CIMAT

Centro de Investigación en Matemáticas, A.C.

Índice Extremo en Procesos Heteroscedásticos

T E S I S

Que para obtener el grado de:

Maestro en Ciencias

Con orientación en:

Probabilidad y Estadística

P r e s e n t a:

Inder Rafael Tecuapetla Gómez

Guanajuato,México. Diciembre de 2006.

Agradecimientos

Aprovecho este espacio para agradecer a todas aquellas personas que han contribuido a la elaboración de este trabajo.

A mi directora de tesis, **Dra. Graciela González Farías**. Es claro que en unas cuantas líneas no podré expresar todas las acciones que me hacen estar agradecido con ella, sin embargo lo intentaré: Por el tema de tesis, por el ambiente de trabajo, por las recomendaciones, por los trabajos, ... Por todo esto, mi reconocimiento.

A mis sinodales, **Dr. Joaquín Ortega Sánchez** y **Dr. Jesús Armando Domínguez Molina**. Al primero por haberme ayudado a dar mis primeros pasos en el amplísimo mundo de los valores extremos. Al segundo por ser parcialmente responsable de que tomara la estadística como mi área de estudio. A ambos por los comentarios generados en la revisión de este trabajo.

Al **Consejo Nacional de Ciencia y Tecnología**, **CONACyT**, por el apoyo económico para realizar estudios de maestría a través de la beca no. 192458 y el apoyo parcial a esta tesis a través del proyecto 45974-F.

Al **Ministerio de Ciencia y Tecnología** español, quien a través del proyecto SEJ2004-04101-ECON, apoyo parcialmente este trabajo.

Al **Dr. José Olmo** y al **Dr. Thomas Mikosch** ya que sus comentarios en la parte final de este proyecto enriquecieron las conclusiones del mismo.

A mi familia: Mi compañera de ruta, **Diana Lizette Jiménez Ortiz**. Mis padres, **Rafael Tecuapetla Huerta** y **Rosalba Gómez Méndez**. Mi hermano: **Gibrán Iván Gómez Méndez**.

A todos mis amigos.

Guanajuato, México — Inder Rafael Tecuapetla Gómez

Diciembre 15, 2006.

Prefacio

La literatura econométrica llama *hechos estilizados* a un conjunto de características compartidas por muchas series de tiempo financieras. Una de estas serie es la de log-returns (*log-rendimientos*), dada su importancia en los mercados financieros. Desde inicios de los sesentas, se ha incrementado el conocimiento de las causas y consecuencias de los hechos estilizados. El vertiginoso crecimiento en la capacidad de almacenamiento y manejo de la información hace posible tengamos trabajos como los de Cont [8], donde hallamos un compendio enciclopédico de los hechos estilizados más comunes.

Quizás el hecho estilizado que ha despertado mayor interés en diversos grupos de investigación, son los cúmulos de variabilidad, *volatility clustering*. Este fenómeno es asociado a la volatilidad de un activo financiero, es decir, la desviación estándar del cambio en el valor de un instrumento financiero con un horizonte temporal específico, junto con los cúmulos que se originan después de un fuerte e inesperado cambio en el nivel de dicho activo. Comentarios como los de Shepard [35], destacan la importancia de describir el cambio en el tiempo del valor de un activo financiero. El manejo de riesgo, por ejemplo, requiere un estimado de la volatilidad del precio de un activo para poder fijar su valor futuro.

La modelación de la volatilidad, ha tenido un desarrollo importante en los último 25 años. A partir del artículo seminal de Engle [12], donde se proponen a los modelos ARCH para describir la varianza no constante de una serie temporal, han surgido modelos capaces de describir apropiadamente este fenómeno Más recientemente, los denominados *modelos de volatilidad estocástica* intentan contribuir a una mejor descripción de este fenómeno.

A pesar de las críticas recibidas a partir de su presentación por Bollerslev [5], los modelos GARCH representan la más útil y popular herramienta para describir la evolución en el tiempo del precio de un activo en los mercados fianancieros. Particular atención es puesta por Mikosch y Starica [28] en la incapacidad de estos modelos, específicamente el GARCH(1,1), para capturar apropiadamente la dependencia en las colas de las series de rendimientos. Una primera lectura a estas líneas no parecen

conectar los cúmulos de variabilidad con la dependencia en las colas de una serie temporal. Sin embargo, la teoría de valores extremos aclara esta duda.

Muchos son los textos que estudian la teoría clásica de valores extremos, es decir, la distribución asintótica del máximo de una sucesión de variables aleatorias bajo el supuesto de independencia. En Leadbetter et al. [23] encontramos un estudio fundamental en la extensión de esta teoría al caso de procesos con dependencia débil. El parámetro clave en la extensión de esta teoría es el denominado *índice extremo*, inicialmente presentado por Loynes [24]. La caracterización dada por Leadbetter [22] de este número, nos permite ligar dependencia en las colas de procesos estacionarios con cúmulos alrededor de un punto de la serie.

Esta tesis está organizada en cuatro capítulos. En el Capítulo 1, damos una breve revisión a los hechos estilizados encontrados en la literatura econométrica. También, incluimos un ejemplo que enfatiza la necesidad de la volatilidad en el manejo de riesgo. El Capítulo 2, contiene los principales modelos heteroscedásticos utilizados en econometría para describir los hechos estilizados. Prestamos especial atención a las representaciones de los modelos ARCH y GARCH. Además, incluimos los resultados más importantes del artículo de Straumann y Mikosch [37], en el cual se establecen estacionariedad, invertibilidad y estimabilidad de los modelos GARCH(1,1) utilizando la teoría de *ecuaciones en recurrencias estocásticas*.

En el Capítulo 3, incluimos los principales resultados relacionados con la teoría de valores extremos. La primera parte de este capítulo se basa en el trabajo de Leadbetter; se presentan condiciones necesarias y suficientes para la existencia del índice extremo de un proceso estrictamente estacionario. En el resto de este tercer capítulo, se presentan los resultados que permiten establecer nuevas familias de estimadores del índice extremo; los estimadores clásicos se basan en la construcción de un proceso puntual de excedencias conglomeradas. Parte importante en esta tesis es el método por *niveles moderadamente altos*, Olmo [32], los principales detalles de este trabajo se incluyen en este tercer capítulo.

Finalmente, el Capítulo 4 presenta los estudios de simulación realizados sobre varias series de log-returns. Establecemos parámetros apropiados para la convergencia del índice extremo. Presentamos estimadores de la función de densidad del estimador del índice extremo basado en el método de Olmo. El resultado más importante de la tesis es dado en este capítulo, los modelos GARCH(1,1) *subestiman la dependencia en la colas de los rendimientos diarios de un activo financiero.*

Una aplicación de la teoría de valores extremos al análisis de series de tiempo, ésta es una buena

manera de resumir nuestro trabajo. Ajustando un GARCH(1,1) a una serie financiera real, y mediante la simulación de datos bajo este modelo, realizamos una procedimiento de inferencia estadística en un proceso estocástico con dependencia. Apartarse de los supuestos clásicos del modelo GARCH, incrementa el costo computacional de calcular el ajuste del modelo, así como también el de calcular el estimador del índice extremo. Nuestras rutinas computacionales pueden extenderse para realizar el mismo estudio de inferencia estadística en procesos más complejos.

Índice general

1. Introducción **1**

1.1. Estudios Empíricos 1

1.2. Importancia de la variabilidad en finanzas 3

2. Modelos Heteroscedásticos **9**

2.1. Modelación 9

2.2. ARCH 10

2.3. GARCH 11

2.4. Extensiones 13

2.4.1. ln ARCH y N-ARCH 13

2.4.2. EGARCH 14

2.4.3. IGARCH 15

2.5. GARCH(1,1) 16

2.5.1. Estacionariedad 16

2.5.2. Invertibilidad 18

2.5.3. Cuasi Máxima Verosimilitud 20

2.6. Hechos estilizados y GARCH(1,1) 23

2.6.1. Análisis exploratorio 23

2.6.2. Validez del supuesto GARCH 25

3. Teoría de Valores Extremos **27**

3.1. Introducción 27

3.2. Resultados bajo la Condición $D(u_n)$ 32

3.2.1. Criterios para existencia de IE 36

3.2.2. Proceso Puntual de Cúmulos 39

3.3. Propuesta de Olmo 43
3.3.1. Estimadores 45
3.3.2. Inferencia Estadística 47

4. Aportaciones **51**
4.1. Estudio del Comportamiento del Estimador de Olmo 51
4.1.1. Para determinar competitividad con otros estimadores 51
4.1.2. Para determinar número de bloques y umbrales 52
4.2. Clustering en GARCH 56
4.3. Distribución del Índice Extremo 57

Capítulo 1

Introducción

1.1. Estudios Empíricos

En la literatura econométrica se reportan algunas características comunes a múltiples series de tiempo. Para ejemplificar estas característica consideraremos la serie de tasas de cambio entre el yen japonés y el dólar estadounidense (JPY-USD). La Figura 1.1 exhibe el comportamiento de dicha serie. A priori utilizaremos técnicas típicas del análisis de series de tiempo. El evidente comportamiento errático de esta serie nos hace cuestionar incluso del supuesto básico de media constante. En este caso, la media muestral de esta serie es 156.44. Es usual utilizar transformaciones o filtros para estabilizar las series de tiempo a estudiar. A continuación describiremos una transformación ampliamente utilizada en el ámbito financiero.

Sea X_t el valor de la serie en el punto t. Típicamente el rendimiento del precio de un activo del tiempo $t-1$ al t se calcula como

$$\frac{X_t}{X_{t-1}} - 1 = -\left(1 - \frac{X_t}{X_{t-1}}\right)$$

Si asumimos que el precio del activo no varía mucho del tiempo $t-1$ al t, $X_t/X_{t-1} \approx 1$. Del cálculo se sabe que $\ln(1-x) \approx -x$, cuando $x \to 0$. Entonces

$$\ln\left(\frac{X_t}{X_{t-1}}\right) \approx \frac{X_t - X_{t-1}}{X_{t-1}}.$$

Esta es la denominada transformación de *rendimientos* o *returns*. La Figura 1.2 muestra los rendimientos de la serie JPY-USD. De aquí en adelante en nuestro trabajo utilizaremos la serie de rendimientos y los denotaremos como R_t.

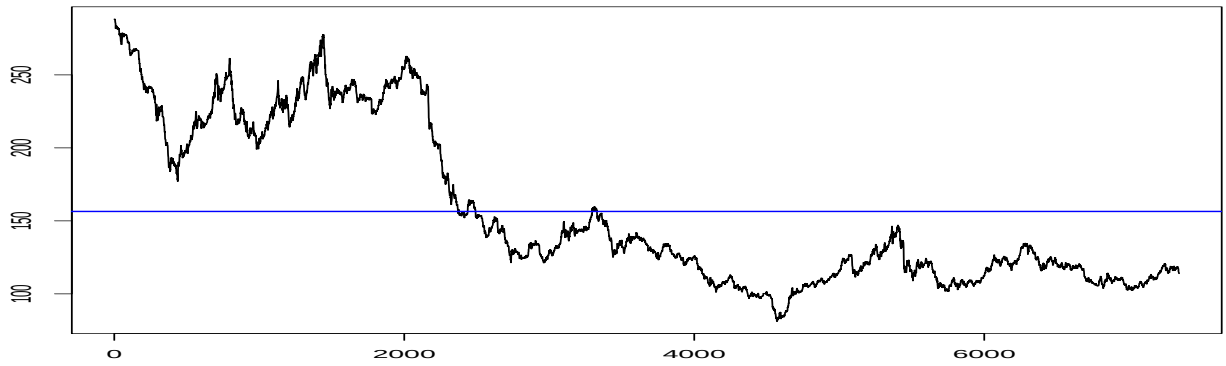


Figura 1.1: Tasas de cambio diarias entre yen japonés y dólar estadounidense. Ventana de observación: 1 de Febrero de 1971 a 1 de Mayo de 2006. Número de observaciones: 7344. Nótese que al inicio de la serie se observan precios (valores) altos, posteriormente valores bajos. La línea recta representa la media muestral del proceso.

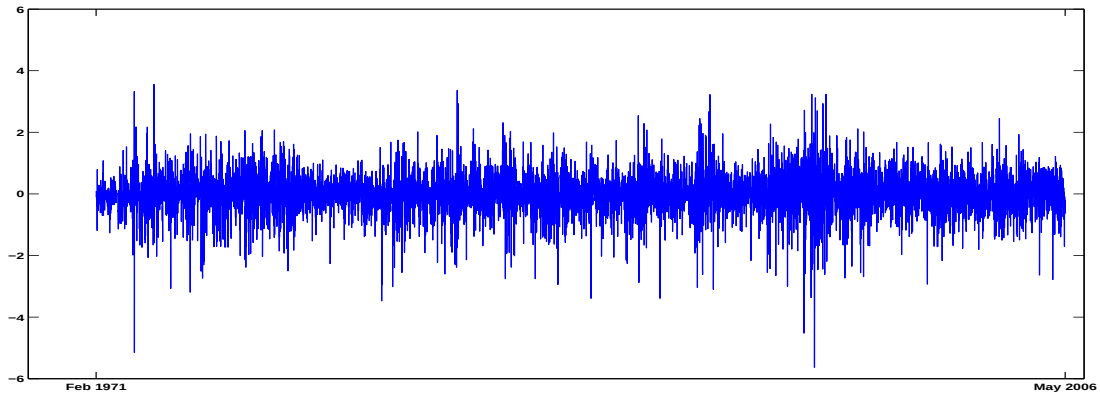


Figura 1.2: Ésta es la figura clásica de estudio en el análisis de series de tiempo financieras. La transformación utilizada se denomina en algunas ocasiones como log-returns.

Comparar los cuantiles teóricos de una normal estándar con los cuantiles muestrales de los log-returns puede darnos una primera idea de qué clase de función de probabilidad describe apropiadamente el comportamiento de estos últimos. La Figura 1.3 muestra este comparativo. Muchos estudios reportan esta gráfica y al igual que aquí, se manifiesta la necesidad de una distribución de colas más pesadas que la distribución normal. Véase Cont [8]. Esta característica se denomina Colas Pesadas en la densidad de los rendimientos.

La función de autocorrelación muestral nos permite dar un primer vistazo a la estructura de dependencia en una serie de tiempo. La Figura 1.4 muestra tal gráfico para los rendimientos (izquierda) y para los rendimientos al cuadrado (derecha). El gráfico de la izquierda nos sugiere que los rezagos

9 y 10 tiene correlaciones significativamente (bandas del 95 %) distintas de cero. Sin embargo, el gráfico de la derecha nos dice que los rendimientos al cuadrado tienen todos sus rezagos (hasta 20) significativamente distintos de cero. Estos dos gráficos nos sugieren la existencia de una estructura de dependencia fuerte, en el sentido de que la serie original es dependiente de sus rezagos grandes. Esto suele reportarse como la existencia de DEPENDENCIA A LARGO PLAZO en los rendimientos.

Figura 1.3: Los extremos de este gráfico sugieren que la distribución marginal de los log-returns posee cola más pesada que la normal.

Características tales como las colas pesadas, la dependencia a largo plazo y la varianza no constante, son llamadas en la literatura como HECHOS ESTILIZADOS. El efecto apalancamiento, correlación negativa entre un activo y su rendimiento, la intermitencia, la irregularidad en los periodos de variabilidad , y los *conglomerados en la varianza* son también conocidos como hechos estilizados. Aparentemente los rendimientos de un activo generan muchos retos para las herramientas clásicas en el análisis de series de tiempo. Dentro de los hechos estilizados comentados en la literatura, estamos interesados en la volatilidad o varianza conglomerada. A continuación damos una justificación de la importancia de este fenómeno en los estudios financieros.

1.2. Importancia de la variabilidad en finanzas

En esta sección seguiremos las ideas presentadas por Shepard [35] para proveer una correcta justificación de la importancia de estudiar la variabilidad en una serie de tiempo financiera. Re-

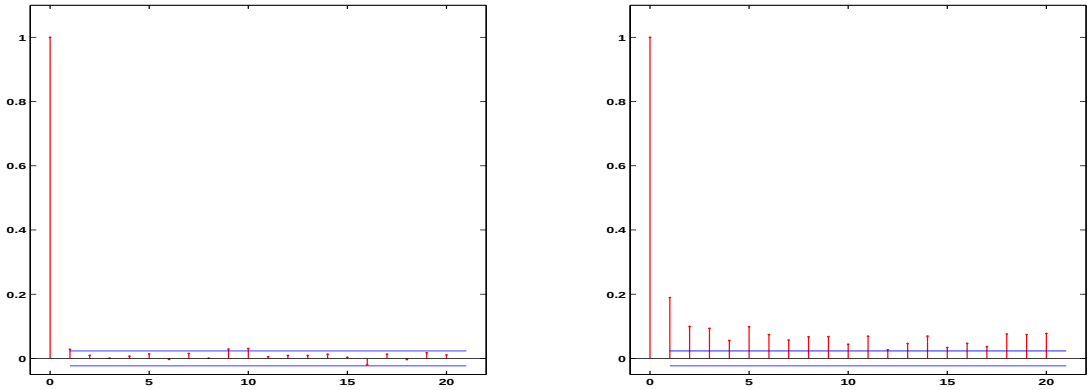


Figura 1.4: Comportamiento de las funciones de autocorrelación de las series de rendimientos y de rendimientos al cuadrado. De acuerdo a esta gráfica, existe una dependencia a largo plazo en ambas series. Las líneas horizontales corresponden a las bandas de confianza al 95 %.

comendamos ampliamente la lectura de este artículo dada la clara exposición de la problemática alrededor de los hechos estilizados, además de las muchas referencias encontradas ahí.

Supongamos que los cambios en los precios ocurren como resultado de un número aleatorio n_t de movimientos en los activos a lo largo de un día de actividades. Por tanto $R_t = \sum_{i=1}^{n_t} x_{it}$, donde x_{it} están idéntica e independientemente distribuidas (*i.i.d.*) y n_t es algún proceso de Poisson. Supongamos que n_t es independiente del tiempo, esta característica la heredará R_t. Se puede complicar el problema asumiendo que n_t sea dependiente del tiempo, lo cual podrá producir aglomeración en varianza en distintos puntos de la serie de rendimientos.

Para ejemplificar consideremos el valor de algún activo de seguros S. Supongamos que S sigue una difusión geométrica $dS = \mu S dt + \sigma S dz$, así que $d \ln S = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dz$. Es posible definir un activo C, como función del precio considerado S. Los economistas llaman a tales activos *contingentes* o *derivados*. El ejemplo clásico de un activo derivado es una *opción*, aquí se permite al dueño de la opción tener derecho, pero no obligación, de comprar el activo correspondiente a un precio en el futuro. Consideremos la opción *europea*; el dueño compra la opción al tiempo T y tiene el derecho de comprar el activo a precio fijo K en la fecha de expiración $T + v$. Un posible K sucede cuando éste iguala a $S(T)$, el precio de hoy. La dependencia de S sobre el tiempo se muestra a continuación.

El valor de la opción europea en la fecha de expiración es

$$C(T + v) = max\{S(T + v) - K, 0\}.$$

Mientras esta ecuación determina el valor de la opción al instante $T + v$, la opción será comprada

al tiempo T, debemos determinar el valor de compra. Una solución muy sencilla es calcular el valor esperado de la opción,

$$\exp(-rv)\,\mathbb{E}_{S(T+v)|S(T)}[c(T+v)],$$

donde r es una tasa de interés *sin* riesgo. Sin embargo, esto pasa por alto el hecho que los posibles compradores esperan rendimientos más altos en activos de riesgo alto que en activos sin riesgo. Por tanto, el mercado típicamente no valora activos por medio de su valor esperado. Esto sugiere la introducción de una función de utilidad que permita mayor flexibilidad en la determinación del precio del activo.

La complejidad agregada a una función de utilidad puede ser evitada usando algunas propiedades de difusión y suponiendo que las transacciones son continuas y libres de costo. Para esto construimos un portafolio π constituido por θ de las propiedades compartidas y por compartir un activo derivado C. Entonces el valor del portafolio se vuelve

$$\begin{aligned} d\pi &= \theta dS - dC \\ &= \theta\,(\mu S dt + \sigma S dz) - (C_s \mu S + C_t + \frac{1}{2} C_{ss}\sigma^2 S^2)\,dt - C_s \sigma S dz \\ &= (\theta - C_s)(\mu S dt - \sigma S dz) - (C_t + \frac{1}{2}\,C_{ss}\sigma^2 S^2)\,dt, \end{aligned}$$

usando el Lema de Itô, donde $C_t = \partial C/\partial t$ y $C_s = \partial C/\partial S$. Seleccionando $\theta = C_s$ en cada periodo de tiempo el inversionista asegura que $d\pi$ sea libre de riesgo instantáneamente, eliminando toda dependencia de dz. Este es el resultado debido a Black y Scholes, en el que se hace a $d\pi$ una función determinística del tiempo.

Como este portafolio es libre de riesgo, su rendimiento debe ser la tasa de interés libre de riesgo r, de otro modo los compradores tendrán oportunidad de *arbitraje* y de beneficios instantáneos sin riesgo. La tasa de interés sin riesgo puede ser tomada como el rendimiento en una ventana muy corta de tiempo. Así, el portafolio sin riesgo está dado por

$$\begin{aligned} d\pi = r\pi dt &= r(C_s S - C)\,dt \\ &= -(C_t + \frac{1}{2} C_{ss}\sigma^2 S^2)dt, \end{aligned}$$

implicando que el activo derivado sigue la ecuación diferencial estocástica

$$C_t + \frac{1}{2} C_{ss}\sigma^2 S^2 + rC_s S = rC, \text{con condición final } C = \text{máx}\{S - K, 0\}.$$

Nótese que esta ecuación *no* depende de μ o de las preferencias al riesgo de los compradores.

Definamos una nueva difusión con media rS^* y varianza $\sigma^2\, S^{*2}$, a saber, $dS^* = rS^*\, dt + \sigma\, S^*\, dz$. Este es el proceso de riesgo neutral. Usando la log-normalidad de la difusión tenemos

$$\ln S^*(T+v) \mid \ln S(T) \sim \mathrm{N}(\ln(S(T) + (r - \sigma^2/2)v, \sigma^2\, v).$$

De donde se sigue el resultado de Black-Scholes sobre la valuación de la opción v periodos adelante del instante de compra de la opción, a saber:

$$b_{s_v}(\sigma^2) = \exp(-rv)\, \mathbb{E}[\text{máx}\{S^*(T+v) - K, 0\} \mid S(T)],$$

la cual es

$$b_{s_v}(\sigma^2) = S(T)\Phi(d) - K\exp(-rv)\Phi(d - \sigma\sqrt{v}), \tag{1.1}$$

donde

$$d = \frac{\ln(S(T)/K) + (r + \sigma^2/2)v}{\sigma\,\sqrt{v}}. \tag{1.2}$$

Notemos que v y K son dados por normas institucionales, $S(T)$ y r son observados, dejando sólo a σ^2 como desconocida. En realidad, los precios de opciones están valuando variabilidad. Esta es una de las principales razones por la cual la variabilidad será el objeto central de estudio en los análisis financieros y no el efecto del promedio de las observaciones como en el análisis clásico de series de tiempo.

Empíricamente, hay dos formas de usar (1.1). La primera es estimar σ^2 y luego calcular el precio de la opción. La segunda consiste en observar precios de las opciones para calcular el valor de σ^2. Este último método es llamado variabilidad estimada implicada.

La principal dificultad con estos análisis es el supuesto inicial del proceso, que los rendimientos siguen una difusión geométrica. Dejando de lado la imperfección inherente a cualquier modelo matemático, debe quedar clara la utilidad de contar con algún modelo capaz de describir apropiadamente tanto los hechos estilizados como la variabilidad observada empíricamente en distintas series financieras.

Para finalizar esta parte introductoria debemos puntualizar cuál es el principal objetivo de nuestro trabajo. Comencemos diciendo que típicamente los estudiosos en finanzas intentan explicar o racionalizar un movimiento específico del mercado a partir de eventos o anuncios políticos o económicos, esto se conoce como el *enfoque tipo evento* y consideramos que es una manera sensata de describir una serie de tiempo financiera desde el punto de vista cualitativo. No resulta difícil imaginar que diferentes activos, influenciados por diferentes eventos, exhibirán diferentes propiedades. Sin

embargo, como vimos a lo largo de la primera sección de este capítulo existen propiedades similares desde el punto de vista estadístico: las variaciones aleatorias similares en los valores de los activos sí comparten ciertas propiedades estadísticas.

Debido a que los hechos estilizados representan un común denominador entre las propiedades exhibidas por muchas series de activos, vale la pena contar con modelos capaces de describirlos. Haciendo esto, ganamos en generalidad, pero perdemos en precisión de las declaraciones que podemos hacer acerca de los activos. Así los hechos estilizados son formulados en términos de *propiedades cualitativas* de los activos y pueden no ser lo suficientemente precisos para distinguir entre diferentes modelos paramétricos. Estas propiedades son tan restringidas que aún no es fácil obtener un proceso estocástico *ad hoc* que posea el mismo conjunto de propiedades que los activos reales y uno tiene que estar dispuesto a reproducirlos con un modelo.

De acuerdo al enfoque tipo evento, un anuncio político o económico puede impactar de tal manera a un activo financiero, que su precio dependerá de este anuncio mucho tiempo después de la realización de tal anuncio. Estas líneas pueden traducirse a un gráfico, observando la Figura 1.2 es evidente la existencia de ciertos cúmulos de puntos a lo largo de la serie. La presencia de dichos cúmulos es *variable* dentro de la serie. Llamaremos a este fenómeno Conglomeración en Varianza. Éste será nuestro principal tema de estudio, específicamente, bajo ciertos supuestos, proveeremos de un estimador asintótico del tamaño esperado de los cúmulos. Como un segundo problema a tratar en este trabajo, estudiamos la conveniencia de estimar los cúmulos de una serie a partir de estimar los cúmulos del modelo estadístico que mejor describe los hechos estilizados y la conglomeración en varianza: los modelos Garch. Antes de abordar esos problemas, describiremos a grosso modo el camino que ha seguido la modelación de la variabilidad de una serie de tiempo financiera.

Capítulo 2

Modelos Heteroscedásticos

2.1. Modelación

De acuerdo a Cox [10], una división conceptualmente útil de los modelos heteroscedásticos está dada por los modelos *guiados por observación* y los modelos *guiados por parámetros*. Concentrémonos en modelos paramétricos con la siguiente estructura

$$y_t \mid W_t \sim \mathrm{N}(0, \sigma_t^2).$$

Modelos Guiados por Observación. Un modelo tal, hace que W_t sea una función de valores rezagados de y_t. De acuerdo a Shepard [35], modelos construidos bajo este enfoque son exitosos por las siguientes tres razones

Primera Desde el punto de vista estadístico, combinando las densidades dadas en esta forma, permite construir la función de verosimilitud vía una *descomposición de predicción*. Esto hace tanto la estimación como las pruebas de hipótesis directas, al menos en principio.

Segunda Desde el punto de vista económico, la teoría financiera frecuentemente se especifica usando momentos *un paso adelante*.

Tercera Desde el punto de vista práctico, toda la teoría desarrollada para los modelos autorregresivos y de medias móviles son aplicables. Sin embargo, tengamos en mente que éstos fueron propuestos para describir la media condicional de un proceso, aquí estamos hablando de la varianza condicional, por tal razón algunos autores proponen nuevos enfoques.

Modelos Guiados por Parámetro. Estos modelos también son conocidos como *modelos en el espacio de estados*, y son tales que permiten que W_t sea una función de algún componente

no observado o *latente*. Estos modelos son más difíciles de manejar estadísticamente que los análogos guiados por observación, sin embargo existen razones para trabajar con ellos:

1. Sus propiedades son fáciles de encontrar, entender, manipular y generalizar al caso multivariado.
2. También, éstos tienen representación más simple en *tiempo continuo*, lo cual es importante dado que mucho de la teoría moderna en finanzas emplea *difusiones*.

Dentro de los modelos guiados por observación, encontramos a los modelos ARCH. Desde su presentación hasta la fecha, han sido objeto de estudios teóricos y empíricos. Enseguida damos la representación de estos modelos. También presentamos su generalización *natural*, los llamados modelos GARCH. Comentaremos brevemente los problemas alrededor de estos dos modelos básicos, los cuales generaron múltiples extensiones, algunas de éstas son descritas a continuación. Se ha incluido este apartado, porque es conveniente conocer una familia de modelos ampliamente usada por la comunidad econométrica debido a su eficiencia para describir apropiadamente los hechos estilizados. Por otro lado, hacemos énfasis en que no son, por mucho, los únicos modelos creados para este fin y debe ser claro que encontrar un modelo general para describir a detalle todos los hechos estilizados es tarea difícil y en proceso. Los siguientes pertenecen a los modelos guiados por observación.

2.2. ARCH

La representación original de los modelos ARCH fue dada por Engle [12]. Supongamos que Y_t está dado por

$$Y_t = \mu_t + X_t, \quad t = 1, \ldots, n, \tag{2.1}$$

donde $\mu_t \in \mathbb{R}$. Sea $\mathcal{F}_{t-1} = \sigma(Y_{t-1}, \mu_{t-1}, Y_{t-2}, \mu_{t-2}, \ldots)$, la σ-álgebra generada por la información disponible al momento $t-1$, y supongamos que

$$X_t \mid \mathcal{F}_{t-1} \sim \mathrm{N}(0, h_t^2) \tag{2.2}$$

donde

$$h_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_q X_{t-q}^2, \tag{2.3}$$

con

$$\alpha_0 > 0, \tag{2.4}$$
$$\alpha_i \geq 0,\ i = 1, \ldots, q,$$

para asegurar que la varianza condicional sea no negativa. La ecuación (2.3) define un proceso ARCH(q).

En el modelo de regresión clásico, un impacto es representado por una desviación de Y_t respecto de la media condicional, en este caso de μ_t, o equivalentemente, un valor grande en valor absoluto de X_t. En (2.1), la varianza del error actual X_t, condicionado sobre los valores realizados de los errores rezagados X_{t-i}, $i = 1, \ldots, q$ es una función creciente de la magnitud de los errores rezagados sin importar el signo. Por tanto, errores grandes de signo positivo o negativo tienden a estar seguidos por un error grande de signo positivo o negativo. Análogamente para los errores pequeños. El orden de los rezagos q determina la longitud del tiempo por el cual un impacto persiste en el condicionamiento de la varianza de errores subsecuentes. Mientras más grande sea q, más largo serán los episodios de volatilidad.

Este modelo incorpora el pasado de las variables en cuestión para predecir el futuro de las mismas. Obviamente esto representa nuestra habilidad de predecir valores futuros como función del tiempo. Esta es quizás la mayor fortaleza, al menos al nivel interpretativo, de este modelo. Concretando: El modelo ARMA expresa la media condicional como función de la información disponible, el modelo ARCH hace lo mismo para la varianza condicional, y por lo tanto cambia la varianza en la predicción sobre el tiempo. Nuestro uso del término 'ARCH' no sólo se refiere al introducido por el modelo de Engle. Entenderemos por 'ARCH', el fenómeno de heteroscedasticidad condicional en general y todos los modelos que capturan tal fenómeno.

2.3. GARCH

Mientras más grande es el orden de los rezagos, más parámetros incorporamos al modelo dado por (2.3). Trabajos empíricos sugieren que rezagos de orden grande son requeridos en la función de varianza condicional. Por tanto, resulta conveniente contar con algún modelo capaz de representar la variabilidad de una serie financiera similarmente a como lo hace el modelo ARCH, pero que reduzca el espacio paramétrico usado en la especificación de la función de varianza condicional. La familia

de modelos ARCH generalizados o GARCH, dados por Bollerslev [5] surgieron como una respuesta a este problema.

Mantengamos en mente la ecuación (2.1) y consideremos ahora

$$X_t \mid \mathcal{F}_{t-1} \sim \mathrm{N}(0, \sigma_t^2), \tag{2.5}$$

donde

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \tag{2.6}$$

con $\alpha_0 > 0,\ \alpha_i \geq 0,\ i = 1, \ldots, p,\ \beta_i \geq 0,\ i = 1, \ldots, q$. Aquí se captura la parte autorregresiva de los errores en el modelo de regresión más los rezagos de la función de varianza condicional. Las ecuaciones (2.5) y (2.6) definen un modelo GARCH(p,q).

La generalización de ARCH a GARCH es similar a la generalización de un proceso AR a un ARMA. La idea fundamental es que un GARCH puede representar parsimoniosamente un proceso ARCH de orden alto, como se muestra en el Teorema 2.1. Dada la importancia del operador de rezagos en series de tiempo, incluimos la siguiente definición.

Definición 2.1 (Operador de Rezagos). El *operador de rezagos* B, está dado por

$$B(y_t) = y_{t-1}. \tag{2.7}$$

En general,

$$B^p(y_t) = y_{t-p},\ p \in \mathbb{Z}, \tag{2.8}$$

donde $B^0(y_t) = y_t$.

Un *polinomio en el operador de rezagos* tiene la forma

$$\theta(B) = \theta_1 B + \ldots + \theta_q B^q, \tag{2.9}$$

donde $\theta_1, \cdots, \theta_q$ son parámetros o constantes. Las *raíces* de tal polinomio están definidas como los q valores de B que satisfacen

$$\theta(B) = 0.$$

Se dice que una *raíz yace fuera del círculo unitario* si su módulo es mayor que uno.

Teorema 2.1. *Un proceso* GARCH *(p,q), es un proceso* ARCH*(∞) con una estructura* racional *sobre los coeficientes de los rezagos, siempre que las raíces de* $1 - \beta(B)$ *yazcan fuera del círculo unitario.*

Demostración. Escribamos (2.6) como

$$\sigma_t^2 = \alpha_0 + \alpha(B)\, X_t^2 + \beta(B)\, \sigma_t^2, \tag{2.10}$$

donde $\alpha(B)$, $\beta(B)$ son polinomios en el operador de rezagos de orden p y q, respectivamente. Si las raíces de $1 - \beta(B)$ yacen fuera del círculo unitario, entonces aún podemos escribir

$$\sigma_t^2 = \frac{\alpha_0}{1 - \beta(1)} + \frac{\alpha(B)}{1 - \beta(B)}\, X_t^2 \tag{2.11}$$

y entonces

$$\sigma_t^2 = \alpha_0 \left(1 - \sum_{k=1}^{p} \beta_k\right)^{-1} + \sum_{k=1}^{\infty} \delta_k\, X_{t-k}^2, \tag{2.12}$$

donde δ_k es el correspondiente coeficiente en la expansión de $\alpha(B)/(1 - \beta(B))$. ■

2.4. Extensiones

Para Engle en su exposición original de los modelos ARCH, resultó natural suponer que la varianza condicional debía ser una combinación lineal de los errores al cuadrado y que la distribución condicional de éstos debía ser normal. Muchos estudios empíricos efectuados hasta ahora dan evidencia de que estos supuestos no son apropiados en la mayoría de los casos. A continuación discutiremos algunos de las representaciones alternativas para la varianza y distribución condicionales de los procesos considerados en este trabajo. Dichas representaciones han probado ser útiles en la práctica.

2.4.1. ln ARCH y N-ARCH

Al estimar un modelo ARCH, frecuentemente se encontraban coeficientes estimados α_i's negativos. Para evitar este problema Geweke [13], Pantula [33] y Milhoj [27] sugieren el modelo ln ARCH, cuya representación está dada por

$$\ln(\sigma_t^2) = \alpha_0 + \alpha_1 \ln(X_{t-1}^2) + \cdots + \alpha_q \ln(X_{t-q}^2). \tag{2.13}$$

Así, tomando exponencial a ambos lados de (2.13), σ_t^2 sea no negativa, y por tanto, las restricciones sobre las α_i's para asegurar que la varianza condicional sea no negativa positiva ya no son requeridas.

Para determinar si el modelo lineal (2.3) o el logarítmico (2.13) proporcionan un mejor ajuste sobre los datos disponibles, Higgins y Bera [18] propusieron un modelo ARCH no lineal (NARCH). La representación de este modelo está dada por

$$\sigma_t^2 = \left[\phi_0\, (\varpi^2)^\delta + \phi_1\, (X_{t-1}^2)^\delta + \cdots + \phi_q\, (X_{t-q}^2)^\delta\right]^{1/\delta}, \tag{2.14}$$

donde, $\varpi^2 > 0,\ \phi_i \geq 0, \delta > 0$ y los ϕ_i's son tales que $\sum_{i=0}^{q} \phi_i = 1$

Escribamos lo anterior como

$$\frac{\sigma_t^{2\delta} - 1}{\delta} = \phi_0 \frac{\varpi^{2\delta} - 1}{\delta} + \phi_1 \frac{X_{t-1}^{2\delta} - 1}{\delta} + \cdots + \phi_q \frac{X_{t-q}^{2\delta} - 1}{\delta}, \tag{2.15}$$

para hacer evidente que el modelo N-ARCH consiste en aplicar una transformación potencia de Box-Cox a ambos lados del modelo ARCH original. Es claro que $\delta = 1$ en (2.15) es equivalente a (2.3), por otro lado, cuando $\delta \to 0$, (2.15) aproxima el modelo ln ARCH. Higgins y Bera estimaron (2.14) en series de tasas de cambio semanal y encontraron que δ fue significativamente menor que uno y mucho más cercano a cero, indicando que aquellos datos favorecían un modelo ln ARCH sobre el ARCH lineal. Las posibles extensiones de estos procedimientos al modelo GARCH son directas.

2.4.2. EGARCH

Una posible limitación de las representaciones dadas arriba es que la varianza condicional σ_t^2 es simétrica en los X_t's rezagados. Nelson [30] sugirió que la simetría en (2.3) puede ser inapropiada para representar ciertas series financieras y basó sus críticas en estudios empíricos conocidos donde se encontró evidencia de que la varianza de algunos rendimientos están negativamente correlacionados con rendimientos pasados. El modelo ARCH clásico no puede capturar este fenómeno porque la magnitud de σ_t^2 no es afectada por el signo de ϵ_t, y por tanto, σ_t^2 no está correlacionada con los errores pasados. Nelson definió X_t en términos de la ecuación (2.6). Sugirió que en la representación ARCH general

$$\sigma_t^2 = \sigma^2(X_{t-1}, \ldots, X_{t-p}, \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2),$$

σ_t^2 puede verse como un proceso estocástico en el cual X_t es capaz de unir a la varianza condicional y a los errores. Su propuesta está dada por

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=i}^{p} \alpha_i\, g(X_{t-i}) + \sum_{i=i}^{q} \ln(\sigma_{t-i}^2), \tag{2.16}$$

donde

$$g(X_t) = \theta\, X_t + \gamma\, (|X_t| - E[|X_t|])\,. \tag{2.17}$$

Las ecuaciones (2.16) y (2.17), definen un modelo GARCH exponencial (EGARCH). Es fácil ver que la sucesión $\{g(X_t)\}$ es independiente y con media cero, y la varianza, de existir, será constante. Así, (2.16) representa un modelo lineal ARMA para $\ln(\sigma_t^2)$ con errores $g(X_t)$. Algunas propiedades del modelo EGARCH son

1. El error de la varianza condicional es lineal a pedazos en X_t, con pendientes $\alpha_i(\theta+\gamma)$ cuando X_t es positiva y $\alpha_i(\theta-\gamma)$ cuando X_t es negativa. Esto produce la asimetría en la varianza condicional.

2. El primer término en (2.17) permite que exista correlación entre el error y futuras varianzas condicionales. Por ejemplo, supongamos que $\gamma = 0$ y que $\theta < 0$. Entonces un X_t negativo causará que el error sea negativo y que el error actual en la varianza del proceso sea positivo.

3. El segundo término en (2.17) produce el llamado *efecto* ARCH. Supongamos que $\theta = 0$ y que $\gamma > 0$. Cada vez que la magnitud absoluta de X_t exceda su valor esperado, el error $g(X_t)$ será positivo. Así, impactos grandes incrementan la varianza condicional.

Trabajando con una serie de índice de precios, Nelson estimó un modelo EGARCH y encontró θ significativamente negativo y γ significativamente positivo. Estos resultados verifican la presencia del efecto ARCH así como también de una relación negativa entre rendimientos y varianzas futuras.

2.4.3. IGARCH

De acuerdo a Chan [6], cuando se considera un modelo GARCH(1,1) en la práctica, frecuentemente se encuentra que $\alpha_1 + \beta_1 \approx 1$. Cuando $\alpha_1 + \beta_1 = 1$, el proceso X_t ya no es estacionario y recibe el nombre de modelo GARCH(1,1) integrado o simplemente IGARCH(1,1). Una de las interpretaciones de este modelo es que la volatilidad es *persistente*. Para ver esto claramente, consideremos

$$\begin{aligned}
\mathbb{E}[\sigma_{t+1}^2 \mid \mathcal{F}_{t-1}] &= \mathbb{E}[\alpha_0 + \alpha_1 X_t^2 + \beta_1 \sigma_t^2 \mid \mathcal{F}_{t-1}] \\
&= \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 = \alpha_0 + \sigma_t^2. \\
\mathbb{E}[\sigma_{t+2}^2 \mid \mathcal{F}_{t-1}] &= \mathbb{E}[\mathbb{E}[\sigma_{t+2}^2 \mid \mathcal{F}_t] \mid F_{t-1}] \\
&= \mathbb{E}[\alpha_0 + \sigma_t^2 \mid \mathcal{F}_{t-1}] \\
&= 2\alpha_0 + \sigma_t^2.
\end{aligned}$$

En general, repitiendo el anterior argumento, tenemos

$$\mathbb{E}[\sigma_{t+j}^2 \mid \mathcal{F}_{t-1}] = j\sigma_0 + \sigma_t^2. \tag{2.18}$$

Nótese que se ha condicionado sobre $\mathcal{F}_{t-1}$ en lugar de $\mathcal{F}_t$ porque σ_{t+1} es medible con respecto a la σ-álgebra más pequeña $\mathcal{F}_t$ así que

$$\mathbb{E}[\sigma_{t+1}^2 \mid \mathcal{F}_t] = \sigma_{t+1}^2.$$

De acuerdo a la ecuación (2.18), la volatilidad de hoy afecta la predicción de la volatilidad de mañana y este efecto continúa perturbando la serie en el futuro. Por lo tanto, cualquier impacto sobre X_t^2 o sobre σ_t^2 será acarreado en el futuro, es decir, será persistente. Aunque el modelo IGARCH(1,1) puede describir apropiadamente un fenómeno como la variabilidad de una serie de tiempo, es recomendable tomar precauciones de acuerdo a lo expuesto en esta sección.

2.5. GARCH(1,1)

Esta sección presentará el modelo heteroscedástico más parsimonioso capaz de describir apropiadamente la variabilidad en una serie de tiempo financiera, el modelo GARCH(1,1). A diferencia de los presentados previamente, definiremos al modelo GARCH(1,1) como un modelo guiado por parámetros. Hacemos esto para ejemplificar las ventajas que tiene este enfoque. La presente sección está basada en el artículo de Straumann y Mikosch [37].

2.5.1. Estacionariedad

Se dice que una serie de tiempo X_t es un proceso GARCH(1,1) si ésta satisface

$$X_t = \sigma_t Z_t, \quad t \in \mathbb{Z},$$

y la ecuación de recurrencia estocástica

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad t \in \mathbb{Z},$$

donde $\{Z_t\}$ es una sucesión de vaiid simétricas con $\mathbb{E}[Z_t^2] = 1$. Las α's y β's son no negativas para garantizar la no negatividad de la volatilidad al cuadrado σ_t^2.

Observemos que la representación de un GARCH(1,1) sigue la siguiente *ecuación de recurrencia estocástica*

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{B}_t, \tag{2.19}$$

donde

$$\mathbf{X}_t = \begin{pmatrix} X_t^2 \\ \sigma_t^2 \end{pmatrix}, \quad \mathbf{A}_t = \begin{pmatrix} \alpha_1 Z_t^2 & \beta_1 Z_t^2 \\ \alpha_1 & \beta_1 \end{pmatrix}, \quad \mathbf{B}_t = \begin{pmatrix} \alpha_0 Z_t^2 \\ \alpha_0, \end{pmatrix}. \tag{2.20}$$

Esta representación corresponde a la de un modelo en el espacio de estados. Además, σ_t^2 al no ser observada nos sitúa en el enfoque de los modelos guiados por parámetros. Aquí, estamos prescindiendo del supuesto típico de normalidad.

El proceso de variabilidad σ_t^2 también satisface una ecuación de recurrencia como (2.19). Basta hacer

$$\mathbf{X}_t = \sigma_t^2, \quad \mathbf{A}_t = \alpha_1 Z_{t-1}^2 + \beta_1, \quad \mathbf{B}_t = \alpha_0. \tag{2.21}$$

Una propiedad importante en todo proceso estocástico es la *estacionariedad*. En el contexto de series de tiempo, este supuesto suele relajarse a estacionariedad de segundo orden, esto es, que ninguno de los dos primeros momentos sean función del tiempo. Damos una definición general de este concepto dada su importancia a lo largo de este escrito.

Definición 2.2 (Estacionariedad)**.** Se dice que el proceso estocástico $\{X_t\}_{t\in T}$ es *estacionario* si para cualesquiera $t_1, \ldots, t_k$ números reales y cualquier Δ entero, se cumple que

$$(X_{t_1}, \ldots, X_{t_k}) \stackrel{\mathcal{D}}{=} (X_{t_1+\Delta}, \ldots, X_{t_k+\Delta}),$$

donde $\stackrel{\mathcal{D}}{=}$ denota igualdad en distribución.

De acuerdo a Brandt [4] se sabe que una ecuación como (2.19) tiene una solución estacionaria si $\mathbb{E}\ln|\mathbf{A}_n| < 0$, $\mathbb{E}\ln^+|\mathbf{B}_n| < \infty$ y si el exponente *alto* de Lyapunov definido como

$$\hat{\gamma} = \text{ínf}\left\{\frac{1}{n}\,\mathbb{E}\ln\|\mathbf{A}_0\cdots\mathbf{A}_n\|,\ n\in\mathbb{N}\right\},$$

es negativo. Aquí $|\cdot|$ es cualquier norma en $\mathbb{R}^n$, $\|\mathbf{A}\| = \sup_{|\mathbf{x}|=1}|\mathbf{A}\mathbf{x}|$.

Por (2.20) tenemos

$$\mathbf{A}_n\cdots\mathbf{A}_0 = \mathbf{A}_n\prod_{i=0}^{n-1}(\alpha_1 Z_i^2 + \beta_1),$$

así

$$\begin{aligned}\hat{\gamma} &= \text{ínf}\left\{\frac{1}{n}\left(\mathbb{E}\ln\mathbf{A}_n + \sum_{i=0}^{n-1}\mathbb{E}\ln(\alpha_1 Z_i^2+\beta_1)\right),\ n\in\mathbb{N}\right\}\\ &= \text{ínf}\left\{\frac{1}{n}\mathbb{E}\ln\mathbf{A}_n + \mathbb{E}\ln(\alpha_1 Z_0^2+\beta_1),\ n\in\mathbb{N}\right\}\\ &= \mathbb{E}\ln(\alpha_1 Z_0+\beta_1).\end{aligned}$$

De acuerdo a estos cálculos, es suficiente que $\mathbb{E}\ln(\alpha_1 Z_0+\beta_1) < 0$, $\mathbb{E}\ln Z_0 < \infty$ para que el proceso GARCH(1,1) sea estacionario. Usando la herramienta típica, Nelson [29] establece que $\mathbb{E}\ln(\alpha_1 Z_0 + \beta_1) < 0$ es una condición necesaria para la estacionariedad de un GARCH(1,1).

Recordando que $\sigma_t^2 = \alpha_0+\alpha_1 X_{t-1}^2+\beta_1\sigma_{t-1}^2$, donde $\alpha_0 > 0,\ \alpha_1,\ \beta_1 \geq 0$. Si $g(x,s) = \alpha_0+\alpha_1 x^2+\beta_1 s^2$ entonces $|g(sz,s^2) - g(\tilde{s}z,\tilde{s}^2)| = (\alpha_1 z^2+\beta_1)|s^2-\tilde{s}^2|$, haciendo $C_1(z) = \alpha_1 z^2+\beta_1$, la condición

$\mathbb{E}\ln C_1(Z_0) < 0$ del Teorema 3.1 en [37], hace que $\mathbb{E}\ln(\alpha_1 Z_0^2 + \beta_1) < 0$, sea condición necesaria y suficiente para estacionariedad.

Más aún,

$$\tilde{\sigma}_{t,m}^2 = \alpha_0 \left(1 + \sum_{k=1}^{m-1} \prod_{i=1}^{k} (\alpha_1 Z_{t-i}^2 + \beta_1)\right) + c_0^2 \prod_{k=1}^{m} (\alpha_1 Z_{t-i}^2 + \beta_1),\ m > 0.$$

Como $\mathbb{E}\ln(\alpha_1 Z_0^2 + \beta_1) < 0$, elegimos p tal que $\mathbb{E}\ln C_1(Z_0) < p < 0$. Entonces la ley fuerte de los grande números implica que

$$\frac{1}{m}\sum_{i=1}^{m} \ln C_1(Z_i) < p \Leftrightarrow \prod_{i=1}^{m} C_1(Z_i) < e^{mp},$$

así $\prod_{k=1}^{m}(\alpha_1 Z_{t-k}^2 + \beta_1) \overset{\text{e.c.s.}}{\to} 0,\ m \to 0.$

Finalmente, por el teorema antes mencionado $\tilde{\sigma}_{t,m}^2 \overset{\text{c.s.}}{\to} \sigma_t^2$, tenemos

$$\sigma_t^2 = \alpha_0 \left(1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\alpha_1 Z_{t-i}^2 + \beta_1)\right) c.s.$$

2.5.2. Invertibilidad

Mencionamos anteriormente que en un modelo guiado por parámetros, tratamos con variables latentes. En el caso de un modelo GARCH(1,1), la variabilidad σ_t^2 no es observada. En tal caso, es natural aproximar la varianza de la siguiente manera:

1. Sea $\tilde{\sigma}_0^2 = c_0$, $c_0 \geq 0$ es una constante arbitraria.

2. Definir $\tilde{\sigma}_t^2 = g(X_{t-1}, \tilde{\sigma}_{t-1}^2)$, $1 \leq t \leq n$.

En nuestro modelo, *no lineal*, diremos que la solución estacionaria $\{X_t\}$ a la ecuación (2.19) es invertible si

$$|\tilde{\sigma}_t^2 - \sigma_t^2| \overset{\mathbb{P}}{\to},\ t \to \infty.$$

En otras palabras, invertibilidad garantiza que el algoritmo de arriba converge.

Hay una segunda interpretación, la cual clarifica la relación entre invertibilidad en los modelos ARMA y la noción dada arriba. Haciendo

$$\tilde{\sigma}_{t,m}^2 = \begin{cases} c_0^2 & m = 0 \\ g(X_{t-1}, \tilde{\sigma}_{t-1,m-1}^2) & m \geq 1, \end{cases} \tag{2.22}$$

tenemos la relación $\hat{\sigma}_t^2 = \hat{\sigma}_{t,t}^2$, para $t \geq 0$. Más aún, como suponemos que $\{X_t\}$ es estacionaria, $(\sigma_{t,m}^2, \sigma_t^2)_t$ es estacionaria para cada $m \geq 0$ fijo en virtud de la Proposición 2.3 en [37]. Así

$$\sigma_{t,m}^2 - \sigma_t^2 \stackrel{\mathrm{d}}{=} \sigma_{m,m}^2 - \sigma_m^2 = \hat{\sigma}_m^2 - \sigma_m^2,$$

para todo $m \geq 0$. Por lo tanto invertibilidad es equivalente a

$$\sigma_{t,m}^2 \xrightarrow{\mathbb{P}} \sigma_t^2, \ m \to \infty, \tag{2.23}$$

para todo $t \in \mathbb{Z}$ fijo. Como esto implica la existencia de una subsucesión que converge c.s., σ_t^2 es medible respecto a la σ-álgebra generada por las observaciones pasadas $\{X_{t-k} \mid k \geq 1\}$, i.e., para todo $t \in \mathbb{Z}$,

$$\sigma_t^2 = f(X_{t-1}, X_{t-2}, \ldots), \quad \text{c.s.},$$

para una función medible f. Si $\sigma_t^2 > 0$ c.s. para el modelo (2.19) la ecuación $\sigma_t^2 = f(X_{t-1}, X_{t-2}, \ldots)$ también implica

$$Z_t = \tilde{f}(X_t, X_{t-1}, X_{t-2}, \ldots), \quad \text{c.s.}$$

para todo $t \in \mathbb{Z}$, donde $\tilde{f}(x_0, x_1, x_2, \ldots) = x_0/(f(x_1, x_2, \ldots))^{1/2}$. Comparemos esta representación con la noción de invertibilidad en un ARMA. Recordemos que un modelo ARMA(p,q) con parámetros $(\phi_1, \ldots, \phi_p; \theta_1, \ldots, \theta_q)$ es un proceso estocástico $\{X_t\}$ que obedece a la siguiente ecuación de recurrencia

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_j Z_{t-j}, \quad t \in \mathbb{Z},$$

donde $\{Z_t\}$ es ruido blanco, con varianza σ^2, i.e., para todo $t \in \mathbb{Z}$, $\mathbb{E}Z_t = 0$ y $\mathrm{Var}Z_t = \sigma^2$ y $\mathrm{Cov}(Z_t, Z_s) = 0, \ t \neq s$. En este caso, invertibilidad estará definida como sigue: existe una sucesión de constantes absolutamente sumable $\{\pi_j\}_{j \geq 0}$ tal que

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \text{c.s.}$$

o, en otras palabras, la *perturbación* al tiempo t es un funcional lineal de las observaciones pasadas y presentes $\{X_{t-k} \mid k \geq 0\}$. Si los dos polinomios característicos $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ y $\theta(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$ no tienen raíces comunes, entonces el proceso ARMA es invertible si y sólo si $\theta(z) \neq 0$ para todo $z \in \mathbb{C}$ tal que $|z| < 1$.

Para modelos de series de tiempo *no lineales*, sin embargo, la invertibilidad puede ser un problema difícil. La noción de invertibilidad dada aquí es una adaptación a la noción introducida por Granger y Andersen [14] en el contexto general de un modelo autorregresivo de medias móviles no lineal.

Para Straumann y Mikosch [37], la aportación más importante de su trabajo es explícitamente formular y resolver problemas relevantes haciendo uso de ecuaciones estocásticas en recurrencia. Uno de tales problemas es determinar la invertibilidad del modelo GARCH(1,1), es decir, el hecho que

$$|\sigma_t^2 - \hat{\sigma}_t^2| \xrightarrow{\mathbb{P}} 0, \ t \to \infty,$$

ésta es una condición fundamental para la estimación del modelo.

En el modelo GARCH(1,1), la constante de Lipschitz de la Proposición 3.7 en [37] puede ser elegida como $C_2(x) = \beta_1$, y por lo tanto $\beta_1 < 1$ es suficiente para la invertibilidad del GARCH(1,1). Esta condición ya se tenía de la estacionariedad porque $\ln \beta_1 \leq \mathbb{E} \ln(\alpha_1 Z_0^2 + \beta_1) < 0$. Por lo tanto todo proceso GARCH(1,1) estacionario es invertible.

Podemos obtener

$$\sigma_{t,m}^2 = \alpha_0(1 + \beta_1 + \cdots + \beta_1^{m-1}) + \alpha_1 \sum_{k=1}^{m} \beta_1^{k-1} X_{t-k}^2 + \beta_1^m c_0^2.$$

Como $\beta_1 < 0$ y $\mathbb{E} \ln^+ X_0^2 < \infty$, la sucesión $\{\sigma_{t,m}^2\}_{m \geq 0}$ converge con probabilidad uno por el Lema 2.1 en [37]. Combinando este resultado con la relación (2.23), concluimos

$$\sigma_t^2 = \alpha_0(1 - \beta_1)^{-1} + \alpha_1 \sum_{k=1}^{\infty} \beta_1^{k-1} X_{t-k}^2, \quad \text{c.s.} \tag{2.24}$$

la cual es la representación de σ_t^2 en términos de las observaciones pasadas. Este resultado ya fue establecido en el capítulo anterior.

2.5.3. Cuasi Máxima Verosimilitud

Por el momento supongamos conocido el parámetro $\theta = (\alpha_0, \alpha_1, \beta_1)$ de un modelo GARCH(1,1) estacionario. El hecho que $\beta_1 < 1$, nos permite aproximar la función de varianza no observada, por medio de la recursión

$$\hat{\sigma}_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \hat{\sigma}_{t-1}^2, \ t \geq 1,$$

con valor inicial $\hat{\sigma}_0^2 \geq 0$. El siguiente cálculo nos permite conocer la tasa de convergencia de esta aproximación,

$$|\hat{\sigma}_t^2 - \sigma_t^2| = \beta_1 |\hat{\sigma}_{t-1}^2 - \sigma_{t-1}^2| = (\beta_1)^t |\hat{\sigma}_0^2 - \sigma_0^2|,$$

es decir, $|\hat{\sigma}_t^2 - \sigma_t^2| \to 0$, cuando $t \to \infty$.

Notemos que para cualquier $\theta = (u_1, u_2, u_3), \ u_3 \in [0, 1)$, podemos definir

$$\hat{h}_t(\theta) = u_1 + u_2 X_{t-1}^2 + u_3 \hat{h}_{t-1}(\theta),$$

y condición inicial $\hat{h}_0(\theta) = 0$. De este modo, la función de verosimilitud condicional de $(X_1, \ldots, X_n)$ dado (X_0, σ_0), bajo la hipótesis de que $\{Z_t\}$ es sucesión de vaiid con distribución $\mathrm{N}(0,1)$, es aproximadamente igual a

$$\hat{L}_n(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_t^2}{\hat{h}_t(\theta)} + \ln \hat{h}_t(\theta)\right).$$

Se define el *estimador cuasi máximo verosímil (ECMV)* de $\hat{\theta}_n$ como el valor de θ que maximiza $\hat{L}_n(\cdot)$. Como $u_3 \in [0,1)$, la sucesión de funciones aleatorias $\{\hat{h}_t\}_{t\geq 0}$ puede ser aproximada por un proceso ergódico estacionario $\{h_t\}_{t\geq 0}$ tal que el error $\|h_t - \hat{h}_t\|$ decae exponencialmente rápido casi seguramente. Resulta que $\{h_t\}$ está caracterizado como la única solución estacionaria de la SRE

$$h_t(\theta) = u_1 + u_2 X_{t-1}^2 + u_3 h_{t-1}(\theta),\ t \geq 1.$$

Berkes et al. [3] usan el decaimiento exponencial de $\|h_t - \hat{h}_t\|$ para probar que el máximo $\tilde{\theta}_n$ de

$$\hat{L}_n(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_t^2}{h_t(\theta)} + \ln h_t(\theta)\right),$$

es asintóticamente equivalente a $\hat{\theta}_n$. Por lo tanto, es suficiente estudiar $\tilde{\theta}_n$, el cual es un problema ligeramente más sencillo debido a que $\{X_t^2/h_t + \ln h_t\}$ es una sucesión estacionaria ergódica.

Consistencia

Supongamos que hemos observado $X_0, X_1, \ldots, X_n$ generado por el modelo (2.19) con θ_0 como parámetro verdadero.

Recordemos la función de verosimilitud condicional,

$$\hat{L}_n(\theta) = -\frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{\hat{h}_t(\theta)} + \ln \hat{h}_t(\theta)\right), \tag{2.25}$$

donde la sucesión $\hat{h}_t(\theta)$ definida por la ecuación de recurrencia $\hat{h}_t(\theta) = g_\theta(X_{t-1}, \hat{h}_{t-1}(\theta))$ junto con la condición inicial $\hat{h}_0(\theta) = c_0^2$ sirve como un estimador de la varianza bajo la hipótesis parametral θ.

El ECMV $\hat{\theta}$ maximiza $\hat{L}_n$ en K, donde $K \subset \Theta$ es un conjunto compacto elegido apropiadamente, i.e.,

$$\hat{\theta}_n = \mathrm{argmax}_{\theta\in K}\hat{L}_n(\theta). \tag{2.26}$$

Supongamos que las condiciones de la Proposición 3.14 en [37] se satisfacen, lo cual implica que $\|\hat{h}_t - h_t\|_K \overset{\text{e.c.s.}}{\to} 0$ cuando $t \to \infty$, donde h_t es la única solución estacionaria de la ecuación de

recurrencia $h_t = g_\theta(X_{t-1}, h_{t-1})$. Entonces siguiendo las ideas presentadas previamente, definimos

$$L_n(\theta) = -\frac{1}{2}\sum_{t=1}^{n}\left(\frac{X_t^2}{h_t(\theta)} + \ln h_t(\theta)\right) \tag{2.27}$$

junto con

$$\tilde{\theta} = \text{argmax}_{\theta\in K} L_n(\theta). \tag{2.28}$$

Es más conveniente trabajar con $\{L_n\}$ porque $\{X_t^2/h_t + \ln h_t\}$ es estacionaria ergódica mientras $\{X_t^2/\hat{h}_t + \ln \hat{h}_t\}$ no lo es. En seguida damos un conjunto de condiciones las cuales implicarán la consistencia fuerte de $\hat{\theta}_n$:

C1 Las condiciones del Teorema 3.1 se satisfacen para $g = g_\theta$ y $\{X_t\}$ es la única solución ergódica estacionaria para la ecuación de recurrencia 2.22 con $\theta = \theta_0$.

C2 $K \subset \Theta$ es un conjunto compacto con $\theta_0 \in K$ y tal que las condiciones de la Proposición 3.14 son observadas.

C3 La clase de funciones $\{g_\theta \mid \theta \subset K\}$ es uniformemente acotada por abajo, i.e., existe una constante λ tal que para todo $(x, s) \in \mathbb{R} \times [0, \infty)$ y para todo $\theta \in K$,

$$g_\theta(x, s) \geq \lambda.$$

C4 La siguiente condición de identificabilidad es cierta en K: para todo $\theta \in K$,

$$h_0(\theta) \equiv \sigma_0^2 \text{ c.s. si y sólo si } \theta = \theta_0.$$

C5 El elemento aleatorio σ_0^2/h_0 tiene esperanza en norma finita:

$$\mathbb{E}\left\|\frac{\sigma_0^2}{h_0}\right\|_K < \infty.$$

Teorema 2.2 (4.1 en [37]). *Bajo las condiciones* **C1-C4** *el ECMV* $\hat{\theta}_n$ *es consistente fuerte, i.e.*

$$\hat{\theta}_n \overset{\text{c.s.}}{\to} \theta_0, \ n \to \infty.$$

Para aplicar correctamente el teorema anterior, debemos exhibir el conjunto compacto K pedido en **C2**. Denotemos por $\theta = (u_1, u_2, u_3)$ y supongamos que $\theta_0 = (\alpha_0, \alpha_1, \beta_1)$ es el parámetro verdadero. Supongamos que θ_0 admite una única solución ergódica estacionaria $\{X_t\}$ a las ecuaciones de un GARCH(1,1), lo cual es equivalente a $\mathbb{E}\ln(\alpha_1 Z_0 + \beta_1) < 0$. Supongamos además que $\alpha_1 > 0$, porque

de otro modo $\sigma_t^2 = \alpha_0(1-\beta_1)^{-1}$ c.s., esto se tiene de la representación (2.24); $\alpha_1 = 0$ implicaría que uno no puede discriminar entre α_0 y β_1. Sea K un subconjunto compacto de $(0,\infty)\times[0,\infty)\times[0,1)$ conteniendo el parámetro verdadero θ_0. Verifiquemos las condiciones **C1**-**C4**.

La condición **C1** es satisfecha por la definición del modelo GARCH(1,1). Para **C2**, bastará verificar los supuestos de la Proposición 3.14. Sea $\bar{u}_i = \text{máx}_{\theta\in K}\, u_i$ y $\tilde{u}_i = \text{mín}_{\theta\in K}\, u_i,\ i = 1,2,3$. Como K es compacto, tenemos que $\bar{u}_i < \infty,\ i = 1,2$, y $u_3 < 1$ y $\tilde{u}_1 > 0$.

Recordemos que g_θ tiene la forma $g_\theta(x,s) = u_1 + u_2 x + u_3 s^2$, así que las funciones f_x definidas en la Proposición 3.14 son continuas. Más aún, $|g_\theta(x,s) - g_\theta(x,\tilde{s})| = u_3|s-\tilde{s}| \geq \bar{u}_3|s-\tilde{s}|$, para todo $x\in\mathbb{R},\ s,\tilde{s}\geq 0$ y para todo $\theta\in K$. Por lo tanto la constante de Lipschitz puede ser elegida como $C_2(x) = \bar{u}_3 < 1$. Finalmente, como $\|g_\bullet(X_0, c_0^2)\|_K < \infty$ para $\tilde{q} > 0$ suficientemente pequeña. La condición **C3** se satisface con $\lambda = \tilde{u}_1 > 0$. La condición **C4** se satisface bajo el supuesto que la distribución de Z_0 no se concentra en dos puntos. La prueba de esta condición se sigue del Lema 5.3 en [37]. Por lo tanto por una aplicación del Teorema 2.2, el ECMV en un GARCH(1,1) es consistente fuerte.

2.6. Hechos estilizados y GARCH(1,1)

2.6.1. Análisis exploratorio

A continuación presentamos algunas rutinas implementadas en MATLAB que nos permiten realizar un análisis primario a la serie de rendimientos de un activo financiero dado, así como estimar un modelo GARCH(1,1). Para ejemplificar utilizaremos de nueva cuenta la serie JPY-USD.[1]

El primer paso será *cargar* los datos a trabajar, en nuestro caso tecleamos

```
datos = load('C:\MATLAB7\toolbox\garch\garchdemos\jpy-usd.mat');
serie = datos.jpyusd(1514:8857);.
```

Generar una gráfica de `serie` resulta conveniente, hacemos esto con

```
plot([1514:8857],jpyusd)
title('TASA DE CAMBIO YEN JAPONES - DOLAR ESTADOUNIDENSE')
set(gca,'XTick',[1514 8857])
set(gca,'XTickLabel',{'Feb 1971' 'May 2006'})
```

[1]Una detallada descripción de los comando usados aquí está disponible en el Toolbox Garch de MATLAB.

este código generó la Figura 1.1.

Implementar un modelo GARCH asume la transformación rendimientos, tal función ya está implementada en MATLAB:

```
rendimientos = price2ret(jpyusd);.
```

La presencia de heteroscedaticidad, mostrada en el análisis previo, indica que un modelo GARCH es apropiado. Usamos la función `garchfit` para estimar los parámetros del modelos. MATLAB tiene como modelo predeterminado el GARCH(1,1). Sólo requerimos especificar la serie de rendimientos de interés como argumento de la función `garchfit`.

```
[coeff,errors,LLF,innovations,sigmas,summary] = garchfit(rendimientos);
```

Ahora que hemos completado la estimación, podemos desplegar los parámetros estimados y sus errores estándar usando la función garchdisp

```
garchdisp(coeff,errors)
```

Obtenemos la siguiente salida

```
 Mean: ARMAX(0,0,0); Variance: GARCH(1,1)

  Conditional Probability Distribution: Gaussian
  Number of Model Parameters Estimated: 4

                                Standard          T
  Parameter        Value          Error       Statistic
 -----------    -----------   ------------   -----------
           C     -0.00754       0.0072211       -1.0442
           K     0.012932       0.0011786       10.9731
    GARCH(1)     0.91188        0.0050023      182.2909
     ARCH(1)     0.060321       0.0031949       18.8805
```

Sustituyendo estos parámetros en la representación dada por Bollerslev, el proceso de estimación

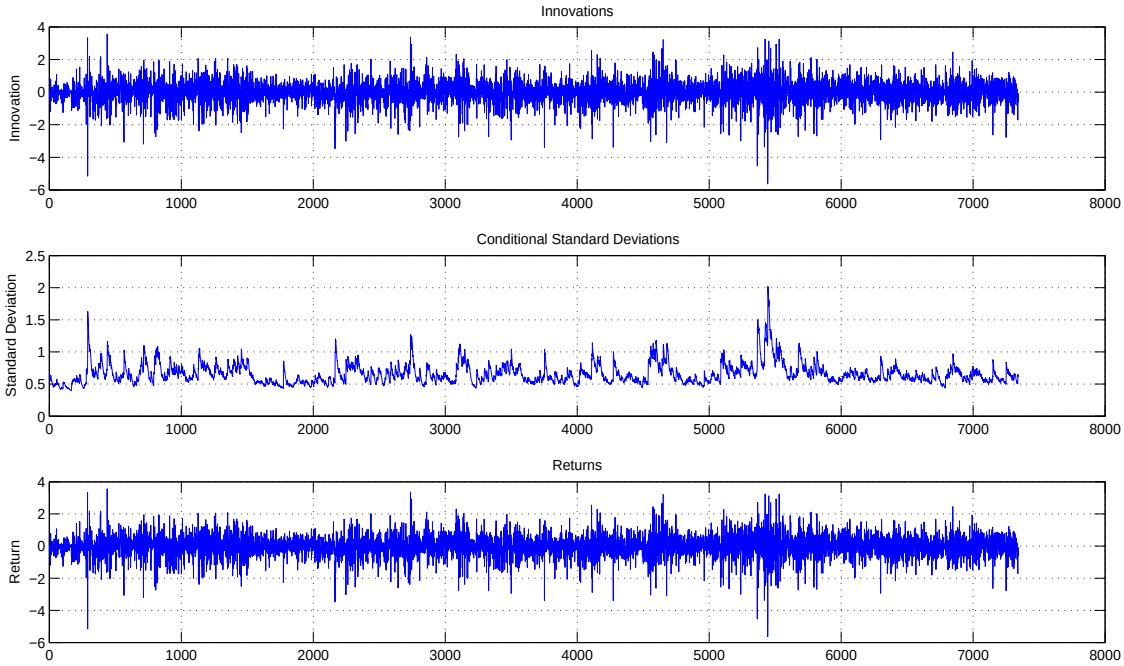
implica que el modelo de varianza condicional que mejor ajusta a los datos observados es

$$Y_t = -0.00754 + \epsilon_t,$$
$$\sigma_t^2 = 0.012932 + 0.060321\epsilon_{t-1}^2 + 0.91188\sigma_{t-1}^2.$$

Además de los parámetros estimados y errores estándar, `garchfit` también regresa el valor de la log-verosimilitud típica optimizada (`LLF`), los residuos (`innovations`) y las desviaciones estándar condicionales (`sigmas`). Usamos la función `garchplot` para observar la relación entre los residuos derivados del modelo ajustado, las desviaciones estándar correspondientes y los rendimientos observados.

```
garchplot(innovations,sigmas,dem2gbp),
```

produciendo la siguiente gráfica



2.6.2. Validez del supuesto GARCH

Hasta este punto, hemos presentado las características compartidas por las series de log returns de algunas series de índole financiera. También hemos dado un breve repaso a los modelos más

usados para repreentar de la mejor manera los hechos estilizados empíricos hallados. El ejercicio que realizamos en esta sección final aplica las herramientas visuales descritas en el capítulo introductorio a un proceso GARCH(1,1). Los parámetros de este modelo fueron ajustados en la sección anterior. La Figura 2.1 estudia si es sensato el uso de éstos modelos para capturar los cúmulos de variabilidad, las colas pesadas y la dependencia a largo plazo exhibidas por las series de log returns.

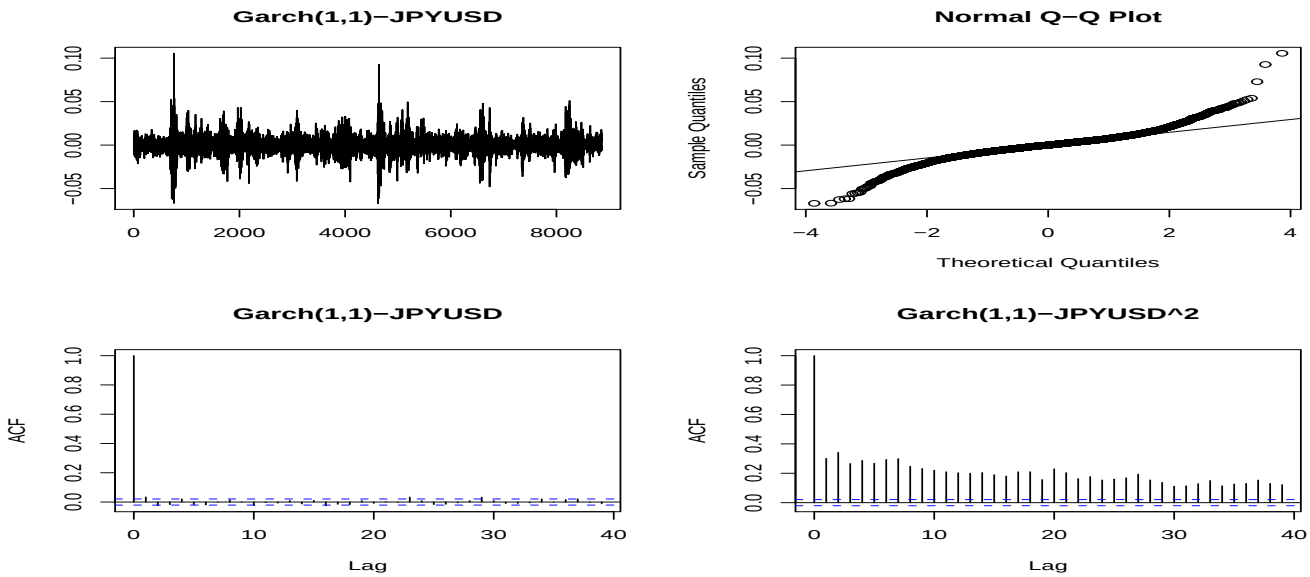


Figura 2.1: De arriba a abajo y de izquierda a derecha: Gráfica de un proceso GARCH(1,1) simulado con $\alpha_0 = 0.0129$, $\alpha_1 = 0.9118$ y $\beta_1 = 0.0603$, gráficamente podemos argumentar existencia de cúmulos de variabilidad; qq-plot normal de los datos simulados, se sostiene el argumento de la cola pesada; FAC de datos simulados, valores muy pequeños; FAC de datos simulados al cuadrado, existe diferencia significativa de cero, sin embargo es notable el hecho de que esta FAC decrece más lentamente que la FAC calculada a la serie de log returns al cuadrado.

A partir de estas gráficas, identificamos que las características descritas por el modelo GARCH(1,1) pueden mejorarse; colas pesadas, lento decaimiento de la FAC muestral en el caso $\alpha_1 + \beta_1 \approx 1$. Dentro de los muchos descubrimientos que contribuyen a incrementar el número de resultados que enfatizan las diferencias entre el comportamiento de herramientas estadísticas a la luz de colas pesadas y dependencia, nosotros agregamos uno más. ¿Los cúmulos de excedencias de umbrales altos o bajos son correctamente capturados por el modelo GARCH(1,1)?, esta duda será resuelta en los siguientes capítulos.

Capítulo 3

Teoría de Valores Extremos

Los capítulos previos han servido para describir algunos problemas relacionados con las series de tiempo financieras, así como para identificar una herramienta de modelación muy útil: Los procesos GARCH. En particular, este trabajo estudia los *cúmulos de variabilidad* presente en las series de log returns de ciertas series financieras, e.g. tasas de cambio e índices de mercado. Haremos esto auxiliados de la Teoría de Valores Extremos, concretamente del llamado *Índice Extremo*, parámetro clave para extender la Teoría Clásica a sucesiones estacionarias con dependencias débiles. A lo largo de este capítulo nos avocaremos en presentar a detalle la justificación teórica de éste parámetro, un método para estimarlo a partir de las observaciones hechas a un proceso estacionario. Más aún, proveemos una explicación de la importancia del valor estimado del índice extremo como herramienta para realizar inferencia sobre el tamaño de los clusters. Este capítulo estará basado en los textos de Leadbetter et. al. [23], Resnick [34] y en el artículo de Leadbetter [22].

3.1. Introducción

El siguiente teorema es bien conocido en el campo del estudio de eventos extremos, de hecho podemos considerarlo el pilar de la Teoría Clásica de Valores Extremos: Sean $\{X_i\}_{i=1}^n$ variables aleatorias independientes e idénticamente distribuidas (*v.a.i.i.d.*) con función de distribución (*Fd*) común F y sea

$$M_n = \max_{1\leq i\leq n}\{X_i\}.$$

Teorema 3.1 (Familia de Distribuciones de Valor Extremo (DVE)). *Supongamos que existen cons-*

tantes $a_n \in \mathbb{R}^+$, $b_n \in \mathbb{R}$, *tales que*

$$\lim_{n\to\infty} \Pr(a_n(M_n - b_n) \leq x) \xrightarrow{D} G(x), \tag{3.1}$$

donde G *es propia y no está concentrada en un punto. Entonces* G *pertenece a una de las siguientes tres familias de distribución:*

1. Gumbel (Tipo I): $\Lambda(x) = e^{-e^{-x}}\mathbf{1}_{(-\infty,\infty)}(x)$.

2. Fréchet (Tipo II): $\Phi_\alpha(x) = e^{-x^{-\alpha}}\mathbf{1}_{[0,\infty)}(x),\ \alpha > 0$.

3. Weibull (Tipo III): $\Psi_\alpha(x) = e^{-(-x)^{\alpha}}\mathbf{1}_{(-\infty,0]}(x) + \mathbf{1}_{[0,\infty)}(x),\ \alpha > 0$.

Nótese que este teorema *asume* que $(a_nM_n - b_n)$ tiene una Fd límite no degenerada G, y luego se prueba que G deberá tener una de las tres formas descritas. Se pueden construir sucesiones $\{X_n\}$ i.i.d. para los que tal G *no* existe. Por ejemplo, introduciendo la siguiente notación

$$x_F = \sup\{x \mid F(x) < 1\},\ (\leq \infty).$$

esto es, $F(x) < 1$ para todo $x < x_F$ y $F(x) = 1$ para todo $x \geq x_F$, y suponiendo que cada X_n tiene una Fd F tal que $x_F < \infty$ y que F tiene un salto en x_F, i.e. $F(x_{F^-}) < 1 = F(x)$. Entonces puede probarse que si $\{u_n\}$ es cualquier sucesión y $\Pr(M_n \leq u_n) \to \rho$, cuando $n \to \infty$, entonces $\rho \in \{0, 1\}$. De modo que si $\Pr(a_nM_n - b_n) \to G$, se sigue, tomando $u_n = x/a_n + b_n$, que $G(x) \in \{0, 1\}$, esto es, G es degenerada. Véase [34].

En (3.1), la convergencia debe darse para toda $x \in \mathcal{C}$, el conjunto de puntos de continuidad. Por otra parte, es interesante considerar sucesiones $\{u_n\}$ las cuales puedan no depender de algún parámetro x o puedan ser funciones más complicadas que las combinaciones lineales consideradas hasta ahora. Incluimos el siguiente teorema debido a que proporciona condiciones necesarias y suficientes para garantizar (3.1), además será objeto de referencia en secciones posteriores.

Teorema 3.2 (Aproximación de Poisson)**.** *Dado* $\tau \in [0, \infty]$ *y una sucesión* $\{u_n\} \subset \mathbb{R}$, *las siguientes dos relaciones son equivalentes*

$$n(1 - F(u_n)) \overset{n\to\infty}{\to} \tau, \tag{3.2}$$

$$\Pr(M_n \leq x) = F^n(u_n) \overset{n\to\infty}{\to} e^{-\tau} \tag{3.3}$$

Demostración. Para $0 \leq \tau$. Si (3.2) vale, entonces

$$\begin{aligned}\Pr(M_n \leq u_n) &= F^n(u_n) = (1 - \bar{F}(u_n))^n \\ &= \left(1 - \frac{\tau}{n} + \sigma(\frac{1}{n})\right)^n \\ &\to e^{-\tau}.\end{aligned}$$

Recíprocamente, si (3.3) vale, entonces $\bar{F}(u_n) \to 0$. En efecto, supongamos que para alguna subsucesión $\{u_{n_k}\} \subset \Re$, $\bar{F}(u_{n_k})$ no converge a cero, esto implica que $\bar{F}(u_{n_k})$ está acotado en un punto lejos del cero y de este modo $\Pr(M_n \leq u_n) = (1 - \bar{F}(u_{n_k}))^n \to 0$, cuando $n_k \to \infty$, esto es una contradicción, por lo tanto $\bar{F}(u_{n_k}) \to 0$.

Tomando logaritmos en (3.3), tenemos:

$$\begin{aligned}\ln F^n(u_n) &\to -\tau \\ n \ln F(u_n) &\to -\tau \\ n \ln(1 - \bar{F}(u_n)) &\to -\tau \\ -n\bar{F}(u_n) &\to -\tau,\end{aligned}$$

en el tercer paso se ha usado que $-\ln(1 - z) \sim z$, cuando $z \to 0$.

Finalmente, para $\tau = \infty$. Si (3.2) vale, pero (3.3) no, entonces debe existir $\{n_k\} \subset \Re$ tal que $\Pr(M_{n_k} \leq u_{n_k}) \to e^{-\tau'}$, cuando $n_k \to \infty$, para algún $\tau' < \infty$. Pero lo demostrado en el caso finito, implica que $n_k \, \bar{F}^n(u_{n_k}) \to e^{-\tau'}$, lo cual es una contradicción. ■

Nótese que si las v.a.'s ya no se suponen independientes, entonces no hemos dado condiciones que garanticen que

$$\Pr(a_n(M_n - b_n) \leq x) \overset{n\to\infty}{\to} G(x).$$

Si bien aún no establecemos una motivación para estudiar esta teoría probabilística, debe quedar claro que los procesos generadores de las series de tiempo consideradas hasta ahora poseen alguna estructura de dependencia.

Existen varias maneras en las que la noción de una sucesión i.i.d. puede generalizarse para permitir dependencia, o permitir que las X_n's tengan diferentes distribuciones, o ambas. Continuemos suponiendo que las X_n's tienen distribución común. En el capítulo anterior estudiamos las condiciones necesarias y suficientes para que un proceso GARCH(1,1) sea estacionario, siguiendo esta línea consideraremos sucesiones *estacionarias*.

Asumamos que la dependencia entre X_i y X_j decae de alguna manera específica cuando $|i-j|$ crece. El ejemplo más simple de este tipo de restricción es la *m-dependencia*, la cual requiere que X_i y X_j sean independientes si $|i-j| > m$. Una restricción más usada para sucesiones estacionarias es la llamada *mezcla fuerte*.

Definición 3.1 (Mezcla Fuerte)**.** Una sucesión $\{X_n\}$ se dice que satisface el supuesto de *mezcla fuerte* si hay una función $g(\cdot)$, la *función de mezcla*, tal que

$$|\Pr(A \cap B) - \Pr(A)\Pr(B)| < g(k),\ g(k) \to 0, \text{cuando } k \to \infty, \tag{3.4}$$

donde, $A \in \mathcal{F}(X_1, \ldots, X_p)$ y $B \in \mathcal{F}(X_{p+k+1}, X_{p+k+2}, \ldots)$ para cualquier p y k; $\mathcal{F}(\cdot)$ denotará la σ-álgebra generada por las v.a.'s indicadas.

Esto se interpreta de la siguiente manera: bajo la condición de mezcla fuerte, cualquier evento A *basado en el pasado hasta el tiempo p será 'casi independiente'* de cualquier evento B *basado en el futuro a partir del tiempo* $p+k+1$, todo esto, cuando k es grande.

La *correlación* entre X_i y X_j también puede proporcionar una medida de dependencia entre estas variables. Por lo tanto otra restricción de dependencia del mismo tipo es $|\mathrm{Cor}(X_i, X_j)| \leq g(|i-j|)$, donde $g(k) \to 0$, cuando $k \to \infty$. Esta restricción resulta muy útil si las X_n's forman una sucesión normal.

Varios resultados de la teoría de valores extremos han sido extendidos aplicando algunas de las restricciones mencionadas arriba. Loynes [24] prueba una cantidad considerable de resultados, inclusive la versión para sucesiones estacionarias del Teorema 3.2, bajo el supuesto de mezcla fuerte. La condición de dependencia establecida por tal autor si bien es útil en muchos contextos, resulta muy restrictiva. A continuación incluiremos una condición más relajada que nos permite establecer resultados similares a los antes mencionados, así como el homólogo al Teorema 3.1 para sucesiones estacionarias.

Para debilitar la condición de mezcla fuerte, debemos notar que los eventos de interés en teoría de valores extremos son típicamente aquellos de la forma $\{X_j \leq u\}$ o sus intersecciones. Por ejemplo, el evento $\{M_n \leq u\} = \bigcap_{i=1}^{n}\{X_i \leq u\}$. Por lo tanto, uno debe proponer una condición para ser satisfecha por eventos de este tipo.

Notación. Escribiremos $F_{i_1,\ldots,i_n}(u)$ para denotar $F_{i_1,\ldots,i_n}(u, \ldots, u)$ si $F_{i_1,\ldots,i_n}(x_1, \ldots, x_n)$ denota la Fd conjunta de $X_{i_1}, \ldots, X_{i_n}$.

Definición 3.2. Diremos que se satisface la condición de mezcla D si para cualesquiera enteros $i_1 < \cdots < i_p$ y $j_1 < \cdots < j_{p'}$ para los cuales $|j_1 - i_p| \geq l$, y cualquier real u,

$$|F_{i_1,\ldots,i_p,j_1,\ldots,j_{p'}}(u) - F_{i_1,\ldots,i_p}(u)F_{j_1,\ldots,j_{p'}}(u)| \leq g(l),$$

donde $g(l) \to 0$ cuando $l \to \infty$.

Aunque esta condición es una reducción significativa de los requerimientos impuestos por una mezcla fuerte, aún podemos mejorarla y seguir obteniendo resultados tipo el Teorema de DVE. Consideraremos una condición, la cual incluirá un requerimiento como el anterior pero aplicado únicamente a una cierta clase de sucesiones $\{u_n\}$, ya no necesariamente a toda u. Tal condición está dada por la siguiente

Definición 3.3 (Condición de Leadbetter)**.** Decimos que se satisface la Condición $D(u_n)$ para la sucesión $\{u_n\} \subset \mathbb{R}$ si para cualesquiera enteros

$$1 \leq i_1 < \cdots < i_q < j_1 < j_2 < \cdots < j_p \leq n,$$

donde $j_1 - i_q > l$, tenemos

$$|F_{i_1,\ldots,i_q,j_1,\ldots,j_p}(u_n) - F_{i_1,\ldots,i_q}(u_n)F_{j_1,\ldots,j_p}(u_n)| \leq \alpha_{n,l_n}, \tag{3.5}$$

donde $\alpha_{n,l_n} \to 0$, cuando $n \to \infty$ para alguna sucesión $l_n = o(n)$.

Un proceso que satisface la Condición $D(u_n)$ será *asintóticamente independiente a pedazos.* Obsérvese que los enteros $1 \leq i_1 < \cdots < i_q < j_1 < j_2 < \cdots < j_p \leq n$ dividen el proceso $X_1, \ldots, X_n$ en *pedazos* y el hecho que (3.5) se cumpla para estos enteros garantiza, de acuerdo a la noción de mezcla fuerte, que los eventos $\{X_{i_1} \leq u_n, \ldots, X_{i_q} \leq u_n\}$ y $\{X_{j_1} \leq u_n, \ldots, X_{j_p} \leq u_n\}$ son *aproximadamente independientes* debido a que la función α_{n,l_n} decae a cero cuando $n \to 0$.

Pueden hacerse modificaciones convenientes a la Condición $D(u_n)$. Por el Lema 3.2.1 en [23], α_{n,l_n} puede escogerse tal que sea no creciente en l_n para cada n, además para tal α_{n,l_n}, la condición $\alpha_{n,l_n} \to 0$, cuando $n \to \infty$ puede reescribirse como $\alpha_{n,[n\lambda]} \to 0$, para cada $\lambda > 0$. Aquí $[\cdot]$ denota la función parte entera.

La dificultad de establecer si (3.5) se satisface para un proceso dado estriba en la necesidad de conocer la distribución conjunta y exhibir una función α_{n,l_n} con las propiedades mencionadas arriba. Esta es una limitante de este enfoque, a continuación argumentaremos las ventajas del mismo.

3.2. Resultados bajo la Condición $D(u_n)$

La idea básica para poder extender la teoría clásica de valores extremos al caso estacionario es mostrar que

$$\Pr(M_n \leq u_n) - \Pr^k(M_{r_n} \leq u_n) \overset{n\to\infty}{\to} 0, \tag{3.6}$$

para cada $k = 1, 2, \ldots$ cuando la Condición $D(u_n)$ $D(u_n)$ se satisface y $r_n = [n/k]$.

Esta es una idea brillante. Tenemos la teoría suficiente para estudiar la distribución del máximo de todo el proceso, llamado por nosotros el 'máximo global'. Usando la propiedad de estacionariedad del proceso, podemos dividir el mismo en pedazos de igual longitud y estudiar la distribución del máximo en estos *bloques*, llamamos a estos 'máximos locales'.

Notación. Si E es cualquier conjunto de enteros, $M(E)$ denotará el $\text{máx}\{X_j \mid j \in E\}$, nótese que $M(E) = M_n$, si $E = \{1, \ldots, n\}$. En este contexto, un *intervalo* será cualquier conjunto E de enteros consecutivos digamos $\{j_1, \ldots, j_2\}$; su *longitud* será $j_2 - j_1 + 1$ y E, F están *separados* por $k_1 - j_2$, si $F = \{k_1, \ldots, k_2\}$ es otro intervalo con $k_1 > j_2$.

Lema 3.1. *Supongamos que la Condición $D(u_n)$ $D(u_n)$ se satisface para alguna sucesión $\{u_n\}$. Sean n, r y k enteros fijos y $E_1, \ldots, E_r$ subintervalos de $\{1, \ldots, n\}$ tales que E_i y E_j están separados por al menos k cuando $i \neq j$. Entonces*

$$\left| \Pr\left(\bigcap_{j=1}^{r} \{M(E_j) \leq u_n\} \right) - \prod_{j=1}^{r} \Pr(M(E_j) \leq u_n) \right| \leq (r-1)\alpha_{n,k}.$$

Demostración. ■

Este lema muestra un grado de independencia en los máximos sobre intervalos separados, lo cual será la base para establecer la prueba del siguiente

Teorema 3.3 (Lema 2.1 en [22]). *Sea u_n una sucesión de constantes y supongamos que la Condición $D(u_n)$ $D(u_n)$ es satisfecha por la sucesión estacionaria X_n. Sea k_n una sucesión de constantes tal que $k_n = o(n)$, $k_n l_n = o(n)$, $k_n \alpha_{n,l_n} \to 0$. Entonces*

$$\Pr(M_n \leq u_n) - \Pr^{k_n}(M_{r_n} \leq u_n) \overset{n\to\infty}{\to} 0,$$

donde $r_n = [n/k_n]$.

Desmostración. Sea l_n como en la definición de $D(u_n)$. Dividamos los enteros $1, \ldots, n$ en intervalos $I_1, I_1^*, \ldots, I_{k_n-1}, I_{k_n-1}^*$ de la siguiente manera

$$I_j = \{(j-1)r_n + 1, \ldots jr_n - l_n\}, \quad I_j^* = \{jr_n - l_n + 1, \ldots, jr_n\}, \quad j = 1, \ldots, r_n - 1,$$

y

$$I_{k_n} = \{k_n l_n - 2l_n + 1, \ldots, k_n r_n - l_n\}, \quad I^*_{k_n} = \{k_n r_n - l_n + 1, \ldots, n\}.$$

Los intervalos I_j contienen $r_n - l_n$ enteros, los intervalos I^*_j y el intervalo I_{k_n} contienen l_n enteros, el intervalo $I^*_{k_n}$ contiene $n - k_n r_n + l_n \leq k_n + l_n$.

Notemos que

$$\{M_n \leq u_n\} \subset \bigcap_{j=1}^{k_n} \{M(I_j) \leq u_n\}, \tag{3.7}$$

de donde se sigue que

$$0 \leq \Pr\left(\bigcap_{j=1}^{k_n} \{M(I_j) \leq u_n\}\right) - \Pr(M_n \leq u_n)$$

Utilizando (3.7) podemos construir la diferencia de estos conjuntos,

$$\{M_n > u_n\} \cap \bigcap_{j=1}^{k_n} \{M(I_j) \leq u_n\}.$$

De esta relación tenemos los siguientes puntos

- $M_n > u_n$ y $M(I_i) \leq u_n$ para $i = 1, \ldots, k_n - 1$. Esto implica que

$$M(I_j) \leq u_n < M(I^*_j), \text{ para algún } j \in \{1, \ldots, k_n\}.$$

- Si $X_j \leq u_n$ para todo $1 \leq j \leq k_n r_n$, pero $X_j > u_n$ para algún $j \in \{k_n r_n + 1, \ldots, k_n(r_n + 1)\}$, esto porque $l_n > k_n$ y por lo tanto $k_n(r_n + 1) \leq k_n r_n + l_n$.

Los puntos anteriores y la estacionariedad muestran que

$$\begin{aligned} 0 &\leq \Pr\left(\bigcap_{j=1}^{k_n} M(I_j) \leq u_n\right) - \Pr(M_n \leq u_n) \\ &\leq (k_n + 1)\Pr(M(I_1) \leq u_n < M(I^*_1)). \end{aligned} \tag{3.8}$$

Análogamente podemos establecer que

$$\begin{aligned} 0 &\leq \Pr(M(I_1) \leq u_n) - \Pr(M_{r_n} \leq u_n) \\ &= \Pr(M(I_1) \leq u_n < M(I^*_1)). \end{aligned} \tag{3.9}$$

Escribiendo, $y = \Pr(M(I_1) \leq u_n)$, $x = \Pr(M_{r_n} \leq u_n)$ y recordando que

$$y^t - x^t \leq t(y - x), \text{ cuando } x, y \in (0, 1),$$

tenemos que

$$\Pr^{k_n}(M(I_1) \leq u_n) - \Pr^{k_n}(M_{r_n} \leq u_n) \leq k_n(\Pr(M(I_1) \leq u_n < M(I_1^*))). \tag{3.10}$$

Por el Lema 3.1, haciendo $I_j = E_j$ y notando que $\Pr(M(I_j) \leq u_n)$ es independiente de j, se tiene que

$$\left|\Pr\left(\bigcap_{j=1}^{k_n}\{M(I_j) \leq u_n\}\right) - \Pr^{k_n}(M(I_1) \leq u_n)\right| \leq (k_n - 1)\alpha_{n,l_n} \tag{3.11}$$

Por (3.8), (3.10) y (3.11),

$$\begin{aligned}
&\left|\Pr(M_n \leq u_n) - \Pr^{k_n}(M_{r_n} \leq u_n)\right| \\
&\qquad \leq \left|\Pr\left(\bigcap_{j=1}^{k_n}\{M(I_j) \leq u_n\}\right) - \Pr(M_n \leq u_n)\right| \\
&\qquad + \left|\Pr^{k_n}(M(I_1) \leq u_n) - \Pr\left(\bigcap_{j=1}^{k_n}\{M(I_j) \leq u_n\}\right)\right| \\
&\qquad + \left|\Pr^{k_n}(M_{r_n} \leq u_n) - \Pr^{k_n}(M(I_1) \leq u_n)\right| \\
&\qquad \leq (2k_n + 1)\Pr(M(I_1) \leq u_n < M(I_1^*)) + (k_n - 1)\alpha_{n,l_n}.
\end{aligned}$$

Por el Lema 3.3.2 en [23] se sabe que $\Pr(M(I_1) \leq u_n < M(I_1^*)) \to 0$ cuando $n \to$, esto y la definición de α_{n,l_n} garantizan el resultado. ■

Supongamos de ahora en adelante que para cada $\tau > 0$ podemos definir una sucesión $\{u_n(\tau)\}$ tal que

$$n[1 - F(u_n(\tau))] \overset{n\to\infty}{\to} \tau. \tag{3.12}$$

Esto impone una restricción sobre la marginal de F de las X_n, pero ésta, más que la excepción es la constante en las aplicaciones. Podemos definir $u_n(\tau)$ para garantizar igualdad en (3.12), a saber

$$u_n(\tau) = F^{\leftarrow}\left(1 - \frac{\tau}{n}\right).$$

En [23] se establece que es necesario y suficiente para satisfacer (3.12) que

$$\frac{1 - F(x^-)}{1 - F(x)} \overset{x\to\infty}{\to} 1, \tag{3.13}$$

esto es, la cola de la distribución F debe ser 'regular'.

Comentario 3.1. Si existe $u_n(\tau')$ que satisface (3.12) para un $\tau' > 0$ fijo, entonces existe un $u_n(\tau) > 0$ que satisface (3.12) para toda τ. Por ejemplo, si hacemos $u_n(1) = F^{\leftarrow}(1 - 1/n)$ para un τ dado basta definir $u_n(\tau) = u_{[n/\tau]}(1)$.

Teorema 3.4. *Sea $\{X_n\}$ una sucesión estacionarias y $u_n(\tau)$ constantes que satisface* (3.12) *y tal que $D(u_n(\tau_0))$ se satisface para algún $\tau_0 > 0$. Entonces existen constantes $\theta \leq \theta'$ tales que*

$$\limsup_{n\to\infty} \Pr(M_n \leq u_n(\tau)) = e^{-\theta\tau}$$
$$\liminf_{n\to\infty} \Pr(M_n \leq u_n(\tau)) = e^{-\theta'\tau}.$$

Por lo tanto si $\Pr(M_n \leq u_n(\tau))$ converge para algún τ, entonces $\theta = \theta'$ y $\Pr(M_n \leq u_n(\tau)) \to e^{-\theta\tau}$ para todo τ.

Demostración. Sabemos que para un entero k, fijo,

$$\Pr(M_n \leq u_n(\tau)) - \Pr^k(M_{n'} \leq u_n(\tau)) \to 0, \text{ cuando } n \to \infty,$$

donde $n' = [n/k]$. Por lo tanto si $\limsup_{n\to\infty} \Pr(M_n \leq u_n(\tau)) = \psi(\tau)$, se sigue que

$$\limsup_{n\to\infty} \Pr(M_n \leq u_n(\tau)) = \psi^{1/k}(\tau). \tag{3.14}$$

Consideremos el caso $u_n(\tau) \geq u_n(\tau/k)$, $u_n(\tau) < u_n(\tau/k)$ que

$$\begin{aligned}
|\Pr(M_{n'} \leq u_n(\tau)) - \Pr(M_{n'} \leq u_{n'}(\tau/k))| &\leq n' \left| F(u_n(\tau)) - F\left(u_{n'}\left(\frac{\tau}{k}\right)\right)\right| \\
&= n' \left| \left(1 - F\left(u_{n'}\left(\frac{\tau}{k}\right)\right)\right) - (1 - F(u_n(\tau)))\right| \\
&= n' \left| \frac{\tau/k}{n'}(1 + o(1)) - \frac{\tau}{k}(1 + o(1))\right| \\
&= o(1),
\end{aligned}$$

porque $n' \sim n/k$. Como $\limsup_{n\to\infty} \Pr(M_{n'} \leq u_{n'}(\tau/k)) = \psi(\tau/k)$ se sigue que $\limsup_{n\to\infty} \Pr(M_{n'} \leq u_{n'}(\tau)) = \psi(\tau/k)$, que junto con (3.14) muestra que $\psi(\tau/k) = \psi^{1/k}(\tau)$, para todo $\tau > 0,\ k = 2, 3, \ldots$. Ahora, si $\tau' < \tau$ es claro que $u_n(\tau') > u_n(\tau)$ cuando n es suficientemente grande, así que $\psi(\tau/k)$ es no creciente y estrictamente positiva debido a que $\liminf_{n\to\infty} \Pr(M_n \leq u_n(\tau)) \geq e^{-\tau}$. Pero es bien conocido que la única solución a esta ecuación funcional es $\psi(\tau) = e^{-\theta\tau}$ donde $\theta > 0$.

Similarmente, se obtiene que $\liminf_{n\to\infty} \Pr(M_n \leq u_n(\tau)) \to e^{-\theta'\tau}$ donde $0 \leq \theta' \leq 1$, y claramente $\theta' > \theta$, lo que completa la prueba. ∎

Para simplificar enunciados en resultados posteriores, conviene tener una terminología para el anterior teorema.

Definición 3.4 (Índice Extremo)**.** Si $\Pr(M_n \leq u_n(\tau)) \to e^{-\theta\tau}$ para cada $\tau > 0$ con $u_n(\tau) > 0$ satisfaciendo (3.12), diremos que X_n tiene Índice Extremo (IE) θ.

La importancia del índice extremo está en que es el parámetro que permite extender los resultados clásicos en teoría de valores extremos del caso independiente a un proceso estacionario. El IE pertenece al intervalo $[0,1]$. Por definición $0 \leq e^{-\theta\tau} \leq 1$, o equivalentemente, $0 \leq \theta\tau \leq \infty$ lo que implica que $\theta \geq 0$. Por otro lado,

$$\begin{aligned}\Pr(M_n \leq u_n(\tau)) &= 1 - \Pr(M_n \leq u_n(\tau)) = 1 - \Pr(\cup_{i=1}^{n} \{X_i > u_n(\tau)\}) \\ &\geq 1 - n\left[1 - F(u_n(\tau))\right],\end{aligned}$$

y por definición de IE,

$$e^{-\theta\tau} \geq 1 - \tau, \forall \tau > 0,$$

y esto último implica que $\theta \leq 1$.

Usando el teorema y definición recientes tenemos el siguiente

Corolario 3.1. *Sea $\{X_n\}$ estacionaria que satisface $D(u_n(\tau))$ para cada $\tau > 0$, definido por* (3.12). *Si para algún $\tau_0 > 0$ $\Pr(M_n \leq u_n(\tau_0))$ converge a un límite α, entonces $\{X_n\}$ tiene IE $\theta = -\tau_0^{-1}\ln(\alpha)$. Así $\Pr(M_n \leq u_n) \to e^{-\theta\tau}$ para todo $\tau > 0$.*

3.2.1. Criterios para existencia de IE

En la teoría clásica de valor extremo, condiciones como la dada por la ecuación (3.12) son necesarias y suficientes para garantizar convergencia en distribución del máximo de una sucesión de variables aleatorias. A continuación damos algunas importantes relaciones que permitirán determinar cuándo un proceso estacionario tiene índice extremo.

El primer resultado tiene, en principio, un interés más teórico que práctico pero sirve como un medio para extender la condición $D(u_n)$ a casos donde $\theta < 1$.

Notación. Escribimos $n' = [n/k]$ para k fijo, $n = 1, 2, \ldots$.

Teorema 3.5 (Teorema 3.1 en [22]). *Sea $\{X_n\}$ una sucesión estacionaria que satisface la Condición $D(u_n)$ $D(u_n(\tau))$ para cada $\tau > 0$ y $u_n(\tau)$ satisface* (3.12). *Entonces X_n tiene índice extremo θ si y sólo si*

$$k \limsup_{n\to\infty} \left|1 - F_{1,\ldots,n'}(u_n) - \theta\frac{\tau_0}{k}\right| \overset{k\to\infty}{\to} 0, \tag{3.15}$$

para algún $\tau_0 > 0$. Equivalentemente esto se satisface si y sólo si

$$1 - F_{1,\ldots,n'}(u_n) \overset{n\to\infty}{\to} \theta\frac{\tau_0}{k} + \lambda_k, \tag{3.16}$$

donde $k\lambda_k \to 0$, cuando $k \to \infty$.

Demostración. Como X_n tiene IE y (3.6) se cumple, entonces

$$F_{1,\ldots,n'}(u_n) = \Pr(M_{n^1} \leq u_n) \approx \Pr^{1/k}(M_n \leq u_n) \to e^{-\theta\frac{\tau}{k}},$$

Usando la expansión de Taylor, tenemos que

$$F_{1,\ldots,n'}(u_n) = 1 - \theta\frac{\tau}{k} + o(\tau/k), \quad n \to \infty$$

De donde se sigue (3.16).

Recíprocamente, si (3.15) se satisface entonces

$$\begin{aligned}\limsup_{n\to\infty} \Pr(M_n \leq u_n) &= \limsup_{n\to\infty}[F_{1,\ldots,n'}(u_n) - 1 + \theta\tau/k] + 1 - \theta\tau/k \\ &\leq 1 - \theta\frac{\tau}{k} + \limsup_{n\to\infty}\left|1 - F_{1,\ldots,n'}(u_n) - \theta\tau/k\right| \\ &= 1 - \theta\frac{\tau}{k} + o(\tau/k).\end{aligned}$$

Nuevamente por (3.6) se tiene

$$\limsup_{n\to\infty} \Pr(M_n \leq u_n) \leq \left(1 - \theta\frac{\tau}{k} + o(\tau/k)\right)^k,$$

para todo k dado. Haciendo $k \to \infty$,

$$\limsup_{n\to\infty} \Pr(M_n \leq u_n) \leq e^{-\theta\tau}.$$

Similarmente se puede establecer que

$$e^{-\theta\tau} \leq \liminf_{n\to\infty} \Pr(M_n \leq u_n),$$

lo que demuestra el teorema. ∎

En algunas situaciones es conveniente establecer los resultados anteriores en términos de un sólo límite. La condición $D(u_n)$ requiere que la cantidad $\alpha_{n,l_n} \to 0$, cuando $n \to \infty$, para alguna $l_n = o(n)$. Es posible obtener $k_n \to \infty$ tal que tanto

$$k_n l_n \to 0 \tag{3.17}$$

como

$$k_n l_n = o(n) \tag{3.18}$$

se satisfagan ($k_n = \min\{\alpha_{n,l_n}^{-1/2}, (n/l_n)^{1/2}\}$). Usando una sucesión k_n tal, incluimos el siguiente teorema sin demostración el cual es una variante del Teorema 3.5.

Teorema 3.6 (Teorema 3.4 en [22]). *Sea X_n una sucesión estacionaria que satisface la Condición $D(u_n)$ $D(u_n(\tau))$ para cada $\tau > 0$ donde $u_n(\tau)$ satisface* (3.12). *Para algún $\tau_0 > 0$ sea $k_n \to \infty$ tal que* (3.17) *y* (3.18) *se satisfacen con $u_n = u_n(\tau_0)$. Si escribiendo $r_n = [n/k_n]$,*

$$k_n[1 - F_{1,\ldots,r_n}(u_n)] \overset{n\to\infty}{\to} \theta\tau_0, \tag{3.19}$$

entonces X_n tiene índice extremo θ. Recíprocamente, si X_n tiene índice extremo θ entonces (3.19) *se cumple para cada $\tau_0 > 0$ y cada $k_n \to \infty$ que satisface* (3.17) *y* (3.18) *con $u_n = u_n(\tau_0)$.*

Concluimos esta sección con un ejemplo de la literatura que nos muestra algo del rango de posibles conductas asintóticas de M_n. Esto ejemplo está enfocado al caso donde el índice θ es menor que uno.

Ejemplo 3.1. Este ejemplo debido a Chernick (1981) considera una sucesión autorregresiva de primer orden estrictamente estacionaria

$$X_n = \frac{1}{r}X_{n-1} + \epsilon_n,$$

donde $r \geq 2$ es un natural, $\{\epsilon_n\}$ son i.i.d. distribuidas uniformemente sobre $\{0, 1, \ldots, (r-1)/r\}$, ϵ_n independiente de X_{n-1} y la distribución de X_n es uniforme en $[0, 1]$.

Por la uniformidad de X_n, $u_n(\tau)$ puede definirse como $1 - \tau/n$. Chernick muestra que con $u_n = u_n(\tau)$, $\tau > 0$, la condición $D(u_n)$ se cumple, pero $D'(u_n)$ no se satisface. Luego, obtiene con argumentos directos que, para $x > 0$,

$$\Pr\left(M_n \leq 1 - \frac{x}{n}\right) \to \exp\left(-\frac{r-1}{r}x\right). \tag{3.20}$$

Escribiendo $u_n(\tau) = 1 - \tau/n$, (3.20) puede escribirse como

$$\Pr(M_n \leq u_n(\tau)) \to e^{-\theta\tau},$$

mostrando que el índice extremo de la sucesión $\{X_n\}$ es $\theta = (r-1)/r$. Como $r \geq 2$, $0 < \theta < 1$, y r toma los valores $2, 3, \ldots$, el índice θ toma valores en $(0, 1)$.

Es interesante notar que si $\epsilon_{n+1} = (r-1)/r$, con probabilidad $1/r$, entonces $X_{n+1} > X_n$ y por lo tanto sin importar lo grande que X_n sea, hay probabilidad fija $1/r$ de que X_{n+1} sea más grande y entonces hay probabilidad $1/r^2$ de que X_{n+2} sea aún más grande. Así valores grandes tienden a ocurrir en *clusters*, ocasionando que excedencias sucesivas de u_n estén *tan relacionadas* que permiten que la condición $D'(u_n)$ se cumpla y en consecuencia $\theta < 1$.

3.2.2. Proceso Puntual de Cúmulos

A diferencia de [23], donde se estudia la conducta de los conglomerados sobre un nivel específico u_n bajo condiciones más generales que $D(u_n)$, aquí estamos interesados únicamente en estudiar los efectos de la Condición $D(u_n)$, principalmente porque los procesos estacionarias de interés son aquellos con dependencia.

Un medio muy simple de definir conglomerados de excedencias es tomar una sucesión r_n y considerar que ocurren eventos separados una distancia r_n uno de otro y pertenecientes al mismo conglomerado. Podemos elegir r_n tal que sea al menos tan grande como las 'longitudes' de los conglomerados, pero pequeño comparado con la 'separación' de los mismos. Para muchas situaciones usuales esto aún permite flexibilidad para elegir r_n, esto será muy útil para la implementación de métodos computacionales para estimar el índice extremo.

Más específicamente, supondremos que $D(u_n)$ se satisface para $u_n = u_n(\tau)$ satisfaciendo (3.12), una sucesión $k \to \infty$ se debe elegir tal que satisfaga (3.17) y (3.18), $r_n = [n/k_n]$. El proceso puntual N_n de interés estará definido sobre el intervalo $(0, 1]$ de la siguiente manera. Si para $s = 1, \ldots, k_n$ existe una excedencia de u_n por X_j para al menos un j tal que $(s-1)r_n < j \leq sr_n$, entonces N_n tiene un solo evento en el punto $t = sr_n/n$. Esto es, cualquier grupo de excedencias en $((s-1)r_n, sr_n]$ es reemplazado por un solo evento - después del tiempo reescalado - en sr_n/n, 'representando' el grupo original. Llamaremos a N_n el 'proceso puntual de posiciones conglomeradas'. Con esta construcción, el siguiente resultado, incluido sin demostración, se satisface.

Teorema 3.7 (Teorema 4.1 en [22]). *Sea $\{X_n\}$ una sucesión estacionaria satisfaciendo $D(u_n(\tau))$ para cada $\tau > 0$ donde $u_n(\tau)$ satisface* (3.12). *Sea $k_n \to \infty$ elegido tal que satisface* (3.17) *y* (3.18) *y supongamos que $\{X_n\}$ tiene índice extremo. Entonces el proceso puntual N_n de posiciones conglomeradas para excedencias de $u_n(\tau)$ converge en distribución a un Proceso de Poisson N sobre* $(0, 1]$ *con parámetro de intensidad $\theta\tau$.*

Demostración. Por un teorema en Kallenberg [21] solo se necesita probar que

$$\mathbb{E}[N_n(a, b]] \to \mathbb{E}[N(a, b]], \ 0 < a < b \leq 1 \tag{3.21}$$

y

$$\Pr(N_n(E) = 0) \to \Pr(N(E) = 0) \tag{3.22}$$

para cada unión disjunta finita E de conjuntos de la forma (] contenidos en $(0, 1]$.

Si v_n denota el número de intervalos $((s-1)r_n, sr_n]$ completamente contenido en $([na],[nb]]$ es claro que $v_n \sim nr_n^{-1}(b-a) \sim k_n(b-a)$ y además que

$$\mathbb{E}[N_n(a,b]] \sim v_n \Pr\left(\bigcup_{i=1}^{r_n}\{X_i > u_n\}\right)$$
$$k_n(b-a)[1-F_{1,\ldots,r_n}(u_n)] \to (b-a)\theta\tau$$

por (3.19). Pero esto es $\mathbb{E}(a,b]$ así que (3.21) se cumple.

Para mostrar (3.22) escribamos $E = \cup_{j=1}^{p}(a_j,b_j]$ y escribamos B_j para los enteros en $([na_j],[nb_j]]$. Entonces es fácil ver que

$$\Pr(N_n(E)=0) = \Pr\left(\bigcap_{j=1}^{p}\{M(B_j)\le u_n\}\right) + o(1) \tag{3.23}$$

$$\prod_{j=1}^{p}\Pr(M(B_j)\le u_n) + \left[\Pr\left(\bigcup_{j=1}^{p}\{M(B_j)\le u_n\}\right) - \prod_{j=1}^{p}\Pr(M(B_j)\le u_n)\right] + o(1). \tag{3.24}$$

Por inducción, la diferencia en los corchetes cuadrados no excede $p\alpha_{n,n\lambda}$ en valor absoluto donde λ es la separación mínima de los intervalos $(a_j,b_j]$ (λ puede ser tomado distinto de cero porque intervalos adjuntos pueden combinarse). Pero $\alpha_{n,l}$ puede tomarse no creciente en l (véase [23]) y se sigue de $D(u_n)$ que $\alpha_{n,n\lambda} \to 0$ cuando $n \to 0$. Como X_n tiene índice extremo se tiene que $\Pr(M(B_j)\le u_n) \to e^{-\theta\tau(b_j-a_j)}$ y por lo tanto

$$\Pr(N_n(E)=0) \to \prod_{j=1}^{p}\Pr(N_n(a_j,b_j]=0)$$
$$= \Pr(N(E)=0),$$

probando (3.22). ■

Debemos notar que las distribuciones límite de los estadísticos de orden serán afectadas de una manera más complicada por la conglomeración que la distribución del máximo. Estas distribuciones surgirán de un resultado límite más complicado para excedencias individuales. Sin embargo, el uso del resultado Poisson dado arriba resultará en las distribuciones para las alturas del 'k -ésimo conglomerado más alto' en lugar del k-ésimo estadístico de orden extremo, de una manera clara. Finalmente, en este teorema se está pidiendo que (3.19) se cumpla para algún τ_0. Esto, permite dar una interpretación del índice extremo.

Supongamos que $\{X_i\}_{i=1}^n$ tiene índice extremo θ, es decir que podemos dividir la serie en k_n bloques de tamaño r_n con $k_n = o(n)$, $k_n\, l_n = o(n)$, l_n como en (3.5) y $r_n = [n/k_n]$, tal que

$$k_n(1 - F_{1,\ldots,r_n}(u_n(\tau))) \overset{n\to\infty}{\to} \theta\,\tau, \tag{3.25}$$

donde $u_n(\tau)$ satisface (3.12). De ahora en adelante $u_n \equiv u_n(\tau)$.

Consideremos el número de excedencias de u_n dentro de un bloque de tamaño r_n. Este evento define la variable aleatoria

$$B_{r_n}^{u_n} = \sum_{i=1}^{r_n} \mathbf{1}_{\{X_i > u_n\}},$$

de la cual podemos calcular

$$\begin{aligned}
E[B_{r_n}^{u_n} \mid B_{r_n}^{u_n} \geq 1] &= \sum_{j=1}^{r_n} j \ \Pr(B_{r_n}^{u_n} = j \mid B_{r_n}^{u_n} \geq 1) \\
&= \frac{\sum_{j=1}^{r_n} j \ \Pr(B_{r_n}^{u_n} = j)}{\Pr(B_{r_n}^{u_n} \geq 1)} \\
&= \frac{E[B_{r_n}^{u_n}]}{\Pr(\cup_{j=1}^{r_n} \{X_j \geq u_n(\tau)\})} \\
&= \frac{r_n\,[1 - F(u_n)]}{1 - F_{1,\ldots,r_n}(u_n)} \\
&\approx \frac{n\,[1 - F(u_n)]}{k_n\,[1 - F_{1,\ldots,r_n}(u_n)]}
\end{aligned}$$

recordando (3.12) y (3.19) tenemos

$$E[B_{r_n}^{u_n} \mid B_{r_n}^{u_n} \geq 1] \overset{n\to\infty}{\to} \theta^{-1}. \tag{3.26}$$

A partir del resultado anterior proponemos la siguiente interpretación. El tamaño medio de un conglomerado *en posición* puede ser interpretado como el número medio (en el límite) de excedencias en un intervalo de longitud r_n, dado que al menos ocurre una excedencia en ese intervalo. Asumiendo el enfoque basado en eventos este parámetro nos puede dar una primera idea del tiempo que un evento está presente en la serie, debido a que existirán excedencias a contar al tiempo t siempre que la perturbación causada por el evento siga presente en ese instante.

Aprovechemos lo construido hasta ahora. Consideremos un nivel $v_n(\tau)$, *nivel extremo*, tal que

$$E[B_{r_n}^{v_n(\tau)} \mid B_{r_n}^{u_n} \geq 1] \overset{n\to\infty}{\to} 1, \tag{3.27}$$

se sigue que $v_n(\tau) \geq u_n$ y por el Lema 3.6.2 en [23], $D(v_n(\tau))$ se satisface.

Análogamente a (3.26), se establece

$$\frac{r_n\,[1 - F(v_n(\tau))]}{1 - F_{1,\ldots,r_n}(u_n)} \overset{n\to\infty}{\to} 1,$$

como hemos supuesto que (3.19) se satisface para u_n, el anterior límite implica que

$$n\,[1 - F(v_n(\tau))] \overset{n\to\infty}{\to} \theta\,\tau.$$

y de acuerdo al Teorema 3.6 esto es suficiente para que

$$k_n[1 - F_{1,\ldots,r_n}(v_n(\tau))] \overset{n\to\infty}{\to} \theta^2\tau.$$

Proponemos un primer estimador del índice extremo. Sea $\hat{F}_{1,\ldots,r_n}(\cdot)$ un estimador de $F_{1,\ldots,r_n}(\cdot)$, la función de distribución conjunta de $\{X_1, \cdots, X_{r_n}\}$ entonces:

$$\theta_{\hat{F}} = \frac{1 - \hat{F}_{1,\ldots,r_n}(v_n(\tau))}{1 - \hat{F}_{1,\ldots,r_n}(u_n)}. \tag{3.28}$$

Al margen del estimador anterior, estudiaremos otros estimadores con propiedades estadísticas establecidas. Básicamente, estos estimadores se basan en 'adelgazar' el proceso $\{X_i\}_{i=1}^n$ por las sucesiones k_n y r_n. Es claro que podemos definir otros adelgazamientos, basta considerar variables aleatorias distintas a $B_{r_n}^{u_n}$ y a $B_{r_n}^{v_n(\tau)}$.

Sean k_n y r_n como arriba, definamos

$$M_j^{k_n} = \text{máx}\,\{X_k : (j-1)\,r_n + 1 \leq k \leq j\,r_n\} \quad j = 1, \ldots, k_n,$$

los máximos 'locales' del proceso $\{X_i\}_{i=1}^n$ a través de la partición de k_n bloques de tamaño r_n. Ahora el proceso de adelgazamiento consiste en considerar estos máximos locales (*block maxima*) que excedan el nivel u_n. Así, definimos la variable aleatoria

$$Z_{r_n}^{u_n} = \sum_{i=1}^{k_n} \mathbf{1}_{\{M_j^{k_n} > u_n\}}.$$

Observemos que este procedimiento nos conduce a un *proceso puntual de conglomerados por posición* $N_t^{u_n(\tau)}$ sobre el intervalo $(0, 1]$ consistente de los elementos de $Z_{r_n}^{u_n}$ indexado por $t = j/k_n$, $j = 1, \ldots, k_n$. Por el Teorema 2.2 este proceso puntual converge a un proceso de Poisson N con parámetro de intensidad $\theta\,\tau$. Análogamente, definimos el proceso $N_t^{v_n(\tau)}$, considerando la variable aleatoria $Z_{r_n}^{v_n(\tau)}$, dado que se cumplen las condiciones del Teorema 2.2 este proceso convergerá a otro proceso de Poisson N' de media $\theta^2\,\tau$.

Por tanto el índice extremo estará caracterizado por el siguiente resultado

$$\lim_{n\to\infty} \frac{E[N_t^{v_n(\tau)}]}{E[N_t^{u_n(\tau)}]} = \theta. \tag{3.29}$$

3.3. Propuesta de Olmo

Los estimadores dados arriba, necesitan que n sea grande, lo cual implicará que los umbrales (niveles) u_n y $v_n(\tau)$ deban ser altos. Sin embargo se pueden definir estimadores a través de *niveles moderadamente altos* como se muestra a continuación. Básicamente las siguientes secciones estarán tomadas de Olmo [31] y Olmo [32].

Empecemos con la siguiente

Definición 3.5. Sea u_n^* similar a u_n, asumamos que este umbral satisface lo siguiente:

1. $D(u_n^*)$.
2. $k_n\,[1 - F_{1,\ldots,r_n}(u_n^*)] \to \infty$, cuando $n \to \infty$.
3. $1 - F_{1,\ldots,r_n}(u_n^*) = [1 - F_{1,\ldots,r_n}(u_n)]\, s_n$, con $s_n = o\left(\frac{1}{1-F_{1,\ldots,r_n}(u_n)}\right)$.
4. $1 - F(u_n^*) = [1 - F(u_n)]\, s_n^{'}$, con $s_n = o\left(\frac{1}{1-F(u_n)}\right)$.
5. $\frac{s_n^{'}}{s_n} \to 1$, cuando $n \to \infty$.

Una sucesión $\{u_n^*\}$ que satisface lo anterior se denomina de niveles moderadamente altos.

Proposición 3.1. *Sea $\{u_n^*\}$ una sucesión de niveles moderadamente altos, y sea c_n una realización de $Z_{r_n}^{u_n^*}$. Si c_n satisface*

$$\frac{c_n - s_n\,\theta\,\tau}{(\theta\,\tau\,s_n\,F_{1,\ldots,r_n}(u_n^*))^{1/2}} \overset{n\to\infty}{\to} \lambda, \tag{3.30}$$

entonces

$$\Pr\left(Z_{r_n}^{u_n^*} \le c_n\right) \overset{n\to\infty}{\to} \Phi(\lambda), \tag{3.31}$$

donde $\Phi(\cdot)$ es la función de distribución normal estándar.

Sea $M_{j:k_n}$ el j-ésimo estadístico de orden de la sucesión $\left\{M_j^{k_n}\right\}_{j=1}^{k_n}$, convenimos en que

$$M_{1:k_n} \ge M_{2:k_n} \ge \cdots \ge M_{k_n:k_n}.$$

Es claro que

$$\left\{Z_{r_n}^{u_n^*} \leq c_n\right\} = \left\{M_{c_n+1:k_n} \leq u_n^*\right\}, \tag{3.32}$$

por lo que

$$\Pr\left(M_{c_n+1:k_n} \leq u_n^*\right) \overset{n\to\infty}{\to} \Phi(\lambda), \tag{3.33}$$

con λ como en Proposición 3.1.

El umbral $v_n(\tau)$, tendrá su homólogo moderadamente alto. Sea v_n^* similar a $v_n(\tau)$, asumamos que este umbral satisface lo siguiente:

1. $k_n\left[1 - F_{1,\ldots,r_n}(v_n^*)\right] \to \infty$, cuando $n \to \infty$.

2. $1 - F_{1,\ldots,r_n}(v_n^*) = \left[1 - F_{1,\ldots,r_n}(v_n(\tau))\right] t_n$, con $t_n = o\left(\frac{1}{1-F_{1,\ldots,r_n}(v_n(\tau))}\right)$.

3. $1 - F(v_n^*) = \left[1 - F(v_n(\tau))\right] t_n^{'}$, con $t_n = o\left(\frac{1}{1-F(v_n(\tau))}\right)$.

4. $\frac{s_n^{'}}{s_n} \to 1$, cuando $n \to \infty$.

5. Finalmente:

$$E\left[B_{r_n}^{v_n^*} \mid B_{r_n}^{u_n^*} \geq 1\right] \overset{n\to\infty}{\to} 1. \tag{3.34}$$

Observemos que (3.34) implica que

$$\frac{n\left[1 - F(v_n^*)\right]}{k_n\left[1 - F_{1,\ldots,r_n}(u_n^*)\right]} \overset{n\to\infty}{\to} 1.$$

Esto produce que $\frac{t_n^{'}}{s_n} \to 1$ y en consecuencia que $\frac{t_n}{s_n} \to 1$ cuando $n \to \infty$. Por tanto

$$\frac{1 - F_{1,\ldots,r_n}(v_n^*)}{1 - F_{1,\ldots,r_n}(u_n^*)} \overset{n\to\infty}{\to} \theta.$$

De estas observaciones se sigue que $v_n^* \geq u_n^*$ y $D(v_n^*)$ se cumple. Consideremos la variable aleatoria $Z_{r_n}^{v_n^*}$.

Proposición 3.2. *Sea u_n^* como en Proposición 3.1 y supongamos que c_n es tal que* (3.30) *se satisface. Sea v_n^* como arriba, entonces*

$$\Pr\left(Z_{r_n}^{v_n^*} \leq \theta\, c_n\right) \overset{n\to\infty}{\to} \Phi(\lambda), \tag{3.35}$$

donde $\Phi(\cdot)$ es la función de distribución normal estándar.

En las siguientes secciones será clara la inserción de los umbrales moderadamente altos, por el momento sólo presentamos un estimador del índice extremo por umbrales moderadamente altos. A saber,

$$\lim_{n\to\infty} \frac{E\left[Z_{r_n}^{v_n^*}\right]}{E\left[Z_{r_n}^{u_n^*}\right]} = \theta^*. \tag{3.36}$$

3.3.1. Estimadores

Por lo dicho anteriormente, el índice extremo provee una medida de la *conglomeración* de las observaciones más grandes de una sucesión estacionaria. Bajo la Condición $D(u_n)$ la distribución límite de M_n es $F^{n\theta}(u_n)$ en lugar de $F^n(u_n)$. Este resultado generó uno de los primeros estimadores del índice extremo. Para sucesiones apropiadas k_n, r_n se satisface que $\Pr^{k_n}(M_{r_n} \leq u_n) \approx \Pr(M_n \leq u_n)$, tomando logaritmos tenemos, $k_n \ln \Pr(M_{r_n} \leq u_n) \approx n\,\theta\, \ln F(u_n)$, de donde se deduce que

$$\theta \approx \frac{\ln \Pr(M_{r_n} \leq u_n)}{r_n \ln F(u_n)}.$$

Entonces, un estimador natural del índice extremo es

$$\hat{\theta}_n^{(1)} = \frac{\ln(1 - Z_{r_n}^{u_n}/k_n)}{r_n \ln(1 - B_{r_n}^{u_n}/n)}. \tag{3.37}$$

Este estimador se denomina el *método logs*.

Alternativamente, el concepto de índice extremo introducido en [22], θ^{-1} como el límite del número promedio de excedencias, produce el *método por bloques*

$$\hat{\theta}_n^{(2)} = \frac{Z_{r_n}^{u_n}}{B_{r_n}^{u_n}}. \tag{3.38}$$

Este estimador puede ser considerado como una aproximación de $\hat{\theta}_n^{(1)}$ usando las expansiones de primer orden del logaritmo en el numerador y denominador de (3.37).

La caracterización dada en Hsing [20] motiva un método diferente para estimar tal parámetro. Concretamente,

$$\bar{\theta} = \frac{W_{r_n}^{u_n}}{B_{r_n}^{u_n}}, \tag{3.39}$$

donde

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

Éste es el conocido *método de corridas*.

El estimador del índice extremo por *niveles altos* está dado por

$$\hat{\theta} = \frac{Z_{r_n}^{v_n(\tau)}}{Z_{r_n}^{u_n}}. \tag{3.40}$$

Análogamente, el estimador del índice extremo por *niveles moderadamente altos* está dado por

$$\hat{\theta}^* = \frac{Z_{r_n}^{v_n(\tau)^*}}{Z_{r_n}^{u_n^*}}. \tag{3.41}$$

Desarrollaremos en detalle el primer estimador. Con la notación de arriba la contraparte empírica de (3.27) es

$$\frac{1}{Z_{r_n}^{u_n(\tau)}} \sum_{j=1}^{k_n} \sum_{i=k}^{j\,r_n} \mathbf{1}_{\{X_k > v_n(\tau)\}} \overset{n\to\infty}{\to} 1. \tag{3.42}$$

Escribamos

$$\Lambda_{k_n}^{v_n(\tau)} = \sum_{j=1}^{k_n} \sum_{i=k}^{j\,r_n} \mathbf{1}_{\{X_k > v_n(\tau)\}}.$$

Construimos el estimador dado en (3.40) a partir de los siguientes puntos

A.1 Considérense sucesiones k_n y r_n apropiadas en el sentido de la ecuación (3.5).

A.2 Construir la sucesión $\left\{M_j^{k_n}\right\}_{j=1}^{k_n}$ a partir de $\{X_i\}_{i=1}^n$.

A.3 Haciendo $u_n = M_{c+1:k_n}$ para un c fijo (c pequeño), por (3.32) se tiene que $Z_{r_n}^{u_n} = c$

A.4 Para cumplir exactamente (3.42), hacemos $\Lambda_{k_n}^{\tau} = X_{c+1:n}$.

A.5 Calculamos $Z_{r_n}^{v_n} = \sum_{j=1}^{k_n} \mathbf{1}_{\{M_j^{k_n} > v_n\}}$.

A.6 Finalmente, obtenemos

$$\hat{\theta} = \frac{Z_{r_n}^{v_n}}{Z_{r_n}^{u_n}} \tag{3.43}$$

Este estimador puede ser interpretado como un refinamiento del método de bloques donde el nivel u_n en (3.38) es sustituido por v_n.

El procedimiento para $\hat{\theta}^*$ es similar. En este caso $B_{r_n}^{\tilde{v}_n} = Z_{r_n}^{\tilde{u}_n}$ y $\tilde{v}_n = X_{Z_{r_n}^{\tilde{u}_n}+1:n}$ con $X_{Z_{r_n}^{\tilde{u}_n}+1:n}$ un estadístico de orden de $\{X_i\}$. El nivel $\tilde{u}_n$ está determinado por las condiciones A.1 - A.6. Una elección adecuada de este nivel es $M_{c_n+1:k_n}$ con $c_n \to \infty$ y $c_n = o(k_n)$.

En la práctica la elección correcta de los niveles u_n y u_n^* no es tan importante como escoger apropiadamente los niveles v_n y v_n^* para satisfacer (3.19). La diferencia entre estos estimadores de θ yace en la distribución límite de sus componentes. En pocas palabras, $B_{r_n}^{u_n}$ converge a una distribución de Poisson mientras que $B_{r_n}^{\tilde{u}_n}$ satisface el Teorema de Límite Central. Estos resultados son importantes para la inferencia efectuada sobre θ.

3.3.2. Inferencia Estadística

Los métodos estándar para estimar el índice extremo se basan en la elección de un nivel extremo u_n. Para particiones apropiadas de la sucesión estacionaria este nivel determina el tamaño del cluster *block cluster size*. Por definición del nivel extremo el número de excedencias dentro de cada cluster es aproximadamente constante cuando n crece. *Más aún por las propiedades de la distribución de Poisson la varianza del tamaño del cluster converge a una constante diferente de cero.* Por lo tanto, estimadores basados en estos niveles no tienen éxito en proveer estimadores más precisos del índice extremo cuando n es grande. Esto, junto con la presencia de dependencia en $\{X_n\}$ hace difícil encontrar la distribución de los estimadores para θ comúnmente usados.

Por otro lado, la extensión de estos estimadores a niveles moderadamente altos no es trivial. Hsing [19] muestra que la distribución de los clusters de excedencias definidas para niveles extremos converge a una distribución geométrica. Si bajamos el nivel para conseguir consistencia, el número de excedencias dentro del cluster se incrementa con n y ya no converge a una función de distribución. Para resolver este problema se incluyó un nivel más bajo definido por una sucesión, digamos s_n, que convergía a infinito. Incrementando el tamaño de los clusters determinado por el nivel más bajo son estandarizados por y_n para obtener una v.a. Se propuso una variante del estimador por el método de bloques para θ que es asintóticamente normal.

La caracterización de θ en [31] como el límite de un cociente determinado por dos niveles hace posible determinar inferencia estadística sobre el parámetro. Bajo $D(u_n)$ o alternativamente $D(\tilde{u}_n)$ y para n suficientemente alto, numerador y denominador de $\hat{\theta}_n$ y $\hat{\theta}_n^*$ son bien aproximados por una distribución binomial. Estos estimadores solo difieren en su conducta límite.

De acuerdo a la sección anterior sabemos que $Z_{r_n}^{u_n}$ y $Z_{r_n}^{v_n}$ son asintóticamente Poisson de parámetros $\theta\tau$ y $\theta^2\tau$, respectivamente. Se puede probar que la distribución condicional de $Z_{r_n}^{v_n}$ dado $Z_{r_n}^{u_n}$ es asintóticamente Binomial$(p_n, Z_{r_n}^{u_n})$, donde

$$p_n = \frac{1 - F_{1,\ldots,r_n}(v_n)}{1 - F_{1,\ldots,r_n}(u_n)}$$

Para encontrar los primeros momentos no condicionales de $\hat{\theta}$ calcularemos el valor esperado y

varianza condicionales. Esto es inmediato de la distribución condicional de $Z_{r_n}^{v_n}$ dado $Z_{r_n}^{u_n}$. Entonces

$$\begin{aligned}\mathbb{E}[\hat{\theta}] &= \mathbb{E}\left[\mathbb{E}\left[\frac{Z_{r_n}^{v_n}}{Z_{r_n}^{u_n}} \mid Z_{r_n}^{u_n}\right]\right] \\ &= \frac{1}{Z_{r_n}^{u_n}}\mathbb{E}\left[Z_{r_n}^{v_n} \mid Z_{r_n}^{u_n}\right] \\ &= \frac{1}{Z_{r_n}^{u_n}} Z_{r_n}^{u_n} \frac{1-F_{1,\ldots,r_n}(v_n)}{1-F_{1,\ldots,r_n}(u_n)}.\end{aligned} \tag{3.44}$$

Análogamente, se calcula

$$\begin{aligned}\mathbb{V}[\hat{\theta}] &= \mathbb{V}[\mathbb{E}[Z_{r_n}^{v_n} \mid Z_{r_n}^{u_n}]] + \mathbb{E}[\mathbb{V}[Z_{r_n}^{v_n} \mid Z_{r_n}^{u_n}]] \\ &= \mathbb{E}\left[\frac{1}{{Z_{r_n}^{u_n}}^2} Z_{r_n}^{u_n} \frac{1-F_{1,\ldots,r_n}(v_n)}{1-F_{1,\ldots,r_n}(u_n)}\right] \\ &= \left(\frac{1-F_{1,\ldots,r_n}(v_n)}{1-F_{1,\ldots,r_n}(u_n)}\right)\left(1-\frac{1-F_{1,\ldots,r_n}(v_n)}{1-F_{1,\ldots,r_n}(u_n)}\right)\mathbb{E}\left[\frac{1}{Z_{r_n}^{u_n}}\right].\end{aligned}$$

Utilizando expansión de Taylor de $\mathbb{E}[1/Z_{r_n}^{u_n}]$ alrededor de $\mathbb{E}[Z_{r_n}^{u_n}]$ obtenemos que

$$\mathbb{V}\left[\hat{\theta}\right] = \frac{1-\theta}{\tau} + O(1). \tag{3.45}$$

Esto es, la varianza converge a una constante diferente de cero para τ constante. Aunque $\hat{\theta}$ es asintóticamente insesgado, éste estimador no es consistente. La definición de $\hat{\theta}^*$ como el cociente del número de excedencias de niveles moderadamente altos fue motivado por la falta de consistencia de $\hat{\theta}$. Nótese que los cálculos para obtener los dos primeros momentos de este estimador, son análogos a los presentados previamente. Por tanto podemos decir que

$$\mathbb{E}[\hat{\theta}^*] = p_n^* = \frac{1-F_{1,\ldots,r_n}(v_n^*)}{1-F_{1,\ldots,r_n}(u_n^*)} \to \theta,$$

y

$$\mathbb{V}[\hat{\theta}^*] = p_n^*(1-p_n^*)\left(\frac{1}{\mathbb{E}[Z_{r_n}^{\tilde{u}_n}]} + \frac{\mathbb{V}[Z_{r_n}^{\tilde{u}_n}]}{\mathbb{E}[Z_{r_n}^{\tilde{u}_n}]^3}\right).$$

Por definición del nivel $\tilde{u}_n$ tenemos

$$\mathbb{V}[\hat{\theta}^*] \overset{n\to\infty}{\to} 0.$$

Por lo tanto,

$$\hat{\theta}^* \overset{P}{\to} \theta.$$

En la práctica para evitar incertidumbre sobre $Z_{r_n}^{\tilde{u}_n}$ el nivel $\tilde{u}_n$ se supone un estadístico de orden intermedio $\tilde{u}_n = M_{c_n+1:k_n}$ con $c_n \to \infty$ y $c_n = o(k_n)$. La distribución binomial de $Z_{\tilde{v}_n} \mid Z_{\tilde{u}_n}$ es bien aproximada por una distribución normal $N(p_n^* c_n, p_n^*(1-p_n^*)c_n)$.

Por lo tanto para n suficientemente grande

$$\hat{\theta}^* \overset{D}{\sim} N\left(\theta, \frac{\theta(1-\theta)}{c_n}\right) \tag{3.46}$$

Capítulo 4

Aportaciones

Este capítulo contiene estudios relacionados con los temas antes expuestos. Primero mostraremos la eficiencia del estimador del índice extremo por el método de Olmo [32] vía simulaciones. Posteriormente, usaremos dicho estimador para estudiar el tamaño medio de los clusters en diversas series de tiempo financieras. Además, trataremos de establecer el hecho de que los modelos GARCH(1,1) subestiman el tamaño real de los clusters. Consideramos que los múltiples resultados establecidos en el Capítulo 2 acerca de la estacionariedad ergódica de un proceso GARCH(1,1), nos permiten establecer una correcta inferencia estadística, indirectamente si se quiere, acerca del fenómeno de estudio en este trabajo: la volatilidad conglomerada. Finalmente, producto de los estudios de simulación contamos con una población de tamaño apropiado de los estimadores del índice extremo para las series consideradas y nos resultó interesante estudiar la distribución de estas poblaciones.

4.1. Estudio del Comportamiento del Estimador de Olmo

4.1.1. Para determinar competitividad con otros estimadores

En la Sección 3.3.1, hicimos un repaso de los diferentes estimadores utilizados para calcular el índice extremo. Dependiendo de la caracterización utilizada para determinar el estimador, en aquella sección incluimos: Método Logs, método por Bloques y el método de corridas. El objetivo de esta parte de la tesis es presentar un comparativo gráfico de la eficiencia del estimador del índice extremo por el método de Olmo, o de los $n$iveles moderadamente altos contra los métodos clásicos de estimación de este parámetro.

Para esto, utilizamos el proceso definido en el Ejemplo 3.1, el Proceso de Chernick de parámetro

r. Simulamos 2500 series correspondientes a esta clase de procesos, cada una con longitud de 8000 observaciones, decidimos hacer $r = 4$ y el número de bloques utilizados fue de 40. Esto último pudo establecer a través de previos estudios de simulación, en donde variando el número de bloques fuimos capacez de observar que en el rango de $\{25, \ldots, 50\}$ para el parámetro de bloques, la convergencia de cada uno de los estimadores resultaba adecuada. Las gráficas presentadas en la Figura 4.1 estudiarán el comportamiento de los estimadores previamente citados únicamente en función de los umbrales utilizados en el cálculo del índice extremo. Los mismos estudios de simulación realizados nos sugieren utilizar porcentajes del tamaño de las observaciones como umbrales apropiados. Por ejemplo, si decimos que estamos usando umbrales del 1 %, esto significa que el número total de excedencias a utilizar es el 1 % del tamaño de la población en estudio.

4.1.2. Para determinar número de bloques y umbrales

En el capítulo anterior mostramos las propiedades estadísticas del llamado estimador por niveles moderadamente altos. Antes de presentar el algoritmo que nos permite calcular tal estimador, recordaremos algunos conceptos ya mencionados y que son fundamentales para establecer el estimador.

Los resultados establecidos en el capítulo anterior dependen fuertemente de la Condición de Leadbetter. Para establecer tal condición bastaba encontrar k_n y r_n tal que los k_n bloques de tamaño r_n sean 'asintóticamente independientes'. El establecer estos valores es un problema abierto, principalmente por la dificultad en establecer la distribución conjunta del proceso bajo consideración. Por otro lado, el estimador de Olmo, requiere el conocimiento de un nivel u_n a partir del cual se define un proceso puntual de excedencias. Debemos ligar estos tres números de alguna forma.

Consideremos un conjunto $\{X_t\}_{t\geq 0}$ de datos de tamaño N. Trabajemos con los datos ordenados de menor a mayor, i.e. $\{X_{(t)}\}_{t\geq 0}$, éstos son los estadísticos de orden. Como primer parámetro de la función estimadora tendremos al número de excedencias a contar, digamos E. A partir de esto definimos $u_n = X_{(N-E)}$. El segundo parámetro será el número de bloques a considerar, digamos B. Esto hace que tanto el número total de bloques $k_n = B$ como el tamaño $r_n = [N/B]$ queden determinados.

Consideremos la parte operativa del estimador. Necesitamos contar las excendencias que se tienen del nivel u_n, llamemos a esta cantidad T_1. Posteriormente, construir el conjunto de máximos 'locales' definido anteriormente:

$$M_j^B = \text{máx}\ \{X_k : (j-1)\,[N/B] + 1 \leq k \leq j\,[N/B]\} \quad j = 1, \ldots, B,$$

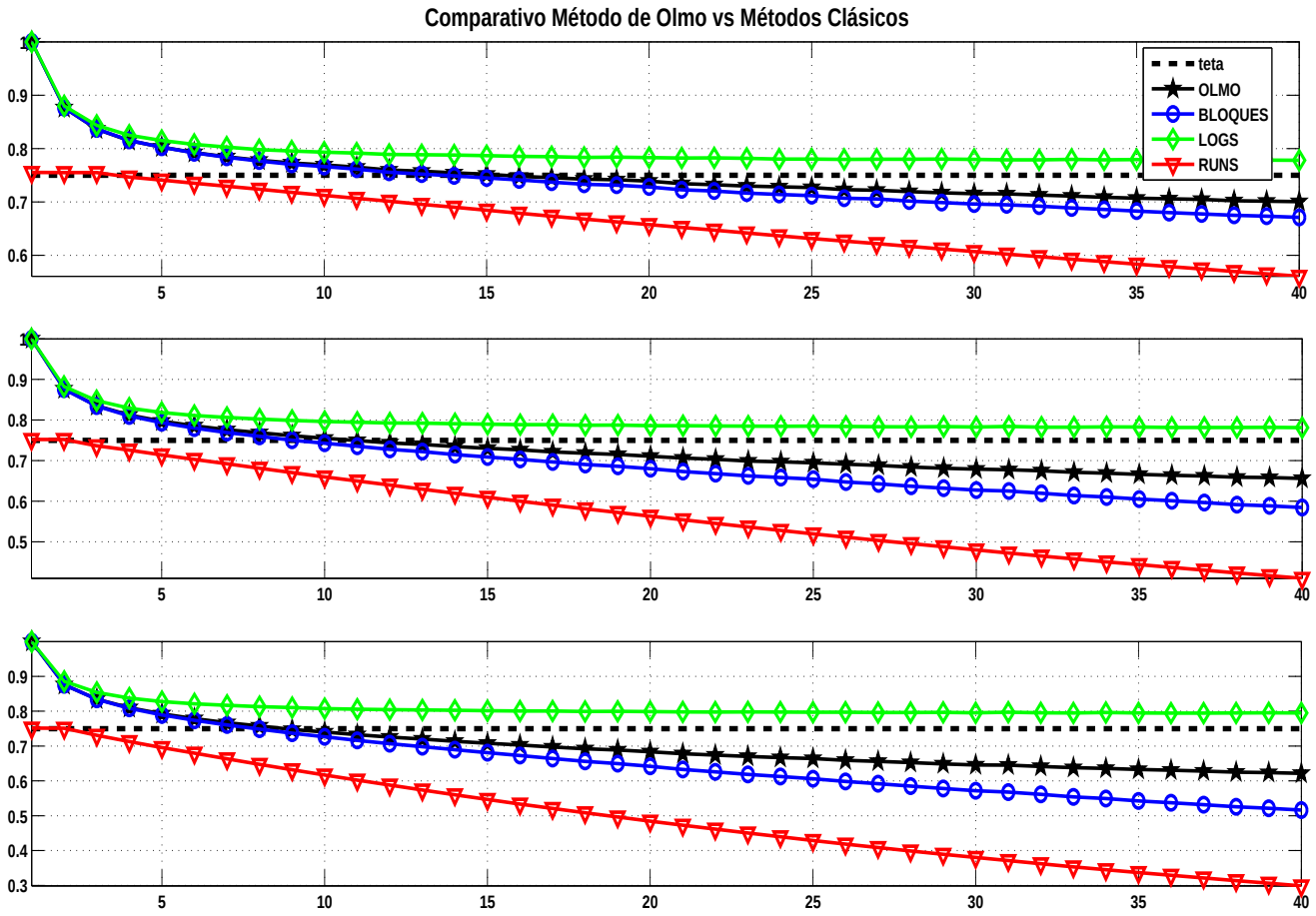


Figura 4.1: De arriba hacia abajo: Considerando el 1 %, los métodos de Olmo y Logs presentan convergencia adecuada al valor real del índice extremo, 0.75. Estudiando excedencias al 2 %, el método Logs presenta sobre estimación del parámetro buscado, el comportamiento de convergencia del método de Olmo supera al de los métodos de Corridas y Bloques. Finalmente, al 3 %, nuevamente el método Logs sobre estima el índice extremo teórico, el método de Olmo presenta convergencia adecuada y mejora las trayectorias de los métodos Corridas y Bloques.

del cual contamos todos los M_j^B's que exceden el nivel u_n, digamos T_2. Enseguida se define el siguiente nivel moderadamente alto de acuerdo a la definición dada en el capítulo anterior. Sea $v_n = X_{(N-T_2)}$ tal número. Contamos el número de máximos locales que exceden este nivel, digamos T_3. Finalmente, el valor del estimador del índice extremo con E excedencias y B bloques está dado por $\frac{T_2}{T_3}$.

Podemos resumir lo anterior en los siguientes puntos

B.1 Se introducen los parámetros excedencias E y bloques B.

B.2 Calculamos los estadísticos de orden $X_{(\cdot)}$.

B.3 Definimos el nivel u_n, $u_n = X_{(N-E)}$.

B.4 Calculamos el total de excedencias del proceso X sobre el nivel u_n, digamos T_1.

B.5 Definimos el proceso de máximos locales del proceso M_n.

B.6 Calculamos el total de excedencias del proceso M_n sobre el nivel u_n, digamos T_2.

B.7 Definimos el nivel v_n, $v_n = X_{(N-T_2)}$.

B.8 Calculamos el total de excedencias del proceso M_n sobre el nivel v_n, digamos T_3.

B.9 Finalmente $\hat{\theta} = \frac{T_2}{T_3}$.

Una lectura rápida a la Sección 3.3.1 nos dejará claro que el algoritmo anterior garantiza los resultados estadísticos expuestos también en aquella sección. La rutina computacional para calcular el índice extremo de un proceso estacionario fue implementado en el lenguaje MATLAB.

Consideremos el proceso definido en el Ejemplo 3.1. Recordemos que se estableció que el índice extremo de tal proceso está dado por $\theta = (r-1)/(r)$, donde $r \in \mathbb{N}$. La Figuras (4.2) (4.3) y (4.4) muestran el desempeño del estimador de Olmo en función de r, N, E y B, parámetro del proceso, número de datos simulados, número de excedencias y número de bloques, respectivamente.

Simulamos un proceso de Chernick [7] con $r = 4$. Teóricamente el índice extremo de este proceso es igual a $3/4 = 0.75$. La Figura 4.2 muestra el comportamiento del estimador del índice extremo para este proceso con $N = 1900$, $B = 35$. Si consideramos un proceso similar pero con $N = 3000$, $B = 50$, la Figura 4.3 muestra el comportamiento del estimador de su índice extremo. Finalmente, la Figura 4.4 despliega el comportamiento del mismo objeto para el mismo proceso con $N = 8000$ y $B = 100$.

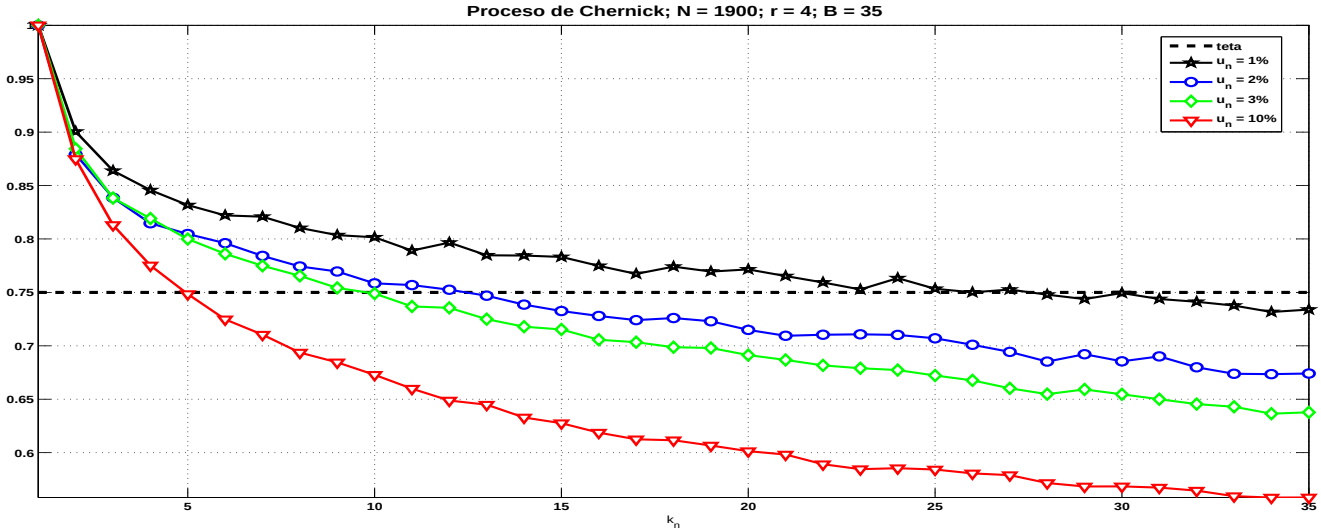


Figura 4.2: La trayectoria del estimador de índice extremo con umbral de 1 % converge apropiada y rápidamente al valor teórico del índice extremo de este proceso, la línea punteada representa dicho valor teórico. Observemos que para los valores de umbrales del 2 %, 3 % y 10 %, la convergencia de las trayectorias a los respectivos índices extremos no es aparente, aunque debemos considerar que la escala de la figura sugiere un buen comportamiento del estimador del índice extremo en estos casos.

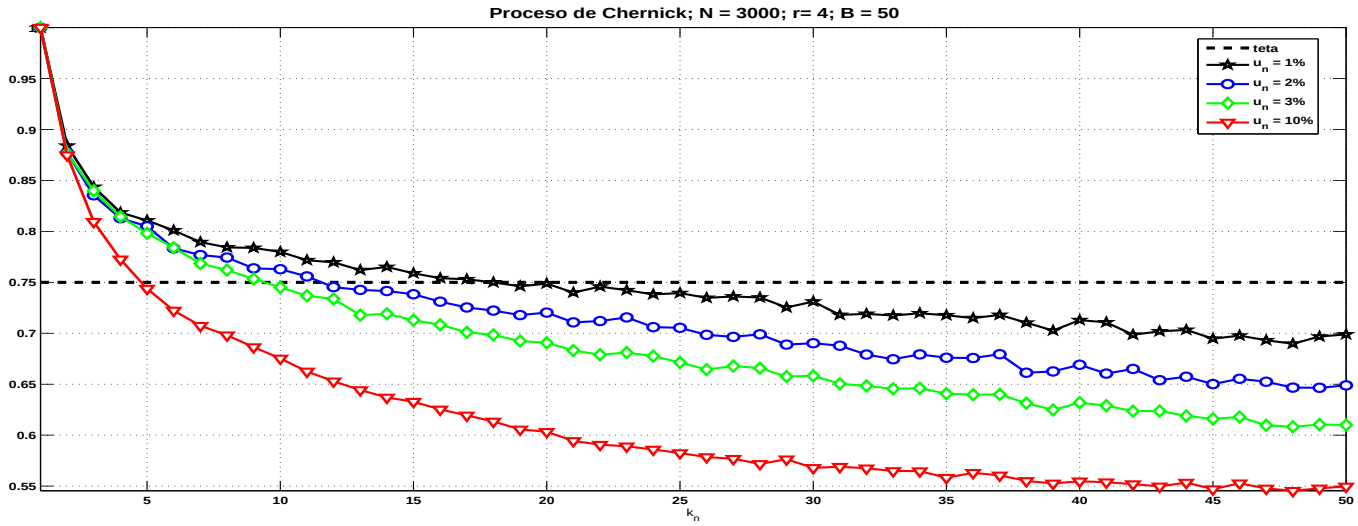


Figura 4.3: Esta figura estudia el comportamiento de un proceso de Chernick de parámetro $r = 4$, y con $N = 3000$. Otra modificación respecto a la figura anterior es el número de bloques utilizados, 50 en este caso. Nuevamente, para umbral de 1 % se presenta una adecuada convergencia. Respetando la escala en la que las trayectorias son presentadas, el comportamiento para el resto de umbrales presentados, resulta moderadamente adecuado.

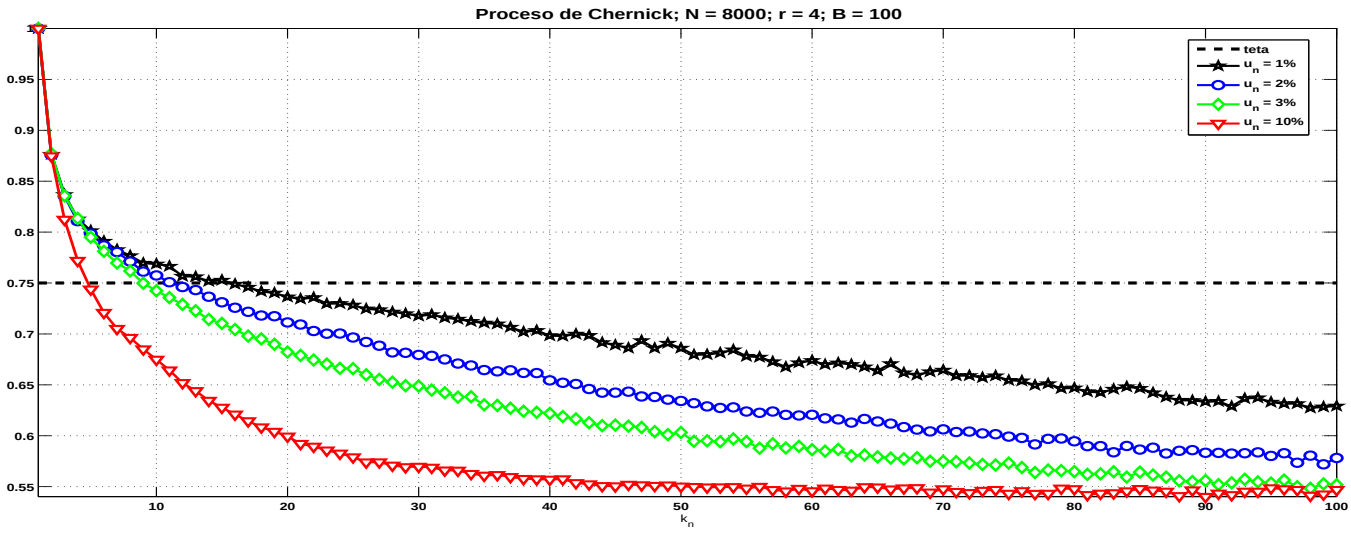


Figura 4.4: En esta última gráfica que estudia el comportamiento del estimador del índice extremo por el método de umbrales moderadamente altos, presentamos las trayectorias de tal estimador para un proceso de Chernick de parámetros $r = 4$; se usan $N = 8000$ y $B = 100$, como parámetros del método de estimación del índice extremo. La convergencia de estas trayectorias es poco satisfactoria, para cualquiera de los umbrales considerados, lo cual nos revela que trabajar con 100 bloques posiblemente no es una buena elección de este parámetro.

4.2. Clustering en GARCH

Calculamos el índice extremo a una serie de datos reales. Por otro lado, ajustamos un GARCH(1,1), obtenemos α_1 y β_1, y simulamos N series y a cada una de éstas les calculamos el índice extremo. Presentamos el promedio de estas estimaciones. Utilizaremos los siguientes conjuntos de datos:

- ◇ DEMGBP: contiene observaciones diarias del Deutschmark/British Pound Foreign Exchange Rate. El periodo de muestra es de Enero 2 de 1984 a Diciembre 31 de 1991, para un total de 1975 observaciones. Realizamos 1000 simulaciones. De acuerdo a la Figura 4.5, en el límite, el índice extremo calculado sobre los modelos GARCH es mayor que los homólogos en la serie real. Esto implicaría una subestimación del tamaño medio de los conglomerados en esta serie.

- ◇ JPYUSD: contiene observaciones diarias del Japan Yen/United States Dollar Foreign Exchange Rate. El periodo de muestra es de Febrero 1 de 1977 a Mayo 1 de 2006, para un total de 7344 observaciones. Realizamos 1000 simulaciones. De acuerdo a la Figura 4.6, en el límite, el índice extremo calculado sobre el modelo GARCH(1,1) es mayor que los homólogos en la serie real. Esto implicaría una subestimación del tamaño medio de los conglomerados en esta serie. Esta serie fue estudiada por Mikosch y Starica [28], en aquel estudio obtiene un resultado contrario

al nuestro. Sin embargo, tal conducta puede obedecer a la escala de medición tomada, los autores mencionados consideraron una escala de tiempo más pequeña a mediciones diarias. En un contexto en donde nos interese estudiar la serie diaria, nuestro resultados parecen sensatos.

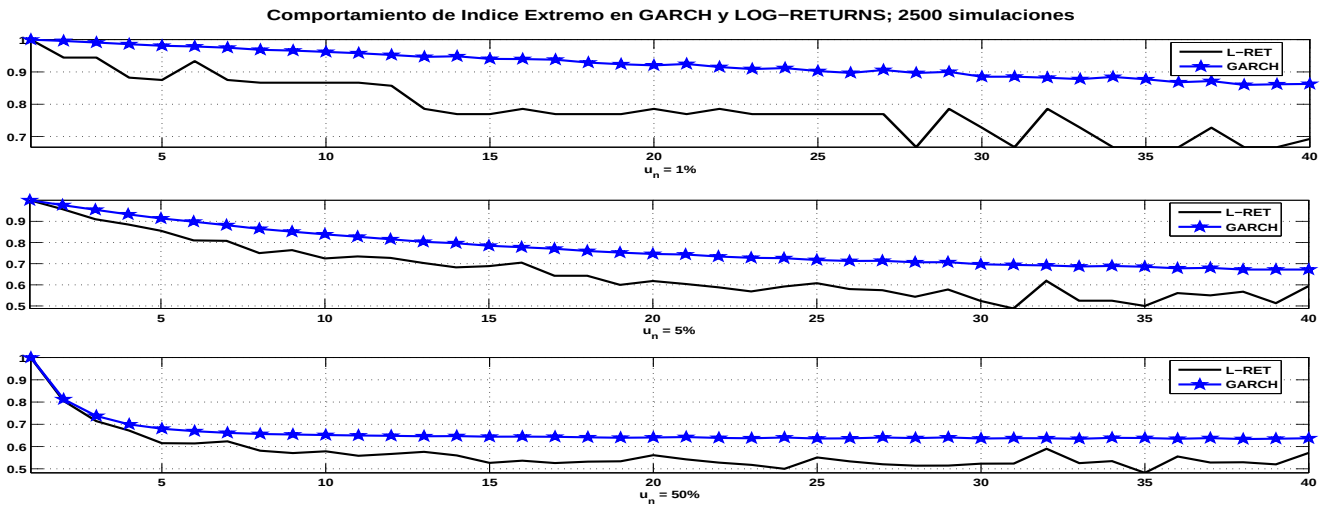


Figura 4.5: De arriba a abajo: umbrales de 1 %, 5 % y 50 %. La trayectoria marcada con estrellas responde al comportamiento del índice extremo de los GARCH(1,1) ajustados a la serie DEMGBP. La línea sin marcar refiere al comportamiento del índice extremo calculado sobre la serie original. En todos los puntos en los que se calcula el estimador, los estimados sobre el GARCH son mayores que los estimado sobre la serie original. De acuerdo al estudio de la sección anterior, utilizar un número de bloques en el conjunto $\{25, \ldots, 50\}$ presentará una convergencia adecuada. Aquí usamos 40 bloques para calcular los estimados.

4.3. Distribución del Índice Extremo

Para finalizar este capítulo, mostramos una serie de gráficas obtenidas a partir de las simulaciones de la sección anterior. Estas gráficas son los histogramas de los estimadores del índice extremo para diversos tamaños de bloques.

La Figura 4.8 presenta un estimador de la densidad del índice extremo. Básicamente, tomamos las 2500 simulaciones presentadas en la sección previa y fijamos un tamaño de bloque específico, digamos $k_n = 55$ y graficamos el histograma de esa población de estimadores. De acuerdo a los coeficientes de kurtosis y sesgo obtenidos, aún para el nivel más bajo, el 1 %, tenemos una densidad de colas tan livianas como la cola de la normal estándar y prácticamente simétrica. De acuerdo a estas gráficas,

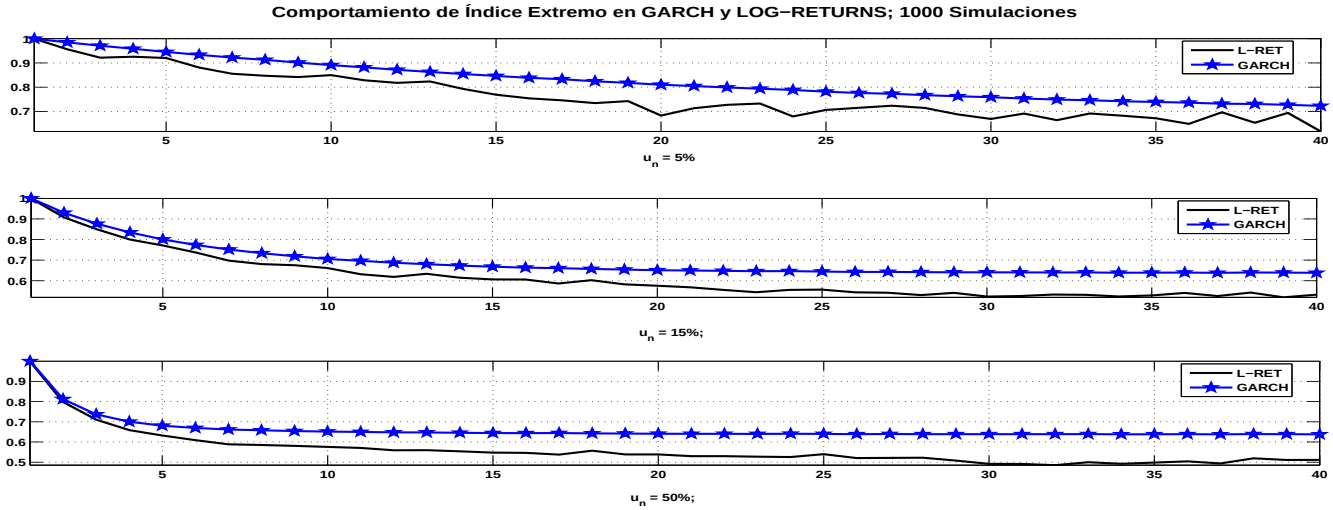


Figura 4.6: De arriba a abajo: umbrales de 5 %, 15 % y 50 %. De nueva cuenta, es evidente que los estimados del índice extremo calculados sobre los correspondientes GARCH(1,1) ajustados a la serie JPYUSD son mayores a los calculado sobre la serie original. De acuerdo al estudio de la sección anterior, utilizar un número de bloques en el conjunto $\{25, \ldots, 50\}$ presentará una convergencia adecuada. Aquí usamos 40 bloques para calcular los estimados.

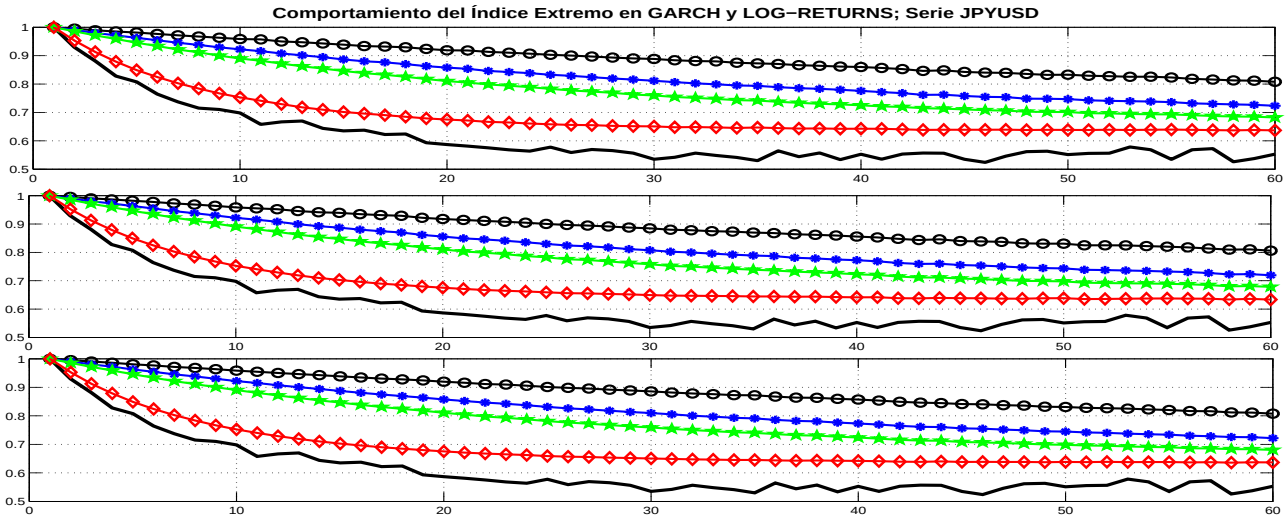


Figura 4.7: La figura reporta el comportamiento del estimador del índice extremo por el método de Olmo, con $k_n = 60$, niveles del 1 % ($\odot$), 2 % ($*$), 3 % ($\star$) y 10 % ($\diamond$). La línea sin marcar presenta el mismo estudio pero calculando el índice de la serie original. De arriba a abajo: 500 simulaciones, 1000 simulaciones y 2500 simulaciones por punto. Al igual que en la serie anterior, es claro que aún al nivel 10 %, el estimador del índice extremo sobre los modelos GARCH(1,1) ajustados, es mayor que el correspondiente estimador en la serie original.

la función de densidad del estimador del índice extremo por el método de niveles moderadamente altos, es de colas ligeras, simétrica alrededor del punto 0.6451.

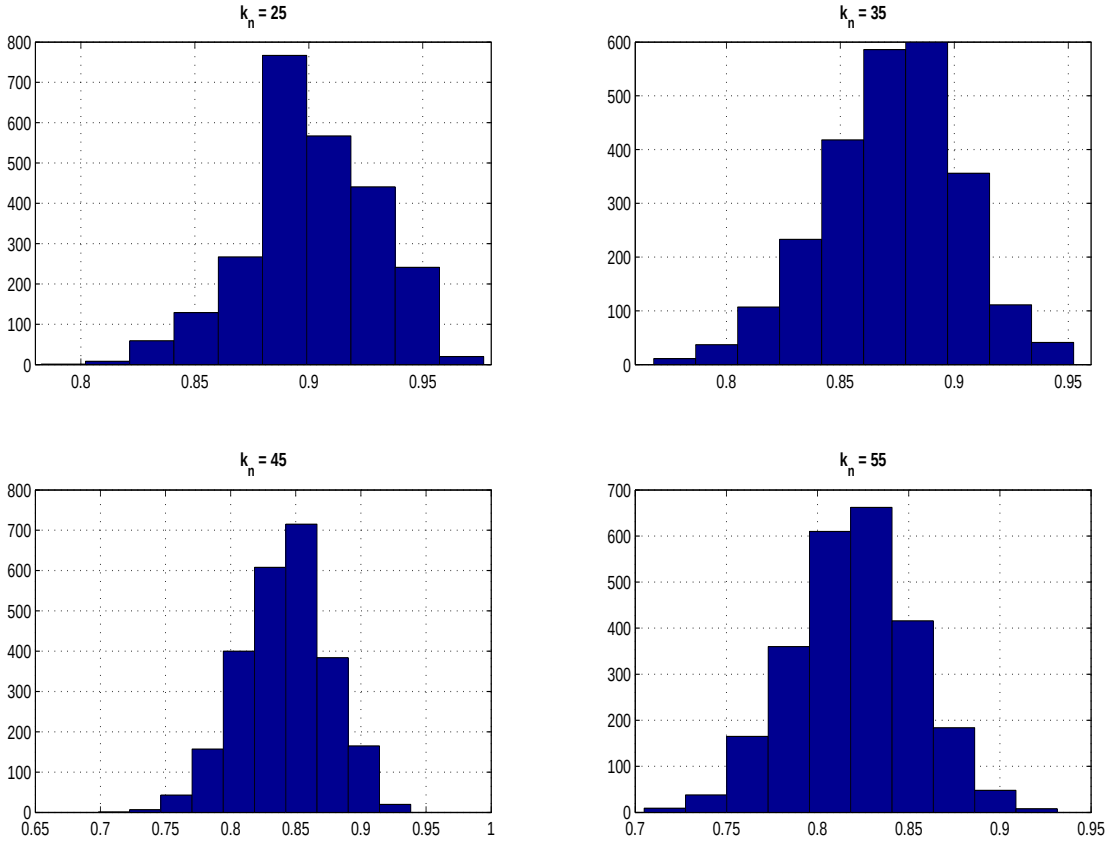


Figura 4.8: $k_n = 25$: kurtosis = 3.0179, sesgo = -0.3281. $k_n = 35$: kurtosis = 2.9354, sesgo = -0.2251. $k_n = 45$: kurtosis = 3.0387, sesgo. $k_n = 55$: kurtosis = 2.9590, sesgo = -0.0634

Bibliografía

[1] Barndorff-Nielsen O.E. y Shepard N. (2002). *Econometric analysis of realized olatility and its use in estimating stochastic volatility models.* Journal of Royal Statistical Society. **64** Part 2. 253-280.

[2] Bera A. K. y Higgins M. L. (1992). *A survey of Arch Models: Properties, estimation and testing.* Working Paper No. 92-0117. College of Commerce and Business Administration. University of Illinois at Urbana-Champaign.

[3] Berkes I., Horváth L. y Kokoszka P. (2001) *GARCH processes: Structure and estimation.* Preprint University of Utah.

[4] Brandt A. (1986). *The stochastic equation* $Y_{n+1} = A_n Y_n + B_n$ *with stationary coefficients.* Adv. in App. Probab. **18**. 211–220.

[5] Bollerslev T. (1986). *Generalized autoregressive conditional heteroscedasticity.* Journal of Econometrics. **31**. 307–327.

[6] Chan N. H. (2004). *Time Series: Applications to finance.* Wiley-Interscience.

[7] Chernick M.R. (1981). *A limit theorem for the maximum of autoregressive processes with uniform marginal distribution.* Annals of Probability. **9**. 145–149.

[8] Cont R. (2001). *Empirical properties of asset returns: Stylized facts and statistical issues.* Quant. Finance. **1**. 223-236.

[9] Cont R. (2004). *Long range dependence in financial markets.* –.

[10] Cox D. R. (1981). *Statistical analysis of time series: Some recent developments.* Scandinavian Journal of Statistics. **8**. 93–115.

[11] Davis R. A. and Mikosch T. (2001). *Point process convergence of stochastic volatility processes with applications to sample autocorrelation.* J. Appl. Probab. **38A**. 103–114.

[12] Engle R. F. (1982). *Autoregressive conditionally heteroscedasticity with estimates of the variance of U.K. inflation.* Econometrica. **50**. 987–1008.

[13] Geweke J. (1986). *Modelling the persistence of conditional variances: Comment.* Econometric Reviews. **5**. 57–61.

[14] Granger C. W. J.; Andersen A. (1978). *On the invertibility of time series models.* Stochastic Process. Appl. **8**(1). 87–92.

[15] Granger C. W. J. (2001). *Macroeconometrics – Past and future.* J. Econometrics. **100**. 17–19.

[16] Granger C. W. J. (2000). *Current perspectives on long memory processes.* Academia Economic Papers. **28**. 1. 1–16.

[17] Hall P. and Yao Q. (2003). *Inference in ARCH and GARCH Models with heavy-tailed errors.* Econometrica. **71**. 1. 285–317.

[18] Higgins M. and Bera A. (1992). *A class of nonlinear ARCH Models.* International economic review, forthcoming.

[19] Hsing T. (1988). *On the extreme order statistics for a stationary sequence.* Stoch. Process. Appl. **29**. 155–169.

[20] Hsing T. (1993). *Extremal index estimation for a weakly Dependent stationary sequence.* Ann. Statist. **21**. 4. 2043–2071.

[21] Kallenberg O. (1976). *Random measures.* Akademie-Verlag Berlin. Academic Press New York.

[22] Leadbetter M. (1983). *Extremes and local dependence in stationary sequences.* Z. Wahrscheinlichkeitstheorie verw. Gebiete. **65**. 291–306.

[23] Leadbetter M., Lindgren G. and Rootzen H. (1983). *Extremes and related properties of random sequences and processes.* Springer-Verlag. New York.

[24] Loynes R. M. (1965). *Extreme values in uniformly mixing stationary stochastic processes.* Ann. Math. Statist. **36**. 3. 993–999.

[25] Mandelbrot B. (1963). *The derivation of certain speculative prices.* Journal of Business **36**. 392–417.

[26] Marshall A. W. and Olkin I. (1983). *Domains of attraction of multivariate extreme value distributions.* Ann. Probab. **11**. 1. 168–177.

[27] Milhoj A. (1987). *A multiplicative parameterization of ARCH Models.* Research Report. Institute of Statistics. University of Copenhagen.

[28] Mikosch T. and Starica C. (2000). *Limit Theory for the sample autocorrelations and extremes of a GARCH(1,1) process.* Ann. Statist. **28**. 5. 1427–1451.

[29] Nelson D. B. (1991). *Stationarity and persistence in the GARCH(1,1) Model.* Econometric Theory. **6**. 318–334.

[30] Nelson D. B. (1991). *Conditional heteroscedasticity in asset returns: A new approach.* Econometrica. **59**. 347–370.

[31] Olmo J. (2005). *Extreme value theory in risk management.* Ph. D. dissertation. Department of Economics. Universidad Carlos III, Madrid.

[32] Olmo J. (2006). *A new family of estimators for the extremal index.* Working Paper. Department of Economics. City University, London.

[33] Pantula S. G. (1986). *Modelling the persistence of conditional variances: Comment.* Econometric Reviews. **5**. 79–97.

[34] Resnick S. (1987). *Extreme values, regular variation and point processes.* Springer-Verlag. New York.

[35] Shepard N. (1996). *Statistical aspects of ARCH and stochastic volatility.* Capítulo 1 en Time Series Models In econometrics, finance and other fields. Editado por D.R. Cox, D.V. Hinkley y O.E. Barndorff-Nielsen. Chapman&Hall.

[36] Smith R. y Weissman I. (1994). *Estimating the extremal index.* Journal of the Royal Statistical Society. Series B (Methodological). **56** No. 3. 515–528.

[37] Straumann D. and Mikosch T. (2003). *Quasi maximum likelihood estimation in heteroscedastic time series: A stochastic recurrence approach.*