REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000–000
https://doi.org/00.00000/revstat.v00i0.000

Where does the Heaviness Start?

Authors: Zhuojing Zhang
– Department of Economics, University of Waterloo, Canada
z793zhan@uwaterloo.ca

Tao Chen ✉
– Big Data Research Lab, University of Waterloo, Department of Economics, University of Waterloo, Canada
t66chen@uwaterloo.ca

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- Datasets with a heavy-tailed histogram tend to have a large number of outliers, which provide important information. As a result, the bulk part and tail part of the dataset with this feature have different characteristics. Then, the choice of a threshold that separates these two parts is important. We propose a novel approach based on the Empirical Likelihood method to estimate this threshold. Because the transition between the bulk and tail parts cannot be fully disjointed in many cases, we allow the threshold to be a random variable instead of a fixed number. In addition, the threshold is relative to a benchmark since heaviness is a relative concept.

Keywords:

- *empirical likelihood; extreme optimal sample fraction; heavy tails; Pareto index; quantile.*

AMS Subject Classification:

- 62G32, 62G30.

✉ Corresponding author

1. INTRODUCTION

Datasets with a heavy-tailed histogram tend to have a large number of outliers that provide important information. In many fields, datasets are characterized by this feature, such as Psychology (Barabási, 2005; Malmgren et al., 2008), Economics (Rossi-Hansberg and Wright, 2007; Giesen and Sudekum, 2010), Finance (Mandelbrot, 1963; Gabaix et al., 2003; Gabaix, 2009), Statistics (Richardson, 1948; Clauset and Woodard, 2013; Cirillo and N., 2016) and Hydrology (Anderson and Meerschaert, 1998; Katz et al., 2002; Bernardara et al., 2007), to name a few. One important challenge for analyzing datasets with this feature is that the behaviours of the bulk part and tail part are different. For example, it is generally accepted that Pareto distributions are useful when describing the distributions of high incomes, which are represented in the tail parts of income datasets. However, Pareto distributions perform poorly over the whole range of incomes (Reed, 2003). Pareto distributions have infinite variance when the shape parameter is smaller than two, and the usual least square method cannot be used directly when datasets have infinite variance (Kanter and Steiger, 1974). Thus, the properties of one part of the dataset can affect the choice of method and mislead the analysis of the other part or the overall pattern, leading to a question about which part of the dataset should be dropped to study non-extreme events or be used to predict rare events. We loosely use the word "threshold" to denote the solution.

We might borrow approaches from Extreme Value Theory (EVT) to estimate the threshold. In EVT, a tail parameter k, which is the number of upper observations used to estimate the tail part of the dataset, is a possible threshold. It is possible to choose k by detecting the change of slope in the mean excess plot (Embrechts et al., 1997) or the first "stable" region of the hillplot which is based on the hill estimator (Hill, 1975 and Drees et al., 2000). However, the most critical issue of these graphical diagnostics is that the results are subjective. A glance ahead to Figure 1 will indicate why this is so. To avoid these problems, completely programmed estimators that can automatically choose the k are widely studied (Caeiro and Gomes, 2016). Hall and Welsh (1985) derived a formula by minimizing the asymptotic mean squared error (AMSE) of the Hill estimator to find the optimal k. However, it requires extra knowledge of an unknown second-order parameter. The estimation of the second-order parameters has been studied in several papers, like Gomes and Pestana (2007) and Caeiro and Gomes (2014). Bootstrap methods based on the minimization of the AMSE criterion are also developed due to the need to know the prior knowledge of the second order parameter (Hall, 1990; Draisma et al., 1999; Danielsson et al., 2001; Gomes and Oliveira, 2001; Gomes et al., 2012). However, the goal of these AMSE minimization approaches is to find an optimal k that balances the bias and variance of the tail index estimator which focuses on the tail part. They are not designed to estimate the transition region between the bulk and tail parts.

Besides these, some papers compared the empirical distribution of data above a threshold k with the fitted generalized Pareto distribution (GPD) by using some goodness-of-fit tests (Northrop and Coleman, 2014; Bader et al., 2018; Schneider et al., 2021), others minimized the distance between the L-moments of the datasets and the fitted GPD (Silva Lomba and Fraga Alves, 2020) or a standard Exponential distribution (Kiran and V., 2021). But, they aim to find a k such that the empirical distribution of data above the k certainly fits the GDP. Moreover, Papastathopoulos and Tawn (2013) introduced extensions of the GPD with

a discussion about the suitable selection of k. However, its ultimate goal is to improve the estimation of the tail index. Additionally, Behrens et al. (2004) had used Bayesian methods to analyze extreme events and discuss the uncertainty of k. Lee et al. (2015) investigated Bayesian measures of surprise to determine suitable k for extreme value models. But, their primary goal is also not to study the transition region or threshold selection. Some works in EVT are even interested in describing the dataset bypassing the threshold selection. For example, Naveau et al. (2016) proposed an extended GPD to jointly model low, moderate, and extreme observations without the need for threshold selection. de Carvalho et al. (2022) proposed a Bayesian regression model for the conditional left and right tail of a possibly heavy-tailed response excluding the requirement for threshold selection.

In EVT, k is reasonable as long as the tail part gives enough information to estimate the tail index, so the bulk part is ignored and a fixed k is acceptable. But, the threshold is a fixed number only if the bulk part and tail part can be fully separated, which is much less likely to happen. Otherwise, we cannot find a fixed number threshold, because the transition between the two parts is gradual. Although extreme value mixture models, which combine a model of the bulk distribution with an extreme value tail model, treat k as a parameter and use the information from the bulk part, the major purpose of these models is also not to study the threshold. Moreover, these models might lack of robustness of the bulk and tail fits to each other and are the computational complexity and implementation difficult (Scarrott and MacDonald, 2012).

To study the transition, we propose an approach based on the Empirical Likelihood (EL) method. In the transition region, the likelihood of a variable falling in the tail part increases with the value of the variable, whereas the likelihood of a variable falling in the bulk part increases as the value of the variable decreases. Naturally, the threshold can be viewed as a random variable K with a density function representing the likelihood of possible values of K that separate the bulk and tail parts.

EL, as a non-parametric method (Owen, 1988; Owen, 1990; Imbens (2002)) is a good candidate for analyzing datasets with a heavy-tailed histogram, since these datasets are usually not amenable to a known distribution. It has been applied in heavy-tailed distribution to construct confidence intervals for the tail index Lu and Peng (2002), the mean Peng (2004) and high quantiles Peng and Qi (2006). Additionally, Einmahl and Segers (2009) proposed an estimator for the spectral measure of an extreme-value distribution based on EL. A review of EL in extreme-value statistics is referred to Qi (2008), while a comprehensive review of EL is referred to Lazar (2021).

EL assigns weights to each observation of the sample dataset without parametric assumptions through an empirical likelihood ratio function under certain constraints. By imposing the right restrictions, we expect these weights to shed light on the transition region and indicate where the heaviness of a distribution starts. In addition, we view the heaviness as a relative concept. Many concepts only make sense when relative to a benchmark. For example, what height is considered tall? The answer to this question is very subjective without a benchmark. Also, the choice of benchmark should align with the specific characteristics and goals of the research. Therefore, our approach allows researchers to choose benchmarks based on their specific research questions. For example, in environmental science, the choice of benchmark may depend on the type of pollutant or environmental variable being studied. In finance, different benchmarks might be used to assess or compare tail risk in various

financial products or asset classes.

The rest of the paper is structured as follows. Section 2 describes the methodology proposed. Section 3 reports and analyzes the simulation results. Section 4 presents an application of our method to an empirical dataset. Section 5 concludes the study.

2. EL-BASED ALGORITHM

We model the transition between the bulk and tail parts through the following empirical likelihood ratio function,

$$\max_{w_1,\ldots,w_n} \left\{ \prod_{i=1}^{n} w_i \left| \sum_{i=1}^{n} w_i m(X_i,\theta) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right. \right\},$$

where $m(X,\theta)$ is a set of appropriately chosen restrictions that link the dataset of interest to the benchmark, $X_1,\ldots,X_n \in \mathbb{R}$ are independent random variables with distribution F, and w_i's are the imputed weights attached to X_i's. w_i can be found by the Lagrange multiplier method, which gives the weight function

$$w_i = n^{-1}\left(1+\lambda m(X_i,\theta)\right)^{-1},$$

where λ is the Lagrange multiplier. Now, we should provide the intuition behind our method: by aligning the "center" and "spread" of the benchmark and dataset of interest, the weights implied by the pre-specified constraints $m(X,\theta)$ disclose the relative rareness of realizations with similar values.

As variance measure the "spread" of the distribution, the variance constraint is used in this paper. This algorithm does not require F itself to have μ and/or σ because the sample mean and variance, which are well-defined, can serve as the same device. Other constraints related to the "spread" of the datasets also work. For example, we can replace the "variance restriction" with restrictions in terms of a combination of low and high quantiles, which always exist.

Next, the choice of the benchmark should be based on researchers' interests and sample datasets. As an exposition, we pick the benchmark to be an Exponential random variable and we only focus on the right tail. To align the benchmark and dataset of interest, we let the median of the benchmark sample equals to the median of the dataset. Then, the threshold is assumed to be bigger than the median of the dataset.

Once we have the benchmark sample, we compare the weights of dataset and benchmark. To make them comparable, we focus the sign of λ in the weight function of the dataset equals to the sign of λ in the weight function of the benchmark sample. Our method can be summarized as follows:

a) From a random sample of F, we obtain its estimated mean ($\hat{\mu}$) and variance ($\hat{\sigma}^2$).

b) Under the "variance restriction", the empirical likelihood ratio function becomes

$$\max_{w_1,\ldots,w_n} \left\{ \prod_{i=1}^{n} w_i \left| \sum_{i=1}^{n} w_i \left(X_i-\hat{\mu}\right)^2 = \hat{\sigma}^2, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right. \right\},$$

and the weight function for X_i's are

$$w_i = n^{-1}\left(1+\lambda\left[(X_i-\hat{\mu})^2-\hat{\sigma}^2\right]\right)^{-1}, \tag{2.1}$$

where the Lagrange multiplier λ can be found by numerical search. Denote the weights of X_i's by $\left\{w^{X_i}\right\}_{i=1}^n$.

c) Simulate a random sample $Y_1,\ldots,Y_n$ from an Exponential distribution with mean being $\hat{Q}/\ln 2$, where $\hat{Q}$ is the sample median of X_i's.

d) Calculate weights of Y_i's by substituting Y_i's, its true mean, $\hat{Q}/\ln 2$ and true variance, $(\hat{Q}/\ln 2)^2$, into the left side of equation (2.1), denoted as $\left\{w^{Y_i}\right\}_{i=1}^n$.

e) Sort $\left\{w^{X_i}\right\}_{i=1}^n$ and $\left\{w^{Y_i}\right\}_{i=1}^n$ in ascending order, respectively. Denote the sorted weights as $\left\{w^X_{(i)}\right\}_{i=1}^n$ and $\left\{w^Y_{(i)}\right\}_{i=1}^n$.

f) Find the set of crossing points I^c between $\left\{w^X_{(i)}\right\}_{i=1}^n$ and $\left\{w^Y_{(i)}\right\}_{i=1}^n$ by collecting the index numbers j such that the sign of $w^X_{(j)}-w^Y_{(j)}$ differs from the sign of $w^X_{(j+1)}-w^Y_{(j+1)}$, for $j=1,\ldots,n-1$.

g) Select the minimum value k^c of I^c that is bigger than $n/2$.

h) Repeat (c)-(g) m times.

This procedure gives $\{k^c_1,\ldots,k^c_m\}$ that is the set of all possible values of the threshold K which models the transition region, denoted as $R(K)$. In many studies, researchers would prefer a simplified version of K, e.g., a representative number, denoted by τ, from K. There are many possible choices for τ and in the current paper, τ is defined to be the average of K. Further inference is now feasible because τ has variance attached to it.

3. SIMULATION STUDIES

We focus on the heaviness of two types of distributions: Pareto distribution and mixed distribution, which is a linear combination of Pareto and Normal distributions. A Pareto distribution is defined by its scale and shape parameters. Here, we fix the scale parameter to be 1 and will only change the value of the shape parameter.

We consider four cases, each with 10000 observations. Of the first two, one involves Pareto distribution where the mean and variance exist (shape = 6) and the other focuses on truncated Pareto distribution. The truncated Pareto distribution is a truncated version of the Pareto distribution, which is truncated at 0.99 quantile of the Pareto distribution with shape = 1.5. We denote them as Pareto(6) and Pareto(1.5), respectively. For a Pareto distribution, the tail index is the inverse of the value of the shape parameter. Thus, the tail index is around 0.1667 for Pareto(6) and around 0.667 for Pareto(1.5). The other two cases are mixed distributions: 90% of a Pareto(6) or Pareto(1.5) and 10% of a Normal random variable. Using Pareto(6) as an example, since we are interested in the right tail, we let the 40^{th} and 45^{th} percentile values of the Normal distribution equal the same percentiles of

the Pareto(6) so that the shape of the right part of the mixed distribution is not distorted. We denote this mixed distribution as Mixed(6). Similarly, Mixed(1.5) for the other mixed distribution.

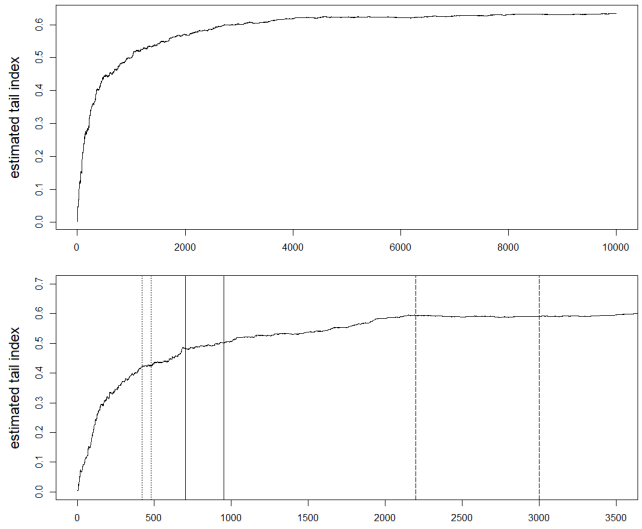

Figure 1: The top panel shows the hillplot for all k and the bottom panel limits the range of k.

Figure 1 provides an example of a Pareto(1.5) hillplot to explain why graphical methods can be subjective. Three intervals, $[420, 550]$, $[700, 950]$ and $[2200, 3000]$, are indicated in the bottom panel of Figure 1 with different types of lines. All of these three intervals can be selected as the first "stable" region under different standards. Thus, the result computed from the first "stable" region is not objective. With substantial expertise, it is possible to choose an acceptable unique solution. For example, the minimum value of the largest intervals $[2200, 3000]$ can be selected. However, it requires expert experience and is time-consuming when there are many datasets.

The results for the four cases can be found in the first four rows of Table 1. To make the value of τ easier to interpret, we present it in percentage format. For example, τ is 86.1% for Pareto(6), meaning that the heaviness starts at the top 13.9% of Pareto(6). The difference between the values of τ for Pareto and mixed distributions is very small (less than 0.31%), since the heaviness is only driven by the Pareto component. As a side result, the table also presents the estimated tail-index, $\hat{\gamma}$, by using τ as the tail parameter in Hill's tail-index estimator. Given the true value of the tail index, we calculate the mean squared error (MSE) of $\hat{\gamma}$. The last column of Table 1 shows that the MSE for all cases is fairly small. Though it is not our goal in this paper to propose another tail parameter estimator, which is one of the focal points in EVT, our approach can be viewed as an addition to that literature.

Next, we set Pareto(6) instead of an Exponential distribution as the benchmark for Pareto(1.5). In Table 1, the value of τ is 84.97% when the target is Pareto(1.5) and benchmark is an Exponential distribution, whereas the value of τ is 86.67% when the benchmark is Pareto(6). The tail of Pareto(6) is heavier than the tail of Exponential distribution, therefore, the heaviness starts at a higher quantile. Then, the $\hat{\gamma}$ for Pareto(1.5) is calculated by setting the tail parameter in Hill's tail-index estimator be 84.97% and be 86.67%, respectively.

Target distribution	Benchmark	τ	$\hat{\gamma}$	MSE
Pareto(6)	Exponential	86.10%	0.1694	0.00002
Mixed(6)	Exponential	86.24%	0.1562	0.00018
Pareto(1.5)	Exponential	84.97%	0.5477	0.01428
Mixed(1.5)	Exponential	85.28%	0.5185	0.02205
Pareto(1.5)	Pareto(6)	86.67%	0.4865	0.03258
Beta(2,5)	Exponential	81.20%	0.0603	0.00436

Table 1: Simulation results

More interestingly, when the benchmark is an Exponential distribution, the estimated shape parameter is 1.826, which implies that the distribution has an infinite variance; whereas, the estimated shape parameter is 2.055, meaning that the distribution has a finite variance when the benchmark is Pareto(6).

Note that if we force Pareto(1.5) to be the benchmark and Pareto(6) be the distribution of interest, we get 86.67% back. The "role reversal" property of our algorithm insures relative heaviness. In addition, the tail index of the target distribution and benchmark are both allowed to be non-positive. For example, we can force a Beta distribution with two shape parameters $(2, 5)$ which has a negative tail index, denoted by Beta(2,5), be the target distribution and an Exponential distribution which has a zero tail index be the benchmark. Then, it gives the threshold K of the Exponential distribution with benchmark Beta(2,5) by the "role reversal" property, since the tail of the Exponential distribution is heavy relative to a Beta distribution. The results are shown in the last row of Table 1. The value of τ is 81.20% and $\hat{\gamma}$ is 0.0603 which is close to 0. Note that the $\hat{\gamma}$ in this case is the estimated tail-index of the Exponential distribution and is calculated by using the Moment tail-index estimator proposed by Dekkers et al. (1989), since the Hill estimator only works for $\gamma > 0$.

Target distribution	Benchmark	τ	$\hat{\gamma}$	MSE
Pareto(6)	Exponential	66.68%	0.1678	0.00001
Mixed(6)	Exponential	67.61%	0.1529	0.00018
Pareto(1.5)	Exponential	72.04%	0.5899	0.26473
Mixed(1.5)	Exponential	71.18%	0.5496	0.26474
Pareto(1.5)	Pareto(6)	79.39 %	0.5711	0.00926
Beta(2,5)	Exponential	71.05%	0.0822	0.00724

Table 2: Simulation results for the multiple restriction

The choice of restriction of the EL ratio function is also based on the specific research questions and will affect the result. We add a median restriction to the empirical likelihood ratio function and the results for all cases by using the median and variance restriction can be found in Table 2. Again, the difference between the values of τ for Pareto and mixed distributions is very small. The values of τ by using the median and variance restriction are all smaller than the values of τ by using only variance restriction, since a median restriction is not sensitive to the presence of outliers.

Although τ and k are two different concepts, we still present the value of k in Table 3. Two approaches, denoted as DAMSE and GC, presented in Caeiro and Gomes (2016) are

considered. DAMSE selects k by minimizing the AMSE of Hill estimator, that is,

$$k := \arg\min_k AMSE(\hat{\gamma}(k)).$$

Then, DAMSE suggests the estimator

$$\hat{k} := \left\lfloor \left(\frac{(1-\hat{\rho})^2 n^{-2\hat{\rho}}}{-2\hat{\rho}\hat{\beta}^2} \right)^{1/(1-2\hat{\rho})} \right\rfloor,$$

where $\lfloor x \rfloor$ denote the integer part of x, $\hat{\beta}$ and $\hat{\rho}$ are the estimated second-order parameters. The ρ–estimators and β–estimators refer to Fraga Alves et al. (2003) and Gomes and Martins (2002), respectively. GC selects k by minimizing AMSE of an auxiliary statistic, which is written as

$$T_{k,n} := \widehat{\gamma}([k/2]) - \widehat{\gamma}(k), \quad k = 2, ..., n-1.$$

The observed values of $T_{k,n}$ are obtained by the double bootstrap procedure. More detail refers to Gomes et al. (2012). From Table 3, we find that DAMSE and GC perform best for Pareto(6) which is a standard Pareto distribution with finite variance, because DAMSE and GC depend on the AMSE minimization of the Hill estimator.

Target distribution	Method	$\hat{k}$	$\hat{\gamma}$
Pareto(6)	DAMSE	81.32%	0.1660
Mixed(6)	DAMSE	50.84%	0.1525
Pareto(1.5)	DAMSE	98.38%	0.2756
Mixed(1.5)	DAMSE	96.92%	0.3634
Pareto(6)	GC	90.26%	0.1659
Mixed(6)	GC	25.30%	0.1528
Pareto(1.5)	GC	94.85%	0.4343
Mixed(1.5)	GC	78.07%	0.5557

Table 3: Threshold selection by approaches from EVT

4. AN EMPIRICAL EXAMPLE

We apply our approach to a U.S. household asset value dataset, which draws from the 2018 Survey of Income and Program Participation (SIPP) and contains 743,753 observations. The SIPP, sponsored by the U.S. Census Bureau, collects information of households' economic status, such as assets and liabilities. Table 4 presents descriptive statistics of the household-level total asset values. The top panel of Figure 2 shows a histogram of the complete dataset and the bottom panel presents the values up to 6 million to make the histogram's trends easier to see.

Minimum	Mean	Maximum	Skewness	Kurtosis
1	690,824	142,631,700	21	897

Table 4: Summary Statistics of SIPP Total Asset Values

From the table and histograms, we see that total asset values tend to have a large number of outliers and its kurtosis is fairly large; that is, the dataset is heavy-tailed. Figure 3

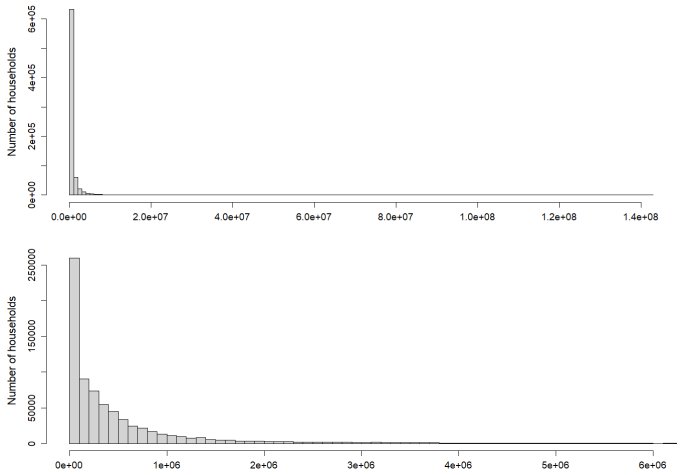

Figure 2: Histograms of household-level total asset values

shows the log-log plot of household-level total asset values. The horizontal axis is the log value of household-level total asset values and the vertical axis is the log value of its survival probability. The tail part of the log-log plot is almost a straight line, which suggests that this portion of the dataset is a member of the Pareto distribution family.

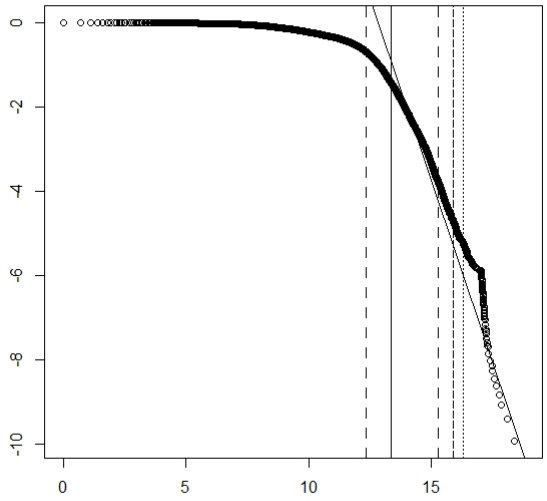

Figure 3: The log-log plot of household-level total asset values with K, τ, k_{dotted} and k_{dashed}.

Min.	1st Qu.	τ	Median	3rd Qu.	Max.
50.04%	75.65%	75.66%	76.28%	76.57%	97.65%

Table 5: Summary Statistics of K

Again, we set Exponential distribution to be the benchmark and align the median. The summary statistics of K are reported in Table 5 with $m = 1000$ and Figure 4 is the corresponding hillplot. Two intervals, $[2700, 4200]$ and $[5500, 7000]$, are chosen as the "first"

possible stable region, where one is between dotted lines, and the other is between two dashed lines. The two intervals are $[99.44\%, 99.64\%]$ and $[99.06\%, 99.26\%]$ in terms of percentiles. Any numbers in those two intervals are admissible choices of the threshold for EVT method. If we denote the middle points by $k_{\text{dotted}} = 99.54\%$ and $k_{\text{dashed}} = 99.16\%$ and together with $\tau = 75.56\%$ and $R(K) = [50.04\%, 97.65\%]$, five vertical lines (dotted, dashed, solid and two longdash) are added to the aforementioned log-log plot, as shown in the Figure 3. In addition, the k is 99.50% by using DAMSE and is 97.52% by using GC.

An encouraging finding is that the solid line and longdash lines, which represent τ and K are close to the turning point, beyond which Pareto distribution becomes a good approximation.

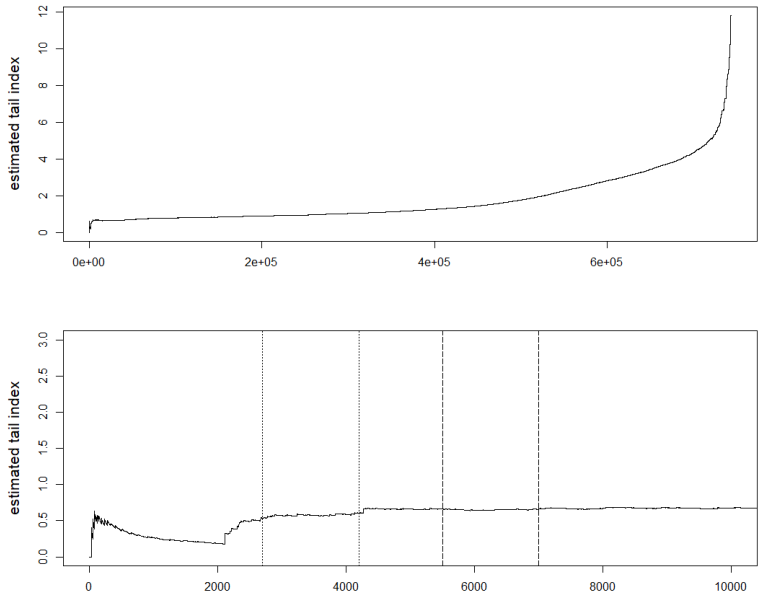

Figure 4: The top panel shows the hillplot of the complete SIPP total assets and the bottom panel limits its range.

5. CONCLUSION

We propose a novel approach based on EL to characterize the transition between the bulk and tail parts of a dataset. There are many desirable features attached to this method. It remains to show the convergence rate and limiting distribution of τ (or some other important statistics of K). We defer those topics to future research projects.

COMPETING INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

The authors contributed equally.

DATA AVAILABILITY STATEMENT

The empirical data underlying this article are publicly available on the Census Bureau, at https://www.census.gov/programs-surveys/sipp/data/datasets/2018-data/2018.html. The computer code used to generate the simulated datasets and the results are available upon request from the corresponding author.

REFERENCES

Anderson, P. L. and Meerschaert, M. M. (1998). Modeling river flows with heavy tails. *Water Resources Research*, 34(9):2271–2280.

Bader, B., Yan, J., and Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics*, 12(1):310–329.

Barabási, A. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211.

Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3):227–244.

Bernardara, P., Schertzer, D., Sauquet, E., Tchiguirinskaia, I., and Lang, M. (2007). The flood probability distribution tail: how heavy is it? *Stochastic Environmental Research and Risk Assessment*, 22(1):107–122.

Caeiro, F. and Gomes, M. I. (2014). A semi-parametric estimator of a shape second-order parameter. In Pacheco, A., Santos, R., Oliveira, M. d. R., and Paulino, C. D., editors, *New Advances in Statistical Modeling and Applications*, pages 137–144. Springer.

Caeiro, F. and Gomes, M. I. (2016). Threshold selection in extreme value analysis. In Dey, D. and Yan, J., editors, *Extreme Value Modeling and Risk Analysis: Methods and Applications*, pages 69–86. Chapman and Hall/CRC Press.

Cirillo, P. and N., T. (2016). On the statistical properties and tail risk of violent conflicts. *Physica A*, 452:29–45.

Clauset, A. and Woodard, R. (2013). Estimating the historical and future probabilities of large terrorist events. *The Annals of Applied Statistics*, 7(4):1838–1865.

Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76(2):226–248.

de Carvalho, M., Pereira, S., Pereira, P., and de Zea Bermudez, P. (2022). An extreme value bayesian lasso for the conditional left and right tails. *Journal of Agricultural, Biological, and Environmental Statistics*, 27(2):222–239.

Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17(4):1833–1855.

Draisma, G., de Haan, L., Peng, L., and Pereira, T. T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes*, 2(4):367–404.

Drees, H., de Haan, L., and Resnick, S. (2000). How to make a hill plot. *The Annals of Statistics*, 28(1):254–274.

Einmahl, J. H. J. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics*, 37(5B):2953–2989.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*. Springer, Berlin.

Fraga Alves, M. I., Gomes, M. I., and de Haan, L. (2003). A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60(2):193–214.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, 1(1):255–294.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267–270.

Giesen, K. and Sudekum, J. (2010). Zipf's law for cities in the regions and the country. *Journal of Economic Geography*, 11(4):667–686.

Gomes, M. I., Figueiredo, F., and Neves, M. M. (2012). Adaptive estimation of heavy right tails: resampling-based methods in action. *Extremes*, 15(4):463–489.

Gomes, M. I. and Martins, M. J. (2002). Asymptotically unbiased estimators of the tail index based on external estimation of the second. *Extremes*, 5(1):5–31.

Gomes, M. I. and Oliveira, O. (2001). The bootstrap methodology in statistics of extremes – choice of the optimal sample fraction. *Extremes*, 4(4):331–358.

Gomes, M. I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (var) estimator. *Journal of the American Statistical Association*, 102(477):280–292.

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.

Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1):331–341.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174.

Imbens, G. W. (2002). Generalized method of moments and empirical likelihood. *Journal of Business & Economic Statistics*, 20(4):493–506.

Kanter, M. and Steiger, W. L. (1974). Regression and autoregression with infinite variance. *Advances in applied probability*, 6(4):768–783.

Katz, R. W., B., P. M., and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8):1287–1304.

Kiran, K. G. and V., S. V. (2021). A mahalanobis distance-based automatic threshold selection method for peaks over threshold model. *Water Resources Research*, 57(1).

Lazar, N. A. (2021). A review of empirical likelihood. *Annual Review of Statistics and its Application*, 8(1):329–344.

Lee, J., Fan, Y., and Sisson, S. A. (2015). Bayesian threshold selection for extremal models using measures of surprise. *Computational Statistics & Data Analysis*, 85(1):84–99.

Lu, J. and Peng, L. (2002). Likelihood based confidence intervals for the tail index. *Extremes*, 5:337–352.

Malmgren, R. D., Stouffer, D. B., Motter, A. E., and Amaral, L. A. N. (2008). Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences - PNAS*, 105(47):18153–18158.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769.

Northrop, P. J. and Coleman, C. L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17(2):289–303.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Papastathopoulos, J. and Tawn, J. A. (2013). Extended generalised pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143.

Peng, L. (2004). Empirical-likelihood-based confidence interval for the mean with a heavy-tailed distribution. *The Annals of Statistics*, 32(3):1192–1214.

Peng, L. and Qi, Y. (2006). Confidence regions for high quantiles of a heavy tailed distribution. *The Annals of Statistics*, 34(4):1964–1986.

Qi, Y. (2008). Bootstrap and empirical likelihood methods in extremes. *Extremes*, 11(1):81–97.

Reed, W. J. (2003). The pareto law of incomes – an explanation and an extension. *Physica A*, 319:469–486.

Richardson, L. F. (1948). Variation of the frequency of fatal quarrels with magnitude. *Journal of the American Statistical Association*, 43(244):523–546.

Rossi-Hansberg, E. and Wright, M. L. J. (2007). Urban structure and growth. *The Review of Economic Studies*, 74(2):597–624.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *Revstat-Statistical Journal*, 10(1):33–60.

Schneider, L., Krajina, A., and Krivobokova, T. (2021). Threshold selection in uni-variate extreme value analysis. *Extremes*, 24:881–913.

Silva Lomba, J. and Fraga Alves, M. (2020). L-moments for automatic threshold selection in extreme value analysis. *Stochastic Environmental Research and Risk Assessment*, 34(3):465–491.