = \sqrt{2n}.$$



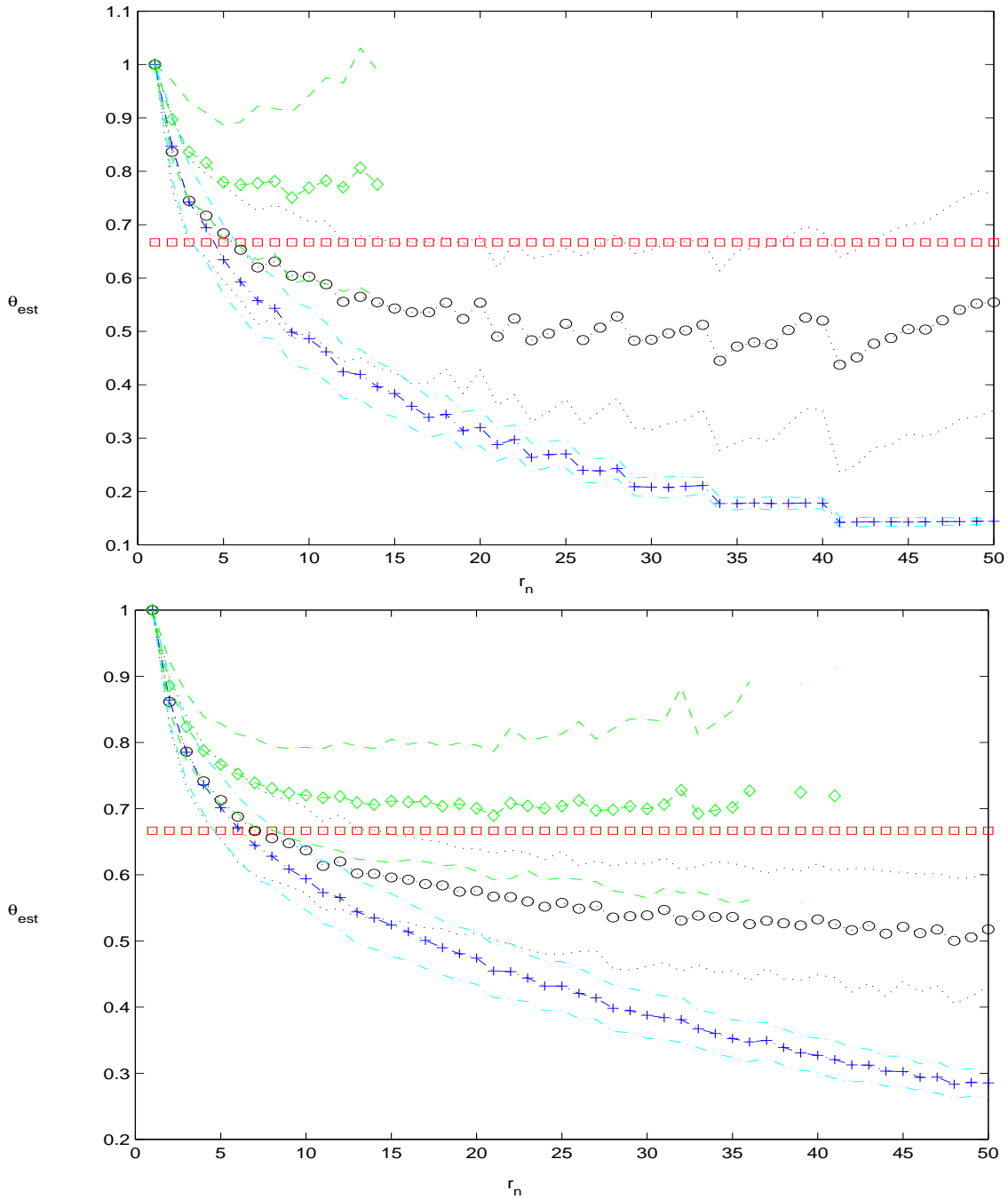
**Figura 3.7.2.** Valores estimados de  $\theta$  para el modelo doble estocástico con  $\psi = 0,9$  y  $\eta = 0,7$ . El extremal index es  $\theta = 0,137$  y se representa por una línea  $\square$ . Las particiones  $r_n$  consideradas están en el rango  $[1, 50]$ .  $\tilde{\theta}_n$  se representa por  $(\dots)$  y  $o$ ; la desviación estándar correspondiente se representa con  $(\dots)$ . El método de los logaritmos  $\hat{\theta}_n^{(1)}$  se representa con  $(- - -)$  y  $\diamond$ . La desviación estándar con  $(- - -)$ . El método de los bloques  $\hat{\theta}_n^{(2)}$  con  $(\cdot - \cdot -)$  y  $+$ , y  $(\cdot - \cdot -)$  para la desviación estándar. Los tamaños de muestra son  $n = 200$  y  $n = 1000$  respectivamente.  $m = 100$  simulaciones se utilizan. El umbral es  $\hat{u}_n = x_{(n-k)}$



con  $k = \sqrt{2n}$ .

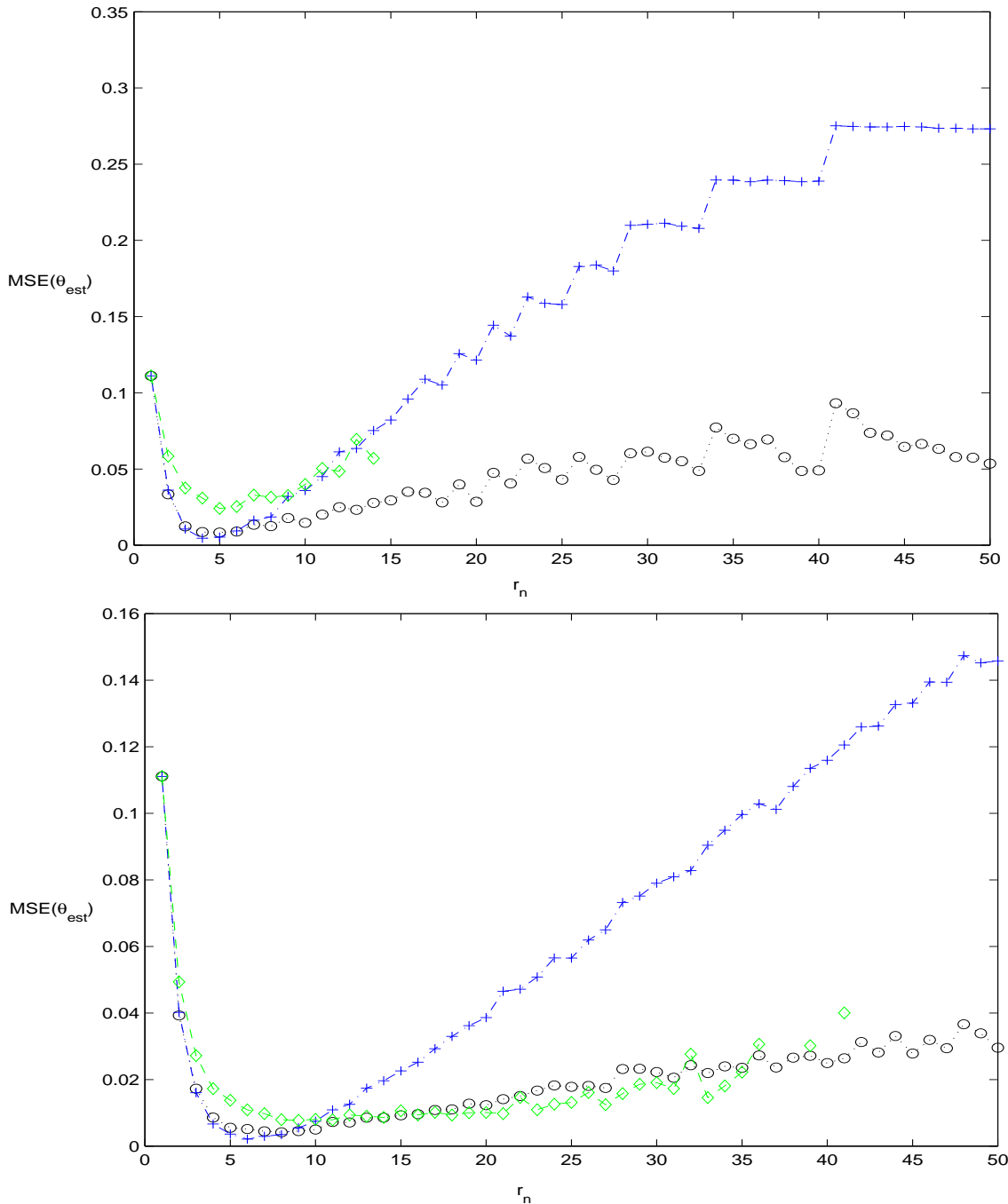


**Figura 3.7.3.** Error cuadrático medio simulado ( $MSE$ ) de los estimadores de  $\theta$  para el modelo doble estocástico con  $\psi = 0,9$  y  $\eta = 0,7$ . Las particiones  $r_n$  consideradas están en  $[1, 50]$ .  $m = 100$  simulaciones del modelo se generan.  $\tilde{\theta}_n$  se representa por  $(\dots)$  y  $o$ ,  $\hat{\theta}_n^{(1)}$  con  $(- - -)$  y  $\diamond$ , y  $(\cdot - \cdot -)$  y  $+$  para  $\hat{\theta}_n^{(2)}$ . Los tamaños de muestra son  $n = 200$  y  $n = 1000$  respectivamente. La secuencia umbral es  $\hat{u}_n = x_{(n-k)}$  con  $k = \sqrt{2n}$ .

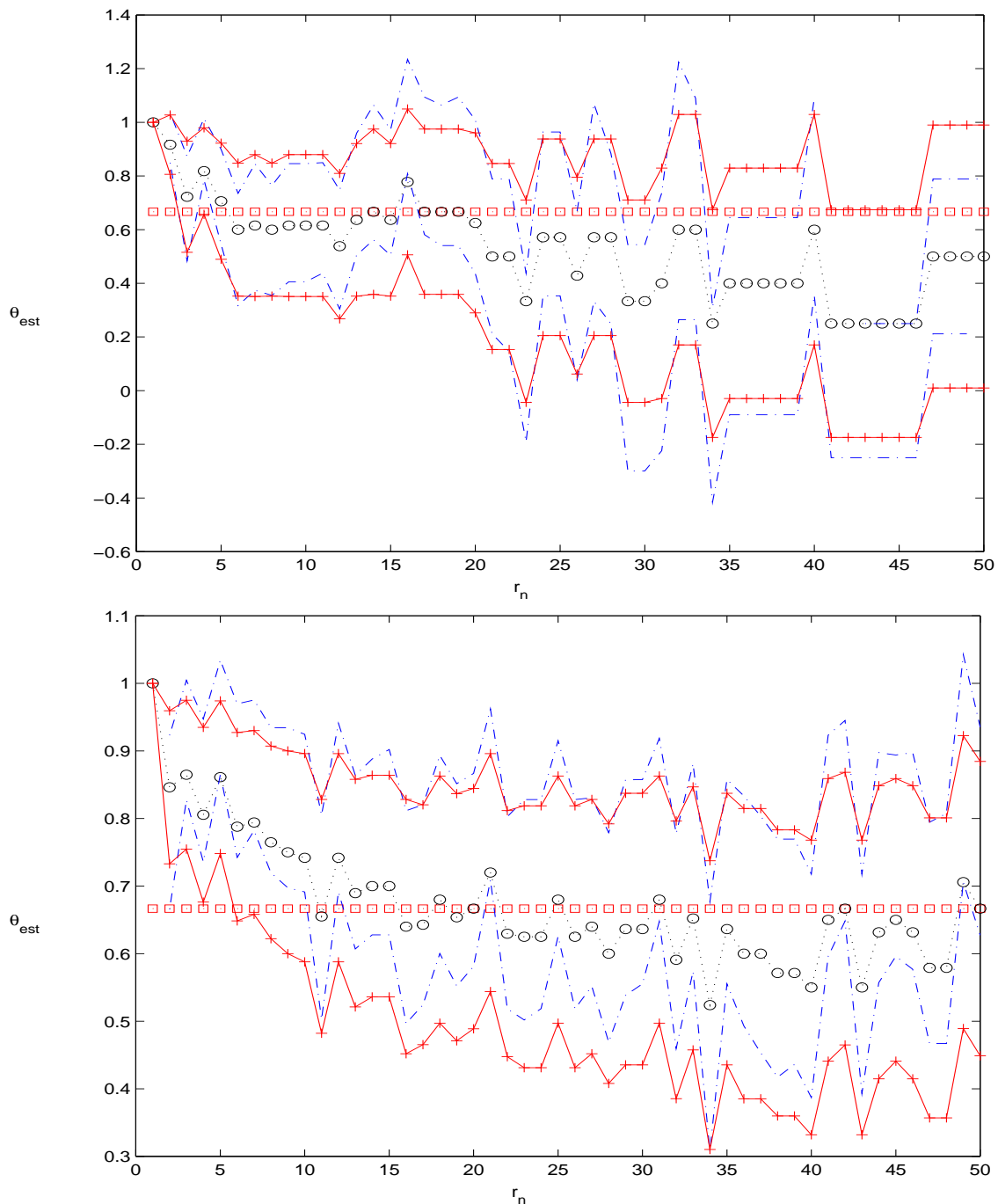


**Figura 3.7.4.** Valores estimados de  $\theta$  para el modelo doble estocástico con  $\psi = 0,5$  and  $\eta = 0,5$ . El extremal index es  $\theta = 0,66$  y se representa por una línea  $\square$ . Las particiones  $r_n$  consideradas están en el rango  $[1, 50]$ .  $\tilde{\theta}_n$  se representa por  $(\dots)$  y  $o$ ; la desviación estándar correspondiente se representa con  $(\dots)$ . El método de los logaritmos  $\hat{\theta}_n^{(1)}$  se representa con  $(- - -)$  y  $\diamond$ . La desviación estándar con  $(- - -)$ . El método de los bloques  $\hat{\theta}_n^{(2)}$  con  $(\cdot - \cdot -)$  y  $+$ , y  $(\cdot - \cdot -)$  para la desviación estándar. Los tamaños de muestra son  $n = 200$  y  $n = 1000$  respectivamente.  $m = 100$  simulaciones se utilizan. El umbral es  $\hat{u}_n = x_{(n-k)}$

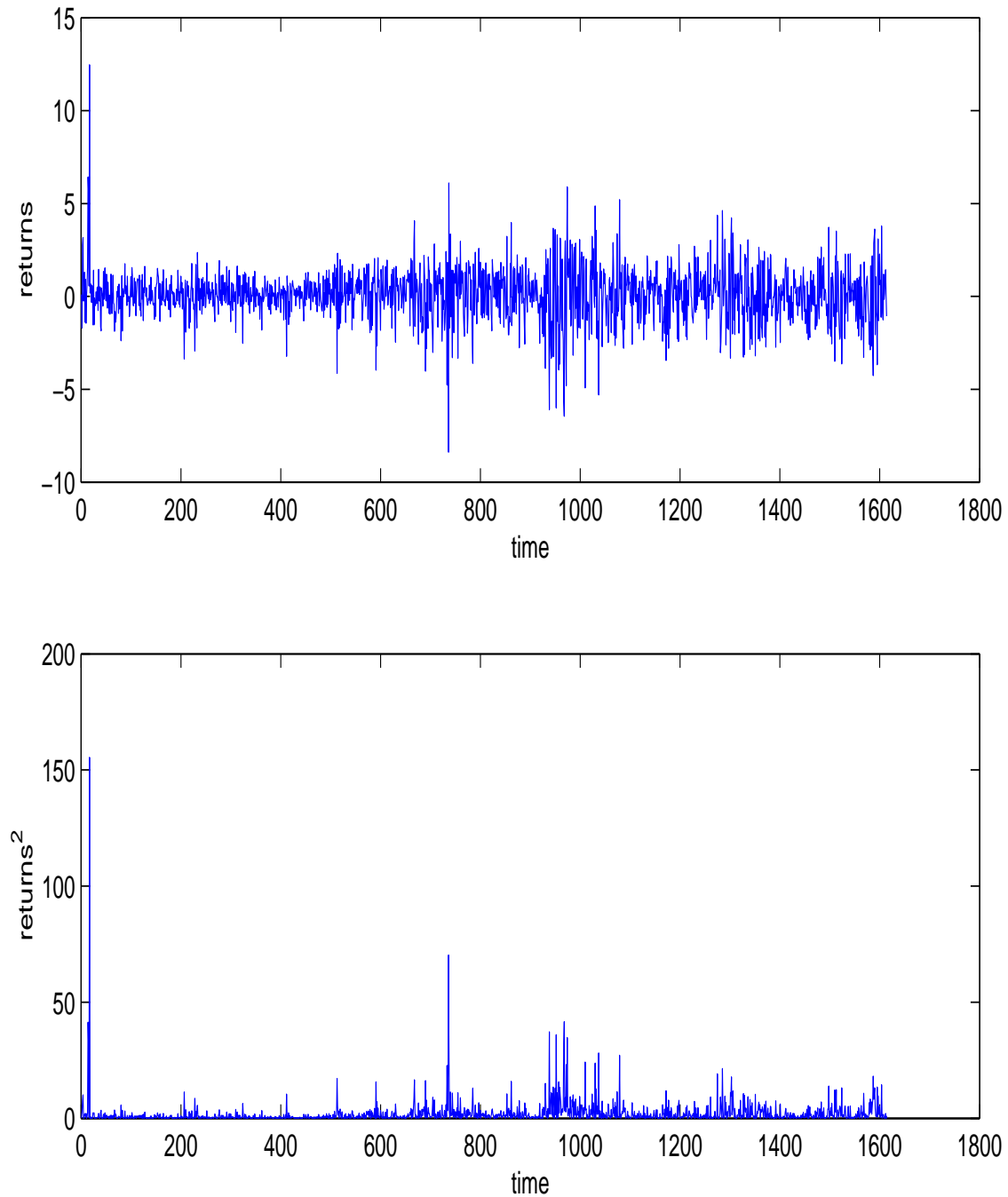
con  $k = \sqrt{2n}$ .



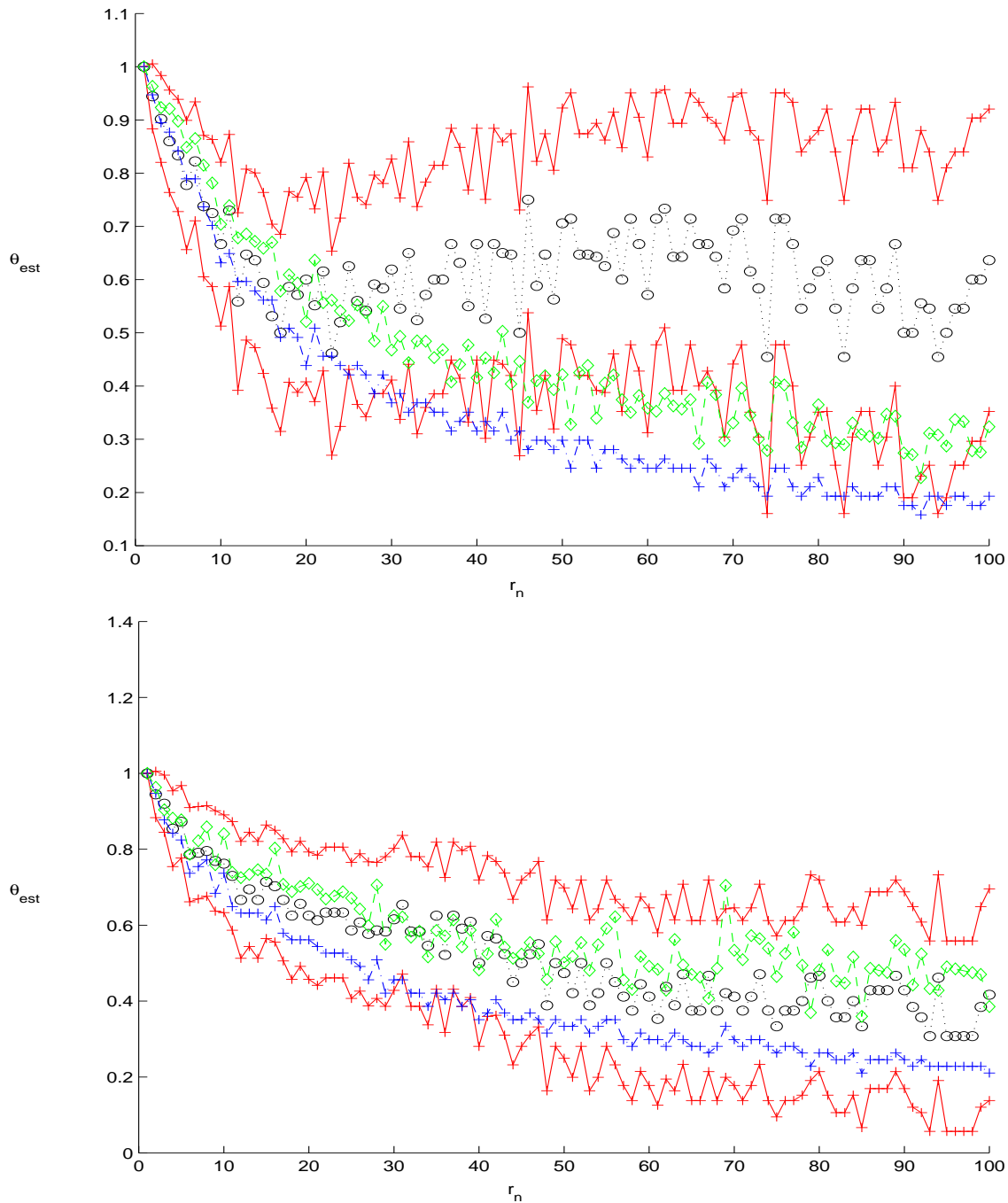
**Figura 3.7.5.** Error cuadrático medio simulado (MSE) de los estimadores de  $\theta$  para el modelo doble estocástico con  $\psi = 0,5$  y  $\eta = 0,5$ . Las particiones  $r_n$  consideradas están en  $[1, 50]$ .  $m = 100$  simulaciones del modelo se generan.  $\tilde{\theta}_n$  se representa por  $(\dots)$  y  $\hat{\theta}_n^{(1)}$  con  $(- - -)$  y  $\diamond$ , y  $(\cdot - \cdot -)$  y  $+$  para  $\hat{\theta}_n^{(2)}$ . Los tamaños de muestra son  $n = 200$  y  $n = 1000$  respectivamente. La secuencia umbral es  $\hat{u}_n = x_{(n-k)}$  con  $k = \sqrt{2n}$ .



**Figura 3.7.6.** Valores estimados de  $\theta$  para el modelo doble estocástico con  $\psi = 0,5$  and  $\eta = 0,5$ . El extremal index es  $\theta = 0,66$  y se representa por una línea  $\square$ . Las particiones  $r_n$  consideradas están en el rango  $[1, 50]$ .  $\tilde{\theta}_n$  se representa por  $(\cdot\cdot\cdot)$  y  $o$ ; la línea discontinua describe el intervalo de confianza bootstrap con  $B = 1000$  y  $(+-)$  se emplea para los intervalos asintóticos. El nivel de significación es  $\alpha = 0,05$ . Los tamaños de muestra son  $n = 200$  y  $n = 1000$  respectivamente. La secuencia umbral es  $\hat{u}_n = x_{(n-k)}$  con  $k = \sqrt{2n}$ .

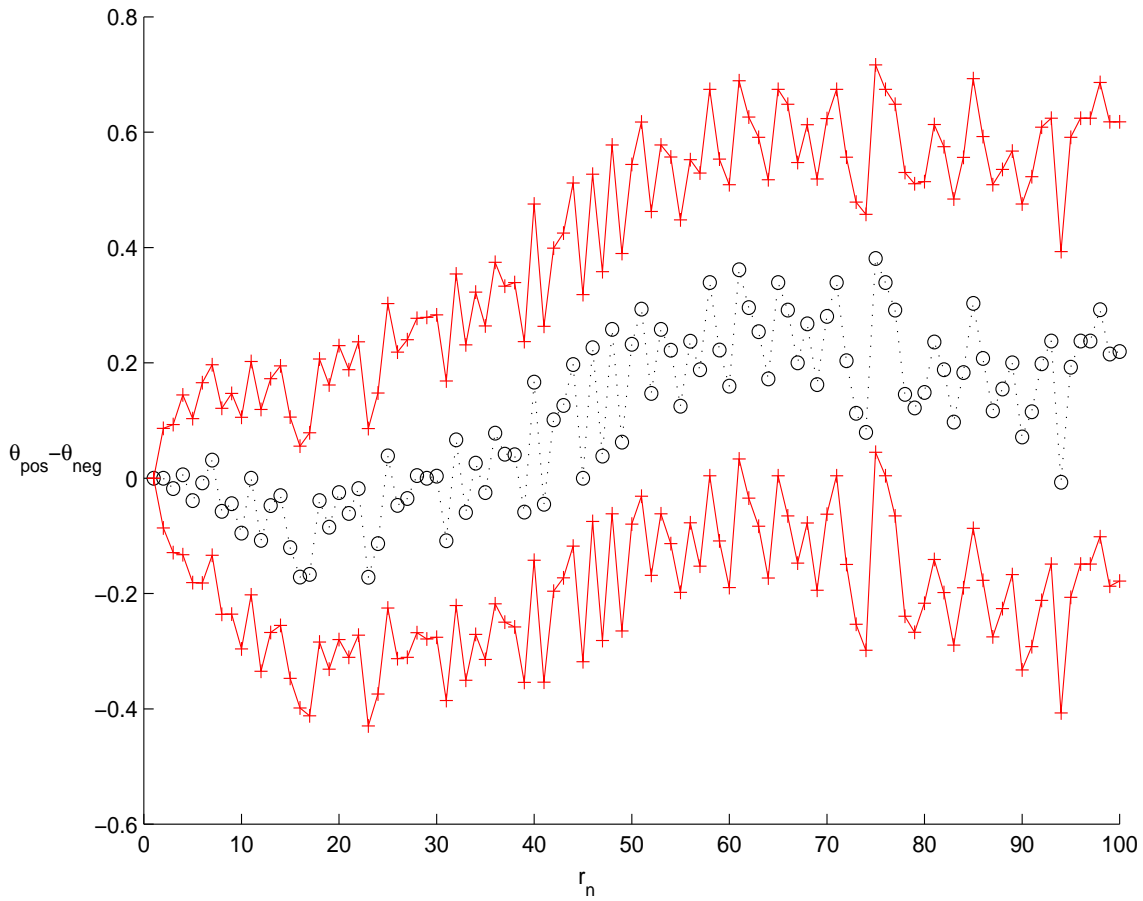


**Figura 3.7.7.** Los rendimientos del índice DaX se representan en el panel superior. Los rendimientos al cuadrado mostrando los patrones de volatilidad se representan en el panel inferior. El periodo de muestra es 19/12/1994 – 20/04/2001 ( $n = 1614$  observaciones).

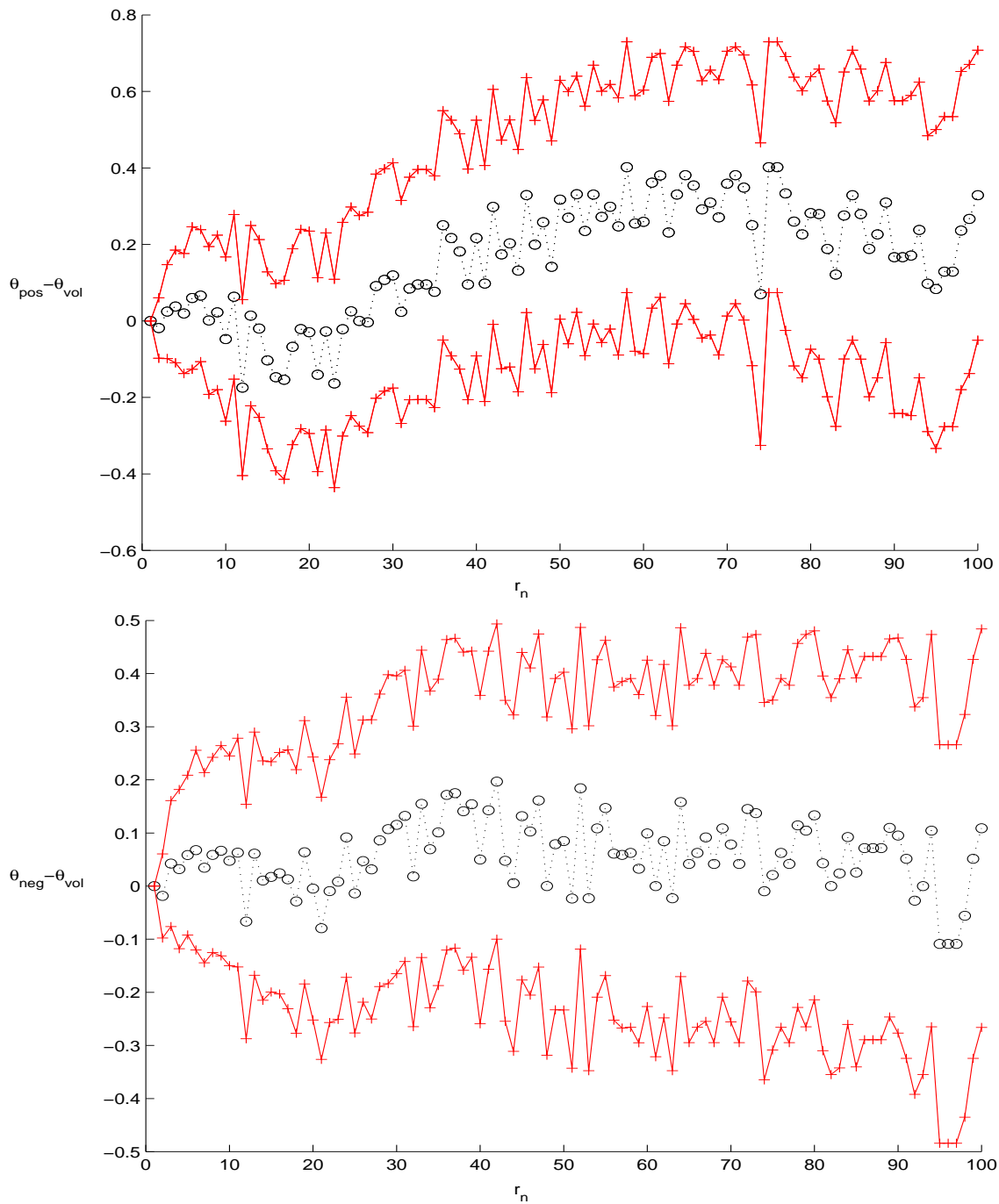


**Figura 3.7.8.** Valores estimados de  $\theta_{pos}$  y  $\theta_{neg}$  para los rendimientos del índice DaX durante el periodo 19/12/1994 – 20/04/2001 ( $n = 1614$ ). El panel superior presenta estimaciones de  $\theta_{pos}$  y el panel inferior de  $\theta_{neg}$ .  $r_n \in [1, 100]$ .  $\tilde{\theta}_n$  es representado por  $(\dots)$  y  $o$ ;  $(+-)$  describe los intervalos de confianza asintóticos con  $\alpha = 0,05$ .  $\hat{\theta}_n^{(1)}$  con  $(- - -)$  y  $\diamond$ , y  $(- \cdot -)$  y  $+$  para  $\hat{\theta}_n^{(2)}$ .  $\hat{u}_{n,pos} = x_{(n-k)}$  y  $\hat{u}_{n,neg} = x_{(k)}$  con  $k = \sqrt{2n}$  son los umbrales correspondientes.





**Figura 3.7.9.** Valores estimados de  $\theta_{pos} - \theta_{neg}$  para los rendimientos del índice DaX durante el periodo 19/12/1994 – 20/04/2001 ( $n=1614$ ).  $r_n \in [1, 100]$ .  $\tilde{\theta}_{n,pos} - \tilde{\theta}_{n,neg}$  se representa por  $(\dots)$  y  $o$ ;  $(+-)$  describe los intervalos de confianza asintóticos con  $\alpha = 0,05$ .  $\hat{u}_{n,pos} = x_{(n-k)}$  y  $\hat{u}_{n,neg} = x_{(k)}$  con  $k = \sqrt{2n}$  son los umbrales correspondientes.



**Figura 3.7.10.** Valores estimados de  $\theta_{\text{pos}} - \theta_{\text{vol}}$  (panel superior) y  $\theta_{\text{neg}} - \theta_{\text{vol}}$  (panel inferior) para los rendimientos del índice DaX durante el periodo 19/12/1994 – 20/04/2001 ( $n = 1614$ ).  $r_n \in [1, 100]$ .  $\tilde{\theta}_{n,\text{pos}} - \tilde{\theta}_{n,\text{vol}}$  y  $\tilde{\theta}_{n,\text{neg}} - \tilde{\theta}_{n,\text{vol}}$  se representa por  $(\dots)$  and  $o$ ;  $(+-)$  describe los intervalos de confianza asintóticos con  $\alpha = 0,05$ .  $u_{n,\text{pos}} = x_{(n-k)}$  es el umbral que determina los valores extremos positivos y  $u_{n,\text{neg}} = x_{(k)}$  para los valores extremos negativos, con  $k = \sqrt{2n}$ .

## Capítulo 4

# Transmisión del Riesgo en Mercados Financieros: El Efecto Contagio

No hay duda de que los mercados financieros están relacionados (interdependencia). Lo que no está tan claro es si existe contagio entre ellos o no, la intensidad, y su dirección de causalidad. El objetivo de este trabajo es definir adecuadamente el término contagio (diferente de interdependencia) y presentar un contraste formal para determinar su existencia, la magnitud de su intensidad y su dirección. Nuestra definición de contagio se basa en medidas de dependencia en las colas y es operacional mediante su equivalencia con algunas propiedades de las funciones cópula. Para hacer esto, definimos una nueva cópula, una variante del tipo Gumbel, con la suficiente flexibilidad como para describir diferentes patrones de dependencia, así como es capaz de modelizar efectos asimétricos de las correspondientes variables (algo no permitido con los modelos estándar de cópulas). Finalmente, estimamos nuestro modelo de cópulas para contrastar la intensidad y la dirección de la causalidad en los extremos entre los mercados de bonos y de acciones (en particular, el fenómeno de vuelo hacia la calidad) durante periodos de crisis. Encontramos evidencia de un efecto de sustitución entre el índice de bonos de empresas del Dow Jones con madurez 2 años y el índice de acciones de empresas que forman el Dow Jones, cuando alguno de ellos entra en etapas de crisis. Por el contrario, si ambos están atravesando periodos de crisis se observa

un efecto contagio. El análisis correspondiente para los bonos con madurez 30 *años* y el mercado de acciones refleja efectos independientes en las innovaciones.

## 4.1. Introducción

Hay acuerdo común sobre el concepto de crisis, esto es, todo el mundo detecta una crisis cuando la está atravesando. Sin embargo, las definiciones de este fenómeno son muy diferentes dependiendo de las características de la economía que se estudian. Por ejemplo un directivo de una empresa preocupado por los niveles de producción puede considerar que una empresa está atravesando una crisis si detecta una pérdida en la productividad para determinados niveles de trabajo y de capital. Por otro lado si la mayoría del negocio de la empresa está basado en las exportaciones al extranjero el gerente de la empresa estará preocupado por posibles apreciaciones de su moneda local frente a la moneda donde se comercia.

Estos ejemplos ilustran la necesidad de encontrar una definición general de crisis que agrupe los diferentes tipos de crisis sin tener en cuenta la causa. Así pues, una forma general y simple de definir una crisis en una economía puede venir dado por un umbral que representa un nivel de tolerancia. Las preguntas que surgen aquí son cómo determinar este nivel de tolerancia y cómo un valor que excede el nivel umbral puede afectar el nivel de tolerancia de otras economías o mercados relacionados en el mundo.

Esta última cuestión claramente indica que una crisis es algo más que un fenómeno aislado que afecta a mercados independientes (financieros, de crédito, de moneda). Una crisis en un mercado se caracteriza por el colapso no solo de esa economía sino por los efectos negativos que se producen en otros mercados. Por lo tanto parece natural pensar en los canales de transmisión que conectan los mercados. Desde un punto de vista económico esto implica el análisis de diferentes mecanismos que afectan el sistema: fundamentos económicos, innovaciones propias del mercado, el impacto de malas noticias, o efectos psicológicos (comportamiento de *rebaño*). La discusión surge aquí en la dirección e intensidad de la dependencia entre los mercados en periodos tumultuosos. Hay una literatura extensa sobre estas características de la dependencia. Por ejemplo Forbes y Rigobon (2001), o Corsetti, Pericoli, Sbracia (2002) donde los conceptos de interdependencia y contagio se analizan en detalle. Desde el punto de vista de la intensidad de la dependencia el contagio implica que

los vínculos entre mercados son más fuertes después de un shock en un mercado, mientras que la interdependencia implica que no hay cambios significativos en estas relaciones. Desde el punto de vista de la dirección el contagio implica que el colapso de un mercado produce la caída de otros mercados mientras la interdependencia implica que ambos mercados colapsan porque ambos están influenciados por los mismos factores.

Desde un punto de vista estadístico, los vínculos entre los mercados normalmente se miden mediante la correlación de Pearson. Baig y Goldfajn (1998) comparan la correlación entre dos mercados para un periodo anterior a la crisis y un periodo posterior a la crisis determinado por un shock. Encuentran que hay un incremento en la correlación cruzada después de una crisis y por tanto existe un efecto de contagio. Esta correlación condicional, sin embargo, no conlleva información adecuada sobre el incremento en la dependencia. Forbes y Rigobon (2001) proponen una medida de correlación ajustada que corrige el problema de condicionar a periodos de crisis donde la correlación entre los mercados es positivamente sesgada porque la volatilidad en la variable en la que condicionas (mercado en crisis) es mayor, incluso si los vínculos entre los mercados se mantienen constantes. Estos autores encuentran que la dependencia cruzada entre los mercados apenas se ve alterada después de un shock, así que hay interdependencia pero no hay contagio. Corsetti, Pericoli y Sbracia (2002) encuentran algo en el medio, algunas veces contagio y otras interdependencia. Consideran que la ausencia de contagio encontrada en Forbes y Rigobon (2001) se puede atribuir a fallos en el procedimiento del contraste.

La correlación, por tanto, puede llevar a resultados erróneos o al menos a diferentes interpretaciones dependiendo de la manera de usarla. Esta medida sólo presenta un cuadro completo de la estructura de dependencia entre los mercados cuando sus correspondientes variables aleatorias son gaussianas conjuntamente. Bajo este supuesto las correlaciones cruzadas son suficientes para describir completamente la estructura de dependencia entre las variables aleatorias. En este marco los modelos GARCH multivariantes son suficientes para describir las dinámicas (co-movements) del vector de variables aleatorias. Hay muchas especificaciones de estos modelos, sin embargo una especificación natural viene dada por la extensión de los modelos GARCH univariantes, esto es, las varianzas y covarianzas son

funciones lineales de los cuadrados y de los productos cruzados de los datos. Engle y Kroner (1995) proponen el modelo *vec* que en el caso de orden uno es,

$$vec(\Sigma_t) = vec(\Omega) + A vec(X_{t-1} X'_{t-1}) + B vec(\Sigma_{t-1}),$$

donde  $A, B$  son  $m^2 \times m^2$  matrices con ciertas restricciones, con  $m$  el número de variables aleatorias. Para  $m = 2$ ,  $vec(\Sigma_t) = (\sigma_{1t}^2, \sigma_{12t}, \sigma_{21t}, \sigma_{2t}^2)$ ,  $\sigma_{it}$ ,  $i = 1, 2$  son las volatilidades condicionadas y  $\sigma_{12t}$ ,  $\sigma_{21t}$  las varianzas condicionales en el momento  $t$ . Engle y Kroner (1995) también introducen los modelos BEKK, que en el caso de primer orden se pueden escribir como

$$\Sigma_t = \Omega + A X_{t-1} X'_{t-1} A' + B \Sigma_{t-1} B',$$

donde  $A, B$  son matrices  $m \times m$ . Estos modelos son realmente complejos: el número de parámetros a estimar para los modelos *vec* de orden 1 es  $2m^4$ , y para el modelo BEKK es  $2m^2$ . Además, a menos que las observaciones sean conjuntamente gaussianas las correlaciones cruzadas no son capaces de describir completamente los patrones de dependencia y por lo tanto cierta dependencia no está recogida en estos modelos. Consideremos por ejemplo los vínculos asimétricos correspondientes a la cola izquierda y derecha, encontrados entre la mayoría de los rendimientos de series financieras. Estos hechos estilizados están lejos de ser explicados por estos modelos.

Engle (1999) propone modelos de correlación dinámica (DCC) que extienden los modelos de correlación condicional constante (CCC) introducidos en Bollerslev (1990). La vocación de los DCC es modelizar la estructura de dependencia entre un vector de variables aleatorias ( $m=2$ ) por medio de la correlación condicionada que se permite que varíe en el tiempo. Primero se modeliza de forma individual la dependencia temporal de cada variable mediante GARCH, volatilidad estocástica (SV), y luego se modeliza la dependencia cruzada entre las innovaciones mediante otro modelo univariante (suavizado exponencial, GARCH, etc.)

$$\left. \begin{aligned} X_{i,t} &= \epsilon_{it}\sigma_{it}, \quad i = 1, 2, \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i X_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2, \end{aligned} \right\} \text{ y } \rho_t^2 = \omega_o + \alpha_o \epsilon_{1,t-1} \epsilon_{2,t-1} + \beta_o \rho_{t-1}^2,$$

con  $\rho_t$  la correlación condicional, y  $\omega_i$ ,  $\alpha_i$ ,  $\beta_i$ ,  $i = o, 1, 2$  los parámetros correspondientes de los procesos GARCH.

La propiedad de martingala se impone en el vector de innovaciones, *i.e.*  $E[\epsilon_{it} | \mathfrak{S}_{i,t-1}] = 0$ ,  $i = 1, 2$ , con  $\mathfrak{S}_{i,t-1}$  el conjunto de información disponible en el momento  $t - 1$  para cada variable aleatoria.

Estos supuestos no excluyen el caso  $E[\epsilon_{it} | \mathfrak{S}_{1,t-1} \cup \mathfrak{S}_{2,t-1}] \neq 0$  (causalidad de Granger, Granger (1969)) y el tipo de especificaciones para la correlación condicional considerados en Engle (1999) no son suficiente para explicar estos vínculos entre las variables aleatorias. Por lo tanto se requieren modelos más complejos tal que las innovaciones satisfagan  $E[\epsilon_{it} | \mathfrak{S}_{1,t-1} \cup \mathfrak{S}_{2,t-1}] = 0$ ,  $i = 1, 2$ . Sin embargo este supuesto no nos libra de diferentes formas de dependencia temporal en el vector de innovaciones  $(\epsilon_{1t}, \epsilon_{2t})$ . En cambio, deberíamos analizar la estructura global de dependencia entre las innovaciones. Esto es dado por las funciones cópula derivadas de la distribución bivalente  $H_t(\epsilon_{1t}, \epsilon_{2t})$ , y por su correspondiente condicional obtenido de  $H_t(\epsilon_{1t}, \epsilon_{2t} | \mathfrak{S}_{1,t-1} \cup \mathfrak{S}_{2,t-1})$ , ver Patton (2001) ó Granger, Terasvirta y Patton (2002).

La definición de cópula se debe a Sklar (1959). Esta función proporciona la estructura completa de dependencia entre las variables aleatorias después de tener en cuenta las distribuciones marginales correspondientes. En particular, el modelo introducido en Engle (1999) se puede considerar una cópula gaussiana donde las dinámicas de dependencia vienen dadas por la correlación condicional. Se podría decir que el conjunto de cópulas disponibles no tiene fin dando lugar a diferentes elecciones ajustándose al problema dado. Ejemplos de cópulas son presentados en Longin and Solnik (2001) donde utilizan la cópula gaussiana para describir la dependencia en los rendimientos de activos financieros, cópulas de  $t$ -Student (Mashal y Zeevi, (2002)) que se ajustan mejor a las colas de estas secuencias, la cópula de Joe-Clayton introducida en Joe (1997) o su variante, la cópula simétrica de Joe-Clayton introducida en Patton (2001) para la dependencia entre mercados de cambio.



Es interesante también analizar los vínculos entre las variables aleatorias en las regiones de la cola. Su función de distribución conjunta en la región de la cola se deriva de la teoría de valores extremos multivariante, ver Resnick (1987). Diferentes aplicaciones de estas distribuciones multivariantes de valores extremos a ejemplos relacionados con las colas aparecen en Ledford y Tawn (1996) ó en de Haan y de Ronde (1998). El análisis de la dependencia en los extremos proporciona una alternativa interesante a la correlación para medir la fuerza de los vínculos entre variables aleatorias cuando éstas se hacen más extremas (diferencias entre interdependencia y contagio).

La vocación de este capítulo es modelizar la dependencia encontrada entre las variables aleatorias que representan economías y mercados financieros diferentes. Esta dependencia se divide en dos clases de acuerdo a su origen. Primero, los vínculos debidos a fundamentos económicos (dependencia racional) y segundo, los co-movimientos de las correspondientes innovaciones (dependencia irracional). Nuestro enfoque se encuentra en la última forma de dependencia y los conceptos de interdependencia y contagio. Para modelizar esta forma de dependencia (dependencia cruzada en las secuencias de las innovaciones) introducimos una innovadora función cópula derivada de la teoría de valores extremos que incorpora suficiente flexibilidad para describir diferentes patrones de dependencia, en particular efectos asimétricos entre las variables no recogidas por las cópulas estándar. Más aún, los conceptos de interdependencia y contagio se revisan y las definiciones propuestas en la literatura se adaptan para ser expresadas como medidas de dependencia en las colas, y por tanto propiedades de las funciones cópula envolviendo a las colas de las distribuciones marginales. Finalmente, la intención de los autores es aplicar esta metodología para contrastar el fenómeno de vuelo hacia la calidad, esto es, salidas del capital de los mercados de acciones a los mercados de bonos cuando los primeros afrontan periodos de crisis.

Este capítulo se estructura como sigue. La Sección 2 introduce la función cópula derivada de la teoría de valores extremos multivariante. La sección siguiente propone medidas de dependencia en la cola como una alternativa a la correlación; estas medidas se utilizan para definir contagio e interdependencia. Los casos de dependencia asintótica e independencia se estudian también en este apartado. Finalmente la sección estudia los aspectos estadísticos

del modelo y proporciona un contraste para la existencia de estos efectos. En la Sección 4, esta innovadora función cópula así como las nuevas medidas de dependencia se aplican para analizar la estructura de dependencia entre los mercados de bonos y acciones (fenómeno de vuelo hacia la calidad). La Sección 5 presenta las conclusiones.

## 4.2. El modelo

Consideremos el modelo

$$\left. \begin{aligned} X_{1,t} &= g_1(X_{1,t-1}, \dots, X_{m,t-1}) + \varepsilon_{1,t}, \\ &\dots \quad \dots \quad \dots \\ X_{m,t} &= g_m(X_{1,t-1}, \dots, X_{m,t-1}) + \varepsilon_{m,t}, \end{aligned} \right\} \quad (4.1)$$

y supongamos que  $(\varepsilon_{1,1}, \dots, \varepsilon_{m,1}), \dots, (\varepsilon_{1,t}, \dots, \varepsilon_{m,t})$  son vectores independientes, esto es, la dependencia multivariante entre las innovaciones viene dada por  $H_t((\varepsilon_1, \dots, \varepsilon_m) | \mathfrak{S}_{t-1})$ , con  $\mathfrak{S}_{t-1} = \mathfrak{S}_{X_{1,t-1}} \cup \dots \cup \mathfrak{S}_{X_{m,t-1}}$ . Notemos que la estructura de dependencia varía en el tiempo aunque las distribuciones marginales de las observaciones no dependan del tiempo. De otro modo si las innovaciones verificaran la propiedad de martingala, las distribuciones marginales no estarían libres del índice  $t$ , esto es, la función de distribución conjunta tomaría esta forma

$$H_t((\varepsilon_{1,t} | \mathfrak{S}_{t-1}, \dots, \varepsilon_{m,t} | \mathfrak{S}_{t-1}) | \mathfrak{S}_{t-1}),$$

con  $\varepsilon_{i,t} | \mathfrak{S}_{t-1}$ ,  $i = 1, \dots, m$  las variables aleatorias condicionales.

Ambas funciones de distribución, sin embargo, dan lugar al tipo de cópulas condicionales introducidas en Patton (2001) donde la dinámica de la función de distribución conjunta está dirigida por un parámetro que varía en el tiempo. En cambio, para funciones adecuadas  $g_1, \dots, g_m$  proponemos una función de distribución multivariante  $H(\varepsilon_1, \dots, \varepsilon_m)$  que no depende del tiempo y que está motivada por la dependencia encontrada entre el vector de máximos de las correspondientes variables aleatorias.

### 4.2.1. La estructura de dependencia: La función cópula

Esta sección estudia la estructura de dependencia entre  $m$  variables aleatorias vía la función cópula. El concepto de cópula se debe a Sklar (1959) y se refiere a la clase de funciones de distribución multivariante definidas en el cubo unidad con marginales uniformes.

**Definición 4.2.1.** *Una función  $C : [0, 1]^m \rightarrow [0, 1]$  es una cópula  $m$ -dimensional si verifica las siguientes propiedades:*

- Para todo  $u_i \in [0, 1]$ ,  $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ .
- Para todo  $u \in [0, 1]^m$ ,  $C(u_1, \dots, u_m) = 0$  si al menos una de sus coordenadas es cero.
- El volumen de toda caja contenida en  $[0, 1]^m$  es no negativo, esto es,  $V_C([u_1, \dots, u_m] \times [v_1, \dots, v_m])$  es no negativo. Para  $m = 2$ ,  $V_C([u_1, u_2] \times [v_1, v_2]) = C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0$  para  $0 \leq u_i, v_i \leq 1$ .

La cópula  $C(u_1, \dots, u_m)$  es la función de distribución conjunta de las transformaciones integrales de la función de probabilidad para cada una de las variables  $X_1, \dots, X_m$  con respecto a las distribuciones marginales  $F_1, \dots, F_m$ . Se puede ver como el componente de la función de distribución multivariante de un vector de variables aleatorias que captura la estructura de dependencia.

**Teorema 4.2.1.** *(teorema de Sklar): Dada una función de distribución  $m$ -dimensional  $H$  con distribuciones marginales continuas  $F_1, \dots, F_m$ , existe una única cópula  $C : [0, 1]^m \rightarrow [0, 1]$  tal que*

$$H(x_1, \dots, x_m) = C(F_1(x_1), \dots, F_m(x_m)), \quad \forall x_1, \dots, x_m \in \mathbb{R} \cup \{\infty\}. \quad (4.2)$$

La implicación en la otra dirección también se cumple, esto es, si  $C(u_1, \dots, u_m)$  es una función de distribución  $m$ -dimensional con marginales uniformes y  $F_1, \dots, F_m$  son funciones de distribución univariante continuas para las variables aleatorias  $X_1, \dots, X_m$ , entonces la

función  $H$  definida en (4.2) es una función de distribución  $m$ -dimensional con marginales  $F_1, \dots, F_m$ .

Es inmediato ver que si tenemos un modelo para la distribución conjunta de las  $m$  variables aleatorias y tenemos las distribuciones marginales de las  $X_i$  y son continuas, la estructura completa de dependencia de las variables aleatorias correspondientes es conocida,

$$C(u_1, \dots, u_m) = H(F_1^{-1}(u_1), \dots, F_m^{-1}(u_m)), \quad (4.3)$$

con  $F_i^{-1}(u) = \inf\{x \in \mathbb{R} | F_i(x) \geq u\}$ , para todo  $0 \leq u \leq 1$ .

Esta medida de dependencia extiende las nociones de correlación lineal (Pearson) y correlación de los rangos (Spearman). Más importante, soluciona los problemas típicos de estas medidas escalares. Embrechts, McNeil y Straumann (1999) dan un excelente resumen acerca de las propiedades y problemas de estas medidas de dependencia.

Se muestra que bajo condiciones muy generales sobre las funciones de distribución marginales la estructura de dependencia de cualquier distribución multivariante se describe mediante una función cópula. En particular este interesante resultado se encuentra para la distribución conjunta del máximo de un vector de variables aleatorias. Más aún, existe una función cópula que dirige esta dependencia en los extremos y cuya expresión se deriva de la teoría de valores extremos.

Consideremos  $M_n = (M_{n1}, \dots, M_{nm})$  el vector compuesto por los máximos, es decir los componentes son  $M_{ni} = \max\{X_{1i}, \dots, X_{ni}\}$ , y el vector de secuencias  $a_n = (a_{n1}, \dots, a_{nm})$  con cada  $a_{ni} > 0$ , y  $b_n = (b_{n1}, \dots, b_{nm})$ . Bajo cierta condición de suavidad en la cola de  $F_i$ , Leadbetter, Lindgren y Rootzén (1983) muestran que

$$\lim_{n \rightarrow \infty} F_i^n(a_{ni}x_i + b_{ni}) = G_i(x_i), \quad i = 1, \dots, m, \quad (4.4)$$

donde  $G_i(x_i)$  es una distribución de valores extremos de uno de los siguientes tres tipos posibles, Gumbel, Weibull y Fréchet. La distribución  $F_i$  se dice que pertenece al dominio de atracción del máximo de  $G_i$ , ver también Embrechts, Klüppelberg y Mikosch (1997).

Denotemos la distribución del máximo multivariante por

$$H^n(a_{n1}x_1 + b_{n1}, \dots, a_{nm}x_m + b_{nm}) = P\{a_{ni}^{-1}(M_{ni} - b_{ni}) \leq x_i, i = 1, \dots, m\}, \quad (4.5)$$

donde  $H(x_1, \dots, x_m) = P\{X_1 \leq x_1, \dots, X_m \leq x_m\}$ . El resultado fundamental de la teoría de valores extremos multivariante es que (4.4) puede ser extendido a

$$\lim_{n \rightarrow \infty} H^n(a_{n1}x_1 + b_{n1}, \dots, a_{nm}x_m + b_{nm}) = G(x_1, \dots, x_m), \quad (4.6)$$

con  $G$  una función de distribución extrema multivariante no degenerada (*mevdf*). La clase de estas distribuciones particulares es precisamente la clase de distribuciones max-estables (Resnick (1987), proposición 5,9). Estas distribuciones se definen por esta propiedad

$$G^t(tx_1, \dots, tx_m) = G(\alpha_1x_1 + \beta_1, \dots, \alpha_mx_m + \beta_m), \quad (4.7)$$

para todo  $t > 0$ , y ciertas constantes  $\alpha_i > 0$  and  $\beta_i$ .

Las funciones de distribución marginales de  $G$  son las distribuciones extremas univariantes  $G_i(x_i)$ . Por el teorema de Sklar, (4.6) se puede escribir como

$$\lim_{n \rightarrow \infty} H^n(a_{n1}x_1 + b_{n1}, \dots, a_{nm}x_m + b_{nm}) = C(G_1(x_1), \dots, G_m(x_m)), \quad (4.8)$$

con  $C$  una función cópula.

Se puede ver haciendo simple álgebra que  $C$  también describe la estructura de dependencia de las observaciones más grandes. Nuestro objetivo en las siguientes líneas es derivar una expresión analítica que se ajuste a esta función cópula. Para hacer esto, las distribuciones marginales se transforman para obtener versiones idénticas y libres de parámetros de estas distribuciones univariantes, en particular funciones de distribución del tipo Fréchet  $\Psi_\alpha(z) = \exp(-z^{-\alpha})$  con  $\alpha = 1$ .

Sea  $Z_i = 1/\log \frac{1}{F_i(X)}$  tal transformación, y denotemos  $P\{Z_i \leq z\} = F_i^*(z)$  con  $z = 1/\log \frac{1}{F_i(a_{ni}x + b_{ni})}$ . Esta distribución satisface estas interesantes propiedades:  $F_i^*(z) = \Psi_1(z)$ ,

$F_i^*(z) = F_i(a_{ni}x + b_{ni})$  y  $F_i^{*n}(nz) = \Psi_1(z)$ . Notemos que estas condiciones implican que  $F_i^{*n}(nz) = F_i^*(z)$  y

$$\lim_{n \rightarrow \infty} H^{*n}(nz_1, \dots, nz_m) = C(\Psi_1(z_1), \dots, \Psi_1(z_m)), \quad (4.9)$$

con  $H^*(z_1, \dots, z_m) = H(a_{n1}x_1 + b_{n1}, \dots, a_{nm}x_m + b_{nm})$ . Esta condición se cumple para todo vector  $(z_1, \dots, z_m)$  en  $[z_{o1}, \infty) \times \dots \times [z_{om}, \infty)$ , con  $(z_{o1}, \dots, z_{om})$  un vector umbral. La función  $C$  se llama cópula extrema porque verifica esta propiedad,

$$C^t(\Psi_1(z_1), \dots, \Psi_1(z_m)) = C(\Psi_1^t(z_1), \dots, \Psi_1^t(z_m)), \quad t > 0, \quad (4.10)$$

donde las marginales son distribuciones de tipo extremo. La demostración sale de forma inmediata de la fórmula (4.7). Esta condición conlleva una propiedad de invarianza dada por los logaritmos de las correspondientes distribuciones, esto es,

$$t \log C(\Psi_1(tz_1), \dots, \Psi_1(tz_m)) = \log C(\Psi_1(z_1), \dots, \Psi_1(z_m)).$$

Entonces, para  $n$  y  $(z_1, \dots, z_m)$  suficientemente grande,

$$\lim_{n \rightarrow \infty} n(1 - H^*(nz_1, \dots, nz_m)) = -\log C(\Psi_1(z_1), \dots, \Psi_1(z_m)), \quad (4.11)$$

y

$$\lim_{n \rightarrow \infty} \frac{H^*(nz_1, \dots, nz_m)}{1 + \log C(\Psi_1(nz_1), \dots, \Psi_1(nz_m))} = 1. \quad (4.12)$$

Otro interesante resultado derivado de la propiedad de invarianza es

$$P \left\{ Z_1 \leq nz_1, \dots, Z_m \leq nz_m \mid \bigcup_{i=1}^m Z_i > nz_{0i} \right\} = P \left\{ Z_1 \leq z_1, \dots, Z_m \leq z_m \mid \bigcup_{i=1}^m Z_i > z_{0i} \right\}.$$

El término de la izquierda de (4.11) se puede considerar como una secuencia de medidas que converge a una constante dado el vector  $(z_1, \dots, z_m)$ , ver Resnick (1987) ó de Haan y de Ronde (1998) para diferentes transformaciones de las distribuciones marginales. La ex-

presión (4.12) ofrece la función de distribución conjunta de las observaciones más grandes, esto es, para  $n$  suficientemente grande el denominador se puede aproximar por la función cópula  $C$ . Por lo tanto

$$P\{Z_1 \leq z_1, \dots, Z_m \leq z_m\} = C(\Psi_1(z_1), \dots, \Psi_1(z_m)), \quad (4.13)$$

para el vector  $(z_1, \dots, z_m)$  suficientemente grande.

Esta última expresión es prometedora en el sentido que  $C$  es una buena aproximación de la estructura de dependencia de las observaciones más grandes. Sin embargo, el reto de elegir un umbral adecuado que determine la región en la que se verifica la condición (4.11) todavía se mantiene.

Por otro lado la propiedad de invarianza implica que la cópula  $C$  debe ser de tipo exponencial. Hay diferentes caracterizaciones de esta distribución. Una expresión general para  $m = 2$  viene dada en la forma de la representación de Pickands, (Pickands, 1981) esto es,

$$C(u_1, u_2) = \exp^{D(t)\log(u_1 u_2)}, \quad (4.14)$$

donde  $u_1 = \Psi_1(z_1)$ ,  $u_2 = \Psi_1(z_2)$ ,  $t = \frac{\log(u_1)}{\log(u_1 u_2)}$ , y  $D(t)$  es una función convexa en  $[0, 1]$  tal que  $\max(t, 1 - t) \leq D(t) \leq 1$  para todo  $0 \leq t \leq 1$ . Esta familia de distribuciones está incluida en la clase de cópulas arquimedianas (Nelsen 1999, capítulo IV). La dependencia en estas cópulas esta dirigida por una única variable  $t$  para  $m = 2$ . La familia Gumbel-Hougaard está dentro de esta clase de distribuciones y satisface la propiedad de invarianza. Se representa por

$$C_G(u_1, u_2; \theta) = \exp^{-[(-\log u_1)^\theta + (-\log u_2)^\theta]^{1/\theta}}, \quad \theta \geq 1. \quad (4.15)$$

Esta función de distribución se conoce normalmente como la cópula bivalente logística de Gumbel. El principal problema que surge si asumimos que  $C$  se modeliza por una distribución de Gumbel  $C_G$  en (4.13) es la selección del umbral. La condición (4.11) puede ser relajada (que no se verifique) para umbrales bajos donde la teoría de valores

extremos no es una técnica fiable. Otro problema de modelizar  $C$  como  $C_G$  es la asimetría, las variables aleatorias modelizadas por la cópula Gumbel son intercambiables y por lo tanto no es posible cuantificar las diferentes contribuciones de las correspondientes variables aleatorias. Para tener en cuenta la posibilidad de dependencia asimétrica proponemos una versión de  $C_G$  capaz de describir estos efectos. Esta función se denota por  $\tilde{C}_G(u_1, u_2; \Theta)$ , con  $\Theta = \{\theta, \gamma, \eta\}$ , y toma la siguiente expresión

$$\tilde{C}_G(u_1, u_2; \Theta) = \exp^{-D(u_1, u_2; \gamma, \eta)[(-\log u_1)^\theta + (-\log u_2)^\theta]^{1/\theta}}, \quad (4.16)$$

con

$$D(u_1, u_2; \gamma, \eta) = \exp^{\gamma(1-u_1)(1-u_2)^\eta}, \quad \gamma \geq 0, \quad \eta > 0. \quad (4.17)$$

**Teorema 4.2.2.** *La función  $\tilde{C}_G : [0, 1] \times [0, 1] \rightarrow [0, 1]$  definida en (4.16) y (4.17) es una función cópula si los parámetros en  $\Theta$  cumplen que  $\tilde{c}_G(u_1, u_2; \Theta) > 0$ ,  $\forall (u_1, u_2) \in [0, 1] \times [0, 1]$ , con  $\tilde{c}_G(u_1, u_2; \Theta) = \frac{\delta^2 \tilde{C}_G(u_1, u_2; \Theta)}{du_1 du_2}$  la función de densidad de la cópula  $\tilde{C}_G$ .*

*Proof.-* Denotemos  $A(u_1, u_2; \theta) = [(-\log u_1)^\theta + (-\log u_2)^\theta]^{1/\theta}$ . Las condiciones relacionadas al contorno de  $\tilde{C}_G$  se verifican de forma inmediata de las propiedades del contorno de las funciones  $D(u_1, u_2; \gamma, \eta)$  y  $A(u_1, u_2; \theta)$ . La demostración de que  $\tilde{C}_G$  es 2-creciente implica el uso de más álgebra. Consideremos  $V_{\tilde{c}_G}([u_{o1}, u_{11}] \times [u_{o2}, u_{12}]) = \tilde{C}_G(u_{11}, u_{12}; \Theta) - \tilde{C}_G(u_{11}, u_{o2}; \Theta) - \tilde{C}_G(u_{o1}, u_{12}; \Theta) + \tilde{C}_G(u_{o1}, u_{o2}; \Theta)$ , y definamos  $V'(u_1) = \tilde{C}_G(u_1, u_{12}; \Theta) - \tilde{C}_G(u_1, u_{o2}; \Theta)$ . Entonces,

$V_{\tilde{c}_G}([u_{o1}, u_{11}] \times [u_{o2}, u_{12}]) = V'(u_{11}) - V'(u_{o1})$ . Notemos que  $V'(u_1) \geq 0$ ,  $\forall u_1 \in [0, 1]$ , con  $u_{o2} < u_{12}$ . Esta función se puede escribir como

$$V'(u_1) = \exp^{-D(u_1, u_{o2}; \gamma, \eta)A(u_1, u_{o2}; \theta)} \left[ \exp^{-D(u_1, u_{12}; \gamma, \eta)A(u_1, u_{12}; \theta) - D(u_1, u_{o2}; \gamma, \eta)A(u_1, u_{o2}; \theta)} - 1 \right],$$

que es mayor que 0 si y solo si  $D(u_1, u_2; \gamma, \eta)A(u_1, u_2; \theta)$  es decreciente en  $u_2$ . La única condición que queda por ver es que  $V'(u_1)$  es no decreciente. Esta condición se verificará si la función  $\frac{\delta \tilde{C}_G(u_1, u_2; \Theta)}{du_1}$  es no decreciente en  $u_2$ , que equivale a ver  $\tilde{c}_G(u_1, u_2; \Theta) > 0$ ,  $\forall (u_1, u_2) \in$



$[0, 1] \times [0, 1]$ .  $\square$

La elección del umbral en (4.13) se soluciona *añadiendo* la función  $D(u_1, u_2; \gamma, \eta)$ . Esta función por medio del par  $(u_1, u_2)$  y del parámetro  $\gamma$  mide la sensibilidad de la estructura de dependencia a alejamientos de la propiedad de invarianza. En otras palabras, o los valores de las funciones marginales se encuentran en las colas de la derecha ( $u_1, u_2 \rightarrow 1$ ) o  $\gamma \cong 0$  la función cópula  $\tilde{C}_G$  está muy próxima a  $C_G$  y la propiedad de invarianza se verifica. De esta forma el función de distribución conjunta para el rango entero de las variables aleatorias  $Z_1, Z_2$  es

$$P \{Z_1 \leq z_1, Z_2 \leq z_2\} = \tilde{C}_G(\Psi_1(z_1), \Psi_1(z_2)), \quad (4.18)$$

donde  $z_i = 1/\log \frac{1}{F_i(x_i)}$  en este caso. Esta función de distribución es dirigida por los parámetros  $\theta, \gamma, \eta$ . La constante  $\gamma$  dice hasta qué punto la propiedad de invarianza se cumple. El parámetro  $\theta$  describe el nivel de dependencia asintótica en la cola de las variables aleatorias. El caso de independencia perfecta está cubierto por  $\theta = 1, \gamma = 0$ . Finalmente  $\eta$  mide el nivel de asimetría ó intercambiabilidad de las variables.

La lista siguiente enumera las ventajas más importantes de nuestra función cópula  $\tilde{C}_G$ .

1. Esta función cópula se deriva de la teoría de valores extremos multivariante, en contraste con selecciones ad-hoc para modelizar la estructura de dependencia.
2. La función  $D(u_1, u_{12}; \gamma, \eta)$  y en particular el parámetro  $\gamma$  extienden los resultados de la teoría de valores extremos multivariante sobre la distribución de las observaciones más grandes al rango entero de las variables aleatorias.
3.  $\tilde{C}_G$  es capaz de explicar los efectos asimétricos de las variables para  $\eta \neq 1$ . Se puede considerar como una alternativa al modelo asimétrico logístico de Tawn (1988).
4. Esta función cópula tiene suficiente flexibilidad para describir diferentes formas de dependencia y de dependencia asintótica, como se muestra abajo.

### 4.3. Contagio: tipos y definiciones

Las medidas lineales de dependencia no son suficientes para describir los patrones de dependencia entre un vector de variables aleatorias. La popular medida de correlación de Pearson tiene toda una serie de problemas y dificultades, ver Embrechts, McNeil y Straumann (1999). Algunas de ellas son que una correlación de cero no implica independencia si las distribuciones marginales no son elípticas, y segundo, la correlación no es invariante bajo transformaciones de las variables aleatorias. La correlación de Spearman (correlación en los rangos) por ejemplo soluciona este último problema, sin embargo, también falla en proporcionar medidas de independencia fuera del mundo elíptico.

En el marco bivalente medidas naturales de dependencia diferentes de las tradicionales medidas de correlación están dadas por la dependencia en las colas. Ledford y Tawn (1997) y Coles, Heffernan y Tawn (1999) definen la medida de dependencia asintótica en las colas  $\aleph$ ,

$$\aleph = \lim_{t \rightarrow \infty} P\{Z_2 > t | Z_1 > t\}. \quad (4.19)$$

Esta medida toma el valor cero si las variables aleatorias son asintóticamente independientes. Hay dos clases de dependencia en los extremos, dependencia asintótica e independencia asintótica. Ambas formas de dependencia permiten dependencia para valores moderadamente grandes de las variables, sin embargo la probabilidad de sucesos extremos conjuntos bajo independencia asintótica converge a cero cuando los sucesos son más extremos. La probabilidad de que una variable sea extrema dado que la otra es extrema es cero en el límite. La cópula  $\tilde{C}_G$  admite ambas formas de dependencia asintótica. Se puede ver que  $\aleph_{\tilde{C}_G} = 2 - 2^{1/\theta}$ , que recoge independencia asintótica para  $\theta = 1$  y dependencia asintótica en cualquier otro caso.

La definición en (4.19) se puede extender al rango entero de las variables aleatorias. Lehman (1966) definió dos variables aleatorias  $Z_1, Z_2$  como positivamente dependientes en el cuadrante (*PQD*) si para todo  $(z_1, z_2) \in \mathbb{R}^2$ ,

$$P\{Z_1 > z_1, Z_2 > z_2\} \geq P\{Z_1 > z_1\}P\{Z_2 > z_2\}, \quad (4.20)$$

o equivalentemente

$$P\{Z_1 \leq z_1, Z_2 \leq z_2\} \geq P\{Z_1 \leq z_1\}P\{Z_2 \leq z_2\}. \quad (4.21)$$

De la misma forma dependencia negativa en el cuadrante (*NQD*) se define dando la vuelta a las desigualdades en ambas expresiones.

**Definición 4.3.1.** *Decimos que dos variables aleatorias son interdependientes si son PQD. En consecuencia interdependencia se caracteriza por movimientos conjuntos en la misma dirección (co-movements) de las correspondientes variables aleatorias.*

Si  $Z_1$  y  $Z_2$  son *NQD* un valor grande en una variable es correspondido por un valor grande de la misma magnitud en la dirección opuesta para la otra variable. Económicamente, interdependencia significa que los lazos en periodos de crisis (definidos en las colas de la distribución) son sólo consecuencia de los mismos vínculos existentes en periodos de calma.

En el caso de que las variables aleatorias sean continuas estas definiciones son una propiedad de la cópula. De la teoría de la probabilidad elemental

$$P\{Z_1 > z_1, Z_2 > z_2\} = \tilde{C}_G(u_1, u_2) - (u_1 + u_2) + 1, \quad (4.22)$$

con  $u_i = \Psi_1(z_i)$ ,  $i = 1, 2$ .

Definamos la función  $g(u_1, u_2)$  como la diferencia entre las probabilidades en (4.20) en términos de la función cópula,

$$g(u_1, u_2) = \tilde{C}_G(u_1, u_2) - u_1 u_2. \quad (4.23)$$

Si esta función es positiva para todo  $(u_1, u_2) \in [0, 1] \times [0, 1]$  las definiciones anteriores para dependencia cruzada son válidas, esto es,  $Z_1$  and  $Z_2$  son *PQD*.

La función  $g(u_1, u_2)$  por si misma no es suficiente para determinar la fuerza de los vínculos entre las variables. Una condición más fuerte se requiere para medir la canti-

dad de dependencia para valores diferentes de las variables aleatorias. Esta condición es monotonicidad en las colas, esto es, la función (4.23) es no-creciente ó no-decreciente en sus argumentos. En particular, monotonicidad creciente en la cola para la función  $P\{Z_1 > z_1, Z_2 > z_2\} - P\{Z_1 > z_1\}P\{Z_2 > z_2\}$  caracteriza la existencia de contagio en las colas de la derecha entre las variables aleatorias. Así pues, contagio en este contexto se puede definir como un incremento significativo en la intensidad de la dependencia entre las variables  $Z_1, Z_2$  cuando éstas toman valores extremos.

**Definición 4.3.2.** *Supongamos  $Z_1, Z_2$  con función común de distribución Fréchet  $\Psi_1$  y consideremos  $z$  un umbral que determina los extremos en la cola derecha de ambas variables. Entonces, existe un efecto contagio entre  $Z_1$  y  $Z_2$  si  $g(u_1, u_2)$  es una función creciente para ambas variables aleatorias, y para  $u_1, u_2 \geq u$  con  $u = \Psi_1(z)$ .*

Por otro lado contagio en intensidad para las colas de la izquierda se caracteriza por monotonicidad decreciente en las colas para la función  $P\{Z_1 \leq z_1, Z_2 \leq z_2\} - P\{Z_1 \leq z_1\}P\{Z_2 \leq z_2\}$ . En términos de cópulas las condiciones en la definición 4.3.2 de contagio equivale a ver estas propiedades,

$$h_1(u_1, u_2) = \frac{\delta \tilde{C}_G(u_1, u_2)}{du_1} - u_2 > 0, \quad h_2(u_1, u_2) = \frac{\delta \tilde{C}_G(u_1, u_2)}{du_2} - u_1 > 0. \quad (4.24)$$

La presencia de monotonicidad en la cola para el rango entero de las variables aleatorias indica algo más fuerte que contagio. Estas propiedades, llamadas Cola creciente a la derecha, Right Tail Increasing (*RTI*), y Cola creciente a la izquierda, Left Tail Decreasing (*LTD*), en Esary y Proschan (1972), implican que  $P\{Z_2 > z_2 | Z_1 > z_1\} > P\{Z_2 > z_2\}$  y  $P\{Z_2 \leq z_2 | Z_1 \leq z_1\} > P\{Z_2 \leq z_2\}$  respectivamente, para cualquier par  $(z_1, z_2)$ , y por lo tanto son sinónimos de contagio e interdependencia. Notemos sin embargo que estos fenómenos no aparecen necesariamente juntos. Puede existir contagio en los extremos entre dos variables aleatorias sin que sean interdependientes, y por otro lado, dos variables aleatorias interdependientes pueden mostrar lazos más débiles (aunque más fuertes que siendo independientes) en periodos de nerviosismo que en periodos de calma.

Los conceptos de contagio e interdependencia introducidos hasta ahora tienen en cuenta la intensidad en la dependencia, esto es, la fuerza de los lazos entre las variables cuando éstas se mueven lejos hacia las colas. Sin embargo, existen otras formas de contagio que tienen en cuenta la dirección de la dependencia, en este marco, la probabilidad condicional de (4.19) se interpreta como una relación de causalidad. Contagio en este contexto ocurre cuando una variable está influenciando a la otra, esto es, un valor grande en una variable está aumentando la probabilidad de que ocurra un valor grande en la otra variable. Entonces la relación entre las variables debe ser asimétrica, o de otra forma sólo hay un incremento en la intensidad de la dependencia (contagio como es definida en 4.3.2). Notemos sin embargo que una condición del tipo  $P\{Z_2 > z_2 | Z_1 > z_1\} > P\{Z_2 > z_2\}$  equivale a (4.20). Más aún, la única diferencia de la anterior con una condición como  $P\{Z_1 > z_1 | Z_2 > z_2\} > P\{Z_1 > z_1\}$  viene dada por las distribuciones marginales. En el caso de  $H^*(z_1, z_2)$  donde las distribuciones marginales son idénticas Fréchet, ambas probabilidades marginales son idénticas.

Centrémonos en cambio en las siguientes para la existencia de contagio,

$$P\{Z_2 > z_2 | Z_1 > z_1\} \geq P\{Z_1 > z_2 | Z_2 > z_1\}, \quad (4.25)$$

para las colas de la derecha, con  $z_2 \geq z_1$ , y

$$P\{Z_2 \leq z_2 | Z_1 \leq z_1\} \geq P\{Z_1 \leq z_2 | Z_2 \leq z_1\}, \quad (4.26)$$

para las colas de la izquierda, donde  $z_2 \leq z_1$ . Estas condiciones equivalen a ver si  $\tilde{C}_G(u_1, u_2; \Theta) > \tilde{C}_G(u_2, u_1; \Theta)$ . Consideremos  $z_1$  un valor umbral que determina los sucesos extremos, entonces esta desigualdad implica que la probabilidad de que  $Z_2$  sea extremo dado que  $Z_1$  es extremo es mayor que la probabilidad de que  $Z_1$  sea extremo con  $Z_2$  extremo. En otras palabras,  $Z_1$  está causando que  $Z_2$  alcance valores extremos. El caso particular de la igualdad en las últimas expresiones representa simetría de las variables y económicamente concierne a interdependencia en la dirección (ambas economías están afectadas por los mismos fac-

tores externos y de la misma forma).

Para formalizar la idea de contagio direccional definimos  $gd_v(u) = \tilde{C}_G(u, v) - \tilde{C}_G(v, u)$  e introducimos la siguiente definición.

**Definición 4.3.3.** *Supongamos  $Z_1, Z_2$  con función de distribución común Fréchet  $\Psi_1$  y consideremos  $z$  un umbral que determina los extremos de ambas variables aleatorias. Entonces,  $Z_1$  está influyendo en que  $Z_2$  alcance valores extremos (efecto contagio) si  $gd_v(u)$  es estrictamente positiva para todo  $v > u$  para la cola de la derecha, y para todo  $v < u$  para la cola de la izquierda, con  $u = \Psi_1(z)$ .*

Esta definición es análoga para  $Z_2$  influyendo  $Z_1$  pero invirtiendo los signos de desigualdad. En términos de los parámetros de la cópula  $\tilde{C}_G$ , hay contagio de  $Z_1$  hacia  $Z_2$  si  $\eta > 1$ , y de  $Z_2$  hacia  $Z_1$  si  $\eta < 1$ .

La definición se puede hacer más fuerte imponiendo una condición de monotonicidad en  $gd_v(u)$ . La intensidad de este tipo de contagio se puede medir mediante esta condición de monotonicidad. En particular,

**Definición 4.3.4.** *Supongamos las condiciones de la definición 4.3.3, y  $Z_1, Z_2$  tal que existe contagio positivo de  $Z_1$  a  $Z_2$ . Entonces,  $Z_1$  está afectando fuertemente  $Z_2$  en los valores extremos (efecto contagio en el sentido fuerte) si  $gd_v(u)$  es una función creciente en  $v$  para todo  $v > u$ .*

Una caracterización de esta definición es

$$sc(u, v) = \frac{\delta \tilde{C}_G(u, v; \Theta)}{dv} - \frac{\delta \tilde{C}_G(v, u; \Theta)}{dv} > 0. \quad (4.27)$$

La interpretación económica que hay detrás se basa en incrementos irracionales en la probabilidad de que  $Z_2$  sea extrema dado que  $Z_1$  ha alcanzado observaciones extremas (recorremos que las variables representan innovaciones).

### 4.3.1. Estimación de la Cópula: Contraste del Efecto Contagio

En general, para estimar el conjunto de parámetros que miden la dependencia  $\Theta$  de cualquier función de distribución multivariante se pueden adoptar dos estrategias. Si las distribuciones marginales son conocidas o se pueden estimar por medio de modelos paramétricos válidos la función de probabilidad para los datos se deriva fácilmente. Si la distribución multivariante es  $H(x_1, x_2) = C(F_1(x_1), F_2(x_2); \Theta)$  la función de verosimilitud es

$$\mathcal{L}(\Theta; x_1, x_2) = \sum_{i=1}^n \log f_1(x_{i,1}) + \sum_{i=1}^n \log f_2(x_{i,2}) + \sum_{i=1}^n \log c(F_1(x_{i,1}), F_2(x_{i,2}); \Theta), \quad (4.28)$$

con  $f_i$  la función de densidad marginal de  $F_i$ , y  $c(F_1(x_1), F_2(x_2); \Theta)$  la densidad bivariente de la cópula. Las estimaciones de los parámetros que gobiernan la dependencia son dependientes de las marginales, así como lo son los parámetros de las correspondientes distribuciones marginales. Por otro lado las estimaciones de  $\Theta$  están libres de estos efectos para estimaciones no-paramétricas de las marginales. Genest, Ghoudi, y Rivest (1995) muestran que las estimaciones derivadas de una estimación mediante pseudo-verosimilitud son consistentes y asintóticamente normales. Este método se implementa en dos etapas. Primero, las estimaciones de las distribuciones marginales se estiman mediante las distribuciones empíricas no-paramétricas respectivas. De este modo  $u_i$  se obtiene como  $u_i = \hat{F}_{i,n}(x)$ , con  $\hat{F}_{i,n}(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}$ , y la log-verosimilitud para  $C$  es

$$\mathcal{L}(\Theta; u_1, u_2) = \sum_{i=1}^n \log c(u_{i,1}, u_{i,2}; \Theta). \quad (4.29)$$

En nuestro caso,  $H^*(z_1, z_2) = \tilde{C}_G(\Psi_1(z_1), \Psi_1(z_2); \Theta)$ , aunque las funciones de distribución marginales sean conocidas, estándar Fréchet, las distribuciones reales  $F_1, F_2$  no lo son, así que es preferible considerar el caso no-paramétrico. Notemos que  $u_i = \Psi_1(z_i)$  se reduce a  $u_i = F_{i,n}(x_i)$  por construcción de  $Z_i$ . La función de log-verosimilitud  $H^*$  se calcula como en (4.29). Esta función, sin embargo, no toma una expresión fácil de manejar en logs, y la función “score” tampoco adopta por tanto una forma cerrada. En cambio métodos

numéricos de optimización se emplean para maximizar la verosimilitud.

Una propiedad interesante de la cópula  $\tilde{C}_G$  es su carácter anidado. Es inmediato ver que  $\gamma = 0$  es la distribución Gumbel estándar, que representa la clase de distribuciones extremas bivariantes. El caso de independencia asintótica para la cola derecha es dada por  $\theta = 1$  mientras que independencia perfecta se describe por  $\theta = 1$  y  $\gamma = 0$ . Finalmente,  $\eta$  mide el nivel de asimetría (intercambiabilidad) de las variables. Como resultado, es muy sencillo implementar contrastes para las correspondientes hipótesis sobre la dependencia por medio de contrastes del ratio de verosimilitudes. El estadístico del contraste es

$$\Lambda_n = 2 \log \frac{\sup_{\Theta} \prod_{i=1}^n \tilde{c}_G(u_{i,1}, u_{i,2}; \Theta)}{\prod_{i=1}^n \tilde{c}_G(u_{i,1}, u_{i,2}; \Theta_0)}, \quad (4.30)$$

con  $\Theta_0$  el conjunto de parámetros bajo la hipótesis nula. La distribución asintótica de  $\Lambda_n$  es chi-cuadrado con número de grados de libertad igual a la diferencia de las dimensiones entre  $\Theta$  y  $\Theta_0$ .

El carácter anidado de la cópula hace inmediato el contrastar dependencia así como contrastar la existencia del efecto contagio en los datos. Los contrastes de hipótesis correspondientes son  $H_0 : \theta = 1, \gamma = 0$  vs  $H_1 : \theta > 1$  ó  $\gamma > 0$ , y  $H_0 : \eta = 1$  vs  $H_1 : \eta \neq 1$  para  $\gamma > 0$ . Mientras tanto la existencia del contagio en la intensidad y del contagio en la dirección en el sentido fuerte se reducen a estudiar las condiciones (4.24) y (4.27) respectivamente sustituyendo los parámetros estimados.

#### 4.4. Aplicación: Vuelo hacia la calidad frente a Contagio

Las crisis financieras se caracterizan por caídas notables en los precios de los activos en los mercados de referencia. La caída de los precios de estos activos desatan una serie de efectos negativos en los precios del resto de los activos comerciados en el mercado por



diferentes razones: un notable peso del activo en la composición de las carteras, comercio bilateral, ó un efecto psicológico o de contagio.

Parece lógico entonces pensar que los inversores a la hora de evitar los efectos negativos de las crisis emigren hacia mercados más seguros: el mercado de bonos. Sin embargo, algunas veces no está claro el tipo de mercado que falla y origina la crisis, ya que la estructura económica entera se colapsa. En esta situación el refugio en los mercados de bonos no proporciona la cobertura deseada contra las pérdidas. El fenómeno de emigración de los mercados de acciones hacia los mercados de bonos se conoce como vuelo hacia la calidad. Medir este efecto es importante desde diferentes puntos de vista: refleja los efectos de los lazos entre estos mercados, en casos de crisis es útil identificar sus fuentes (financieras frente a otros tipos), o la causalidad de la relación, esto es, si mercados de acciones a la baja implican mercados de bonos al alza, o hay algún tipo de factor común que produzca co-movimientos (*ej.* tipos de interés bajos).

En este apartado se contrasta este fenómeno para dos pares diferentes de índices financieros: índice de bonos de empresas del Dow Jones con madurez 2 años (DJBI02) y el índice de acciones de empresas que forman el Dow Jones (DJSI), y el índice de bonos de empresas del Dow Jones con madurez 30 años (DJBI30) frente al índice de acciones de empresas que forman el Dow Jones antes citado. Estas series se estudian para el periodo 02/01/1997 – 24/09/2004. Los datos para los bonos se han tomado de la página web oficial del índice Dow Jones y el índice del precio de acciones del Dow Jones de *www.freelunch.com*. Las observaciones de la muestra total correspondiente a vacaciones públicas y datos faltantes en cualquiera de las series se han omitido del análisis para todas las series para evitar distorsiones y la incorporación de rendimientos nulos espúreos y de dependencias aberrantes. Esto deja  $n=1942$  observaciones. Las observaciones consideradas para el análisis son los rendimientos logarítmicos en términos porcentuales y denotados como  $r_t$ ,

$$r_t = 100 (\log P_t - \log P_{t-1}),$$

con  $P_t$  los precios originales en el momento  $t$ .

La metodología seguida en este trabajo empírico comienza por el filtrado de los datos por modelos univariantes como se presenta en (4.1) y analizando luego los patrones de dependencia para el vector de innovaciones resultante  $(\varepsilon_1, \varepsilon_2)$  por medio de la cópula  $\tilde{C}_G$ . Esta cópula es suficiente para contrastar la existencia de efectos de contagio, co-movimientos, o efectos opuestos en las colas que se reflejan en el conjunto de parámetros  $\Theta$  de la cópula  $\tilde{C}_G$ .

Las Tablas 4.6.1, 4.6.2 muestran que el índice *DJBI02* está bien modelizado por un modelo AR(1)-GARCH(1,1) como sigue,

$$X_{1,t} = 0,00025 + 0,089X_{1,t-1} + \sigma_{1,t}\varepsilon_{1,t}, \text{ con } \varepsilon_{1,t} \text{ i.i.d. } (0, 1), \text{ y}$$

$$\sigma_{1,t}^2 = 6,194 \cdot 10^{-8} + 0,071\varepsilon_{1,t-1}^2 + 0,903\sigma_{1,t-1}^2.$$

El índice *DJSI* es modelizado mediante el siguiente modelo GARCH(1,1) (tablas 4.6.3, 4.6.4),

$$X_{2,t} = \sigma_{2,t}\varepsilon_{2,t}, \text{ con } \varepsilon_{2,t} \text{ i.i.d. } (0, 1), \text{ y } \sigma_{2,t}^2 = 3,0012 \cdot 10^{-6} + 0,096\varepsilon_{2,t-1}^2 + 0,887\sigma_{2,t-1}^2.$$

La secuencia bivalente de las innovaciones  $(\varepsilon_{1,t}, \varepsilon_{2,t})$  se representa en la figura 4.6.1. Un primer vistazo al cuadro proporciona cierta guía hacia la existencia de un efecto de vuelo hacia la calidad entre las innovaciones de *DJSI* y las innovaciones correspondientes a *DJBI02*. El análisis de la correlación cruzada (figura 4.6.2) confirma la existencia de shocks opuestos en la secuencia de innovaciones así como valida los modelos univariantes propuestos para verificar los supuestos (4.1).

La función cópula  $\tilde{C}_G$  introducida en este capítulo se estima numéricamente. Las estimaciones de los parámetros para este ejemplo son  $\hat{\theta}_n = 1,031$ ,  $\hat{\eta}_n = 1$  y  $\hat{\gamma}_n = 0,175$ . Este modelo se ajusta bien a los datos para diferentes secciones de la cópula como se puede ver para ambas marginales (ver figura 4.6.3). Los siguientes cuadros se derivan de  $\tilde{C}_G$  estimada de los datos. De esta forma las figuras 4.6.4 y 4.6.5 muestran interdependencia negativa entre las variables aleatorias en la cola izquierda, que se hace más fuerte en la mitad de la

distribución bivalente y se vuelve positiva en la cola derecha. Ambos gráficos son idénticos indicando la ausencia de contagio direccional, esto es, efectos asimétricos entre las variables. Esto se describe también en la figura 4.6.6. Por otro lado también es notable la presencia de contagio en la intensidad en la cola izquierda (figura 4.6.5). Un análisis más profundo de las figuras 4.6.4 y 4.6.5 muestra movimientos opuestos en la mitad del dominio, que decrecen cuando las variables toman valores más grandes en valor absoluto. Este fenómeno es más pronunciado para los valores extremos negativos que tienden a moverse juntos, o al menos no en direcciones opuestas (contagio sin interdependencia).

Es conveniente no confundir el fenómeno de contagio que se acaba de ilustrar, que aparece cuando ambas variables toman simultáneamente valores extremos en la misma cola, del efecto de vuelo hacia la calidad. Este fenómeno ocurre cuando los valores extremos ocurren en las colas opuestas, en particular cuando  $DJBI02$  toma valores extremos positivos y  $DJSI$  toma valores extremos negativos. La figura 4.6.7 claramente describe la existencia de este fenómeno en ambas colas, que se puede interpretar como un efecto de sustitución entre estas series financieras cuando alguna de ellas están atravesando un periodo de crisis.

El análisis para el par bonos a 30 años del Dow Jones ( $DJBI30$ ) y el índice Dow Jones ( $DJSI$ ) da resultados diferentes. El  $DJBI30$  es modelizado por un modelo AR(1)-GARCH(1,1) donde  $DJSI$  también entra en la ecuación. Las estimaciones del parámetro se muestran en las tablas 4.6.5 y 4.6.6 y se pueden resumir como sigue,

$$X_{1,t} = 0,00037 + 0,063X_{1,t-1} + 0,048X_{2,t-1} + 0,028X_{2,t-2} + \sigma_{1,t}\varepsilon_{1,t}, \text{ con } \varepsilon_{1,t} \text{ i.i.d. } (0, 1), \text{ y}$$

$$\sigma_{1,t}^2 = 1,375 \cdot 10^{-6} + 0,056\varepsilon_{1,t-1}^2 + 0,905\sigma_{1,t-1}^2.$$

El par  $(\varepsilon_{1,t}, \varepsilon_{2,t})$  se representa en la figura 4.6.8. La función de correlación cruzada (figura 4.6.9) indica la presencia de correlación lineal entre cualquier combinación de los retardos. Este gráfico también da validez al modelo univariante propuesto para describir las dinámicas de  $DJBI30$ . Las estimaciones de los parámetros de  $\tilde{C}_G$  en este caso son  $\hat{\theta}_n = 1,01$ ,  $\hat{\eta}_n = 1$  y  $\hat{\gamma}_n = 0,0003$ . Este modelo se ajusta muy bien a los datos (figura 4.6.10). La cópula ajustada muestra que las variables aleatorias son débilmente interdependientes (figura 4.6.11), esto es, ambas secuencias de innovaciones, aunque cerca de la independencia, se mueven en la

misma dirección. Más formalmente, el contraste del ratio de verosimilitudes introducido en (4.30) calculado para  $H_0 : \theta = 1, \gamma = 0$  vs  $H_1 : \theta > 1$  ó  $\gamma > 0$  no rechaza la hipótesis de independencia. Notemos del valor de  $\gamma$  que las variables aleatorias son simétricas y por tanto no hay contagio direccional. Más aún, la figura (4.6.12) describe la ausencia de contagio en la intensidad en cualquiera de las colas reflejando un debilitamiento de los vínculos entre las variables en las colas. En el límite estas variables aleatorias son asintóticamente independientes en ambas colas. Finalmente el fenómeno de vuelo hacia la calidad no está presente para estas series como se muestra en la figura 4.6.13.

## 4.5. Conclusiones

Contagio e interdependencia son conceptos diferentes. En este capítulo el contagio se relaciona a los sucesos extremos o en la cola. Por medio de la teoría de cópulas somos capaces de analizar y contrastar la existencia de contagio, su intensidad, así como su dirección causal. Esto se hace creando una nueva cópula, derivada de la teoría de valores extremos multivariante que es suficientemente flexible tanto para describir los diferentes patrones de dependencia como para modelizar posibles efectos asimétricos entre los mercados.

Esta cópula también se ha aplicado a estudiar los vínculos entre mercados seguros y mercados de riesgo representados por el índice de bonos a 2 años y el índice Dow Jones de acciones (*DJSI*). Desde el punto de vista de fundamentos económicos, este último índice es independiente de *DJBI*, mientras que los índices de bonos, *DJBI02* y *DJBI30*, tienen comportamientos diferentes dependiendo de su madurez. El precio de *DJBI02* es independiente de la evolución de los mercados de riesgo, de hecho el precio medio condicional está dirigido sólo por su precio pasado, mientras que la varianza condicional es bien modelizada por un modelo GARCH(1,1). Por otro lado *DJBI30* es positivamente influenciado por la evolución de *DJSI* reflejando la salud del sistema económico entero.

Teniendo en cuenta los lazos irracionales entre los mercados recogidos en las secuencias de las innovaciones y modelizado por la función cópula introducida en este capítulo, las

Conclusiones son también diferentes para las pares correspondientes analizados de secuencias financieras. Los shocks entre *DJBI02* y *DJSI* están negativamente relacionados. En particular, el efecto del vuelo hacia la calidad está presente indicando un efecto de sustitución entre ambos instrumentos financieros cuando alguno de ellos está atravesando periodos de crisis. También es notable la existencia de un efecto contagio en la intensidad de la dependencia en situaciones de crisis en ambos mercados, shocks negativos comunes. Por otro lado, las secuencias de innovaciones de *DJSI* y *DJBI30* son casi independientes. No hay contagio o vuelo hacia la calidad.

La conclusión sobre la dependencia en estas series financieras es que mientras que *DJBI02* puede servir de refugio para los inversores que emigran de las crisis atribuidas a los shocks en los mercados de riesgo, *DJBI30* refleja la salud de la economía entera, incluyendo los mercados de acciones y se usan por un tipo de inversores no preocupados por las fluctuaciones puntuales de los precios en los mercados de acciones.

**4.6. Apéndice: Tablas y Figuras**

Parámetro	Valor	Error Estándar	T-estadístico
C	0,00023981	$3,2242e - 005$	7,4378
AR(1)	0,075112	0,023976	3,1327
AR(2)	-0,0060864	0,023504	-0,2590
Regress(1)	0,0033575	0,0023594	1,4231
Regress(2)	0,001902	0,0023953	0,7940
Regress(3)	-0,0032545	0,0027541	-1,1817
K	$4,6857e - 008$	$1,0825e - 008$	4,3286
GARCH(1)	0,92472	0,0088449	104,5485
ARCH(1)	0,055603	0,0053162	10,4591

**Tabla 4.6.1.** *Estimaciones de los parámetros para el índice DJCB02 durante el periodo 02/01/1997 – 24/09/2004.*

Parámetro	Valor	Error Estándar	T-estadístico
C	0,000254010	$3,1119e - 005$	8,1626
AR(1)	0,089148	0,024137	3,6934
K	$6,1945e - 008$	$1,4083e - 008$	4,3985
GARCH(1)	0,90361	0,011627	77,7188
ARCH(1)	0,071044	0,0066707	10,6502

**Tabla 4.6.2.** *Estimaciones de los parámetros para el índice DJCB02 durante el periodo 02/01/1997 – 24/09/2004.*

Parámetro	Valor	Error Estándar	T-estadístico
C	0,00049681	0,00024775	2,0053
AR(1)	-0,0078747	0,026044	-0,3024
AR(2)	-0,011078	0,023067	-0,4803
Regress(1)	0,0060823	0,0409134	0,1487
Regress(2)	0,035009	0,043498	0,8048
Regress(3)	-0,056755	0,038002	-1,4935
K	$2,9658e - 006$	$8,5241e - 007$	3,4793
GARCH(1)	0,88763	0,011124	79,7956
ARCH(1)	0,095891	0,0092913	10,3205

**Tabla 4.6.3.** *Estimaciones de los parámetros para el índice DJSP durante el periodo 02/01/1997 – 24/09/2004.*

Parámetro	Valor	Error Estándar	T-estadístico
C	0,00049118	0,00024454	2,0086
K	$3,0012e - 006$	$8,5464e - 007$	3,5116
GARCH(1)	0,88719	0,010683	83,0474
ARCH(1)	0,096116	0,0084832	11,3302

**Tabla 4.6.4.** *Estimaciones de los parámetros para el índice DJSP durante el periodo 02/01/1997 – 24/09/2004.*

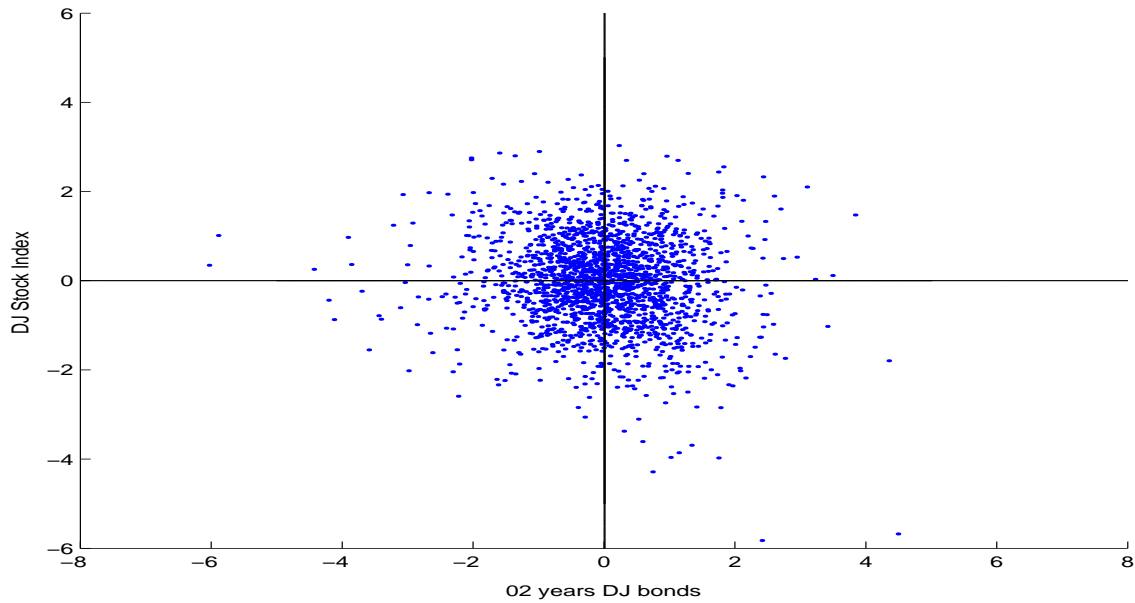
Parámetro	Valor	Error Estándar	T-estadístico
C	0,0003741	0,000129	2,9001
AR(1)	0,063049	0,023987	2,6284
AR(2)	-0,0075111	0,023612	-0,3181
Regress(1)	0,04884	0,0097296	5,0197
Regress(2)	0,028364	0,01017	2,7891
Regress(3)	0,0058778	0,011312	0,5196
K	$1,3042e - 006$	$4,3895e - 007$	2,9712
GARCH(1)	0,90783	0,019353	46,9090
ARCH(1)	0,055566	0,0098548	5,6385

**Tabla 4.6.5.** *Estimaciones de los parámetros para el índice DJCB30 durante el periodo 02/01/1997 – 24/09/2004.*

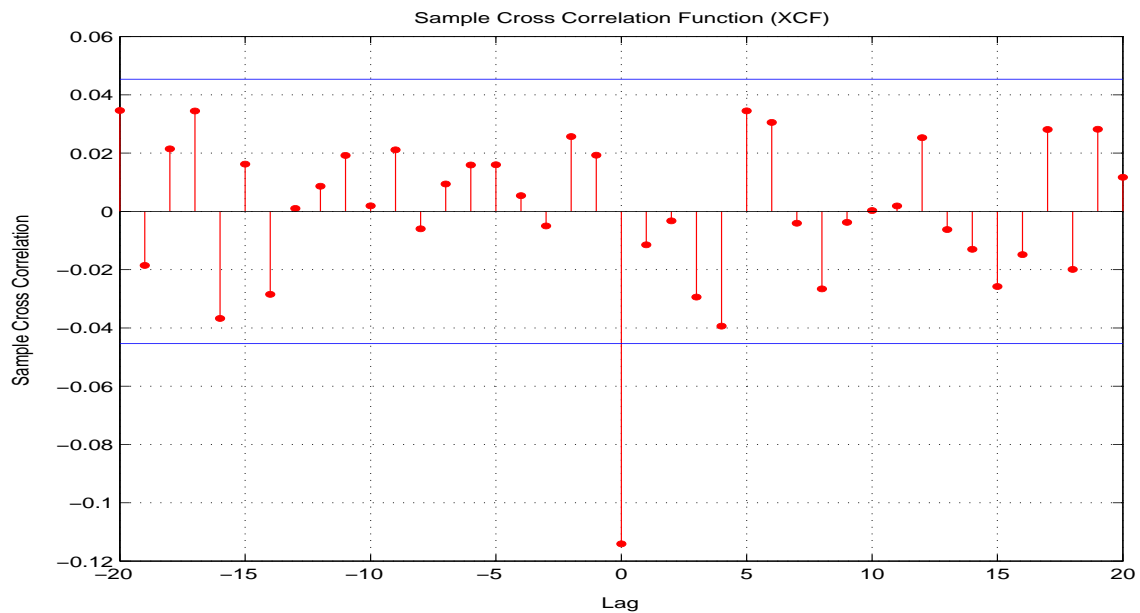
Parámetro	Valor	Error Estándar	T-estadístico
C	0,00037386	0,00012836	2,9127
AR(1)	0,063416	0,023811	2,6633
Regress(1)	0,048314	0,0097464	4,9572
Regress(2)	0,028526	0,010177	2,8031
K	$1,3752e - 006$	$4,5783e - 007$	3,0038
GARCH(1)	0,90551	0,020033	45,2007
ARCH(1)	0,055866	0,010029	5,5704

**Tabla 4.6.6.** *Estimaciones de los parámetros para el índice DJCB30 durante el periodo 02/01/1997 – 24/09/2004.*

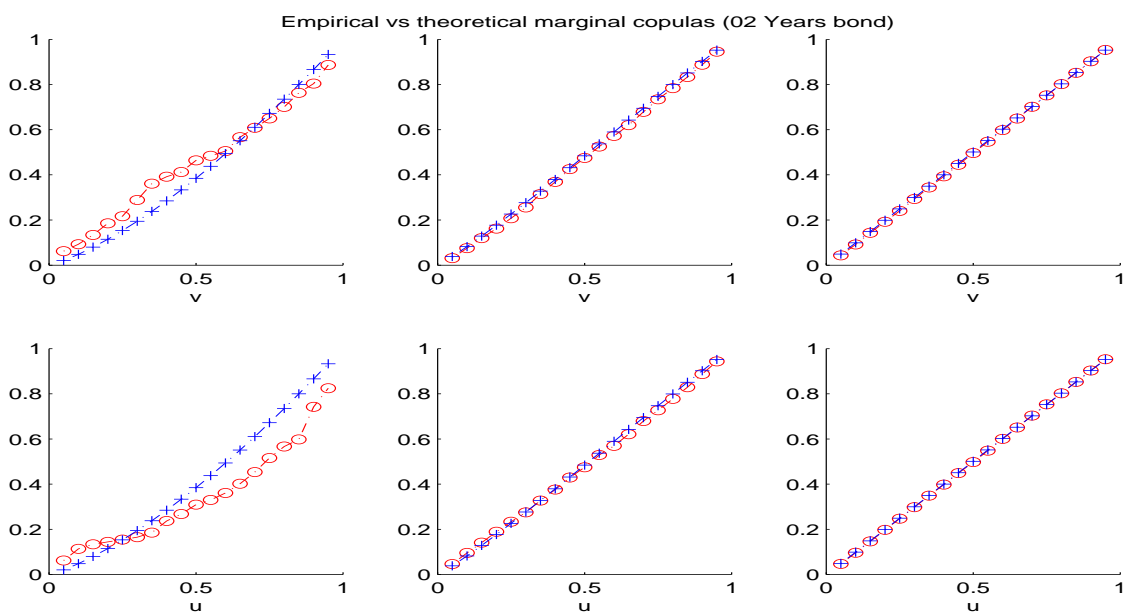




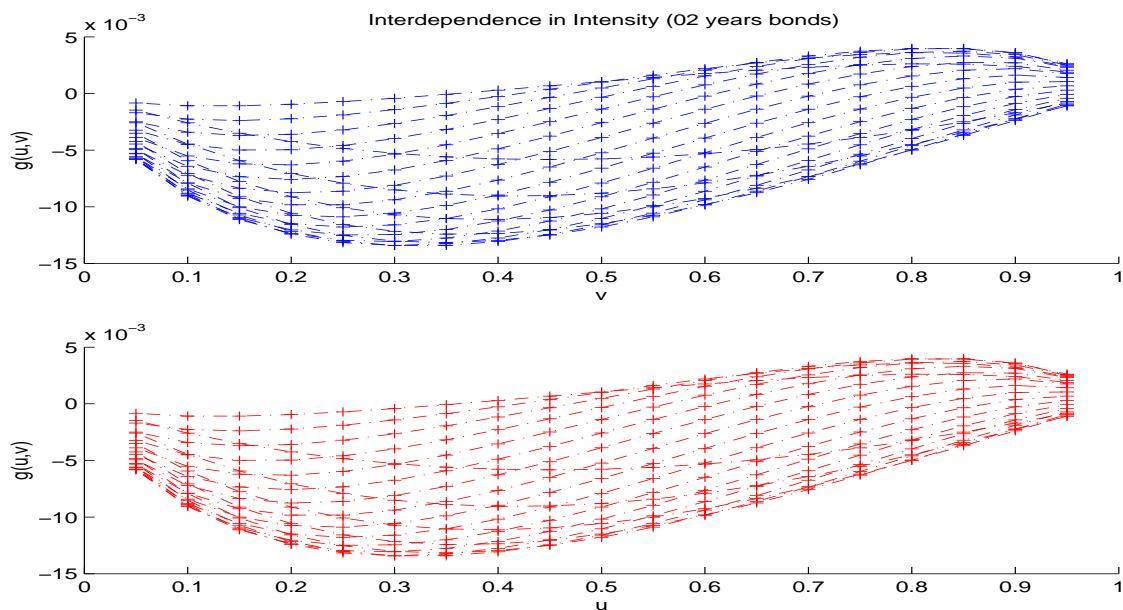
**Figura 4.6.1.** Gráfico bivalente de la secuencia de innovaciones de los bonos a 2 años y del índice de acciones Dow Jones. Las observaciones se extienden durante el periodo 02/01/1997 – 24/09/2004,  $n = 1942$  observaciones.



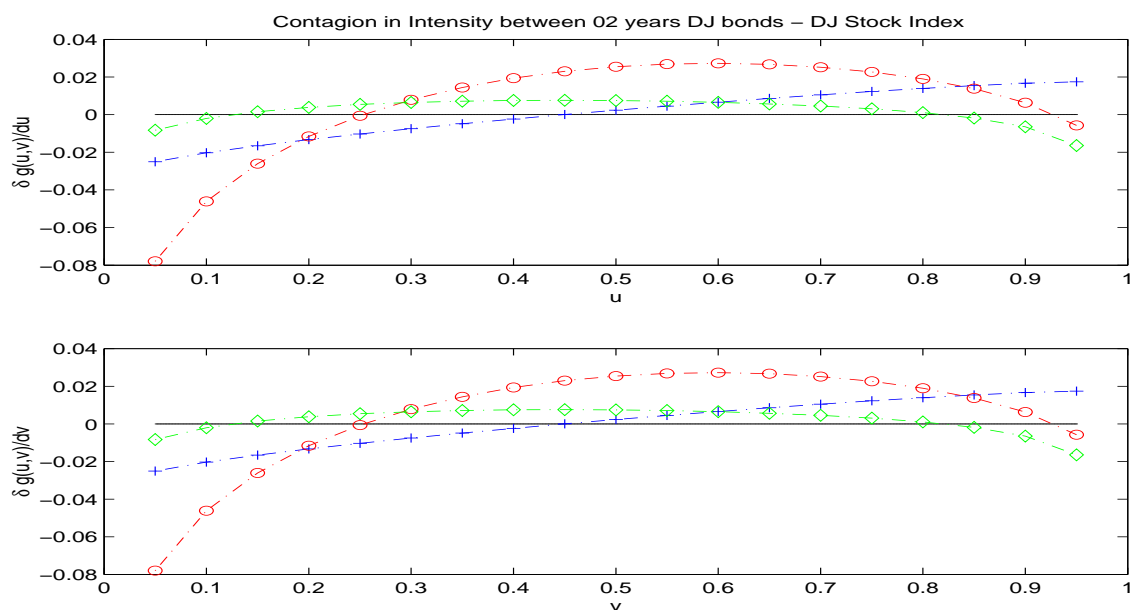
**Figura 4.6.2.** Correlación cruzada para diferentes retardos de la secuencia de innovaciones bivalente, bonos del Dow Jones a 2 años y del índice de acciones Dow Jones, que se extienden durante el periodo 02/01/1997 – 24/09/2004,  $n = 1942$  observaciones.



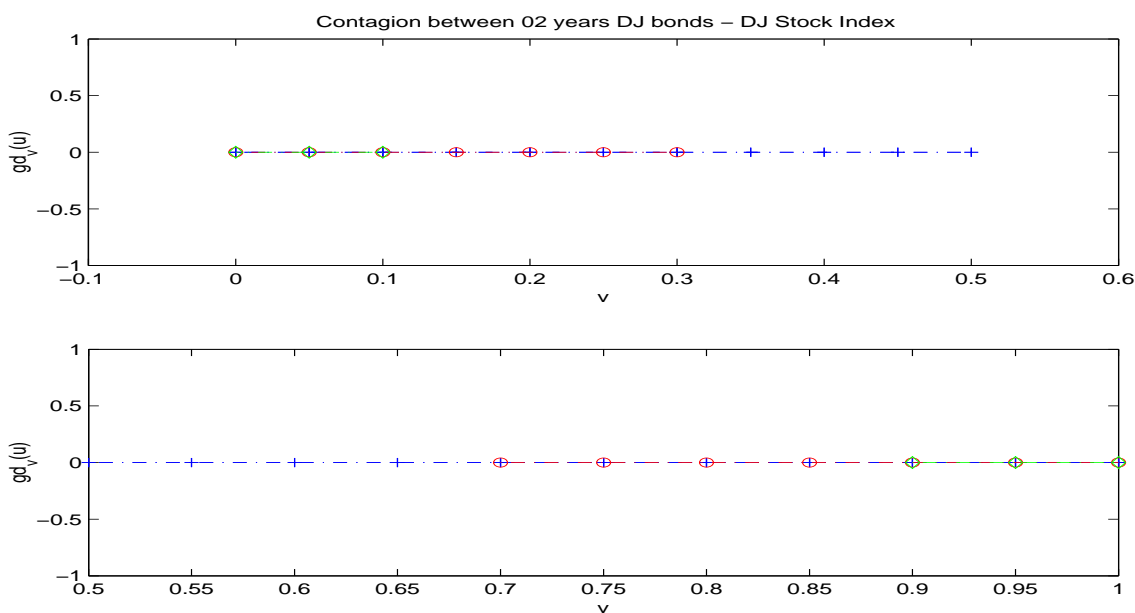
**Figura 4.6.3.** Distribución marginal Empírica (o-) y teórica (+-) de la distribución bivalente. El panel superior describe las secciones verticales y el panel inferior la sección horizontal. Los paneles de la izquierda representan el cuantil 0,05, los del medio el cuantil 0,50 y los de la derecha el cuantil 0,95.



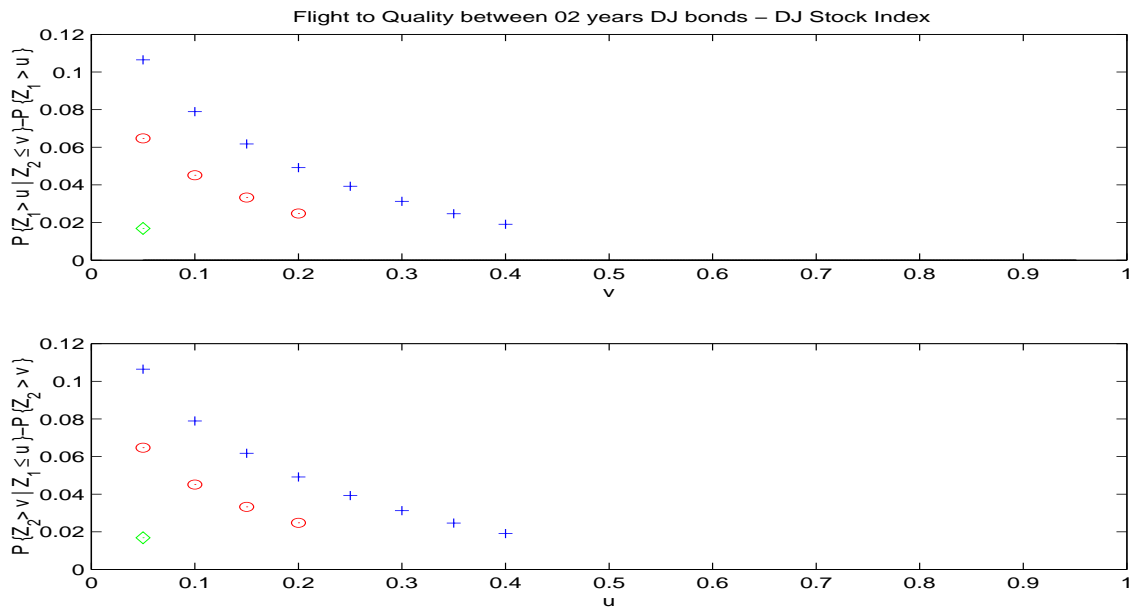
**Figura 4.6.4.** El panel superior describe la función  $g(u,v)$  como está definida en (4.23) representada contra las innovaciones de DJSI. El panel inferior  $g(u,v)$  se dibuja contra las innovaciones de DJBI02.



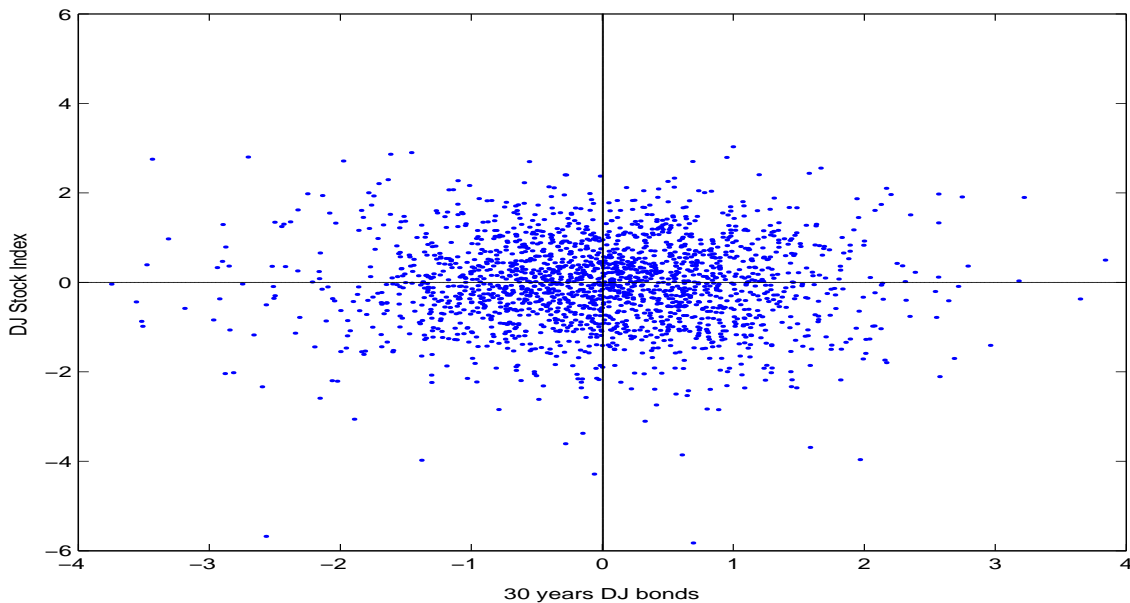
**Figura 4.6.5.** El panel superior describe la función  $h_1(u, v)$  como está definida en (4.24) representada contra las innovaciones de DJBI02 y el panel inferior describe  $h_2(u, v)$  contra las innovaciones de DJSI. (+-) representa el cuantil 0.05, (o-) el cuantil 0.50 y ( $\diamond$ -) el cuantil 0.95.



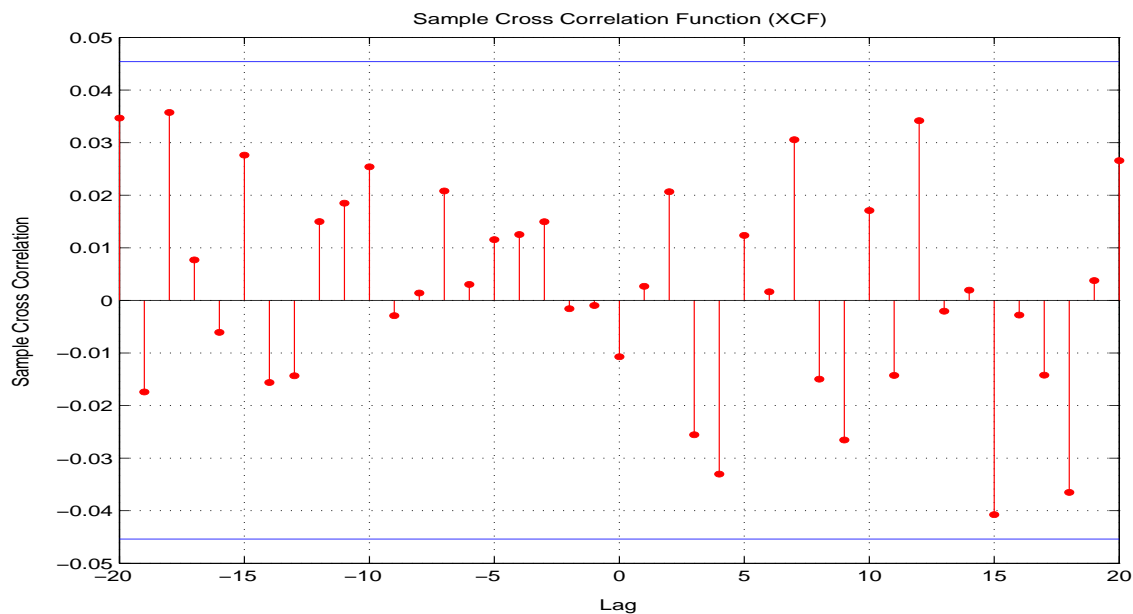
**Figura 4.6.6.** El panel superior describe  $gd_v(u) = \tilde{C}_G(u, v) - \tilde{C}_G(v, u)$  para la cola de la izquierda ( $v \leq u$ ). (+-) representa  $u = 0,50$ , (o-) representa  $u = 0,30$  y ( $\diamond$ -) for  $u = 0,10$ . El panel inferior  $gd_v(u)$  para la cola de la derecha ( $v > u$ ). (+-) representa  $u = 0,50$ , (o-) representa  $u = 0,70$  y ( $\diamond$ -) para  $u = 0,90$ .



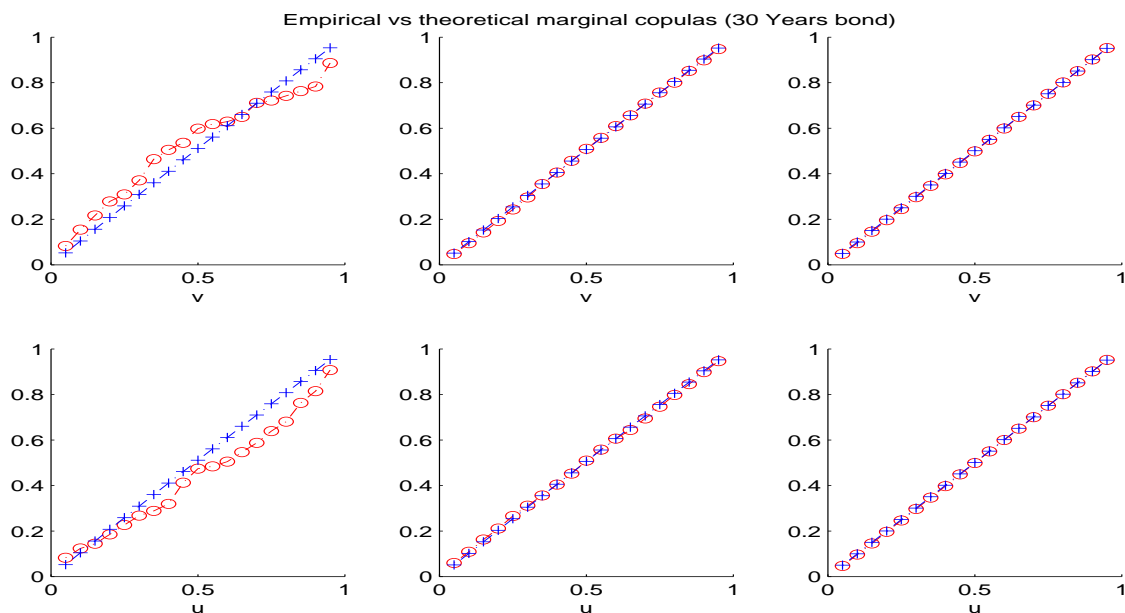
**Figura 4.6.7.** El panel superior describe el vuelo hacia la calidad de DJSI hacia DJBI02. (+-) representa  $u = 0,60$ , (o-) representa  $u = 0,80$  y (◇-) para  $u = 0,95$ . El panel superior describe el vuelo hacia la calidad de DJBI02 hacia DJSI. (+-) representa  $v = 0,60$ , (o-) representa  $v = 0,80$  y (◇-) para  $v = 0,95$ .



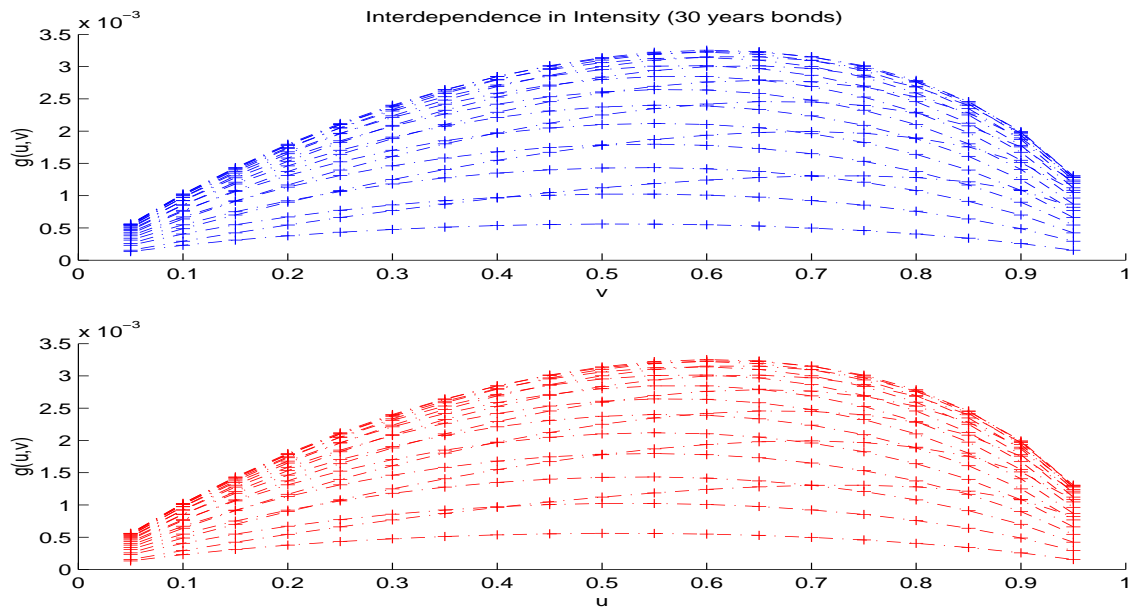
**Figura 4.6.8.** Gráfico bivalente de la secuencia de innovaciones de los bonos a 30 años y del índice de acciones Dow Jones. Las observaciones se extienden durante el periodo 02/01/1997 – 24/09/2004,  $n = 1942$  observaciones.



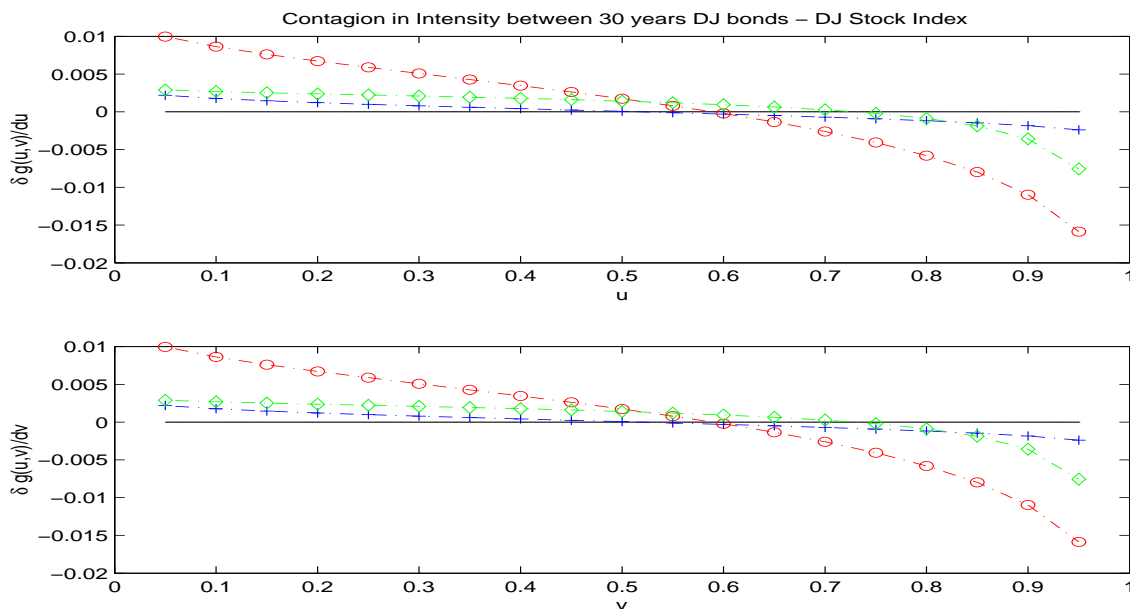
**Figura 4.6.9.** *Correlación cruzada para diferentes retardos de la secuencia de innovaciones bivalente, bonos del Dow Jones a 30 años y del índice de acciones Dow Jones, que se extienden durante el periodo 02/01/1997 – 24/09/2004,  $n = 1942$  observaciones.*



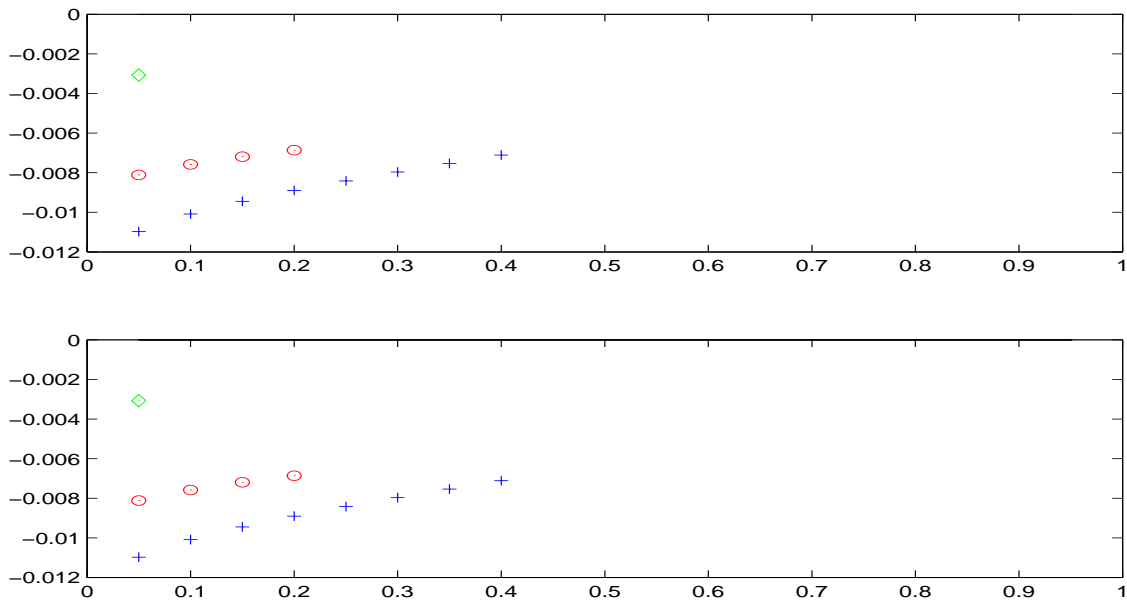
**Figura 4.6.10.** *Distribución marginal Empírica (o—) y teórica (+—) de la distribución bivalente. El panel superior describe las secciones verticales y el panel inferior la sección horizontal. Los paneles de la izquierda representan el cuantil 0,05, los del medio el cuantil 0,50 y los de la derecha el cuantil 0,95.*



**Figura 4.6.11.** El panel superior describe la función  $g(u, v)$  como está definida en (4.23) representada contra las innovaciones de DJSI. El panel inferior  $g(u, v)$  se dibuja contra las innovaciones de DJBI30.



**Figura 4.6.12.** El panel superior describe la función  $h_1(u, v)$  como está definida en (4.24) representada contra las innovaciones de DJBI30 y el panel inferior describe  $h_2(u, v)$  contra las innovaciones de DJSI. (+) representa el cuantil 0.05, (o) el cunatil 0.50 y ( $\diamond$ ) el cuantil 0.95.



**Figura 4.6.13.** El panel superior describe el vuelo hacia la calidad de DJSI hacia DJBI30.  $(+-)$  representa  $u = 0,60$ ,  $(o-)$  representa  $u = 0,80$  y  $(\diamond-)$  para  $u = 0,95$ . El panel inferior describe el vuelo hacia la calidad de DJBI30 hacia DJSI.  $(+-)$  representa  $v = 0,60$ ,  $(o-)$  representa  $v = 0,80$  y  $(\diamond-)$  para  $v = 0,95$ .





# Capítulo 5

## Futuras Líneas de Investigación

Este capítulo introduce brevemente las posibles extensiones de esta tesis. La discusión se centra en tres temas. Primero, la transmisión del riesgo entre economías y los efectos del contagio en la diversificación. Segundo, la transmisión del riesgo en series financieras que implican dependencia temporal en las observaciones extremas y que son generadas por colas pesadas, y finalmente el uso de técnicas de *EVT* para la detección de observaciones aberrantes (atípicos) fuera del rango de las observaciones disponibles y lejos de ser explicadas por los valores extremos de la distribución de los datos.

Los efectos de contagio en los diferentes aspectos de la economía es un área prometedora de investigación que está acumulando interés tanto de académicos como de la empresa privada. Esta tesis ha explorado el vuelo hacia la calidad desde los mercados de acciones hacia el mercado de bonos. Otros interesantes resultados que envuelven efectos de contagio se encuentran en el análisis de las crisis bancarias, crisis monetarias y las crisis de carteras. El objetivo siguiente de esta tesis es la gestión del riesgo de las carteras. En particular los efectos negativos del contagio en las estrategias de diversificación.

La metodología introducida en Markowitz para la selección de carteras (Journal of Finance, 1952, 1991) se basa en minimizar la combinación lineal del vector de pesos de los rendimientos individuales y de la correspondiente matriz de varianzas y covarianzas. De esta forma la cartera óptima es la combinación lineal de mínima varianza. Esta medida estadística, sin embargo es una herramienta adecuada para medir el riesgo sólo bajo ciertas condiciones restrictivas: una distribución log-normal de los precios de los activos (esto implica una distribución normal de los rendimientos), o utilidades cuadráticas. En este mundo la correlación juega un papel principal para cuantificar la dependencia. Por otro lado, una de las contribuciones de esta tesis ha sido resaltar la necesidad de medidas estadísticas alternativas para modelizar la dependencia cuando la distribución de los rendimientos está lejos de ser explicada por la distribución normal. Por tanto, las decisiones de elección de cartera basadas en medidas lineales como la varianza y la correlación pueden ser erróneas. La asimetría observada en los rendimientos financieros no es recogida tampoco en estas medidas. La asimetría negativa implica una mayor probabilidad de los rendimientos negativos grandes que de los valores correspondientes positivos. En este contexto, es importante desarrollar una metodología que sea capaz de tener en cuenta el riesgo (cola negativa), y la probabilidad de beneficios (cola positiva) de una forma separada.

Las técnicas de la teoría de valores extremos son adecuadas para estos problemas por su capacidad para describir la distribución de los rendimientos en las colas. Junto con estas técnicas la función cópula bivalente desarrollada en Gonzalo y Olmo (2005) será el principal ingrediente para modelizar relaciones asimétricas entre los elementos de la cartera. La intención de esta investigación es cambiar el criterio de la decisión de la elección de

carteras de minimizar la varianza a minimizar el riesgo, medido como la probabilidad en la cola negativa de la distribución de la cartera.

El segundo tema de investigación considera la transmisión del riesgo en series temporales. Las teorías basadas en la dependencia en los segundos momentos han sido enormemente populares desde que surgieron en la mitad de la década de los ochenta. En particular los modelos GARCH(1,1), debido a su simplicidad y tratabilidad en la práctica. Los inversores privados han sido muy aficionados a estas metodologías también. Conforme el tiempo pasaba y las deficiencias y limitaciones de esta metodología surgían, diferentes modificaciones del modelo GARCH(1,1) básico surgieron recogiendo los hechos estilizados no descritos por el GARCH(1,1) original. Algunos de estos hechos estilizados son la asimetría negativa de los rendimientos o un mayor clustering de las observaciones negativas. Diferentes modelos como EGARCH, IGARCH, etc. se *diseñaron* para describir estas características de los datos financieros. Todas estas metodologías asumen rendimientos condicionales gaussianos donde la volatilidad varía en el tiempo y es estacionaria (bajo algunas restricciones más) reflejando las dinámicas encontradas en los datos.

En los últimos *años* sin embargo, la creencia común de que las observaciones están normalmente distribuidas ha ido encontrando más y más enemigos. Los primeros signos del desencuentro con estos modelos fue la necesidad de asumir distribuciones *t*-Student para las observaciones en vez de distribuciones normales. Esta familia de distribuciones tampoco describe muy bien las características empíricas de los datos financieros. Otros autores, en cambio, han propuesto el uso de distribuciones más pesadas como por ejemplo las distribuciones  $\alpha$ -stables de varianza infinita. La intención de mi futura investigación se encuentra en mixturas de modelos gaussianos y de colas pesadas. La condición de Berman dice que

$$Cov(\varepsilon_t, \varepsilon_{t-j}) \log j \rightarrow 0 \Rightarrow \theta = 1,$$

cuando  $j \rightarrow \infty$ , con  $\varepsilon_t$  las observaciones estandarizadas. El rechazo estadístico de la hipótesis nula  $\theta = 1$  frente a  $\theta < 1$  implica que o la covarianza no es  $o(1/\log j)$  o  $\varepsilon_t$  no se distribuye como una normal. Dado que la correlación de los rendimientos financieros

es cero no hay duda de que rechazar  $\theta = 1$  equivale a rechazar el supuesto gaussiano.

Bajo esta evidencia parece apropiado considerar mixturas de distribuciones como proceso generador de los datos. La distribución normal para periodos de calma donde las observaciones son *iid*, y distribuciones de colas pesadas para periodos de estrés. La dependencia sería causada por las observaciones más grandes que generarían un efecto de perturbación (pánico o euforia) durante un periodo de tiempo (clusters). No hay duda de que esta teoría de transmisión del riesgo en las series financieras está en sus albores y que necesita mucha atención y base teórica para construir una teoría fuerte y consistente pero los primeros pasos se han dado con el contraste de hipótesis para el extremal index y la condición de Berman como es comentado en Olmo (2005a).

De la misma manera que los procesos generando clusters de observaciones significativamente grandes pueden estar mal especificadas y modelizadas por modelos de volatilidad condicionada heteroscedástica, el resultado contrario es también notable. Atípicos aislados pueden esconder verdadera heteroscedasticidad condicionada y ser confundidos con procesos *iid*. Este problema, la detección de atípicos en series financieras es también de interés en la gestión del riesgo. Unas pocas observaciones pueden sesgar las estimaciones de la cola de la distribución marginal de los datos. Para el caso condicional estas observaciones grandes significativas pueden afectar las estimaciones de los parámetros que dirigen la dependencia. Es importante entonces tener en cuenta estas observaciones y desarrollar una metodología capaz de detectar atípicos en ambos casos, tanto para secuencias *iid* como para modelos de volatilidad condicional heteroscedástica.

Una definición basta de atípico es una observación que no es generada por el proceso generador de los datos. En consecuencia, el primer reto es conocer el proceso. Bajo supuestos *iid* esto se reduce a conocer  $F$  si la distribución de los datos tiene soporte finito. De otra forma la definición anterior puede llevar a equívocos y necesita de supuestos adicionales. Por otro lado, para propósitos de gestión del riesgo parece más interesante la detección de atípicos en series temporales. En este caso el proceso generador no es solo  $F$  sino el proceso aleatorio que genera la dependencia. La literatura actual en contrastes para la detección de atípicos para series temporales se basa en contrastes del ratio de verosimilitudes y en

modelos de regresión entre los residuos y las innovaciones verdaderas, ver Tsay (1988) y Chen y Liu (1993) para un resumen de los diferentes métodos.

Estos contrastes requieren cierta información sobre el proceso incluyendo el tipo y situación del atípico. Además, las innovaciones del proceso generador deben ser gaussianas. Bajo estos supuestos, la distribución asintótica del estadístico del contraste para la hipótesis nula es normal estándar. Si la situación del atípico no se conoce, sin embargo, la distribución asintótica bajo la nula es altamente no estándar y debe ser calculada por simulación.

El objetivo de esta investigación futura es aprovecharnos de las propiedades de *EVT* para detectar atípicos. En particular de los resultados asintóticos de la distribución del máximo muestral. El contraste de hipótesis y las reglas de detección para observaciones atípicas basadas en *EVT* se esboza en los siguientes resultados.

Las tres distribuciones extremas están relacionadas. Dada una variable aleatoria  $Y$  con una función de distribución Fréchet ( $\Phi_\xi$ ) existen ciertas transformaciones de  $Y$  que siguen una distribución Gumbel ( $\Lambda$ ), y una distribución Weibull ( $\Psi_\xi$ ).

$$Y \sim \Phi_\xi \Leftrightarrow \ln Y^{1/\xi} \sim \Lambda \Leftrightarrow -Y^{-1} \sim \Psi_\xi, \quad (5.1)$$

con  $\xi$  el tail index de  $F$ .

Supongamos  $\{x_n\} = (x_1, \dots, x_n)$  una secuencia aleatoria *iid*, y denotemos  $\gamma_n = M_n(\{x_n\})$  para su máximo muestral. El estadístico utilizado para contrastar la presencia de atípicos es  $a_n^{-1}(\gamma_n - b_n)$ . Bajo la hipótesis nula de no atípicos este estadístico sigue asintóticamente una distribución extrema de Fréchet, Weibull o del tipo Gumbel. Denotando  $G_\xi$  para la distribución extrema, entonces la regla de rechazo para la hipótesis nula de no atípicos estará dada por

$$P \{a_n^{-1}(\gamma_n - b_n) < G_\xi\} < \alpha, \quad (5.2)$$

con  $\alpha$  el nivel de significación. Esta regla equivale a que la probabilidad de que  $\gamma_n$  esté generada por  $F$  es despreciable. Esto se mide por la probabilidad de que  $\gamma_n$  estandarizado pertenezca al rango de la distribución de valores extremos correspondiente.

Notemos que de esta forma la información relativa a la situación y al tipo de atípico

y la forma de la distribución  $F$  no son necesarias. Las únicas dificultades surgen ahora de la estimación de  $\xi$ ,  $a_n$ , y  $b_n$ . No obstante por medio de (5.1) el estadístico del contraste  $a_n^{-1}(\gamma_n - b_n)$  se puede transformar para que siga asintóticamente la distribución Gumbel ( $\Lambda$ ), esto es que sea libre de parámetros. En consecuencia, el estadístico del contraste es un pivot, aunque la estimación del tail index y de las secuencias  $a_n$ ,  $b_n$  se debe considerar todavía. Las observaciones de la secuencia influyendo en la estimación de estos parámetros se denotarán como observaciones influyentes más que como atípicos.

Para propósitos de gestión del riesgo la extensión del contraste a procesos dependientes del tipo GARCH también se considera. Más detalles sobre la detección de atípicos en el caso *iid* así como para series temporales se pueden encontrar en Olmo (2005b).

# Bibliografía

- [1] Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2001). “The Distribution of Realized Exchange Rate Volatility,” *Journal of American Statistical Association*.
- [2] Anderson, T.W., and D.A. Darling (1952). “Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes,” *Annals of Mathematical Statistics* 23, 193 – 212.
- [3] Baig, T., and I. Golfajn (1998). “Financial markets contagion in the asian crises,” IMF Mimeo.
- [4] Barndorff-Nielsen, O.E. and N. Shephard (2002). “Econometric analysis of realised volatility and its use in estimating stochastic volatility models,” *Journal of the Royal Statistical Society, B*, 64 (Forthcoming).
- [5] Beran, R. (1986). “Simulated Power Functions,” *Annals of Statistics* 14, 151 – 173.
- [6] Berman, S.M. (1964). “Limit theorems for the maximum term in stationary sequences,” *Annals of Mathematical Statistics* 35, 502-516.
- [7] Black, F. and M. Scholes (1973). “The pricing of options and corporate liabilities,” *Journal of Political Economy* 81, 637-653.
- [8] Bollerslev, T. (1990). “Modelling the Coherence in Sort-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model,” *Review of Economics and Statistics*, 72, 498-505.

- [10] Chen, C. and L. Liu (1993). "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association*, 88, 421, Theory and Methods.
- [10] Chernick, M.R. (1981). "A limit theorem for the maximum of autorregressive processes with uniform marginal distribution," *Annals of Probability* 9, 145 – 149.
- [11] Coles, S., J. Heffernan, and J.A. Tawn (1999). "Dependence measures for Extreme Value Analysis," *Extremes* 2:4, 339-365.
- [12] Corsetti, G., M. Pericoli, and M. Sbracia (2002). "Some contagion, some interdependence: More pitfalls in Tests of Financial Contagion," Working Paper, Bank of Italia.
- [13] Davis, R.A., and S.I. Resnick (1984). "Tail Estimates Motivated by Extreme Value Theory," *Annals of Statistics* 12, 1467 – 1487.
- [14] Diebold, F.X., T. Schuerman and J.D. Strouhair (1998). "Pitfalls and opportunities in the use of Extreme Value Theory in Risk Management," Working Paper Series, Wharton School, Univeristy of Pennsylvania.
- [15] de Haan, L. (1976). "Sample extremes: an elementary introduction," *Statist. Neerlandica*, 30, 161 – 172.
- [16] de Haan, L., and J. de Ronde (1998). "Sea and Wind: multivariate extremes at work," *Extremes*, 1, 7-45.
- [17] Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.
- [18] Embrechts, P., A. McNeil, and D. Straumann (1999). "Correlation: Pitfalls and Alternatives," *Risk*, 12 (5), 11-21.
- [19] Engle, R. (1999). "Dynamic Conditional Correlation- A Simple Class of Multivariate GARCH Models." Manuscript.



- [20] Engle, R.F., and T. Bollerslev (1986). “Modelling the persistence of conditional variances,” *Econometric Reviews* 5, 1 – 50 (with discussion).
- [21] Engle, R., and K. Kroner (1995). “Multivariate Simultaneous GARCH,” *Econometric Theory*, 11, 122-150.
- [22] Essary, J.D., and F. Proschan (1972). “Relationships among some concepts of bivariate dependence,” *Annals of Mathematical Statistics* 43, 651-655.
- [23] Ferro, C.A., and J. Segers (2003). “Inference for clusters of extremes,” *Journal of the Royal Statistical Society B*, 65, 545 – 556.
- [24] Forbes, K., and R. Rigobon (2001). “No contagion, only interdependence: Measuring stock markets comovements,” *Journal of Finance* 57 (5), 2223-2261.
- [25] Genest, C., K. Ghoudi, and L.P. Rivest (1995). “A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions,” *Biometrika* 82, 3, 543-552.
- [26] Gnedenko, B.V. (1943). “ Sur la distribution limit du terme d’une série aléatoire,” *Annals of Mathematics* 44, 423 – 453.
- [27] Goldie, C.M., and R.L. Smith (1987). “ Slow Variation with Remainder: Theory and Applications,” *Quart. J. Math. Oxford Ser. (2)*, 38, 45 – 71.
- [28] Gonzalo, J., and J. Olmo (2005). “Contagion versus Flight to Quality in Financial Markets,” Working paper, Department of Economics, Universidad Carlos III de Madrid.
- [30] Granger, C.W.J. (1969). “Investigating Causal Relationships by Econometric Models and Cross-spectral Methods,” *Econometrica* 37, 24-36.
- [30] Granger, C.W.J. (2002). “ Some Comments on Risk,” *Journal of Applied Econometrics* 17, 447 – 456.
- [31] Granger, C.W.J., T. Terasvirta, and A. Patton (2002). “Common Factors in Conditional Distributions,” SSE/EFI Working Paper Series in Economics and Finance.

- [32] Hall, P. (1982). “On some simple estimates of an exponent of regular variation,” *Journal of Royal Statistical Society B*, 44, 37 – 42.
- [33] Hartmann, P., S. traetmans, and C. de Vries (2001). “Extreme market linkages in crisis periods,” Discussion Paper Series 2916, Center for Economic Policy Research, [www.cerp.org](http://www.cerp.org).
- [34] Hill, B.M. (1975). “A Simple General Approach to Inference about the Tail of a Distribution,” *Annals of statistics* 5, 1163 – 1174.
- [35] Hsing, T. (1991). “Estimating the parameters of rare events. *Stochastic Processes and Applications*,” 37, 117 – 139.
- [36] Hsing, T. (1993). “Extremal Index Estimation for a Weakly Dependent Stationary Sequence,” *Annals of Statistics*, 21, 2043 – 2071.
- [37] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.
- [38] Kallenberg, O. (1976). *Random Measures*. Akademie Verlag, Berlin and Academic Press, New York.
- [39] Knight, F.H. (1921). “Risk, Uncertainty and Profit,” Boston: Houghton Mifflin, 1921.
- [40] Kolmogorov, A.N. (1933). “Sulla Determinazione Empirica di una Legge di Distribuzione,” *Giornale del Istituto Italiano degli Attuari* 4, 83 – 91.
- [41] Leadbetter, M. R. (1983). “Extremes and Local Dependence in Stationary Sequences,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 65, 291 – 306.
- [42] Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer-Verlag.
- [43] Ledford, A.W., and J.A. Tawn (1996). “Statistics for near independence in multivariate extreme values,” *Biometrika* 83, 1, 169-187.

- [44] Ledford, A.W., and J.A. Tawn (1997). “Modelling Dependence within joint tail regions,” *Journal of Royal Statistical Society B*, 49, 475-499.
- [45] Lehmann, E.L. (1966). “Some Concepts of Dependence,” *Ann. Math. Statist.* 37, 1137-1153.
- [46] Longin, F., and B. Solnik (2001). “Extreme correlation of international equity markets,” *Journal of Finance* 56 (2), 649-676.
- [47] Loynes, R.M. (1965). “Extreme Values in Uniformly Mixing Stationary Stochastic Processes,” *Annals of Mathematical Statistics*, 36, 993 – 999.
- [48] Mashal, R., and A. Zeevi (2002). “Beyond Correlation: Extreme Comovements Between Financial Assets,” Paper prepared for the Conference in Extremal Events in Finance held in Montreal, October 2002.
- [49] Markowitz, H. (1952). “Portfolio Selection,” *Journal of Finance* 7 (1), 77-91.
- [50] Markowitz, H. (1991). “Foundations of Portfolio Theory,” *Journal of Finance* 46, 469-477.
- [51] Mason, D., and J.H. Schuenemeyer (1983). “A modified Kolmogorov-Smirnov test sensitive to tail alternatives,” *Annals of Statistics* 3, 933 – 946.
- [52] Nelsen, R.B. (1999). *An introduction to copulas*. Lecture Notes in Statistics 139, New York: Springer-Verlag.
- [53] O’Brien, G.L. (1974). “The maximum term of uniformly mixing stationary sequences,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 30, 57 – 63.
- [54] O’Brien, G.L. (1987). “Extreme Values for Stationary and Markov Sequences,” *Annals of Probability*, 15, 281 – 291.
- [55] Olmo, J. (2005a). “Testing the Existence of Clustering in the Extreme Values,” Submitted.

- [56] Olmo, J. (2005b). “Outliers Detection with Extreme Value Theory,” Work in progress.
- [57] Patton, A.J. (2001). “Modelling Time Varying Exchange Rate Dependence Using the Conditional Copula,” Working Paper 2001-09, Department of Economics, University of California.
- [58] Pickands III, J. (1975). “Statistical Inference Using Extreme Order Statistics,” *Annals of Statistics* 1, 119 – 131.
- [59] Pickands III, J. (1981). “Multivariate extreme value distributions,” In *Proceedings of the 43rd Session of the International Statistical Institute* 2 859-878. ISI, Voorburg, Netherlands.
- [60] Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.
- [61] Poon, S.H., M. Rockinger, and J.A. Tawn (2003). “Modelling Extreme-Value Dependence in International Stock Markets,” *Statistica Sinica* 13, 929-953.
- [62] Press, W. H. (1992). *Numerical Recipes in C: the art of scientific computing*. Cambridge University Press, Ed. 2.
- [63] Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag.
- [64] Romano, J. P. (1988). “A Bootstrap Revival of Some Nonparametric Distance Tests,” *Journal of the American Statistical Association* 83, 698 – 708.
- [65] Rosenblatt, M. (1956). “A central limit theorem and a strong mixing condition,” *Proc. Nat. Acad. Sci. USA*, 42, 43 – 47.
- [66] Shiryaev, A. N. (2001). *Essentials of Stochastic Finance: Facts, Models, Theory*. Singapore: World Scientific Publishing.
- [67] Sklar, A. (1959). “Fonctions de répartition à n dimensions et leurs marges,” *Publ. Inst. Statis. Univ. Paris*, 8, 229-231.

- [68] Smith, R.L. (1985). “Maximum Likelihood Estimation in a Class of Non-Regular Cases,” *Biometrika* 72, 67 – 90.
- [69] Smith, R.L. (1987). “Estimating Tails of Probability Distributions,” *Annals of Statistics* 15, 1174 – 1207.
- [70] Smith, R.L., and I. Weissman (1994). “Estimating the Extremal index,” *Journal of the Royal Statistical Society B*, 56, 515 – 528.
- [71] Tawn, J.A. (1988). “Bivariate Extreme Value Theory: Models and Estimation,” *Biometrika* 75, 397-415.
- [72] Tsay, R.S. (1988). “Outliers, Level Shifts, and Variance Changes in Time Series,” *Journal of Forecasting*, 7, 1 – 20.
- [73] von Mises, R. (1936). “La Distribution de la Plus Grande de  $n$  Valeurs,” *Selected papers of Richard von Mises* 2, 271 – 294. American Mathematical Society. Providence.