REVISTA INNOVAR JOURNAL

Un estudio comparativo de algunos estimadores del índice de cola ξ

Andrés Mora
Profesor investigador del Colegio de Estudios Superiores de Administración (CESA).
Correo electrónico: amora@cesa.edu.co

RESUMEN: Se presenta un análisis comparativo de manera teórica y práctica de algunos métodos elegidos para estimar el índice de cola para distribuciones tipo-Pareto. Como era de esperar, los resultados de la simulación muestran que no existe un método óptimo. Sin embargo, la técnica GPD extendida (EGPD, por sus siglas en inglés) ofrece mejoras comparado con el estimador de Hill. También se aplican los métodos elegidos al caso de datos de fuego en Dinamarca (*Danish Fire data*) para obtener su índice de cola y un estimador de cuantil alto.

PALABRAS CLAVE: teoría de valor extremo, índice de cola, estimador de Hill, GPD extendida.

Introducción

La regulación moderna de administración de riesgos requiere estimados de pérdidas potenciales con altos niveles de confiabilidad. Por tanto, se requieren modelos matemáticos sofisticados y herramientas computacionales. Es entonces cuando surge la teoría del valor extremo (EVT, por sus siglas en inglés) como una disciplina estadística, la cual desarrolla técnicas y modelos para describir los resultados inesperados, anormales o extremos. EVT ha sido aplicada en áreas de ingeniería, y recientemente se ha convertido en herramienta fundamental en aplicaciones para finanzas y seguros. Esta teoría brinda modelos para extrapolar datos observados a niveles no observados, debido a que los valores extremos son escasos. De esta manera, se puede decir que EVT se enfoca en el análisis de las colas de la distribución de pérdidas para medir grandes pérdidas que no son tan frecuentes.

Existen dos clases de modelos para tratar valores extremos: máximo por bloques (*block maxima*) y picos sobre el umbral (POT, por sus siglas en inglés). El método POT es la técnica más usada para analizar la cola de una función de distribución. Estas dos técnicas están basadas en modelos distribucionales obtenidos a partir de teorías asintóticas.

El problema por resolver es la estimación de los parámetros de estas distribuciones límites, en particular el parámetro de forma (representado por ξ), el cual determina el comportamiento de los valores extremos. Algunos autores denominan el parámetro de forma como el índice de valor extremo (*extreme value index*), mientras que otros lo denominan el índice de cola (*tail index*). Para evitar confusiones, se usará la expresión "índice de cola" cuando se refiera al parámetro de forma[1]. Establecer el umbral mediante el método POT conlleva a un *trade-off* entre sesgo y varianza en la estima-

A COMPARATIVE STUDY OF CERTAIN ESTIMATORS FOR THE TAIL INDEX ξ

ABSTRACT: We present a theoretical and practical comparative analysis of various selected methods to estimate the tail index for Pareto-type distributions. As it was expected, the simulation results show that there is no best method. However, the extended GPD (EGPD) method offers improvements compared to the Hill estimator. We also apply the selected methods to Danish Fire data in order to obtain its tail index and a high quintile estimator.

KEYWORDS: Extreme Value Theory (EVT), tail index, Hill estimator, Extended GPD (EGPD).

ÉTUDE COMPARATIVE DE CERTAINS ESTIMATEURS DE L'INDICE DE QUEUE ξ

RÉSUMÉ : Présentation d'une analyse comparative de manière théorique et pratique de certaines méthodes choisies pour estimer l'indice de queue pour des distributions type-Pareto. Comme prévu, les résultats de la simulation démontrent l'inexistence d'une méthode optimum. Cependant, la technique EGPD permet des améliorations en comparaison à l'estimateur d'Hill. Les méthodes choisies s'appliquent aussi dans le cas de *Danish Fire data* pour obtenir l'indice de queue et un estimateur de quantile élevé.

MOTS-CLEFS : théorie de valeur extrême, indice de queue, estimateur d'Hill, GPD étendue.

UM ESTUDO COMPARATIVO DE ALGUNS ESTIMADORES DO ÍNDICE DE CAUDAS ξ

RESUMO: Apresenta-se uma análise comparativa de maneira teórica e prática de alguns métodos eleitos para estimar o índice de caudas para distribuições tipo-Pareto. Como era de se esperar, os resultados da simulação mostram que não existe um método ótimo. Sem embargo, a técnica GPD estendida (EGPD, por suas siglas em inglês) oferece melhoras comparado ao estimador de *Hill*. Também aplicam-se os métodos eleitos *Danish Fire Data* para obter seu índice de caudas e um estimador de quantil elevado.

PALAVRAS CHAVE: Teoria do Valor Extremo, índice de caudas, estimador de Hill, GPD Estendida.

CLASIFICACIÓN JEL: C15, C16, G32.

RECIBIDO: enero de 2009 **APROBADO:** diciembre de 2010

CORRESPONDENCIA: Diagonal 35, No. 5-23, CESA - Casa Biblioteca, Bogotá (Colombia).

CITACIÓN: Mora, A. (2011). Un estudio comparativo de algunos estimadores del índice de cola ξ. *INNOVAR, 21*(40), 17-34.

[1] El autor agradece la recomendación por parte del evaluador en la traducción de este término, para evitar la confusión con otro concepto que es el *extremal index*, usado para dependencia serial en los extremos.

ción de los parámetros de la función de distribución que se asume para ajustar los valores extremos. Usar métodos basados en cuantiles (por ejemplo, el estimador de Hill), también depende de la elección *apropiada* de estadísticos ordenados superiores. Elegir el umbral óptimo (en el método POT) conduce al mismo problema de escoger el número óptimo k de estadísticos ordenados superiores (en el estimador de Hill). Si se utiliza el método máximo por bloques, sesgo se presenta con bloques muy pequeños y varianza en el caso contrario.

De esta manera, la estimación del parámetro de forma se convierte en un problema importante para estimar de manera confiable cuantiles altos como una medida de riesgo, siguiendo la técnica EVT. Sin embargo, la selección de k (o del umbral) para estimar el parámetro de forma no es una tarea fácil. El propósito de este documento es, entonces, revisar algunos métodos escogidos para estimar este parámetro, con el fin de mitigar los problemas anteriormente mencionados.

El documento está organizado como sigue: la sección 1 describe brevemente el marco conceptual de EVT. La sección 2 revisa los estimadores elegidos para modelos tipo-Pareto. La sección 3 presenta y comenta los resultados de las simulaciones al calcular los estimadores elegidos a varias funciones de distribución. En la sección 4 aparecen los resultados de los métodos aplicados a un caso, datos de fuego en Dinamarca (*Danish Fire data*). Finalmente, la sección 5 concluye el documento.

Introducción a la teoría del valor extremo (EVT)

Se puede decir que los precursores de la teoría del valor extremo son Gnedenko (1943) y Gumbel (1958)[2]. Textos que tratan el tema de EVT de manera muy completa y con enfoque más probabilístico son Embrechts *et al.* (1997), Resnick (1987) y de Haan y Ferreira (2006). Desde el punto de vista estadístico están los textos de Beirlant *et al.* (2004), Coles (2001), Falk *et al.* (2004) y Reiss & Thomas (1997). Textos más aplicados a finanzas y riesgos son McNeil *et al.* (2005), Malevergne y Sornette (2006) y Moix (2001). Un texto más reciente de Balkema y Embrechts (2007) presenta un enfoque geométrico de valores extremos.

[2] *Sur la distribution limité du terme maximum d'une série aléatoire* y *Statistics of Extremes*, respectivamente.

La teoría del valor extremo (EVT) ha cobrado importancia en el campo de riesgos financieros, puesto que los administradores de riesgo están interesados en estimar probabilidades en las colas y cuantiles de distribuciones de pérdidas, dado que los datos financieros presentan colas pesadas (*fat tails*).

La teoría del valor extremo dice que el valor más grande o más pequeño de un conjunto de valores tomados de la misma distribución original tiende a una distribución asintótica que solo depende de la cola de la distribución original. Se puede decir entonces que EVT es el estudio de las colas de las distribuciones. EVT ha sido utilizado para modelar eventos catastróficos en seguros y otros eventos financieros, como pérdidas inesperadas en crédito.

A continuación se describen brevemente los métodos usados en EVT, que son el máximo por bloques y picos sobre el umbral. Para más detalles, se recomienda ver, por ejemplo, el capítulo 7 de McNeil *et al*. (2005) y los capítulos 3 y 6 de Embrechts *et al.* (1997).

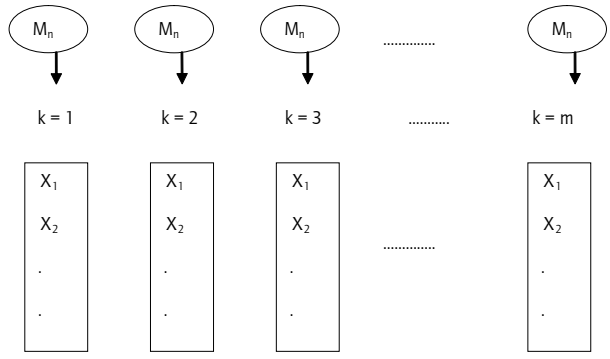
El método máximo por bloques

Suponga que se tienen datos de una distribución subyacente F desconocida y que los datos pueden ser divididos en m bloques de tamaño n. Se asume que la distribución extrema de los datos es de tipo valor extremo generalizado (GEV, por sus siglas en inglés) con algunos parámetros desconocidos ξ, μ, σ. Se pueden estimar estos parámetros mediante máxima verosimilitud (las observaciones de los máximos por bloque se asumen ser independientes). Pero la selección del tamaño del bloque conlleva un *trade-off* entre sesgo y varianza. Bloques pequeños generan sesgo y bloques muy grandes generan alta varianza en la estimación de los parámetros (ver el ejemplo 7.12 de McNeil *et al.* (2005) para una aplicación del método máximo por bloques a retornos del S&P). La figura 1 representa de manera gráfica este método.

Picos sobre el umbral (POT)

El método POT es útil para grandes observaciones que exceden un umbral alto u. Este método es más útil que el máximo por bloques en aplicaciones prácticas, debido al uso más eficiente de los datos en valores extremos (Gilli y Kellezi, 2003, p. 4). Desde el punto de vista de un administrador de riesgos, se está interesado en las pérdidas que exceden un umbral u. Este enfoque ha sido estudiado por Smith (1989), Davison y Smith (1990) y Leadbetter (1991), quienes muestran el uso práctico de POT en la teoría del valor extremo.

FIGURA 1. Representación gráfica del método máximo por bloques.



Fuente: elaboración propia.

El método POT.

Dado unos datos de pérdida X_1, ..., X_n de una función de distribución (desconocida) F, existe un número aleatorio N_u de pérdidas que excederá el umbral u. Estos datos se nombran como: $\widetilde{X}_1, \ldots, \widetilde{X}_{N_u}$. Para cada uno de estos excesos se calcula la cantidad $Y_j = \widetilde{X}_j - u$ de las pérdidas en exceso. Entonces, se desea estimar los parámetros de una distribución de Pareto generalizada (GPD, por sus siglas en inglés), ajustando esta distribución a los N_u pérdidas en exceso. Se estiman estos parámetros mediante máxima verosimilitud (se asume que los datos en exceso son independientes).

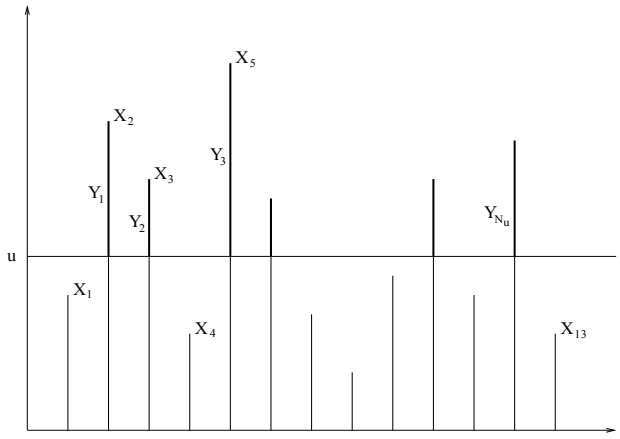
Sin embargo, como ya se anotó, la selección del umbral conlleva un *trade-off* entre sesgo y varianza en la estimación. Valores muy bajos del umbral generan sesgo en la estimación, mientras que valores muy altos del umbral generan alta varianza en la estimación.

Ver el ejemplo 7.23 de McNeil *et al*. (2005) de una aplicación del método POT a un caso de seguros, y el caso de datos de pérdida por fuego en Dinamarca (*Danish fire loss data*), el cual se retoma en la sección 4 de este documento. El ejemplo 7.24 de McNeil *et al.* (2005) aplica POT a datos de pérdida de la acción AT&T. Una representación gráfica del método POT se ilustra en la figura 2.

En el caso de datos de fuego en Dinamarca se calcularán también cuantiles altos como medidas de riesgo. Una medida de riesgo muy conocida es el VaR, cuya definición aquí presentada corresponde a la definición 2.10 de McNeil *et al*. (2005):

Definición 1: *(Value at Risk - VaR). Dado algún nivel de confiabilidad $\alpha \in (0, 1)$, el VaR de un portafolio al nivel de confiabilidad α está dado por el número más pequeño l, tal que la probabilidad de que la pérdida L exceda l no es más grande que $(1 - \alpha)$. Formalmente,*

FIGURA 2. Gráfico de los excesos de datos sobre un umbral u.



Fuente: tomado de Embrechts *et al.* (1997).

$$\mathrm{VaR}_{\alpha} = \inf\{l \in \mathrm{R} : P(L > l) \leq (1-\alpha)\} = \inf\{l \in \mathrm{R} : F_L(l) \geq \alpha\}.$$

En términos probabilísticos, VaR es simplemente un cuantil de la distribución de pérdidas. VaR fue desarrollado por J. P. Morgan en 1994 y rápidamente se convirtió en estándar de medida de riesgo para los reguladores y administradores de riesgo.

Estimación del índice de cola ξ

El método basado en GPD no es la única aproximación para estimar los parámetros, en específico el parámetro de forma, de la cola de una distribución. Para la selección del umbral, Smith (1987) desarrolló herramientas estadísticas que tienen en cuenta el *trade off* entre sesgo y varianza al estimar los parámetros. Otro método es el conocido enfoque de Hill que se describe a continuación.

El estimador de Hill

El *estimador de Hill* es el estimador más popular del parámetro de forma o el índice de cola, el cual está restringido al caso cuando $\xi > 0$. Es decir, que es aplicable para distribuciones que pertenecen al dominio máximo de atracción de Fréchet[3].

Dados unos estadísticos ordenados $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$, el estimador de Hill toma la siguiente forma:

$$\hat{\xi}_{k,n}^{(H)} = \frac{1}{k}\sum_{j=1}^{k} \ln X_{n-j+1,n} - \ln X_{n-k,n},\ k = 1,\ldots,n-1, \qquad (2.1)$$

donde $X_{n\text{-}k,n}$ puede ser visto como el umbral para el método POT. Los valores por encima del umbral se denotan como $X_{n-j+1,n}$ y j toma valores 1,..., k (k = 1,..., n − 1).

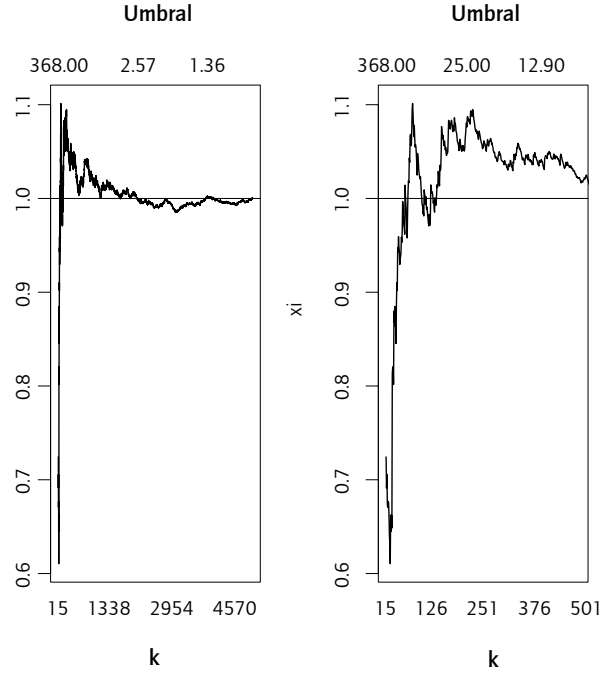
[3] Estas distribuciones son las más estudiadas en EVT por ser distribuciones de colas pesadas. Ver, por ejemplo, Embrechts *et al.* (1997) para más detalles.

Embrechts *et al.* (1997) muestran varios métodos para obtener el estimador de Hill. Beirlant *et al.* (2004) también obtienen el estimador de Hill mediante enfoques de cuantiles, de probabilidad, entre otros.

Por lo general, se usa el gráfico de Hill para estimar ξ, el cual es un gráfico de $\hat{\xi}_{k,n}^{(H)}$ versus k, y se escoge el estimado del índice de cola donde el gráfico es estable para algún valor de k no muy pequeño ni muy grande. Usualmente, se observa alta variabilidad en el gráfico para valores pequeños de k. Esto se debe a que hay muy pocas observaciones de datos extremos (pérdidas) y gran diferencia entre ellos. Si se escoge un valor grande de k, donde se muestran casi todos los datos, se puede observar sesgo en el gráfico de Hill.

La figura 3 muestra un ejemplo del gráfico de Hill para 5000 datos iid de una distribución de Pareto (α = 1) con ξ =1.

FIGURA 3. Gráfico de Hill para 5000 datos iid (izquierda) de una distribución de Pareto (α = 1) con ξ = 1, y su versión expandida (derecha) hasta 600 estadísticos ordenados. La línea horizontal muestra el verdadero valor de ξ.



Fuente: elaboración propia.

El estimador de Hill parece dar una buena aproximación del verdadero valor de ξ para el ejemplo anterior.

De acuerdo con Drees *et al.* (2000), el gráfico de Hill es de gran ayuda cuando la distribución de los datos es (o muy cercana a) una distribución Pareto. Puesto que el estimador de Hill es el estimador máximo verosímil de una distribución Pareto, se espera que el gráfico de Hill se aproxime

al verdadero valor de ξ en el lado derecho del gráfico. Esta conclusión se puede observar en la figura 3.

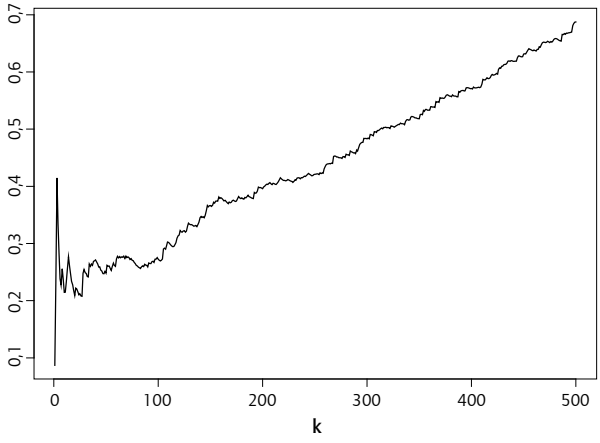
Para datos iid, el estimador de Hill es (débilmente) consistente[4] para ξ.

Si $k, n \to \infty$, y $k/n \to 0$, entonces:

$$\hat{\xi}_{k,n}^{(H)} \xrightarrow{P} \xi$$

Cabe preguntarse, ¿funciona el estimador de Hill en la realidad? La siguiente figura se tomó del artículo de Matthys y Beirlant (2000), y representa los estimados mediante el enfoque de Hill para caídas en porcentaje del S&P 500. El periodo de observación es enero de 1980 a octubre 16 de 1987.

FIGURA 4. Gráfico de Hill para las caídas del S&P 500.



Fuente: Matthys y Beirlant (2000).

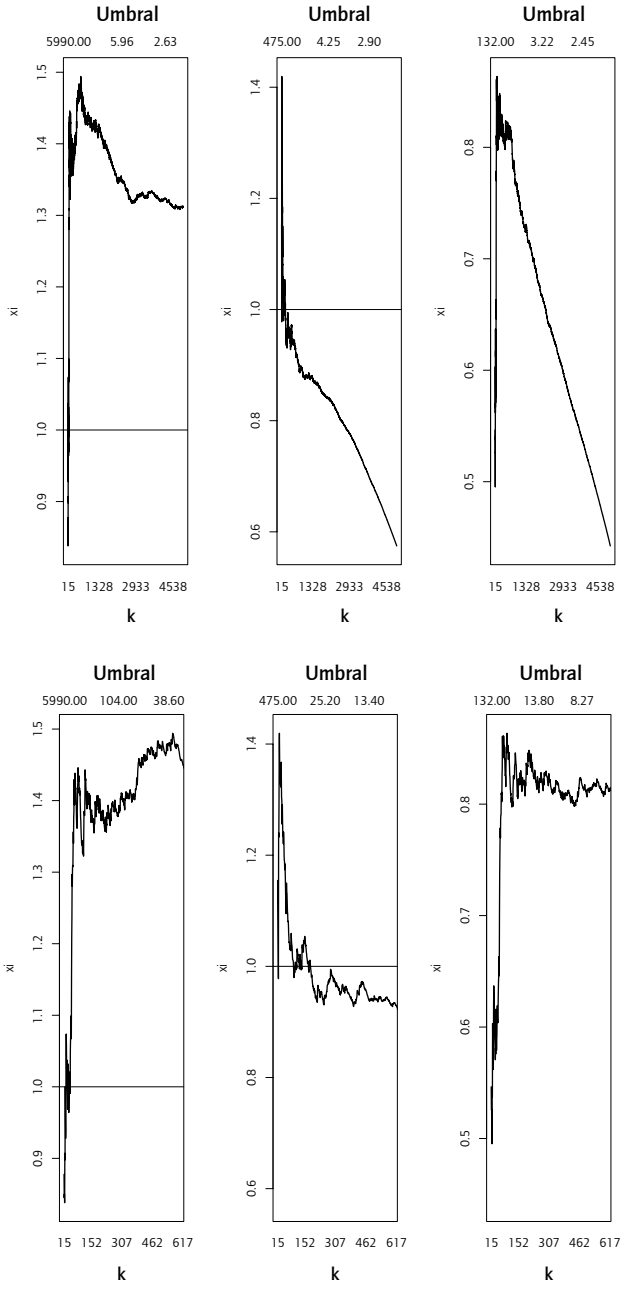
En este ejemplo real no es claro qué valor de k escoger para estimar el índice de cola. Para pequeños valores de k se nota gran varianza del estimador, pero si k es relativamente alto, se observa un sesgo significativo en la estimación del índice de cola.

La figura 5 muestra otros ejemplos con funciones conocidas.

En los ejemplos de la figura 5, los gráficos de Hill no son útiles para escoger k, y por consiguiente no se puede tener un estimado confiable de ξ. Los datos en la figura 5 corresponden a las siguientes funciones de supervivencia $(1 - F(x))$:

- $x^{-1/\xi}(1 + x^{-0.5})$ (izquierda)
- $x^{-1/\xi}(1 \,/\, \log x)$ (centro)
- $x^{-1/\xi}(1 \,/\, \sqrt{\log x})$ (derecha)

FIGURA 5. Gráficos de Hill para 5000 datos iid de algunas distribuciones tipo-Pareto con $\xi = 1$ (arriba), y su versión expandida (abajo). La línea horizontal indica el verdadero valor de ξ.



Fuente: elaboración propia.

La función de supervivencia de la distribución de Pareto es $cx^{-1/\xi}$, donde la función de variación lenta L[5], es igual a la

[4] Para más detalles de propiedades del estimador de Hill, ver Embrechts *et al*. (1997).

[5] Una función positiva, Lebesgue-medible L en $(0,\infty)$ es de variación lenta en ∞ si:

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

Como ejemplos de funciones de variación lenta están las constantes, logarítmicas, potencias de logaritmos y funciones de logaritmos iterados.

constante c. El problema central radica en que una distribución tipo-Pareto no significa que la cola es exactamente como la de una distribución de Pareto, debido a que la función de variación lenta no es constante.

Como se mencionó al comienzo del documento, el problema de escoger k en el gráfico de Hill es el mismo problema de escoger el umbral u para el método POT.

Lo anterior se observa no solo en la figura 5 (donde se conoce L), sino también en la figura 4 (donde *no* se conoce L) para el caso de las caídas del S&P.

Como ya se mencionó, la selección del umbral desempeña un papel importante en la convergencia de la distribución de los excesos a una GPD. La adecuada selección del umbral conlleva a estimadores confiables del parámetro de forma y de cuantiles altos (como medida de riesgo). Pero la calidad de estos estimadores depende mucho del comportamiento de segundo orden de la función de variación lenta L.

Bajo condiciones de variación regular de segundo orden, el estimador de Hill es asintóticamente normal con varianza asintótica ξ^2 para datos iid. Esta condición dice:

Para $\xi > 0, \rho \leq 0$

$$\lim_{x\to\infty} \frac{\dfrac{\overline{F}(tx)}{\overline{F}(x)} - t^{-1/\xi}}{a(x)} = t^{-1/\xi}\,\frac{t^{\rho}-1}{\rho},$$

donde $a(x) \to 0$ cuando $x \to \infty$ y $t > 0$. El parámetro ρ es denominado *el parámetro de segundo orden* y mide la velocidad de convergencia en EVT. En otras palabras, ρ gobierna la tasa de convergencia de $\overline{F}(tx)/\overline{F}(x)$ a $t^{-1/\xi}$. Asumiendo distribuciones tipo-Pareto[6], ver (2.2) en nota 6, la condición anterior se reescribe como:

$$\lim_{x\to\infty} \frac{\dfrac{L(tx)}{L(x)} - 1}{a(x)} = \frac{t^{\rho}-1}{\rho},$$

Cuando $\rho = 0$, el lado derecho de esta expresión se lee como log t, por regla de L'Hôpital. Equivalentemente, el parámetro de segundo orden gobierna la convergencia de $L(tx)/L(x)$ a 1. Esta condición será utilizada en el método EGPD (ver Condición 1 más adelante).

[6] Una distribución tipo-Pareto o de colas pesadas es una distribución F, la cual satisface:

$$1 - F(x) = x^{-1/\xi} L(x), x \to \infty, \xi > 0, \qquad (2.2)$$

donde L es una función de variación lenta para todo $x > 0$. Distribuciones tipo-Pareto también se denominan "distribuciones con cola de Pareto". Ejemplos de estas distribuciones son Pareto, gamma inversa, t-Student, loggamma, F y Burr.

Para más detalles de propiedades de segundo orden, ver la sección 3.12.1 de Bingham *et al.* (1987), que está basada en el artículo de Goldie y Smith (1987). Otras referencias útiles son: de Haan y Resnick (1998), Hall (1979), Smith (1982) y el artículo de Fraga Alves *et al.* (2007).

La condición de variación regular de segundo orden tiene dos objetivos: 1) establecer la normalidad asintótica de los estimadores de ξ y 2) estudiar las tasas de convergencia a distribuciones de valor extremo. Observe que si ρ es un valor muy grande negativo, el lado derecho de la condición tiende a cero, obteniendo convergencia de $L(tx)/L(x)$ a 1, lo que no sucede cuando $\rho = 0$. Esta observación es importante tenerla en cuenta en la sección de simulación, puesto que estimaciones de ξ se complican cuando $\rho = 0$.

Entonces, ¿cómo solucionar el problema de *trade-off* entre sesgo y varianza? Algunos autores han propuesto suavizar el gráfico de Hill. La idea es mitigar el problema de alta volatilidad observada en el gráfico de Hill (para valores relativamente pequeños de k).

Otros autores han intentado reducir el nivel de sesgo observado en el gráfico de Hill. Como ejemplos están los trabajos de Peng (1998), Feuerverger y Hall (1999), quienes proponen dos métodos para estimar el índice de cola pero que aumenta varianza comparado con métodos tradicionales. Beirlant *et al.* (1999) se basan en un modelo de regresión para los espaciados de los estadísticos ordenados superiores de una distribución tipo-Pareto y construir un estimador de ξ. Gomes *et al.* (2000) usan estimadores generalizados de *Jackknife* para reducir sesgo, y Gomes y Martins (2002) determinan el parámetro de segundo orden de manera externa para obtener un estimado de índice de cola con menor sesgo.

Otras metodologías incluyen técnicas de *bootstrap,* como Dacorogna *et al.* (1995), Danielsson y de Vries (1997), quienes se basan principalmente en el enfoque propuesto por Hall (1990). En su artículo, Hall muestra una aplicación de remuestreos de menor tamaño a la muestra original para el estimador de Hill; sin embargo, este procedimiento requiere que el parámetro de segundo orden sea conocido y el tamaño de muestra sea muy grande.

Otro método que pareciera menos sofisticado, es escoger el umbral como un porcentaje de los excesos. En su tesis doctoral, Chavez-Demoulin (1999) sugiere escoger un umbral equivalente al 10% de los datos en exceso, basado en estudios de simulación. Por ejemplo, Dutta y Perry (2006) seleccionan un umbral del 5% y 10% de los excesos cuando aplican POT a pérdidas por riesgo operativo.

La primera revisión en este artículo es el estimador de *average* Hill propuesto por Resnick y Stărică (1997), quienes

reducen la varianza asintótica del estimador de Hill, pero no sesgo. Luego se revisa el método de Zipf, seguido por una técnica de reducción de sesgo (modelo EGPD), pero que aumenta en varianza con respecto al estimador de Hill.

Entonces, ¿cuál es el mejor método para usar con el fin de estimar el parámetro de forma o el índice de cola si existe? Esta es la pregunta que se intenta resolver en todo el documento. A continuación se presenta una breve descripción de tres métodos alternativos para la estimación del índice de cola.

El estimador *average* Hill

Un método para resolver el problema de alta variabilidad en el gráfico de Hill es el propuesto por Resnick y Stărică (1997). La idea consiste en promediar los valores del estimador de Hill correspondiente a diferentes valores p de los estadísticos ordenados. Los estadísticos ordenados son los datos de una muestra aleatoria ordenados de menor a mayor.

El modelo es:

$$\hat{\xi}_{k,n,c}^{(avH)} = \frac{1}{(c-1)k} \sum_{p=k+1}^{ck} \hat{\xi}_{p,n}^{(H)}, \quad k = 1,\ldots,n-1,$$

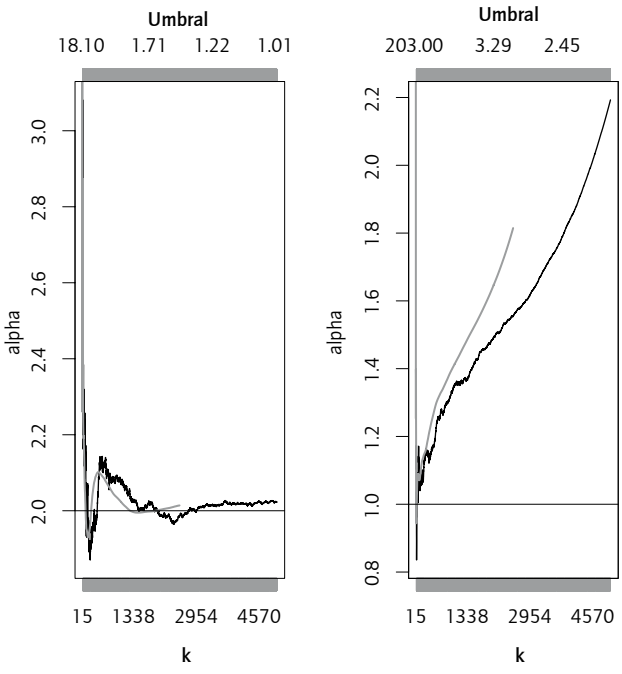
donde $c > 1$, y $\hat{\xi}_{p,n}^{(H)}$ corresponde al estimador de Hill. El estimador *average* Hill funciona bien cuando los datos representados en el gráfico de Hill exhiben alta volatilidad, pero no cuando se presenta alto sesgo, como se observa en el panel derecho de la figura 6.

La variable c puede ser interpretada como el parámetro promediador de los valores del estimador de Hill. Un resultado importante es que la varianza del estimador *average* Hill decrece con el valor de c. Adicionalmente, existe una reducción de la varianza con respecto al estimador de Hill. Por ejemplo, si $c = 2$, la varianza del estimador *average* Hill es $0{,}653\xi^2$ (puesto que la varianza está dada por $2\xi^2/c$ $[1 - (\log c)/c]$), es decir, una reducción del 34,7% de la varianza con respecto al estimador de Hill. Por consiguiente, se desearía escoger un valor de c tan grande como sea posible para reducir la varianza de este estimador. Sin embargo, al incrementar el valor de c se obtiene una reducción de datos por ser graficados, y esto dificulta la selección del valor verdadero de ξ. Como una solución a este problema, Resnick y Stărică proponen utilizar c entre $n^{0,1}$ y $n^{0,2}$, donde n es el tamaño de los datos.

El estimador de Zipf

Una herramienta gráfica muy útil para un análisis descriptivo de datos es el QQ-plot (*Quantile-Quantile plot*). Esta

FIGURA 6. Estimador de Hill (línea negra) y average Hill (línea gris) para 5000 datos iid de una distribución Pareto ($1 - F(x) = x^{-1/\xi}$) con $\alpha = 2$ (izquierda), y para $1 - F(x) = x^{-1/\xi}$ ($1 / \sqrt{\log x}$) con $\alpha = 1$ (derecha); $\alpha = 1/\xi$ y $c = 2$ para ambos casos. La línea horizontal representa el verdadero valor de α.



Fuente: elaboración propia.

herramienta es valiosa para determinar si una serie de datos de tamaño n proviene de una población con una distribución paramétrica específica. Si los cuantiles de una distribución paramétrica están linealmente relacionados con los cuantiles de los datos de una muestra en el QQ-plot, se obtiene un ajuste de los datos de la muestra a dicha distribución paramétrica. La distribución normal se ha empleado en muchas aplicaciones estadísticas, pero por lo general, los datos financieros (y de seguros) exhiben colas pesadas que no son adecuadamente modelados por una distribución normal. Para este tipo de datos es mejor utilizar el gráfico de cuantiles de Pareto (*Pareto-Quantile plot*, también denominado *Zipf plot*) que se construye utilizando los siguientes valores para los ejes $\{x,y\}$:

$$\left\{\log\left(\frac{n+1}{j}\right), \log X_{n-j+1}\right\}, \qquad j = 1,\ldots,n. \qquad (2.3)$$

Si los datos (X_i) siguen una distribución estricta de Pareto, la curva obtenida siguiendo (2.3), en el gráfico de cuantiles de Pareto se aproxima a una línea recta. Además, su pendiente es aproximadamente ξ. Si los datos siguen (2.2) la curva graficada en el gráfico de cuantiles de Pareto es aproximadamente lineal, pero para valores pequeños de j.

La pendiente de la línea resultante de aplicar mínimos cuadrados a los puntos obtenidos de (2.3) es un estima-

dor denominado el "estimador-qq" (Kratz y Resnick, 1996). Puesto que Zipf usó este estimador desde finales de la década de los años 1920, el estimador-qq también es conocido como el "estimador de Zipf". Schultze y Steinebach (1996) también propusieron el mismo estimador usando un procedimiento similar (ver Csörgő y Viharos, 1998). El modelo es:

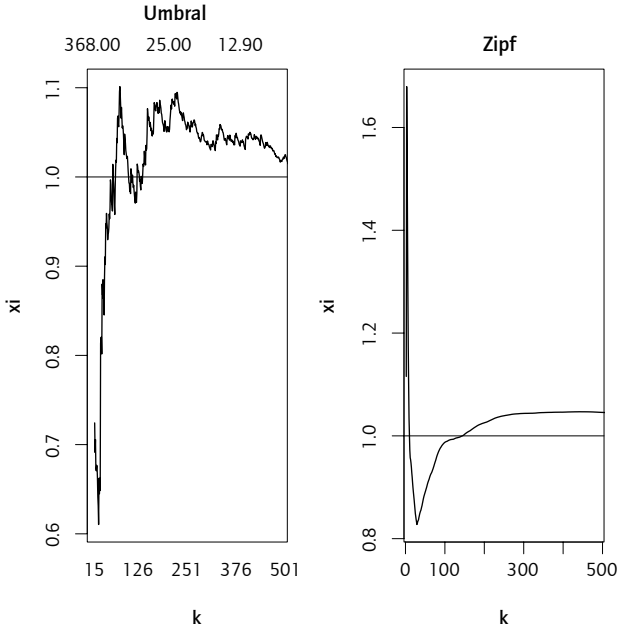
$$\hat{\xi}_{k,n}^{(Z)} = \frac{\sum_{j=1}^{k}\log\left(\frac{k+1}{j}\right)\log X_{n-j+1,n} - \frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{k+1}{j}\right)\sum_{j=1}^{k}\log X_{n-j+1,n}}{\sum_{j=1}^{k}\left(\log^2\left(\frac{k+1}{j}\right)\right) - \frac{1}{k}\left(\sum_{j=1}^{k}\log\left(\frac{k+1}{j}\right)\right)^2}, \quad k=1,\ldots,n-1.$$

El estimador de Zipf es consistente si $k \to \infty$ y $k/n \to 0$. Bajo otras condiciones de segundo orden en $F(x)$ y restricciones sobre $k(n)$, este estimador es asintóticamente normal con media asintótica ξ y varianza asintótica $2\xi^2$, (ver Kratz y Resnick, 1996).

Por consiguiente, el estimador de Hill presenta menor varianza que el estimador de Zipf. Sin embargo, una ventaja del estimador de Zipf con respecto al estimador de Hill es que los residuos de la curva obtenida en el gráfico de cuantiles de Pareto contienen información que potencialmente podría ser usada para mitigar el sesgo en los estimados cuando las colas de la distribución empírica de los datos no siguen una distribución Pareto.

La figura 7 muestra una comparación del estimador de Zipf y Hill para datos simulados de una distribución Pareto.

FIGURA 7. Estimador de Hill (izquierda) y estimador de Zipf (derecha) para 5000 datos iid de una distribución Pareto (α = 1). La línea horizontal representa el verdadero valor de ξ.



Fuente: elaboración propia.

El método de la distribución generalizada de Pareto extendida (EGPD)

La definición de una distribución tipo-Pareto puede ser reescrita como:

$$\lim_{u\to\infty}\frac{1-F(ux)}{1-F(u)} = x^{-1/\xi},$$

para cualquier valor $x > 1$, y esta puede ser interpretada como:

$$P(X/u > u \mid X > u) \approx x^{-1/\xi}, \qquad (2.4)$$

para valores grandes de u y $x > 1$.

La relación (2.4) dice que se puede aproximar la distribución de los excesos relativos condicionados en $X_i > u$ (de ahora en adelante, la distribución condicional de los excesos relativos) a una distribución estricta de Pareto.

La distribución condicional de los excesos relativos en el caso de las distribuciones tipo-Pareto satisface:

$$P(X/u > u \mid X > u) = x^{-1/\xi}\frac{L(ux)}{L(u)}, \qquad (2.5)$$

para todo $x \geq 1$ y L siendo una función de variación lenta. Puesto que $L(ux) / L(u) \to 1$ para todo $x > 0$, se obtiene (2.4) a partir de (2.5).

Condición 1: *Existe una constante real $\rho < 0$ y una función b que satisface $b(u) \to 0$, cuando $u \to \infty$, tal que para todo $x \geq 1$,*

$$\frac{\frac{L(ux)}{L(u)}-1}{b(u)} \to \frac{x^{\rho}-1}{\rho}, \text{ cuando } u\to\infty \qquad (2.6)$$

Remplazando el término $\frac{L(ux)}{L(u)}$ de (2.6) en (2.5) se obtiene:

$$P(X/u > u \mid X > u) = x^{-1/\xi}\left(1 + b(u)\frac{x^{\rho}-1}{\rho} + o(b(u))\right), \text{ cuando } u\to\infty.$$

Eliminando el término remanente, es decir $o(b(u))$, y estableciendo $\delta = b(u) / \rho$, se obtiene:

$$P(X/u > u \mid X > u) := F_{\xi,\delta,\rho} = (1-\delta)\left\{1 - x^{-1/\xi}\right\} + \delta\left\{1 - x^{-1/\xi+\rho}\right\}. \qquad (2.7)$$

El rango de los parámetros es $1/\xi\rho \leq \delta < 1$, $\xi > 0$ y $\rho < 0$. La ecuación (2.7) puede ser vista como una mezcla de dos distribuciones de Pareto donde el parámetro ponderador δ puede ser negativo (puesto que ρ es negativo).

Cuando $\delta = 0$, se obtiene el modelo estricto de Pareto para los excesos relativos. Cuando $\rho = -1$, la distribución condicional de los excesos relativos, asintóticamente, equivale al modelo de distribución de Pareto generalizada (GPD).

De esta mezcla de dos distribuciones de Pareto, se obtiene un nuevo estimador para ξ, mediante máxima verosimilitud. Esta nueva aproximación de la mezcla de distribucio-

nes se denomina el "modelo de distribución de Pareto generalizada extendida" (EGPD).

Tanto ξ como δ se estiman mediante máxima verosimilitud con una previa estimación de ρ. Ver Remark 3.2 de Beirlant *et al*. (2005) y las referencias allí contenidas para detalles de estimación externa de ρ.

Este modelo es difícilmente identificable cuando ρ es muy cercano a cero. La varianza asintótica de este nuevo estimador es igual a:

$$\left(\frac{1-\rho}{\rho}\right)^2 \frac{\xi^2}{k}.$$

La figura 8 compara los estimados de Hill y modelo EGPD para datos simulados de una distribución t con tres grados de libertad (con $\rho = -2/3$). El gráfico para el nuevo estimador luce más estable alrededor del verdadero valor de ξ que para el estimador de Hill en este caso.

La selección de k

El gráfico $\{k, \hat{\xi}_{k,n}\}$ es una herramienta adecuada para escoger k y así decidir el estimado de ξ. Sin embargo, la selección óptima de k para elegir el verdadero valor de ξ no es una tarea fácil, como se vio anteriormente.

Existen varios métodos adaptativos para escoger k; ver por ejemplo la sección 4.7 de Beirlant *et al*. (2004) y las referencias allí contenidas. Un método puede ser minimizar el *error cuadrático medio asintótico* (AMSE, por sus siglas en inglés) del estimador de Hill, que está dado por:

$$AMSE\left(\hat{\xi}_{k,n}^{(H)}\right) = A\,\mathrm{var}\left(\hat{\xi}_{k,n}^{(H)}\right) + ABias\left(\hat{\xi}_{k,n}^{(H)}\right) = \frac{\xi^2}{k} + \left(\frac{b_{n,k}}{1-\rho}\right)^2.$$

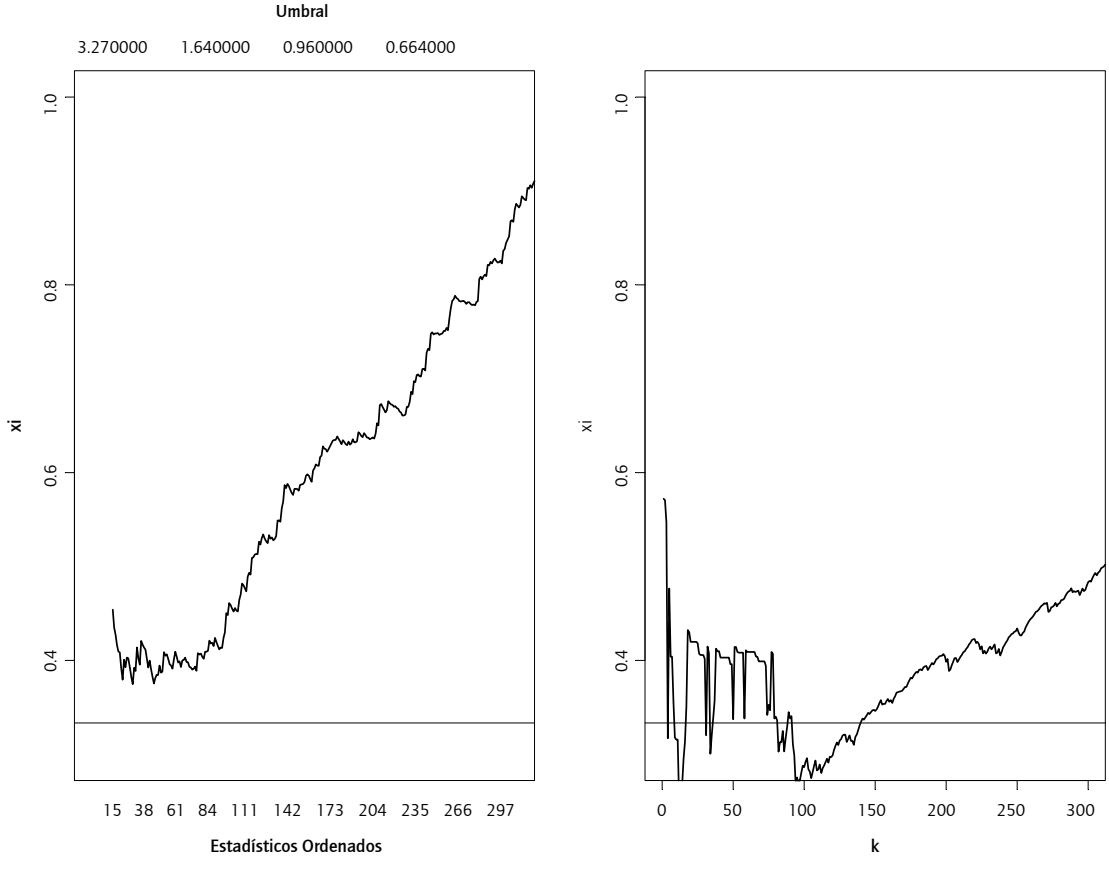
La idea es estimar ξ mediante el estimador de Hill (ver por ejemplo la sección 4 de Beirlant *et al*., 1999), donde:

$$\hat{k} = \underset{k;k>10}{\arg\min}\left(\frac{\hat{\xi}_k^2}{k} + \left(\frac{\hat{b}_{n,k}}{1-\hat{\rho}_k}\right)^2\right),$$

$\hat{\xi}$, $\hat{b}$ y $\hat{\rho}$ denotan los estimadores de máxima verosimilitud de ξ, b y ρ respectivamente, posiblemente obtenidos mediante EGPD o por otro método alternativo.

En la sección 4 de este documento se aplicarán los métodos para estimar el índice de cola al caso de los datos de

FIGURA 8. Estimador de Hill (izquierda) y modelo EGPD (derecha) para 1000 datos iid de una distribución *t*-Student con $\nu = 3$. La línea horizontal representa el verdadero valor de ξ.



fuego en Dinamarca, siguiendo la idea de Reiss y Thomas (1997) para escoger k gráficamente. Para un número k intermedio, existe un balance entre varianza y sesgo del estimador; entonces un *plateau* (o meseta) se vuelve visible en el gráfico de Hill. De esta manera, se busca una región estable en los gráficos, cuando se aplican los diferentes métodos para estimar el índice de cola, y con este valor decidir cuál es el valor de ξ.

En la siguiente sección se realiza una simulación para probar los métodos de estimación de ξ a funciones de distribución comúnmente usadas en administración de riesgos.

Simulación

En esta sección se compara el desempeño de los métodos anteriormente revisados, mediante simulación Monte Carlo. Los estimadores usados en la simulación son:

- El estimador *average* Hill
- El estimador de Zipf
- El método EGPD.

Se escogen cuatro tipos de datos con distribuciones donde $\xi > 0$ y dos tipos de datos con distribuciones donde $\xi = 0$ (ver tabla 1).

También se pueden clasificar las distribuciones de acuerdo con el parámetro de segundo orden ρ (parámetro que mide la velocidad de convergencia en EVT):

- "Casos complicados" donde $\rho \in (-1, 0)$; (distribución t-Student con $\nu = 3, 4, 8$),
- "Casos *muy* complicados" donde $\rho = 0$; (distribuciones lognormal, Weibull y loggamma),
- "Casos normales" donde $\rho < -1$; (distribuciones Cauchy estándar y Pareto).

Esta clasificación corresponde a los casos en que el estimador de Hill puede funcionar bien o no, y se desea saber cuál es el comportamiento de los tres estimadores.

TABLA 1. Distribuciones usadas en la simulación.

Distribuciones	Parámetros	ξ	ρ
t - Student	$\nu = 3, 4, 8$	1/3, 1/4, 1/8	−2/3, −1/2, −1/4
Lognormal	$(\mu, \sigma) = (0,1)$	0	0
Weibull	$\tau = 0{,}5$	0	0
Loggamma	$(\alpha, \beta) = (1, 2); (1,10)$	1; 1	0
Cauchy estándar		1	− 2
Pareto	$\alpha = 1/\xi$	1.2, 1.1, 1.0, 0.7, 0.5, 0.3	$-\infty$

Fuente: elaboración propia.

Para cada una de estas distribuciones, se generaron 1000 muestras de tamaño 1000. El desempeño de cada estimador[7] se evalúa en términos de la raíz del error cuadrático medio relativo (RRMSE, por sus siglas en inglés) basado en los k estadísticos ordenados superiores para la mayoría de las distribuciones. Para las distribuciones lognormal y Weibull se usa el error cuadrático medio (MSE, por sus siglas en inglés).

$$RRMSE\left(\hat{\xi}\right) = \frac{\sqrt{MSE\left(\hat{\xi}\right)}}{\xi},$$

donde $MSE\,(\hat{\xi}) = E(\hat{\xi} - \xi)^2$. El verdadero valor de ξ se obtiene de la tercera columna de la tabla 1, mientras que el valor esperado del estimador se obtiene de calcular la media de los estimados que arroje cada método en la simulación Monte Carlo para cada tipo de distribución.

La distribución t − Student

La distribución t − Student ha sido empleada en el estudio de administración de riesgos, en especial para ajustar datos empíricos de retornos de títulos de renta variable para el cálculo de riesgo de mercado, puesto que la forma de la distribución t es similar a la distribución normal estándar. Recuerde que la distribución normal estándar fue originalmente usada para estimación de VaR (valor en riesgo), por su facilidad de cálculo y tratamiento. Aunque la distribución t es simétrica, esta tiene colas más amplias que la normal, lo que permite ajustar mejor datos extremos que la normal.

Para este caso el estimador *average* Hill tiende a ser, en general, preferido para valores relativamente pequeños de k. Cuando se incrementa el número de grados de libertad, el método *average* Hill es mejor comparado con EGPD, pero el rango de estadísticos ordenados para su optimalidad decrece ($k \leq 51$ para $\nu = 3$, y $k \leq 32$ para $\nu = 8$). Sin embargo, el método EGPD tiene el mínimo RRMSE para $\nu = 3, 4$ (9,8% en $k = 160$ para $\nu = 3$, y 16,5% en $k = 102$ para $\nu = 4$). El estimador *average* Hill tiene el mínimo RRMSE (33,2%) para $\nu = 8$ (pero $k = 7$). El estimador de Zipf no es muy útil comparado con los otros dos métodos.

[7] Los siguientes resultados se obtuvieron utilizando los códigos en S-PLUS elaborados por J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, D. de Waal, C. Ferro y B. Vandewalle, y se pueden encontrar en ucs.kuleuven.be/Wiley/index.html (Chapter 4). Los archivos que se utilizaron fueron: Zipf.SSC, Hill2oPV.SSC para los métodos Zipf y EGDP, respectivamente. La selección de los valores de los parámetros y algunas características de las funciones de distribución están principalmente basadas en McNeil y Saladin (1997), quienes realizaron un estudio similar aplicando el método POT.

TABLA 2.

t, $\nu = 3$ ($\xi = 1/3$, $\rho = -2/3$)	
Método	**RRMSE mín.**
EGPD	0,098
Average Hill	0,167
Zipf	0,303

t, $\nu = 4$ ($\xi = 1/4$, $\rho = -1/2$)	
Método	**RRMSE mín.**
EGPD	0,165
Average Hill	0,206
Zipf	0,430

t, $\nu = 8$ ($\xi = 1/8$, $\rho = -1/4$)	
Método	**RRMSE mín.**
Average Hill	0,332
EGPD	0,492
Zipf	0,952

Fuente: elaboración propia.

La distribución lognormal

La distribución lognormal ha sido utilizada para modelar datos de severidad en el sector de seguros o en riesgo operativo. La distribución lognormal también se ha utilizado como supuesto para modelar el precio de títulos de renta variable. Por ejemplo, el modelo de Black-Scholes, usado para modelar el precio de opciones, supone que el precio del activo financiero subyacente sigue una distribución lognormal, es decir, que los retornos logarítmicos de este activo se distribuyen normalmente (una variable aleatoria X se distribuye lognormal si su logaritmo natural, $Y = \log X$ se distribuye normal).

El método *average* Hill no se desempeña tan bien, y de acuerdo con las simulaciones, este es el método con mayor MSE. Para todos los valores de k, el modelo EGPD es el que mejor se desempeña.

TABLA 3.

Lognormal ($\xi = 0$, $\rho = 0$)	
Método	**MSE mín.**
EGPD	0,057
Zipf	0,161
Average Hill	2,286

Fuente: elaboración propia.

La distribución Weibull

En una distribución Weibull, cuando el índice de cola es igual a 1, esta distribución se reduce a la exponencial, y cuando es igual a 2 se aproxima a la distribución Rayleigh. Cuando el índice de cola es igual a 3,5, la distribución Weibull se aproxima a una normal. Es una *stretched-exponential* cuando el índice de cola es menor que uno y decae más lentamente que una exponencial. Por tal razón se estudia el caso donde su parámetro de forma, $\tau < 1$.

La distribución Weibull tiene también un ξ igual a cero. De nuevo, el método de *average* Hill es el que tiene mayor MSE. El método Zipf se desempeña mejor que el método EGPD para valores de $k \leq 49$.

TABLA 4. Weibull ($\xi = 0$, $\rho = 0$)

Método	**MSE mín.**
EGPD	0,165
Zipf	0,178
Average Hill	0,788

Fuente: elaboración propia.

La distribución loggamma

La distribución loggamma también es comúnmente usada para modelar datos de pérdidas o severidades en riesgo operativo y seguros.

Para ambos casos de la distribución loggamma ($\beta = 2$ y 10), el mejor método es Zipf para valores relativamente pequeños de k; sin embargo, el desempeño del estimador por el método EGPD es muy similar al del estimador de Zipf. Cuando se incrementa el valor de β, en general, los métodos presentan mayores valores de RRMSE, lo que muestra un peor desempeño de los métodos a mayor contaminación por la función de variación lenta de la distribución loggamma. Nuevamente, el estimador *average* Hill tiene los valores más altos de RRMSE, independientemente del valor de β, y por tal razón no se incluye en los siguientes resultados.

TABLA 5.

LG, $\alpha = 1$, $\beta = 2$ ($\xi = 1$, $\rho = 0$)	
Método	**RRMSE mín.**
EGPD	0,248
Zipf	0,747
Average Hill	0,956

LG, $\alpha = 1$, $\beta = 10$ ($\xi = 1$, $\rho = 0$)	
Método	**RRMSE mín.**
EGPD	0,242
Zipf	0,879
Average Hill	1,286

Fuente: elaboración propia.

La distribución Cauchy estándar

La distribución Cauchy estándar es una distribución t–Student con un grado de libertad. El estimador *average* Hill es preferido para valores pequeños de k ($k \leq 151$). La técnica EGPD se desempeña mejor para valores relativamente altos de k y es el método que menor RRMSE presenta.

TABLA 6.

Cauchy estándar ($\xi = 1$, $\rho = -2$)	
Método	**RRMSE mín.**
EGPD	0,041
Average Hill	0,083
Zipf	0,120

Fuente: elaboración propia.

La distribución Pareto

La distribución Pareto es la más comúnmente usada para simular distribuciones de pérdidas que presentan colas pesadas. En este caso, $E[X^j] = \infty$ para $\xi \geq 1 / j$. Por tal razón,

se prueban los métodos de estimación de ξ para varios valores de parámetro α ($\xi = 1 / \alpha$), de la distribución Pareto. Para valores de $\xi \geq 0{,}5$, no es posible calcular momentos de segundo orden ni superiores. Para $\xi \geq 1$ (1,0, 1,1 y 1,2 en este estudio), no es posible calcular ningún momento (incluso la media no es finita). En este último caso no sería posible calcular el *Expected Shortfall* para las distribuciones de pérdidas con estas características, denominado como los *modelos de media-infinita* (ver Nešlehová *et al*., 2006, y las referencias allí contenidas para una discusión más detallada de los problemas y consecuencias de estimación de altos cuantiles con este tipo de modelos). Para $\xi \geq 1/3$, no es posible calcular el tercer momento ni superiores, pero sí la media y el segundo momento. Para ello, se tiene un caso donde $\xi = 0{,}3$, en el cual sí se puede calcular el tercer momento, pero no el cuarto.

El estimador *average* Hill es el mejor método para valores relativamente pequeños de k para los diferentes valores de ξ estudiados. Sin embargo, el estimador *average* Hill no es el método con menor RRMSE. Para valores pequeños de ξ, el rango del número de estadísticos ordenados superiores disminuye con respecto a la preferencia del método *average* Hill comparado con los otros dos estimadores, como se muestra a continuación:

$\xi =$	0,3	0,5	0,7	1,0	1,1	1,2
$k \leq$	27	57	94	144	168	180

El estimador de Zipf también se desempeña bien para todos los casos. Cuando $\xi = 0{,}3$, 0,5 y 1,1, el método de Zipf tiene el menor RRMSE (0,085, 0,082 y 0,085, respectivamente). El método EGPD tiene el menor RRMSE cuando ξ = 0,7, 1,0 y 1,2 (0,052, 0,036 y 0,030, respectivamente).

Comentarios

Si el criterio de selección de un estimador fuese el mínimo valor de RRMSE (o MSE dependiendo del caso), para estimar el índice de cola, en general, se recomienda el método EGPD en conjunto con el estimador *average* Hill si los datos se distribuyen t – Student (casos donde $\xi < 1$). Para el caso Pareto se recomienda usar el estimador de Zipf en conjunto con el método EGPD. El método EGPD tiene el mínimo RRMSE para datos que se distribuyen Cauchy estándar y loggamma (donde $\xi = 1$). Los métodos anteriormente revisados están diseñados para desempeñarse bien en casos cuando $\xi > 0$. Sin embargo, para los casos en que los datos se distribuyen lognormal y Weibull (donde $\xi = 0$), el método EGDP presenta el menor RRMSE.

Recuerde que la distribución subyacente de los datos es desconocida en la práctica, pero de acuerdo con la simulación, el método EGPD es la técnica que mejor se desempeña cuando $-1 < \rho \leq 0$. En otras palabras, se corrobora la teoría de que este método es una buena técnica de reducción de sesgo, presentado en el gráfico de Hill, y podría ser usado en el área de seguros y riesgo operativo, cuando se asume que las severidades siguen una distribución loggamma, lognormal o Weibull con fines de simulación y estimación de cuantiles altos (como VaR al 99% o 99,9%). Otra manera para saber qué método utilizar en datos de pérdida históricos es realizar una prueba de bondad de ajuste y observar cuál distribución paramétrica es la que más se aproxima a los datos empíricos. De esta manera, se selecciona el estimador que mejor se desempeña, según las observaciones realizadas en este documento, y están basadas en los resultados obtenidos de la simulación.

TABLA 7.

Pareto, ($\xi = 0{,}3$, $\rho = -\infty$)			**Pareto, ($\xi = 0{,}5$, $\rho = -\infty$)**			**Pareto, ($\xi = 0{,}7$, $\rho = -\infty$)**	
Método	**RRMSE mín.**		**Método**	**RRMSE mín.**		**Método**	**RRMSE mín.**
Zipf	0,085		Zipf	0,082		EGPD	0,052
Average Hill	0,237		EGPD	0,087		Zipf	0,086
EGPD	0,243		*Average* Hill	0,160		*Average* Hill	0,126

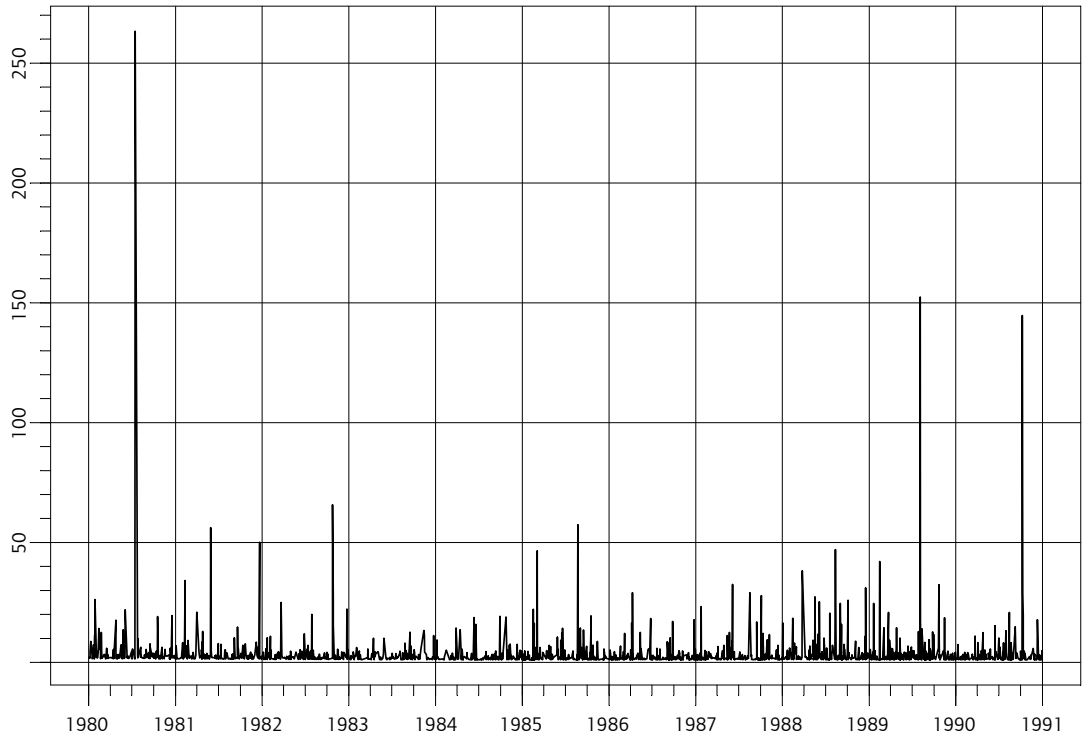
Fuente: elaboración propia.

TABLA 8.

Pareto, ($\xi = 1{,}0$, $\rho = -\infty$)			**Pareto, ($\xi = 1{,}1$, $\rho = -\infty$)**			**Pareto, ($\xi = 1{,}2$, $\rho = -\infty$)**	
Método	**RRMSE mín.**		**Método**	**RRMSE mín.**		**Método**	**RRMSE mín.**
EGPD	0,036		Zipf	0,085		EGPD	0,030
Zipf	0,081		*Average* Hill	0,087		*Average* Hill	0,081
Average Hill	0,097		EGPD	0,140		Zipf	0,084

Fuente: elaboración propia.

FIGURA 9. Datos de fuego en Dinamarca.



Fuente: elaboración propia.

Caso: datos de fuego en Dinamarca

Se aplican los métodos revisados a un caso particular. Los datos corresponden a 2167 reclamos en seguro contra el fuego en Dinamarca (*Danish Fire data*), donde las pérdidas se expresan en millones de coronas danesas (DKM), desde marzo 1 de 1980 hasta diciembre 12 de 1990. Se inicia con un análisis exploratorio de los datos[8], y la figura 9 muestra la serie de tiempo.

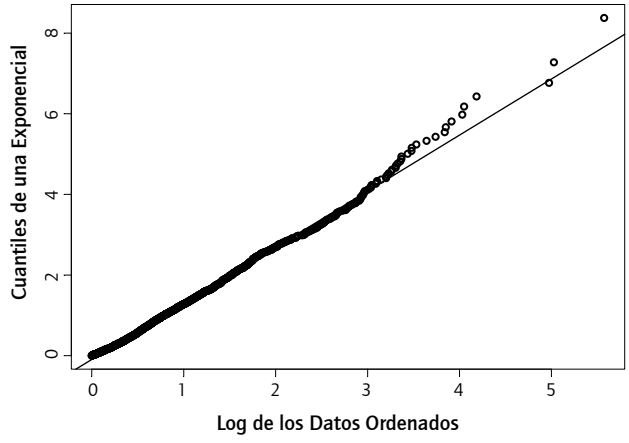
A continuación se muestra el resumen del análisis. Se observa que la distribución de las pérdidas es sesgada a la derecha y presenta alta curtosis (buen ejemplo para estudio de colas pesadas).

	Pérdida en DKM
Mín	1,00
Primer cuartil	1,32
Mediana	1,77
Media	3,38
Sesgo	18,76
Curtosis	483,76
Tercer cuartil	2,96
Máx	263,25
$\hat{Q}(0.99)$	26,04

El cuantil al 99% de los datos empíricos es igual a 26,04 DKM. Más adelante se calculan cuantiles a este nivel con los diferentes métodos estudiados, para compararlo con este dato.

La figura 10 muestra el gráfico de cuantiles de Pareto.

FIGURA 10. Gráfico de cuantiles de Pareto para los datos de fuego en Dinamarca.



Fuente: elaboración propia.

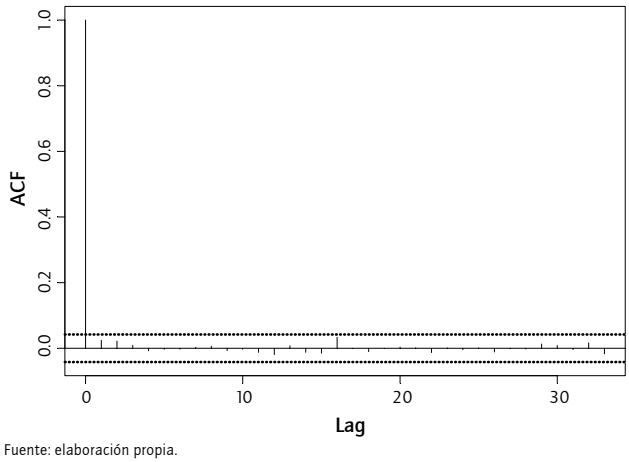
El gráfico de cuantiles de Pareto dice que los datos exhiben colas pesadas. Se tendería a pensar que la distribución subyacente es muy cercana a una distribución Pareto, debido a que la curva mostrada en el gráfico de cuantiles de

[8] Los datos están disponibles en www.ma.hw.ac.uk/~mcneil/ y han sido previamente estudiados por varios autores bajo el marco de EVT.

Pareto es muy cercana a una recta en su parte derecha. La pendiente de la línea ajustada en el gráfico de cuantiles de Pareto es una buena aproximación de ξ cuando se asume que los datos siguen una distribución Pareto (como se había mencionado anteriormente). Para este caso, la pendiente es igual a 0,722, como primera aproximación de ξ para los datos de fuego en Dinamarca. De todas maneras, no es seguro que la distribución subyacente de los datos siga una distribución Pareto. A continuación se observa el gráfico de función de autocorrelación (ACF, por sus siglas en inglés).

Como se observa en la figura 11, los datos no exhiben correlación serial. Esto sucede a menudo con datos del sector seguros (y pérdidas por riesgo operativo) y se puede presumir que son observaciones iid; sin embargo, esto no es muy común con los datos financieros. Ver Resnick (1997) para pruebas adicionales de independencia aplicados a estos datos. La figura 12 muestra los gráficos de los métodos aplicados a los datos de fuego en Dinamarca para estimar ξ.

FIGURA 11. ACF de los datos por analizar.

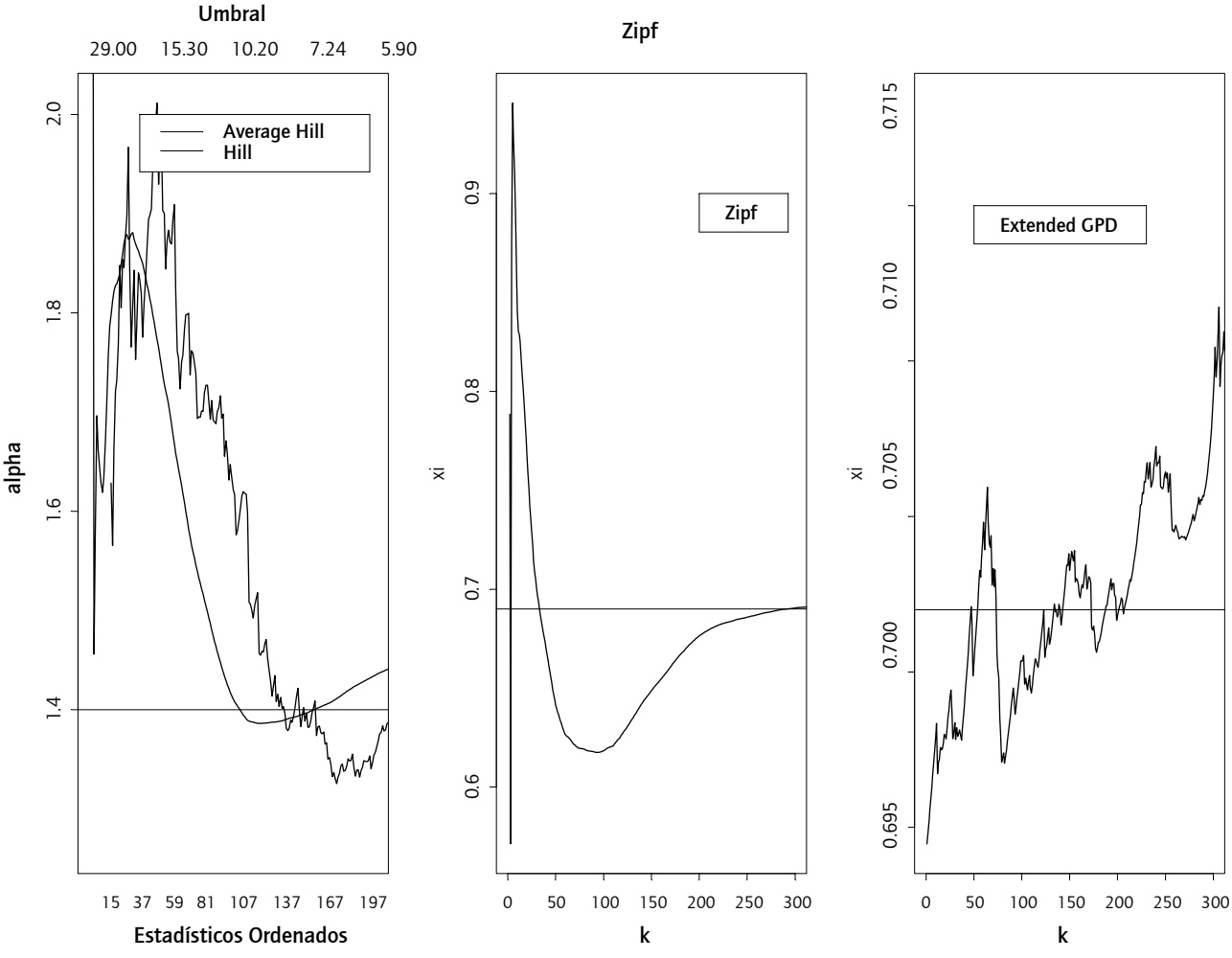


Fuente: elaboración propia.

La estabilidad en los gráficos para los métodos *average* Hill, Zipf y EGPD dan los siguientes resultados:

$\hat{\alpha}^{(avH)}$ = 1,385, $\hat{\xi}^{(avH)}_{120,2167}$ = 0,722 (k = 120),

FIGURA 12. Los métodos aplicados al caso analizado.



Fuente: elaboración propia.

$\hat{\xi}^{(Z)}_{290,2167}$ = 0,690, (k =290),
$\hat{\xi}^{(EGPD)}_{290,2167}$ = 0,702 (k =72).

También se estima un cuantil alto $Q(1-p)$ para los datos analizados, donde p es un número muy pequeño entre 0 y 1. Para ello, se usa el estimador de Weissman en la mayoría de los casos (excepto para el método EGPD).

- *Estimador de Weissman.* Está dado por la siguiente ecuación:

$$\hat{Q}(1-p) \quad = X_{n-k,n}\left(\frac{k+1}{(n+1)p}\right)^{\hat{\xi}^{(H)}_{k,n}}, \quad k=1,\ldots,n-1.$$

- *Estimador de cuantil para el método EGPD.* Se calcula mediante la siguiente ecuación:

$$\hat{p}_{k,x} = \frac{k+1}{n+1}\overline{G}_P\left(x/X_{n-k},\hat{\xi},\hat{\delta},\hat{\rho}\right),$$

donde $\overline{G}_P$ denota la función de supervivencia de la distribución Pareto. Fijando $\hat{p}_{k,x}$ en un valor muy pequeño (por ejemplo 0,01 para estimar $\hat{Q}(0{,}99)$), se puede resolver numéricamente para x, y así obtener un estimador de cuantil extremo.

- *Estimador de cuantil*[9] *para el método POT.* Se utiliza la siguiente relación:

$$Q(1-p) = u + \frac{\beta}{\xi}\left(\left(\frac{p}{\overline{F}(u)}\right)^{-\xi} - 1\right)$$

Se requiere un estimado de $\overline{F}(u)$, entonces se propone usar el estimador Nu/n (ver McNeil *et al.,* 2005 y las referencias allí contenidas). ξ y β son los parámetros de forma y escala de la GPD.

Se usa la siguiente información (que corresponde a $X_{n-k,n}$) para los métodos *average* Hill, Zipf y EGPD, respectivamente, para estimar Q (0,99) o $\text{VaR}_{99\%}$ para los datos de fuego en Dinamarca:

- $X_{2047,2167}$ = 8,72
- $X_{1887,2167}$ = 4,61
- $X_{2095,\,2167}$ = 13,35, $\hat{\delta}$ = −0,15, $\hat{\rho}$ = −4,6.

De esta manera se calcula el cuantil para cada uno de los métodos:

- $\hat{Q}^{(avH)}(0{,}99) = 30{,}19$
- $\hat{Q}^{(Z)}(0{,}99) = 15{,}09$
- $\hat{Q}^{(EGPD)}(0{,}99) = 24{,}90$

Comentarios

Los métodos analizados en este documento arrojan un ξ entre 0,690 y 0,722. Mediante el método POT, McNeil (1997) obtuvo un estimado de ξ igual a 0,684 (con error estándar de 0,27) para un umbral de 20 DKM y k = 36. Csörgő y Viharos (1998) también estimaron ξ mediante varios métodos para los datos de fuego en Dinamarca y obtuvieron los siguientes resultados: 0,716, 0,713, 0,719, 0,717 y 0,719 (con M = 0,000323).

Los resultados obtenidos en este documento para la estimación de ξ son similares a los alcanzados por Csörgő y Viharos y también a los de McNeil cuando usa un umbral de 20 DKM. Sin embargo, cuando el umbral es 10 DKM, McNeil obtiene un estimado de ξ igual a 0,497 (con error estándar de 0,14).

Al usar la información del artículo de McNeil (1997), se estima Q (0,99) usando el método POT. Para un umbral de 20 DKM es 25,8 DKM; y 27,3 DKM, para un umbral de 10 DKM. Recuerde que al principio de esta sección se estimó el cuantil empírico de 26,04 DKM, y el método analizado en este documento que más se aproxima es el EGDP, con un valor de 24,9 DKM. Degen *et al.* (2007) utilizaron la aproximación de las pérdidas en exceso a una distribución g-h (en lugar de una distribución GPD como lo hace el método POT) y estimaron Q (0,99) para los datos de fuego en Dinamarca en 27,32 DKM. Para más detalles de esta novedosa aproximación, ver el artículo de Degen *et al*., y las referencias allí contenidas.

Conclusiones y futura investigación

Los administradores de riesgo no solamente están interesados en estimar el capital regulatorio sino también el capital económico. Las pérdidas esperadas deberían ser cubiertas con capital regulatorio, mientras que el capital económico cubriría pérdidas inesperadas por eventos extremos. Un estimado confiable del parámetro del índice de cola conlleva a estimados confiables de cuantiles altos (como medida de riesgo) y por ende, estimados razonables de cargos de capital.

Los administradores de riesgo deben tener en cuenta que los eventos extremos pueden ocurrir y causar grandes pérdidas. El problema es que los valores extremos escondidos en una cola de una distribución no son fáciles de detectar. Una herramienta para solucionar este problema es EVT.

[9] Para estimar los cuantiles por el método POT se utiliza el complemento QRMlib (*Quantitative Risk Management library*) para S-PLUS, que se puede encontrar en www.ma.hw.ac.uk/~mcneil/book/QRMlib.html con un registro previo.

Se sugiere entonces a los reguladores y profesionales en riesgo comenzar a probar modelos EVT para cuantificar riesgo operativo y en el sector de seguros. Para series de tiempo financieras es frecuente ver aplicaciones de medidas condicionales de riesgo (de mercado), puesto que un hecho estilizado de los retornos de activos financieros es que la varianza de los retornos muestra autocorrelación positiva. McNeil y Frey (2000) presentan una aplicación de EVT condicional (específicamente el uso de POT a las innovaciones de un modelo GARCH, las cuales se asumen ser iid) para estimar medidas condicionales de riesgo, y mediante pruebas de *backtesting* muestran la superioridad de esta metodología.

Sin embargo, si EVT es la herramienta elegida para tratar los valores extremos del negocio y estimar medidas de riesgo, el área de administración de riesgos debe considerar el problema de la estimación del "parámetro de forma o índice de cola", un problema aún no resuelto.

La técnica más comúnmente usada para estimar el parámetro de forma, o el índice de cola, es el estimador de Hill. Aunque el estimador de Hill es un estimador consistente para ξ para datos iid (bajo ciertas condiciones también para datos no iid), este método presenta algunos problemas:

1) Escoger el valor de k del gráfico de Hill no es una tarea fácil.

2) El gráfico de Hill puede mostrar alguna volatilidad, y si la distribución subyacente de los datos es muy diferente a una Pareto, el gráfico puede exhibir sesgo.

3) El estimador de Hill no es invariante en localización.

En el punto 2) se revisó el estimador *average* Hill como una posible solución para la alta volatilidad. Resnick y Stărică (1997) proponen el uso de este método como una base para técnicas de *boostrap* para corregir el problema de sesgo. El estimador de Zipf puede ser visto también como una técnica de suavización del gráfico de Hill para estimar el índice de cola. Sin embargo, su varianza asintótica es el doble de la varianza del estimador de Hill. No obstante, una ventaja de este método es que los residuos del gráfico de cuantiles de Pareto contienen información que podría ser usada para mitigar el sesgo en los estimados cuando la cola de la distribución no es Pareto. Los estimadores de Hill y de Zipf pertenecen a una clase más grande de kernel, el cual podría ser interesante para posteriores estudios. Respecto al sesgo, se probó el método EGPD, el cual reduce sesgo comparado con el estimador de Hill.

Infortunadamente, no se puede responder a la pregunta de "cuál es el mejor método por usar, si existe, para calcular el parámetro de forma o el índice de cola". La selección *óptima* del umbral (o selección de k) no está resuelta aún, como se mencionó anteriormente. Esta selección óptima puede ser obtenida sólo bajo algunas propiedades precisas de segundo orden en la función de variación lenta L, la cual no se puede inferir de los datos.

Sin embargo, se puede brindar una guía cuando se trabaja con datos si se conocen sus funciones de distribución. Por ejemplo, en el sector de seguros generales es común usar las distribuciones loggamma, lognormal y Weibull para modelar las distribuciones de pérdidas. Se sugiere usar la técnica de EGPD (técnica de reducción de sesgo) en estos casos para estimar ξ. Como era de esperarse, este es el mejor método cuando se presenta alto sesgo en la estimación del índice de cola (casos en que $\rho = 0$). En general, cuando no se conoce con certeza la distribución de los datos, se recomienda usar el método EGDP en conjunto con el estimador de Zipf. A partir de los resultados de la simulación de acuerdo con el menor RRMSE (o MSE según el caso), el mejor método es el EGPD en la mayoría de distribuciones de colas pesadas y semipesadas. Cuando la cola de las distribuciones no es tan pesada, en la mayoría de los casos el mejor método es el de Zipf.

Al aplicar los métodos analizados en este documento al caso de datos de fuego en Dinamarca, se obtienen estimadores del índice de cola muy similares a estudios previos. El rango de la estimación del índice de cola está entre 0,690 y 0,722 en este estudio. Csörgő y Viharos (1998) encuentran un rango entre 0,713 y 0,719, mientras que McNeil (1997) encuentra valores de 0,497 y 0,684, dependiendo de la selección del umbral.

Existe también la posibilidad de analizar los métodos para estimar el índice de cola cuando $\xi \in \mathbb{R}$. De igual manera, hay métodos robustos para estimación del índice de cola que se podrían probar en simulaciones como las mostradas en este documento. Finalmente, un método novedoso en EVT es la aproximación de las pérdidas en exceso a una distribución g-h más que a una GPD para calcular valores de cuantiles altos como medida de riesgo.

Agradecimientos

El autor agradece las valiosas sugerencias y recomendaciones de los evaluadores anónimos que ayudaron significativamente a mejorar la versión previa de este artículo.

Referencias bibliográficas

Balkema, G. & Embrechts, P. (2007). *High Risk Scenarios and Extremes. A Geometric Approach.* Zurich Lectures in Advanced Mathematics. Zürich, Switzerland: European Mathematical Society Publishing House.

Beirlant, J., Dierckx, G., Goegebeur, Y. & Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes,* 2, 177-200.

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

Beirlant, J., Joossens, E. & Segers, J. (2005). Unbiased tail estimation by an extension of the generalized Pareto distribution. *CentER Discussion Paper* 2005-112.

Bingham, N., Goldie, C. & Teugels, J. (1987). *Regular Variation.* Cambridge: Cambridge University Press.

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

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer-Verlag,

Csörgő, S. & Viharos, L. (1998). Estimating the tail index. In B. Szyszkowicz (Ed.). *Asymptotic Methods in Probability and Statistics* (pp. 833-881). Amsterdam: North-Holland.

Dacorogna, M. M., Müller, U. A., Pictet, O. V. & de Vries, C. G. (1995). *The distribution of extremal foreign exchange rate returns in extremely large data sets*. Zürich: O&A Preprint.

Danielsson, J. & de Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance, 4,* 241-257.

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

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

Drees, H., de Haan, L. & Resnick, S. I. (2000). How to make a Hill plot. *Annals of Statistics,* 28, 254-274.

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

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlín: Springer.

Falk, M., Hüssler, J. & Reiss, R.-D. (2004). *Laws of small numbers: extremes and rare events* (2nd ed.). Basel: Birkhäuser.

Feuerverger, A. & Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics, 27,* 760-781.

Fraga Alves, M. I., Gomes, M. I., Haan, L. de & Neves, M. (2007). A Note on Second Order Conditions in Extreme Value Theory: Linking General and Heavy Tails Conditions. Notas e Comunicações CEAUL. *Revstat, 5*(3), 285-305.

Gilli, M. & Kellezi, E. (2003). *An application of extreme value theory for measuring risk*. Geneva: Dept. of Economics and FAME, University of Geneva, 23.

Goldie, C. M. & Smith, R. L. (1987). Slow variation with remainder? A survey of the theory and its applications. *Quart. J. Math. Oxford*, *38*(2), 45-71.

Gomes, M. I. & Martins, M. J. (2002). "Asymptotically unbiased" estimators of the tail index. *Extremes,* 5, 1-31.

Gomes, M. I., Martins, M. J. & Neves, M. (2000). Alternatives to a semiparametric estimator of parameters of rare events—the Jackknife methodology. *Extremes, 3*, 207-229.

Gnedenko, B. V. (1943). Sur la distribution limité du terme d'une série aléatoire. *Ann. Math.*, *44*, 423-453.

Gumbel, E. J. (1958). *Statistics of Extremes*. New York: Columbia University Press,

Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl Probab.*, *16,* 433-439.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameters in nonparametric problems. *Journal of Multivariate Analysis, 32,* 177-203.

Haan, L. de & Resnick, S. I. (1998). On asymptotic normality of the Hill estimator. *Commun. Statist. Stochastic Models, 14,* 849-866.

Haan, L. de & Ferreira, A. (2006). *Extreme value theory. An introduction*. Springer Series in Operational Research and Financial Engineering. New York: Springer-Verlag.

Kratz, M. & Resnick, S. I. (1996). The qq-estimator and heavy tails. *Stochastic Models*, 12, 699-724.

Leadbetter, M. R. (1991). On a basis for "peaks over thresholds". *Statistics and Probability Letters, 12,* 357-362.

Malevergne, Y. & Sornette, D. (2006). *Extreme financial risks*. Berlín: Springer-Verlag.

Matthys, G. & Beirlant, J. (2000). Adaptative threshold selection in tail index estimation. In P. Embrechts (Ed.). *Extremes and Integrated Risk Management* (pp. 37-48). London: Risk Books.

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin,* 27, 117-137.

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

McNeil, A. J. & Saladin, T. (1997). The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions. In Proceedings of XXVIIth International *ASTIN Colloquium*, pp. 23-43, Cairns, Australia.

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

Moix, P.-Y. (2001). *The measurement of market risk*. Lecture Notes in Economics and Mathematical Systems 504. Berlin: Springer-Verlag.

Nešlehová, J., Chavez-Demoulin, V. & Embrechts, P. (2006). Infinite-mean models and the LDA for operational risk. *Journal of Operational Risk,* pp. 3-25.

Peng, L. (1998). Asymptotically unbiased estimators for extreme value index. *Statistics & Probability Letters, 38,* 107-115.

Reiss, R.-D. & Thomas, M. (1997). *Statistical Analysis of Extreme Values*. Basel: Brikhäuser.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* New York: Springer.

Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *Astin Bulletin,* 27, 139-151.

Resnick, S. I. & Stărică, C. (1997). Smoothing the Hill estimator. *Advances in Applied Probability,* 29, 271-293.

Schultze, J. & Steinebach, J. (1996). On least squares estimates of an exponential tail coefficient. *Statist and Decisions*, *14*, 353-372.

Smith, R. L. (1982). Uniform rates of convergence in extreme-value theory. *Adv. Appl. Probab.*, *14*, 600-622.

Smith, R. L. (1987). Estimating Tails of Probability Distributions. *The Annals of Statistics, 15,* 1174-1207.

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