

# Nuevos modelos y técnicas estadísticas para el estudio de datos financieros

Memòria presentada per aspirar  
al grau de Doctora en Matemàtiques.

Departament de Matemàtiques  
Universitat Autònoma de Barcelona

Director:  
Dr. Joan del Castillo Franquet

CERTIFICO que la present Memòria  
ha estat  
realitzada per na Jalila Daoudi,  
sota la direcció de la Dr. Joan del  
Castillo Franquet.

Bellaterra, maig del 2009

Signat: Dr. Joan del Castillo Franquet



إلى روح أبي الطاهرة

إلى أمي العظيمة



# Índice general

<b>Introducción</b>	<b>9</b>
<b>1. Distribuciones de colas pesadas. Fundamentos</b>	<b>13</b>
1.1. Definiciones básicas en supervivencia y fiabilidad . . . . .	14
1.2. Distribuciones subexponenciales . . . . .	15
1.3. Distribuciones de variación regular . . . . .	16
1.4. Distribuciones estables . . . . .	17
1.5. Teoría del valor extremo . . . . .	19
1.6. Media de una distribución truncada . . . . .	23
1.7. Relación entre las clases de colas pesadas y la teoría de la supervivencia .	26
<b>2. Mixtura de distribuciones normales truncadas</b>	<b>29</b>
2.1. Caracterización económica del modelo de intercambio . . . . .	30
2.2. Distribución normal truncada . . . . .	32
2.3. Modelo del intercambio bajo la hipótesis de normalidad . . . . .	34
2.4. Estimador de máxima verosimilitud (EMV). . . . .	39
2.5. Estudios empíricos . . . . .	42
<b>3. Estimación de la distribución de Pareto Generalizada</b>	<b>51</b>
3.1. Propiedades de la distribución de Pareto; coeficiente de variación . . . . .	52
3.2. Existencia del EMV en la distribución de Pareto . . . . .	53
3.3. Extensión al caso general . . . . .	57
3.4. Modelo alternativo TNP . . . . .	62
3.5. Ejemplo de aplicación . . . . .	64
<b>4. El coeficiente de variación en el estudio de las colas</b>	<b>71</b>
4.1. Coeficiente de variación . . . . .	72

## Índice general

---

4.2. Caracterización a través del coeficiente de variación condicionado . . . . .	75
4.3. Estimación de las colas. CV-Plot . . . . .	79
4.4. Análisis exploratorio . . . . .	82
4.5. Inferencia del peso de la cola y el coeficiente de variación . . . . .	87
4.6. Test de exponencialidad . . . . .	90
<b>5. Apéndices</b>	<b>95</b>

# Agraïments

It is said: “ All knowledge is, in the final, analysis, history.  
All sciences are, in the abstract, mathematics.”

I venture to add: “All methods of acquiring knowledge are  
essentially through statistics.”

C.R.Rao.

En primer lloc, voldria expressar el meu agraïment pel suport i dedicació al meu director Joan del Castillo que ha fet possible aquest treball. No només m’ha guiat i orientat en la realització de la tesi, sinó també m’ha transmès coneixements i destresa en insistir fins aconseguir els objectius. Un agraïment més personal el dirigeixo a la seva paciència i comprensió, així com el fet, de donar-me els ànims i la total seguretat que un necessita en tot moment.

No voldria oblidar el recolzament i la disponibilitat del professor Pere Puig des d’un principi donant respostes a tots els meus dubtes inclosos els més evidents. També, li vull agrair els seus consells que han estat sempre de gran utilitat.

Als meus companys de doctorat amb qui he compartit moltes estones enriquidores i divertides, tant per els coneixements que m’han transmès, com pel ambient que ha fet que em sentís com a casa. Als professors amb els qui he après tot impartint assignatures, sense oblidar de donar les gràcies al Departament de Matemàtiques per oferir-me la oportunitat de ser professora associada (falsa o verdadera, no importa el nom), i així poder aprendre donant classes.

Per últim, el darrer agraïment el dedico a la meva família que sempre m’encoratja, m’ajuda i em motiva. Ells no acaben d’entendre el que faig però, saben lo important

## Índice general

---

que és per a mi. Especialment, a la meva mare que ha sabut injectar-me energies, ànims des de la distància i moltes coses més que no cabien en aquestes pàgines. Com no vull deixar-me ningú, agraeixo també a tots a qui he atabalat amb les meves preguntes, dubtes, pensaments, suggeriments.....

“ Doubt grows with knowledge ”  
Johann Wolfgang von Goethe

# Introducción

Nuestra línea de investigación se ha desarrollado en el ámbito de la estadística aplicada a las finanzas. Nuestro objetivo es encontrar y analizar nuevos modelos estadísticos para ajustar los datos financieros y nuevas técnicas para estudiar el comportamiento de las colas. Una aplicación destacada de este trabajo es el estudio del riesgo operacional.

En los últimos años, la industria bancaria ha cambiado profundamente por los procesos de liberalización, innovación financiera y tecnológica. Esto, ha generado en las entidades financieras una evolución en el ámbito de la modelización de procesos para la medición y gestión del riesgo. El riesgo financiero se define como el impacto adverso en el rendimiento debido a diferentes fuentes de incertidumbre. En la actualidad, y desde una perspectiva avanzada de riesgos, se identificarían y se cuantificarían tres tipos de riesgo: riesgo de mercado, riesgo de crédito y riesgo operacional.

La administración del riesgo de mercado consiste en controlar los riesgos derivados de fluctuaciones en las tasas de interés, tipos de cambio, precios y otros factores de riesgo en los mercados de dinero, cambios y bursátil. Las transacciones con clientes crean exposición a riesgo de crédito, el cual se refiere al grado de incertidumbre de la solvencia de la contraparte para cumplir con los compromisos derivados de una operación de intermediación. La administración del riesgo de crédito es reducir la pérdida que pudiera resultar porque una contraparte no cumpla con los términos convenidos.

A diferencia de los anteriores, el riesgo operacional es un riesgo que no es producto de la toma de una posición de riesgo, tiene su origen en sucesos que no pueden ser adscritos a riesgo de mercado o a riesgo de crédito, se define como la pérdida potencial por deficiencias en los controles, por los errores en el procesamiento y almacenamiento de las operaciones o en la transmisión de información, así como por resoluciones admin-

istrativas y judiciales adversas, fraudes, robos o por factores externos.

El método más reciente para la cobertura de riesgo operacional es el método de medición avanzado (AMA) que consiste en la modelización de la distribución agregada de pérdidas (Loss Distribution Approach o LDA) que se ha utilizado con éxito en el ámbito de seguros. Una excelente referencia es el libro de [HK04]. Bajo el supuesto de que las severidades son independientes entre si e independientes de la frecuencia de los sucesos, la metodología LDA requiere la modelización por separado de la frecuencia y de la severidad. El capital regulatorio se calcula como el percentil de la distribución agregada de pérdidas para un nivel de probabilidad del 99,9 %.

Debido a la escasez de datos históricos internos, se utiliza la base externa de eventos operacionales acaecidos a entidades financieras diferentes a la propia entidad para tener información adicional que permite contrastar las evaluaciones y mediciones realizadas internamente, tales como, los eventos relevantes pero infrecuentes y la base cualitativa que corresponde a los cuestionarios de auto-evaluación de los expertos. Por ello, las distribuciones que mejor ajustan son las distribuciones de colas semi-pesadas o pesadas.

En el Capítulo uno, haremos una descripción de las distribuciones de colas pesadas que son distribuciones que tienen colas más pesadas que la distribución exponencial. Históricamente se han utilizado en el mundo de seguros, específicamente las distribuciones subexponenciales, las del máximo y las características de los dominios de atracción de Von Misses. En la última década, esta metodología se ha trasladado al mundo de las finanzas. Esto viene detallado en [EKM97].

De forma más amplia los mercados financieros están bajo una permanente tensión por la interacción entre la oferta y la demanda, lo cual implica fluctuaciones pronunciadas en los precios. El modelo clásico para estudiar la evolución de los precios es el modelo de Black Scholes que supone normalidad en la distribución de los precios. Los estudios empíricos, que presentan una curtosis elevada y valores extremos que no se pueden ajustar por la distribución normal, muestran que este modelo está lejos de ser adecuado.

Suponiendo normalidad de la distribución de la oferta y de la demanda, y si hay

tantas ofertas como demandas (mercado ideal), las transacciones siguen una mixtura de normales. En caso contrario, cuando no se aceptan de la misma manera las ofertas y las demandas las transacciones pueden seguir una mixtura de normales truncadas. El caso simétrico de esta distribución corresponde al trabajo de [DP04].

En el Capítulo dos, proponemos la mixtura de normales truncadas para ajustar los datos de tipo de cambio [CD09b]. Es un modelo muy apropiado dado que en la práctica nos permite estudiar la no-normalidad de los datos teniendo en cuenta la asimetría de los mismos. Para ello, primero desarrollamos las propiedades de la distribución propuesta y a continuación demostramos que la función de verosimilitud tiene un máximo único y que este máximo depende del coeficiente de variación.

El enfoque basado en la modelización de la distribución de severidad mediante distribuciones de colas semi pesadas tales que la distribución lognormal, la inversa gaussiana y la mixtura de normales proporcionan estimaciones robustas estables del capital regulatorio, es decir, las cifras de capital entre dos periodos sólo pueden variar por su exposición al riesgo. El enfoque basado en la teoría de valor extremo que se caracteriza por el ajuste de los valores que superan un determinado umbral con la distribución de Pareto generalizada es de mucha importancia dado que las entidades se basan únicamente sobre las pérdidas elevadas, aunque la elección de manera eficiente del umbral a partir del cual se realiza el ajuste a una distribución de Pareto es crucial para obtener valores estables de las medidas de riesgo [CS06].

Varios autores estudiaron la estimación de la distribución de Pareto generalizada mediante el método de máxima verosimilitud. El comportamiento asintótico del estimador de máxima verosimilitud ha sido estudiado en [DS90] y en [Smi85]. No obstante, los métodos numéricos no siempre tienen soluciones sobretodo cuando las muestras son pequeñas, por ello se han propuesto otros métodos tales como el método de los momentos ponderados que sólo se puede utilizar cuando el momento de orden dos es finito [HW87], y el método que consiste en estimar los parámetros a través de los estadísticos de orden [CH97].

En el Capítulo tres, explicaremos los problemas destacados de la no convergencia en algunos casos de la función *profile* verosimilitud de la distribución de Pareto general-

izada [DS90]. Luego, demostraremos que la función *profile* verosimilitud se caracteriza por el coeficiente de variación empírico. Por último, probaremos que en el caso de la distribución de Pareto, la función de verosimilitud tiene un máximo global cuando el coeficiente de variación empírico es mayor que uno [CD09a]. Por otro lado, ilustramos con un ejemplo que la función de verosimilitud de la distribución de Pareto generalizada puede no tener soluciones.

En el Capítulo cuatro, utilizamos la caracterización de la distribución de Pareto generalizada a través del coeficiente de variación condicionado para desarrollar una metodología previa y complementaria a los estudios paramétricos y contrastar el modelo desde un punto de vista empírico. Es un método alternativo a los métodos clásicos ME-Plot y el Hill-Plot para estudiar las colas. Además nos permite encontrar de manera eficiente el umbral a partir del cual podemos ajustar los datos por una distribución de Pareto y estimar el parámetro del peso de la cola. Por otro lado, proponemos un test de exponencialidad contra las alternativas de cola Pareto.

Una de las dificultades de la distribución de Pareto generalizada es que al incluir distribuciones de soporte acotado, existen problemas de convergencia de los estimadores. Además las ecuaciones de verosimilitud para muestras pequeñas pueden no tener soluciones. Por ello, proponemos la distribución TNP que es la unión de la normal truncada, la distribución exponencial y la distribución de Pareto como alternativa a la distribución GPD.

Los resultados recogidos en los Capítulos 2 y 3 han sido publicadas en *Journal of Statistical Planning and Inference*, y *Statistics and Probability Letters*, respectivamente. Los textos en inglés se encuentran en el Apéndice A y en el Apéndice B. Este trabajo nos permite seguir en la misma línea de investigación dada la novedad del tema y también nos abre nuevos horizontes en el mundo de finanzas.

# Capítulo 1

## Distribuciones de colas pesadas.

### Fundamentos

Las fluctuaciones pronunciadas de precios conducen a serias inestabilidades en los mercados financieros. Estas perturbaciones llevan a problemas en la gestión del riesgo. En este contexto, es necesaria la modelización del comportamiento de estos precios que alcanzan valores extremos. La distribución normal no determina con suficiente precisión dicho comportamiento, lo cual obliga a recurrir a otro tipo de distribuciones de colas pesadas o semi pesadas. La creciente disponibilidad de datos financieros hace que un número cada vez mayor de investigadores se dedique al análisis empírico de las múltiples cuestiones teóricas de interés relativas a los mercados financieros.

Los modelos de colas pesadas son una de las clases más importantes de las familias de distribuciones en Estadística. Son apropiadas cuando ocurren valores extremos con probabilidades pequeñas que no podemos ignorar.

Una clase especial de estas distribuciones son las distribuciones subexponenciales. La primera noción aparece en [Chi64]. El nombre proviene del hecho que estas distribuciones presentan colas más pesadas que la distribución exponencial. Esta clase es adecuada para modelar fluctuaciones muy grandes. En la práctica, las distribuciones subexponenciales son muy útiles en seguros, especialmente en la modelización de las probabilidades de ruina. Otra clase de importancia son las distribuciones de variación regular que son mucho más manejables debido al comportamiento asintótico.

## 1.1. Definiciones básicas en supervivencia y fiabilidad

En primer lugar, introduciremos algunas definiciones utilizadas en el análisis de la supervivencia véase [BP81].

Sea  $X$  una variable aleatoria positiva con función de densidad  $f(x)$  y función de distribución  $F(x)$ .

La función de *supervivencia* se define como:

$$\bar{F}(x) = 1 - F(x) = P\{X \geq x\} = \int_x^{\infty} f(t) dt.$$

La función de *riesgo* se define como:

$$r(x) = \frac{f(x)}{\bar{F}(x)},$$

puede interpretarse como la probabilidad de fallo durante un intervalo de tiempo muy pequeño suponiendo que la variable ha sobrevivido hasta el comienzo del intervalo.

$$r(t) = \lim_{\delta(t) \rightarrow 0} P(X \in [t, t + \delta(t)] | X \geq t) / \delta(t).$$

A esta función, asociamos la función de riesgo acumulada que definimos como,

la función de *riesgo acumulado* se define como:

$$R(x) = \int_0^x r(t) dt.$$

Las funciones  $f(x)$ ,  $\bar{F}(x)$ ,  $r(x)$ ,  $R(x)$  caracterizan de manera equivalente la distribución de  $X$ , conociendo una de ellas, se puede deducir fácilmente las otras. La función de riesgo acumulada esta relacionada con la función de supervivencia de la manera siguiente:

$$R(x) = \int_0^x r(t) dt = \int_0^x \frac{f(t)}{\bar{F}(t)} dt = - \int_0^x \frac{d}{du} \log \bar{F}(t) = - \log(\bar{F}(x)).$$

Cuando la función de riesgo es monótona, podemos distinguir dos clases;

Diremos que una distribución o una variable aleatoria es de la *clase IFR* (Increasing failure rate) si su función de riesgo  $r(x)$  es creciente para todo  $t \geq 0$ . Análogamente diremos que es de la *clase DFR* (Decreasing failure rate) si su función de riesgo  $r(x)$  es decreciente.

Más adelante, relacionaremos los modelos DFR con las clases de colas pesadas.

## 1.2. Distribuciones subexponenciales

Una variable aleatoria  $X$  se llama *subexponencial* cuando la probabilidad de la suma y del máximo son comparables. Concretamente; sea  $\{X_1, X_2, \dots, X_n\}$  una muestra de  $X$  y sea  $x$  tal que;  $F(x) < 1$ , se dice que  $X$  es subexponencial si,

$$\lim_{x \rightarrow \infty} \frac{P\{X_1 + X_2 + \dots + X_n > x\}}{P\{\max(X_1, X_2, \dots, X_n) > x\}} = 1 \quad n \geq 2$$

Embrechts [Emb80] demostró que es suficiente probarlo para un  $n$  mayor o igual a dos.

Las colas de las distribuciones de la suma y del máximo tienen asintóticamente el mismo orden, esto indica la gran influencia de la observación máxima en el total de las observaciones.

A continuación, presentamos algunas propiedades de esta distribución que serán útiles más adelante.

**Proposición 1.1.** *Supongamos que  $F$  y  $G$  son dos funciones de distribución. Si  $F$  sigue una distribución subexponencial y  $F, G$  son comparables en el infinito, es decir;  $\lim_{x \rightarrow \infty} \bar{F}(x)/\bar{G}(x) = c$ , donde  $c$  es una constante positiva. Entonces  $G$  sigue una distribución subexponencial.*

*Demostración.* véase [EKM97] □

**Teorema 1.1.** *Una distribución subexponencial:*

- a) *Tiende más lentamente al infinito que la distribución exponencial, es decir  $\lim_{x \rightarrow \infty} \bar{F}(x)/e^{-\lambda x} = \infty$  cuando  $x$  tiende al infinito para cualquiera  $\lambda$  positiva.*
- b) *No tiene función generatriz de momentos.*

*Demostración.* véase [GK98] □

Por la propiedad (a), una distribución que tiene colas menos pesadas que la distribución exponencial no pertenece a las distribuciones subexponenciales. Por ejemplo, la distribución normal que es una distribución con colas ligeras.

En general, utilizando la definición no resulta fácil comprobar que una distribución es subexponencial, por lo tanto, una propiedad de gran interés es la siguiente.

**Observación 1.1.** Sea  $X$  una variable aleatoria de función de distribución  $F$  y función de densidad  $f$ , y tal que, la función de riesgo  $r$  es decreciente hacia 0. Entonces, si la función  $e^{xr(x)}f(x)$  es integrable en  $(0, \infty)$ ,  $F$  es una distribución subexponencial.

### 1.3. Distribuciones de variación regular

Las distribuciones de variación regular, son mucho más manejables que las distribuciones subexponenciales debido al comportamiento asintótico, una referencia, es [BGT89].

Diremos que  $L$  es de *variación lenta* si para todo  $\gamma > 0$

$$\lim_{x \rightarrow \infty} \frac{L(\gamma x)}{L(x)} = 1. \quad (1.1)$$

**Ejemplo 1.3.1.** Podemos coger como función de variación lenta una constante positiva o  $L(x) = \log(x)$ .

Diremos que una función de distribución,  $F$ , es de *variación regular* de índice  $-\alpha$  si:

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(xt)}{\bar{F}(x)} = t^{-\alpha}. \quad t > 0$$

De forma equivalente  $F$  se puede expresar como

$$\bar{F}(x) = P\{X > x\} = x^{-\alpha}L(x), \quad \alpha > 0$$

donde  $L$  es una función de variación lenta.

**Proposición 1.2.** Supongamos que  $F$  es tal que:  $F(x) < 1$  para cualquiera  $x$  positiva. Sea  $X$  una variable aleatoria de distribución de variación regular de índice  $-\alpha$ , entonces  $X$  tiene momentos de orden  $\beta$  cuando  $\beta$  es inferior a  $\alpha$ , en caso contrario los momentos son infinitos.

*Demostración.* véase [EKM97] □

Observamos que cuanto más pequeño es el orden de los momentos, las colas son más pesadas, esto se explica por la desigualdad de Chebychev.

Dada una variable aleatoria  $X$ , si el momento de orden  $\alpha$  existe, la desigualdad de Chebychev dice que

$$\bar{F}(t) = P\{X \geq t\} \leq \frac{E(X^\alpha)}{t^\alpha}.$$

**Proposición 1.3.** Sean  $F_1$  y  $F_2$  dos funciones de variación regular de índice  $-\alpha$  tal que,  $\bar{F}_i(x) = x^{-\alpha}L_i(x)$ ,  $i = 1, 2$  donde  $L_i$  es una función de variación lenta. Entonces la convolución  $F_1 * F_2$  es de variación regular de índice  $-\alpha$ , es decir,  $1 - \overline{F_1 * F_2}(x) \sim x^{-\alpha}(L_1(x) + L_2(x))$ .

*Demostración.* véase [Fel81] □

Esta proposición reemplaza al teorema central del límite al proporcionar información relativa a las colas.

**Proposición 1.4.** Las distribuciones de variación regular son subexponenciales.

*Demostración.* véase [Fel81] □

Una idea de la demostración es la siguiente, por un lado,

$$\begin{aligned} P\{\max\{X_1, X_2, \dots, X_n > x\}\} &= 1 - F(x)^n \\ &= \bar{F}(x) \sum_{k=0}^{n-1} F^k(x). \end{aligned}$$

Cuando  $x$  tiende a infinito, la última expresión es asintóticamente equivalente a  $n\bar{F}(x)$ .

Por otro lado, la función de supervivencia de la suma de las variables aleatorias iid es la convolución de la función de supervivencia de cada una de ellas [Fel81], y como las familias de variación regular son cerradas bajo convolución. Entonces, la función de supervivencia de la suma es igual a  $n\bar{F}(x)$ . Por lo tanto, la probabilidad de la suma y del máximo son comparables y el límite cuando  $n$  tiende a infinito da uno.

Como caso particular de las distribuciones de variación regular, podemos citar las distribuciones estables.

## 1.4. Distribuciones estables

Las distribuciones estables tienen un gran importancia en la teoría del valor extremo. El hecho de tener varianzas infinitas no las impide ser útiles en los seguros. Mandelbrot [Man63] propuso esta familia de distribuciones como alternativa a la distribución normal para estimar el peso de la cola de la distribución empírica. Una recomendable referencia es el libro de [Zol86].

Supongamos que  $X_1, \dots, X_n$  son variables aleatorias idénticamente distribuidas de función de distribución  $F$ .

$F$  es *estable* si existe  $a_n > 0$  y  $b_n \in \mathbb{R}$ , tal que;

$$\forall n; \quad a_n(X_1 + X_2 + \dots + X_n) + b_n \xrightarrow{d} X_1 \quad (1.2)$$

donde  $\xrightarrow{d}$  es la convergencia en distribución.

A parte de unos casos particulares, la densidad de una distribución estable es desconocida analíticamente, solo se conoce la expresión de su función característica.

La función característica de una función estable se expresa como,

$$\Phi_X(t) = Ee^{iXt} = e^{i\gamma t - c|t|^\alpha(1 - i\beta \operatorname{sign}(t)z(t, \alpha))} \quad t \in \mathbb{R}$$

donde  $\gamma$  es una constante,  $c > 0$ ,  $\alpha \in (0, 2]$ ,  $\beta \in [-1, 1]$  y

$$z(t, \alpha) = \begin{cases} \tan(\frac{\pi\alpha}{2}) & \text{si } \alpha \neq 1 \\ -\frac{2}{\pi} \ln |t| & \text{si } \alpha = 1 \end{cases}$$

$\alpha$  es el parámetro más importante porque determina el peso de la cola, y por tanto el orden de los momentos y el comportamiento asintótico de la suma (teorema central del límite generalizado).

$\alpha = 2, \beta = 0$  es el caso de la normal y  $\alpha = 1, \beta = 0$  es la distribución de Cauchy.

**Definición 1.1. (Dominio de atracción).**  $F$  pertenece al dominio de atracción de una distribución estable  $S$  de parámetro  $\alpha$  si para cualquiera  $X_1, X_2, \dots, X_n$  variables aleatorias idénticamente distribuidas de función de distribución  $F$  existen unas constantes

$a_n > 0, b_n \in \mathbb{R}$  tal que;

$$a_n(X_1 + X_2 + \dots + X_n) - b_n \xrightarrow{d} Y, \quad (1.3)$$

donde  $Y$  es una variable aleatoria de función de distribución  $S$  de parámetro  $\alpha$

**Observación 1.2.** Cada función estable pertenece a su dominio de atracción.

A partir de la ecuación 1.2, si  $X$  es una variable aleatoria que sigue una distribución estable de parámetro  $\alpha$ , entonces bajo normalización y centralización, la suma tiene la misma distribución con el mismo  $\alpha$ .

Diremos que la variable aleatoria  $X$  satisface el teorema del límite generalizado con límite  $Y$  si se cumple la ecuación 1.3.

**Observación 1.3.** A partir de la ecuación 1.3, si la función de distribución de  $X_k, k = 1, \dots, n$  pertenece al dominio de atracción de una distribución estable  $S$  de parámetro  $\alpha$ , entonces  $a_n(X_1 + \dots + X_n) - b_n$  es asintóticamente estable de parámetro  $\alpha$ .

**Proposición 1.5.** Si  $F$  pertenece al dominio de atracción de una distribución estable de parámetro  $\alpha$ , entonces  $F$  es de variación regular de índice  $-\alpha$ .

Este resultado viene expresado en la siguiente proposición.

**Proposición 1.6.**  $F$  pertenece al dominio de atracción de una distribución estable de parámetro  $\alpha$  para  $\alpha < 2$  sii

$$F(-x) = \frac{c_1 + o(1)}{x^\alpha} L(x) \quad \bar{F}(x) = \frac{c_2 + o(1)}{x^\alpha} L(x)$$

donde  $L$  es una función de variación lenta y  $c_1, c_2$  son constantes positivas tal que;  $c_1 + c_2 > 0$ .

Las distribuciones estables son de variación regular, puesto que cada función estable pertenece a su dominio de atracción.

## 1.5. Teoría del valor extremo

La teoría del valor extremo es una rama de la teoría de la Probabilidad que se centra en explicar, por medio de modelos de distribuciones, los sucesos o casos raros, es decir

poco probables, pero que de ocurrir pueden ocasionar grandes pérdidas. Esta teoría ha sido muy utilizada en hidrología y aplicada al diseño de sistemas de protección del medio ambiente [Col01]. Recientemente ha sido utilizado en seguros y finanzas.

**Teorema 1.2. (Fisher-Tipett).** *Sea  $X$  una variable aleatoria, y sea  $X_i, i = 1, \dots, n$  una muestra de  $X$ . Si existen  $c_n > 0$  y  $d_n \in \mathbb{R}$  tal que*

$$c_n^{-1}(\max(X_1, X_2, X_3, \dots, X_n) - d_n) \xrightarrow{d} H.$$

*Entonces,  $H$  es una de las tres distribuciones siguientes:*

$$\begin{aligned} \Phi_\alpha(x) &= \exp(-x^{-\alpha}) \quad \text{si } x > 0 \\ \Psi_\alpha(x) &= \begin{cases} \exp(-(-x)^{-\alpha}) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \quad \alpha > 0 \\ \Lambda(x) &= \exp(-\exp^{-x}) \quad \text{si } x \in \mathbb{R} \end{aligned}$$

$\Phi_\alpha$  corresponde a la distribución de Fréchet,  $\Psi_\alpha$  corresponde a la distribución de Weibull y  $\Lambda$  corresponde a la distribución de Gumbel.

En primer lugar, según la teoría de los dominios de atracción se podría pensar en el método clásico basado en escoger uno de los tres límites y luego estimar los parámetros. Distinguiremos entre escoger el modelo y estimarlo o hacerlo todo a la vez. Hacer una selección previa del modelo implica llegar a conclusiones sensibles a dicho modelo, entonces, es más adecuado unificar primero los modelos.

Von Misses fue el primero en proponer la unificación de los tres modelos observando lo siguiente,

$$(x - a)/b = 1 + (x - \mu)/(\alpha\sigma),$$

entonces, cuando  $\alpha \rightarrow \infty$

$$\left(\frac{x - a}{b}\right)^\alpha = \left(1 + \frac{1}{\alpha} \left(\frac{x - \mu}{\sigma}\right)\right)^\alpha \rightarrow \exp\left(\frac{x - \mu}{\sigma}\right).$$

Por lo tanto, la distribución de Fréchet como la distribución de Weibull tienden a la distribución de Gumbel cuando  $\alpha$  tiende a  $\infty$ .

Para simplificar, tomamos  $\xi = 1/\alpha$  de tal manera que unificamos los tres modelos en la distribución generalizada de los valores extremos definida por la siguiente expresión,

$$G(z) = \exp\left(-\left(1 + \xi\left(\frac{z - \mu}{\sigma}\right)_+\right)^{-1/\xi}\right),$$

si  $\xi > 0$ , el modelo corresponde a la distribución de Fréchet, el modelo corresponde a la Gumbel si  $\xi = 0$  y si  $\xi < 0$  el modelo corresponde a la distribución de Weibull.

Los parámetros satisfacen  $-\infty < \mu < \infty$ ,  $\sigma > 0$  y  $-\infty < \xi < \infty$ , respecto al soporte  $1 + \xi\frac{z-\mu}{\sigma} > 0$ .

**Definición 1.2.** *La función de densidad de la función generalizada de valores extremos viene definida por,*

$$f(z) = G(z) \frac{1}{\sigma} \left(1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right)^{\frac{-1+\xi}{\xi}}$$

siempre que  $1 + \xi\left(\frac{z-\mu}{\sigma}\right) > 0$ .

**Proposición 1.7.**  *$F$  pertenece al dominio de atracción de Fréchet de parámetro  $\alpha$ , si y sólo si  $F$  es de variación regular de índice  $-\alpha$ .*

*Demostración.* véase [EKM97]. □

Este resultado implica que el dominio de atracción del máximo de la distribución de Fréchet viene dado por funciones de distribución cuya cola decrece como una función potencial.

Este tipo de distribuciones son las más importantes en seguros ya que las cuantías grandes al tener probabilidades significativas pueden ser una amenaza para la solvencia de la empresa.

Algunos ejemplos de este tipo de distribuciones; la distribución de Pareto que estudiaremos con más detalles más adelante y las distribuciones estables con exponente  $\alpha < 2$ .

**Condición de Von Mises.** Sea  $F$  una función de distribución con función de densidad  $f$  definida en un intervalo finito  $(z, x_F)$ . Si

$$\lim_{x \rightarrow x_F} \frac{(x_F - x)f(x)}{\bar{F}(x)} = \alpha > 0,$$

entonces,  $F \in MDA(\Psi_\alpha)$

Es importante resaltar que las distribuciones de MDA ( $\psi_\alpha$ ) tienen el soporte acotado a la derecha y esto no es adecuado para la modelización de los datos.

El conjunto de distribuciones, pertenecientes al dominio de atracción del máximo de la distribución de Gumbel, está formado por distribuciones cuya cola derecha es de forma exponencial, como ejemplo, podemos citar la distribución gamma, la normal y la log-normal.

Las distribuciones que habitualmente se utilizan para modelizar datos en finanzas y seguros pertenecen a los dominios de atracción del máximo de las distribuciones de Fréchet y de Gumbell como indican [HK04] y [Col01]

**Proposición 1.8.** *Si una distribución pertenece al dominio de atracción de Fréchet (Gumbel), la distribución del exceso respecto de un umbral  $u$  suficientemente alto es una distribución de Pareto generalizada (exponencial) respectivamente.*

*Demostración.* véase [Col01]

Un esquema de la demostración es el siguiente, para  $n$  grande

$$F^n(x) \approx \exp \left\{ - \left( \frac{x - \mu}{\sigma} \right)^{-\alpha} \right\}, \quad \alpha > 0$$

$$\log(F(x)) \approx \frac{-1}{n} \left( \frac{x - \mu}{\sigma} \right)^{-\alpha},$$

como  $\log(F(x)) \approx -(1 - F(x))$  entonces,

$$P(X - u > y | X > u) = \left( 1 + \frac{y}{u - \mu} \right)^{-\alpha}.$$

En el caso de la distribución de Gumbell, para  $n$  grande

$$F^n(x) \approx \exp \left\{ \exp \left\{ - \left( \frac{x - \mu}{\sigma} \right) \right\} \right\},$$

utilizando la misma aproximación que antes, tendremos

$$P(X - u > y | X > u) = \exp \left\{ - \frac{y}{\sigma} \right\}.$$

□

El teorema que presentamos a continuación demuestra que si una variable aleatoria tiene una distribución que pertenece al dominio de atracción del máximo de una distribución del valor extremo  $G_\xi(x)$ , entonces la distribución condicionada de la variable aleatoria exceso respecto de un umbral elevado sigue una distribución de Pareto generalizada, con el mismo parámetro de forma  $\xi$  que  $G_\xi(x)$ .

**Teorema 1.3.** *Sea  $X_n, n \in \mathbb{N}$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de distribución  $F$  y sea  $M_n = \max\{X_n\}$ . Supongamos que la distribución de  $M_n$  es tal que,*

$$P\{M_n \leq z\} \approx G_\xi(z) \quad \mu > 0$$

Entonces, la distribución condicionada converge a:

$$P((X - u) \leq y | X > u) \sim W_\xi(y) = 1 - \left(1 + \frac{\xi y}{\sigma'}\right)^{-1/\xi},$$

$$y > 0 \text{ y } 1 + \frac{\xi y}{\sigma'} > 0, \quad \sigma' = \sigma + \xi(\mu - u)$$

$W_\xi(y)$  se denomina distribución de Pareto Generalizada.

*Demostración.* véase [Col01] □

Es importante destacar que  $\sigma'$  depende del umbral  $u$  y como indican [EKM97], esto significa que en las aplicaciones, las estimaciones de  $\xi$  y de  $\sigma'$  dependen del umbral escogido.

Se puede comprobar que la distribución del valor extremo y la distribución de Pareto generalizada se relacionan de la siguiente manera.

$$W_\xi(x) = 1 + \log G_\xi(x).$$

El Teorema 1.3 implica que si el máximo tiene una distribución asintótica que pertenece a la familia del valor extremo,  $G_\xi$ , entonces el exceso sobre un determinado umbral tiene una distribución de Pareto generalizada. Además, los parámetros de la distribución de Pareto generalizada están unívocamente determinados por los parámetros de la distribución del valor extremo correspondientes.

El Teorema 1.3 permite estimar los parámetros utilizando los máximos muestrales o los valores que exceden un determinado umbral suficientemente elevado.

Por último, podemos concluir que al estudiar máximos, si el tamaño muestral es suficientemente grande, el parámetro  $\xi$  permanece estable en cambio los parámetros  $\mu$  y  $\sigma$  cambian pero mantienen la relación  $\sigma' = \sigma + \xi(\mu - u)$ .

## 1.6. Media de una distribución truncada

Sea  $X$  una variable aleatoria, la media de la distribución truncada en  $u$  es  $e(u) = E(X - u | X > u)$ .

**Proposición 1.9.** *Sea  $X$  una variable aleatoria de función de distribución de variación regular de índice  $-\alpha$  y función de variación lenta  $L$ , entonces el comportamiento de  $e(u)$  en el infinito es como  $u/(\alpha - 1)$ .*

*Demostración.* Tenemos que,

$$e(u) = \frac{1}{\overline{F}(u)} \int_u^\infty \overline{F}(x) dx = \frac{1}{\overline{F}(u)} \int_u^\infty x^{-\alpha} L(x) dx.$$

Aplicando el Teorema de Karámata, probamos que  $e(u)$  es asintóticamente equivalente a  $u/(\alpha - 1)$ . □

En las distribuciones de variación regular de índice  $-\alpha$ , la media de las distribuciones truncadas tiende a infinito cuando  $u$  tiende a infinito.

La varianza condicionada viene definida por

$$\sigma^2(u) = \text{Var}(X - u | X > u) = \int_0^\infty (x - u)^2 f_u(x) dx - (E(x - u | x > u))^2.$$

**Teorema 1.4. (Karámata).** *Sea  $L$  una función de variación lenta localmente acotada en  $[x_0, \infty)$  para  $x_0 \geq 0$ , entonces para  $\alpha > -1$*

$$\int_{x_0}^x t^\alpha L(t) dt \sim (\alpha + 1)^{-1} x^{\alpha+1} L(x) \quad x \rightarrow \infty$$

para  $\alpha < -1$

$$\int_x^\infty t^\alpha L(t) dt \sim -(\alpha + 1)^{-1} x^{\alpha+1} L(x) \quad x \rightarrow \infty$$

*Demostración.* véase [BGT89] □

A continuación, presentamos los métodos clásicos ME-Plot y Hill-Plot para estudiar el comportamiento de las colas.

### ME-Plot

La media de los excesos es un concepto muy importante en la teoría del valor extremo. Recientemente, esta cantidad comienza a utilizarse en finanzas al igual que el VaR. Es una medida del riesgo económico. Además, se reduce por la diversificación de carteras sin necesidad de linealidad. En cambio, el VaR necesita linealidad. El ME-Plot consiste en el gráfico  $\{X_{k,n}, e_n(X_{k,n})\}$  donde:

$$\begin{aligned} e_n(u) &= \int_u^\infty \overline{F}_n(x) dx / \overline{F}_n(u) \\ &= (\text{card} \Delta_n(u))^{-1} \sum_{i \in \Delta_n(u)} (X_i - u) \quad u \geq 0, \end{aligned}$$

donde:

$$\Delta_n(u) = \{i \mid i = 1, \dots, n, X_i > u\}$$

La media de los excesos empírica se utiliza para aproximar la correspondiente teórica pero no es estable en las colas. Por otro lado, la función exceso medio de una distribución de Pareto es lineal con pendiente positiva. La función empírica también lo es, lo que permite estimar el parámetro del peso de la cola mediante un ajuste por mínimos cuadrados.

### Hill-Plot

Sea  $x$  una variable aleatoria que verifica:

$$\overline{F}(x) = (u/x)^{-\alpha} \quad \text{para } x \geq u > 0,$$

donde  $\alpha$  es el parámetro del peso de la cola. Se supone que el umbral  $u$  es conocido, entonces, el estimador de máximo verosimilitud es,

$$\hat{\alpha}_n = \left( \frac{1}{n} \sum_{j=1}^n \ln(x_{j,n}) - \ln(u) \right)^{-1}.$$

Supongamos ahora que el umbral es el  $k$ -ésimo valor de la muestra ordenada.

Sea  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$  la muestra ordenada procedente de una población cuya función de distribución verifica:

$$\bar{F}(x) = Cx^{-\alpha}, \quad x > u.$$

y sea

$$K = \{i | X_{(n-i)} > u, i = 1 \dots n\}.$$

Se trata de obtener el estimador máximo verosímil del parámetro  $\alpha$  de la distribución de Pareto pero condicionado al suceso  $\{K = k\}$ . Por ello, se maximiza la función de densidad conjunta del  $k$ -ésimo estadístico de orden dada por

$$f(x_{(n-k)}, \dots, x_{(n)}) = \frac{n!}{(n-k)!} (1 - Cx_{(n-k)}^{-\alpha})^{n-k} C^k \alpha^k \prod_{i=1}^k X_{n-i+1}^{-(\alpha+1)}.$$

El valor de  $\alpha$  que hace máxima la expresión anterior es el estimador de HILL [Hil75] y viene dado por,

$$\left( \frac{1}{k} \sum_{j=1}^k \ln(X_{j,n}) - \ln(X_{k,n}) \right)^{-1},$$

donde  $k = k(n) \rightarrow \infty$ ,  $k$  es el punto a partir del cual se estabiliza el gráfico.

Se observa que el estimador de Hill tiene la misma forma que el estimador de máximo verosímil de la distribución de Pareto sólo que el umbral  $u$  determinista se sustituye por el umbral aleatorio  $x_{(n-k)}$ .

Como indican los autores en [EKM97], el estimador de Hill muestra grandes variaciones para valores de  $k$  pequeños y pequeñas variaciones para valores grandes de  $k$  que conducen a una estimación sesgada.

## 1.7. Relación entre las clases de colas pesadas y la teoría de la supervivencia

**Proposición 1.10.** *Sea  $x$  una variable aleatoria de función de densidad  $f(x)$ , De la caracterización de Karámata,  $f$  es de variación regular de exponente  $\alpha - 1$  si y sólo si*

$$\lim_{x \rightarrow \infty} x f(x) / \bar{F}(x) = \alpha. \tag{1.4}$$

*Demostración.* véase [EKM97] □

Diremos que una distribución o una variable aleatoria  $X$  es de las *Clases IVRL (DVRL)*. (Increasing o Decreasing failure rate en average) si la varianza condicionada de  $X$  es creciente (decreciente) respectivamente.

**Corolario 1.1.** *La función de riesgo de las distribuciones de colas pesadas decrecen asintóticamente como una hipérbola:*

$$r(x) \sim \frac{\alpha}{x}.$$

*Demostración.* Aplicando la Proposición 1.10 obtenemos el resultado. □

**Proposición 1.11.** *Para una variable aleatoria no negativa de función de distribución, cada una de las propiedades implica la siguiente:*

- *El logaritmo de la función de densidad es una función convexa (cóncava).*
- *$r(x)$  es una función monótona decreciente (creciente).*
- *El coeficiente de variación es mayor (menor) que uno.*

*Demostración.* véase [BP81] □

Las distribuciones con el logaritmo de la función de densidad cóncava, como la normal truncada siempre tienen el coeficiente de variación menor que uno y las distribuciones con el logaritmo de la función de densidad convexa como la distribución de Pareto siempre tienen el coeficiente de variación mayor que uno.

Las distribuciones de variación regular cubren una clase importante de las distribuciones de colas pesadas y su función de riesgo es monótona decreciente. Con lo cual, el coeficiente de variación de las distribuciones de colas pesadas es mayor que uno a partir de un cierto umbral.

$$cv(x - u | x > u) > 1$$

Este resultado nos permite estudiar las colas a partir del coeficiente de variación condicionado. En el Capítulo cuatro, trataremos este resultado con más detalle.

De estas clases de distribuciones vienen derivadas otras, que nos permiten relacionar el coeficiente de variación condicionado con la teoría de fiabilidad.

**Proposición 1.12.** *Las distribuciones son de clase IVRL(DVRL) si y sólo si el coeficiente de variación condicionado es mayor(menor) que uno.*

*Demostración.* véase [GK00]

□

## Capítulo 2

# Mixtura de distribuciones normales truncadas

Para ajustar datos financieros, Barndorff Nielsen [BN98] introdujo la normal inversa gaussiana, Hull y White [HW98] propusieron la mixtura de normales. Los dos modelos ofrecen un buen ajuste de las log-rentabilidades porque ajustan los momentos de orden superior. No obstante, los análisis empíricos demuestran que estos modelos no explican bien la no-normalidad debida a los valores extremos y a la asimetría.

La fluctuación pronunciada de los precios se debe a la interacción entre la oferta y la demanda, lo cual implica la no-normalidad de los datos. Esta última se ha tratado de maneras diferentes. Algunos autores suponen que la volatilidad sigue un proceso independiente de las rentabilidades, (véase [Tay05], Capítulo 11). Otros proponen los modelos mencionados anteriormente. Nuestro punto de vista va desde la utilización de la mixtura de normales truncadas cuando hay poca diferencia entre la oferta y la demanda hasta la mixtura de distribuciones de Pareto.

En este Capítulo, expondremos los resultados de nuestro artículo “the mixture of left-right truncated normal distributions” [CD09b]. En primer lugar, a diferencia del artículo, definimos el modelo de las transacciones sin suponer la normalidad de la oferta y de la demanda, luego proponemos la mixtura de normales truncadas cuando las distribuciones de la oferta y de la demanda son normales. Por otra parte, desarrollamos las propiedades de la distribución propuesta y demostramos que la función de verosimilitud tiene un máximo y que este máximo depende del coeficiente de variación incluso cuando

el punto de corte no es conocido. Por último, presentamos unos estudios empíricos donde comprobamos que el modelo propuesto es el más adecuado para ajustar los datos de tipo de cambio. Los estudios empíricos de los datos analizados en este Capítulo indican que la no-normalidad puede producirse además por el gran número de valores alrededor de la media "inliers" dando lugar a que la distribución de Laplace ofrezca mejor aproximación que la distribución normal.

## 2.1. Caracterización económica del modelo de intercambio

Los precios se determinan por las fuerzas de la oferta y de la demanda que actúan en el mercado. La interacción de los precios de la oferta, que se definen por, el precio de equilibrio más la variación de la oferta, y de los precios de la demanda, definidos por, el precio de equilibrio más la variación de la demanda da lugar a la fluctuación de los precios. A partir de los modelos de la oferta y de la demanda, vamos a definir un modelo del intercambio cuando la oferta y demanda reaccionan de manera similar (se aceptan de la misma manera las ofertas y las demandas) y cuando reaccionan de manera distinta. Este último se asemeja más a la realidad del mercado de los activos financieros.

Sea  $z$  una variable aleatoria de función de distribución  $F$  que representa el intercambio entre la demanda y la oferta. Sea  $b$  (bid) la variable aleatoria que representa la oferta con función de distribución  $F_b$ . Sea  $a$  (ask) la variable aleatoria que representa la demanda con función de distribución  $F_a$ .

Sea  $c$  una constante tal que si el demandante está dispuesto a comprar por encima de  $c$ , se produce el intercambio y si el oferente acepta a vender a un precio menor que  $c$ , también se produce el intercambio. Entonces, la distribución de la transacción es una mixtura de las distribuciones truncadas de la oferta y de la demanda. En la Figura 2.1, presentamos el caso simétrico de las distribuciones de la oferta y de la demanda.

A continuación, vamos a desarrollar los detalles para generar el modelo del intercambio.

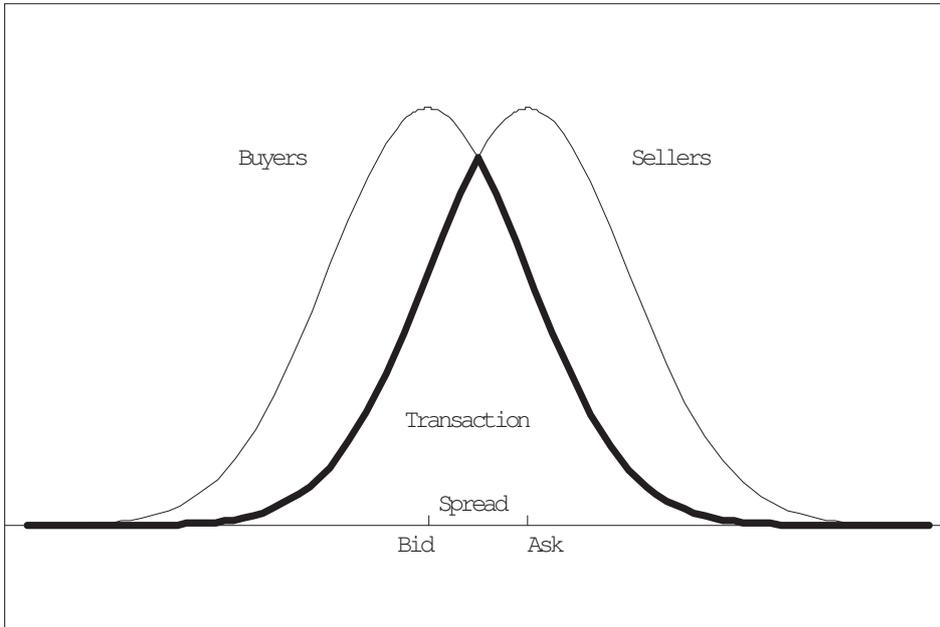


Figura 2.1: Caso simétrico de las transacciones

En primer lugar, se genera una observación  $d$  de distribución Bernoulli de probabilidad  $p$ , donde  $p$  es la proporción de la oferta. En segundo lugar, si  $d = 1$  generamos  $b$  de distribución  $F_b$ , sino generamos  $a$  de distribución  $F_a$ . Entonces, si  $d = 1$  y  $b > c$  tomamos  $z = b$  en caso contrario volvemos al paso anterior, Por último, si  $d = 0$  y  $a < c$  tomamos  $z = a$ , sino volvemos al paso anterior.

Por lo tanto,  $z$  puede escribirse como,

$$z = \begin{cases} b & \pi_b \\ a & \pi_a \end{cases}$$

donde  $a$  y  $b$  son las variables aleatorias truncadas.  $\pi_a$  y  $\pi_b$  son las proporciones de oferta y demanda aceptadas, que calcularemos en el resultado siguiente.

**Teorema 2.1.** *Con el algoritmo anterior, el intercambio  $z$  tiene la siguiente función de densidad*

$$f_z(t) = \pi f_b(t) + (1 - \pi)f_a(t),$$

donde  $f_b$  es la función de densidad de  $(b|b > c)$  y  $f_a$  es la función de densidad de  $(a|a < c)$ .

$$\pi = \pi_b = pp_b/(1 - p_3), p_a = Pr\{a > c\}, p_b = Pr\{b > c\} \text{ y } p_3 = (1 - p)p_a + p(1 - p_b)$$

*Demostración.* Si  $d = 1$  y  $b_1 > c$  entonces  $P\{z = b_1\} = pP\{b_1 > c\} = p_1$ , de manera similar si  $d = 0$  y  $a_1 < c$  entonces  $P\{z = a_1\} = (1 - p)P\{a_1 < c\} = p_2$ . Si no ocurre ninguna de las anteriores volvemos a generar  $d$  de manera que

$$P\{z = d\} = p(1 - p_b) + (1 - p)(1 - p_a) = p_3 \text{ tal que } p_1 + p_2 + p_3 = 1.$$

De manera iterativa, consideramos la sucesión  $\{a_n, b_n, d_n\}$  tales que

$$P\{z = b_n\} = p_1 + p_1p_3 + \dots + p_1p_3^{n-1}, \quad P\{z = a_n\} = p_2 + p_2p_3 + \dots + p_2p_3^{n-1} \text{ y por último } P\{z = d_n\} = p_3^n.$$

En el límite, la sucesión  $\{a_n, b_n, d_n\}$  converge a  $\{a, b, d\}$  tales que,

$$P\{z = b\} = pp_b/(1 - p_3) = \pi_b = \pi, \quad P\{z = a\} = (1 - p)(1 - p_a)/(1 - p_3) = \pi_a \text{ y } P\{z = d_n\} \text{ tiende a } 0. \quad \square$$

En un mercado ideal, cuando todas las ofertas se cubren por las demandas y todas las demandas se cubren por las ofertas, la distribución de las transacciones es una mixtura de las dos distribuciones de la oferta y de la demanda. En el caso de que las transacciones se aceptan a partir de un determinado nivel de oferta y a partir de un determinado nivel de demanda, la distribución de las transacciones es una mixtura de las dos distribuciones de la oferta y de la demanda truncadas.

Antes de definir el modelo de intercambio cuando la oferta y la demanda siguen una distribución normal, vamos a describir algunas propiedades de la normal truncada que nos serán de utilidad más adelante.

## 2.2. Distribución normal truncada

La distribución normal truncada ha sido utilizada en economía, especialmente en el truncamiento selectivo por Hechman [Hec79] (premio nobel en economía, 2000).

Denominamos por  $TN^+(a, \mu, \sigma^2)$  la función de densidad de una variable aleatoria normal  $X \sim N(\mu, \sigma)$ , con truncamiento simple por izquierda en  $a$ ;

$$f_a(x; \mu, \sigma) = \frac{\varphi((x - \mu)/\sigma)}{\sigma(1 - \Phi(\alpha))} \quad x > a$$

donde,

$$\varphi(t) = \exp\{-t^2/2\}/\sqrt{2\pi}, \quad \alpha = (a - \mu)/\sigma$$

y  $\varphi = \Phi'$  es la función de densidad de la normal estándar.

Si  $a = 0$ , la familia normal truncada es un caso particular de modelos exponenciales bajo cambio de escala [CP99], con densidad dada por;

$$f(x; \alpha, \beta) = c(\alpha, \beta) \exp(-\alpha x - \beta x^2); \quad x > 0$$

donde,

$$c(\alpha, \beta) = \sqrt{\beta}/\sqrt{\pi} \exp\{\alpha^2/(4\beta)\} (1 - \Phi(\frac{\alpha}{\sqrt{2\beta}})).$$

Observamos que esta parametrización es equivalente a la anterior con  $\alpha = \alpha/\sigma$  y  $\beta = 1/(2\sigma^2)$ .

La condición necesaria y suficiente para que las ecuaciones de verosimilitud tengan solución ha sido estudiada por [Cas94]. Sea  $\bar{x}$  la media de una muestra de  $X$  y  $cv$  el coeficiente de variación empírico. Las ecuaciones de verosimilitud se reducen a resolver

$$\frac{\alpha^2 + 1 - \alpha\lambda(\alpha)}{(\lambda(\alpha) - \alpha)^2} = 1 + cv^2 \quad (2.1)$$

donde  $\lambda(\alpha) = \varphi(\alpha)/(1 - \Phi(\alpha))$ . Los estimadores de máximo verosimilitud son

$$\hat{\mu} = \frac{y\alpha}{\alpha - \lambda(\alpha)}, \quad \hat{\sigma} = \frac{\bar{y}}{\lambda(\alpha) - \alpha} \quad (2.2)$$

La ecuación de verosimilitud (2.1) tiene solución si  $cv < 1$  y el  $cv$  es una función biyectiva de  $\alpha$ . Entonces, la familia de distribución normal truncada puede reparametrizarse en función del coeficiente de variación y del parámetro de escala  $\sigma$ .

De la misma manera, denominamos por  $TN^-(a, \mu, \sigma^2)$  la función de densidad de una normal con truncamiento a la derecha en  $x = a$ , viene definida por;

$$\bar{f}_a(x; \mu, \sigma) = \frac{\varphi((x - \mu)/\sigma)}{\sigma\Phi(\alpha)} \quad x < a$$

**Observación 2.1.** Sea  $x$  una variable aleatoria y  $a$  el punto de truncamiento, si  $x$  sigue una distribución  $TN(\mu, \sigma^2)$ ,  $(x + a)$  sigue una distribución  $TN^+(a, \mu + a, \sigma^2)$  y  $(x - a)$  sigue una distribución  $TN^-(a, a - \mu, \sigma^2)$ . En particular, si  $x$  sigue una  $TN^-(a, \mu, \sigma^2)$  entonces  $(-x)$  sigue una  $TN^+(-a, -\mu, \sigma^2)$ .

## 2.3. Modelo del intercambio bajo la hipótesis de normalidad

Suponemos en este apartado que los precios de la oferta y demanda siguen una distribución normal.

**Proposición 2.1.** *Supongamos que el precio de la oferta,  $b$ , sigue una  $N(\mu_b, \sigma_b^2)$  y el precio de la demanda,  $a$ , sigue una  $N(\mu_a, \sigma_a^2)$ . El intercambio  $z$  tiene la siguiente función de densidad*

$$f(z; \mu_a, \mu_b, \sigma_a, \sigma_b, c, \pi) = \pi f_c(z; \mu_b, \sigma_b) + (1 - \pi) \bar{f}_c(z; \mu_a, \sigma_a) \quad (2.3)$$

donde,

$$\pi = pp_b / (1 - p_3), \quad p_3 = p(1 - p_b) + (1 - p)p_a$$

*Demostración.* Es un caso particular del Teorema 2.1 cuando la distribución de la oferta y de la demanda son normales.  $\square$

Observamos que la distribución de  $(b|b > c)$  es  $TN^+(c, \mu_b, \sigma_b^2)$  y la distribución de  $(a|a \leq c)$  es  $TN^-(c, \mu_a, \sigma_a^2)$ , con lo cual, la distribución propuesta es una mixtura de normales truncadas asimétrica. La denominamos por MTN.

La idea básica de las familias asimétricas fue introducida por Fechner [Fec97] que juntó dos distribuciones normales truncadas en cero.

El dominio de los parámetros de la distribución MTN es el conjunto.

$$\Omega = \{(\mu_a, \mu_b, \sigma_a, \sigma_b, c); (c, \mu_a, \mu_b) \in \mathbb{R}, (\sigma_a, \sigma_b) \in \mathbb{R}^+\}$$

El modelo MTN incluye a varios sub-modelos, cuando  $\pi = 1/2$ ,  $\sigma_a = \sigma_b$  y  $\mu_a = \mu_b = c$ , el modelo corresponde a la distribución normal. Si  $\pi = 1/2$ ,  $\sigma_a = \sigma_b$  y  $(\mu_a + \mu_b)/2 = c$ , el modelo corresponde a la distribución normal truncada simetrizada ([DP04]).

En general,  $0 < \pi < 1$  y el soporte de la distribución es  $\mathbb{R}$ . Si  $\pi = 0$  o  $\pi = 1$  el soporte es  $(0, \infty)$ . Cuando  $\pi = 1$  y  $\mu_b$  tiende  $\infty$ , el modelo correspondiente es la distribución

exponencial.

Desde un punto de vista económico, podemos restringir el estudio a.

$$\{\mu_b < c < \mu_a\}$$

porque si  $c < \mu_b$  implica que más del 50 % de los precios de compra son mayores que  $c$ . Entonces,  $c$  no es un buen precio para el vendedor, y si  $\mu_a < c$ ,  $c$  no es un buen precio para el comprador.

**Proposición 2.2.** *Una condición necesaria y suficiente para que el modelo MTN sea simétrico respecto a  $c$  es  $c = (\mu_a + \mu_b)/2$ ,  $\sigma_a = \sigma_b$  y  $\pi = 1/2$ . Entonces, la distribución de  $z$  es una normal truncada simetrizada.*

*Demostración.* Sea  $x > 0$ , para que la distribución MTN sea simétrica, se debe cumplir la siguiente igualdad

$$\pi f_c(c+x; \mu_b, \sigma_b) = (1-\pi) \bar{f}_c(c-x; \mu_a, \sigma_a)$$

A partir de la ecuación anterior tenemos,  $\varphi((c+x-\mu_b)/\sigma_b)$  y  $\varphi((c-x-\mu_a)/\sigma_a)$  son proporcionales. Entonces, tomando logaritmos de estos últimos términos, tendremos que

$$((c+x-\mu_b)/\sigma_b)^2 \text{ y } ((c-x-\mu_a)/\sigma_a)^2$$

son iguales. Derivando éstos términos respecto a  $x$ , tendremos que  $\sigma_a = \sigma_b$ ,  $c = (\mu_a + \mu_b)/2$  y  $\pi = 1/2$ . La distribución resultante es normal truncada simetrizada.  $\square$

**Proposición 2.3.** *Si  $\mu_a \leq c \leq \mu_b$ , la función de distribución MTN es bimodal, de modas  $\mu_a$  y  $\mu_b$ .*

*Demostración.* Si  $c \leq \mu$ , la función de densidad de la normal truncada a la izquierda,  $f_c(z; \mu, \sigma)$  es creciente en  $(c, \mu)$  y decrece para valores grandes de  $z$ . Por otro lado,  $\bar{f}_c(z; \mu, \sigma)$  es decreciente en  $(\mu, c)$  y crece en  $z$  en el intervalo  $(-\infty, c)$ . Entonces, para  $\mu_a \leq c \leq \mu_b$ , el modelo MTN tiene dos modas en  $\mu_a$  y en  $\mu_b$ .  $\square$

En este caso, la distribución es similar a la mixtura de normales.

**Proposición 2.4.** *Si  $\mu_b \leq c \leq \mu_a$ , la función de distribución MTN es unimodal, de moda  $c$ .*

*Demostración.* la función de densidad de la normal truncada izquierda,  $f_c(z; \mu, \sigma)$  es monótona decreciente en  $z$  cuando  $\mu \leq c$ , y  $\bar{f}_c(z; \mu, \sigma)$  es monótona creciente en  $z$  cuando  $\mu \geq c$ . Entonces, como  $\mu_b \leq c$   $f_c(z; \mu_b, \sigma_b)$  es decreciente en  $z$  en el intervalo  $(c, \infty)$ , y como  $c < \mu_a$ ,  $\bar{f}_c(z; \mu_a, \sigma_a)$  es creciente en  $z$  en el intervalo  $(-\infty, c)$ .  $\square$

la Figura 2.2, nos muestra el caso de la distribución MTN cuando  $\pi = 1/2$ ,  $c = (\mu_a + \mu_b)/2$  y  $\sigma_a = 1,5\sigma_b$ . En este caso la distribución MTN es asimétrica.

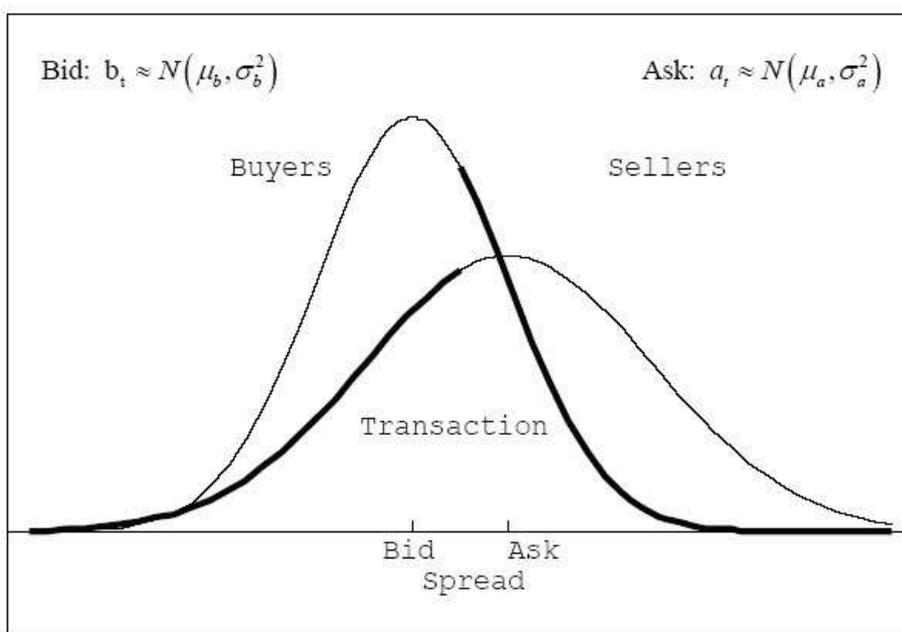


Figura 2.2: Caso asimétrico de las transacciones. Las líneas delgadas definen la distribución de los precios de la oferta y de la demanda y las líneas gruesas definen la distribución de las transacciones

El modelo propuesto corresponde a la "intersección" de dos muestras con diferentes medias, de manera similar la unión de la distribución de la oferta y de la distribución de la demanda corresponde a la mixtura de ambas. La función resultante en cualquier caso debe normalizarse para obtener una función de densidad.

**Proposición 2.5.** Sea  $x$  una variable aleatoria con función de distribución  $TN(\mu, \sigma)$ .

La función generatriz de momentos se define por,

$$M_x(t) = E(e^{tx}) = \frac{\Phi(\mu/\sigma + t\sigma)}{\Phi(\mu/\sigma)} \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$$

*Demostración.* Similar al caso de la distribución normal teniendo en cuenta las constantes normalizadoras, para más detalles (véase [Cas94]).  $\square$

Entonces, la esperanza y la varianza de  $x$  pueden expresarse como,

$$\begin{aligned} E(X) &= \mu + \sigma\lambda(\alpha) \\ \text{Var}(x) &= \sigma^2(1 - \lambda(\alpha^2 + \alpha\lambda(\alpha))) \end{aligned}$$

donde

$$\alpha = -\mu/\sigma \quad \text{y} \quad \lambda(\alpha) = \frac{\varphi(\alpha)}{1 - \Phi(\alpha)}$$

$\lambda(\alpha)$  es la inversa de la función ratio de Mills.

El coeficiente de variación es,

$$cv(x) = \sqrt{\pi - 2}/2 \approx 0,755$$

**Proposición 2.6.** Si  $y_b$  es una variable aleatoria con función de distribución  $TN^+(c, \mu_b, \sigma_b)$ .

La función generatriz de momentos es

$$M_b(t) = \frac{\Phi((\mu_b - c)/\sigma_b + t\sigma_b)}{\Phi((\mu_b - c)/\sigma_b)} \exp\left\{\mu_b t + \frac{\sigma_b^2 t^2}{2}\right\}$$

De manera similar, Si  $y_a$  es una variable aleatoria con función de distribución  $TN^-(c, \mu_a, \sigma_a)$ .

La función generatriz de momentos es

$$\overline{M}_a(t) = \frac{\Phi((c - \mu_a)/\sigma_a + t\sigma_a)}{\Phi((c - \mu_a)/\sigma_a)} \exp\left\{\mu_a t + \frac{\sigma_a^2 t^2}{2}\right\}$$

*Demostración.* Consideramos  $x_b = y_b - c$ ,  $x_b$  sigue una  $TN(\mu_b - c, \sigma_b)$ , según la Proposición 2.5 y como  $E(e^{ty_b}) = e^{tc} E(e^{tx_b})$ , obtenemos el resultado.  $\square$

De manera similar, sea  $x_a = c - y_a$ ,  $x_a$  sigue una  $TN(c - \mu_a, \sigma_a)$ , de la Proposición 2.5 y el resultado  $E(e^{ty_a}) = e^{tc} E(e^{-tx_a})$ , se deduce el resultado.

**Corolario 2.1.** Si  $y$  es una variable aleatoria con función de distribución MTN. La función generatriz de momentos es

$$M(t) = \pi M_b(t) + (1 - \pi)\overline{M}_a(t)$$

$\pi/\lambda_b$	1	2	3	4	5	6	7	8	9	10
0.5	6.00	7.50	9.00	9.95	10.56	10.98	11.28	11.51	11.69	11.83
0.4	6.12	8.24	10.25	11.56	12.41	12.99	13.41	13.72	13.96	14.16
0.3	6.50	9.49	12.39	14.32	15.59	16.47	17.10	17.57	17.93	18.22
0.2	7.17	11.54	16.20	19.50	21.74	23.31	24.44	25.28	25.93	26.44
0.1	8.16	14.65	23.75	31.44	37.18	41.39	44.50	46.85	48.67	50.09

Cuadro 2.1: Curtosis de la la mixtura de exponenciales

*Demostración.* La función generatriz de momentos se calcula como,

$$\int_{-\infty}^{\infty} e^{ty} ((\pi f_b(y; \mu_b, \sigma_b) + (1 - \pi) f_a(y; \mu_a, \sigma_a)) dy = \pi M_b(t) + (1 - \pi) \overline{M}_a(t)$$

□

Consideramos el caso límite cuando  $\mu$  tiende a infinito, es decir cuando la distribución normal truncada tiende a la distribución exponencial. En este caso la distribución mixtura de normales truncadas MTN puede ser aproximada a la mixtura de una exponencial de media  $\lambda_b$  y otra exponencial de media  $\lambda_a$ . Entonces, la función generatriz de momentos es

$$M_{\infty}(T) = \frac{\pi}{1 - \lambda_b t} + \frac{1 - \pi}{1 + \lambda_a t}$$

En la Tabla 2.1, presentamos como varia la curtosis de la mixtura de exponenciales suponiendo  $\lambda_a = 1$  para diferentes valores del parámetro  $\pi$  y diferentes valores de  $\lambda_b$ .

Observamos que la curtosis toma valores entre 6 que es el caso de la distribución de laplace a 50. Dependiendo de los parámetros, la distribución resultante puede ser de cola muy pesada.

A continuación, presentamos el estimador de máximo verosimilitud de la distribución MTN que aunque es una distribución con cinco o seis parámetros, dependiendo de si el parámetro de corte es conocido o no, tiene un máximo único.

## 2.4. Estimador de máxima verosimilitud (EMV).

Consideramos una muestra  $\{x_1, x_2, \dots, x_n\}$  proveniente de una distribución normal truncada asimétrica. Entonces, la función de log-verosimilitud está dada por:

$$l(\theta) = \sum \log f(x_i, \theta)$$

Hay dos casos a destacar, cuando el punto de corte  $c$  es constante y cuando  $c$  es variable.

**Teorema 2.2.** *La función de verosimilitud del modelo MTN puede expresarse como suma de tres modelos exponenciales sin ningún parámetro en común. Con lo cual, el estimador de máxima verosimilitud existe y es único.*

*Demostración.* Recordemos que si  $y \leq c$  entonces  $f_c(y_t, \mu_b, \sigma_b) = 0$  y si  $y > c$ , entonces  $f_c(y_t, \mu_a, \sigma_a) = 0$ . Sea  $n_c$  el tamaño de  $\{y_t > c\}$ , entonces la función de verosimilitud viene definida por,

$$\begin{aligned} l_c(\pi, \mu_b, \sigma_b, \mu_a, \sigma_a) &= \sum \log(\pi f_b(y_t, \mu_b, \sigma_b) + (1 - \pi)f_a(y_t, \mu_a, \sigma_a)) \\ &= \sum_{y_t \leq c} \log(1 - \pi)f_a(y_t, \mu_a, \sigma_a) + \sum_{y_t \geq c} \log(\pi f_b(y_t, \mu_b, \sigma_b)) \\ &= (n - n_c)\log(1 - \pi) + n_c \log(\pi) + \sum_{y_t \leq c} \log f_a(y_t, \mu_a, \sigma_a) + \\ &\quad \sum_{y_t > c} \log f_b(y_t, \mu_b, \sigma_b) \end{aligned}$$

La función

$$l\pi = (n - n_c)\log(1 - \pi) + n_c \log(\pi)$$

es la función de verosimilitud de la distribución Binomial de  $n_c$  de tamaño muestral  $n$  y probabilidad  $\pi$ . La función

$$lb = \sum_{y_t \geq c} \log(f_b(y_t, \mu_b, \sigma_b))$$

es la función de verosimilitud de la normal truncada a la izquierda de la muestra  $\{y_t > c\}$  y

$$la = \sum_{y_t \leq c} \log(f_a(y_t, \mu_a, \sigma_a))$$

es la función de verosimilitud de la normal truncada a la derecha de la muestra  $\{y_t \leq c\}$ . Las tres distribuciones son modelos exponenciales sin parámetros comunes. Por lo tanto el estimador existe y es único.  $\square$

**Proposición 2.7.** *Dada una muestra  $\{y_t\}$  del modelo MTN con parámetro de corte  $c$  conocido,  $n_c$  es el tamaño muestral de  $\{y_t > c\}$ . Consideramos las muestras  $\{y_t - c | y_t > c\}$  y  $\{c - y_t | y_t < c\}$  de media y coeficiente de variación  $(\bar{y}_+, cv_+)$  y  $(\bar{y}_-, cv_-)$  respectivamente. Entonces, el estadístico  $\{n_c, \bar{y}_+, cv_+, \bar{y}_-, cv_-\}$  es un estadístico suficiente del modelo MTN.*

*Demostración.* Las muestras  $\{y_t - c | y_t > c\}$  y  $\{c - y_t | y_t < c\}$  son la parte derecha y la parte izquierda de la muestra  $\{y_t\}$ . Las dos partes son positivas. A partir del Teorema 2.2 y las ecuaciones (2.1) y (2.2),  $\{n_c, \bar{y}_+, cv_+, \bar{y}_-, cv_-\}$  es un estadístico suficiente de los parámetros del modelo MTN.  $\square$

El estadístico suficiente es una nueva herramienta para el estudio de los datos financieros. La media de la parte derecha y de la parte izquierda tienen el mismo orden que las desviaciones estándar.

El coeficiente de variación es una nueva herramienta para estudiar la no-normalidad de las colas. Esta medida sólo utiliza los dos primeros momentos a diferencia de la curtosis que utiliza los cuatro primeros. Por lo tanto el coeficiente de variación proporciona más estabilidad que la curtosis.

**Observación 2.2.** *Dada una muestra  $\{y_t\}$  con función de distribución  $TN^+(c, \mu, \sigma^2)$  con  $c$  conocido, el EMV de  $\{y_t\}$  es el mismo que el EMV de la función de distribución  $TN(\mu_0, \sigma^2)$  correspondiente a la muestra  $\{y_t - c\}$  con  $\mu = c + \mu_0$ . De manera similar, dada una muestra  $\{y_t\}$  con función de distribución  $TN^-(c, \mu, \sigma^2)$ , el EMV de  $\{y_t\}$  es el mismo que el EMV de la función de distribución  $TN(\mu_0, \sigma^2)$  correspondiente a la muestra  $\{c - y_t\}$  con  $\mu = c - \mu_0$ .*

A continuación, demostramos que el máximo verosimilitud del modelo MTN existe cuando el punto de corte  $c$  no es constante.

Recordemos que dada una muestra de tamaño  $n$ , denominamos por  $\{y_{(t)}\}$  los estadísticos de orden tal que  $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ . En general, en el caso de la distribución normal el máximo verosimil no existe para una muestra de tamaño uno. De manera similar, si la muestra es cero, el máximo verosimil de una distribución exponencial no está definido. Estos casos ocurren en el modelo MTN cuando  $y_{(1)} \leq c \leq y_{(2)}$  y cuando

$y_{(n-1)} \leq c \leq y_{(n)}$ . La probabilidad de estos intervalos tiende a cero cuando aumenta el tamaño muestral.

**Teorema 2.3.** *Dada una muestra  $y_1, y_2, \dots, y_n$ , el EMV del modelo MTN existe siempre que exista  $\epsilon > 0$  tal que  $y_{(2)} + \epsilon \leq c \leq y_{(n-2)} - \epsilon$ , o  $c < y_{(1)} - \epsilon$  o  $c > y_{(n)} + \epsilon$ .*

*Demostración.* El Teorema 2.2 demuestra que la función de verosimilitud del modelo MTN puede expresarse en función de  $l\pi$ ,  $la$  y  $lb$ .  $l\pi$  corresponde a la distribución binomial,  $l\pi$  sólo puede tomar números finitos porque la muestra es finita. De hecho,  $l\pi$  es uniformemente acotado. Los otros términos pueden ser reducidos aplicando la Observación 2.2 a, la distribución normal truncada, con función de densidad  $f_0(y; \mu, \sigma)$  para las muestras  $\{y_t - c | y_t > c\}$  y  $\{y_t - c | y_t \leq c\}$ . Por simetría, es suficiente demostrar que  $f_0(y; \mu, \sigma)$  es uniformemente acotada en  $\{y_t - c | y_t > c\}$ .

Suponiendo  $y_{(2)} + \epsilon \leq c \leq y_{(n-2)} - \epsilon$ , entonces, el coeficiente de variación  $cv_s = s_c/m_c$  está en un compacto de  $\mathbb{R}^+$ . En particular,  $cv_c$  es uniformemente acotado para los valores de  $m_c$  distintos de cero.

Si  $cv_c < 1$ , el EMV depende de las muestras a través de las funciones continuas (2.1) y (2.2). La función de densidad,  $f_0(y, \mu, \sigma)$  alcanza el máximo en  $y = \mu$  o en  $y = 0$ , y viene dado por los siguientes valores.

$$\frac{1}{\sqrt{2\pi}\sigma(1 - \Phi(\alpha))}, \quad \frac{\varphi(\alpha)}{\sqrt{2\pi}\sigma(1 - \Phi(\alpha))}.$$

Donde  $\alpha$  es la solución de (2.1). Entonces, tenemos que comprobar que  $1 - \Phi(\alpha)$  y  $\lambda(\alpha) - \alpha$  son uniformemente acotadas.

El termino izquierdo de (2.1) es estrictamente creciente y aplica  $(-\infty, \infty)$  en  $(1, 2)$  y  $1 + cv_c^2$  está en un compacto en este intervalo, entonces, la solución  $\alpha$  de (2.1) está en un compacto de  $\mathbb{R}$ . Además como  $1 - \Phi(\alpha)$  y  $\lambda(\alpha) - \alpha$  son funciones continuas estrictamente decrecientes y tienden a cero, entonces son uniformemente acotadas en un conjunto compacto en  $\alpha$ .

Si  $cv_c \geq 1$ , el máximo verosimil corresponde a la distribución exponencial, en este caso, el máximo viene dado por  $1/m_c$ , lo cual es uniformemente acotado. Además, si  $cv_c$  tiende a uno,  $f_0(y; \mu, \sigma)$  tiende uniformemente a la distribución exponencial [CP99]. Entonces, el máximo es uniformemente acotado.

□

**Observación 2.3.** *Si  $c < y_{(1)} - \epsilon$  o  $c > y_{(n)} + \epsilon$ , El modelo MTN se reduce a la normal truncada. Si  $c$  pertenece a un compacto, los procedimientos anteriores son adecuados. Si  $c$  tiende a  $-\infty$ ,  $TN^+(c, \mu, \sigma^2)$  tiende a la distribución normal  $N(\mu, \sigma^2)$ . De manera similar, si  $c$  tiende a  $\infty$ ,  $TN^-(c, \mu, \sigma^2)$  tiende a la distribución normal. Entonces, la función de verosimilitud es uniformemente acotada.*

De la teoría del valor extremo véase ([Col01] y [EKM97]), si la distribución es de variación regular, la distribución límite de los valores que exceden un determinado umbral converge a la distribución de Pareto. Entonces, siguiendo la misma idea de la dispersión entre las ordenes de la demanda y oferta. Consideramos como modelo posible la distribución de Pareto.

## 2.5. Estudios empíricos

El ejemplo que presentamos corresponde a unos datos diarios de tipo de cambio de Dólar-Euro entre el 1 de enero de 1999 y el 1 de enero de 2003. Los datos están en OANDA Corporaton <http://www.oanda.com/convert/fxhistory>. Consideramos la serie de las rentabilidades,

$$y_t = 100 \log(x_t/x_{t-1})$$

donde  $x_t$  es el valor de un Euro en Dólar en el tiempo  $t$ .

Tenemos 1825 datos diarios (1824 rentabilidades). La Tabla 2.2 nos muestra la media, la desviación estándar, la asimetría y la curtosis de las log-rentabilidades para los datos diarios, los datos cada dos días, los datos cada tres y los datos cada cuatro días. También, nos muestra la media y el coeficiente de variación de los datos positivos.

Observamos que el coeficiente de variación de los valores absolutos tiende al coeficiente de variación teórico de los valores absolutos de la distribución normal y el exceso de curtosis tiende a cero, lo cual indica normalidad de las log-rentabilidades a medida que aumenta el tiempo. El coeficiente de variación de los valores positivos está entre 0.756 distribución normal y 1 que corresponde a la distribución de Laplace.

Tiempo	tamaño	media	Sd	Sk	ku-3	media(abs)	cv(abs)
diarios	1824	0.0037	0.5545	0.1019	1.90002	0.3829	1.0471
dos dias	912	0.0074	0.7727	0.7972	0.7972	0.5764	0.8923
tres dias	608	0.0111	0.9550	0.5408	0.5408	0.7280	0.8481
cuatro dias	456	0.0148	1.0694	0.1380	0.1380	0.8436	0.7777

Cuadro 2.2: Estadísticos descriptivos de los datos, Dólar-Euro

Suponemos simetría respecto a cero de las distribuciones que consideramos a continuación, la distribución normal, la mixtura de normales, la distribución normal inversa gaussiana. Las rentabilidades están centralizadas siendo este punto frecuente cuando se trata de modelizar rentabilidades.

La distribución normal se determina por la desviación estándar, la distribución de Laplace se determina por la media de los valores absolutos de las rentabilidades.

En la Figura 2.3, presentamos el histograma de la distribución empírica de los datos de tipo de cambio y las distribuciones de densidad estimadas por la distribución normal y la distribución de laplace. Alrededor de cero, la función de densidad empírica es puntiaguda, es decir, hay más valores centrales que los predichos por la distribución normal. Parece que La distribución de Laplace se ajusta mejor a estos datos.

Siguiendo Hull y White [HW98], consideramos la mixtura de normales como un promedio ponderado de dos normales de media cero, la función de distribución se define por,

$$G(y) = p\Phi\left(\frac{y}{u\sigma}\right) + (1 - p)\Phi\left(\frac{y}{v\sigma}\right)$$

donde  $\Phi(y)$  es la distribución estándar acumulada de la distribución normal.  $p$  corresponde al peso que damos a cada una de las normales y  $u, v$  vienen definidos tal que  $pu^2 + (1 - p)v^2 = 1$ .

La estimación de los parámetros se ha realizado mediante el método de máxima verosimilitud, Hull y White [HW98] sugieren el método de los fractiles de la distribución. Hemos tomado como puntos iniciales, para calcular el estimador de máxima verosimilitud, los obtenidos mediante el método de los fractiles.

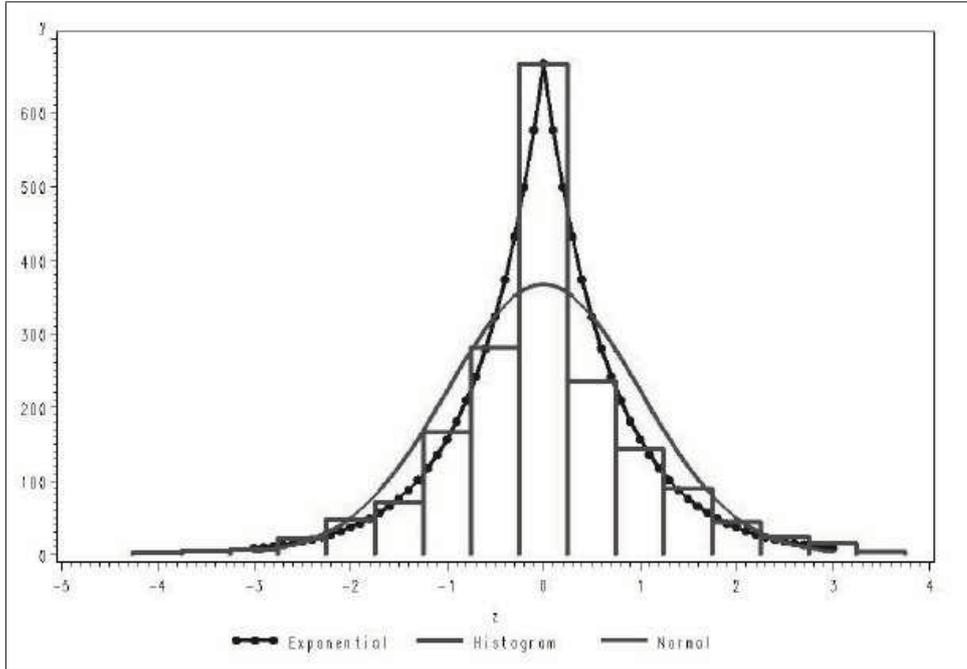


Figura 2.3: Histograma de los datos Dólar-Euro y las distribuciones normal y Laplace estimadas

Las distribuciones infinitamente divisibles son una alternativa a la distribución normal, éstas distribuciones generalizan los procesos tipo paseo aleatorio a los procesos de Lévy. Una distribución de importancia tanto en el mundo de seguros como en el mundo de finanzas infinitamente divisible cerrada por convolución es, la distribución normal inversa gaussiana denominada por *NIG*, y introducida por Barndorff Nielsen [BN98]. La función de densidad de la *NIG* se define por,

$$f(y; \phi, \omega) = \frac{\omega \exp \omega}{\pi \sqrt{y^2 \omega \phi^{-1} + \omega^2}} K_1(\sqrt{y^2 \omega \phi^{-1} + \omega^2})$$

donde  $K_1$  es la función de Bessel de índice 1. La varianza y la curtosis de  $y$  vienen dados por [BS03]

$$\text{var}[y] = \phi, \quad \text{ku}[y] = 3/\omega.$$

La estimación de los parámetros mediante el método de momentos devuelve la varianza y la curtosis de los datos, con lo que, es fácil interpretar dichos parámetros. Además, los parámetros se pueden estimar fácilmente por el método de máxima verosimilitud. En el

caso de los modelos de volatilidad multivariante, Castillo y Lee [CL08] han propuesto La h- verosimilitud.

A continuación, estimamos los parámetros de la distribución empírica de los datos mediante la distribución MTN, la distribución normal, la distribución de Laplace y la distribución normal inversa gaussiana y calculamos el estadístico  $\chi^2$  y los p-valores correspondientes para comprobar cual de las distribuciones ajusta mejor estos datos.

La agrupación de los datos se realiza en categorías. Los puntos de corte se expresan en términos de la desviación estándar. Tenemos 99 ceros, tomamos la mitad en la parte positiva y la otra mitad en la parte negativa.

La estimación de los parámetros se ha realizado mediante el método de máxima verosimilitud, con lo cual, la distribución asintótica está entre la distribución  $\chi^2$  con  $k - 1$  grados de libertad y la distribución  $\chi^2$  con  $k - 1 - p$  grados de libertad, donde  $p$  es el número de parámetros estimados. El verdadero p-valor pertenece a este intervalo.

Para comparar el modelo MTN con los modelos anteriores y restringiendonos al caso simétrico con  $c = 0$ , el coeficiente de variación es igual a 1.0471. Cuando el  $cv \geq 1$ , el máximo verosimilitud corresponde a la distribución exponencial y la distribución MTN se reduce a la distribución de Laplace.

Para valorar la bondad del ajuste vamos a realizar el test  $\chi^2$ . Las distribuciones consideradas son la distribución normal, la distribución de Laplace, la distribución mixtura de normales y la distribución inversa gaussiana. En la Tabla 2.3, calculamos los estadísticos del test y sus p-valores correspondientes.

Observamos que para los datos diarios el mejor ajuste de las frecuencias observadas viene dado mediante la distribución de Laplace dado que el p-valor es mayor que 0.01. La distribución de Lapace es la única que no tenemos evidencias para rechazar con un 99% tal como lo indica la Tabla 2.3.

La mixtura de normales y la distribución normal inversa gaussiana mejoran el ajuste respecto a la distribución normal. Utilizando el método de los momentos, la distribución

Puntos de corte	Observados	Mix_Nor	NIG	Laplace	Normal
-4.0	2	0.4	8.3	2.8	0.1
-3.5	2	1.4	4.6	3.0	0.4
-3.0	8	5.1	7.6	6.1	2.0
-2.5	15	14.8	13.0	12.6	8.9
-2.0	26	36.3	23.3	25.9	30.2
-1.5	63	74.3	44.8	53.5	80.4
-1.0	102	127.1	94.8	110.4	167.5
-0.5	212	182.1	225.3	227.8	273.4
0.0	495	470.5	490.4	470.0	349.2
0.5	477	470.5	490.4	470.0	349.2
1.0	172	182.1	225.3	227.8	273.4
1.5	125	127.1	94.8	110.4	167.5
2.0	59	74.3	44.8	53.5	80.4
2.5	38	36.3	23.3	25.9	30.2
3.0	18	14.8	13.0	12.6	8.9
3.5	4	5.1	7.6	6.1	2
4.0	3	1.4	4.6	3.0	0.4
>4.0	3	0.4	8.3	2.8	0.1
chi-square statistic		59.36	59.79	31.30	473.08
Número de parámetros		3	2	1	1
P-value		0.000	0.000	0.012-0.018	0.000

Cuadro 2.3: Frecuencias observadas y estimadas de los datos de tipo de cambio diarios, los puntos de corte se expresan en función de  $\sigma$ . Las distribuciones por las que ajustamos son: La distribución mixtura de normales, la distribución NIG, la distribución de Laplace y la distribución normal.

normal inversa gaussiana tendrá la misma varianza y la misma curtosis que los datos.

En las tablas 2.4 y 2.5 se considera el caso simétrico del modelo MTN, los parámetros estimados corresponden a los datos cada dos y cada tres días, con lo cual, el coeficiente de variación es menor que uno. En este caso la distribución MTN no coincide con la distribución de Laplace como es el caso de los datos diarios.

En la Tabla 2.4, para los datos cada dos días, La distribución MTN es el modelo que arroja el p-valor más grande, si bien no podemos rechazar el modelo NIG debido a que otorga un p-valor más grande que 0.05. La distribución de Laplace mejora el ajuste con respecto a la distribución normal pero se rechaza al 95 %. Observamos que no tenemos evidencias para rechazar la mixtura de normales, no obstante, el estadístico es elevado y tiene un parámetro de más respecto a la distribución MTN.

En la Tabla 2.5, para los datos cada tres días, la distribución MTN es la que presenta el p-valor más grande. También, se acepta el ajuste mediante la distribución NIG y la mixtura de normales.

Las distribuciones mixtura de normales y la inversa gaussiana mejoran el ajuste respecto a la distribución normal sin embargo hay una diferencia significativa entre la distribución de Laplace, la mixtura de normales y la distribución NIG. Las distribuciones normal y Laplace se rechazan al 95 %.

Aunque la función de distribución NTN en el caso simétrico ajusta mejor que otras alternativas, el ajuste es mejorable eliminando la restricción en  $\pi$  y  $c$ .

La función *profile* log verosimilitud cada dos días en el caso simétrico cuando  $c = 0$ ,  $\pi = 0,5$  devuelve  $l_0 = -1033,78$ . El máximo verosimilitud de la distribución MTN cuando  $c = 0$  y sin ninguna restricción sobre  $\pi$ , devuelve  $\hat{\pi} = 0,486$  y la función de log verosimilitud es igual a  $l_1 = -1030,22$ .

El máximo verosimilitud sin ninguna restricción se obtiene utilizando la función *profile* de verosimilitud, los resultados obtenidos son  $\hat{c} = 0,11$ ,  $\hat{\pi} = 0,390$  y el logaritmo de la función *profile* verosimilitud devuelve  $l_2 = -1026,99$ .

Puntos de corte	Observados	MTN	Mix_Nor	NIG	Laplace	Normal
-4.0	0	0.5	0.1	1.0	2.1	0.0
-3.5	0	1.0	0.4	1.2	2.1	0.2
-3.0	5	2.7	1.8	2.7	4.0	1.0
-2.5	6	6.7	6.2	6.1	7.8	4.4
-2.0	14	15.4	17.1	13.9	15.3	15.1
-1.5	32	33.1	38.9	31.4	29.8	40.2
-1.0	80	65.9	71.8	68.4	58.3	83.8
-0.5	103	121.8	108.9	132.7	113.9	136.7
0.0	229	209.1	210.9	198.7	222.7	174.6
0.5	191	209.1	210.9	198.7	222.7	174.6
1.0	118	121.8	108.9	132.7	113.9	136.7
1.5	72	65.9	71.8	68.4	58.3	83.8
2.0	34	33.1	38.9	31.4	29.8	40.2
2.5	19	15.4	17.1	13.9	15.3	15.1
3.0	6	6.7	6.1	6.2	7.8	4.4
3.5	1	2.7	1.8	2.7	4.0	1.0
4.0	2	1.0	0.4	1.2	2.1	0.2
> 4.0	0	0.5	0.1	1.0	2.1	0.0
chi-square statistic		17.36	20.45	24.29	28.58	70.16
Número de parámetros		2	3	2	1	1
P-value		0.298-0.430	0.117-0.252	0.060-0.112	0.027-0.039	0.0000

Cuadro 2.4: Frecuencias observadas y estimadas de los datos de tipo de cambio cada dos días, los puntos de corte se expresan en función de  $\sigma$ . Las distribuciones por las que ajustamos son: La distribución MTN, la distribución mixtura de normales, la distribución NIG, la distribución de Laplace y la distribución normal.

Cut points	Observed	MTN	Mix_Nor	NIG	Laplace	Normal
-4.0	0	0.2	0.1	0.4	1.6	0.0
-3.5	0	0.5	0.3	0.6	1.5	0.1
-3.0	3	1.4	1.2	1.5	2.9	0.7
-2.5	2	4.1	4.0	3.8	5.5	3.0
-2.0	11	10.5	11.1	9.5	10.6	10.1
-1.5	32	23.6	25.4	22.7	20.5	26.8
-1.0	37	47.3	47.7	49.2	39.4	55.8
-0.5	86	84.0	81.2	90.0	75.9	91.1
0.0	136	132.4	133.2	126.5	146.2	116.4
0.5	129	132.4	133.2	126.5	146.2	116.4
1.0	78	84.0	81.2	90.0	75.9	91.1
1.5	56	47.3	47.7	49.2	39.4	55.8
2.0	19	23.6	25.4	22.7	20.5	26.8
2.5	14	10.5	11.1	9.5	10.6	10.1
3.0	3	4.1	4.0	3.8	5.5	3.0
3.5	2	1.4	1.2	1.5	2.9	0.7
4.0	0	0.5	0.3	0.6	1.5	0.1
> 4.0	0	0.2	0.1	0.4	1.6	0.0
chi-square statistic		14.11	13.98	17.92	28.83	29.17
Number of parameters		2	3	2	1	1
P-value		0.517-0.659	0.451-0.669	0.267-0.394	0.025-0.036	0.023-0.033

Cuadro 2.5: Frecuencias observadas y estimadas de los datos de tipo de cambio cada tres días, los puntos de corte se expresan en función de  $\sigma$ . Las distribuciones por las que ajustamos son: La distribución MTN, la distribución mixtura de normales, la distribución NIG, la distribución de Laplace y la distribución normal.

$\pi$	$c$	$\mu_b$	$\sigma_b$	$\mu_a$	$\sigma_a$	log-verosimilitud
0.5000	0.000	-2.400	1.407	2.400	1.407	-1035.596
0.4857	0.000	-1.654	1.265	3.551	1.602	-1033.784
0.3904	0.110	-0.290	0.931	3.287	1.557	-1026.990

Cuadro 2.6: Función *profile* log-verosimilitud de la distribución MTN

La Tabla 2.6 que definimos a continuación nos muestra los parámetros de la distribución MTN estimados mediante el método de máxima verosimilitud con valores prefijados en  $\pi$  y  $c$  y sin ningún tipo de restricción sobre las mismas.

El test de razón de verosimilitud muestra que el ajuste por la distribución MTN es mejor eliminando las restricciones.

En la Figura 2.4, dibujamos la función *profile* log-verosimilitud de la distribución MTN para estimar el parámetro  $c$ .

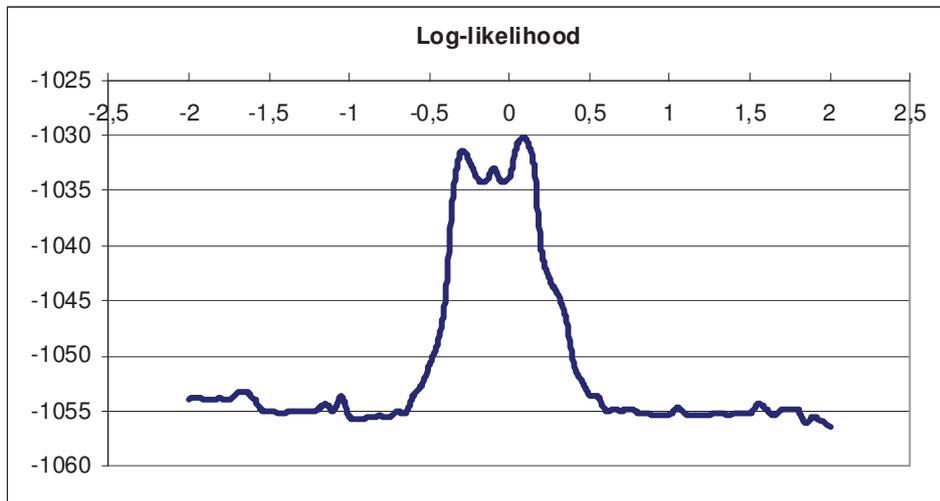


Figura 2.4: *Profile* de verosimilitud de la distribución MTN para estimar  $c$

## Capítulo 3

# Estimación de la distribución de Pareto Generalizada

La distribución de Pareto generalizada denominada por GPD ha sido utilizada en hidrología, seguros, finanzas y también ha sido aplicada al diseño de sistemas medioambientales [FR03]. Hosking y Wallis [HW87] utilizaron la GPD en la modelización de los máximos anuales de los flujos del río de Hunsingore en Inglaterra. La distribución de Pareto es un caso particular de la distribución GPD y es muy apropiada cuando hay grandes pérdidas véase Arnold [Arn83].

La distribución de Pareto generalizada, se puede aplicar en cualquier situación en la cual la distribución exponencial puede ser utilizada y donde se requieren métodos robustos para contrastar frente a colas pesadas o colas ligeras (véase [DS90]).

En este Capítulo, expondremos los resultados de nuestro artículo “ Estimation of the generalized Pareto distribution ”. A diferencia del artículo, en primer lugar desarrollamos algunas propiedades de la distribución de Pareto y demostramos que bajo el supuesto de que el coeficiente de variación es mayor que uno, las ecuaciones de verosimilitud siempre tienen solución. En segundo lugar, extendemos la estimación de los parámetros a través del coeficiente de variación al caso de la distribución de Pareto generalizada y probamos que el comportamiento de la función *profile likelihood* depende del coeficiente de variación empírico. Luego, ilustramos con un ejemplo que la función *profile* verosimilitud puede no tener soluciones. Finalmente, la distribución GPD al incluir distribuciones de soporte compacto implica problemas de convergencia de los estimadores. Para ello,

proponemos un modelo alternativo que denominamos por TNP donde remplazamos las distribuciones GPD de soporte compacto por la distribución normal truncada.

### 3.1. Propiedades de la distribución de Pareto; coeficiente de variación

**Definición 3.1.** *La función de densidad de la distribución de Pareto se define por*

$$f(x; \sigma, \alpha) = \frac{\alpha}{\sigma} \left( \frac{\sigma}{x + \sigma} \right)^{\alpha+1} \quad x \geq 0, \quad (3.1)$$

*y la función de supervivencia se define por*

$$\bar{F}(x; \sigma, \alpha) = \left( \frac{\sigma}{x + \sigma} \right)^{\alpha} \quad x \geq 0.$$

*donde  $\sigma$  es el parámetro de escala y  $\alpha$  es el parámetro del peso de la cola.*

La distribución de Pareto es de variación regular,

$$\bar{F}(x; \sigma, \alpha) = x^{-\alpha} L(x), \quad \text{donde } L(x) = \sigma^{\alpha} (1 + \sigma/x)^{-\alpha}.$$

Los momentos de orden  $r$  sólo existen cuando  $r < \alpha$  y vienen dados por,

$$E((1 + (X/\sigma))^r) = (1 - r/\alpha)^{-1}.$$

La media y el coeficiente de variación son,

$$E(x) = \sigma/(\alpha - 1) \quad \text{y} \quad \zeta^2(x) = \alpha/(\alpha - 2).$$

Es importante destacar que para la distribución de Pareto, el coeficiente de variación es función biyectiva de  $\alpha$  y es mayor que uno. De hecho, el modelo de Pareto puede parametrizarse en función del coeficiente de variación y del parámetro de escala de la siguiente manera,

$$\alpha = 2cv^2(x)/(cv^2(x) - 1), \quad \sigma = (\alpha - 1)E(x).$$

la distribución de Pareto con  $\alpha \leq 2$ , asintóticamente tiene las mismas colas que las distribuciones estables. Las cuales han sido introducidas en finanzas por Mandelbrot [Man63], también han sido utilizadas por Mittnik y Rachev [MR93]. Una amplia

discusión sobre el uso de estas distribuciones se encuentra en [CLM97] y [BG74]. Sin embargo, a diferencia de los estudios empíricos el parámetro del peso de la cola se mantiene constante en el tiempo, con lo que dichas distribuciones no son apropiadas para la modelización de los datos.

Por otro lado, la varianza de la distribución de Pareto con  $\alpha \leq 2$  no es finita, lo cual contradice los estudios que consideran la necesidad de por lo menos los dos primeros momentos. Se suma a lo anterior que el coeficiente de variación no está definido, por lo tanto nos restringimos en este Capítulo al caso de la distribución de Pareto con  $\alpha > 2$ .

**Proposición 3.1.** *La distribución de Pareto tiende a la distribución exponencial cuando  $\alpha$  tiende a infinito y  $\sigma = a\alpha + b$*

*Demostración.* La distribución exponencial es el caso límite cuando  $\alpha$  y  $\sigma$  tienden al infinito con una restricción lineal. Sea  $\alpha = 1/k$ ,  $k > 0$ , y  $\sigma = a\alpha + b$ ,  $a > 0$  entonces  $f(x; \alpha, \sigma)$  se puede expresar como;

$$\begin{aligned} f(x; a, k) &= \frac{1}{a + bk} \exp \left\{ \left( \frac{1}{k} + 1 \right) \ln \left( \frac{xk}{a + bk} + 1 \right) \right\} \\ &\sim \frac{1}{a} \exp \left\{ -\frac{x}{a} \right\} \quad k \rightarrow 0 \end{aligned}$$

por lo tanto,  $f$  es una exponencial de valor esperado  $a$ . □

**Corolario 3.1.** *Si consideramos una Pareto( $\alpha, \sigma$ ) con esperanza constante, Esta distribución tiende a la exponencial con la misma esperanza cuando  $\alpha$  tiende a infinito.*

Para estimar los parámetros de la distribución de Pareto, [HW87] propusieron el método de los momentos ponderados como alternativa al método de máxima verosimilitud dado que las ecuaciones de verosimilitud no siempre tienen soluciones, aquí proponemos que bajo la condición de que el coeficiente de variación es mayor que uno, estas ecuaciones siempre tienen soluciones.

## 3.2. Existencia del EMV en la distribución de Pareto

De la ecuación 3.1, observamos que no existe un estadístico suficiente para los parámetros  $\alpha$  y  $\sigma$ , entonces, en la estimación de los parámetros, debemos considerar

todos los datos de la muestra, en cada paso del método iterativo.

Dada una muestra  $x_1, \dots, x_n$  del modelo 3.1 definido anteriormente, la función de log-verosimilitud es

$$L(\alpha, \sigma, x_1, \dots, x_n) = n \log \alpha + n\alpha \log \sigma - (\alpha + 1) \sum_{i=1}^n \ln(x_i + \sigma).$$

El estimador de máxima verosimilitud  $(\tilde{\alpha}, \tilde{\sigma})$ , será aquel que satisfaga las ecuaciones de verosimilitud.

$$\frac{\delta L}{\delta \alpha} = \frac{n}{\alpha} + n \ln \sigma - \sum_{i=1}^n \log x_i + \sigma$$

$$\frac{\delta L}{\delta \sigma} = \frac{n\alpha}{\sigma} - (\alpha + 1) \sum_{i=1}^n \frac{1}{x_i + \sigma}$$

De estas ecuaciones, hallamos  $\hat{\alpha}$  y  $\xi(\sigma)$ .

$$\begin{cases} \hat{\alpha} &= \xi(\sigma)^{-1} \\ \xi(\sigma) &= n^{-1} \left( \sum_{i=1}^n \ln \left( \frac{x_i + \sigma}{\sigma} \right) \right) \end{cases}$$

De este modo, la función *profile* verosimilitud es

$$l_p(\sigma) = -\log(\sigma \xi(\sigma)) - \xi(\sigma) - 1. \quad (3.2)$$

Si  $l_p(\sigma)$  tiene un máximo en un determinado valor  $\hat{\sigma}$  finito, entonces  $(\hat{\alpha}, \hat{\sigma})$  es el estimador de máxima verosimilitud. Por otro lado,  $l_p(\sigma)$  puede ser una función monótona creciente en  $(0, \infty)$  y entonces no tendrá máximo, esto se debe a que el coeficiente de variación de la distribución de Pareto es siempre mayor que uno, aunque empíricamente puede ser menor que uno.

A continuación, definimos la función *profile* verosimilitud que requiere menos condiciones para encontrar el EMV.

**Definición 3.2.** La función *profile* verosimilitud para un determinado parámetro  $\alpha$  es la función

$$L_p = L_p(\sigma) = L(\hat{\alpha}, \sigma),$$

donde  $\hat{\alpha}$  es el estimador de máxima verosimilitud de  $\alpha$  del modelo restringido para un valor prefijado de  $\sigma$ .

De manera similar, definimos la *profile log verosimilitud* como

$$l_p = l_p(\sigma) = \log(L_p(\sigma)) = \log(L(\hat{\alpha}, \sigma)).$$

La función *profile verosimilitud*  $L_p$  puede ser usada como si fuese una verdadera función de verosimilitud, con lo cual, el *profile* estimador de máxima verosimilitud de  $\alpha$  es igual al estimador de máxima verosimilitud general  $\hat{\alpha}$  basado en  $L(\alpha, \sigma)$ . De hecho tenemos

$$L(\hat{\alpha}, \sigma) = \sup_{\alpha} L(\alpha, \hat{\sigma}) = \sup_{\alpha} L(\alpha, \hat{\sigma}) = \sup_{\alpha} L_p(\sigma) = L_p(\hat{\sigma})$$

Los estimadores  $(\hat{\alpha}, \hat{\sigma})$  preservan las propiedades asintóticas del EMV, es decir son asintóticamente eficientes e insesgados.

La proposición que presentamos a continuación nos permite el uso de unas herramientas para demostrar que el comportamiento de  $l_p(\sigma)$  se caracteriza por el coeficiente de variación empírico.

**Proposición 3.2.** *Tenemos que,*

$$\lim_{\sigma \rightarrow \infty} \log(1 + x/\sigma) = x, \quad \lim_{\sigma \rightarrow \infty} \sigma x/(\sigma + x) = x \quad \text{y} \quad \lim_{\sigma \rightarrow \infty} \sigma^2(\log(1 + x/\sigma) - x/(\sigma + x)) = x^2/2$$

*Demostración.* Se utilizan los cálculos estándar de series. □

**Proposición 3.3.** *Sea  $l_p(\sigma)$  la función *profile verosimilitud* definida por 3.2 de la variable aleatoria  $X$  y sea  $\bar{x}$  la media de una muestra de  $X$ , entonces,*

$$\lim_{\sigma \rightarrow 0} l_p(\sigma) = -\infty, \quad \lim_{\sigma \rightarrow \infty} l_p(\sigma) = -\log(\bar{x}) - 1.$$

*Demostración.* Utilizando la Proposición 3.2, podemos comprobar que  $\xi(\sigma)$  es una función decreciente tales que,

$$\lim_{\sigma \rightarrow 0} \xi(\sigma) = \infty \quad \lim_{\sigma \rightarrow \infty} \xi(\sigma) = 0,$$

por otro lado, tenemos que  $\lim_{\sigma \rightarrow 0} \sigma \xi(\sigma) = \bar{x}$ , lo que implica el resultado en el segundo límite. Cuando  $\sigma$  tiende a 0,  $l_p(\sigma)$  tiende a  $-\infty$ . □

**Observación 3.1.** *El límite de la función *profile verosimilitud*,  $l_p(\sigma)$ , corresponde a la función *log-verosimilitud* de la familia exponencial  $\theta \exp -\theta x$ , su función *log-verosimilitud* es  $\log(\theta) - \theta \bar{x}$  y el estimador de máxima verosimilitud es  $\hat{\theta} = 1/\bar{x}$  y  $l(\hat{\theta}) = -\log(\bar{x}) - 1$ .*

De hecho, la distribución exponencial puede ser una buena aproximación de las muestras que provienen de la distribución Pareto. La distribución exponencial es el caso límite de la distribución de Pareto.

**Proposición 3.4.**  $l_p(\sigma)$  tendrá un máximo global si y sólo si existe  $\sigma$  tal que  $l_p(\sigma) > -\text{Log}(\bar{x}) - 1$ .

**Teorema 3.1.** Si el coeficiente de variación empírico  $cv > 1$  entonces  $l_p(\sigma)$  es una función monótona decreciente, y si el  $cv < 1$ ,  $l_p(\sigma)$  es una función monótona creciente para  $\sigma$  suficientemente grande.

*Demostración.* Al derivar la expresión 3.2 respecto a  $\sigma$  obtenemos,

$$l'_p(\sigma) = -(\xi(\sigma) + \sigma\xi'(\sigma) + \sigma\xi(\sigma)\xi'(\sigma))/(\sigma\xi(\sigma)), \quad (3.3)$$

como  $\sigma\xi(\sigma)$  es mayor que 0 entonces  $l'_p(\sigma)$  tiene el mismo signo que el numerador de la expresión 3.3, que puede expresarse de la siguiente manera

$$-\frac{1}{n}\sum_{i=1}^n(\log(1 + x_i/\sigma) - x_i/(\sigma + x_i)) + \left(\frac{1}{n}\sum_{i=1}^n(\log(1 + x_i/\sigma))(x_i/(\sigma + x_i))\right).$$

Aplicando la Proposición 3.2, podemos demostrar que

$$\lim_{\sigma \rightarrow \infty} \sigma^2 \text{num}(\sigma) = -\left(\frac{1}{2n} \sum_{i=1}^n x_i^2 - \bar{x}^2\right).$$

Remarcamos además que  $\frac{1}{2n} \sum_{i=1}^n x_i^2 - \bar{x}^2 > 0$  es igual a  $(\mu_2 - \mu_1^2)/\mu_1^2$ , lo cual es equivalente a  $cv > 1$ . □

**Corolario 3.2.** Sea  $\{x_i\}$  una muestra de valores positivos de coeficiente de variación empírico mayor que uno ( $cv > 1$ ), entonces, la función de verosimilitud tiene un máximo global.

La condición es suficiente y creemos que es necesaria debido a que si el  $cv < 1$  la función de verosimilitud no tiene solución.

Hosking y Wallis [HW87] estudiaron la estimación de la distribución de Pareto generalizada y comentaron que en algunos casos no existían soluciones. Esta dificultad de encontrar el valor del parámetro  $\xi$  se hace más difícil a medida que decrece  $\xi$ , especialmente en el caso de las distribuciones GPD con soporte acotado. En la próxima sección, generalizamos los resultados comentados anteriormente al caso de la distribución de Pareto generalizada y explicaremos los problemas de no convergencia destacados en algunos casos en la estimación de los parámetros [CH97] y [DS90].

### 3.3. Extensión al caso general

**Definición 3.3.** La distribución de Pareto generalizada viene definida por,

$$F(x) = 1 - (1 + \xi x/\psi)^{-1/\xi},$$

donde,  $\psi$  es el parámetro de escala, y  $\xi$  es el parámetro peso de la cola.

Si  $\xi > 0$  la GPD corresponde a la distribución de Pareto, si  $\xi < 0$ , y  $0 < x < \psi/|\xi|$  la GPD tiene soporte acotado. El caso límite  $\xi = 0$  corresponde a la distribución exponencial.

La parametrización que usamos en este Capítulo para incluir el caso anterior de la distribución de Pareto es la siguiente

$$f(x; \sigma, \xi) = \frac{1}{\sigma|\xi|} \left(1 + \frac{sx}{\sigma}\right), \quad (3.4)$$

donde

$$\sigma = \psi/|\xi|, \quad |\xi| = s\xi \quad \text{y} \quad s = \text{sign}(\xi)$$

La función de log verosimilitud de la distribución de Pareto generalizada viene dada por,

$$l(\sigma, \xi) = -\log(s\xi\sigma) - \frac{1 + \xi}{\xi n} \sum_{i=1}^n \log(1 + sx_i/\sigma). \quad (3.5)$$

Si  $\xi < 0$ ,  $\sigma > M = \max(x_i)$ , en caso contrario la función de verosimilitud es cero. Cuando  $\sigma$  tiende a  $M$ , el máximo verosimilitud tiende a infinito, por lo que se toma el máximo local de 3.5 porque a medida que  $\sigma$  se acerca a  $M$ , la función *profile* verosimilitud puede crecer de manera arbitraria.

Davison [Dav84] y Smith [Smi85] estudiaron la estimación de los parámetros mediante el máximo verosimilitud, en particular para muestras grandes. De hecho, el estimador es asintóticamente normal y es eficiente siempre que  $-0,5 < \xi$ .

Habitualmente, se asume la restricción  $-0,5 < \xi < 0,5$  debido a que el caso  $\xi < -0,5$  corresponde a las distribuciones de soporte compacto y el caso  $\xi > 0,5$  a las distribuciones con varianza infinita.

Cuando se utiliza la distribución GPD como alternativa a la distribución exponencial, los valores de  $\xi$  próximos a cero son de interés porque la distribución exponencial es un caso particular de la GPD cuando  $\xi = 0$ .

El coeficiente de variación teórico es,

$$\varsigma = \sqrt{1/(1 - 2\xi)}, \quad (3.6)$$

por lo tanto es mayor que uno, en cambio el coeficiente de variación empírico no siempre es mayor que uno sobretodo si las muestras son pequeñas .

Al derivar la expresión 3.5 respecto a  $\sigma$  obtenemos,

$$\xi(\sigma) \equiv \xi(\sigma, s) = \frac{1}{n} \sum_{i=1}^n \log(1 + sx_i/\sigma),$$

de este modo, la función *profile* verosimilitud viene dada por,

$$l_p(\sigma, s) = -\log[s\xi(\sigma)\sigma] - \xi(\sigma) - 1. \quad (3.7)$$

Observamos que cuando  $s = 1$ ,  $l_p(\sigma, s) = l_p(\sigma, 1) = l_p(\sigma)$  que corresponde a la función *profile* verosimilitud en el caso de la distribución de Pareto.

**Proposición 3.5.** *Sea  $l_p(\sigma, s)$  la función profile de verosimilitud definida por 3.7 de la variable aleatoria  $X$  y sea  $\bar{x}$  la media de una muestra de  $X$  entonces,*

$$\lim_{\sigma \rightarrow \infty} \xi(\sigma) = 0, \quad \lim_{\sigma \rightarrow \infty} \sigma\xi(\sigma) = s\bar{x} \quad y \quad \lim_{\sigma \rightarrow \infty} l_p(\sigma, s) = -\log(\bar{x}) - 1 \equiv l_0.$$

*Demostración.* Si utilizando la Proposición 3.3 y la Proposición 3.2, podemos probar que

$$\lim_{\sigma \rightarrow \infty} \sigma\xi(\sigma) = \lim_{\sigma \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma \log(1 + sx_i/\sigma) = s\bar{x}.$$

□

**Proposición 3.6.** *Sea  $\{x_i\}$  una muestra con función de distribución GPD de soporte acotado, sea  $M = \max(x_i)$  entonces,*

$$\lim_{\sigma \rightarrow M} \xi(\sigma) = \infty, \quad \lim_{\sigma \rightarrow M} l_p(\sigma) = \infty.$$

*Demostración.* en este caso, podemos expresar  $\xi(\sigma)$  de la manera siguiente,

$$\xi(\sigma) = \frac{1}{n} \sum_{i=1}^{n-1} \log(1 - x_i/\sigma) - \frac{1}{n} (1 - M/\sigma),$$

cuando  $\sigma$  tiende a  $M$ , sólo afecta el último termino.  $\square$

De manera similar que en el caso de la distribución de Pareto cuando  $\xi > 0$ , depende de si el coeficiente de variación es menor o mayor que uno, la función *profile* verosimilitud es una función monótona creciente o decreciente respectivamente.

**Teorema 3.2.** *Para las distribuciones GPD con  $\xi < 0$ , si el  $cv > 1$ ,  $l_p(\sigma, -1)$  es una función monótona creciente y si el  $cv < 1$ ,  $l_p(\sigma, -1)$  es una función monótona decreciente para  $\sigma$  suficientemente grande.*

*Demostración.* En este caso, consideramos la expresión,

$$-sl'_p(\sigma, s) = (\xi(\sigma) + \sigma\xi'(\sigma) + \sigma\xi(\sigma)\xi'(\sigma))/(\sigma|\xi(\sigma)|),$$

y utilizamos los mismos pasos que el Teorema 3.1.  $\square$

**Observación 3.2.** *Si  $l_p(\sigma, -1)$  tiene un mínimo local y el coeficiente de variación empírico es menor que uno, también tendrá un máximo local debido a que  $l_p(\sigma, -1)$  crece en el mínimo y es monótona decreciente para  $\sigma$  grande.*

Como consecuencia del Teorema 3.2, dado que  $l_p(\sigma, s)$  tiende a 0 y es una función monótona para  $\sigma$  grande, dependiendo del  $cv$ ,  $l_p(\sigma, s)$  es mayor o menor que  $l_0$ .

**Proposición 3.7.** *Si el coeficiente de variación empírico  $cv > 1$ , entonces, para  $\sigma$  suficientemente grande tenemos,*

$$l_p(\sigma, 1) > l_0 > l_p(\sigma, -1),$$

*y si el coeficiente de variación empírico  $cv < 1$  entonces,*

$$l_p(\sigma, 1) < l_0 < l_p(\sigma, -1).$$

*Demostración.* Si  $cv > 1$ ,  $l_p(\sigma, 1)$  es una función monótona decreciente y si el  $cv < 1$ ,  $l_p(\sigma, -1)$  es una función monótona creciente para  $\sigma$  grande.  $\square$

**Observación 3.3.** *Si el coeficiente de variación empírico es mayor que uno, son más adecuadas las distribuciones GPD con  $\xi > 0$  y si el coeficiente de variación empírico es menor que uno, son más apropiadas las GPD con  $\xi < 0$ .*

Hosking y Wallis [HW87] indicaron que en algunos casos no existía el máximo local. De hecho, podemos confirmar que si el coeficiente de variación empírico es menor que uno, la función *profile* verosimilitud de la distribución GPD puede no tener máximo. El ejemplo que presentamos a continuación ilustra este caso.

**Ejemplo 3.3.1.** Consideramos una muestra de tamaño dos,  $\{x_1, x_2\} = \{1, 2\}$ . Probaremos que el máximo verosimilitud no existe en el caso de la distribución de Pareto, para ello vamos a demostrar que la función *profile* verosimilitud  $l_p(\sigma) = l_p(\sigma, 1)$  es una función monótona para  $\sigma > 0$ .

La derivada de la función *profile* verosimilitud viene dada por,

$$num(x_1, x_2) = 2x_1x_2 + 2\bar{x}\sigma - (\bar{x} + \sigma)(\sigma \log(1 + x_1/\sigma) + \sigma \log(1 + x_2/\sigma)), \quad (3.8)$$

con denominador una función positiva, entonces  $l_p(\sigma)$  y  $num(x_1, x_2)$  tienen el mismo signo.

Tenemos que comprobar que  $num(x_1, x_2)$  es positivo. La expresión 3.8 puede expresarse como,

$$num(x_1, x_2) = 2x_1x_2 - \bar{x}(f_{x_1} + f_{x_2}) + (g_{x_1} + g_{x_2}),$$

donde  $f_x = \sigma \log(1 + x/\sigma)$  y  $g_x = \sigma(x - f_x)$ .

En lo que sigue, vamos a comprobar que  $f_x$  y  $g_x$  son funciones monótonos crecientes, los resultados en el caso de  $f_x$  son los siguientes,

$$\lim_{\sigma \rightarrow \infty} f_x = x, \quad \lim_{\sigma \rightarrow \infty} f'_x = 0, \quad f''_x = \frac{-x^2}{\sigma(\sigma + x)^2}.$$

Por lo tanto,  $f'_x$  es una función monótona creciente y  $f'_x > 0$  con lo cual,  $f_x$  es una función monótona creciente y  $f_x < x$ . De manera similar, los resultados en el caso de la función  $g_x$  son los siguientes,

$$\lim_{\sigma \rightarrow \infty} g'_x = x, \quad \lim_{\sigma \rightarrow \infty} g''_x = 0, \quad g'''_x = \frac{2x^3}{\sigma(\sigma + x)^3}.$$

Entonces, la segunda derivada  $g''_x$  es una función monótona creciente y  $g''_x < 0$ . Con lo que,  $g'_x$  es una función monótona creciente y  $g'_x > 0$ , por lo tanto,  $g_x$  es una función monótona creciente.

Finalmente sólo falta probar que,

$$3(f_1 + f_2)/2 - 4 < g_1 + g_2,$$

o de manera equivalente que,

$$(3/2 + \sigma)(f_1 + f_2) < 4 + 3\sigma. \quad (3.9)$$

Para ello, primero suponemos que  $\sigma < 4$ , en este caso es fácil comprobar que

$$(f_1 + f_2) < 2,52,$$

con lo cual obtendremos el resultado 3.9. Ahora si  $\sigma > 4$  entonces, podemos verificar que

$$3(f_1 + f_2)/2 - 4 < 0,5 < 1,94 < (g_1 + g_2).$$

Por lo tanto,  $num(1, 2)$  es positivo y  $l_p(\sigma)$  es una función monótona creciente para cualquier  $\sigma$ .

De manera similar con la misma muestra podemos demostrar que la función *profile* verosimilitud no tiene máximo local [CD09a].

Seguidamente, indicamos los problemas que presenta la GPD por contener distribuciones de soporte acotado.

Simulamos datos de la distribución GPD con  $\xi > 0$  de tamaño  $n = 15, 25, 50, 100$  con parámetro del peso  $\xi = 0,1, 0,2, 0,3, 0,4$  y con parámetro de escala 1 dado que el modelo 3.4 es invariante por cambio de escala. Para ello, para cada combinación de  $n$  y  $\xi$ , generamos 50.000 datos de distribución Pareto que corresponde a la distribución GPD con  $\xi > 0$ . A continuación haremos un recuento de los  $\hat{\xi}$  positivos, los  $\hat{\xi}$  negativos y también un recuento de las veces que no converge el algoritmo. En la Tabla 3.1, presentamos los resultados.

Observamos que en el caso  $\xi = 0,3$  que son distribuciones que tienen curtosis infinita, para  $n = 15$ , el 29% de los casos llevan a la decisión errónea  $\xi < 0$ , y el 4,8% de los casos no convergen. Este problema se mantiene en las muestras grandes, para  $\xi = 0,1$  y  $n = 100$ , el 24,9% de los casos llevan a  $\xi < 0$ .

$\xi$	0.4			0.3			0.2			0.1		
$n$	$\hat{\xi} > 0$	$\hat{\xi} < 0$	Nc	$\hat{\xi} > 0$	$\hat{\xi} < 0$	Nc	$\hat{\xi} > 0$	$\hat{\xi} < 0$	Nc	$\hat{\xi} > 0$	$\hat{\xi} < 0$	Nc
15	73.4	23.1	3.5	66.2	29.0	4.8	57.1	36.4	6.5	46.8	44.5	8.7
25	84.5	15.3	0.2	77.2	22.4	0.3	66.9	32.6	0.5	53.5	45.7	0.9
50	95.8	4.2	0.0	90.7	9.3	0.0	80.5	19.5	0.0	63.8	36.2	0.0
100	99.6	0.4	0.0	98.0	2.0	0.0	92.3	7.7	0.0	75.1	24.9	0.0

Cuadro 3.1: Recuento de los valores de  $\hat{\xi}$  positivos, negativos y de las veces en las cuales no hay convergencia

A diferencia de lo anterior y sin considerar el caso GPD de las distribuciones de soporte acotado, simulando datos de la distribución de Pareto, las muestras con el  $cv > 1$  llevan a la distribución de Pareto y las muestras con  $cv < 1$  llevan a la distribución exponencial, las dos distribuciones tienen soporte en  $(0, \infty)$ .

Asumiendo que las distribuciones de colas pesadas tienen soporte en  $(0, \infty)$ , un modelo alternativo para las muestras que presentan el  $cv$  menor que uno es la distribución normal truncada, para las muestras con el  $cv > 1$ , la distribución de Pareto, el caso límite es la distribución exponencial con el  $cv = 1$ .

En la próxima sección proponemos el modelo TNP que sólo depende del coeficiente de variación y que es la unión de las tres distribuciones anteriores.

### 3.4. Modelo alternativo TNP

La distribución condicionada de cualquier variable aleatoria para un umbral suficientemente elevado es una distribución GPD [Pic75], pero en la práctica el umbral se elige suficientemente alto restringiendo el estudio a pocos datos cuando el modelo básico en finanzas corresponde a la distribución normal [Jor05]y [BO02]. Lo más natural es considerar un modelo de las colas que incluye la distribución normal.

Por un lado, el comportamiento de las colas de la distribución GPD es muy diferente de un modelo a otro, además la distribución GPD con  $\xi < 0$  al incluir distribuciones de soporte acotado no es adecuada para la modelización en finanzas.

Por otro, desde un punto de vista práctico, los modelos asimétricos son de mucha importancia, porque excepto los datos de tipo de cambio, los datos financieros siempre presentan el “leverage effect” es decir los datos negativos presentan más impacto que los datos positivos [BNS01].

Por todo lo anterior, sugerimos la unión de la distribución de Pareto y de la distribución normal truncada en un modelo que denominamos TNP para cada valor del coeficiente de variación definido en  $(0, \infty)$ .

Para cada valor del parámetro de escala  $\sigma$  y del coeficiente de variación  $0 < cv$ , tenemos diferentes modelos para los valores no negativos.

El modelo propuesto viene dado por,

$$TNP(x, \sigma, cv) = \begin{cases} Par(\alpha, \sigma) & cv > 1 \\ Exp(\sigma) & cv = 1 \\ TN(\alpha, \sigma) & cv < 1 \end{cases}$$

donde,  $TN$  es la función de densidad de la distribución normal truncada definida en el Capítulo 2.

Entonces, tenemos una representación del modelo para cada valor del coeficiente de variación.

Las distribuciones Pareto y la normal truncada son complementarias a través del coeficiente de variación. Para la primera el coeficiente de variación es mayor que uno, para la segunda el coeficiente de variación es menor que uno, el caso límite corresponde a la exponencial con el coeficiente de variación igual a uno.

**Proposición 3.8.** *El modelo  $TNP(cv, \sigma)$  está bien definido y es invariante por cambio de escala.*

*Demostración.* Las tres distribuciones,  $Par(\alpha, \sigma)$ ,  $TN(\mu, \sigma^2)$  y la distribución exponencial  $Exp(\sigma)$  están bien definidas. Los diferentes parámetros corresponden a los tres modelos.  $cv = 1$  corresponde a la distribución exponencial y  $cv < 1$  corresponde a

la distribución normal truncada y  $cv > 1$  corresponde a la distribución de Pareto. Además, el modelo es invariante por cambio de escala  $\sigma$ . Si  $X \sim TNP(cv, \sigma)$  entonces  $\lambda X \sim TNP(cv, \lambda\sigma)$   $\square$

**Proposición 3.9.** *El estimador de máxima verosimilitud del modelo  $TNP(\zeta, \sigma)$  existe, es asintóticamente normal y es eficiente.*

*Demostración.* En el caso de  $TNP(\zeta, \sigma)$  con  $\zeta > 1$ , el resultado viene dado por [Smi85] y el caso  $TNP(\zeta, \sigma)$  con  $\zeta < 1$  es un resultado clásico de las familias exponenciales [Cas02]  $\square$

Si hay simetría de los datos respecto a cero, podemos aplicar el modelo directamente al valor absoluto de los datos, en caso contrario, siempre podemos incluir un parámetro de localización.

### 3.5. Ejemplo de aplicación

El ejemplo que presentamos a continuación corresponde a los datos diarios de tipo de cambio entre el 1 de enero 1999 hasta el 31 de diciembre 2002.

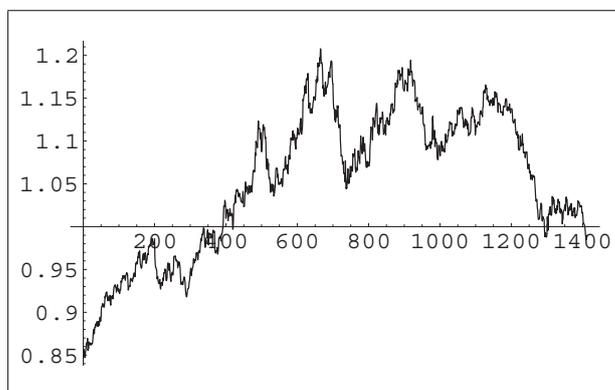


Figura 3.1: Evolución de los precios (lo que vale un Dólar en Euro)

A continuación, calculamos los estadísticos de los datos de tipo de cambio Dólar-Euro, Dólar-Yen y Dólar-Libra.

Observamos que la media de los valores positivos es igual a la media de los valores negativos. Las dos medias son iguales a la media de los valores absolutos, además, la asimetría es pequeña. Con lo que, podemos suponer simetría de los datos.

1999-2002	Dólar-Euro	Dólar-Yen	Dólar-Libra
media(abs)	0.00383	0.00373	0.00273
cv(abs)	1.06013	1.09966	1.10425
media(pos)	0.00383	0.00374	0.00275
media(neg)	0.00382	0.00372	0.00271
cv(pos)	1.08874	1.12283	1.11910
cv(neg)	1.03124	1.06841	1.08979
número de datos	1460	1460	1460

Cuadro 3.2: Estadísticos de los datos de tipo de cambio positivos y negativos por separado, Dólar-Euro, Dólar-Yen y Dólar-Libra

Euro	diarios	dos días	tres días	semanales
media	-0.0001	-0.0002	-0.0002	-0.0006
sd	0.0056	0.0079	0.0096	0.0146
curtosis	2.0594	1.8642	0.6415	0.0604
media(abs)	0.0038	0.0056	0.0074	0.0117
cv(abs)	1.0601	0.9760	0.8278	0.7430
número datos	1460	729	486	208

Cuadro 3.3: Estadísticos de los datos Dólar-Euro diarios, Dólar-Euro cada dos días, Dólar-Euro cada tres días y Dólar- Euro semanales

Datos diarios		Datos cada dos días		Datos cada tres días	
Mix-normales		Mix-normales		Mix-normales	
u	0.57980	u	0.74997	u	0.35570
p	0.54715	p	0.77092	p	0.28776
TNP	Pareto	TNP	TrunNor	TNP	TrunNor
$\sigma$	0.06574	$\mu$	-0.20080	$\mu$	-0.00921
$\alpha$	18.14629	$\sigma$	0.03457	$\sigma$	0.01262

Cuadro 3.4: Estimación de los parámetros de la mixtura de normales, Pareto y normal truncada

Observamos que el coeficiente de variación de los datos diarios positivos y de los datos diarios negativos es mayor que uno. Además, la curtosis es mayor que cero, lo cual indica que los datos no son normales.

El coeficiente de variación de los datos cada dos días y cada tres días tiende al coeficiente de variación de la distribución normal truncada. De manera similar, la curtosis de los datos cada dos días y cada tres días tiende a cero, lo cual indica normalidad de los datos a medida que aumenta el tiempo de diarios a semanales.

A continuación, estimamos los parámetros de la distribución empírica de los datos mediante la distribución mixtura de normales, la distribución normal truncada y la distribución de Pareto.

Para estudiar los valores negativos, consideramos  $x_{-}(-x)_{+}$ , con lo cual nos reducimos al estudio de las variables positivas.

Bajo la hipótesis de simetría que suponemos a continuación, la distribución normal se define por la desviación estándar, la distribución de Laplace se define por la media de los positivos. La cual es un caso particular de la distribución TNP cuando el coeficiente es igual a uno. La distribución normal truncada y la distribución de Pareto se determinan por la media y el coeficiente de variación de los datos positivos. La distribución mixtura de normales definida en el Capítulo 2, se define por el parámetro del peso y

los parámetros  $u$  y  $v$  que relacionan la varianza teórica con la correspondiente empírica. Dicha distribución en las colas decrece de manera rápida igual que la distribución normal pero su curtosis es significativamente elevada respecto a la normal.

Recordemos que cuando hay simetría de los datos, la distribución de las log-rentabilidades se define a partir de la distribución de los valores positivos, que denominamos por  $F_{ab}$ , de la siguiente manera,

$$F(x) = \begin{cases} \bar{F}_{ab}(-x)/2 & x < 0 \\ 1/2 + F_{ab}(x)/2 & x \geq 0 \end{cases}$$

A continuación, para valorar la bondad del ajuste de los datos diarios, realizamos el test  $\chi^2$ . Las distribuciones consideradas son la distribución de Pareto, la distribución mixtura de normales, la distribución de Laplace y la distribución normal.

Observamos que en el caso de los datos diarios, la distribución de Laplace es el modelo que arroja el p-valor más grande, si bien no tenemos evidencias para rechazar la distribución de Pareto.

De manera similar, en el caso de los datos cada dos días, la distribución de Laplace es la que otorga el mejor ajuste, no obstante, no tenemos evidencias para rechazar la distribución normal truncada.

La distribución de Laplace es la más apropiada en el caso de los datos diarios y en el caso de los datos cada dos días porque haciendo el test de exponencialidad utilizando el estadístico de Grennwood, se acepta la exponencialidad de los datos.

Para los datos cada tres días, el cv es igual a 0.828, dicho valor está entre el valor teórico de la distribución exponencial ( $\zeta = 1$ ) y el coeficiente de variación de la normal truncada ( $\zeta = 0,755$ ). En este caso, no se acepta la  $H_0$  de los datos. No tenemos evidencias para rechazar las cuatro distribuciones, si bien los valores centrales se infra-estiman por la distribución normal y se sobre-estiman por la distribución de Laplace. En términos de  $\chi^2$ , la distribución  $TN$  es la que otorga el mejor ajusta.

Puntos de corte	Observados	Pareto	Mix_Nor	Laplace	Normal
-4.0	2	3.6	1.0	2.1	0.0
-3.5	2	2.9	2.1	2.3	0.3
-3.0	5	5.4	5.4	4.8	1.6
-2.5	13	10.3	12.3	9.9	7.1
-2.0	23	20.1	24.6	20.5	24.1
-1.5	55	40.5	45.8	42.4	64.3
-1.0	79	83.4	93.5	88.0	134.1
-0.5	175	176.9	205.1	182.3	218.8
0.0	394	387.0	340.3	377.8	279.5
0.5	386	387.0	340.3	377.8	279.5
1.0	135	176.9	205.1	182.3	218.8
1.5	97	83.4	93.5	88.0	134.1
2.0	48	40.5	45.8	42.4	64.3
2.5	23	20.1	24.6	20.5	24.1
3.0	15	10.3	12.3	9.9	7.1
3.5	3	5.4	5.4	4.8	1.6
4.0	2	2.9	2.1	2.3	0.3
>4.0	3	3.6	1.0	2.1	0.0
chi-square statistic		25.27	54.82	25.07	479.59
Número de parámetros		2	3	1	1
P-value		0.0465-0.088	0.0000	0.0686-0.093	0.0000

Cuadro 3.5: Frecuencias observadas y estimadas de los datos de tipo de cambio diarios, los puntos de corte se expresan en función de  $\sigma$ . Las distribuciones por las que ajustamos son: La distribución de Pareto, la distribución mixtura de normales, la distribución de Laplace y la distribución normal.

Puntos de corte	Observados	TruncNor	Mix_Nor	Laplace	Normal
-4.0	0	1.0	1.1	1.4	0.0
-3.5	2	1.2	1.3	1.4	0.1
-3.0	2	2.6	2.5	2.8	0.8
-2.5	4	5.4	4.5	5.5	3.5
-2.0	10	11.2	8.9	11.2	12.0
-1.5	25	23.0	21.5	22.5	32.0
-1.0	50	46.4	54.1	45.3	66.9
-0.5	105	92.3	110.2	91.1	109.3
0.0	203	181.4	160.4	183.5	139.8
0.5	158	181.4	160.4	183.5	139.8
1.0	69	92.3	110.2	91.1	109.3
1.5	51	46.4	54.1	45.3	66.9
2.0	28	23.0	21.5	22.5	32.0
2.5	9	11.2	8.9	11.2	12.0
3.0	5	5.4	4.5	5.5	3.5
3.5	6	2.6	2.5	2.8	0.8
4.0	1	1.2	1.3	1.4	0.1
>4.0	1	1.0	1.1	1.4	0.0
chi-square statistic		22.45	36.78	22.85	164.73
Número de parámetros		2	3	1	1
P-value		0.0966-0.1680	0.0008-0.0036	0.1178-0.1541	0.0000

Cuadro 3.6: Frecuencias observadas y estimadas de los datos de tipo de cambio cada dos días, los puntos de corte se expresan en función de  $\sigma$ . Las distribuciones por las que ajustamos son: La distribución normal truncada, la distribución mixtura de normales, la distribución de Laplace y la distribución normal.

Puntos de corte	Observados	TruncNor	Mix_Nor	Laplace	Normal
-4.0	0	0.1	0.1	1.3	0.0
-3.5	0	0.3	0.4	1.2	0.1
-3.0	1	1.0	1.3	2.4	0.5
-2.5	2	3.1	3.8	4.5	2.4
-2.0	10	8.4	9.3	8.6	8.0
-1.5	16	19.4	19.3	16.5	21.4
-1.0	41	39.1	33.7	31.7	44.6
-0.5	77	68.2	58.8	60.6	72.8
0.0	109	103.4	116.4	116.1	93.1
0.5	101	103.4	116.4	116.1	93.1
1.0	64	68.2	58.8	60.6	72.8
1.5	35	39.1	33.7	31.7	44.6
2.0	13	19.4	19.3	16.5	21.4
2.5	10	8.4	9.3	8.6	8.0
3.0	5	3.1	3.8	4.5	2.4
3.5	2	1.0	1.3	2.4	0.5
4.0	0	0.3	0.4	1.2	0.1
>4.0	0	0.1	0.1	1.3	0.0
chi-square statistic		8.85	15.55	18.74	20.26
Número de parámetros		2	3	1	1
P-value		0.8852-0.9448	0.3417-0.5559	0.2822-0.3436	0.2086-0.2611

Cuadro 3.7: Frecuencias observadas y estimadas de los datos de tipo de cambio cada tres días, los puntos de corte se expresan en función de  $\sigma$ . Las distribuciones por las que ajustamos son: La distribución normal truncada, la distribución mixtura de normales, la distribución de Laplace y la distribución normal.

## Capítulo 4

# El coeficiente de variación en el estudio de las colas

Se sabe que la media condicionada sobre un umbral arbitrario determina de manera única cualquier función de distribución. La extensión de este resultado al caso multivariante fue demostrada en [JM82]. El coeficiente de variación condicionado se puede expresar en función de la media condicionada [GK00]. De hecho, el coeficiente de variación también caracteriza la función de distribución. Este resultado fue establecido por Sullo y Rutherford [SR77]. En este contexto, si  $x$  es una variable aleatoria de función de distribución exponencial. La distribución de los valores que exceden un determinado umbral también es una exponencial con la misma esperanza. La distribución de Pareto generalizada, que incluye la distribución exponencial, es el caso simple con el coeficiente de variación condicionado constante. Este resultado nos permite relacionar la teoría del valor extremo con el coeficiente de variación condicionado. Las aplicaciones de la teoría del valor extremo a la gestión del riesgo es cada vez de más interés. La referencia de [EKM97] nos facilita una revisión a la teoría matemática con una orientación hacia usos en seguros y finanzas.

Por otro lado, la teoría clásica del valor extremo aproxima, la distribución del máximo por la distribución GEV y la distribución de los valores extremos por la distribución de Pareto generalizada. Las dos aproximaciones están relacionadas por la idea de Pickands [Pic75] y [FR03]. Sin embargo, aunque en finanzas se dispone de un gran número de datos, la teoría clásica utiliza pocos datos (máximos o valores extremos). Nuestro punto de vista para estudiar las colas es utilizar tantos datos como sea posible y no sólo los

máximos o los valores extremos. Sugerimos utilizar toda la información siempre que la precisión asintótica sea similar. En este Capítulo, desarrollamos una metodología previa y complementaria a los estudios paramétricos para contrastar el modelo desde un punto de vista empírico. Es un método alternativo al ME-Plot y al Hill-Plot para estudiar el comportamiento de las colas y estimar el parámetro del peso de la cola. Por otro lado, proponemos un test de exponencialidad contra las alternativas de cola Pareto a través del coeficiente de variación condicionado.

## 4.1. Coeficiente de variación

La curtosis es la medida clásica para estudiar la normalidad de los datos. Sea  $x$  una variable aleatoria de media  $\mu$  y varianza  $\sigma^2$ , el exceso curtosis viene definido por

$$ku(x) = E((x - \mu)^4)/\sigma^4 - 3.$$

En el caso de la distribución normal, el exceso curtosis es igual a cero. Una curtosis más elevada que la de la distribución normal se debe a la concentración de los datos alrededor de la media “inliers” o a los valores extremos que no se pueden ajustar a la distribución normal.

Por otro lado, el coeficiente de variación se utiliza para comparar la variabilidad de los datos no negativos respecto a la media, viene definido por  $\sigma/\mu$  donde  $\sigma$  es finita y  $\mu \neq 0$ .

El coeficiente de variación es invariante por cambio de escala, en el sentido de que si cada observación es multiplicada por una constante el valor del coeficiente es el mismo.

**Proposición 4.1.** *Dada una variable aleatoria  $x$  de media  $\mu$  y los primeros cuatro momentos finitos, la curtosis puede expresarse en función del coeficiente de variación de la siguiente manera.*

$$ku(x) + 3 = 1 + cv^2(y), \text{ donde } y = (x - \mu)^2.$$

Entonces, toda la información dada por la curtosis la podemos obtener mediante el coeficiente de variación.

Mandelbrot y otros [Man63] y [FR71] estudiaron la distribución empírica de los datos financieros. El exceso curtosis de estos datos respecto a la normal obliga a recurrir a

otras distribuciones de colas pesadas o semi pesadas. Lo mismo lo podemos confirmar basándonos en el coeficiente de variación de  $y$ .

La medida habitual para estudiar las colas es la curtosis basada en el momento de orden cuatro. El coeficiente de variación dado que sólo depende de los dos primeros momentos es una herramienta adecuada frente a la curtosis. Es de mucha utilidad en las situaciones donde la existencia del momento de orden cuatro no es segura, como las descritas por diversos autores [Chi64].

**Proposición 4.2.** *Sea  $x$  una variable aleatoria tal que  $\mu_k = E[x^k] < \infty$  para  $k < 4$ , la distribución asintótica del coeficiente de variación empírico es*

$$\sqrt{n}(cv - \zeta) \rightarrow N(0, \sigma_a^2),$$

donde  $\zeta = \zeta[x]$  es el coeficiente de variación teórico y  $\sigma_a^2$  es la varianza asintótica

$$\sigma_a^2 = (4\mu_2^3 - 4\mu_1\mu_2\mu_3 + \mu_1^2(\mu_4 - \mu_2^2))/\mu_1^6.$$

*Demostración.* Sean  $\mu_1$  y  $\mu_2$  los dos primeros momentos. Con el teorema central del límite, la distribución asintótica de  $(\mu_1, \mu_2)$  es una distribución normal. Aplicando el método delta a  $(\mu_2 - \mu_1^2)\mu_1^2$ , el resultado es una consecuencia directa.  $\square$

El resultado de la Proposición 4.2 puede ser utilizado para diversas distribuciones. En el caso de la distribución exponencial, la distribución asintótica es

$$\sqrt{n}(cv - 1) \rightarrow N(0, 1). \tag{4.1}$$

Sea  $x$  una variable aleatoria de distribución normal de media cero. La parte positiva  $x_+ = (x|x > 0)$  tiene el coeficiente de variación igual a  $\zeta[x_+] = (\pi - 2)/2$  y la varianza igual a

$$\pi^2(\pi - 3)/4(\pi - 2).$$

Entonces, podemos contrastar la hipótesis nula de que  $x$  es una normal truncada utilizando la Proposición 4.2.

En el caso de la distribución de Pareto, suponiendo que los cuatro primeros momentos son finitos, es decir ( $\alpha > 4$ ). El coeficiente de variación es igual a  $\zeta = \alpha/(\alpha - 2)$  y la varianza asintótica es

$$\sigma_a^2 = \frac{(\alpha - 1)^2(\alpha^2 - \alpha + 6)}{(\alpha - 2)^2(\alpha - 3)(\alpha - 4)}.$$

Este resultado permite contrastar la hipótesis de que la distribución es Pareto además permite aproximar la potencia del test de exponencialidad contra la alternativa distribución de Pareto.

$$pow(cv_\xi, n) = 1 - N(\sqrt{n}(crt - cv_\xi)/\sigma_a),$$

donde  $crt$  son los valores críticos.

El coeficiente de variación no caracteriza de manera única la distribución exponencial, entonces si consideramos la expresión 4.4 y hacemos el test de exponencialidad mediante simulación no podemos detectar las alternativas que tienen el coeficiente igual a uno como es el caso de la distribución de Student de los valores positivos con 4 grados de libertad. Con lo cual, introducimos el coeficiente de variación condicionado.

### Coefficiente de variación condicionado

Comenzamos considerando la variable que describe el tiempo de vida restante cuando se ha vivido hasta un cierto momento. Sea  $x$  una variable que supongamos positiva con función de densidad  $f(x)$ . Denominamos por  $x_{|u} = (x - u | x > u)$  los valores que exceden el umbral  $u$ .

Si  $x$  representa el tiempo de vida de un individuo u de un objeto, la distribución de  $x_{|u}$  describe la vida de un individuo que ha vivido un tiempo  $u$ . De aquí vienen los nombres de supervivencia y fiabilidad, también de fiabilidad condicional o fiabilidad de  $x_{|u}$ .

**Proposición 4.3.** *Las funciones de fiabilidad y de densidad de  $x_u$ , vienen dadas por*

$$\bar{F}_u(x) = \bar{F}(u + x)/\bar{F}(u)$$

y

$$f_u(x) = f(u + x)/\bar{F}(u)$$

Podemos interpretar  $x_{|u}$  como el truncamiento de  $x$  a partir de  $u$  desplazado otra vez al origen. El coeficiente de variación de  $x_{|u}$  puede expresarse como

$$\zeta(x_{|u}) = \frac{(E(x_{|u})^2 - (E(x_{|u}))^2)^{1/2}}{E(x_{|u})}$$

donde  $E[.]$  es el valor esperado.

El resultado que presentamos a continuación relaciona el coeficiente de variación y las colas pesadas y nos permite proponer el coeficiente como una nueva herramienta para estudiar el comportamiento de las colas.

## 4.2. Caracterización a través del coeficiente de variación condicionado

El coeficiente de variación condicionado caracteriza de manera única cualquier distribución tanto en el caso univariante como en el caso bivariante [GK00]. Este estadístico puede expresarse en función de la media condicionada de la siguiente manera

$$\zeta^2(x - u | x > u) = \frac{2}{\overline{F}(x)} \int_u^\infty \overline{F}(x)(e(x) - e(u))dx / e^2(u),$$

donde  $e(u)$  es la media truncada en  $u$  definida en el Capítulo 1.

Como el coeficiente depende únicamente de la media truncada, la cual sólo depende de la cola superior. Entonces, el coeficiente condicionado puede expresarse en función de la cola superior, y por lo tanto nos permite representar el comportamiento de las colas.

### Distribuciones de variación regular

En las distribuciones de variación regular de índice  $-\alpha$ , la varianza de las distribuciones truncadas es asintóticamente creciente.

**Proposición 4.4.** *Sea  $x$  una variable aleatoria de función de distribución de variación regular de índice  $-\alpha$  y función de variación lenta  $L$ , entonces el comportamiento de la varianza condicionado denominada por  $\sigma^2(u)$  en el infinito es como*

$$u^2\alpha/(\alpha - 1)^2(\alpha - 2)$$

y el comportamiento de  $\zeta^2(u)$  en el infinito es como  $\alpha/(\alpha - 2)$

*Demostración.* Sabemos que

$$\sigma^2(u) = \int_0^\infty (x - u)^2 f_u(x) dx - (E(x - u | x > u))^2,$$

y si aplicamos la Proposición 1.8 del Capítulo 1, obtenemos que

$$E(x - u | x > u) \sim u / (\alpha - 1)$$

donde

$$f_u(x) = \frac{f(x)}{P\{X > u\}} \quad \text{si } x > u$$

De hecho,

$$\sigma^2(u) = 2 \frac{1}{\bar{F}(u)} \int_u^\infty (x - u) \bar{F}(x) dx - \left( \frac{1}{\bar{F}(u)} \int_u^\infty x^{-\alpha} L(x) dx \right)^2$$

Ahora si aplicando el Teorema de Karámata a  $\sigma^2(u)$ , obtendríamos que  $\sigma^2(u)$  es asintóticamente equivalente a

$$u^2 \alpha / (\alpha - 1)^2 (\alpha - 2).$$

Por lo tanto, el coeficiente condicionado es asintóticamente equivalente a  $\alpha / (\alpha - 2)$ .  $\square$

Las distribuciones de variación regular de índice  $-\alpha > 2$  que tienen las colas más pesadas que la distribución exponencial en el sentido de que,

$$\bar{F}_1 > \bar{F}_2 \quad \text{sii} \quad \lim_{c \rightarrow \infty} \frac{\bar{F}_1(c)}{\bar{F}_2(c)} > 1$$

tienen el coeficiente de variación condicionada mayor que uno.

Por lo que, podemos identificar las distribuciones de colas pesadas por las que tienen el coeficiente de variación condicionado mayor que uno. Las distribuciones de colas ligeras por las que tienen el coeficiente condicionado menor que uno, el caso frontera es el caso de la exponencial con el coeficiente igual a uno. Un caso particular de las distribuciones de colas pesadas es el caso de la distribución de Pareto Generalizada, que además se caracteriza por el hecho de tener el coeficiente condicionado constante.

## Distribución GPD

La distribución de Pareto generalizada definida en el Capítulo 3, decrece de manera polinomial como  $x^{-1/\xi}$  y el parámetro de forma es  $\xi^{-1}$  y es el parámetro que nos indica el peso de la cola. Cuanto más grande sea  $\xi$  más pesada es la cola.

En la siguiente proposición, demostramos que la distribución de Pareto generalizada es la única con el coeficiente de variación condicionado constante.

**Proposición 4.5.** *Supongamos que  $x$  es una variable aleatoria continua con los dos primeros momentos finitos. Si el coeficiente de variación de  $x|_u$  por encima de cualquier umbral  $u > 0$  es constante  $\zeta(x|_u) = \zeta(x)$ , entonces, la función de riesgo es solución de la ecuación diferencial*

$$r' = -\xi r^2$$

donde  $\xi = (\zeta^2 - 1)/(2 * \zeta^2)$  y la única distribución que lo cumple es la distribución de Pareto generalizada.

*Demostración.* Denotamos por  $\mu_1$  y  $\mu_2$  los primeros momentos de  $x$ . De la igualdad  $\zeta(x|_u) = \zeta(x)$  obtenemos,

$$\mu_1^2 \bar{F}(u) \int_u^\infty (x-u)^2 f(x) dx = \mu_2 \left( \int_u^\infty (x-u) f(x) dx \right)^2, \quad (4.2)$$

al derivar 4.2 respecto a  $u$ , obtenemos

$$\int_u^\infty (x-u)^2 f(x) dx = 2 \frac{\mu_2 - \mu_1^2}{\mu_1^2} \int_u^\infty (x-u) f(x) \frac{\bar{F}(u)}{f(u)} dx, \quad (4.3)$$

y al sustituir el primer término de la expresión 4.3 por el valor correspondiente en la primera ecuación 4.2 obtenemos

$$2k \frac{\bar{F}(u)^2}{f(u)} = \int_u^\infty (x-u) f(x) dx,$$

donde  $k = (\mu_2 - \mu_1^2)/\mu_2 = \zeta^2/(1 + \zeta^2)$ . Al derivar de nuevo y dividir por  $\bar{F}^3$ . Obtenemos la ecuación diferencial siguiente,

$$r' = -\xi r^2,$$

donde  $\xi = (2k - 1)/2k$ . Por lo tanto,

$$\bar{F}(x) = \exp \{-R(x)\} = (1 + \xi x / \beta)^{-1/\xi}.$$

□

Como casos particulares de la distribución de Pareto generalizada, tenemos la distribución exponencial cuando  $\xi = 0$ , la distribución uniforme cuando  $\xi = -1$  y la distribución de Pareto cuando  $\xi > 0$ .

Esta caracterización ha sido demostrada por [GK00] que estudiaron la monotonidad de la varianza condicionada demostrando que  $\sigma^2(u)$  es una función creciente (decreciente) si y sólo si el coeficiente de variación es  $\geq$  ( $\leq$ )1 respectivamente. Luego, expresaron el coeficiente de variación en función de la esperanza condicionada, y como esta última determina las distribuciones de manera única, el coeficiente de variación también. Más adelante, veremos que este resultado nos permite decidir entre la distribución de Pareto con un comportamiento polinomial y la distribución exponencial. De una manera más general, el resultado nos permite contrastar la exponencialidad contra las familias de cola Pareto.

Antes de explicar el CV-Plot, vamos a describir el comportamiento de las colas.

La cola superior de una distribución tanto en finanzas como en seguros pertenece al dominio de atracción de la distribución de Fréchet o al dominio de atracción de la distribución de Gumbel véase [Fel81]. De hecho, utilizando la Proposición 1.8 del Capítulo 1, en el caso de las distribuciones de variación regular, los valores que exceden el umbral convergen a una distribución de Pareto generalizada, y los valores que exceden el umbral en las distribuciones que pertenecen al dominio de atracción de la distribución de Gumbel, convergen a la distribución exponencial. Por tanto, decidir entre las distribuciones de colas pesadas y las distribuciones de colas semi-pesadas se reduce a decidir entre la distribución de Pareto con un comportamiento polinomial y la distribución exponencial.

En la práctica, decidir entre colas polinomiales y exponenciales es equivalente a contrastar entre distribuciones con todos los momentos finitos y distribuciones con un número finito de momentos. El resultado que presentamos a continuación da una condición suficiente del comportamiento exponencial de las colas.

**Proposición 4.6.** *Sea  $x$  una variable aleatoria con función generatriz de momentos  $M(t)$ , la cola decrece exponencialmente.*

*Demostración.* La desigualdad de Chebychev permite demostrar

$$Pr\{x > \nu\} = Pr\{e^{tx} > e^{t\nu}\} \leq M(t)e^{-t\nu}$$

□

De hecho, el resultado de caracterización nos permite decidir entre las dos distribuciones basándonos en el coeficiente de variación condicionado. La distribución exponencial con  $\zeta = 1$  que corresponde a la hipótesis nula. La alternativa es la distribución de Pareto con el coeficiente de variación  $\zeta = (\alpha/(\alpha - 2))^{1/2} > 1$  para  $\alpha > 2$  es decir si los dos primeros momentos son finitos.

En la siguiente sección, proponemos un método que denominamos por CV-Plot alternativo a los métodos clásicos Hill-Plot y ME-Plot para estudiar el comportamiento de las colas.

### 4.3. Estimación de las colas. CV-Plot

Como indican [BT96], existen algunos problemas en la estimación de las colas de las distribuciones. En primer lugar, el problema de selección del umbral, precisamente, decidir cuantos datos de un conjunto son relevantes para estimar la cola. Luego, como hacer frente al considerable sesgo debido a la extrapolación y a la gran varianza debida a los pocos datos de los que habitualmente se dispone en este tipo de problemas. De hecho, el objetivo es buscar el umbral a partir del cual podemos ajustar los datos a una distribución de Pareto. Nuestra propuesta es elegir el umbral a través del coeficiente de variación condicionado.

Tal como se ha comentado anteriormente, Sullo y Rutherford [SR77] caracterizaron las distribuciones que tienen el coeficiente de variación condicionado constante. La distribución de Pareto generalizada es el único caso con el coeficiente de variación condicionado constante y la distribución exponencial es cuando el coeficiente de variación es igual a uno. Este resultado llamado resultado de caracterización nos permite estudiar las colas en función del coeficiente de variación y por lo tanto nos permite relacionar el coeficiente de variación con la teoría del valor extremo.

#### CV-Plot

El coeficiente de variación empírico de una muestra de tamaño  $n$ ,  $\{x_j\}$ , viene definido por

$$cv[\{x_j\}] = (\sum x_j^2/n - \bar{x}^2)^{1/2}/\bar{x}.$$

Dada una muestra  $\{x_j\}$  positiva. Denominamos por  $\{x_{(j)}\}$  los estadísticos de orden, tales que  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . El CV-Plot consiste en la representación gráfica

$$i \rightarrow cv[x_j - x_{(i)}; x_j > x_{(i)}], \text{ donde } i = 1, \dots, n_0 \quad (4.4)$$

donde  $n_0$  son potencias de dos,  $n_0 = 2^{11}$  es el valor que consideramos más grande de la muestra. Se trata de calcular el coeficiente de variación condicionado para diferentes umbrales y luego bajo la hipótesis nula de exponencialidad, calculamos los intervalos de confianza de los coeficientes de variación al 90% de las muestras de tamaño  $n_k = 2^k$ ,  $k = 5, \dots, 11$ .

Para definir umbrales aproximadamente iguales bajo la hipótesis nula de exponencialidad, consideramos el siguiente resultado.

**Proposición 4.7.** *Supongamos que  $x$  es una variable aleatoria con función de distribución exponencial de media uno. Si  $q_k = k \text{Log}2$ , entonces*

$$P\{x > q_k\} = 2^{-k}$$

Dada una muestra  $\{x_j\}$  de tamaño  $n_0$  con función de distribución exponencial. Los estadísticos de orden  $x_{(n_k)}$  correspondientes a la muestra  $n_k = n_0 2^{-k}$  son aproximadamente iguales a  $q_k$ . De hecho, los umbrales  $u_k = x_{(n_k)}$  tienen aproximadamente la misma distancia entre si.

Bajo la hipótesis de exponencialidad, calculamos los intervalos de confianza al 90% del coeficiente de variación para los diferentes tamaños de muestra  $2^k$  para  $k = 5, \dots, 11$ . Los intervalos de confianza se han calculado mediante simulación con muestras de tamaño 50.000. Los resultados vienen calculados en la Tabla 4.1.

Sea  $n_k = n_0 2^{-k}$  y denotamos por  $(a_k, b_k)$  los puntos críticos al 90% de las diferentes muestras  $n_k$ . Los puntos  $\{n_0 - n_k, a_k\}$  y  $(n_0 - n_k, b_k)$  para  $k = 0, \dots, 6$  están relacionados por interpolación lineal para tener una referencia de los datos empíricos y controlar las variabilidades que podemos esperar del CV-Plot.

cv	q	n=1024	n=512	n=256	n=128	n=64	n=32
1	0.95	1.050	1.071	1.100	1.139	1.183	1.232
1	0.05	0.950	0.929	0.901	0.864	0.811	0.741

Cuadro 4.1: puntos críticos bajo la  $H_0$

Si el gráfico se mantiene dentro de los límites del intervalo, decimos que los datos son exponenciales. En caso contrario, los datos son de cola Pareto.

Si La distribución de Pareto no se acepta para toda la muestra, el coeficiente de variación puede utilizarse para determinar el umbral  $u$  a partir del cual podemos ajustar los datos restantes mediante la distribución de Pareto. Nuestra propuesta es calcular el parámetro del peso de la cola mediante una ponderación de los coeficientes de variación por encima de los umbrales.

### Comparación del CV-Plot y el ME-Plot

La media de los excesos empírica definida en el Capítulo 1 (ME-Plot) se utiliza para aproximar la correspondiente teórica. la función exceso medio de una distribución de Pareto es lineal con pendiente positiva. La función empírica también lo es, entonces si la gráfica de la función exceso empírica es lineal podría pensarse de que los datos provienen de una distribución de Pareto.

Si una distribución tiene cola más pesada que la distribución exponencial, su función exceso medio es creciente. Mientras que para las distribuciones con colas ligeras la función exceso es decreciente. Por lo tanto, la función exceso proporciona una información importante acerca de la naturaleza de la distribución. Por otro lado, la función exceso medio permite encontrar el umbral a partir del cual los datos se podrían ajustar a una distribución de Pareto.

El problema de cara a la estimación reside en que la variabilidad de los gráficos de la función de exceso medio incrementa en las colas [EKM97]. Además como en general esta función es una media de un número pequeño de términos, en muchos casos, es muy inestable por la existencia de valores extremos. El problema se debe a que el umbral a

partir del cual el exceso se comporta como una distribución de Pareto es alto y hay pocos datos. Por lo que, es importante el enfoque del coeficiente de variación condicionado.

**Proposición 4.8.** *A diferencia de la media condicionada, el coeficiente de variación no depende del parámetro de escala y sólo depende de los dos primeros momentos.*

La media de los valores por encima de los umbrales es una estimación de la media condicionada. Esta estimación cambia linealmente en función de  $u$  y depende del parámetro de escala. De hecho, no resulta fácil interpretar el ME-Plot dado que en algunos casos no es obvio determinar el valor a partir del cual el ME-Plot es lineal. Además, el ME-Plot es muy sensible en los valores extremos y, con pocos datos por encima del umbral para los cuales el modelo puede ser estimado, implican una varianza elevada.

En cambio, una estimación del coeficiente de variación implica una estimación del parámetro del peso de la cola debido a que el coeficiente de variación es invariante por cambio de escala.

Simulando unos datos mediante la distribución de Student que es una distribución de variación regular, la distribución exponencial y la distribución de Pareto. Vamos a hacer una somera descripción del CV-Plot

## 4.4. Análisis exploratorio

La función de densidad de la distribución de Student viene definida por

$$f_n(x) = \frac{1}{\sqrt{\pi n}} \frac{1}{\beta(\frac{n}{2}, \frac{1}{2})} \left( \frac{n}{n+x^2} \right)^{(n+1)/2}.$$

La distribución de Student es de variación regular de índice  $-n$ .

Hemos tomado como ejemplo la distribución de Student de los valores positivos con 4 grados de libertad porque el coeficiente de variación es igual a uno, que es igual al coeficiente de variación de la distribución exponencial. Para ello, para distinguir entre las dos, consideramos el coeficiente de variación condicionado.

Hemos simulado una muestra de 3000 datos con una distribución de Student de parámetro 4, y otra muestra del mismo tamaño con una distribución de Student de

parámetro 3. Hemos realizado el CV-Plot de los 2048 valores mayores de los datos. Observamos que el CV-Plot de los valores absolutos de la distribución de Student con 4 grados de libertad empieza en 1 cuando se toman todos los datos y luego crece para los valores que superan un determinado umbral y al final se estabiliza. Esto es lógico porque la distribución de Student es una distribución de variación regular, por lo tanto la distribución límite por encima de los umbrales es una distribución de Pareto con coeficiente de variación  $\zeta = (n/(n-2))^{1/2}$ , es decir en el caso de la distribución de Student con  $n = 4$ , el CV-Plot se estabiliza en  $cv = 1,414$ . De manera similar, en el caso de la distribución de Student con tres grados de libertad. El CV-Plot se estabiliza en  $cv = 1,72$ .

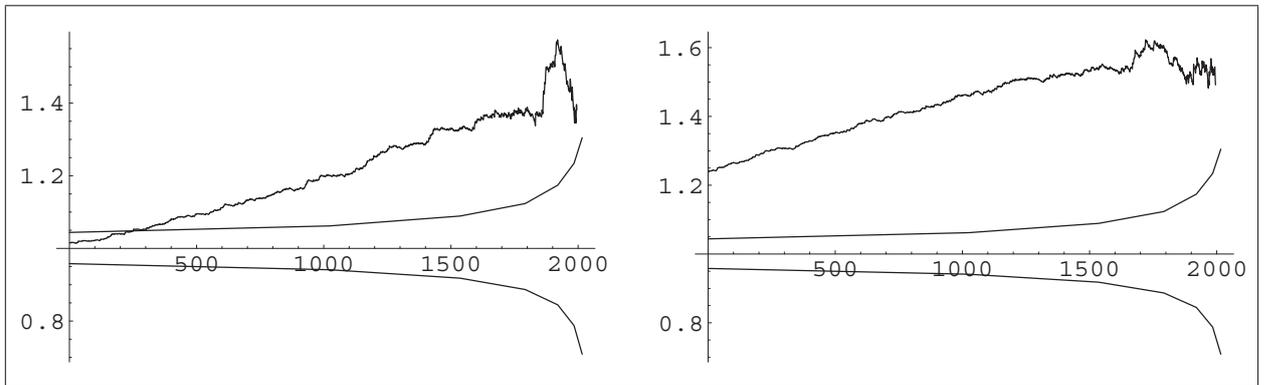


Figura 4.1: CV-Plot de la Student(4) y la Student(3) de derecha a izquierda

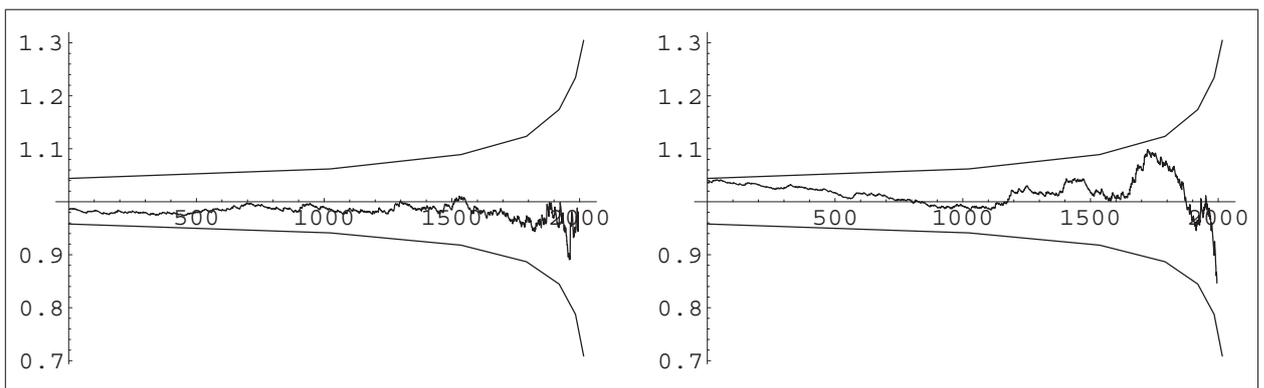


Figura 4.2: CV-Plot de la exponencial(1) y la Pareto con  $cv=1.03$  de derecha a izquierda

Por otra parte, hemos simulado una muestra de 3000 datos con una distribución exponencial de parámetro 1 y otra muestra del mismo tamaño con un distribución de Pareto de parámetro 1.03. Hemos elegido esta última porque una distribución de Pareto con el coeficiente de variación igual a uno es una distribución exponencial.

Observamos que en los dos casos, los datos siguen una distribución exponencial.

A continuación, estudiaremos los datos de tipo de cambio Dólar-Yen y Dólar-Libra desde el 1 de enero, 1979 hasta el 31 de diciembre, 2003.

Observamos que el coeficiente de variación de los datos por separado tanto positivos como negativos es mayor que uno. Remarcamos también que el coeficiente de variación del valor absoluto de los datos es mayor que uno. Tanto la curtosis como el coeficiente de variación indican que los datos no se pueden ajustar con la distribución normal.

En la Figura 4.3, calculamos el coeficiente de variación por encima de los umbrales y representamos el CV-Plot (el coeficiente de variación de los datos por encima de los umbrales y el intervalo de confianza bajo la hipótesis de exponencialidad).

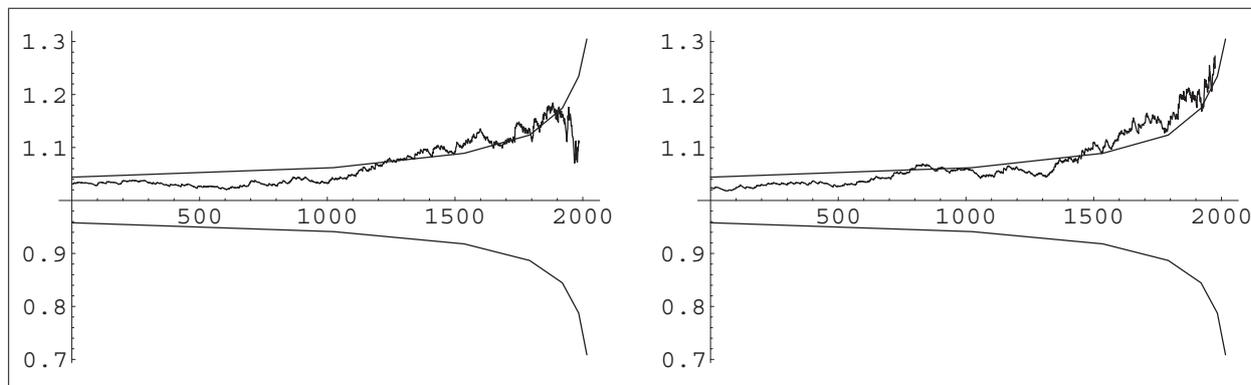


Figura 4.3: CV-Plot de Los datos Dólar-Libra positivos y negativos de derecha a izquierda

Observamos que los Dólar-Yen-positivos se mantienen dentro del intervalo de confianza, lo cual implica que estos datos son exponenciales. En cambio, se rechaza la hipótesis de exponencialidad en el caso de los Dólar- Yen negativos.

Seguidamente en la Tabla 4.3, calculamos el parámetro del peso de la cola para diferentes bases de datos para verificar a partir de que tamaño muestral se estabiliza el

años	coefneg	coefpos	coefabs	curtosisY	coefneg	coefpos	coefabs	CurtosisL
1979	1.5243	1.209	1.3497	6.8677	1.2787	1.3619	1.3210	5.698
1980	1.3051	1.365	1.3681	3.4208	1.3398	1.1951	1.2741	3.1719
1981	1.2632	1.1368	1.2157	2.8487	1.2051	1.2421	1.2278	3.6997
1982	1.248	1.139	1.1903	1.9022	1.167	1.1744	1.1745	1.2879
1983	1.2746	1.2604	1.2676	3.018	1.2146	1.2601	1.24	2.4844
1984	1.5495	1.1426	1.3467	9.1216	1.1881	1.1744	1.1724	1.7019
1985	1.3852	1.2803	1.4265	11.1911	1.3692	1.2212	1.3143	4.4843
1986	1.3164	1.311	1.3142	4.1486	1.1777	1.2313	1.2027	2.6934
1987	1.3395	1.2877	1.3386	3.8781	1.1998	1.3419	1.2558	3.0604
1988	1.2674	1.5837	1.4548	8.4963	1.263	1.2974	1.2867	3.6888
1989	1.5277	1.2985	1.4019	4.947	1.2681	1.3886	1.3368	4.5269
1990	0.9467	0.9314	0.9503	1.3828	1.1969	1.2709	1.2267	4.4657
1991	0.9912	0.9924	0.9966	2.1434	1.0573	0.9872	0.9966	1.9812
1992	1.1737	1.0201	1.1081	7.4994	0.9895	1.1985	1.1406	7.3678
1993	0.9653	1.0654	1.0101	3.2567	1.0333	0.9317	0.9813	1.4497
1994	1.0996	1.0041	1.0582	2.498	1.1177	1.2159	1.1378	4.1027
1995	1.2221	1.071	1.1473	3.8977	1.1569	1.1222	1.1473	2.7819
1996	1.1053	0.8757	0.9782	2.9714	1.0476	1.2105	1.1168	4.6241
1997	1.1747	0.8896	1.0366	5.5185	0.9224	0.9431	0.9343	2.201
1998	1.1629	0.8718	1.0731	9.9919	1.0982	0.8697	0.9946	2.8807
1999	1.019	0.9601	0.9907	2.655	1.1998	1.0027	1.1032	4.9931
2000	1.3099	1.0871	1.1889	6.2857	1.1389	1.049	1.0938	1.853
2001	1.1062	1.096	1.0977	2.2644	1.1439	1.0524	1.0960	1.7056
2002	1.101	1.1192	1.1090	3.7222	1.0547	1.1443	1.0938	2.6082
2003	0.9957	1.098	1.0418	2.555	0.99	1.0621	1.0217	2.0072

Cuadro 4.2: Coeficiente de variación y Curtosis de los datos Dólar-Yen y Dólar-Libra positivos y negativos por separado

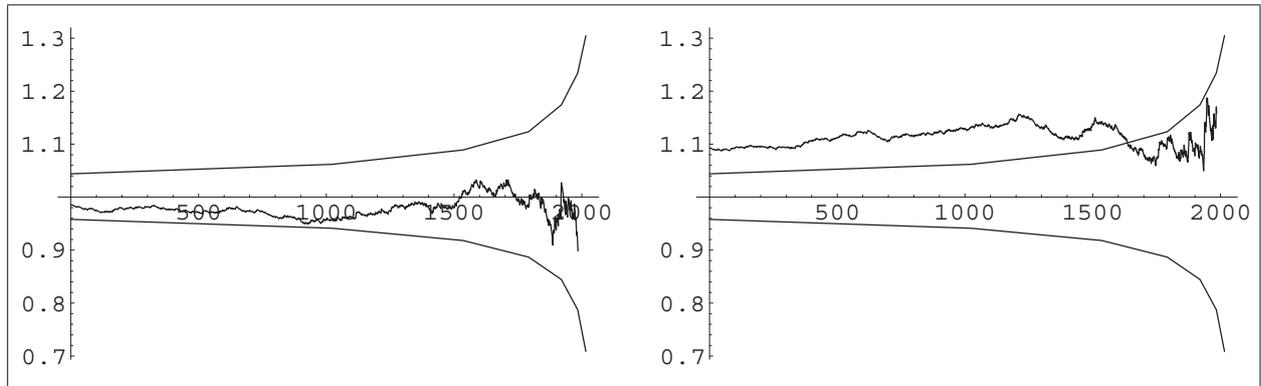


Figura 4.4: CV-Plot de los datos dólar-Yen positivos y negativos de derecha a izquierda

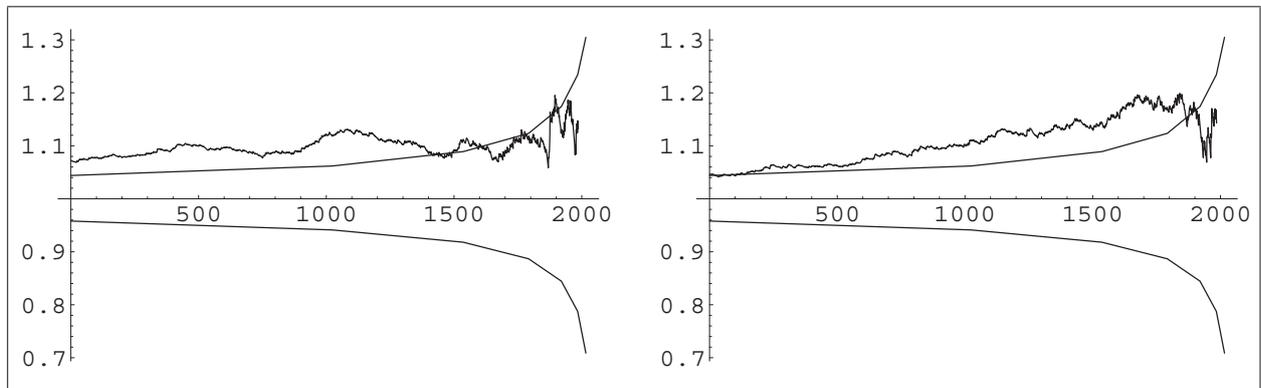


Figura 4.5: CV-Plot del valor absoluto de los Dólar-Yen y de los Dólar-Libra de derecha a izquierda

coeficiente de variación.

De la Tabla 4.3, observamos que los parámetros estimados de la distribución de Pareto para diferentes umbrales coinciden con los valores empíricos del coeficiente de variación condicionado.

Para los 64 o 32 valores mayores de la muestra, observamos que el coeficiente de variación no es estable. Se estabiliza a partir de los 128, 256 valores mayores, no obstante se pierde mucha información al no utilizar más datos. Nuestra propuesta es una ponderación de los coeficientes de variación por encima de los umbrales para los diferentes tamaños de muestra. Los resultados de los Dólar-Libra positivos, Dólar-Libra negativos, Dólar-Yen positivos y Dólar-Yen negativos respectivamente son

	32	64	128	256	512	1024
libra/pos	149.73 (1.01)	46.79 (1.03)	6.89 (1.15)	10.24 (1.12)	11.18 (1.11)	31.10 (1.042)
libra/neg	4.00 (1.22)	5.07 (1.22)	11.45 (1.18)	9.25 (1.10)	13.96 (1.05)	23.00 (1.02)
yen/pos	15.76 (1.05)	372.91 (0.9)	40.24 (1.02)	150.40 (0.99)	63.62 (1.01)	227.00 (0.95)
yen/neg	9.96 (1.18)	8.22 (1.16)	14.52 (1.1)	10.91 (1.11)	7.55 (1.14)	8.80 (1.13)

Cuadro 4.3: Estimación del parámetro de peso de la cola y el cv correspondiente

$$\{1,053, 1,06, 1,11, 0,98\}$$

Consideramos el estadístico,

$$T_m(cv_\epsilon) = \sum_{k=0}^m n_k (\zeta_k - cv_\epsilon)^2, \quad (4.5)$$

donde  $cv_\epsilon$  es el valor esperado de  $\zeta_k$ .

El coeficiente de variación ponderado de los coeficientes por encima de los umbrales se obtiene de manera directa calculando el valor  $cv_\epsilon$  que minimiza el estadístico  $T_m(cv_\epsilon)$ .

El estadístico  $T_m(cv_\epsilon)$  también puede utilizarse para contrastar la hipótesis nula de que los datos provienen de una distribución de Pareto.

En la próxima sección, haremos una comparación de los resultados mediante la distribución de Pareto por encima de los umbrales y la estimación mediante la distribución de valores extremos GEV.

## 4.5. Inferencia del peso de la cola y el coeficiente de variación

Tal como se ha comentado en el Capítulo 1, los parámetros de la distribución de Pareto generalizada están unívocamente determinados por la distribución del valor ex-

tremo GEV. Vamos a aplicar este resultado a los datos para comprobar si el parámetro del peso de la cola estimado de la distribución GEV y de la distribución de Pareto generalizada coinciden.

La función de verosimilitud de la función de distribución GEV viene dada por,

$$l(\mu, \sigma, \xi) = -n \log \sigma - \left(\frac{1}{\xi} + 1\right) \sum_i \log\left(1 + \xi \frac{y_i - \mu}{\sigma}\right) - \sum_i \left(1 + \xi \frac{y_i - \mu}{\sigma}\right)^{-1/\xi}$$

Hemos calculando numéricamente los estimadores de la GEV mediante el método de máxima verosimilitud. Utilizando los máximos anuales de los datos de tipo de cambio que presentamos en la Tabla 4.4, hemos obtenido los resultados del cuadro 4.5.

Observamos que depende de como se tome el umbral, los estimadores calculados de la distribución GEV y los calculados de la distribución de Pareto son o muy similares si el umbral es alto o muy diferentes en caso contrario. De hecho, es mejor utilizar todos los coeficientes de variación por encima de los umbrales mediante una ponderación de los mismos.

Observamos que en el caso de los Dólar-Yen positivos, el peso de la cola es infinito es decir  $\hat{\xi} = 0$  y la distribución que mejor ajusta estos datos es la distribución exponencial, lo cual coincide con lo que sugiere el CV-Plot.

El CV-Plot sugiere modelizar los datos Dólar-Yen negativos por la distribución de Pareto. Entonces, al estimar los parámetros por el método de máxima verosimilitud utilizando la muestra de tamaño  $n_0$  que es el tamaño muestral más grande  $2^{11}$ . Obtenemos  $\hat{\xi}^{-1} = 13,473$  y  $\hat{\mu} = 0,024$  y por lo tanto  $cv_\xi = 1,084$ . Este resultado es muy parecido al encontrado mediante el CV-Plot. De manera similar, estimamos los datos para los diferentes tamaños de muestra  $n_k$  y hallamos unos coeficientes de variación próximos a los encontrados mediante el CV-Plot que son los siguientes,

$$\{1.092, 1.132, 1.142, 1.112, 1.100, 1.168, 1.184, 1.611\}$$

Observamos que el coeficiente de variación de los  $2^4$  valores mayores de la muestra devuelve el valor  $cv_7 = 1,611$ , lo cual es alto pero aparece en la expresión de  $T_m(cv_\xi)$  con un peso bajo. Con lo que el efecto en la expresión de  $T_m(cv_\xi)$  es mínimo.

Año	Dólar-Yen		Dólar-Libra	
	Positivo	Negativo	Positivo	Negativo
1979	0.019	0.031	0.028	0.017
1980	0.019	0.027	0.014	0.018
1981	0.020	0.023	0.038	0.022
1982	0.020	0.025	0.017	0.016
1983	0.016	0.018	0.019	0.016
1984	0.009	0.022	0.019	0.019
1985	0.015	0.034	0.029	0.045
1986	0.020	0.030	0.024	0.019
1987	0.019	0.026	0.023	0.019
1988	0.033	0.021	0.025	0.024
1989	0.023	0.027	0.028	0.026
1990	0.020	0.018	0.019	0.029
1991	0.014	0.019	0.021	0.020
1992	0.014	0.029	0.049	0.023
1993	0.025	0.018	0.021	0.018
1994	0.015	0.022	0.015	0.015
1995	0.027	0.029	0.013	0.013
1996	0.012	0.018	0.014	0.011
1997	0.025	0.029	0.017	0.016
1998	0.024	0.062	0.010	0.017
1999	0.023	0.027	0.014	0.021
2000	0.018	0.030	0.014	0.018
2001	0.021	0.024	0.014	0.014
2002	0.020	0.027	0.015	0.014
2003	0.013	0.017	0.013	0.018

Cuadro 4.4: Los máximos anuales de los datos de tipo de cambio. Dólar-Yen, Dólar-Libra. Positivos y negativos por separado

	Yen/neg.	Yen/pos.	Libra/pos.	Libra/neg.
peso de la cola	5.48	infty	4.014	6.44
$\mu$	0.023	0.017	0.02	0.024
$\sigma$	0.0005	0.0047	0.003	0.007

Cuadro 4.5: Estimación de los parámetros de la GEV

Por otro lado, utilizando la fórmula 3.6, el coeficiente de variación de los datos Dólar-Yen en el caso de la distribución GEV es  $cv_\xi = 1,255$ . El error estándar de  $\hat{\xi}^{-1}$  se ha calculado mediante la matriz de información de fisher observada y da como resultado  $sd(\hat{\xi}^{-1}) = 5,326$ , de hecho el intervalo de confianza al 95 % incluye el parámetro estimado mediante la distribución de Pareto. No obstante, el rango de  $\xi$  es muy amplio y incluye distribuciones que tienen media infinita véase [BG74].

Ahora bien, suponiendo que el valor dado por la estimación del parámetro  $\xi$  de la distribución de Pareto es el valor verdadero. Calculamos el valor del estadístico  $T_m(cv_\xi)$  y nos da  $T_m(1,084) = 9,781$  y el p-valor correspondiente da  $p = 0,398$ .

De manera similar, si consideramos que el valor dado por la estimación del parámetro de la GEV es el valor verdadero. El estadístico  $T_m(1,255)$  en este caso es 9.781 y el p-valor correspondiente es igual a 0.086. El parámetro del peso de la cola estimado de la distribución de Pareto coincide con el valor dado por el CV-Plot, en cambio en el caso de la distribución GEV el parámetro estimado no es igual al dado por el CV-Plot.

Nuestra propuesta para estimar la cola es calcular un  $cv_\xi$  que minimiza  $T_m(cv_\xi)$ , lo cual devuelve  $c\hat{v}_\xi = 1,114$  y  $\hat{\xi}^{-1} = 10,297$ , que se aproximan a los valores estimados de la distribución de Pareto con un umbral bajo y una muestra grande.

## 4.6. Test de exponencialidad

Varios autores han utilizado el coeficiente de variación empírico  $cv$  para contrastar la exponencialidad de los datos. Para el desarrollo de este test, se han basado en el

estadístico de Greenwood, construido en 1946 como,

$$G_{n-1} = \sum_{i=1}^n (U_i - U_{i-1})^2, i = 1, \dots, n \quad U_0 = 0 \text{ y } U_n = 1$$

donde los  $U_i$  son los estadísticos de orden de una muestra de tamaño  $(n - 1)$  distribuidos uniformemente en  $(0,1)$ .

La distribución exacta de  $G_n$  bajo la hipótesis nula de uniformidad es intratable a medida que  $n$  sea grande, de hecho, algunos autores estudiaron diversas aproximaciones, Stephens [Ste81] aproxima los percentiles de  $nG_n$  en las dos colas utilizando la distribución de Pearson con cuatro momentos. En la aproximación de los percentiles en la cola inferior contra la alternativa normal truncada, el mejor resultado es dado por Castillo y Puig[CP99]. Consiste en una transformación del estadístico de verosimilitud mediante la aproximación Saddlepoint a otro estadístico que tiene la misma distribución. Lo que permite la aceleración de la convergencia de este estadístico hacia su distribución asintótica. La manera más simple para contrastar la exponencialidad de los datos es utilizar la Proposición 4.4.

Ahora bien, el test de exponencialidad basado en la expresión 4.4 no proporciona la información suficiente cuando las alternativas son las distribuciones que tienen el coeficiente de variación igual a uno, como es el caso de la distribución de Student con 4 grados de libertad ( $cv[x_+] = cv[|x|] = 1$ ). Para ello, en esta sección desarrollamos un test más consistente para contrastar la exponencialidad considerando el coeficiente de variación de los valores  $x_{|u}$  para diferentes valores de  $u$ .

Sea  $\{x_j\}$  una muestra de una función de distribución exponencial, para cualquier umbral  $u$  tal que  $u_0 < u_1 < \dots < u_m$ , definimos  $cv_k$  por  $cv_k = cv[x_j - u_k; x_j > u_k]$ . Sea  $n_k$  el número de valores que exceden el umbral  $u_k$  en el conjunto  $\{x_j : x_j > u_k\}$ , donde  $0 \leq k \leq m$ .

Consideramos el estadístico

$$T_m = \sum_{k=0}^m n_k (cv_k - 1)^2,$$

de la expresión 4.4,  $n_k (cv_k - 1)^2$  es asintóticamente distribuido como  $\chi_1^2$ . La distribución asintótica de  $T_m$  no depende del parámetro de escala, además es fácil simular  $T_m$ . La esperanza asintótica del estadístico  $T_m$  es  $m + 1$ , la distribución asintótica de  $T_m$  no es  $\chi_{m+1}^2$  debido a que las variables  $cv_k$  no son independientes. Valores menores de  $T_m$

indican exponencialidad cuando el valor esperado de  $cv_k$  es uno. Valores mayores de  $T_m$  implican un desvío de la exponencialidad.

Desde un punto de vista práctico, tomamos una muestra  $\{x_j\}$  de función de distribución desconocida, representamos el CV-Plot y buscamos a partir de este gráfico la muestra para la cual el CV-Plot es constante. Suponemos  $n_0 = 2^{k_0}$  es la muestra más grande que proporciona estabilidad del CV-Plot y  $u_0$  el mínimo de los  $n_0$  valores mayores de la muestra. Utilizamos la Proposición 4.7 para encontrar umbrales aproximadamente iguales bajo la hipótesis nula de exponencialidad y consideramos  $n_k = n_0 2^{-k}$  y los umbrales  $u_k = x_{(n_k)}$  para  $k = 1, \dots, m$ , donde  $m = k_0 - 4$ . La muestra más pequeña que consideramos es  $n_m = 16$ . Entonces, el test consiste en calcular  $T_m$ , el intervalo de confianza y el p-value correspondiente por simulación bajo la hipótesis nula de exponencialidad.

Como ejemplo de aplicación, tomamos los 2048 valores mayores de los positivos y de los negativos de los datos Dólar-Yen respectivamente. El CV-Plot correspondiente está representado en la Figura 4.4.

Tomamos  $k_0 = 11$  y  $m = 7$ . El coeficiente de variación de los valores positivos para las diferentes muestras  $n_k = n_0 2^k$  es el siguiente

{ 0.986,0.957,1.012,0.990,1.024,0.905,1.1059,0.954 }

y para los valores negativos

{1.092,1.132,1.142,1.112,1.100,1.168,1.184,1.611}

En el caso de los datos positivos, el estadístico  $T_m$  y su p-valor correspondiente son  $T_m = 3,18$  y  $p = 0,767$ . En el caso de los valores negativos,  $T_m = 59,07$  y  $p = 0,002$ .

Entonces, se acepta la hipótesis de exponencialidad para los datos Dólar-Yen positivos y se rechaza para los datos Dólar-Yen negativos. En el primer caso, se acepta la exponencialidad con un umbral bajo y una muestra grande. En el segundo caso se rechaza la  $H_0$  por que el umbral no es muy alto y los coeficientes de variación son mayores que uno. Los resultados coinciden con el CV-Plot.

Generalizando el caso de la distribución exponencial al caso de la distribución de Pareto con un coeficiente de variación  $\zeta_\epsilon$ . Introducimos el estadístico

$$T_m(\zeta_\epsilon) = \sum_{k=0}^m n_k (cv_k - \zeta_\epsilon)^2.$$

El estadístico  $T_m$  puede utilizarse para contrastar la hipótesis nula de que la distribución es de Pareto con el coeficiente del peso de cola  $\xi^{-1}$ . Este estadístico es invariante por escala.

Una forma alternativa para estimar el parámetro del peso de la cola es calcular  $\zeta_\epsilon$  que minimiza  $T_m(\zeta_\epsilon)$  es decir, si tomamos los valores superiores de las muestras de mediana en mediana para los puntos de corte ( $u = 32, 64, 128, 256, 512, 1024$ ). Luego, calculamos  $\zeta$  de manera que  $T_m(\zeta_\epsilon)$  sea mínima, obtenemos

$$\zeta_\epsilon = \frac{\sum_{k=0}^m n_k cv_k}{\sum_{k=0}^m n_k}.$$

### Potencia contra las familias de cola Pareto

**Proposición 4.9.** *Dada una distribución con los dos primeros momentos finitos y que no tiene cola exponencial, la potencia del test  $T_m$  contra las familias de cola Pareto tiende a uno cuando  $n$  tiende a infinito y  $\Delta u$  tiende a cero.*

*Demostración.* El estadístico  $T_m$  tiende a uno cuando  $\Delta u \rightarrow 0$  y  $n \rightarrow \infty$  debido a que  $u \rightarrow cv(x|_u)$  es una función continua y la distribución exponencial es la única con el coeficiente de variación igual a uno.  $\square$

La Tabla 4.6 muestra las potencias aproximadas de  $T_m$ ,  $m = 0, \dots, 4$  de las muestras de tamaño  $2^k$ ,  $k = 7, \dots, 11$  contra la alternativa distribución de Student con cuatro grados de libertad. Estos valores en % han sido calculados para  $\epsilon = 0,05$  por simulación tomando 10000 muestras. La potencia de  $T_m$  aumenta al aumentar  $m$ .

Observamos que para  $n_0 = 2048$ , la potencia de  $T_0$  es la dada por el test de exponencialidad utilizando la fórmula 4.4. Para una muestra de 256 La potencia aumenta de 24.6 % a 31.9 %.

$n_k$	$T_0$	$T_1$	$T_2$	$T_3$
128	20.4	21.0	21.3	22.4
256	24.6	29.4	30.7	31.9
512	28.7	44.1	48.4	47.3
1024	33.5	68.9	76.4	75.6
2048	37.6	96.3	98	96.9

Cuadro 4.6: Potencia del test contra la alternativa distribución de Student

$n_k$	$T_0$	$T_1$	$T_2$	$T_3$
128	3.44	5.87	7.63	8.93
256	3.65	6.21	8.53	10.02
512	3.77	6.56	8.91	10.81
1024	3.81	6.82	9.17	11.41
2048	3.84	6.84	9.40	11.66

Cuadro 4.7: Valores críticos del test bajo la hipótesis nula

Los valores críticos correspondientes para un nivel de confianza del 95% bajo la hipótesis nula de exponencialidad para los mismos valores de  $m$  y los mismos tamaños de muestra han sido calculados mediante simulación en la Tabla 4.7. los valores críticos no varían mucho al aumentar el tamaño de la muestra porque  $T_m$  converge en distribución cuando  $m$  tiende a infinito.

# Capítulo 5

## Apéndices



**THE MIXTURE OF LEFT-RIGHT TRUNCATED NORMAL DISTRIBUTIONS**

Joan del Castillo<sup>(1)</sup> and Jalila Daoudi<sup>(2)</sup>

(1) Corresponding author: Departament de Matemàtiques. Universitat Autònoma de Barcelona, 08193 Barcelona, Spain. E-mail: castillo@mat.uab.cat.

(2) Departament de Matemàtiques. Universitat Autònoma de Barcelona, Spain.

ABSTRACT. This paper introduces the mixture of left-right truncated normal distributions, from the spreads between bid and ask prices, as a statistical model for handle non-normality of asset price returns. It has been proved that there is only one maximum for the likelihood function of the new model.

**Keywords:** Coefficient of variation, heavy tailed distributions, return distributions. Normal inverse Gaussian distribution.

1. INTRODUCTION

The Black and Scholes model for the logarithm of asset prices is the Gaussian random walk. This is built by summing independent and identically distributed normal variables (returns). However, from an empirical point of view this assumption is far from perfect. This problem has relevant economic consequences in risk management, asset allocation and pricing derivatives.

Non-normality of the asset price returns has been handled in several ways. A satisfactory method is to assume that the volatility follows a stochastic process independent of the returns process (Taylor, 2005, Ch.11). If the volatility process is a two state Markov chain, then the mixture of normal distributions arises as the marginal distribution for returns. This model was previously proposed by Hull and White (1998). Barndorff-Nielsen (1998) introduce the normal inverse Gaussian distribution by considering a continuous mixture of normal distributions, from the inverse Gaussian distributions as the mixing distribution. These models provide better adjustments to the asset price return distributions because they fit higher order moments. Exploratory data analysis in Section 2, shows that these models do not correctly explain the non-normality.

In Section 3 it is shown that the prices formation mechanism may explain some amount of non-normality. Assuming normality for bid and ask prices, the observed transaction prices may be a mixture of normal distributions or a mixture of left-right truncated normal distributions, the latter case being more likely.

The statistical properties of the mixture of left-right truncated normal distributions are developed in Section 4. It is proved that there is only one maximum for the likelihood function and the maximum is closely related to the coefficient of variation. The results apply, in particular, to the symmetric mixture of truncated distributions characterized by Damilano and Puig (2004), in terms of the asymptotic efficiency of the median and the sample mean. Continuous mixed truncated

normal has been used to generate asymmetric models (see Pedersen and Hwang, 2007). Our results show that continuity at zero of this distribution can be avoided in statistical analysis. Empirical work in Section 5 suggests that in financial data non-normality is also produced for a large number of values close to the mean, here referred to as "inliers". This could explain that the Laplace approximation is often a better approach than normal distribution for daily returns.

## 2. EXPLORATORY DATA ANALYSIS

The behavior of exchange rates between the euro and the US dollar is considered. The data set consists of five years of daily exchange rates from the starting date of the euro, January 1, 1999 (the data set is available from OANDA Corporation at <http://www.oanda.com/convert/fxhistory>).

Throughout the paper we work with the series of returns

$$(2.1) \quad y_t = 100 \log(x_t/x_{t-1})$$

where  $x_t$  is the value of one euro denominated in dollars at time  $t$ . For daily returns used in this section,  $t$  runs from 1 to 1825 days (1824 returns). Table 1 reports the mean, the standard deviation (Sd), the skewness (Sk) and the kurtosis (Ku) for returns on the Euro in US dollar for one, two, three and four days, and the mean and the coefficient of variation for the corresponding absolute values of returns.

In the following we compare the fits of the empirical distribution with the normal distribution, the mixture of normal distributions, the Laplace distribution and the normal inverse Gaussian distribution. The distributions considered in this section are always assumed to be symmetric with respect to zero and returns are mean corrected. This assumption is usual when modelling the returns of financial assets, since the mean value is really a small quantity with respect to the standard deviation (here one hundred times smaller) and large sample size are considered, see Table 1. However, for more general settings a more in-depth analysis may be needed. Hence, the normal distribution is determined by the standard deviation (zero mean). The Laplace distribution is determined by the mean of the absolute values of the returns.

Time	size	mean	Sd	Sk	Ku - 3	mean(abs)	cv(abs)
Daily	1824	0.0037	0.5545	0.1019	1.9002	0.3829	1.0471
Two-days	912	0.0074	0.7727	0.0210	0.7972	0.5764	0.8923
Three-days	608	0.0111	0.9550	-0.0318	0.5408	0.7280	0.8481
Four-days	456	0.0148	1.0694	0.0464	0.1380	0.8436	0.7777

TABLE 1. Descriptive statistics for returns on the Euro in US dollar for one, two, three and four days. The coefficient of variation and the kurtosis show that returns reach normality when time horizon increases.

Figure 1 shows the histogram for exchange rate returns and the fitted-probability density functions for both the normal and Laplace distributions. There is a high spike at zero showing that there are more outcomes closer to the central value (inliers) than are predicted by the normality assumption. The Laplace distribution shows a better fit for the data set.

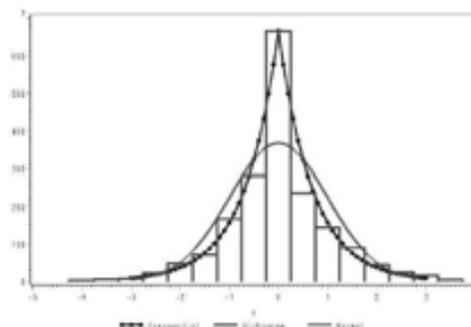


FIGURE 1. The histogram for daily exchange rate returns and the fitted-probability density functions for normal and Laplace distributions.

Following Hull and White (1998), we considered the mixture of normal distributions as a weighted average of two zero mean normal distributions, with the cumulative distribution function

$$(2.2) \quad G(y) = \pi \Phi\left(\frac{y}{u\sigma}\right) + (1 - \pi) \Phi\left(\frac{y}{v\sigma}\right),$$

where  $\Phi(y)$  is the cumulative distribution function of the standard normal distribution and parameter  $\pi$  corresponds to the weight and  $\pi u^2 + (1 - \pi)v^2 = 1$ . The estimation of parameters is done by maximum likelihood. Hull and White (1998) suggest an alternative approach of fitting the fractiles of the distribution, which is used here as a starting point for maximum likelihood.

Infinitely divisible distributions are an alternative to normal distribution that generalize Gaussian random walk to Lévy processes. An appealing family of infinitely divisible distributions closed by convolution is the symmetric normal inverse Gaussian introduced by Barndorff-Nielsen (1998), (*NIG*). The probability density function for a *NIG*( $\omega, \phi$ ) random variable,  $y$ , is

$$(2.3) \quad f(y; \phi, \omega) = \frac{\omega \exp(\omega)}{\pi \sqrt{y^2 + \phi\omega}} K_1\left(\sqrt{y^2\omega\phi^{-1} + \omega^2}\right),$$

where  $K_1$  is the three-order modified Bessel function with index 1. The variance and the kurtosis of  $y$  are given by (see Bibby and Sorensen, 2003)

$$(2.4) \quad \text{var}[y] = \phi, \quad \text{ku}[y] = 3/\omega,$$

where  $\text{ku}[y]$  is the kurtosis excess of  $y$ . A clear interpretation of the parameters  $\omega$  and  $\phi$  are given by (2.4). If the method of moments were used for mean corrected returns, the model exactly fits the variance and the kurtosis of the data set. For *NIG*( $\phi, \omega$ ), the parameters are easily estimated by maximum likelihood, as is done here. However, for multivariate volatility models, Castillo and Lee (2008) suggest using an h-likelihood approach.

Table 2 shows the fitted-probability density functions for exchange rate daily returns. The distributions considered are the normal, the mixture of normal distributions, the Laplace and the normal inverse Gaussian. The histogram is carried out in  $k = 18$  classes. Cut points are expressed in terms of the standard deviation. Since there are 99 zeros in the data set, a half of them are allocated to the positive part

of the histogram. The Pearson goodness of fit statistic is calculated with aggregate categories. Since the parameters are estimated by maximum likelihood from the sample, the asymptotic distribution can be bounded between a chi-squared with  $(k - 1)$  degrees of freedom and a chi-squared with  $(k - 1 - p)$  degrees of freedom, where  $p$  is the number of estimated parameters. The corresponding P-values are reported in Table 2, the true asymptotic P-value belongs to this interval.

As seen in Table 2, the best fitting to the observed frequencies, in terms of Pearson statistic, is provided by the Laplace distribution, its P-value is bigger than 0.01. The Laplace is the only distribution not rejected with 99% confidence. The mixture of both normal distributions and NIG distributions clearly improves the normal fitting. However, there is considerable difference between these and the Laplace distribution. The NIG distribution may have the same variance and kurtosis as the data set (if the method of moments were used). Therefore, the kurtosis does not correctly explain exchange-rate returns.

Cut points	Observed	Mix_Nor	NIG	Laplace	Normal
-4.0	2	0.4	8.3	2.8	0.1
-3.5	2	1.4	4.6	3.0	0.4
-3.0	8	5.1	7.6	6.1	2.0
-2.5	15	14.8	13.0	12.6	8.9
-2.0	26	36.3	23.3	25.9	30.2
-1.5	63	74.3	44.8	53.5	80.4
-1.0	102	127.1	94.8	110.4	167.5
-0.5	212	182.1	225.3	227.8	273.4
0.0	495	470.5	490.4	470.0	349.2
0.5	477	470.5	490.4	470.0	349.2
1.0	172	182.1	225.3	227.8	273.4
1.5	125	127.1	94.8	110.4	167.5
2.0	59	74.3	44.8	53.5	80.4
2.5	38	36.3	23.3	25.9	30.2
3.0	18	14.8	13.0	12.6	8.9
3.5	4	5.1	7.6	6.1	2
4.0	3	1.4	4.6	3.0	0.4
>4.0	3	0.4	8.3	2.8	0.1
chi-square statistic		59.36	59.79	31.30	473.08
Number of parameters		3	2	1	1
P-value		0.000	0.000	0.012-0.018	0.000

TABLE 2. Observed and estimated frequencies for returns of daily exchange rates between the euro and the dollar (cut points are expressed in terms of the standard deviation). The fitted distributions are the mixture of normal distributions, the normal inverse Gaussian distribution, the Laplace distribution and the normal distribution.

### 3. THE MLRTN-MODEL

The mixture of left-right truncated normals (MLRTN) is introduced in this section as a model for the asset price return distributions from a simply economic point of view. From now on, we always assume the logarithmic scale for the prices and their increments. Hence, the more basic model for asset prices assumes that returns  $y$  (price increments) are drawn from independent normal distributed random variables. A normal random variable with mean value  $\mu$  and variance  $\sigma^2$  is denoted by  $N(\mu, \sigma^2)$ .

Let us denote by  $TN^+(c, \mu, \sigma^2)$  and by  $TN^-(c, \mu, \sigma^2)$ , the distribution of a normal random variable  $y$  left truncated and the right truncated at  $y = c$ , respectively. Their probability density functions are given, respectively, by

$$(3.1) \quad \begin{aligned} f_c(y; \mu, \sigma) &= \frac{\varphi((y - \mu)/\sigma)}{\sigma(1 - \Phi(\alpha))}, & y > c, \\ \bar{f}_c(y; \mu, \sigma) &= \frac{\varphi((y - \mu)/\sigma)}{\sigma\Phi(\alpha)}, & y \leq c, \end{aligned}$$

where  $\varphi(y) = \Phi'(y)$  is the probability density function of the standard normal distribution and  $\alpha = (c - \mu)/\sigma$ . Let us denote by  $TN(\mu, \sigma^2)$  a normal random variable with mean  $\mu$  and variance  $\sigma^2$ , left truncated at  $c = 0$ .

**Remark 1.** *From these definitions, given a random variable  $x$  and a constant  $c$ , they are equivalent in saying that  $x$  is distributed as  $TN(\mu, \sigma)$ ,  $x + c$  is distributed as  $TN^+(c, \mu + c, \sigma)$  or  $c - x$  is distributed as  $TN^-(c, c - \mu, \sigma)$ . In particular, if  $y$  is distributed as  $TN^-(c, \mu, \sigma)$  then  $(-y)$  is distributed as  $TN^+(-c, -\mu, \sigma)$ .*

We start by thinking that a price comes from the agreement between supply and demand. Let us assume that the changes of prices proposed by sellers,  $a$ , are drawn from a  $N(\mu_a, \sigma_a^2)$  and changes proposed by buyers,  $b$ , are drawn from a  $N(\mu_b, \sigma_b^2)$ . The mechanism of agreement between ask prices (the last price plus the increment  $a$ ) and bid prices (the last price plus the increment  $b$ ) state the distribution of returns. If all the changes proposed by buyers and sellers are accepted, then the distribution of returns is a mixture of normal distributions.

An alternative way to produce the agreement is now suggested. Let  $c$  be a constant such that if a buyer offers more than this constant ( $b > c$ ), then the transaction is produced and if a seller accepts less ( $a < c$ ) then the transaction is also produced. If  $p$  is the ratio of bid prices then the generating algorithm of transaction prices is described in the following theorem:

**Theorem 1.** *Let us consider the following algorithm of transaction prices:*

- (1) *A Bernoulli distributed random variable,  $d$ , with probability  $p$  is drawn. If  $d = 1$ , then we draw  $b$ , else we draw  $a$ .*
- (2) *If  $d = 1$  and  $b > c$  then  $y = b$ , else a new  $d$  is drawn and the process is repeated.*
- (3) *In the same way, if  $d = 0$  and  $a < c$  then  $y = a$  else a new  $d$  is generated and the process is repeated.*

*With the previous algorithm, the probability density function of  $y$  is given by*

$$(3.2) \quad \pi f_c(y; \mu_b, \sigma_b) + (1 - \pi) \bar{f}_c(y; \mu_a, \sigma_a)$$

where  $\pi = p p_b / (1 - p_3)$ ,  $p_a = \Pr\{a > c\}$ ,  $p_b = \Pr\{b > c\}$  and  $p_3 = (1 - p)p_a + p(1 - p_b)$ .

*Proof.* The step one provides three possibilities. First: if  $d = 1$  and  $b > c$  then  $y = b$ ; this happens with probability  $p_1 = \Pr\{d = 1\} \Pr\{b > c\} = p p_b$ . Second: if  $d = 0$  and  $a < c$  then  $y = a$ ; this happens with probability  $p_2 = (1 - p)(1 - p_a)$ . Third: else, a new  $d$  is drawn; this last case happens with probability  $p_3 = (1 - p)p_a + p(1 - p_b)$ .

Iterating the process, the proportion of offerings accepted is

$$\pi = p_1 + p_1 p_3 + \dots + p_1 p_3^n + \dots = \frac{p p_b}{1 - p_3}$$

Clearly the distribution of  $(b | b > c)$  is  $TN^+(c, \mu_b, \sigma_b^2)$  and the distribution of  $(a | a \leq c)$  is  $TN^-(c, \mu_b, \sigma_b^2)$ . Hence, the result follows.  $\square$

Hereafter, the mixture of left-right truncated normals given by (3.2) will be denoted by MLRTN.

Figure 2 shows a simple case assuming that  $\pi = 1/2$ ,  $c = (\mu_a + \mu_b)/2$  and  $\sigma_a = 1.5\sigma_b$ . In this case the mixture of left-right truncated normals is asymmetric. The model corresponds to the "intersection" of samples of distributions with different means, in the same way that the mixture of distributions corresponds to the "union" of samples. In both cases, the "intersection" or the "union" must be normalized in order to be true probability density functions.

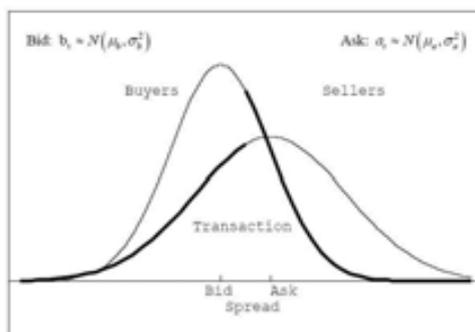


FIGURE 2. Asymmetric MLRTN. The thin lines represent the probability density function for prices of buyers and sellers. The thick line represents the probability density function for transaction prices, up to a normalizing constant.

#### 4. STATISTICAL PROPERTIES

The MLRTN includes a large number of sub-models. When  $\pi = 1/2$ ,  $\sigma_a = \sigma_b$  and  $\mu_a = \mu_b = c$  it is the normal distribution. When  $\pi = 1/2$ , and  $\mu_a$  and  $\mu_b$  tend to  $\infty$ , then the limit density corresponds to the Laplace distribution. When  $\pi = 1/2$ ,  $\sigma_a = \sigma_b$  and  $(\mu_a + \mu_b)/2 = c$ , (3.2) is the symmetrized truncated normal considered in Damilano and Puig (2004). In the general case,  $0 < \pi < 1$ , the support of the distribution is  $\mathbb{R}$ . If  $\pi = 0$  or  $\pi = 1$ , then the support is a half straight line. Hence, for  $\pi = 1$  and  $\mu_b$  tending to  $\infty$ , the model (3.2) tends to the exponential distribution.

**Proposition 1.** *The following results hold:*

- (1) *The necessary and sufficient condition so that MLRTN be symmetrical with respect to  $c$  is  $\pi = 1/2$ ,  $\sigma_a = \sigma_b$  and  $(\mu_a + \mu_b)/2 = c$ .*
- (2) *If  $\mu_b \leq c \leq \mu_a$ , the probability density function (3.2) of the MLRTN is unimodal with its mode at  $c$ .*
- (3) *If  $\mu_a \leq c \leq \mu_b$  then the probability density function (3.2) of the MLRTN is bimodal with modes at  $\mu_a$  and  $\mu_b$ .*

*Proof.* (1) Let  $x > 0$ , from symmetry

$$\pi f_c(c + x; \mu_b, \sigma_b) = (1 - \pi) \bar{f}_c(c - x; \mu_a, \sigma_a).$$

Then, from (3.1),  $\varphi((c+x-\mu_b)/\sigma_b)$  and  $\varphi((c-x-\mu_a)/\sigma_a)$  are proportional. Taking logarithms it follows that  $((c+x-\mu_b)/\sigma_b)^2$  and  $((c-x-\mu_a)/\sigma_a)^2$  are equal but a constant. Taking derivatives it follows that  $\sigma_a = \sigma_b$  and  $(\mu_a + \mu_b)/2 = c$ . Therefore,  $\pi = 1/2$ .

(2) The probability density of the left truncated normal,  $f_c(y; \mu, \sigma)$ , is monotonous decreasing function of  $y$  when  $\mu \leq c$ . The symmetric case happens for the right truncated normal  $\bar{f}_c(y; \mu, \sigma)$ . From  $\mu_b \leq c$ , therefore  $f_c(y; \mu_b, \sigma_b)$  is a decreasing function of  $y$  from  $c$  to  $\infty$ . And from  $c \leq \mu_a$ ,  $\bar{f}_c(y; \mu_a, \sigma_a)$  is an increasing function for  $y$  from  $-\infty$  to  $c$ .

(3) If  $c \leq \mu$ , then the probability density of the left truncated normal,  $f_c(y; \mu, \sigma)$ , is an increasing function from  $c$  to the mode  $\mu$  and it decrease for larger values of  $y$ . The symmetric case happens for the right truncated normal  $\bar{f}_c(y; \mu, \sigma)$ .

From  $\mu_a \leq c \leq \mu_b$ , therefore  $f_c(y; \mu_b, \sigma_b)$  has a mode at  $y = \mu_b$  and  $\bar{f}_c(y; \mu_a, \sigma_a)$  has a mode at  $y = \mu_a$ .  $\square$

From an economic point of view  $\mu_b \leq c \leq \mu_a$ , because  $c < \mu_b$  means that more than 50% of buyer prices are bigger than  $c$ , hence  $c$  is not a good price for sellers. In the same way  $\mu_a < c$  means that  $c$  is not a good price for buyers.

Notice that if  $\mu_a \leq c \leq \mu_b$  then the distribution is similar to a mixture of normal distributions.

**Proposition 2.** *A TN  $(\mu, \sigma)$  distributed random variable  $x$  has the moment-generating function*

$$(4.1) \quad M_x(t) = E(e^{tx}) = \frac{\Phi(\mu/\sigma + t\sigma)}{\Phi(\mu/\sigma)} \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

Hence, the first two moments of  $x$  are given by

$$(4.2) \quad \begin{aligned} E(x) &= \mu + \sigma\lambda(\alpha) \\ var(x) &= \sigma^2(1 - \lambda(\alpha)^2 + \alpha\lambda(\alpha)) \end{aligned}$$

where  $\alpha = -\mu/\sigma$  and  $\lambda(\alpha) = \varphi(\alpha)/(1 - \Phi(\alpha))$  is the inverse of the Mills ratio.

*Proof.* Is the same proof as for the normal distribution, taking care over the normalizing constants. For more details see Castillo (1994).  $\square$

A TN  $(0, \sigma^2)$  distributed random variable  $x$  is known as a *half normal distribution*, its distribution is that of the absolute value of a normal one distribution. In this case, from (4.2),  $E[x] = \sqrt{2/\pi}\sigma$  and the coefficient of variation is

$$(4.3) \quad cv[x] = \sqrt{\frac{\pi-2}{2}} \approx 0.75551.$$

**Corollary 1.** *A TN<sup>+</sup>  $(c, \mu_b, \sigma_b)$  distributed random variable  $y_b$  has the moment-generating function*

$$M_b(t) = \frac{\Phi((\mu_b - c)/\sigma + t\sigma_b)}{\Phi((\mu_b - c)/\sigma_b)} \exp\left(\mu_b t + \frac{\sigma_b^2 t^2}{2}\right).$$

*A TN<sup>-</sup>  $(c, \mu_a, \sigma_a)$  distributed random variable  $y_a$  has the moment-generating function*

$$\bar{M}_a(t) = \frac{\Phi((c - \mu_a)/\sigma - t\sigma_a)}{\Phi((c - \mu_a)/\sigma_a)} \exp\left(\mu_a t + \frac{\sigma_a^2 t^2}{2}\right)$$

A *MLRTN* distributed random variable  $y$ , as in (3.2), has the moment-generating function

$$(4.4) \quad M(t) = \pi M_b(t) + (1 - \pi) \bar{M}_a(t)$$

*Proof.* For the first equation we consider  $x_b = y_b - c$  which is distributed as  $TN(\mu_b - c, \sigma_b)$ . Hence, from (4.1) and  $E(e^{t y_b}) = e^{tc} E(e^{t x_b})$  the first result follows.

For the second equation,  $x_a = c - y_a$  is distributed as  $TN(c - \mu_a, \sigma_a)$ . Hence, using (4.1) and  $E(e^{t y_a}) = e^{tc} E(e^{-t x_a})$  the result follows.

A *MLRTN* distributed random variable  $y$  is a mixture of random variables like the  $y_b$  and the  $y_a$  given before. Hence, its moment-generating function is easily obtained by

$$\int_{-\infty}^{\infty} e^{t y} ((\pi f_b(y; \mu_b, \sigma_b) + (1 - \pi) \bar{f}_a(y; \mu_a, \sigma_a))) dy = \pi M_b(t) + (1 - \pi) \bar{M}_a(t).$$

□

From (4.4), the moments for a *MLRTN* can be obtained easily. For simplicity's sake, we will consider the limit case. Since the distribution of a  $TN(\mu, \sigma)$  random variable with  $\mu$  tending to  $-\infty$  tends towards an exponential distribution, a *MLRTN* distributed random variable can be arbitrarily close to a mixture of an exponential distribution with mean  $\lambda_b$  and another exponential distribution with mean  $\lambda_a$  with the sign changed. Then, as in (4.4), the moment-generating function is

$$(4.5) \quad M_{\infty}(t) = \frac{\pi}{1 - \lambda_b t} + \frac{1 - \pi}{1 + \lambda_a t}$$

Table 3 shows a range of variation of kurtosis for a *MLRTN* in terms of  $\pi$  and  $\lambda_b$ , assuming  $\lambda_a = 1$ .

$\pi \backslash \lambda_b$	1	2	3	4	5	6	7	8	9	10
0.5	6.00	7.50	9.00	9.95	10.56	10.98	11.28	11.51	11.69	11.83
0.4	6.12	8.24	10.25	11.56	12.41	12.99	13.41	13.72	13.96	14.16
0.3	6.50	9.49	12.39	14.32	15.59	16.47	17.10	17.57	17.93	18.22
0.2	7.17	11.54	16.20	19.50	21.74	23.31	24.44	25.28	25.93	26.44
0.1	8.16	14.65	23.75	31.44	37.18	41.39	44.50	46.85	48.67	50.09

TABLE 3. Range of variation of the kurtosis for a mixture of left-right truncated normal distributions in the limit case, in terms of  $\pi$  and  $\lambda_b$ , assuming  $\lambda_a = 1$

Although the MLRTN has six parameters, in general, there exists only one maximum likelihood estimator (MLE), as the following results prove.

**Theorem 2.** *Let  $y_1, y_2, \dots, y_n$  be a sample from the MLRTN and let  $c$  be a known constant. The log-likelihood function splits on three terms, corresponding to the likelihoods of three exponential models, without common parameters. Hence, the MLE exist and is unique, for samples large enough.*

*Proof.* Notice that if  $y \leq c$  then  $f_c(y; \mu_b, \sigma_b) = 0$  and if  $y > c$  then  $\bar{f}_c(y; \mu_a, \sigma_a) = 0$ . Let  $n_c$  be the size of the set  $\{y_t > c\}$ . Hence, the log-likelihood function for the sample is

$$(4.6) \quad \begin{aligned} l_c(\pi, \mu_b, \sigma_b, \mu_a, \sigma_a) = & \sum_{t=1}^n \log(\pi f_c(y_t; \mu_b, \sigma_b) + (1 - \pi) \bar{f}_c(y_t; \mu_a, \sigma_a)) = \\ & \sum_{y_t \leq c} \log((1 - \pi) \bar{f}_c(y_t; \mu_a, \sigma_a)) + \sum_{y_t > c} \log(\pi f_c(y_t; \mu_b, \sigma_b)) = \\ & (n - n_c) \log(1 - \pi) + n_c \log(\pi) + \sum_{y_t \leq c} \log(\bar{f}_c(y_t; \mu_a, \sigma_a)) + \sum_{y_t > c} \log(f_c(y_t; \mu_b, \sigma_b)) \end{aligned}$$

The function  $l\pi = (n - n_c) \log(1 - \pi) + n_c \log(\pi)$  is the log-likelihood of the binomial  $n_c$  with sample size  $n$  and probability  $\pi$ . The function  $lb = \sum_{y_t > c} \log(f_c(y_t; \mu_b, \sigma_b))$  is the log-likelihood of a left truncated normal distribution for the sample  $\{y_t > c\}$  and  $la = \sum_{y_t \leq c} \log(\bar{f}_c(y_t; \mu_a, \sigma_a))$  is the log-likelihood of a right truncated normal distribution for the sample  $\{y_t \leq c\}$ . The three distributions are exponential families without common parameters. From the general theory it follows that the MLE exist and is unique, see also Castillo (1994).  $\square$

**Remark 2.** Given a sample  $\{y_t\}$  from  $TN^+(c, \mu, \sigma^2)$  with  $c$  known, the MLE can be find from the MLE for a  $TN(\mu_0, \sigma^2)$  and the sample  $\{y_t - c\}$ , with  $\mu = c + \mu_0$ . In the same way, given a sample  $\{y_t\}$  from  $TN^-(c, \mu, \sigma^2)$  with  $c$  known, the MLE can be find from the MLE for  $TN(\mu_0, \sigma^2)$  and the sample  $\{c - y_t\}$ , with  $\mu = c - \mu_0$ .

Let  $y_1, y_2, \dots, y_n$  be a sample from  $TN(\mu, \sigma^2)$ , let  $\bar{y}$  be the sample mean, let  $s$  be the standard deviation and  $cv = s/\bar{y}$ , the empirical coefficient of variation. The likelihood equations are reduced to solving

$$(4.7) \quad \frac{\alpha^2 + 1 - \alpha\lambda(\alpha)}{(\lambda(\alpha) - \alpha)^2} = 1 + cv^2,$$

where  $\alpha = -\mu/\sigma$ , and  $\lambda(\alpha) = \varphi(\alpha)/(1 - \Phi(\alpha))$ . The MLE is therefore given by

$$(4.8) \quad \hat{\mu} = \frac{\bar{y}\alpha}{\alpha - \lambda(\alpha)}, \quad \hat{\sigma} = \frac{\bar{y}}{\lambda(\alpha) - \alpha}.$$

Castillo(1994) shows that Equation (4.7) has a solution if and only if  $cv < 1$  and it provides a one-to-one map between  $cv$  and  $\alpha$ . Moreover, the exponential distribution appears as the limit case when  $cv \geq 1$ .

Given a sample  $\{y_t\}$  from an MLRTN with  $c$  known,  $n_c$  denote the size of the set  $\{y_t > c\}$ , the set  $\{y_t - c : y_t > c\}$  is called the *right part* of the sample and the set  $\{c - y_t : y_t \leq c\}$  is called the *left part* of the sample. Both cases are samples of positive random variables. Let  $\bar{y}_+$  and  $cv_+$  be the mean value and the coefficient of variation of the right part of the sample, let  $\bar{y}_-$  and  $cv_-$  be the mean value and the coefficient of variation of the left part. From Theorem 2, (4.7) and (4.8) it follows that

$$(4.9) \quad \{n_c, \bar{y}_+, cv_+, \bar{y}_-, cv_-\}$$

is a sufficient statistic for the parameters of the MLRTN.

The MLRTN model, introduced in Theorem 1, suggests that the sufficient statistic (4.9) has to be relevant for the analysis of prices of financial assets. This quantities are directly observable and can be used beyond the model.

**Remark 3.** *If  $c$  is unknown the profile likelihood*

$$l_p(c) = \max \{l_c(\pi, \mu_b, \sigma_b, \mu_a, \sigma_a)\}$$

*can be used for its estimation. Theorem 2 shows that  $l_p(c)$  is well defined and computationally simple.*

## 5. EMPIRICAL ANALYSIS

First of all, in order to compare the MLRTN model with those considered in Section 2, we restrict (3.2) to the symmetric case  $\pi = 1/2$ ,  $\sigma_a = \sigma_b$  and  $(\mu_a + \mu_b)/2 = c$ . The MLRTN reduces to the symmetrized truncated normal distribution; moreover  $c = 0$ , since returns are mean corrected. The MLE for the MLRTN in this symmetric case reduce to the MLE for  $TN(\mu, \sigma^2)$  with the absolute values of the series. We compare the models for daily, two-day and three-day returns.

For daily returns the coefficient of variation of the absolute values is  $cv = 1.0471$ . When  $cv \geq 1$  the MLE for truncated normal distribution corresponds to the exponential distribution (see Castillo, 1994). Hence, in this case, the MLRTN reduces to the Laplace distribution that provides the best fit to the data set, as is explained for Table 2 in Section 2.

Tables 4 and 5 are constructed as Table 2, also considering the symmetric MLRTN model. Now the fitted-probability density functions for exchange rate corresponds to two-day and three-day returns. Since the coefficient of variation of the absolute returns are less than 1, the MLRTN is not equal to the Laplace, as it is for daily returns.

In Table 4, for two-day returns, the MLRTN provides the best fit to the observed frequencies in terms of the Pearson goodness of fit statistic. The mixture of normal distributions can not be rejected but it has a higher value for the statistic and one more parameter. The NIG is near to being rejected with 90% confidence. The Laplace distribution clearly improves the normal fit, but it is rejected with 95% confidence.

In Table 5, for three-day returns, the MLRTN and the mixture of normal distributions are the best fitting distributions, although the second has one more parameter. The NIG is also acceptable. The normal and the Laplace distributions can be rejected with 95% confidence.

**5.1. New tools for the analysis of financial data.** The mean value of returns is a very small quantity with respect to the standard deviation, see Table 1. Therefore, this mean value only makes sense for long term series. However, the mean value of the absolute returns is of the same order as the standard deviation. The mean value of the absolute daily returns has an economic interpretation as the average of what you will win or lose on a random day in the market. The analysis can be extended beyond the symmetrical model to the right part and the left part of the returns, using the sufficient statistic (4.9). From 5 years considered we have 1824 daily returns, of which 876 are positive, 849 negative and 99 are equal to

Cut points	Observed	MLRTN	Mix_Nor	NIG	Laplace	Normal
-4.0	0	0.5	0.1	1.0	2.1	0.0
-3.5	0	1.0	0.4	1.2	2.1	0.2
-3.0	5	2.7	1.8	2.7	4.0	1.0
-2.5	6	6.7	6.2	6.1	7.8	4.4
-2.0	14	15.4	17.1	13.9	15.3	15.1
-1.5	32	33.1	38.9	31.4	29.8	40.2
-1.0	80	65.9	71.8	68.4	58.3	83.8
-0.5	103	121.8	108.9	132.7	113.9	136.7
0.0	229	209.1	210.9	198.7	222.7	174.6
0.5	191	209.1	210.9	198.7	222.7	174.6
1.0	118	121.8	108.9	132.7	113.9	136.7
1.5	72	65.9	71.8	68.4	58.3	83.8
2.0	34	33.1	38.9	31.4	29.8	40.2
2.5	19	15.4	17.1	13.9	15.3	15.1
3.0	6	6.7	6.1	6.2	7.8	4.4
3.5	1	2.7	1.8	2.7	4.0	1.0
4.0	2	1.0	0.4	1.2	2.1	0.2
> 4.0	0	0.5	0.1	1.0	2.1	0.0
chi-square statistic		17.36	20.45	24.29	28.58	70.16
Number of parameters		2	3	2	1	1
P-value		0.298-0.430	0.117-0.252	0.060-0.112	0.027-0.039	0.0000

TABLE 4. Observed and estimated frequencies for returns of two-day exchange rates between the euro and the dollar (cut points are expressed in terms of the standard deviation). The fitted distributions are the MLRTN, the mixture of normal distributions, the normal inverse Gaussian distribution, the Laplace distribution and the normal distribution.

Cut points	Observed	MLRTN	Mix_Nor	NIG	Laplace	Normal
-4.0	0	0.2	0.1	0.4	1.6	0.0
-3.5	0	0.5	0.3	0.6	1.5	0.1
-3.0	3	1.4	1.2	1.5	2.9	0.7
-2.5	2	4.1	4.0	3.8	5.5	3.0
-2.0	11	10.5	11.1	9.5	10.6	10.1
-1.5	32	23.6	25.4	22.7	20.5	26.8
-1.0	37	47.3	47.7	49.2	39.4	55.8
-0.5	86	84.0	81.2	90.0	75.9	91.1
0.0	136	132.4	133.2	126.5	146.2	116.4
0.5	129	132.4	133.2	126.5	146.2	116.4
1.0	78	84.0	81.2	90.0	75.9	91.1
1.5	56	47.3	47.7	49.2	39.4	55.8
2.0	19	23.6	25.4	22.7	20.5	26.8
2.5	14	10.5	11.1	9.5	10.6	10.1
3.0	3	4.1	4.0	3.8	5.5	3.0
3.5	2	1.4	1.2	1.5	2.9	0.7
4.0	0	0.5	0.3	0.6	1.5	0.1
> 4.0	0	0.2	0.1	0.4	1.6	0.0
chi-square statistic		14.11	13.98	17.92	28.83	29.17
Number of parameters		2	3	2	1	1
P-value		0.517-0.659	0.451-0.669	0.267-0.394	0.025-0.036	0.023-0.033

TABLE 5. Observed and estimated frequencies for returns for three-day exchange rates between the euro and the dollar (cut points are expressed in terms of the standard deviation). The fitted distributions are the MLRTN, the mixture of normal distributions, the normal inverse Gaussian distribution, the Laplace distribution and the normal distribution.

zero. In our analysis, with the symmetrical mixture of left-right truncated normal distributions, the mean value of the absolute returns and the coefficient of variation of the absolute returns are the two relevant quantities.

The coefficient of variation of the absolute returns (or those of the right part and left part of returns) provides a new way of measuring the non-normality of tail distributions for financial data. This measurement uses only two moments and is therefore more stable and more widely applicable than kurtosis. We recall that some papers on high frequency data suggest that the kurtosis could be infinite and only three moments can be used. Table 1 shows that for periods of time of two, three or four days the empirical coefficient of variations of the absolute returns tends towards the theoretical value for the absolute value of a normal distribution (4.3) and the excess of kurtosis of the sample tends towards zero. This fact roughly indicates the time interval where the aggregation to Gaussianity is produced. In our model, if  $\mu_b \leq c \leq \mu_a$ . the coefficient of variation of the absolute returns is between 1 (Laplace distribution) and 0.756 (normal distribution).

**5.2. Asymmetric case.** Although the symmetric MLRTN works better than alternative models, the fit can be improved removing the restrictions on  $\pi$  and  $c$  (losing the symmetric shape).

The log-likelihood for for two-day returns in the symmetric case, with  $c = 0$  and  $\pi = 0.5$ , is  $l_0 = -1033.78$ . The MLE for MLRTN, with  $c = 0$  and without restriction on  $\pi$ , gives  $\hat{\pi} = 0.486$  and the log-likelihood  $l_1 = -1030.22$ .

The MLE for MLRTN without any restriction can be found from the profile likelihood considered in Remark 1. It gives  $\hat{c} = 0.11$ ,  $\hat{\pi} = 0.390$  and the log-likelihood  $l_2 = -1026.99$ . Table 6 shows the estimated parameters by MLE for MLRTN in the three cases. The likelihood ratio test shows that the fit in the symmetric case is significantly improved when removing the restrictions.

$\pi$	$c$	$\mu_b$	$\sigma_b$	$\mu_a$	$\sigma_a$	likelihood
0.5000	0.000	-2.400	1.407	2.400	1.407	-1035.596
0.4857	0.000	-1.654	1.265	3.551	1.602	-1033.784
0.3904	0.110	-0.290	0.931	3.287	1.557	-1026.990

TABLE 6. Estimated parameters fitting the MLRTN for two-day exchange rates between the euro and the dollar, with restriction and without restriction on  $\pi$  and  $c$ .

## 6. CONCLUDING REMARKS

- (1) The mechanism of agreement between ask prices and bid prices state the distribution of returns. The normal distribution, the mixture of normal distributions and the MLRTN distributions can be generated in this way. The latter case being more likely, from an economic point of view.
- (2) The MLRTN is a statistically satisfactory model. The log-likelihood function splits on three terms, corresponding to the likelihoods of three exponential models. The MLE exists and is unique for samples large enough.
- (3) From the empirical point of view, the MLRTN we propose provides satisfactory fits for exchange rates between the euro and the US dollar. These fits improve on those obtained with the normal distribution, the Laplace distribution, the mixture of normal distributions and the normal inverse Gaussian distribution.

- (4) The sufficient statistic of the MLRTN provides new tools for the analysis of financial data. The mean values of the right part and the left part of the returns distribution are of the same order as the standard deviation and have an economic interpretation. The coefficient of variation of the right part and the left part of the returns distribution provides a new way of measuring the non-normality of tail distributions. This measurement uses only two moments and is therefore more stable and more widely applicable than kurtosis.

#### 7. ACKNOWLEDGMENTS

The authors are grateful to the anonymous referees for their valuable comments and suggestions. The research was partially supported by the Spanish Ministerio de Educación y Cultura, grant MTM2006-01477.

#### 8. REFERENCES

- : Barndorff-Nielsen, O.E. (1998). "Processes of Normal Inverse Gaussian Type". *Finance and Stochastics*, vol. 2: 41-61.
- : Bibby, B.M. and Sorensen, M. (2003), Hyperbolic Processes in Finance. in S. Rachev (ed.): *Handbook of Heavy Tailed Distributions in Finance*. Elsevier Science, 2003, pp 211-248.
- : Castillo, J. (1994). "The Singly Truncated Normal Distribution, a Non-Steep Exponential Family". *Annals of the Institute of Mathematical Statistics*. vol. 46: 57-66.
- : Castillo, J. and Lee, Y. (2008) "GLM-methods for volatility models". *Statistical Modelling*. vol 8: 263-283.
- : Damilano, G. and Puig, P. (2004). "Efficiency of Linear Combination of the Median and the Sample Mean: The Double Truncated Normal Distribution. *Scandinavian Journal of Statistics* Vol. 31, 629-637.
- : Hull, J. and White, A. (1998). "Value at Risk When Daily Changes in Marked Variables are not Normally Distributed". *The Journal of Derivatives*, vol. 5, num. 3: 9-19.
- : Pedersen, C. and Hwang,S. (2007). Does downside beta matter in asset pricing ? *Applied Financial Economics*, vol 17, 961-978.
- : Taylor, S. (2005). "Asset Price Dynamics, Volatility and Prediction". *Princeton University Press*.



## ESTIMATION OF GENERALIZED PARETO DISTRIBUTION

JOAN DEL CASTILLO &amp; JALILA DAOUDI

**Abstract**<sup>1</sup>: This paper provides precise arguments to explain the anomalous behavior of the likelihood surface when sampling from the generalized Pareto distribution for small or moderate samples. The behavior of the profile-likelihood function is characterized in terms of the empirical coefficient of variation. A sufficient condition is given for global maximum of the likelihood function of the Pareto distribution to be at a finite point.

**Keywords**: Heavy-tailed inference. Extreme value theory. The coefficient of variation.

## 1. INTRODUCTION

The Pareto distribution has long been used as a model for the tails of another long-tailed distribution, see Arnold (1983). Applications to risk management in finance and economics are now of increasing importance. Since Pickands (1975), it has been well known that the conditional distribution of any random variable over a high threshold is approximately generalized Pareto distribution (GPD), which includes the Pareto distribution, the exponential distribution and distributions with bounded support. These distributions are closely related to the extreme value theory (Coles, 2001, and Embrechts et al. 1997).

The GPD has been used by many authors to model exceedances in several fields such as hydrology, insurance, finance and environmental science, see Finkenstadt and Rootzén (2003), Coles (2001) and Embrechts et al. (1997). In general, GPD can be applied to any situation in which the exponential distribution might be used but in which some robustness is required against heavier tailed or lighter tailed alternatives, see Van Montford and Witter (1985). The asymptotic behavior of maximum likelihood estimators was studied by Davison (1984) and Smith (1985). Nevertheless, there is evidence that numerical techniques for maximum likelihood estimation do not work well in small samples and other estimation methods have been proposed, see Castillo and Hadi (1997) and Hosking and Wallis (1987).

The paper provides precise arguments to explain the anomalous behavior of the likelihood surface when sampling from the GPD distribution (Davison and Smith, 1990, and Castillo and Hadi, 1997). In Section 2, Theorem 1 proves that the behavior of the profile-likelihood for GPD is characterized by the sign of the empirical coefficient of variation of the sample. Corollary 1 proves that the maximum of the likelihood function for the Pareto distribution is at a finite point, for samples in which the coefficient of variation is larger than 1. An example in the Appendix shows that a local maximum for the likelihood function of the GPD could not exist when the condition is not fulfilled.

Monte Carlo simulation in Section 3 raises the problem of mis-specification for small or moderate samples in GPD. It can be also explained from the difference

---

<sup>1</sup>Research partially supported by the Spanish Ministerio de Educacion y Ciencia, grant MTM2006-01477.

between the theoretical and the empirical coefficients of variation for small samples. The practical relevance of these results is also discussed.

## 2. MAIN RESULTS

The cumulative distribution function of GPD is

$$(2.1) \quad F(x) = 1 - (1 + \xi x/\psi)^{-1/\xi},$$

where  $\psi > 0$  and  $\xi$  are scale and shape parameters. For  $\xi > 0$  the range of  $x$  is  $x > 0$  and the GPD is just one of several forms of the usual *Pareto family of distribution* often called the Pareto distribution. For  $\xi < 0$  the range of  $x$  is  $0 < x < \psi/|\xi|$ , then GPD have bounded support. The limit case  $\xi = 0$  corresponds to the exponential distribution.

An alternative parameterization is  $\sigma = \psi/|\xi|$  and  $|\xi| = s\xi$ , where  $s = \text{sign}(\xi)$ . Then, the probability density function for GPD is given by

$$(2.2) \quad f(x; \sigma, \xi) = \frac{1}{\sigma|\xi|} \left(1 + s \frac{x}{\sigma}\right)^{-(1+\xi)/\xi},$$

for  $\xi < 0$ , the range of  $x$  is now  $0 < x < \sigma$ . Using this notation the two families of distributions corresponding to  $\xi > 0$  and  $\xi < 0$  can be studied at the same time.

Given a sample  $\{x_i\}$  of size  $n$ , the log-likelihood function for GPD distribution, divided by the sample size, is

$$(2.3) \quad l(\sigma, \xi) = -\log(s\xi\sigma) - \frac{1+\xi}{\xi n} \sum_{i=1}^n \log(1 + s x_i/\sigma).$$

If  $\xi < 0$  it is assumed  $\sigma > M = \max\{x_i\}$ , otherwise the likelihood is zero. In this case the likelihood may be made arbitrarily large as  $\sigma$  tends to  $M$ , so the maximum likelihood estimators are taken to be the values which yield a local maximum of (2.3), that often appears.

Maximum likelihood estimation of generalized Pareto parameters was discussed by Davison (1984) and Smith (1985). In particular, for large samples, maximum likelihood estimator exist and is asymptotically normal and efficient, provided that  $-0.5 < \xi$ . The restriction  $-0.5 < \xi < 0.5$  is usually assumed for both practical and theoretical reasons, since GPD with  $\xi < -0.5$  have finite end points and the probability density function is strictly positive at each endpoint, and GPD with  $\xi > 0.5$  have infinite variance. When GPD is used as an alternative to the exponential distribution, values of  $\xi$  near 0 will be of greatest interest, because the exponential distribution is a GPD with  $\xi = 0$ .

For moderate or small samples, anomalous behavior of the likelihood surface can be encountered when sampling from the GPD distribution (Davison and Smith, 1990, and Castillo and Hadi, 1997). This will be explained in this paper from the coefficient of variation of the sample. For instance, the coefficient of variation for Pareto distribution ( $\xi < 0.5$ ) is given by

$$(2.4) \quad \zeta = \sqrt{1/(1-2\xi)} > 1,$$

but for small samples the *empirical* coefficient of variation

$$(2.5) \quad cv = \sqrt{m_2 - m_1^2}/m_1,$$

where  $m_k = \sum x_i^k/n$  are the sample moments, may be lower than 1. Theorem 1 and Corollary 1 below in this Section provide more precise arguments.

Equating to zero the derivative of  $l(\sigma, \xi)$  in (2.3), with respect to  $\xi$ , we find  $\hat{\xi} = \xi(\sigma)$ , where

$$(2.6) \quad \xi(\sigma) \equiv \xi(\sigma, s) = \frac{1}{n} \sum_{i=1}^n \log(1 + s x_i/\sigma).$$

The profile-likelihood is given by

$$(2.7) \quad l_p(\sigma, s) = -\log[s \xi(\sigma) \sigma] - \xi(\sigma) - 1.$$

**Proposition 1.** *The following limits hold:*

$$\begin{aligned} \lim_{\sigma \rightarrow \infty} \sigma \log(1 + x/\sigma) &= x, & \lim_{\sigma \rightarrow \infty} \sigma x/(\sigma + x) &= x, \\ \lim_{\sigma \rightarrow \infty} \sigma^2 (\log(1 + x/\sigma) - x/(\sigma + x)) &= x^2/2. \end{aligned}$$

*Proof.* It is an elementary exercise in calculus, using series expansion.  $\square$

**Proposition 2.** *Let  $l_p(\sigma, s)$  be the profile-likelihood, defined by (2.7) and let  $\bar{x}$  be the sample mean, then*

$$\begin{aligned} \lim_{\sigma \rightarrow \infty} \xi(\sigma) &= 0, & \lim_{\sigma \rightarrow \infty} \sigma \xi(\sigma) &= s \bar{x}. \\ l_0 &\equiv \lim_{\sigma \rightarrow \infty} l_p(\sigma, s) = -\log(\bar{x}) - 1. \end{aligned}$$

*Proof.* The first limit is straightforward. From Proposition 1 it follows:

$$\lim_{\sigma \rightarrow \infty} \sigma \xi(\sigma) = \lim_{\sigma \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma \log(1 + s x_i/\sigma) = s \bar{x}$$

This prove the second limit and hence, since  $s^2 = 1$ , the last limit follows.  $\square$

**Remark 1.** *The limit of the profile-likelihood,  $l_p(\sigma, s)$ , as  $\sigma$  tends to infinity, in Proposition 2, corresponds to the log-likelihood of the exponential distribution for the same sample. More precisely, the log-likelihood function for the exponential distribution,  $\sigma e^{-\sigma x}$ , divided by the sample size  $n$ , is  $l(\sigma, 0) = \log \sigma - \sigma \bar{x}$  and the maximum likelihood estimator is  $\hat{\sigma} = 1/\bar{x}$  then,*

$$(2.8) \quad l(\hat{\sigma}, 0) = l_0 = -\log(\bar{x}) - 1.$$

**Theorem 1.** *For the Pareto distribution ( $\xi > 0$ ), if the empirical coefficient of variation is  $cv > 1$  then  $l_p(\sigma, 1)$  is a monotonous decreasing function for sufficiently large  $\sigma$ , and if  $cv < 1$  it is monotonous increasing. For the distributions with bounded support in GPD ( $\xi < 0$ ), if  $cv > 1$  then  $l_p(\sigma, -1)$  is a monotonous increasing function for sufficiently large  $\sigma$ , and if  $cv < 1$  it is monotonous decreasing.*

*Proof.* The derivative of (2.7) with respect to  $\sigma$  is given by

$$-s l'_p(\sigma, s) = (\xi(\sigma) + \sigma \xi'(\sigma) + \sigma \xi(\sigma) \xi'(\sigma)) / (\sigma |\xi(\sigma)|).$$

and the sign of  $-s l'_p(\sigma, s)$  is the same as the sign of  $num(\sigma) = \xi(\sigma) + \sigma \xi'(\sigma) + \sigma \xi(\sigma) \xi'(\sigma)$ , since  $\sigma |\xi(\sigma)| > 0$ .

Taking derivative with respect to  $\sigma$  in (2.6) it follows

$$\sigma \xi'(\sigma) = -\frac{1}{n} \sum_{i=1}^n s x_i / (\sigma + s x_i),$$

hence

$$\begin{aligned} \text{num}(\sigma) &= \frac{1}{n} \sum_{i=1}^n (\log(1 + s x_i / \sigma) - s x_i / (\sigma + s x_i)) - \\ &\quad \left( \frac{1}{n} \sum_{i=1}^n \log(1 + s x_i / \sigma) \right) \left( \frac{1}{n} \sum_{i=1}^n s x_i / (\sigma + s x_i) \right). \end{aligned}$$

From Proposition 1, we have

$$\lim_{\sigma \rightarrow \infty} \{\sigma^2 \text{num}(\sigma)\} = \frac{1}{2n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

Finally, note that  $\frac{1}{2n} \sum_{i=1}^n x_i^2 - \bar{x}^2 > 0$  is equivalent to  $(m_2 - m_1^2) > m_1^2$  and equivalent to  $cv > 1$ .  $\square$

A first consequence of Theorem 1 is obtained immediately.  $l_p(\sigma, s)$  tends to  $l_0$ , from Proposition 2, and  $l_p(\sigma, s)$  is a monotonous function for sufficiently large  $\sigma$  (Theorem 1) then these facts determine whether  $l_p(\sigma, s)$  is greater or less than  $l_0$ . If  $cv > 1$ ,  $l_p(\sigma, 1)$  is a monotonous decreasing function and  $l_p(\sigma, -1)$  is a monotonous increasing function, for sufficiently large  $\sigma$ , then

$$(2.9) \quad l_p(\sigma, 1) > l_0 > l_p(\sigma, -1).$$

In the same way, if  $cv < 1$  then

$$(2.10) \quad l_p(\sigma, 1) < l_0 < l_p(\sigma, -1),$$

for sufficiently large  $\sigma$ .

**Remark 2.** From (2.6), as  $\sigma$  tends to infinite  $|\xi(\sigma)|$  tends to zero. Hence, for  $\xi$  in a neighborhood of zero, the inequalities (2.9) and (2.10) show that if  $cv > 1$  the Pareto distribution is more likely than the exponential distribution and if  $cv < 1$  a bounded support distribution in GPD is more likely than the exponential distribution. These facts are numerically relevant for an algorithm to obtain the maximum likelihood estimator in GPD.

**Corollary 1.** Given a sample  $\{x_i\}$  of positive numbers with an empirical coefficient of variation  $cv > 1$ , the likelihood function for the Pareto distribution has a global maximum at a finite point and the maximum is higher than the maximum for the likelihood function of the exponential distribution.

*Proof.* For small values of  $\sigma$  we write

$$l_p(\sigma, 1) = -(\log \sigma + \xi(\sigma)) - \log[\xi(\sigma)] - 1,$$

$(\log \sigma + \xi(\sigma))$ , tends to  $\frac{1}{n} \sum_{i=1}^n \log(x_i)$  and  $\log[\xi(\sigma)]$  goes to  $\infty$ . This proves

$$\lim_{\sigma \rightarrow 0} l_p(\sigma, 1) = -\infty.$$

Since  $l_p(\sigma, 1)$  is a continuous and monotonous decreasing function for sufficiently large  $\sigma$  (Theorem 1) the last limit prove that a global maximum exists. Inequality (2.9) shows that it is up the maximum for the exponential distribution.  $\square$

When the coefficient of variation of the sample is  $cv < 1$  there may be no maximum likelihood estimator for the Pareto distribution and neither is there a local maximum for the parameter space of the bounded support distributions in the GPD. In the Appendix we give a simple example of this situation.

If  $l_p(\sigma, -1)$  has a local minimum, as is usual, and  $cv < 1$  then Theorem 1 proves that there is a local maximum for the likelihood function, since  $l_p(\sigma, -1)$  increase on the right side of the minimum and is monotonous decreasing for large values of  $\sigma$ .

### 3. DISCUSSION

Analytical maximization of the log-likelihood for GPD is not possible, so numerical techniques are required taking care to avoid numerical instabilities when  $\xi$  near zero (Coles, 2001, pp 81). Theorem 1 clarifies the behavior of the likelihood function in terms of the coefficient of variation of the sample. If  $cv > 1$  the Pareto distribution is more likely than a bounded support distribution in a neighborhood of zero, and if  $cv < 1$  a bounded support distribution is more likely than a Pareto distribution. The numerical algorithms have to consider the  $cv$  sign of the sample

Corollary 1 proves that the likelihood function for the Pareto distribution has a global maximum at a finite point for samples in which  $cv > 1$ , so it is extremely simple to find it numerically. This is a sufficient condition and we also believe, from numerical experiments, that it is necessary, although we are not able to prove it.

Hosking and Wallis (1987, pp 343) say "we conclude that the vast majority of failures of the algorithms are caused by the nonexistence of a local maximum of the likelihood function rather than by failure of our algorithm to find a local maximum". We agree with them. Moreover, the example given in the Appendix proves the nonexistence of a local maximum for a particular sample. Now it is clear that the nonexistence of a local maximum it is possible. This problem increases when  $\xi$  decreases, specially for the bounded support distributions in GPD, as Hosking and Wallis (1987) showed.

Sampling from Pareto distribution in GPD shows another problem. A simulation experiment was run to compute mis-specification for sample sizes  $n = 15, 25, 50, 100$  and shape parameter  $\xi = 0.1, 0.2, 0.3, 0.4$ . The scale parameter  $\sigma$  was set to 1, since the model (2.2) is invariant under scale changes. For each combination of values of  $n$  and  $\xi$ , 50,000 random samples were generated from the Pareto distribution and the number of times the parameter estimates  $\hat{\xi}$  to be positive, negative or that the algorithm does not converge are reported. It is noted from (2.4) that the theoretical coefficient of variation for Pareto distribution is  $\zeta > 1$ , but for small samples the empirical coefficient of variation,  $cv$ , may be lower than 1. Table 1 shows that for  $\xi = 0.3$  (a distribution with infinite kurtosis) and sample size  $n = 15$ , 29% of cases that lead to a wrong decision,  $\hat{\xi} < 0$  (bounded support distribution), while in 4.8% cases the algorithm does not converge. The problem remains for larger samples. For  $\xi = 0.1$  and sample size  $n = 100$ , 24.9% of cases have  $\hat{\xi} < 0$ . However, if we assume Pareto distribution without considering the global GPD model (with the bounded support distributions), then samples with  $cv > 1$  lead to Pareto distribution and samples with  $cv < 1$  lead to the exponential distribution,  $\xi = 0$ , in both cases the support for the distribution is  $(0, \infty)$ .

In the context of heavy-tailed inference assuming that the true distribution has support in  $(0, \infty)$ , an alternative model for samples with  $cv < 1$  may be truncated

normal distribution. Castillo (1994) shows that the likelihood equations for truncated normal distribution have a solution if and only if the empirical coefficient of variation is  $cv < 1$ . Pareto distribution and truncated normal distribution are two complementary families of distributions the former with theoretical coefficient of variation  $\zeta > 1$  and the latter with  $\zeta < 1$ . In both cases the exponential distribution is the limit distribution as  $\zeta$  tends to 1.

$\xi$	0.4			0.3			0.2			0.1		
	$\hat{\xi} > 0$	$\hat{\xi} < 0$	NC	$\hat{\xi} > 0$	$\hat{\xi} < 0$	NC	$\hat{\xi} > 0$	$\hat{\xi} < 0$	NC	$\hat{\xi} > 0$	$\hat{\xi} < 0$	NC
15	73.4	23.1	3.5	66.2	29.0	4.8	57.1	36.4	6.5	46.8	44.5	8.7
25	84.5	15.3	0.2	77.2	22.4	0.3	66.9	32.6	0.5	53.5	45.7	0.9
50	95.8	4.2	0.0	90.7	9.3	0.0	80.5	19.5	0.0	63.8	36.2	0.0
100	99.6	0.4	0.0	98.0	2.0	0.0	92.3	7.7	0.0	75.1	24.9	0.0

Table 1. Random samples generated from the Pareto distribution for sample sizes  $n = 15, 25, 50, 100$  and shape parameters  $\xi = 0.1, 0.2, 0.3, 0.4$ . The frequency with which the parameter estimate  $\hat{\xi}$  is positive, negative or the algorithm does not converge (NC), are reported.

#### 4. BIBLIOGRAPHY

- (1) Arnold, B.(1983). *Pareto distributions*. International Co-operative Publishing House. Fairland, Maryland.
- (2) Castillo, E. and Hadi, A. (1997). "Fitting the Generalized Pareto Distribution to Data". *Journal of the American Statistical Association*, 92, 1609-1620.
- (3) Castillo, J. (1994). "The Singly Truncated Normal Distribution, a Non-Steep Exponential Family". *Annals of the Institute of Mathematical Statistics*. 46, 57-66.
- (4) Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Springer, London.
- (5) Davison, A (1984). Modelling Excesses Over High Thresholds, with an Application, in *Statistical Extremes and Applications*, ed. J.Tiago de Oliveira, Dordrecht: D.Reidel,pp. 461-482.
- (6) Davison, A. and Smith, R. (1990). "Models for Exceedances over High Thresholds". *J.R. Statist. Soc. B*, 52, 393-442.
- (7) Embrechts, P. Klüppelberg, C. and Mikosch, T. (1997). *Modeling Extremal Events for Insurance and Finance*. Springer, Berlin.
- (8) Finkenstadt, B.and Rootzén, H. (edit) (2003). *Extreme values in Finance, Telecommunications, and the Environment*. Chapman & Hall .
- (9) Hosking, J. and Wallis, J. (1987). "Parameter and quantile estimation for the generalized Pareto distribution". *Technometrics* 29 , 339-349.
- (10) Pickands, J. (1975). "Statistical inference using extreme order statistics". *The Annals of Statistics* 3, 119-131.
- (11) Smith, R. (1985). "Maximum likelihood estimation in a class of nonregular cases". *Biometrika* 72, 67-90.
- (12) Van Montford, M. and Witter, J. (1985). Testing Exponentiality Against Generalized Pareto Distribution. *Journal of Hydrology*, 78, 305-315.

## 5. APPENDIX

Let us examine the sample of size two given by  $\{x_1, x_2\} = \{1, 2\}$  for the GPD. First, we will show that the profile-likelihood  $l_p(\sigma, 1)$ , given by (2.7), is a monotonous increasing function for  $\sigma > 0$  and, hence, there is no maximum likelihood estimator for the Pareto distribution with this sample.

The derivative  $l'_p(\sigma, 1)$  is given by

$$num(\sigma) = 8 + 6\sigma - \sigma(3 + 2\sigma)(\log(1 + 1/\sigma) + \log(1 + 2/\sigma)),$$

divided by a positive function. Hence,  $l'_p(\sigma, 1)$  and  $num(\sigma)$  have equal sign and we will see that it is positive.

The following results hold for the function  $num(\sigma)$  and for its derivatives:

$$(5.1) \quad \lim_{\sigma \rightarrow \infty} num(\sigma) = 4, \quad \lim_{\sigma \rightarrow \infty} num'(\sigma) = 0, \quad \lim_{\sigma \rightarrow \infty} num''(\sigma) = 0,$$

$$(5.2) \quad num'''(\sigma) = -\frac{48 + 152\sigma + 162\sigma^2 + 69\sigma^3 + 9\sigma^4}{\sigma^2(2 + 3\sigma + \sigma^2)^3} < 0.$$

Then,  $num''(\sigma)$  is a monotonous decreasing function and, from the limit zero property,  $num''(\sigma) > 0$ . Then,  $num'(\sigma)$  is monotonous increasing and, hence,  $num'(\sigma) < 0$ . Therefore,  $num(\sigma)$  is a monotonous decreasing function, hence  $num(\sigma) > 4$ , and its sign is always positive, as we said.

We will also show that the profile-likelihood  $l_p(\sigma, -1)$  with the same sample,  $\{1, 2\}$ , is a monotonous decreasing function for  $\sigma > 2$  and, hence, does not exist a local maximum for the parameter space of the bounded support distributions in the GPD.

The derivative  $l'_p(\sigma, -1)$  is given by

$$nu(\sigma) = 8 - 6\sigma + \sigma(3 - 2\sigma)(\log(1 - 1/\sigma) + \log(1 - 2/\sigma)),$$

divided by a negative function. We will see that the sign of  $nu(\sigma)$  is positive for  $\sigma$  greater than 2.

The following results hold for the function  $nu(\sigma)$  and for its derivatives:

$$(5.3) \quad \lim_{\sigma \rightarrow \infty} nu(\sigma) = 4, \quad \lim_{\sigma \rightarrow \infty} nu'(\sigma) = 0, \quad \lim_{\sigma \rightarrow \infty} nu''(\sigma) = 0,$$

$$(5.4) \quad nu'''(\sigma) = \frac{48 - 152\sigma + 162\sigma^2 - 69\sigma^3 + 9\sigma^4}{\sigma^2(2 - 3\sigma + \sigma^2)^3}.$$

If  $\sigma > 2$ , the denominator of  $nu'''(\sigma)$  is positive; the greater real root of the numerator is at  $\sigma_1 = 4.36556$ , then  $nu'''(\sigma) > 0$ , for greater values of  $\sigma$ . Therefore, for  $\sigma > \sigma_1$ ,  $nu''(\sigma)$  is a monotonous increasing function and, from the limit zero property,  $nu''(\sigma) < 0$ . Then,  $nu'(\sigma)$  is monotonous decreasing and, hence,  $nu'(\sigma) > 0$ , for  $\sigma > \sigma_1$ . Finally,  $nu(\sigma)$  is a monotonous increasing function for  $\sigma > \sigma_1$ . It is also clear that  $\lim_{\sigma \downarrow 2} nu(\sigma) = \infty$ .

For  $2 < \sigma < \sigma_1$ ,  $nu(\sigma)$  has a minimum at  $\sigma_0 = 2.89221$  and  $nu(\sigma_0) = 3.53456 > 0$ . Then, we conclude that the sign of  $nu(\sigma)$  is positive for  $\sigma > 2$ ,  $l_p(\sigma, -1)$  is a monotonous decreasing function for  $\sigma > 2$  and does not exist a local maximum for the likelihood function of the GPD model with this sample.



# Bibliografía

- [Arn83] B. Arnold, *Pareto distributions*, “Internacional Co-operative Publishing House. Fairland, Maryland”, 1983.
- [BG74] R. Blattberg and N. Gonedes, *A comparison of the stable and student distributions as statistical models for stock prices*, *Journal of Business* **47** (1974), 244–280.
- [BGT89] N. Bingham, C. Goldie, and J. Teugels, *Regular variation*, Cambridge University Press, Cambridge, 1989.
- [BN98] O. Barndorff-Nielsen, *Processes of normal inverse Gaussian type*, *Finance Stoch.* **2** (1998), no. 1, 41–68.
- [BNS01] O. Barndorff-Nielsen and N. Shephard, *Modelling by Lévy processes for financial econometrics*, Lévy processes, Birkhäuser Boston, Boston, MA, 2001, pp. 283–318.
- [BO02] J. Berkowitz and J. O’Brien, *How accurate are value-at-risk models at commercial banks?*, *The Journal of Finance* **57** (2002).
- [BP81] R. Barlow and F. Proshan, *Statistical theory of reliability and life testing*, “To begin with”, 1981.
- [BS03] B. Bibby and M. Sorensen, *Hiperbolic process in finance*, Rachev (ed.):Handbooh of Heavy Tailed Distributions in Finance. Elsevier Science (2003), 211–248.
- [BT96] J. Beirlant and J. Teugels, *Excess functions and estimation of the extreme-value index*, *Bernoulli* **2** (1996), no. 4, 293–318.

- [Cas94] J. Castillo, *The singly truncated normal distribution, a non-steep exponential family*, Annals of the Institute of Mathematical Statistics **46** (1994), 57–66.
- [Cas02] ———, *Distribucions amb cues pesades en dades financeres*, IEC (2002).
- [CD09a] J. Castillo and J. Daoudi, *Estimation of the generalized pareto distribution*, Statistics and Probability Letters. **79** (2009), 684–688.
- [CD09b] ———, *The mixture of left-right truncated normal distributions*, Journal of Statistical Planning and Inference, doi: 10.1016/j.jspi.2009.04.010 (2009).
- [CH97] E. Castillo and A. Hadi, *Fitting the generalized Pareto distribution to data*, J. Amer. Statist. Assoc. **92** (1997), no. 440, 1609–1620.
- [Chi64] V. Chistyakov, *A theorem on sums of independent positive random variables and its applications to branching random processes*, Theory Probab. Appl. **9** (1964), 640–648.
- [CL08] J. Castillo and Y. Lee, “*glm-methods for volatility models*”, Statistical Modelling **8** (2008).
- [CLM97] J. Compbell, A. Lo, and A. MacKinlay, *The econometrics of financial markets.*, Princeton University Press, Princeton, 1997.
- [Col01] S. Coles, *An introduction to statistical modeling of extreme values*, Springer Series in Statistics, Springer-Verlag London Ltd., London, 2001.
- [CP99] J. Castillo and P. Puig, *The best test of exponenciality against singly truncated normal alternatives*, Journal of the American Statistical Association (1999), 529–532.
- [CS06] S. Carrillo and A. Suárez, *Medición efectiva del riesgo operacional*, Estabilidad Financiera n 11 (2006), 61–90.
- [Dav84] A. Davison, *Modeling excesses over high thresholds, with applications, in statistical extremes and applications*, 461–482.
- [DP04] G. Damilano and P. Puig, *Efficiency of a linear combination of the median and the sample mean: the double truncated normal distribution*, Scand. J. Statist. **31** (2004), no. 4, 629–637.

- 
- [DS90] A. Davison and R. Smith, *Models for exceedances over high thresholds*, J. Roy. Statist. Soc. Ser. B **52** (1990), no. 3, 393–442, With discussion and a reply by the authors.
- [EKM97] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling extremal events*, Springer-Verlag, Berlin, 1997, For insurance and finance.
- [Emb80] P. Embrechts, *A characterization of Fourier-Stieltjes transforms*, Nieuw Arch. Wisk.
- [Fec97] G. Fechner, *Fechner-Kollektivmasslehre*, Engleman, Leipzig (1897).
- [Fel81] V. Feller, *Introducción a la teoría de probabilidades y sus aplicaciones, volumen ii*, segunda ed., Limusa, 1981.
- [FR71] E. Fama and R. Roll, *Parameter estimation for symmetric stable distributions*, J. Amer. Statist. Assoc. **66** (1971), 817–836.
- [FR03] B. Finkenstadt and H. Rootzen, *Extreme values in finance*, “Chapman Hall”, 2003.
- [GK98] C. Goldie and C. Klüppelberg, *Subexponential distributions*, A practical guide to heavy tails (Santa Barbara, CA, 1995), Birkhäuser Boston, Boston, MA, 1998, pp. 435–459.
- [GK00] R. Gupta and S. Kirmani, *Residual coefficient of variation and some characterization results*, J. Statist. Plann. Inference **91** (2000), no. 1, 23–31.
- [Hec79] J. Hechman, *Sample selection bias as a specification error*, Econometrica **47** (1979), 153–161.
- [Hil75] B. Hill, *A simple general approach to inference about the tail of a distribution*, Ann. Statist. **3** (1975), 1163–1174.
- [HK04] R. Hogg and S. Klugman, *Loss distribution*, “New York: Wiley”, 2004.
- [HW87] J. Hosking and J. Wallis, *Parameter and quantile estimation for the generalized Pareto distribution*, Technometrics **29** (1987), no. 3, 339–349.
- [HW98] J. Hull and A. White, *Value at risk when daily changes in market variables are not normality distributed*, Journal of Derivatives (1998), 9–19.

- [JM82] P. Jupp and K. Mardia, *A characterization of the multivariate Pareto distribution*, Ann. Statist. **10** (1982), no. 3, 1021–1024.
- [Jor05] P. Jorion, *Financial risk manager handbook*, Wiley Finance, 2005.
- [Man63] B. Mandelbrot, *Modeling excesses over high thresholds, with applications, in statistical extremes and applications*, Econometrica **29** (1963).
- [MR93] S. Mittnik and S. T. Rachev, *Modeling asset returns with alternative stable distributions*, Econometric Rev. **12** (1993), no. 3, 261–389, With comments by P. C. B. Phillips, F. X. Diebold and R. T. Baillie and a reply by the authors.
- [Pic75] J. Pickands, III, *Statistical inference using extreme order statistics*, Ann. Statist. **3** (1975), 119–131.
- [Smi85] R. Smith, *Maximum likelihood estimation in a class of nonregular cases*, Biometrika **72** (1985).
- [SR77] P. Sullo and D. Rutherford, *Characterizations of the power distribution by conditional exceedance*, American Statistical Association (1977).
- [Ste81] M. Stephens, *Further percentage points for greenwood's statistic*, Journal of the Royal Statistical Society.A **42** (1981), 364–366.
- [Tay05] S. Taylor, *Asset price dynamics, volatility and prediction*, “Princeton University Press”, 2005.
- [Zol86] V. Zolotarev, *One-dimensional stable distributions*, American Mathematical Society, Providence, RI, 1986, Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver.