Journal of Banking & Finance 54 (2015) 60–71

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

A parametric alternative to the Hill estimator for heavy-tailed distributions

Joseph H.T. Kim [a], Joocheol Kim [b,*]

[a] Department of Applied Statistics, Yonsei University, South Korea
[b] School of Economics, Yonsei University, South Korea

ARTICLE INFO

Article history:
Received 2 July 2014
Accepted 23 December 2014
Available online 10 January 2015

JEL classification:
C13
C60
F31

Keywords:
Hill estimator
Tail index
Extreme value theory
Generalized Pareto distribution
Scaled Log phase-type distribution

ABSTRACT

Despite its wide use, the Hill estimator and its plot remain to be difficult to use in Extreme Value Theory (EVT) due to substantial sampling variations in extreme sample quantiles. In this paper, we propose a new plot we call the eigenvalue plot which can be seen as a generalization of the Hill plot. The theory behind the plot is based on a heavy-tailed parametric distribution class called the scaled Log phase-type (LogPH) distributions, a generalization of the ordinary LogPH distribution class which was previously used to model insurance claims data. We show that its tail property and moment condition are well aligned with EVT. Based on our findings, we construct the eigenvalue plot from fitting a shifted PH distribution to the excess log data with a minimal phase size. Through various numerical examples we illustrate and compare our method against the Hill plot.

1. Introduction

The Hill (1975) estimator has been widely used in modeling heavy tail phenomenon arising in many disciplines including finance and insurance. Despite its popularity, the Hill plot still remains to be difficult to use because it is not straightforward from the plot to select the tail threshold where the power tail starts, and the estimation of the tail index is highly sensitive to the choice the threshold. It could also induce the bias driven by the unknown slowly varying part of the tail. Thus, as Stărică (1997) put, practitioners find no region in the plot where the Hill estimate is sufficiently stable that reliable estimates can be achieved, and are left to the mercy of inspiration when estimating the tail index. Among extensions and adaptations of the Hill estimator are the smoothed Hill estimator of Resnick and Stărică (1997), the bootstrap version of the Hill estimator of Hall (1990), and an alternative Hill plot using the log scale for the horizontal axis in Drees et al. (2000). Most of these refinements seek to reduce the variability of the Hill plot on the upper side of the data, but there seems to be no clear cut superior estimator, and the need of further research is evident. Other alternative tail index estimators also available including the estimator by Pickands (1975) and the Dekkers–Einmahl–de Haan estimator by Dekkers et al. (1989). However, the performance for these alternative estimators again crucially depend on the interplay between the tail index and the second order parameter, and the Hill estimator seems to be most popular in the literature due to its simplicity.

Recently the Log phase-type (LogPH) distribution class has been proposed to model heavy-tailed datasets. This class is obtained by treating the underlying variable to be such that its (natural) logarithm follows a phase-type (PH) distribution of Neuts (1975), which models the absorption time in a suitably chosen finite state Markov chain. In Ahn et al. (2012), it is shown that the LogPH class has a power tail and asymptotically has a linear mean excess function resembling the popular generalized Pareto distribution (GPD) with positive shape parameter, a well-aligned result to Extreme Value Theory. Their numerical examples show the LogPH class provides a satisfactory fit, along with the estimated tail index, for the heavy-tailed datasets. However, this approach may suffer from some issues which could deteriorate the quality of the tail index estimation. First, by using the whole data, as opposed to the tail part only, the fit is inevitably affected by non-extreme losses which take up the most data count. Second, depending on the shape of the loss data, one may end up with an excessive phase size. Indeed, as

* Corresponding author.
E-mail addresses: jhtkim@yonsei.ac.kr (J.H.T. Kim), joocheol@yonsei.ac.kr (J. Kim).

http://dx.doi.org/10.1016/j.jbankfin.2014.12.020
0378-4266/

illustrated in Ahn et al. (2012), the required phase size can get quite big, leading to impractically many parameters, distorting accurate tail fitting. Controlling the phase size is a generic difficulty of the PH calibration and an area for further research (e.g., Bladt, 2005). Finally, the LogPH class is defined only on $[1,\infty)$, unable to capture any positive losses below 1. An ad hoc remedy such as rescaling or shifting of the loss may be used as suggested in Ahn et al. (2012), but this may distort the estimated parameters.

In the present paper, we propose an alternative method to estimate the tail index of heavy-tailed distribution, inspired by Ahn et al. (2012). In particular, we extend the LogPH distributions class by adding a scale parameter, and use it to estimate the tail index of heavy-tailed loss data. At the same time, we use the excess loss approach with the phase size kept minimal at two, which balances between the flexibility of the model and the stability of the extreme tail fit. It turns out that the Hill estimator can be seen as a special case of our method, corresponding to phase size one. With this approach, we construct a plot we call the eigenvalue plot, a generalised version of the Hill plot. Through various simulated and actual examples we show that the proposed method is more stable and easier to use compared to the Hill estimator.

The paper is organized as follows. In Section 2, after a brief review on the Hill estimator, we introduce the shifted PH and scaled LogPH distributions, a generalization of the ordinary PH and LogPH classes, respectively, and develop theories related to tail index. Equipped with the theories developed, we explain how to use the eigenvalue plot to find the tail index in Section 3. Section 4 presents various numerical examples to illustrate the proposed plot and compare to the Hill counterpart. The examples include simulated heavy-tailed data and actual finance and insurance datasets.

2. Tail index estimation

Let us denote the tail (or survival) function of a distribution by $\overline{F}(y) = 1 - F(y), 0 < y < \infty$, where $F(y)$ is the cumulative distribution function. Then we say that $\overline{F}(y)$ is regularly varying with index $-\alpha < 0$, and write $\overline{F} \in \mathcal{R}_{-\alpha}$, if

$$\lim_{y\to\infty} \frac{\overline{F}(y\lambda)}{\overline{F}(y)} = \lambda^{-\alpha}, \qquad \lambda > 0. \tag{1}$$

When $\alpha = 0$ the tail is called slowly varying, or $\overline{F} \in \mathcal{R}_0$. Using this we can represent a regularly varying distribution as $\overline{F}(y) \sim L(y)y^{-\alpha}$ where $L(\cdot) \in \mathcal{R}_0$. We note that $f(y) \sim g(y)$ means $\lim_{y\to\infty} f(y)/g(y) = 1$. Thus the tail of regularly varying functions can be represented by power functions multiplied by slowly varying functions. In many applications it is of interest to accurately estimate the tail index α from the given dataset; we refer the reader to, e.g., Embrechts et al. (1997), Beirlant et al. (2006), Reiss and Thomas (2007) and Clauset et al. (2009) for further details, where the last reference discusses how to discern and quantify power-law behavior in empirical data.

2.1. The Hill estimator

The method of Hill (1975) is one of the most studied methods to estimate α in the literature. When $F(y)$ is a regularly varying distribution with $\alpha > 0$, the Hill method is designed to find the tail index αusing a sample $Y_1, \ldots, Y_n$ from $F(y)$. If we consider the order statistics $Y_{n,n} \leqslant \cdots \leqslant Y_{1,n}$ of the loss sample, the Hill estimator is defined as

$$\hat{\alpha}_{k,n} = \left(\frac{1}{k} \sum_{j=1}^{k} \log Y_{j,n} - \log Y_{k,n} \right)^{-1}, \quad 2 \leqslant k \leqslant n. \tag{2}$$

After computing $\hat{\alpha}_{k,n}$ for different k values, one can draw the Hill plot $\{(k, \hat{\alpha}_{k,n}^{-1})\}, k = 2, 3, \ldots, n$ on the plane, where $\hat{\alpha}_{k,n}^{-1}$ represents the upper bound of the moment existence. The goal is then to find a stable area in the plot over different but relatively small k values. An alternative expression of (2) shows that the Hill estimator is a minor modification of the reciprocal of the empirical mean excess function of the log-transformed data. That is, by denoting $X_{j,n} = \log Y_{j,n}$, we may rewrite the Hill estimator as

$$\begin{aligned} \hat{\alpha}_{k,n} &= \left(\frac{\sum_{j=1}^{k} (X_{j,n} - d_k) I\{X_{j,n} > d_k\}}{\sum_{j=1}^{n} I\{X_{j,n} > d_k\}} \cdot \frac{k-1}{k} \right)^{-1} \\ &= \left(\hat{e}_X(d_k) \cdot \left(1 - \frac{1}{k}\right) \right)^{-1}, \quad d_k = X_{k,n}, \end{aligned} \tag{3}$$

where $\hat{e}_X(d)$ is the sample version of the mean excess function defined as $e_X(d) = E(X - d | X > d)$. It is known that (e.g., Drees et al., 2000) the Hill plot works most effectively when the underlying loss distribution is Pareto (or very close to Pareto) with

$$F(y) = 1 - \left(\frac{y}{\sigma}\right)^{-\alpha}, \quad y > \sigma > 0. \tag{4}$$

The Hill estimator has indeed minimum mean squared error when the second order parameter governing the rate of convergence of (1) is zero which corresponds to the Pareto case in (4). As the log variable of Pareto is exponentially distributed, we see that the Hill estimator (3) is an approximate MLE for the exponential distribution, which is the sample mean. We also note that the shape parameter α in Pareto becomes the scale parameter for the log loss variable. However, once the underlying distribution departs from Pareto, its performance gets poorer and it becomes less clear what portion of the plot is most accurate. From the log-transformed data viewpoint, this means that the exponential distribution is inadequate to accommodate the shape of the log loss in a unstable manner.

2.2. Generalizing the Hill estimator

One way to improve the tail index estimation, thus generalize the Hill estimator, is to allow for other parametric distributions beyond the exponential distribution, so that various possible shapes of the log data can be accommodated easily even when its shape is somewhat different from the exponential distribution. However, because the range of the candidate models can be too broad without further restrictions, we would like to limit our choices to a distribution class that generalizes the exponential distribution, while maintaining an exponentially decaying tail.

One such a class is the phase-type (PH) distribution class introduced by Neuts (1975), Neuts (1981), which is the distribution of the time until absorption in a continuous time Markov chain with an absorbing state; see, e.g., Latouche and Ramaswami (1999). The distribution function of the PH distribution with parameter (β, T) is given by

$$F_X(x) = 1 - \beta e^{Tx}\mathbf{1}, \quad x > 0, \tag{5}$$

where β and T are matrices of size $p \times 1$ and $p \times p$, respectively, where p is commonly called the phase size. To ensure (5) is a proper distribution function, we require the sum of all elements of β to be one and the real part of each eigenvalue of T to be strictly negative. We denote $-\eta_T < 0$ to be the eigenvalue of T closest to 0. As a generalization of the exponential distribution, the PH distribution class has received attention in the applied literature of, e.g., queues, insurance and reliability due to their useful properties such as closure properties, computational tractability and denseness, where the last property means that any non-negative distribution on $[0,\infty)$ can be approximated by a PH distribution to any desired accuracy. As special cases, the PH distribution includes the exponential distribution (when $p = 1$), its mixtures, and convolutions as well as the Erlang distributions, which are the gamma

distributions with integer shape parameter. See Appendix I for further technical details of the PH distribution.

A useful property relevant to later discussions is that, when $X \sim \mathrm{PH}(\beta, T)$ as defined in (5), the excess of loss $X - d | X > d$ is again PH distributed with parameter (β_d, T) where

$$\beta_d = \frac{\beta e^{dT}}{\beta e^{dT}\mathbf{1}} \tag{6}$$

for any constant $d > 0$, reminiscent of the property of the exponential and the GPD distributions. Furthermore, from (5), we see that $1 - F_X(x) = O(e^{-\eta_T x})$ as $x \to \infty$, from the Jordan decomposition of T, showing that the PH distribution class has an exponentially decaying tail.

In light of these properties, we may fit the log data to the PH distribution to estimate the tail index. However, there are two issues to be resolved. First, even though the class of PH distributions is dense, modeling heavy tails may require an excessive number of states in the Markov chain and thereby a large number of parameters, leading to impractical models. This problem is addressed in Section 3. Second, when we log-transform the original data, all the losses in $(0, 1)$ will result in negative values which are outside the nonnegative support of the PH distribution class. In the next subsection we resolve this issue by extending the support of the PH distribution class with an additional location parameter.

2.3. Scaled LogPH distribution

When X has a PH distribution with (β, T), we may add a location parameter c so that $X - c$ can be shifted PH distributed, denoted by $X - c \sim \mathrm{PH}(\beta, T; -c)$. By shifting the PH distribution, the data points that result in negative values after log-transformation can be handled naturally. Let us now investigate the distribution of the exponentiated variable of $X - c$, which describes the original heavy-tailed loss, and show that it has a power tail with η_T corresponding to the tail index of the regularly varying distribution. When exponentiated, the shifted PH variable yields a new random variable $Y = \exp(X - c) = e^{-c}e^X$, of which the distribution will be termed the scaled LogPH distribution, or simply $Y \sim \mathrm{LogPH}(\beta, T; -c)$. Although c can take any real value, we are primarily interested in $c > 0$ as the log-transformation of positive loss results in a support starting from a negative value. When $c = 0$, it reduces to the standard LogPH class, which was introduced in Ghosh et al. (2011) and further studied by Ahn et al. (2012). The main advantage of $\mathrm{LogPH}(\beta, T; -c)$ with $c > 0$ is that it can handle, by choosing a suitable c, loss data lying in $(0, 1)$, which is impossible under the standard LogPH class of which the support is $[1, \infty)$.

The distributional properties of the scaled LogPH distribution class are similar to those of the standard LogPH class in Ahn et al. (2012), but not identical. We start with the distribution function of Y, which is given by, using (5),

$$\begin{aligned} F_Y(y) &= P(e^{X-c} \leqslant y) = P(X \leqslant \log y + c) \\ &= 1 - \beta e^{T(\log y + c)}\mathbf{1}, \qquad \text{for } y \geqslant e^{-c} \end{aligned} \tag{7}$$

and

$$f_Y(y) = \frac{1}{y}\beta e^{T(\log y + c)}t, \ y \geqslant 1, t = -T\mathbf{1}. \tag{8}$$

To determine the kth moment of Y, recall $Y = e^{X-c}$ with $X \sim \mathrm{PH}(\beta, T)$. Then using (8), we apply the change of variable technique to obtain

$$\begin{aligned} E[(Ye^c)^k] = E[e^{kX}] &= \int_1^\infty y^k \frac{1}{y}\beta e^{T\log y} t dy = \beta\left(\int_1^\infty y^{k-1}e^{T\log y}dy\right)t \\ &= \beta\left(\int_0^\infty e^{(k-1)z}e^{Tz}e^z dz\right)t = \beta\left(\int_0^\infty e^{(kI+T)z}dz\right)t \\ &= -\beta(k\mathbf{I} + T)^{-1}t, \end{aligned} \tag{9}$$

with $\mathbf{I}$ being the identity matrix, which in turn yields

$$E[Y^k] = e^{-ck}\beta(k\mathbf{I} + T)^{-1}T\mathbf{1}. \tag{10}$$

The finiteness and nonnegativity of the equality (9) depends on the eigenvalues of the matrix inside the integrand. More specifically, the standard matrix theory states that (9) is assured when all the eigenvalues of $kI + T$ are negative. We recall that the real parts of all the eigenvalues of T have negative real parts, and that $-\eta_T < 0$ is the eigenvalue of T closest to 0. Thus the kth moment of Y in (10) exists only for $k < \eta_T$, indicating a heavy tail with finite moments.

The tail behavior of the scaled LogPH distribution can be more specific. Referring to Ahn et al. (2012), the tail function of the standard LogPH distribution decays with rate $(\log y)^k y^{-\eta_T}$ for some $k, 0 \leqslant k \leqslant m_T - 1$ where $-\eta_T < 0$ is the eigenvalue of T closest to 0 and m_T is its multiplicity, implying that the standard LogPH has a power law tail. As the scaled LogPH is a scaled version of the ordinary LogPH, it too has a power law tail and, for $Y \sim \mathrm{LogPH}(\beta, T; -c)$, we have

$$\lim_{y\to\infty} \frac{\overline{F}_Y(y\lambda)}{\overline{F}_Y(y)} = \lambda^{-\eta_T}, \qquad \lambda > 0, \tag{11}$$

implying a regularly varying distribution with tail index $\eta_T > 0$.

3. Finding the tail index

In order to find the tail index, we fit the log-transformed data to the shifted PH distribution with parameter set $(\beta, T; -c)$. The estimated η_T is then the tail index estimate $\hat{\alpha}$. However, to ensure the quality the estimated tail index in real applications, we need additional considerations which we elaborate in this section.

3.1. Method

We extend the property (6) that, for $X \sim \mathrm{PH}(\beta, T; 0)$, the excess of loss $X - d | X > d$ is again central PH distributed with parameter $(\beta_d, T; 0)$, to the shifted PH random variable $X - c$. That is, for $X - c$, its excess loss of is shown to be again PH distributed through

$$\begin{aligned} (X - c) - d | (X - c) > d = X - (c + d) | X > (c + d) \\ \sim \mathrm{PH}(\beta_{c+d}, T; 0). \end{aligned} \tag{12}$$

We point out two important aspects of this result at this stage. First, this result says that, for the original variable $Y \sim \mathrm{LogPH}(\beta, T; -c)$, its log excess loss $\log Y - d | \log Y > d$ is $\mathrm{PH}(\beta_{c+d}, T; 0)$ distributed with location parameter 0. This means that we may use a standard PH fitting algorithm, which will be introduced shortly, to fit the log excess loss as the location parameter always vanishes. Second, in doing so, matrix T, hence η_T as well, is invariant to both the excess threshold d and the initial location parameter c. This implies that, for any given threshold, the excess variable of the PH distribution would produce the same T and η_T, which resembles the well-known GPD property that the tail index remains the same over different thresholds. Thus we may repeatedly fit the PH distribution for the log excess data over different d values, and expect that the resulting fits would give similar $\widehat{T}$ and thus $\hat{\eta}_T$ regardless of d, whenever the fit is adequate. Our strategy is then to find the region where $\hat{\eta}_T$ is deemed stable, like in the Hill plot. A natural plot for this is to draw $(n - n_d, \hat{\eta}_T^{-1}(d))$ on the plane over varying d values, where n_d is the number of observations greater than d and $\hat{\eta}_T(d)$ stands for $\hat{\eta}_T$ estimated at threshold d. We call this the eigenvalue plot as $-\eta_T$is the eigenvalue of T closest to 0. Note that the horizontal axis $n - n_d$ represents the number of deletions.

Fitting a PH distribution, however, generally requires the choice of the phase size p, which is an important but challenging decision

because the needed p can get quite large to warrant adequate fit depending on the dataset. Thus, if we want to create a generic statistical tool to estimate the tail index, it is impractical to consider all possible phase sizes for a given data. Because our goal is to obtain the tail index implied from the tail fit, rather than producing a perfect fit to explain the whole data, we instead exploit the fact that T is invariant to the excess loss threshold. We observe that, while the densities of heavy-tailed loss take various shapes, the shape differences are present only in the below-extreme data range; as the loss gets larger, the shape differences are no longer distinguishable in that they all convex decrease by necessity. So, for d large, the log excess loss $X - d|X > d$ must be in a similar shape regardless of the true distribution's density shape. This observation indicates that a minimal number for the phase size, such as two, may suffice for the tail index determination. The result is that, for small d's, the fit is likely to be poor and the estimated η_T will be unreliable as a PH with a small phase size cannot adequately capture differing shapes of the data. However, as d gets larger, the fit will increasingly improve as the excess loss shape becomes convex decreasing for any loss data unbounded to the right, and may be adequately modeled with a PH distribution. This approach is practically advantageous as it allows us to avoid the problem of choosing the optimal phase size for actual datasets with different shapes without losing much information in finding the tail index. In Fig. 1 several PH densities with two phases, denoted $\mathrm{PH}_2(\beta, T)$, are presented to illustrate its ample shape flexibility with different parameters. In fact, Dickson and Hipp (2000) showed that the $\mathrm{PH}_2(\beta, T)$ is either a weighted average of two exponential densities, or a weighted average of an exponential and an Erlang density with the same scale parameter. Therefore, if a full weight is given to the exponential density component, $\mathrm{PH}_2(\beta, T)$ reduces to the exponential distribution, which yields the Hill estimator. In this sense, the Hill plot can be seen as a special case of the eigenvalue plot, corresponding to $p = 1$.

3.2. Eigenvalue plot

The eigenvalue plot is similar to the Hill plot because both use the same excess values of the log data over varying threshold d values; the Hill plot essentially uses the reciprocal of the sample mean and the eigenvalue plot shows $\hat{\eta}_T^{-1}(d)$. However, the two plots behave differently for most cases. Here we put forward some features of the eigenvalue plot along with an additional tool to further ensure the quality of the tail index determination.

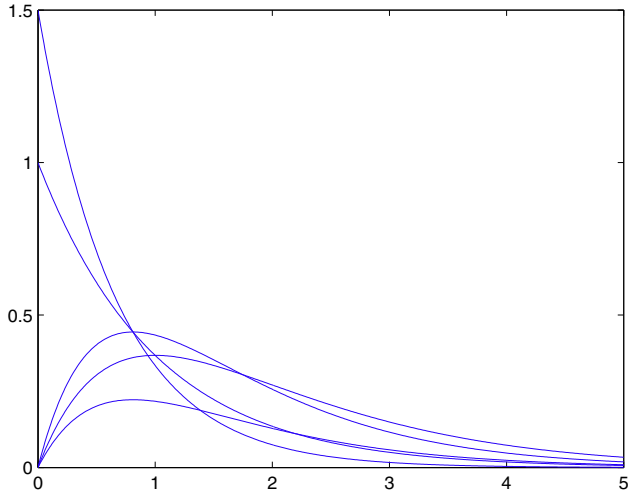

Fig. 1. Densities of PH distribution with two phases.

First, while in the Hill plot one looks into the deep tail area where $\hat{\alpha}_{k,n}$ is stable, the eigenvalue plot typically produces a stable area at a much smaller d, because it fits with a flexible parametric class containing more than one distribution with two phases. We note however that the threshold producing the correct tail index in the eigenvalue plot is not the starting point of the GPD realm; it is rather a threshold where the LogPH fit is adequate, and thus the correct tail index is reasonably assured. This is the fundamental difference between the two plots. While the Hill estimator, relying on the non-robust sample mean of extreme quantiles, aims at finding the GPD realm in the upper tail, our method fits a flexible parametric distribution whose tail asymptotic follows the GPD, with much more data.

Typically the eigenvalue plot $(n - n_d, \hat{\eta}_T^{-1}(d))$ can be roughly split into three consecutive segments based on the plot behavior. In the first and the last segments the plot is relatively unstable due to different reasons and in a different manner; the tail index is identified in the second segment where $\hat{\eta}_T(d)$ is relatively stable. In the first segment where the thresholds are small (or n_d's are large), it is unstable because the fit for the excess loss is often inadequate with a small number of phases, as previously mentioned. In this segment $\hat{\eta}_T(d)$ tends to be a monotone function with virtually no fluctuation. Followed is the second segment where $\hat{\eta}_T(d)$ shows a relatively stable behavior where the tail index is identified. If the plot takes different values in the second segment we advise to look into the right end region of the interval to find the tail index as one should get closer to the tail for accurate tail modeling. The stability however starts to break as the threshold enters the third segment because there will be fewer observations to be fitted to the PH distribution. In this segment $\hat{\eta}_T(d)$ tends to rapidly fluctuate because of high volatility of extreme quantiles, the same reason as the Hill plot does. We suggest to ignore this segment because the result is unreliable under small sample size.

Second feature of the eigenvalue plot is that, as we employ a parametric approach, we are able to utilize some statistical theories to evaluate the goodness of fit of the PH distribution so that the tail index estimation can be carried out more precisely. Among such measures, we consider the Kolmogorov–Smirnov (K–S) test, of which the test statistic for a sample of size n is

$$K(n) = \max |F_n(x) - F(x)| \tag{13}$$

where $F_n(x)$ and $F(x)$ are the empirical and the model distribution, respectively. Common critical values for the K–S test are $1.22/\sqrt{n}$ at a level of confidence 95% when $n \geqslant 40$, though the proper test decision would require a relatively large sample size; see, e.g., Klugman et al. (2008) for further details. In order to include this procedure in the eigenvalue plot, we slightly modify (13) as follows. For a given d, we get the log excess loss and obtain its empirical distribution and the fitted PH_2, respectively, and then compute $K(n_d)\sqrt{n_d}$ to see if the fit violates the critical barrier 1.22. Using the barrier, the plot $(n - n_d, K(n_d)\sqrt{n_d})$ hints the starting point of the threshold where the PH fit becomes adequate, and also tells us where the extreme, thus unreliable, tail starts in the upper tail by showing volatile values. In addition, it allows us to compare the relative goodness of fit over different thresholds. As a general rule, one prefers the value with a small K–S statistic, but the tail index estimate corresponding to the smallest K–S statistic does not always guarantee the best estimate due to sampling variation as will be seen in numerical examples; we suggest looking at the neighborhood around the smallest K–S statistic and ensure its plot is stable like the eigenvalue plot.

3.3. Implementation

Suppose that $y_1, \ldots, y_n$ is a non-negative sample from a heavy-tailed distribution. To construct the eigenvalue plot, we first take

log transformation as $x_i = \log y_i$, $i = 1, \ldots, n$. For the log data $x_1, \ldots, x_n$, we apply threshold $d = x_{n,n}$, the smallest observation, to get excess loss $x - d | x > d$ and fit a PH_2 distribution to this excess loss. Among several methods to fit a PH distribution to the given data, we use the one proposed by Asmussen et al. (1996), coined the EMPHT algorithm. The EMPHT algorithm uses the EM (Expectation–Maximization) procedure, an iterative method for MLE, based on the fact that one knows the observations, but not their entire Markov chain paths. Further details on this algorithm is presented in Appendix II. The fitting procedure is repeated for $d = x_{n-1,n}, x_{n-2,n}, \ldots, x_{3,n}$ to create pairs $(n - n_d, \hat{\eta}_T^{-1}(d))$, which form the eigenvalue plot. The algorithm stops at $x_{3,n}$ because we need at least 3 data points for PH_2 fitting. For comparison reasons, the same horizontal axis $n - n_d$ is used for the Hill plot in the numerical section. With larger phase sizes, separate experiments show that the PH fit for the early thresholds improves, but the tail fluctuation occurs early as well, because there are more parameters to fit; overall, we found no obvious advantage in tail index determination by increasing the phase size.

The process above is applicable to datasets where no modification has been made. The identical procedure can be used for left-truncated datasets provided that the support of the distribution starts from zero, which is sometimes observed in insurance datasets, as seen in the Danish fire data in Section 4.3. If however the true population has a support starting from an unknown but strictly positive number, we suggest to shift the original loss data by the smallest value before using the procedure described above in order to approximately match the support of the data with the support of the scaled LogPH. One such example is a GPD with a positive location parameter. For a sample $y_1, \ldots, y_n$ from such a distribution one should use $\log(y_1 - \min\{y_i\}), \ldots, \log(y_n - \min\{y_i\})$ for fitting. In practice, if the data starts from a strictly positive number, the modeler should be able to tell whether it is due to truncation or strictly positive support.

4. Numerical examples

In this section we present numerical illustrations to compare the performance of the eigenvalue plot and the Hill plot. We carry out the comparison for both simulated datasets from heavy-tailed distributions and the real datasets from finance and insurance. There are variants of the Hill estimator in the literature, though there seems no dominant alternative and they all suffer from the same difficulty of high quantile variations by focusing on the upper tail only. Thus we consider the standard Hill estimator for illustration.

Each figure consists of three subplots numbered from (a) to (c). The Hill plot and the eigenvalue plot are given in (a) and (b), respectively, along with the K–S guideline plot (that is, $K(n_d)\sqrt{n_d}$ for each d) is presented in (c). All plots share the common horizontal scale for an easy comparison. The horizontal axis stands for the number of deletions due to the threshold, i.e., $n - n_d$. For instance, if $n = 2000$, the value 1,800 in the horizontal axis means that there are 200 exceedances at that point.

4.1. Simulated data from GPD and GEV

The GPD and GEV (generalized extreme value) distributions play the central role in EVT and the Hill estimator is essentially designed to determine the tail index in the GPD realm of the data. To compare the performance of the Hill and eigenvalue plots, we consider several GPD and GEV distributions with different choices for parameters ξ, σ and μ. For each GPD and GEV chosen, we simulate samples of size 2000 and present two selected paths. We have tried larger sample sizes and more repetitions, but the conclusions are similar.

As the Hill and eigenvalue plots look for stable regions in different places, we have shaded reasonable regions where the true index value is revealed in each plot to elucidate this difference; in the eigenvalue plot we have shaded the area where the K–S statistic is less than 1 excluding the area where it is volatile in the extreme upper tail, and for the Hill plot we have shaded the upper 10% line for all figures. One should find stable regions within the shaded area in each plot.

Fig. 2–4 compare the two plots from the GPD with different parameters. We see that, depending the value of ξ, the Hill plot has a trend of increasing (Fig. 2) or decreasing (Fig. 3 and 4) over the whole dataset, confirming a common practice that the Hill plot is not to be used for all possible thresholds. Indeed practical experience (e.g., Section 7.2 in McNeil et al., 2005) suggests the 5% upper tail to be used to determine the tail index with the Hill plot. When we focus on this region in each Figure, the Hill plot is sometimes able to reveal the true tail index. However finding the tail index is not always easy as all Hill paths in these three figures hardly show any stable region; this is particularly true in Fig. 4 where the Hill plots deteriorate to an unacceptable range at an extreme speed.

If we turn to the eigenvalue plot, the task is relatively easier. First the plot in each of the three figures has a much wider area where the estimated value remains stable, close to the true value, indicated by a wide shaded area. Also the plot approaches the true value much faster than the Hill counterpart; in all three cases, we see a stable pattern from roughly 800 deletions where the K–S statistic is below 1. However the smallest K–S statistic does not always guarantee the optimal tail index as shown, e.g., in the thinner path of Fig. 2 where the tail index is approximately 0.6 when the K–S statistic achieves its minimum after 1,700 deletions.

Secondly, we observe that, whenever the Hill plot works well, the eigenvalue plot shows a considerably better performance, as shown for the thicker paths in all three figures. As aforementioned, we suggest to use the values around the right end of the shaded areas as higher thresholds are more consistent with tail modeling in general. For the thinner paths in the three figures, the cases where the Hill plots perform poorly, the eigenvalue plots also suffer but their deterioration is clearly less substantial. For example, the thinner path in Fig. 2 seems to indicate a tail index in $[0.5, 0.7]$ based on the Hill plot, but $[0.5, 0.6]$ from the eigenvalue plot. Lastly, as the threshold (or number of deletions) gets close to the extreme upper tail both the eigenvalue and the K–S statistics fluctuate in all figures, and we suggest not to use the information in this region.

For the GEV in Fig. 5, one can draw a similar conclusion. The comparison between the thicker and thinner paths reconfirms the advantage of the eigenvalue plot, with the latter having an acceptable stable range from deletions 400, a substantially early starting point compared to the former.

4.2. Foreign exchange rate data

For financial application, we consider the foreign exchange rate dataset studied recently by Ibragimov et al. (2013), where they show that the exchange rates of emerging and developing countries are typically more heavy-tailed than those of developed countries by applying the log–log rank size regressions with optimal shifts in ranks, as well as the Hill estimator. In determining the tail index using these two methods, they choose two truncation levels, i.e., 5% and 10% of the upper tail, and provide two estimates for the

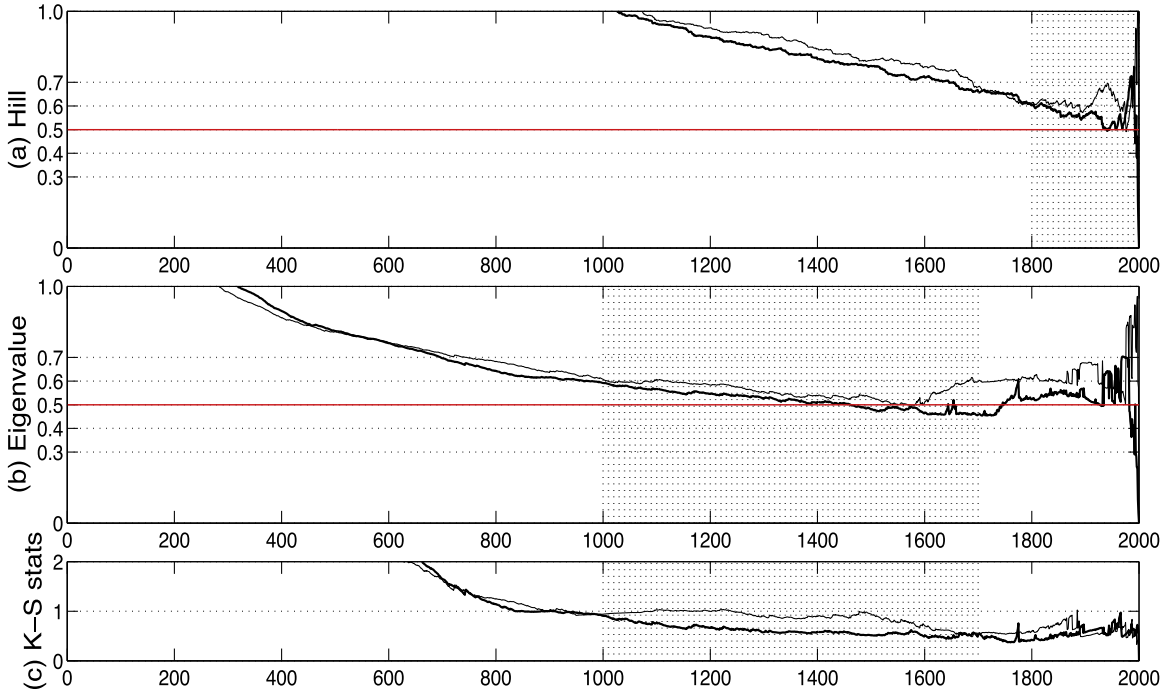

Fig. 2. Comparison of $\hat{\alpha}^{-1}$ for GPD with $\mu = 0, \sigma = 1, \xi = 0.5$.

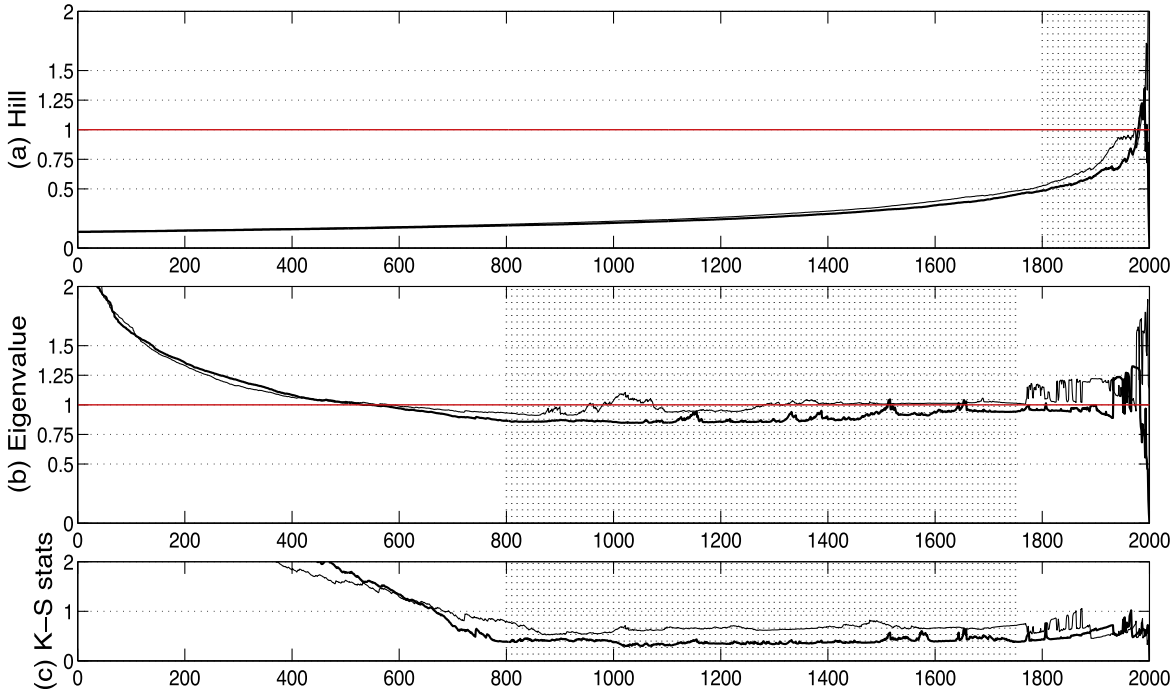

Fig. 3. Comparison of $\hat{\alpha}^{-1}$ for GPD with $\mu = 5, \sigma = 0.2, \xi = 1$.

tail index to account for the uncertainty of the tail threshold location. The choices 5% and 10% are routine but necessarily arbitrary.

We use the identical dataset and attempt to find the tail index using the eigenvalue plot, which in most cases reveals only one tail index estimate without data truncation, unlike the other two methods. Most of our estimates are shown to be consistent with their estimates, with a few exceptions. Our analysis here is not to challenge the results of Ibragimov et al. (2013), but to confirm that our methodology can serve as a good alternative to the Hill estimator. The dataset we analyze consists of daily currency exchange change rates of 17 countries from January 4, 1999 to June 22, 2012 in log returns.[1] According to the classification of Ibragimov et al. (2013), the developed country currencies are: Australian dollar (AUD), Canadian dollar (CAD), Swiss franc (CHF), Danish krone (DKK), Euro (EUR), pound sterling (GBP), Japanese yen (JPY), Norwegian kroner (NOK) and Swedish krona (SEK). The currencies of emerging countries are: Chinese renminbi (CNY), Hong Kong dollar (HKD), Indian rupee (INR), South Korean won (KRW), Malaysian

[1] The data source is the Board of Governors of the Federal Reserve System and can be downloaded at http://www.federalreserve.gov/datadownload. Russian rouble is excluded in our analysis as its data is unavailable in this website.

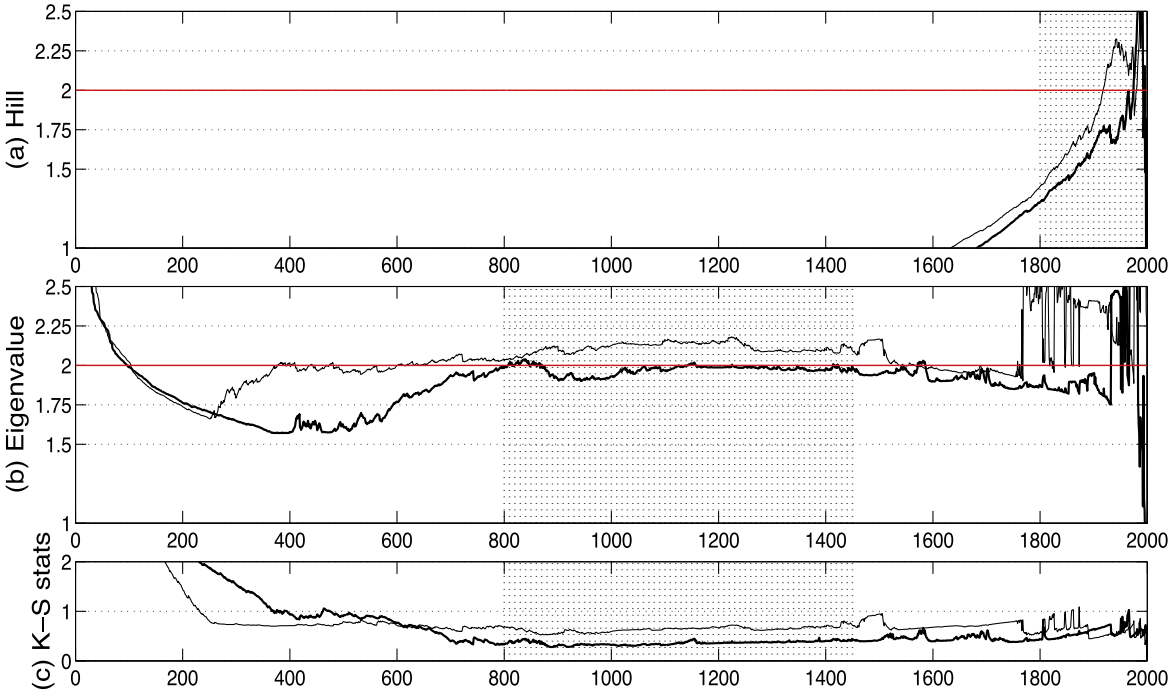

Fig. 4. Comparison of $\hat{\alpha}^{-1}$ for GPD with $\mu = 10, \sigma = 0.1, \xi = 2$.

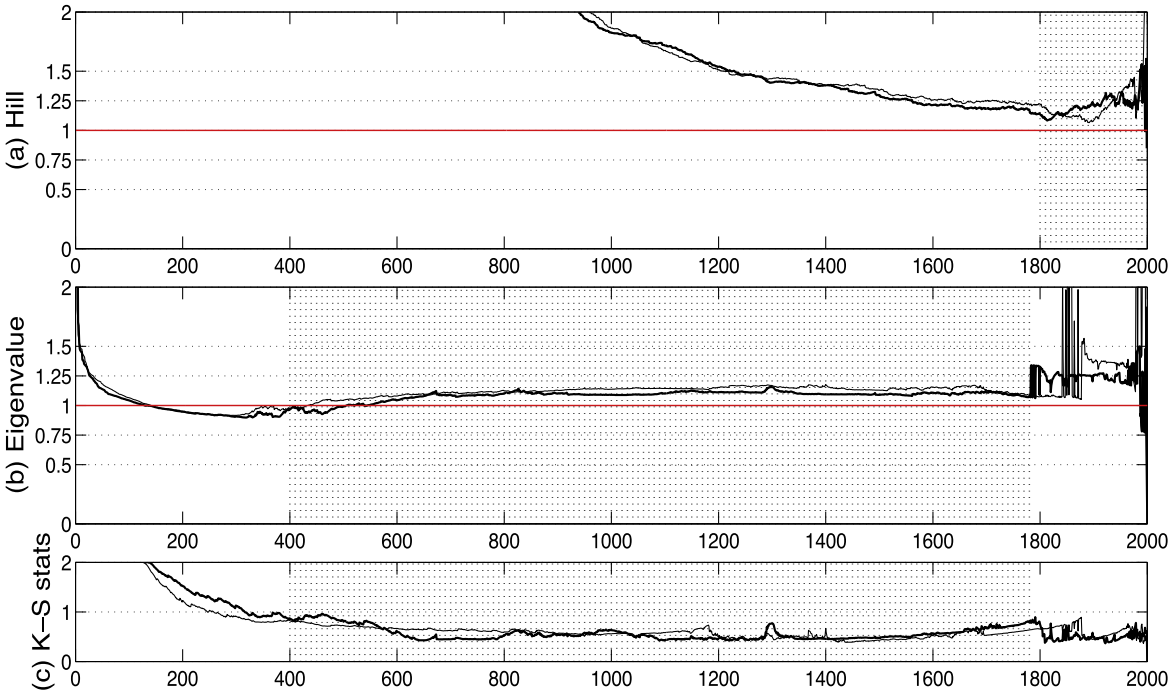

Fig. 5. Comparison of $\hat{\alpha}^{-1}$ for GEV with $\mu = 0, \sigma = 1, \xi = 1$.

ringgit (MYR), Singapore dollar (SGD), Taiwan dollar (TWD) and Thai baht (THB). The base currency is the US dollar (USD). For an easy comparison to Ibragimov et al. (2013), all the numbers in this subsection represent the tail index estimate $\hat{\alpha}$, not $\hat{\alpha}^{-1}$.

Table 1 compares the tail index estimates produced by the log–log rank size regressions (RS(10%), RS(5%)), Hill estimates (Hill(10%), Hill(5%), Hill(7.5%)) and our estimates for the developed countries (top of the table) and the emerging countries (bottom). In the table, the bold faced numbers are provided by the present authors, while all remaining numbers are from Ibragimov et al. (2013). We include Hill(7.5%) in the tables to show that Hill estimates are not monotone, in other words, they may fluctuate so much that Hill(7.5%) may not lie between Hill(5%) and Hill(10%), which is the case for GBP, KRW and THB. The total number of data points in each country is $N = 3390$ except for CNY, INR, MYR and TWD. Because CNY and MYR had been in a fixed rate regime until July 21, 2005, we also included alternative datasets, after eliminating zero returns, so that $N = 1743$ for CNY and $N = 1742$ for MYR.

Table 1
Tail index estimates for exchange rates.

Currency	RS (10%)	RS (5%)	Hill (10%)	Hill (5%)	Hill (7.5%)	Our estimates
AUD	2.80	2.71	2.85	2.88	**2.87**	**2.60 (3200, 5.61%)**
CAD	3.26	3.46	2.99	3.31	**3.05**	**3.32 (3290, 2.95%)**
CHF	3.92	3.90	3.70	4.14	**3.93**	**3.96 (3279, 3.27%)**
DKK	3.86	3.98	3.64	4.01	**3.75**	**4.02 (3243, 4.34%)**
EUR	4.23	4.59	3.77	4.28	**4.04**	**4.42 (3195, 5.75%)**
GBP	3.54	3.34	3.65	3.70	**3.94**	**3.29 (3184, 3.29%)**
JPY	3.39	3.72	3.09	3.54	**3.14**	**3.59 (3302, 2.60%)**
NOK	3.56	3.90	3.24	3.52	**3.33**	**3.74 (3352, 1.12%)**
SEK	3.64	3.98	3.22	3.67	**3.40**	**4.25 (3262, 3.78%)**
CNY	2.18	2.40	1.90	2.36		
(CNY 1743)			**2.33**	**2.54**	**2.38**	**2.29 (1656, 4.99%)**
HKD	2.25	2.57	1.91	2.30	**2.06**	**2.47 (3280, 3.24%)**
INR	2.86	3.16	2.52	2.81	**2.62**	**3.60 (3349, 1.21%)**
KRW	2.32	2.36	2.20	2.23	**2.34**	**2.28 (3334, 1.65%)**
MYR	3.25	4.08	2.56	3.51		
(MYR 1742)			**3.53**	**3.92**	**3.84**	**5.69 (1615, 4.77%)**
SGD	3.12	3.36	2.84	3.07	**2.98**	**3.36 (3178, 6.25%)**
THB	2.66	2.85	2.37	2.47	**2.54**	**2.89 (3326, 1.89%)**
TWD	2.53	2.54	2.35	2.77	**2.53**	**2.29 (3194, 5.70%)**

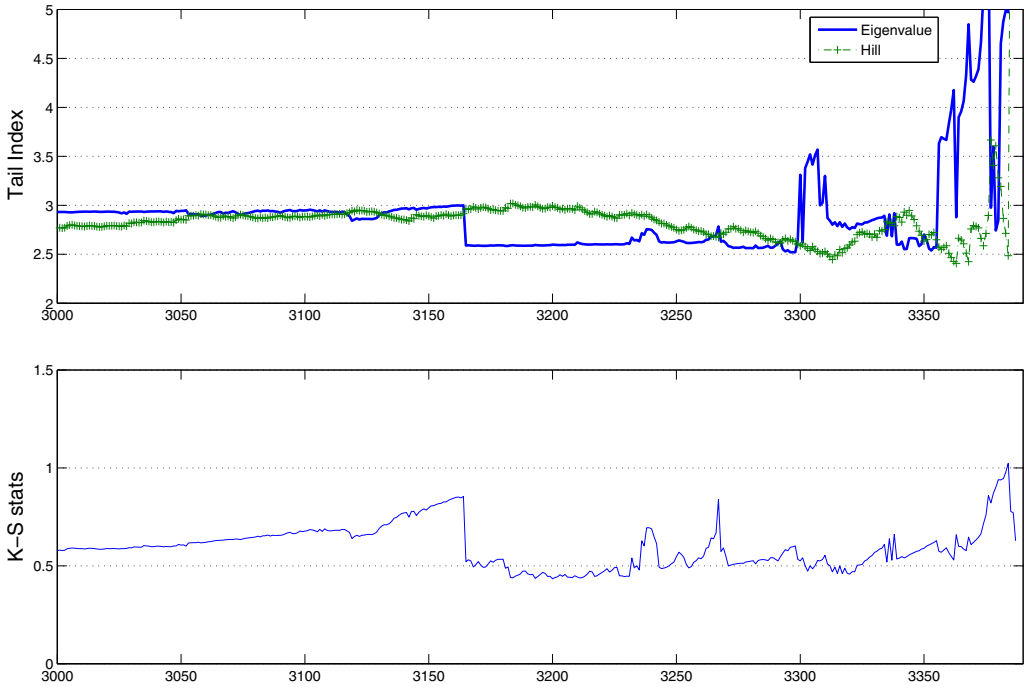

Fig. 6. Comparison of $\hat{\alpha}$ for AUD.

For INR and TWD, there are some missing points, so $N = 3389$ for INR and $N = 3387$ for TWD. For our estimates, we also provide the number of deletions and the upper truncation level, from which the tail index is identified.

For the developed countries, from the table numbers, we see that most of our estimates are within the ranges of the RS and Hill estimates, confirming that the eigenvalue and Hill plot generally agree on the tail index estimation. Two exceptions are AUD and GBP, which are considerably lower than the reported numbers in the original paper. DKK is also outside of the reported numbers, but the difference is marginal. To illustrate these two cases, we present the eigenvalue plots for AUD and GBP in Figs. 6 and 7, respectively, along with the Hill estimates. In these two figures, we see that both the eigenvalue and Hill plots fluctuate very much on the right end, but as we include more non-extreme data, we can identify a stable segment of the eigenvalue plot with relatively small K–S statistics, say less than 0.5. It is interesting to see that in Figs. 6 and 7 both the eigenvalue and K–S statistic plots show sharp jumps in the middle range at the identical thresholds, which seems to indicate a qualitative change in tail thickness at these exact points; these points thus appear to the starting point of the tail, but such a claim warrants further research. Anyway, we would not include the stepped-up area in the tail index estimation as this area belongs to the body of the dataset, as supported by the K–S statistic plots. In summary, for AUD and GBP, the Hill and eigenvalue plots give different opinions, and we believe from the graphs that the eigenvalue plot reveals the tail index more effectively via longer stable segments with the help of the K–S statistics.

For the emerging countries, the numbers are generally smaller than those of the developed countries, indicating heavier tails due to greater risk. Again, the three different methods of the tail index estimation provide similar results for most countries. There

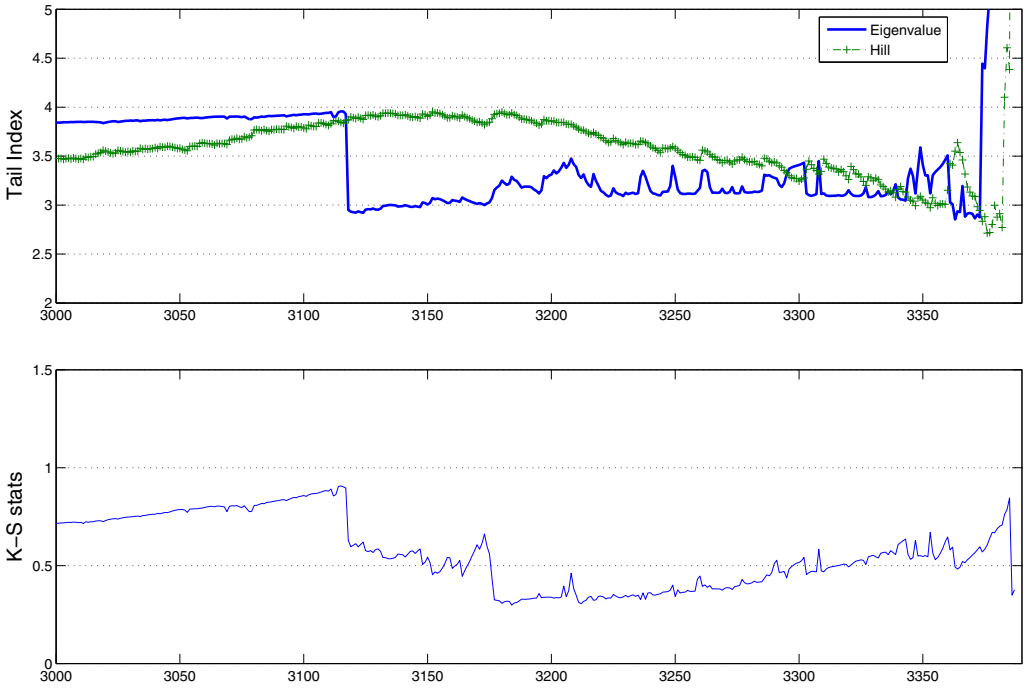

Fig. 7. Comparison of $\hat{\alpha}$ for GBP.

are a few incidences where our estimates are outside the numbers from the previous study, but we believe the differences are marginal, except for INR and MYR. Our analysis[2] reveals that the values of INR and MYR are about 3.6 and 5.69, respectively, considerably higher than any of the considered Hill estimates, but relatively close to 3.16 of RS(5%) and 4.08 of RS(5%), respectively. This implies that INR and MYR, while classified as emerging country currencies, have a tail heaviness well comparable to the developed country group, which contradicts the finding of the previous study. It may be the case that there have been active government interventions in INR and MYR currency exchange markets, respectively. For CNY and MYR, where there are two possible datasets depending on whether one includes the fixed rate regime range (where all returns are zeros) or not, we obtain considerably different Hill estimates at both 5% or 10% truncation level because the actual threshold locations change as the data size does. This implies that a blind application of the routine upper 5% or 10% truncation level, while convenient, should be avoided in reading the Hill plot; our estimates use no such routine levels.

Now we extend our analysis to investigate the impact of the financial crisis of 2008 to the tail index. This is another topic discussed in Ibragimov et al. (2013), where they found that the exchange rates of the developed countries appear to have become more pronouncedly heavy-tailed since the beginning of the crisis, while the changes in the tail index for the emerging countries exchange rates are mixed. As we now have more post-crisis data, it would be interesting to re-examine their finding using our approach. In our dataset we have added more recent data so that the post-crisis period ranges from September 15, 2008 to October 31, 2014, with $N = 1539$. Table 2 presents the results. We again observe that most of our estimates agree to the previous study based on the Hill estimates, for both before and after the crisis periods, with somewhat larger values identified by our approach for the before-crisis period. However, we find noticeable differences

Table 2
Tail index change in the exchange rates due to 2008 crisis.

Currency	Jan. 4, 1999–Sep. 15, 2008	Sep. 15, 2008–Oct. 31, 2014
AUD	3.85	2.40
CAD	5.67	3.21
CHF	6.36	3.73
DKK	4.54	3.83
EUR	5.66	3.95
GBP	6.12	2.89
JPY	3.73	3.30
NOK	5.53	3.53
SEK	7.16	3.54
CNY	1.52	2.68
HKD	2.34	2.59
INR	2.59	4.40
KRW	4.49	2.21
MYR	9.22	4.79
SGD	3.71	3.36
THB	2.81	3.86
TWD	2.99	2.24

in JPY for the after-crisis period and MYR for the before-crisis period. The previous study reports that the tail index of JPY had changed from 3.80 to 2.90 after the crisis, but by extending the post-crisis period we see that the tail index of JPY returns to 3.3, fairly close to its pre-crisis level, a phenomenon we do not see in other countries. For MYR, its moments exist up to the order 9 before the crisis and almost 5 after the crisis, indicating considerably less risk compared to other countries in both periods. This is because the Malaysian government had pegged MYR to USD for nearly seven years starting from 1998, and later switched in 2005 to a managed floating exchange rate regime, which prevents excessive market fluctuation by imposing barriers. For other countries, our analysis shows that, similar to Ibragimov et al. (2013), the changes in the tail index of the exchange rates before and after the crisis are mixed in that, after the crisis, some currencies have become more heavy-tailed, while the others have became less heavy-tailed. Compared to the previous result, we have a perfect agreement on the direction of the tail index change for all the

[2] Available upon request, however it is essentially the same as the figures for AUD and GBP.

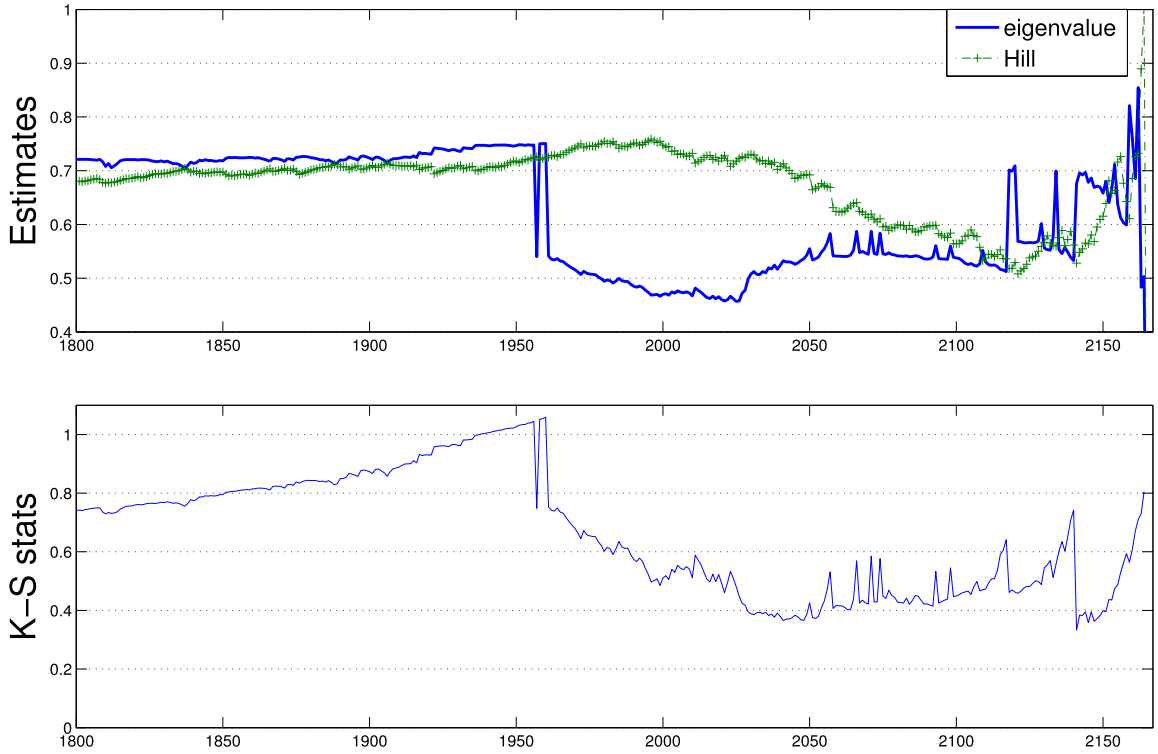

Fig. 8. Comparison of $\hat{\alpha}^{-1}$ for Danish fire data.

countries except MYR. This exception might be due to the blind truncation as explained.

4.3. Danish fire data

The Danish fire data[3] is an insurance claim size data widely studied in the EVT literature, especially regarding the tail index; McNeil (1997), Resnick (2005); and Embrechts et al. (1997). The data, collected at Copenhagen Reinsurance, comprise 2167 fire losses between 1980 and 1990, both years inclusive. The numbers have been adjusted for inflation to reflect 1985 values and are expressed in millions of Danish Krone. The minimum of the data is 1, which has obviously been left truncated, so no adjustment is needed in fitting the scaled LogPH distribution. In Fig. 8 we present the comparison of the Hill and eigenvalue plots for this dataset.

Previously two choices for the tail threshold of the Danish fire dataset have been suggested. According to Section 6.4 and 6.5 of Embrechts et al. (1997), the first threshold is at a loss of $u = 10$ with 109 exceedances, in which case the Hill estimate has a value $\hat{\alpha}^{-1} = 0.618$, indicating that only the first moment exists. For the second choice, the threshold is set at $u = 18$ with 47 exceedances, the Hill estimate is $\hat{\alpha}^{-1} = 0.497$, suggesting the existence of the first two moments. If the GPD is fitted, the corresponding maximum likelihood estimates are $\hat{\alpha}^{-1} = \hat{\xi} = 0.497$ at $u = 10$ and $\hat{\xi} = 0.735$ at $u = 18$. So essentially $\hat{\alpha}^{-1}$ is indicated to be in $[0.5, 0.74]$.

Our analysis from Fig. 8 gives a similar result. From the K–S plot, we observe too local minimums with the first occurring at around 2040 deletions and the second at 2145, where the K–S values are less than 0.4. If we read the tail estimates for these two areas, the eigenvalue plot gives two candidates, 0.53 and 0.68, respectively, which are fairly close to the two numbers suggested in the previous research. The latter value is also close to what Ahn et al. (2012) obtains, 0.69, from fitting a 2-phased PH distribution, but the fit was for the whole dataset rather than the tail part, so its credibility may be limited. While it is hard to assert which one of these two is the true value, we favor 0.53 over 0.68, because, based on our simulation studies in Section 4.1, the tail index tends to be identified with not-so-extreme deletions. In particular, considering that the original loss dataset before truncation, which is unavailable, would have much more data points with all the losses below 1 million Krones included, the remaining points after deleting 2145 would represent a small enough tail proportion (about 0.25% if the complete dataset consists of 10,000 losses) to have extreme volatility that is random. At any rate, if we form a range for the estimate, the range produced by the eigenvalue plot $[0.53, 0.68]$ is consistent with that of the previous studies.

[3] Available in Dr. A. McNeil's webpage www.macs.hw.ac.uk/~mcneil/.

5. Concluding remarks

The Hill plot, despite its wide-spread usage, has been found difficult to use in practice due to its instability and bias in the upper extreme tail. In this article we propose a new alternative to the Hill plot based on the parametric scaled LogPH distribution class with a minimal phase size to identify the tail index of heavy-tailed loss distributions. The scaled LogPH distribution class is shown to be well aligned with the standard EVT methodology in terms of the tail behavior and the mean excess function. By fitting the excess loss of the log-transformed data to the ordinary PH distribution over different thresholds, we construct the eigenvalue plot which generalizes the Hill plot. Through various simulations and the real datasets from finance and insurance, we illustrate that the eigenvalue plot, along with the Kolmogorov–Smirnov statistic as a supplementary guideline, is advantageous over the Hill plot in determining the tail index more accurately.

Acknowledgement

Joseph Kim is grateful for the support of the National Research Foundation of Korea (NRF-2012R1A1A1043439). Joocheol Kim is

grateful for the support of the National Research Foundation of Korea (NRF-2014S1A5A2A01011100).

Appendix I

Consider a continuous time Markov chain with a state space $\{1,\ldots,p,p+1\}$, with the state $p+1$ being an absorbing state, which has an initial probability vector $(\beta,0)$, with β being a p dimensional row vector, $\beta\mathbf{1}=1$ and also an infinitesimal generator

$$Q=\begin{pmatrix} T & t \\ \mathbf{0} & 0 \end{pmatrix},\ t=-T\mathbf{1},$$

where $\mathbf{0}$ and $\mathbf{1}$ are a p dimensional row vector of zeros, and a column vector of ones, respectively. T, of size $p\times p$, is an infinitesimal generator such that the off-diagonal elements of T are non-negative and the diagonal elements of T are strictly negative. We also assume that the real part of each eigenvalue of T is strictly negative and define $-\eta_T<0$ to be the eigenvalue of T closest to 0. These assumptions ensure that the first p states are transient and that absorption occurs almost surely. If we let X be the time until absorption in the Markov chain, then the distribution of X is called the PH-distribution with parameters (β,T), denoted by $\mathrm{PH}(\beta,T)$.

The denseness property of the PH distribution class means that it is dense in the Prohorov metric of weak convergence in the set of all probability distributions on $[0,\infty)$; see Latouche and Ramaswami (1999). The density and distribution function of X, if X is $\mathrm{PH}(\beta,T)$ distributed, are respectively

$$F_X(x)=1-\beta e^{Tx}\mathbf{1},\quad x>0, \tag{14}$$

and

$$f_X(x)=\beta e^{Tx}t,\quad x>0. \tag{15}$$

From this, one can show that the Laplace–Stieltjes transform of $\mathrm{PH}(\beta,T)$ is given by

$$\phi(s)=E[e^{-sX}]=\beta(s\mathbf{I}-T)^{-1}t \text{ for } Re(s)\geqslant 0 \tag{16}$$

where $\mathbf{I}$ is the identity matrix, and the moments

$$E[X^k]=k!\,\beta(-T^{-1})^k\mathbf{1}, \tag{17}$$

which exists for all $k\geqslant 1$.

Appendix II: EMPHT algorithm

Consider n incomplete observations $x_1,\ldots,x_n$ generated by the PH distribution $\mathrm{PH}(\beta,T)$ with a $p+1$ dimensional state space as defined in Section 3. The values are incomplete because x_i's represent only the absorption times. If we assume that we have observed the entire path of the underlying Markov Chain, the complete observation corresponding to a given x_i of the PH distribution can be written as

$$y_i=(j_0^{(i)},\ldots,j_{m_i-1}^{(i)};\ s_0^{(i)},\ldots,s_{m_i-1}^{(i)})$$

where m_i is the number of jumps until the Markov chain hits the absorbing state 0, $j_k^{(i)}$ is the k-th state visited by the Markov Chain with $j_{m_i}^{(i)}=0$, and $s_k^{(i)}$ is the corresponding sojourn time at the k-th state with $s_{m_i}=\infty$, indicating that the last jump leads to the absorbing state. Then, the sojourn times should satisfy $x_i=s_0^{(i)}+\cdots+s_{m_i-1}^{(i)}$, and the density of the i-th complete observation can be represented by

$$f(y_i|\beta,T)=[\beta]_{j_0^{(i)}}\exp\left[[T]_{j_0^{(i)}j_0^{(i)}}s_0^{(i)}\right][T]_{j_0^{(i)}j_1^{(i)}}\cdots\exp\left[[T]_{j_{m_i-1}^{(i)}j_{m_i-1}^{(i)}}s_{m_i-1}^{(i)}\right][t]_{j_{m_i-1}^{(i)}}$$

where we use the notations $[B]_{i,j}$ and $[\mathbf{b}]_i$ to denote (i,j)-th element and i-th element of a given matrix B and a vector $\mathbf{b}$ respectively. Now, if we denote the complete sample of all of n observations by

$$\mathbf{y}=(j_0^{(1)},\ldots,j_{m_1-1}^{(1)};s_0^{(1)},\ldots,s_{m_1-1}^{(1)};\ldots;j_0^{(n)},\ldots,j_{m_n-1}^{(n)};s_0^{(n)},\ldots,s_{m_n-1}^{(n)}),$$

then the density of the complete sample can be written in the form

$$f(\mathbf{y}|\beta,T)=\prod_{i=1}^{p}[\beta]_i^{B_i}\prod_{i=1}^{p}\exp\left[[T]_{i,i}Z_i\right]\prod_{i=1}^{p}\prod_{\substack{j=1\\ j\neq i}}^{p}[T]_{i,j}^{N_{i,j}}[t]_i^{N_{i,0}} \tag{18}$$

where $B_i, i=1,\ldots,p$ denote the number of Markov Chains starting in state i, $Z_i, i=1,\ldots,p$ the total time spent by Markov chains in state i, and $N_{i,j}$ the total number of jumps from state i to state j for $i\neq j, i=1,\ldots,p$ and $j=0,1,\ldots,p$. We see that the density in (18) is a member of the multi-parameter exponential family with sufficient statistics B_i's, Z_i's, and $N_{i,j}$'s. This leads the MLE based on the complete sample $\mathbf{y}$ by

$$\hat{\beta}_i=\frac{B_i}{n};\ [\hat{T}]_{i,j}=\frac{N_{i,j}}{Z_i},\ i\neq j;\ [\hat{t}]_i=\frac{N_{i,0}}{Z_i};\ [\hat{T}]_{ii}=-[\hat{t}]_i-\sum_{\substack{j=1\\ j\neq i}}[\hat{T}]_{i,j}. \tag{19}$$

The EM algorithm then calculates MLE of β and T based on the incomplete data. The EM algorithm is an iterative procedure that maximizes in each step the conditional expectation of the log-likelihood given the observed sample $x_1,\ldots,x_n$. That is, in each iteration, two separate steps called the E- and M-steps, are carried out. In the E-step one calculates the conditional expectation of the log-likelihood given the observations which are absorption times, and it is maximized in the M-step. From the density function in (18), the log-likelihood of the complete sample $\mathbf{y}$ is in the form of

$$\log f(\mathbf{y}|\beta,T)=\sum_{i=1}^{p}B_i\log([\beta]_i)+\sum_{i=1}^{p}[T]_{i,i}Z_i+\sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p}N_{i,j}\log([T]_{i,j})$$
$$+\sum_{i=1}^{p}N_{i,0}\log([t]_i), \tag{20}$$

which we observe to be linear in the sufficient statistics. Hence, in the E-step, calculating the conditional expectation of the log-likelihood given the observations reduces to calculating the conditional expectations of the sufficient statistics. The exact expressions of these quantities as well as the detailed EM procedure can be found in Asmussen et al. (1996).[4]

The EM algorithm for the PH calibration can be summarized as follows. Given an initial parameters (β_0,T_0), in E-step, the conditional expectations of the sufficient statistics given the observed sample $x_1,\ldots,x_n$ are calculated, and are plugged in the log-likelihood function in (20) as the parameters for complete sample $\mathbf{y}$. Then, in M-step, the MLEs are obtained using the equations in (19). These two steps are iterated until convergence. As pointed out by Bladt (2005), the representation of a given PH distribution is not unique, so the estimated values of the parameters could be different depending on the initial values of the parameters though it represents the same distribution function. The minimal representation of the PH distributions still remains an open problem.

References

Ahn, S., Kim, J.H.T., Ramaswami, V., 2012. A new class of models for heavy tailed distributions in finance and insurance risk. Insurance Mathematics and Economics 51, 43–52.

Asmussen, S., Nerman, O., Olsson, M., 1996. Fitting phase-type distributions via the EM algorithm. Scandinavian Journal of Statistics 23 (4), 419–441.

[4] The codes for EMPHT algorithm is available in http://home.imf.au.dk/asmus/pspapers.html.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2006. Statistics of extremes: theory and applications. John Wiley & Sons.
Bladt, M., 2005. A review on phase-type distributions and their use in risk theory. Astin Bulletin 35 (1), 145–167.
Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. SIAM Review 51 (4), 661–703.
Dekkers, A.L., Einmahl, J.H., De Haan, L., et al., 1989. A moment estimator for the index of an extreme-value distribution. Annals of Statistics 17 (4), 1833–1855.
Dickson, D., Hipp, C., 2000. Ruin problems for phase-type (2) risk processes. Scandinavian Actuarial Journal 2000 (2), 147–167.
Drees, H., De Haan, L., Resnick, S., 2000. How to make a Hill plot. Annals of Statistics, 254–274.
Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer, New York.
Ghosh, A., Jana, R., Ramaswami, V., Rowland, J., Shankaranarayanan, N., 2011. Modeling and characterization of large-scale wi-fi traffic in public hot-spots. In: INFOCOM. 2011 Proceedings IEEE. IEEE, pp. 2921–2929.
Hall, P., 1990. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. Journal of multivariate analysis 32 (2), 177–203.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Annals of Statistics 3 (5), 1163–1174.
Ibragimov, M., Ibragimov, R., Kattuman, P., 2013. Emerging markets and heavy tails. Journal of Banking and Finance.
Klugman, S., Panjer, H., Willmot, G., 2008. Loss Models, Third ed. John Wiley, New York.
Latouche, G., Ramaswami, V., 1999. Introduction to matrix analytic methods in stochastic modeling. In: ASA/SIAM Series on Statistics and Applied Probability, Philadelphia.
McNeil, A., 1997. Estimating the tails of loss severity distributions using extreme value theory. ASTIN Bulletin 27 (1), 117–137.
McNeil, A.J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management. Princeton University Press, New Jersey.
Neuts, M., 1975. Probability distributions of phase type. In: Liber Amicorum Prof. Emeritus H. Florin (Ed.). University of Louvain, Belgium, pp. 173–206.
Neuts, M., 1981. Matrix-geometric Solutions in Stochastic Models. An Algorithmic Approach. Johns Hopkins University Press, Baltimore, MD.
Pickands III, J., 1975. Statistical inference using extreme order statistics. Annals of Statistics, 119–131.
Reiss, R., Thomas, M., 2007. Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields. Birkhauser.
Resnick, S., 2005. Discussion of the danish data on large fire insurance losses. Astin Bulletin 27 (1), 139–152.
Resnick, S., Stărică, C., 1997. Smoothing the Hill estimator. Advances in Applied Probability, 271–293.